

Estimation of $1/f$ Noise

Brett Ninness, *Member, IEEE*

Abstract—Several models have emerged for describing $1/f^\gamma$ noise processes. Based on these, various techniques for estimating the properties of such processes have been developed. This paper provides theoretical analysis of a new wavelet-based approach which has the advantages of having low computational complexity and being able to handle the case where the $1/f^\gamma$ noise might be embedded in a further white-noise process. However, the analysis conducted here shows that these advantages are balanced by the fact that the wavelet-based scheme is only consistent for spectral exponents γ in the range $\gamma \in (0, 1)$. This is in contradiction to the results suggested in previous empirical studies. When $\gamma \in (0, 1)$ this paper also establishes that wavelet-based maximum-likelihood methods are asymptotically Gaussian and efficient. Finally, the asymptotic rate of mean-square convergence of the parameter estimates is established and is shown to slow as γ approaches one. Combined with a survey of non-wavelet-based methods, these new results give a perspective on the various tradeoffs to be considered when modeling and estimating $1/f^\gamma$ noise processes.

Index Terms—Flicker noise, fractional Brownian motion, Hurst exponent, maximum-likelihood estimation, $1/f$ noise, wavelet analysis.

I. INTRODUCTION

RECENTLY, there has been significant interest in “long memory” [19] and “multiscale” stochastic processes [3], [4], [27] and their overlap with work on fractals and wavelet analysis [14], particularly through the study of fractional Brownian motion (fBm) and fractional Gaussian noise (fGn) [6], [12], [15], [16], [32], [37], and [40].

A large part of the impetus for such work has been the problem of dealing with so-called $1/f$ stochastic processes which have become of growing importance to physicists and the signal processing community [2], [24], [28], [36], and [41], and more recently, to the control theory community [27].

To be more specific, “ $1/f$ ” noise is the colloquial term given to a stochastic process $\{x_k\}$ whose sample spectral density, or periodogram,¹ $|\hat{x}_N(\omega)|^2$ is of the form $\mathbf{E}\{|\hat{x}_N(\omega)|^2\} \approx \sigma_x^2 |\omega|^{-\gamma}$ for some finite nonzero σ_x and γ and where $\mathbf{E}\{\cdot\}$ denotes expectation. Such processes (which are sometimes also called “flicker noise”) have been empirically observed in a wide variety of physical processes such as [24], [30] currents in semiconductors, oscillation of quartz crystals, geophysical records, rate of insulin uptake, economic data, traffic-flow rates, image texture [28], and heart-rate variability [20].

Manuscript received April 1, 1996; revised June 10, 1997. This work was supported by the Australian Research Council and the Centre for Industrial Control Science, University of Newcastle, Callaghan 2308, Australia.

The author is with the Department of Electrical and Computer Engineering, University of Newcastle, Australia.

Publisher Item Identifier S 0018-9448(98)00002-9.

¹The squared absolute value of the N -point DFT $\hat{x}_N(\omega)$ of $\{x_k\}$.

In these areas, for the purposes of prediction, control, or diagnosis, it is of great interest to be able to estimate the spectral exponent γ from an observed sample path. Many methods to achieve this have been proposed. They range from least squares estimation of the slope of log-axes plots of sample periodograms [18], [26] through to approximate and exact maximum-likelihood (ML) estimation [8], [17], [26], [28] and direct measuring of fractal dimension of observed sample paths [13], [15], [18]. Aside from the ML-based schemes, these methods assume that the sample path observation $\{x_k\}$ is not corrupted by any other noise sources. Various convergence results (which will be briefly surveyed) are then available.

Unfortunately, the ML methods, although able to cope with measurement noise, are computationally intensive to implement. An exception is the work of Wornell and Oppenheim [45], [46] where the “whitening filter” property [12], [40], [44] of the wavelet transform on a $1/f$ processes has been exploited to derive a computationally efficient ML estimation scheme.² Wornell and Oppenheim study the properties of their method empirically via a computer simulation study and conclude that it appears to be consistent and asymptotically efficient for a wide range of spectral exponents $\gamma > 0$.

In contrast, this paper provides a theoretical analysis of the wavelet-based estimation scheme of [45] and [46], and finds that in the presence of measurement noise it is only consistent for a restricted range of γ . Namely, $\gamma \in (0, 1)$. For $\gamma > 1$ this paper shows that in fact the estimator converges with probability one to an incorrect estimate; this conclusion is illustrated by a simulation study. The paper also provides a theoretical analysis for the distributional properties of the estimate, and finds that again the asymptotic efficiency and Gaussianity results suggested empirically in [45] and [46] hold only for $\gamma \in (0, 1)$. Finally, when $\gamma \in (0, 1)$ then the strong consistency and distributional results can be combined to also establish mean-square consistency in such a way as to show how the convergence rate depends on γ ; it slows to zero as γ increases towards one. This effect of slower convergence for larger γ is in accordance with various simulation studies in the literature [18], [28], [45].

Since the wavelet transform is a linear function of the data, it forms a sufficient statistic, so that these results give some indication of the intrinsic difficulty of using a wavelet-based model to estimate γ from noise corrupted observations by any means, ML or not. Combined with the comments of Masry [32], this indicates that some circumspection is required in the

²Throughout this paper the term “ $1/f$ process” will be used in its generic sense to mean a process with spectrum like $1/f^\gamma$ with any γ not necessarily equal to one. When specific ranges of γ are to be discussed, the term “ $1/f^\gamma$ process” will be used.

choice of model for analysis of $1/f$ processes. Tradeoffs exist between computational complexity, theoretical justifiability [32], [46] and, as this paper shows, estimation accuracy.

II. MODELING OF $1/f$ PROCESSES

There are a number of technical difficulties in the modeling of random processes with $1/f^\gamma$ -type spectra, the main difficulty being that for $\gamma > 1$ (so that the process is a so-called “long-memory” process [19], although other authors reserve this for the case $\gamma > 2$) the spectrum is nonintegrable so that no stationary process can be associated with the spectrum.

The first attempts at solving this conundrum [1] involved using the idea of fractional integrals [33] to extend the usual definition of the Wiener process in terms of a stochastic integral of uncorrelated Gaussian increments.

To be more specific, for the case of $\gamma = 0$ the attendant difficulties of defining a process associated with a constant spectrum are traditionally handled by defining the classical nonstationary Brownian motion process $B(t)$ and then considering its increments dB which lead to the required spectra. That is,

$$B(t) = \sigma_B^2 \int_0^t dB(\sigma)$$

where the increments are a stationary process with a white spectrum. Therefore, the (formal) derivative of $B(t)$ can be considered to have a spectral representation

$$\frac{dB}{dt} = \int_{-\infty}^{\infty} e^{j\omega t} d\mu(\omega)$$

where the measure $d\mu$ satisfies

$$\mathbb{E}\{|d\mu(\omega)|^2\} = \sigma_B^2 d\omega^2, \quad \sigma_B^2 < \infty$$

in which case $B(t)$, since it is the integral of $\dot{B}(t)$, should have a spectrum like $\sigma_B^2 \omega^{-2}$. Following this heuristic line of reasoning, integrating again

$$B_2(t) = \sigma_B^2 \int_0^t \int_0^\xi dB(\sigma) d\xi$$

gives a process $B_2(t)$, that being the double integral of \dot{B} , should have a spectrum like $\sigma_B^2 \omega^{-4}$, and so on, so that using Liouville’s formula [22]

$$\begin{aligned} B_n(t) &= \sigma_B^2 \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} dB(\sigma) dt_2 \cdots dt_n \\ &= \frac{\sigma_B^2}{\Gamma(n)} \int_0^t (t - \xi)^{n-1} dB(\xi) \end{aligned} \quad (1)$$

should have a spectrum like $\sigma_B^2 \omega^{-2(n+1)}$.

The contribution in [1] was to note that if the interest is in a process with nonintegral spectral exponent γ , then this can be modeled by using a so-called “fractional integral” of $dB(t)$ in which the exponent n in (1) is nonintegral.

Unfortunately, this does not lead to a process with stationary increments (so that the spectral density interpretation is

difficult). This can be rectified, as was shown by Mandelbrot and Van Ness [31], by modifying (1) slightly to become

$$\begin{aligned} B_H(t) &= \frac{\sigma_B^2}{\Gamma(H + 1/2)} \left\{ \int_0^t |t - \xi|^{H-1/2} dB(\xi) \right. \\ &\quad \left. + \int_{-\infty}^0 (|t - \xi|^{H-1/2} - |\xi|^{H-1/2}) dB(\xi) \right\}. \end{aligned} \quad (2)$$

This latter equation with $H \in (0, 1)$ is known as “fractional Brownian motion” and is by now the pre-eminent model for processes with $1/f^\gamma$ spectra. The parameter H is known as the “Hurst exponent,” in recognition of early work in the area [21]. When $H = 1/2$, ordinary Brownian motion results. When $H = 0$ or 1 , then $B_H(t)$ degenerates into a process that is (respectively) either zero, or a straight line through the origin.

The stationary increments $B_H(t + \tau) - B_H(t)$ of fBm are termed fractional Gaussian noise (fGn) and are zero-mean with variance proportional to τ^{2H-2} so that using the Fourier transform pair

$$\frac{|\tau|^{\gamma-1}}{2\Gamma(\gamma) \cos(\gamma\pi/2)} \leftrightarrow \frac{1}{|\omega|^\gamma}$$

then indeed fGn serves as a model for Gaussian processes with $1/f^\gamma$ spectra with $\gamma \in (0, 1)$.

There are many other interesting properties of fractional Brownian motions. For example, their sample paths are self-similar in the sense that [16]

$$B_H(\lambda t) \stackrel{\mathcal{D}}{=} \lambda^H B_H(t)$$

where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. This makes them a particular example of the multiscale stochastic processes studied recently in [3], [4], and [27]. In fact, fBm’s are the only Gaussian process to display this self-similar property [31], and as might be expected from this property, their sample paths are fractal (with Hausdorff–Besicovitch and box dimension $D = 2 - H$) although they are also mean-square-continuous and continuous on compact sets with probability one.

As well, fractional Brownian motion is a nonstationary process with covariance function

$$R(t_1, t_2) = \frac{\Gamma(2 - 2H) \cos \pi H}{2\pi H(1 - 2H)} \{|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}\}$$

so that for $H > 1/2$ then

$$\mathbb{E}\{B_H(t)[B_H(t + \tau) - B_H(t)]\} > 0$$

and *vice versa* for $H < 1/2$. The implication is that sample paths tend to be increasing or decreasing depending on H [13]. This nonstationarity of fractional Brownian motion means that even though its derivation was predicated on trying to find a useful definition of a random process with $1/f^\gamma$ type spectrum for $\gamma > 1$, this interpretation is still difficult.

Various suggestions have been put forward to circumvent this difficulty. Some involve the use of time-localized transformations such as the Wigner–Ville [15] and wavelet [15], [32] transforms, others make formal the heuristic motivation for fractional integrals outlined above [31]. All approaches arrive at the relationship between the spectral exponent γ and the Hurst parameter H as $\gamma = 2H + 1$. A particularly appealing

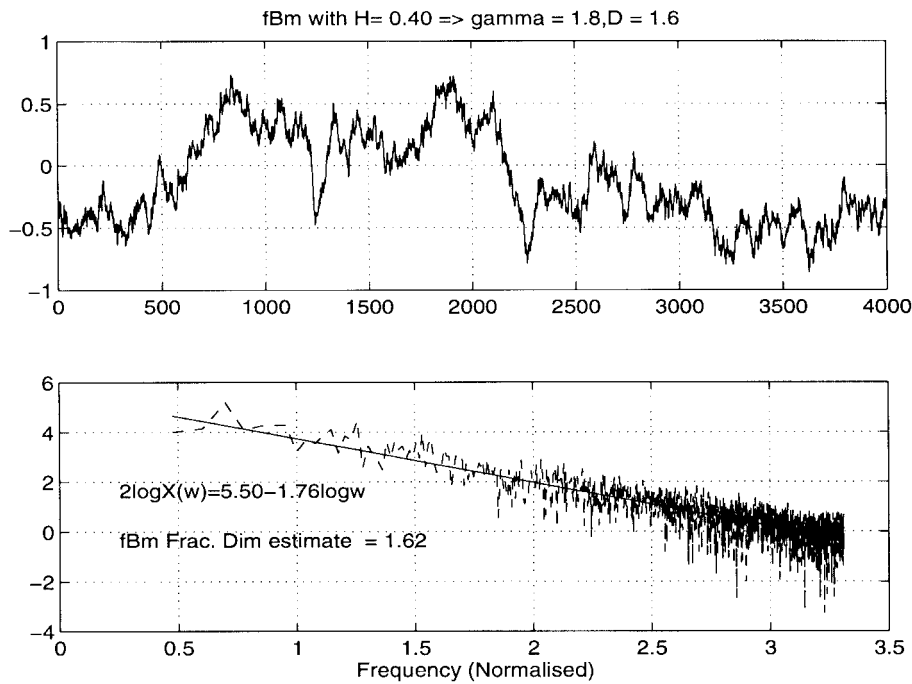


Fig. 1. Illustration of the use of linear regression of a periodogram plot to estimate the spectral exponent. Here the measured sample path is not noise-corrupted, the true $\gamma = 1.8$ and the estimate is found as $\hat{\gamma} = 1.76$.

interpretation is that of Solo [39], where it is proved that in fact fBm does reconcile the theoretical difficulty with empirical observations of $1/f^\gamma$ processes in that the sample periodogram of an fBm process converges to a $1/f^{2H+1}$ shape.

For non-Gaussian $1/f$ -type processes, there are other models available, such as the discrete-time “fractional differencing” model [10], [19], [22], [23]

$$(1 - q^{-1})^d x_k = \varepsilon_k \quad (3)$$

where $d \in (-1/2, 1/2)$ and $\{\varepsilon_k\}$ is uncorrelated and has variance $E\{\varepsilon_k^2\} = \sigma^2$. For d in this range, $\{x_k\}$ is stationary with covariance $R(\tau)$ and spectral density $\phi(\omega)$ given by

$$R(\tau) = \frac{\Gamma(1-d)\Gamma(\tau+d)}{\Gamma(d)\Gamma(\tau+1-d)}$$

$$\phi(\omega) = \frac{\sigma^2}{2\pi(4 \sin^2(\omega/2))^d}$$

so that, as would be expected since (3) is the discrete-time equivalent to a fractional integral, the fGn and fractional differencing models are equivalent in the sense that [19] with $d = H - 1/2$

$$\frac{1}{|\omega|^{2H-1}} = \phi(\omega)\psi(\omega)$$

where $\psi(\omega)$ is a positive continuous function. Another $1/f$ process model which is particularly important to this paper is that proposed by Wornell [43], [44], [46] in which the wavelet series expansion of a stationary in time but uncorrelated in time and scale process is used. More detailed discussion of this model will be deferred to a later section.

III. ESTIMATION OF $1/f$ PROCESSES

As mentioned in the Introduction, there is great interest in estimating the spectral exponent γ from a length N observed sample path realization $\{x_k\}$ of a $1/f^\gamma$ process. For example, in [28] such an estimate provides a measure of image texture, while in [20] it serves as a cardiac health diagnostic instrument. The purpose of this section is to provide a brief overview of previous work on estimation of $1/f$ processes so that the results to be presented in Section V can be seen in context.

The simplest approach to estimating γ is to assume that there are no external noise corruptions on the available measurements and to then calculate the sample periodogram $I_N(\omega)$ defined as

$$I_N(\omega) \triangleq \frac{1}{N} \left| \sum_{k=0}^{N-1} x_k e^{-j\omega k} \right|^2.$$

The estimate $\hat{\gamma}$ of γ is then taken as the least squares estimated slope of $I_N(\omega)$ when plotted on log-log axes; see Fig. 1 for an illustration of this method where the fBm process is simulated using the “random midpoint displacement method” [13]. The most fundamental result on the performance of this technique is that [39] it is a mean-square-consistent estimate if the $1/f$ process can be modeled as an fBm or fGn. More refined information is provided by Leu and Papamarcou [26] who show that the mean-square rate of convergence with increasing observation length N is³

$$\text{Var}\{\hat{\gamma}\} = o\left(\frac{\log^2 N}{N}\right), \quad \text{as } N \rightarrow \infty.$$

³Here and in the sequel the notation $g(N) = o(f(N))$ as $N \rightarrow \infty$ will mean $\lim_{N \rightarrow \infty} g(N)/f(N) = 0$ and the notation $g(N) = O(f(N))$ as $N \rightarrow \infty$ will mean $\lim_{N \rightarrow \infty} g(N)/f(N) = C$ where $0 < C < \infty$.

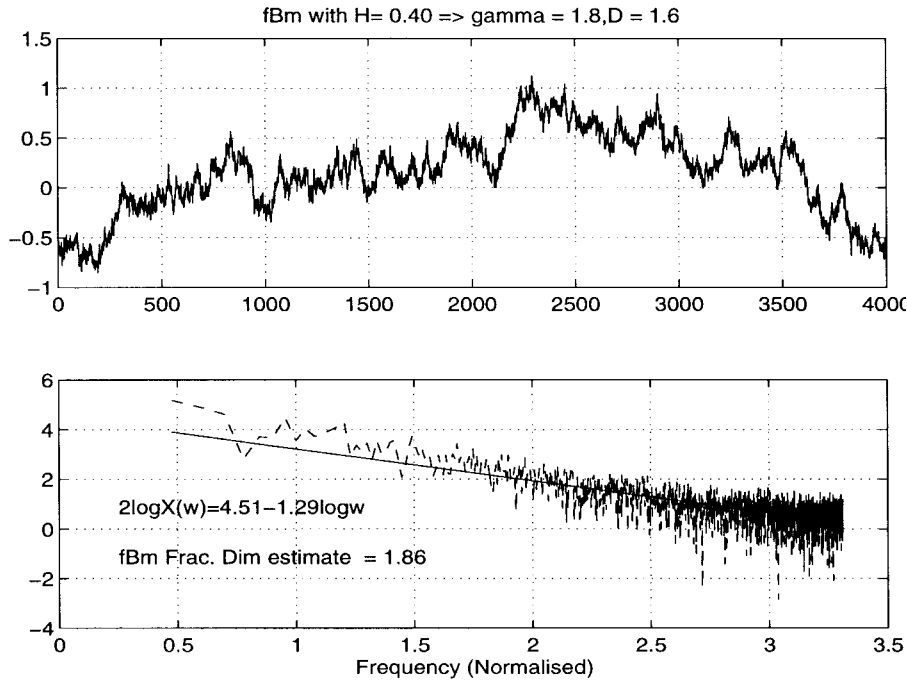


Fig. 2. Illustration of the use of linear regression of a periodogram plot to estimate the spectral exponent. Here the same measured sample path as the previous figure is used, but now it is corrupted by white Gaussian noise of variance 0.001. The true value is still $\gamma = 1.8$ but now the estimate is found as $\hat{\gamma} = 1.29$.

When a fractional differencing model (3) is used to model the $1/f$ process, then the resulting estimate is also known to be (weakly) consistent [19], [23] and asymptotically Gaussian. Note that this periodogram/linear-regression-based method can be seriously affected by the effect of noise corruption of the sample path observations as is illustrated in Fig. 2, where white noise of variance $\sigma_v^2 = 0.001$ has been added to the measurements.

It is also possible to estimate the shape of quite general spectral densities $\phi(\omega, \theta)$ that are defined by a vector of parameters θ by using the methods of maximum likelihood; a special case of this then becomes the method of interest to us where $\theta = [\sigma^2, \gamma, \sigma_v^2]$ and $\phi(\omega, \theta) = \sigma^2 |\omega|^{-\gamma} + \sigma_v^2$.

Two approaches have been analyzed in the literature. First, the methods of Whittle [42] can be used to approximate the log-likelihood function by using the periodogram so that an approximate ML estimate $\hat{\theta}$ is given by

$$\hat{\theta} = \arg \min_{\theta} \left\{ \int_{-\pi}^{\pi} \log \phi(\omega, \theta) + \frac{I_N(\omega)}{\phi(\omega, \theta)} d\omega \right\}.$$

Secondly, at (significantly) more computational expense, the exact log likelihood can be computed to find the exact ML estimate as

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{N} \left\{ \log \det [T_N(\theta)] + \sum_{r=0}^{N-1} \sum_{s=0}^{N-1} [T_N^{-1}(\theta)]_{r,s} x_r x_s \right\}$$

where

$$[T_N(\theta)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(\omega, \theta) e^{j\omega(r-s)} d\omega.$$

For $\gamma \in (0, 1)$, Fox and Taqqu [17] have shown the approximate ML estimator to be strongly consistent, asymptotically Gaussian, and efficient. Leu and Papamorcou [26] have extended these results (by employing stronger assumptions on $\phi(\omega, \theta)$ that disallow estimation with white-noise-corrupted measurements) to also hold for $\gamma \in (1, 2)$ and have estimated the mean-square rate of convergence of the approximate ML method as

$$\text{Var} \{\hat{\gamma}\} = o\left(\frac{1}{N}\right), \quad \text{as } N \rightarrow \infty.$$

For $\gamma \in (0, 1)$, Dalhaus [8] has shown the same distributional results, but only weak consistency for the exact ML scheme. These results would appear to also be of relevance to various ML estimation methods [10], [11] that have been developed around the fractional differencing model (3).

In Section V, the purpose of this paper will be to analyze a particular ML scheme that is approximate for a fBm or fGn model of $1/f$ processes (but exact for the wavelet model of Wornell [44]). This analysis will establish the scheme to be strongly consistent, asymptotically Normal, and efficient, but if and only if $\gamma \in (0, 1)$. The results in Section V will also establish precise mean-square convergence rates that are stronger than those presented above (those above bound the convergence rate, but do not establish what it is) and show an explicit dependence on γ . The results in Section V are not implied by (nor do they imply) the just-surveyed ML results since the latter are predicated on a stationary model for the $1/f$ process whereas the wavelet-based model [44] studied in Section V is a nonstationary one.

Note that other estimation methods based on ideas of ‘‘Allan Variance’’ and measuring fractal dimension are also available [18], [26], but will not be commented on here.

IV. WAVELET TRANSFORMS AND $1/f$ PROCESSES

Given the multiscale nature of the fBm model of $1/f$ processes, and the multiscale motivation of much of wavelet theory [9], [30], it arises that wavelets are a powerful analysis tool for studying $1/f$ processes [6], [12], [16], [29], and [32].

To begin with, given a signal $X(t)$ and wavelet $\psi(t)$, the wavelet transform of $X(t)$ at time t and scale a is defined as [9]

$$(\mathcal{W}X)(a, t) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} X(\sigma) \psi\left(\frac{\sigma - t}{a}\right) d\sigma. \quad (4)$$

A key feature of this transform [16], [32] is that if $X(t) = B_H(t)$, then even though $X(t)$ is nonstationary, its wavelet transform $(\mathcal{W}X)(a, t)$ is stationary in t for fixed a and has spectral density $\phi(\omega)$ given by⁴

$$\phi(\omega) = a |\hat{\psi}(a\omega)|^2 \frac{1}{|\omega|^{2H+1}}$$

which provides yet another interpretation of the spectrum of an fBm being a $1/f^\gamma$ process with $\gamma = 2H + 1$. As well, no Gaussian process other than fBm shares these second-order properties of its wavelet transform [37].

Aside from this, it also provides another method of estimating γ from observed data [30] since the variance of $(\mathcal{W}B_H)(a, t)$ obeys

$$\sigma_a^2 \triangleq \mathbb{E}\{[(\mathcal{W}B_H)(a, t)]^2\} \propto a \int_{-\infty}^{\infty} \frac{|\hat{\psi}(a\omega)|^2}{|\omega|^{2H+1}} d\omega \quad (5)$$

so that for $a' \neq a$

$$\sigma_{a'}^2 \propto a' \int_{-\infty}^{\infty} \frac{|\hat{\psi}(a'\omega)|^2}{|\omega|^{2H+1}} d\omega \propto \left(\frac{a'}{a}\right)^{2H+1} \sigma_a^2. \quad (6)$$

Therefore, since $\{(\mathcal{W}B_H)(a, t)\}$ is stationary, an estimate (Δ is a sampling interval)

$$\hat{\sigma}_a^2 = \frac{1}{N} \sum_{k=0}^{N-1} [(\mathcal{W}B_H)(a, t + k\Delta)]^2$$

of σ_a^2 can be taken for a range of $a = \{a_1, a_2, \dots, a_M\}$ distributed such that for some β , $a_{k+1} = \beta a_k$ and hence by the relationship (6) an intuitive estimate of γ is the slope of a graph of $\hat{\sigma}_{a_k}$ versus a_k with respect to $\log - \log_\beta$ axes.

Due to computational considerations, the wavelet transform is not commonly used in the form (4). Rather, a particular class of wavelets $\psi(t)$ admitting a “multiscale” representation [9] are employed so that a needs only be varied on a so-called “dyadic” scale $a = 1/2^n$. This leads to a more compact notation

$$c_k^n \triangleq (\mathcal{W}B_H)(2^{-n}, 2^{-n}k)$$

where the wavelet coefficients $\{c_k^n\}$ may be efficiently calculated using a bank of filters with impulse responses $\{h_k\}$ and

$\{g_k\}$ and with subsampled outputs

$$c_k^{n-1} = \sum_{\nu=-\infty}^{\infty} g_{\nu-2k} d_k^n$$

$$d_k^{m-1} = \sum_{\nu=-\infty}^{\infty} h_{\nu-2k} d_k^m.$$

This method is commonly initialized with $c_k^n = X(k\Delta)$ for some sampling period Δ , and is also most commonly used with $\psi(t)$ chosen such that the impulse responses $\{h_k\}$ and $\{g_k\}$ are finite (FIR) in which case the above method is very computationally efficient and is known as the “Fast Wavelet Transform” [9].

As well, when $\psi(t)$ is chosen so that the wavelet transform need only be calculated at dyadic points the inverse wavelet transform becomes a series expansion

$$X(t) = \sum_{m=-\infty}^{\infty} 2^{m/2} \sum_{n=-\infty}^{\infty} c_n^m \psi(2^m t - n), \quad (7)$$

Wornell [43], [44], [46] has shown how this can be used to provide a new representation for $1/f^\gamma$ processes over very large ranges of γ . Specifically, if the wavelet coefficients $\{c_n^m\}$ are uncorrelated across time and scale, are zero-mean, and have variance

$$\mathbb{E}\{(c_n^m)^2\} = \frac{\sigma^2}{2^{\gamma m}} \quad (8)$$

then with $\{\psi(t)\}$ an orthonormal wavelet basis $X(t)$ possesses a “time-averaged” spectrum

$$\phi(\omega) = \sigma^2 \sum_{m=-\infty}^{\infty} \frac{|\hat{\psi}(\omega/2^m)|^2}{2^{\gamma m}}$$

that is nearly of $1/f^\gamma$ type

$$\frac{\sigma_L^2}{|\omega|^\gamma} \leq \phi(\omega) \leq \frac{\sigma_U^2}{|\omega|^\gamma}, \quad \sigma_U^2 \geq \sigma_L^2.$$

This applies for γ larger than valid for the fBm or fGn models of $1/f$ processes ($\gamma = 5$, for example), but requires greater regularity of wavelets for higher γ . By this it is meant that the wavelets possess a greater number of vanishing moments

$$\int_{-\infty}^{\infty} t^p \psi(t) dt = 0, \quad p \in [0, P-1].$$

Note that it is known how to design wavelets with an arbitrarily large number P of vanishing moments [9].

Of most interest to this paper is that results in the reverse direction also hold, namely, that given a $1/f$ process, the wavelet transform acts as a “whitening” filter on the process as would be suggested by the above synthesis result. This has been rigorously analyzed for the specific case of fBm, where with $a = 2^n$, $\gamma = 2H + 1$, (5) gives (8) for the variance of the wavelet coefficients and the work in [40] shows that both along and across scales the coefficients are approximately uncorrelated in that for some $C < \infty$

$$\mathbb{E}\{c_k^n c_\ell^m\} \leq \frac{C 2^{2(m+n)(P-H)}}{(2^n k - 2^m \ell)^{2(P-H)}}$$

⁴For $\phi(t)$ a given function, $\hat{\phi}(\omega)$ denotes its Fourier transform.

provided that $|2^{-n}k - 2^{-m}\ell| > \max(2^{-n}\xi, 2^{-m}\xi)$. Similar results are reported in [12] so that for $P \gg 1$ the correlation decay along scales is rapid as the figures in [40] illustrate. These latter figures also show rapid decay across scales, with the rapidity again proportional to the number P of vanishing moments, which as previously mentioned can be made arbitrarily large by the appropriate choice of a wavelet. For example, with the Daubechies scheme of wavelet construction, as the FIR filters in the fast wavelet transform become arbitrarily large P becomes arbitrarily large for the underlying wavelet.

V. WAVELET-BASED ESTIMATORS FOR $1/f$ PROCESSES

In [45], Wornell and Oppenheim proposed that this “whitening” filter property of the wavelet transform be exploited to derive an approximate ML estimator for $1/f$ processes that is much more computationally efficient than the ML methods surveyed in Section III. Note, that as previously mentioned, this wavelet-based ML approach is not equivalent to other ML approaches since it employs a nonstationary rather than a stationary model for the underlying $1/f$ process.

In [45], the ML estimator is derived on the assumption that the observations of the $1/f$ process sample path $X(t)$ may possibly be corrupted by white noise so that the fast wavelet transform of the observations can be written as

$$r_n^m = c_n^m + \nu_n$$

where $\{\nu_k\}$ is white, zero-mean, Gaussian, and has variance $\mathbf{E}\{\nu_k^2\} = \sigma_\nu^2$ and $\{c_n^m\}$ are the wavelet coefficients of the sample path $X(t)$ of the $1/f$ process. With the assumption that $X(t)$ is Gaussian

$$r_n^m \sim \mathcal{N}(0, \alpha\lambda^m + \sigma_\nu^2) = \mathcal{N}(0, \sigma_m^2(\theta))$$

where $\theta^T \triangleq [\alpha, \lambda, \sigma_\nu^2]$, $\lambda \triangleq 1/2^\gamma$, and α is a parameter to scale the magnitude, but not the distribution of energy with frequency of the $1/f$ process.

If the fast wavelet transform is used to calculate $\{r_n^m\}$ then it is necessary to assume a data record of length $N = N_0 2^M$ for some integral N_0 and M in which case with the notation $R_M \triangleq \{r_n^m\}$ the negative log-likelihood function $\ell(R_M|\theta)$ for the data given the parameter vector θ is (modulo an additive constant)

$$\begin{aligned} \ell(R_M|\theta) &= -\log \left(\prod_{m=1}^M \prod_{n=1}^{N_0 2^{m-1}} \frac{1}{\sqrt{2\pi\sigma_m^2(\theta)}} e^{-(1/2)(r_n^m/\sigma_m(\theta))^2} \right) \\ &= \frac{1}{2} \sum_{m=1}^M \sum_{n=1}^{N_0 2^{m-1}} \log \sigma_m^2(\theta) + \left(\frac{r_n^m}{\sigma_m(\theta)} \right)^2. \end{aligned}$$

Equivalently, the formulation suggested by Wornell and Oppenheim of

$$\ell(Z_M|\theta) \triangleq \frac{1}{2} \sum_{m=1}^M N_0 2^{m-1} \log \sigma_m^2(\theta) + \frac{z_m^2}{\sigma_m^2(\theta)} \quad (9)$$

can be used where

$$Z_M \triangleq \{z_1, \dots, z_M\} \quad z_m^2 = \sum_{n=1}^{N_0 2^{m-1}} (r_n^m)^2.$$

Now, define the ML estimate $\hat{\theta}_M$ as

$$\hat{\theta}_M = \underset{\theta \in \Theta}{\operatorname{argmin}} \{Q_M(Z_M|\theta)\} \quad (10)$$

where

$$Q_M(Z_M|\theta) \triangleq \frac{1}{2^M} \ell(Z_M|\theta). \quad (11)$$

Wornell and Oppenheim [45] do not use the normalizing factor of $1/2^M$ in (11) and claim in [45] that “*It is well known that ML estimators are generally asymptotically efficient and consistent. This, specifically, turns out to be the case here.*” Presumably this comment is based on empirical evidence since in [45] Wornell and Oppenheim offer Monte Carlo simulations, but no theoretical analysis. The purpose of this section is to show that if this theoretical analysis is performed, then it can be concluded that when $\sigma_\nu^2 > 0$ the method is only consistent and asymptotically efficient for $\gamma \in (0, 1)$.

As a preliminary comment, note that it is not true that all ML estimators enjoy pleasant properties, it being relatively easy to devise examples wherein the estimates are not consistent or, in fact, even well-defined [5], [38].

As a consequence of this, the convergence of any ML estimate needs to be rigorously examined on an individual basis. Performing this for the case at hand allows the conclusion that the ML estimator defined by (10) and (11) is always convergent at least to a bounded set.

Theorem 1: Suppose that Θ is compact and that $0 \notin \Theta$. Then for $\gamma > 0$

$$\lim_{M \rightarrow \infty} \hat{\theta}_M \in \mathcal{S} \triangleq \{\theta \in \Theta: \overline{Q}(\theta) \leq \overline{Q}(\beta) \forall \beta \in \Theta\}$$

where

$$\overline{Q}(\theta) \triangleq \lim_{M \rightarrow \infty} \mathbf{E}\{Q_M(Z|\theta)\}.$$

Proof: The idea of the proof is to first show that $Q_M(\theta) \rightarrow \mathbf{E}\{Q_M(\theta)\}$ with probability one as $M \rightarrow \infty$ and that this convergence is uniform in θ over compact Θ , and to then show that this uniform convergence on a compact space implies convergence of the minima. The following definitions are required:

$$Q_M(Z|\theta) \triangleq \frac{1}{2^M} \sum_{m=1}^M f(m, \theta, z_m)$$

$$f(m, \theta, z_m) \triangleq N_0 2^{m-1} \log \sigma_m^2(\theta) + \frac{z_m^2}{\sigma_m^2(\theta)}$$

$$\tau_m \triangleq f(m, \theta, z_m) - \mathbf{E}\{f(m, \theta, z_m)\}$$

$$= \frac{z_m^2 - \mathbf{E}\{z_m^2\}}{\sigma_m^2(\theta)}$$

$$S_M \triangleq \sum_{m=1}^M \tau_m$$

$$\mathcal{E}_M(\varepsilon) \triangleq \{\omega \in \Omega: |S_M| > 2^M \varepsilon; \quad \varepsilon > 0\}$$

where in the above all random variables are defined on an underlying probability space $\{\Omega, \mathcal{F}, \mathbf{P}\}$. Therefore, by Chebychev's inequality

$$\mathbf{P}\{\mathcal{E}_M(\varepsilon)\} \leq \frac{\mathbf{E}\{S_M^2\}}{2^{2M}\varepsilon^2}.$$

Now

$$\mathbf{E}\{S_M^2\} = \sum_{m=1}^N \mathbf{E}\{\tau_m^2\} + 2 \sum_{m=1}^M \sum_{\ell>m} \mathbf{E}\{\tau_m \tau_\ell\}$$

and also

$$z_m^2 = \sum_{n=1}^{N_0 2^{m-1}} (r_n^m)^2 = R^T R$$

where

$$R^T \triangleq [r_1^m, \dots, r_{N_0 2^{m-1}}^m], \quad R \sim \mathcal{N}(0, \sigma_m^2(\theta_0)I).$$

Therefore,

$$\frac{1}{\sigma_m^2(\theta_0)} R^T R \sim \chi_{N_0 2^{m-1}}^2 \quad (12)$$

so by Lemma A.5 in the Appendix

$$\mathbf{E}\{z_m^2\} = N_0 2^{m-1} \sigma_m^2(\theta_0). \quad (13)$$

Also, using Lemma A.5 in the Appendix

$$\begin{aligned} \mathbf{E}\{(z_m^2)^2\} &= \mathbf{E}\{(R^T R)^2\} \\ &= [\sigma_m^2(\theta_0) \text{Trace}\{I\}]^2 + 2\sigma_m^4(\theta_0) \text{Trace}\{I\} \\ &= \sigma_m^4(\theta_0) N_0 2^{m-1} [N_0 2^{m-1} + 2] \end{aligned} \quad (14)$$

so that

$$\begin{aligned} \sigma_m^4(\theta) \mathbf{E}\{\tau_m^2\} &= \mathbf{E}\{(z_m^2)^2\} - (\mathbf{E}\{z_m^2\})^2 \\ &= \sigma_m^4(\theta_0) N_0 2^{m-1} [N_0 2^{m-1} + 2] \\ &\quad - N_0^2 2^{2m-2} \sigma_m^4(\theta_0) \\ &= N_0 2^m \sigma_m^4(\theta_0) \end{aligned}$$

to give

$$\mathbf{E}\{\tau_m^2\} = N_0 2^m \left(\frac{\sigma_m^2(\theta_0)}{\sigma_m^2(\theta)} \right)^2.$$

Also

$$\sigma_m^2(\theta) \sigma_\ell^2(\theta) \mathbf{E}\{\tau_m \tau_\ell\} = \mathbf{E}\{z_m^2 z_\ell^2\} - \mathbf{E}\{z_m^2\} \mathbf{E}\{z_\ell^2\}$$

and

$$\mathbf{E}\{z_m^2 z_\ell^2\} = \mathbf{E}\{R^T R W^T W\}$$

where

$$\begin{aligned} R^T &= [r_1^m, \dots, r_{N_0 2^{m-1}}^m] & R &\sim \mathcal{N}(0, \sigma_m^2(\theta_0)I) \\ W^T &= [r_1^\ell, \dots, r_{N_0 2^{\ell-1}}^\ell] & W &\sim \mathcal{N}(0, \sigma_\ell^2(\theta_0)I) \end{aligned}$$

so that since W and R are uncorrelated it may be concluded that

$$\begin{aligned} \mathbf{E}\{z_m^2 z_\ell^2\} &= \text{Trace}\{\sigma_m^2(\theta_0)I\} \text{Trace}\{\sigma_\ell^2(\theta_0)I\} \\ &= \sigma_m^2(\theta_0) \sigma_\ell^2(\theta_0) N_0^2 2^{m+\ell-2}. \end{aligned} \quad (15)$$

But

$$\mathbf{E}\{z_m^2\} \mathbf{E}\{z_\ell^2\} = \sigma_m^2(\theta_0) \sigma_\ell^2(\theta_0) N_0^2 2^{m+\ell-2}$$

so that $\mathbf{E}\{\tau_m \tau_\ell\} = 0$ and hence

$$\mathbf{E}\{S_M^2\} = \sum_{m=1}^M \mathbf{E}\{\tau_m^2\} = N_0 \sum_{m=1}^M 2^m \left(\frac{\sigma_m^2(\theta_0)}{\sigma_m^2(\theta)} \right)^2.$$

By assumption $\theta = 0 \notin \Theta$ so $\exists \kappa < \infty$ such that

$$\sigma_m^2(\theta_0) / \sigma_m^2(\theta) < \kappa$$

on Θ , and so $\mathbf{E}\{S_M^2\} \leq \kappa' 2^M$ for some $\kappa' < \infty$. Therefore,

$$\begin{aligned} \sup_{\theta \in \Theta} \left\{ \sum_{M=1}^{\infty} \mathbf{P}\{\mathcal{E}_M(\varepsilon)\} \right\} &\leq \sum_{M=1}^{\infty} \frac{1}{2^{2M}\varepsilon^2} \mathbf{E}\{S_M^2\} \\ &\leq \frac{\kappa'}{\varepsilon^2} \sum_{M=1}^{\infty} \frac{1}{2^M} < \infty \end{aligned}$$

so that by the Borel–Cantelli Lemma [7]

$$\frac{1}{2^M} \sum_{m=1}^M \tau_m \xrightarrow{\text{a.s.}} 0 \text{ as } M \rightarrow \infty$$

uniformly in $\theta \in \Theta$. By the definition of τ_m this means that $\forall \varepsilon > 0 \exists M_1$ such that

$$|Q_M(Z|\theta) - \mathbf{E}\{Q_M(Z|\theta)\}| < \varepsilon/2 \quad \text{a.s. } \forall M > M_1, \forall \theta \in \Theta. \quad (16)$$

Therefore, for $M > M_1$

$$\mathbf{E}\{Q_M(Z|\theta)\} > Q_M(Z|\theta) - \varepsilon/2 \quad \text{a.s. } \forall \theta \in \Theta.$$

However, by the definition of $\hat{\theta}_M$,

$$Q_M(Z|\theta) \geq Q_M(Z|\hat{\theta}_M) \quad \forall \theta \in \Theta.$$

Therefore,

$$\mathbf{E}\{Q_M(Z|\theta)\} > Q_M(Z|\hat{\theta}_M) - \varepsilon/2 \quad \text{a.s. } \forall \theta \in \Theta. \quad (17)$$

Now, Θ is compact, so \exists a subsequence $\{\hat{\theta}_{M_n}\}$ of $\{\hat{\theta}_M\}$ and a $\theta^* \in \Theta$ such that $\hat{\theta}_{M_n} \rightarrow \theta^*$ as $M_n \rightarrow \infty$. Also, $Q_M(Z|\theta)$ is continuous in θ so there exists a neighborhood \mathcal{D} of θ^* such that

$$|Q_M(Z|\hat{\theta}_{M_n}) - Q_M(Z|\theta^*)| < \varepsilon/4 \quad \forall \theta \in \mathcal{D}.$$

As well, since $Q_M(Z|\theta)$ converges uniformly to $\mathbf{E}\{Q_M(Z|\theta)\}$ then $\exists M_3$ such that $\forall M_n > M_3$

$$|Q_{M_n}(Z|\theta^*) - \mathbf{E}\{Q_{M_n}(Z|\theta^*)\}| < \varepsilon/4.$$

Combining the previous two bounds with (17) and defining M_2 such that $M_n > M_2 \Rightarrow \hat{\theta}_{M_n} \in \mathcal{D}$ means that $\forall M_n >$

$\max(M_1, M_2, M_3)$

$$\mathbf{E}\{Q_{M_n}(Z|\theta)\} > \mathbf{E}\{Q_{M_n}(Z|\theta^*)\} - \varepsilon.$$

Since ε is arbitrary this means that

$$\lim_{M_n \rightarrow \infty} \hat{\theta}_{M_n} = \theta^* \in \mathcal{S} \triangleq \{\theta \in \Theta: \overline{Q}(\theta) \leq \overline{Q}(\beta) \forall \beta \in \Theta\}$$

where

$$\overline{Q}(\theta) \triangleq \lim_{M \rightarrow \infty} \mathbf{E}\{Q_M(Z|\theta)\}.$$

Observing that this applies for any convergent subsequence of $\hat{\theta}_M$ then completes the proof. \square

Note that for any finite M , then (modulo numerical difficulties arising from trying to find the minima of a function possessing very large magnitude) the addition of the $1/2^M$ normalizing factor in (11) makes no difference to the estimate obtained. However, by its definition, consistency is an asymptotic property, not a finite-data property, and for it to exist it is important that the normalizing factor $1/2^M$ is included in (11) since with its perusal of the proof of Theorem 1 shows that

$$\mathbf{E}\{Q_M(Z|\theta)\} = \frac{N_0}{2^{M+1}} \sum_{m=1}^M 2^m \left\{ \log \sigma_m^2(\theta) + \frac{\sigma_m^2(\theta_0)}{\sigma_m^2(\theta)} \right\}$$

which if $0 \notin \Theta$ is obviously convergent to a well-defined function $\overline{Q}(\theta)$ on Θ when $M \rightarrow \infty$. Without the normalizing factor $1/2^M$ then $\mathbf{E}\{Q_M(Z|\theta)\}$ (and hence by (16) $Q_M(Z|\theta)$ as well) diverges uniformly in θ with increasing M .

In any event, for consistency of the ML estimator, it will be necessary that the set \mathcal{S} defined in Theorem 1 consists of only one point θ_0 defined to be the vector containing the true parameter values. In order to study the nature of \mathcal{S} note that a property shared by all ML estimators [25] is that

$$\mathbf{E}\{Q_M(Z|\theta)\} \geq \mathbf{E}\{Q_M(Z|\theta_0)\} \quad \forall \theta \in \Theta$$

so that at least $\theta_0 \in \mathcal{S}$ may be concluded. However, for θ_0 to be the only element in the interior of \mathcal{S} it will be necessary that in the limit as $M \rightarrow \infty$ the Hessian of $\mathbf{E}\{Q_M(Z|\theta)\}$ be positive definite at θ_0 . For finite M , the Hessian is easily calculated as (see [45, eq. (43)])

$$\begin{aligned} & \frac{d^2}{d\theta d\theta^T} \mathbf{E}\{Q_M(Z|\theta_0)\} \\ &= \frac{N_0}{2^{M+2}} \sum_{m=1}^M \frac{2^m}{(\alpha\lambda^m + \sigma_v^2)^2} \\ & \cdot \begin{bmatrix} \lambda^{2m} & (\alpha/\lambda)m\lambda^{2m} & \lambda^m \\ (\alpha/\lambda)m\lambda^{2m} & (\alpha/\lambda)^2 m^2 \lambda^{2m} & (\alpha/\lambda)m\lambda^m \\ \lambda^m & (\alpha/\lambda)m\lambda^m & 1 \end{bmatrix}. \end{aligned} \quad (18)$$

Note that due to its importance in estimation theory this Hessian evaluated at θ_0 is commonly studied under the name

of information matrix $\mathcal{I}_M(\theta_0)$ defined by

$$\begin{aligned} \mathcal{I}_M(\theta_0) &\triangleq \mathbf{E} \left\{ \frac{d^2}{d\theta d\theta^T} \ell(Z_M|\theta_0) \right\} \\ &= 2^M \frac{d^2}{d\theta d\theta^T} \mathbf{E}\{Q_M(Z_M|\theta_0)\}. \end{aligned}$$

Unfortunately, as (18) shows, the entries pertaining to α and λ in the Hessian of $\mathbf{E}\{Q_M(Z|\theta)\}$ tend to zero as M increases if $2\lambda^2 < 1$ so that $\mathbf{E}\{Q_M(Z|\theta)\}$ “flattens out” and \mathcal{S} , instead of being the singleton θ_0 , is a set of values of θ of equal cost. In other words, since $\gamma = |\log_2 \lambda|$, $\overline{Q}(\theta)$ possesses a unique minimum and hence \mathcal{S} is a singleton at θ_0 if and only if $\gamma \in (0, 1)$.

Corollary 1: Define θ_0 to be a vector containing the true values of $[\alpha, \lambda, \sigma_v^2]$. Under the conditions that $\theta_0 \in \text{Interior}\{\Theta\}$, $0 \notin \Theta$, and $\gamma \in (0, 1)$ then

$$\hat{\theta}_M \xrightarrow{\text{a.s.}} \theta_0 \text{ as } M \rightarrow \infty. \quad (19)$$

If $\gamma > 1$ then $\hat{\theta}_M$ still converges almost surely, but to the bounded set \mathcal{S} and not necessarily to θ_0 .

It is common in studies of ML estimation methods to use the information matrix $\mathcal{I}_M(\theta_0)$ via the Cramér–Rao bound as an indication of the quality of the estimates [45]. This arises since even though the bound applies only to unbiased estimates, many ML schemes have the property that their (possibly biased) estimates converge weakly to a Gaussian distribution with covariance equal to the Cramér–Rao bound (in which case the estimators are termed “efficient”). However, to be rigorous it is necessary to prove that this convergence does in fact hold. In the case studied here, it does, but again only for appropriate $1/2^M$ normalization of the log-likelihood function and only for a narrower range of spectral exponents than originally suggested in [45].

Theorem 2: Under the conditions of Theorem 1 and provided Θ is convex, then if and only if $\gamma \in (0, 1)$

$$\sqrt{2^M} P_M^{1/2} (\hat{\theta}_M - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I) \text{ as } M \rightarrow \infty \quad (20)$$

where

$$P_M = \mathbf{E} \left\{ \frac{d^2 Q_M(Z|\theta_0)}{d\theta d\theta^T} \right\}. \quad (21)$$

Proof: By Theorem 1, for large enough M , $\hat{\theta}_M$ lies with probability one within any given open neighborhood of θ_0 which is assumed to be in the interior of Θ . Therefore, for large enough M , $Q_M(Z|\theta)$ is minimized at a $\hat{\theta}_M \in \text{Interior}\{\Theta\}$ so $\hat{\theta}_M$ satisfies the necessary condition for minimization of a differentiable function on an open domain

$$\left. \frac{dQ_M(Z|\theta)}{d\theta} \right|_{\theta=\hat{\theta}_M} = 0 \text{ a.s.}$$

Therefore, using the mean value theorem, for large enough M $\exists \mu \in [0, 1]$ such that

$$\frac{dQ_M(Z|\theta_0)}{d\theta} = R_M(\beta)(\theta_0 - \hat{\theta}_M) \text{ a.s.}$$

where

$$R_M(\beta) \triangleq \left. \frac{d^2 Q_M(Z|\theta)}{d\theta d\theta^T} \right|_{\theta=\beta}$$

$$\beta \triangleq \mu \hat{\theta}_M + (1-\mu)\theta_0.$$

Therefore, using Lemma A.2 in the Appendix

$$\sqrt{2^M} P_M^{-1/2}(\theta_0) R_M(\beta) (\theta_0 - \hat{\theta}_M) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I) \text{ as } M \rightarrow \infty. \quad (22)$$

But, by Lemma A.1 in the Appendix

$$R_M(\beta) \xrightarrow{\text{a.s.}} P_M(\beta) \text{ Uniformly in } \beta \in \Theta \quad (23)$$

and by Theorem 1

$$\hat{\theta}_M \xrightarrow{\text{a.s.}} \theta_0 \text{ as } M \rightarrow \infty \quad (24)$$

so that $\beta \xrightarrow{\text{a.s.}} \theta_0$. Combining (22)–(24) then gives

$$\sqrt{2^M} P_M^{1/2}(\theta_0) (\hat{\theta}_M - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I) \text{ as } M \rightarrow \infty. \quad \square$$

As already mentioned, with this distributional result in hand it is natural to use $\mathcal{I}_M(\theta_0)$ to infer the second-order properties of the ML estimate. However, without further examination, since neither almost-sure nor weak convergence imply mean-square convergence, there is no guarantee that these second-order properties even exist, much less than they are related to the asymptotic distributional variances. Fortunately, when this further examination is undertaken, the hoped for second-order properties of the ML estimate do in fact exist, but as usual only for restricted γ .

Corollary 2: If $\gamma \in (0, 1)$ then

$$\lim_{M \rightarrow \infty} \mathcal{I}_M^{1/2}(\theta_0) \mathbf{E}\{(\hat{\theta}_M - \theta_0)(\hat{\theta}_M - \theta_0)^T\} \mathcal{I}_M^{T/2}(\theta_0) = I.$$

Proof: From the proof of Theorem 2, for M large enough

$$\theta_0 - \hat{\theta}_M = R_M^{-1}(\beta) \frac{dQ_M(Z|\theta_0)}{d\theta}$$

where $\beta = \mu \hat{\theta}_M + (1-\mu)\theta_0$ for some $\mu \in [0, 1]$. It has already been established that for $\gamma \in (0, 1)$ then $\mathbf{E}\{R_M(\theta_0)\} > 0$ so that the continuity of $\mathbf{E}\{R_M(\theta)\}$ in θ implies that $\exists \delta > 0$ such that

$$\|\mathbf{E}\{R_M(\theta)\}\| \geq \varepsilon \quad \forall \theta \text{ such that } \|\theta - \theta_0\| \leq \delta.$$

Now, assuming that all random variables are defined on an underlying probability space $\{\Omega, \mathcal{F}, \mathbf{P}\}$, define the following subsets of the event space Ω :

$$\Omega_M^1 \triangleq \{\omega \in \Omega: \|\hat{\theta}_M - \theta_0\| < \delta\},$$

$$\Omega_M^2 \triangleq \{\omega: \|\mathbf{E}\{R_M(\beta)\}\| \geq \varepsilon/2 \forall \beta \text{ such that } \|\beta - \theta_0\| < \delta\}$$

and put $\Omega_M = \Omega_M^1 \cap \Omega_M^2$. In this case, since perusal of (18) shows that $\exists K < \infty$ such that $\mathbf{E}\{\|R_M(\theta_0)\|\} < K \forall M$ then

$$\begin{aligned} & \mathbf{E}\{\|\mathcal{I}_M^{1/2}(\theta_0)(\hat{\theta}_M - \theta_0)\|^4\} \\ & \leq 4^M \|\mathbf{E}\{R_M(\theta_0)\}\|^2 \mathbf{E}\left\{\|R_M^{-1}(\beta)\|^4 \left\|\frac{dQ_M(Z|\theta_0)}{d\theta}\right\|^4\right\} \\ & \leq \frac{K^2 4^{M+2}}{\varepsilon^4} \int_{\Omega_M} \left\|\frac{dQ_M(Z|\theta_0)}{d\theta}\right\|^4 d\mathbf{P} \\ & \quad + 4^M \int_{\bar{\Omega}_M} \|\hat{\theta}_M - \theta_0\|^4 d\mathbf{P}. \end{aligned}$$

Therefore, since Θ is compact, for some $C < \infty$

$$\begin{aligned} \mathbf{E}\{\|\mathcal{I}_M^{1/2}(\theta_0)(\hat{\theta}_M - \theta_0)\|^4\} & \leq C 4^M \mathbf{E}\left\{\left\|\frac{dQ_M(Z|\theta_0)}{d\theta}\right\|^4\right\} \\ & \quad + C 4^M \mathbf{P}\{\bar{\Omega}_M\}. \end{aligned} \quad (25)$$

Now

$$\|R_M(\theta)\| \geq \|\mathbf{E}\{R_M(\theta)\}\| - \|R_M(\theta) - \mathbf{E}\{R_M(\theta)\}\|$$

and

$$\|\mathbf{E}\{R_M(\theta)\}\| \geq \varepsilon \forall \|\theta - \theta_0\| < \delta$$

so that

$$\bar{\Omega}_M^2 \subset \left\{ \omega: \sup_{\|\theta - \theta_0\| < \delta} \|R_M(\theta) - \mathbf{E}\{R_M(\theta)\}\| > \varepsilon/2 \right\}$$

and so by Lemma A.3 and Chebychev's inequality $\exists K < \infty$ such that

$$\mathbf{P}\{\bar{\Omega}_M^2\} \leq \frac{K}{\varepsilon^4 4^M}.$$

Similarly, since θ_0 is the unique minimizer of $\mathbf{E}\{Q_M(Z|\theta)\}$ then for some $\rho > 0$

$$\begin{aligned} \mathbf{E}\{Q_M(Z|\xi)\} & < \mathbf{E}\{Q_M(Z|\theta)\} - \rho, \\ & \forall \|\theta - \theta_0\| > \delta, \|\xi - \theta_0\| < \delta/2. \end{aligned}$$

Therefore,

$$\bar{\Omega}_M^1 \subset \left\{ \omega: \sup_{\theta \in \Theta} |Q_M(\theta) - \mathbf{E}\{Q_M(\theta)\}| > \rho/2 \right\}$$

so that again by using Lemma A.3 and Chebychev's inequality

$$\mathbf{P}\{\bar{\Omega}_M^1\} \leq \frac{K}{\rho^2 4^M}$$

and since by de Morgan's Theorem $\bar{\Omega}_M = \bar{\Omega}_M^1 \cup \bar{\Omega}_M^2$

$$\mathbf{P}\{\bar{\Omega}_M\} \leq \frac{K}{4^M}.$$

Substituting this into (25) together with the use of Lemma A.3 gives that

$$\mathbf{E}\{\|\mathcal{I}_M^{1/2}(\theta_0)(\hat{\theta}_M - \theta_0)\|^4\} \leq C 4^M \frac{K}{4^M} + C 4^M \frac{K}{4^M} < \infty$$

so that combining Theorem 2 with [7, Theorem 4.5.2] gives the result. \square

This allows an estimate of the mean-square rate of convergence of the ML estimates to be found as follows.

Corollary 3: If $\gamma \in (0, 1)$ then

$$\text{Var}\{\hat{\alpha}\} = O\left\{\frac{1}{N^{1-\gamma}}\right\} \text{ as } N \rightarrow \infty \quad (26)$$

$$\text{Var}\{\hat{\lambda}\} = O\left\{\frac{1}{N^{1-\gamma} \log^2 N}\right\} \text{ as } N \rightarrow \infty \quad (27)$$

$$\text{Var}\{\hat{\sigma}_\nu^2\} = O\left\{\frac{1}{N}\right\} \text{ as } N \rightarrow \infty. \quad (28)$$

Proof: Define

$$\hat{\beta}_M \triangleq \sqrt{2^M} \left[\lambda^M \hat{\alpha}, M\lambda^M \left(\frac{\hat{\lambda}}{\alpha}\right), \hat{\sigma}_\nu^2 \right]^T. \quad (29)$$

Then since by Theorem 2

$$\mathcal{I}_M^{1/2}(\theta_0)(\hat{\theta}_M - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I) \text{ as } M \rightarrow \infty$$

and by the definition (29) $\hat{\beta}_M = \mathcal{T}_M \hat{\theta}_M$ where

$$\mathcal{T}_M \triangleq \sqrt{2^M} \begin{bmatrix} \lambda^M & 0 & 0 \\ 0 & M\lambda^M \left(\frac{\lambda}{\alpha}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

use of Slutsky's Theorem then gives

$$(\hat{\beta}_M - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}^{-1}(\beta_0)) \text{ as } M \rightarrow \infty$$

where

$$\mathcal{I}(\beta_0) = \lim_{M \rightarrow \infty} \mathcal{I}_M(\beta_0)$$

with

$$\mathcal{I}_M(\beta_0) = \mathcal{T}_M^{-1} \mathcal{I}_M(\theta_0) \mathcal{T}_M^{-1}$$

which using (18) may be evaluated as shown at the bottom of this page. Therefore, since by assumption $\gamma \in (0, 1)$ and hence $2\lambda^2 > 1$ then $0 < \mathcal{I}(\beta_0) < \infty$ and hence using Corollary 2

$$\lim_{M \rightarrow \infty} \mathbf{E}\{(\hat{\beta}_M - \beta_0)(\hat{\beta}_M - \beta_0)^T\} = \mathcal{I}^{-1}(\beta_0).$$

Concentrating attention now on \mathcal{T}_M and using the facts that the use of the fast wavelet transform to obtain the data $\{r_n^m\}$ implies that $M = \log_2(N/N_0)$ and remembering that $\gamma = |\log_2 \lambda|$ implies $\lambda^M = (N_0/N)^\gamma$ then completes the proof. \square

This suggests that estimates of α and γ should be more accurate for smaller γ which is in agreement with the many simulation results available in the literature [18], [28], [44], [45], and is also in accordance with intuition.

To expand on this last point of intuitive interpretation, for γ small near 0 the spectrum of the process is nearly flat so its size and slope at high frequencies is highly indicative of its overall behavior. For γ large near 1, the spectrum of the process is much more hyperbolic and hence there is a need to observe it at low frequencies in order to characterize it; observing the spectrum at low frequencies requires observing more data than does observing the spectrum at high frequencies, hence the slower rate of convergence for large γ indicated in (26) and (27). Continuing this spectral argument to also address lack of convergence, note that as γ increases and the hyperbolicity of the 1/f noise spectrum increases, the amount of energy of the 1/f process above the white-noise floor decreases, thereby making estimation of the 1/f process more difficult—eventually, a threshold is reached at $\gamma = 1$ when the white-noise energy dominates and consistency becomes impossible.

An illustration of this phenomenon of nonconvergence can be provided by reviewing a simulation study in which the ML scheme using Daubechies order ten wavelets was tested for 1/f processes with exponents of $\gamma = 0.6, 1.3,$ and 1.8 and with $\alpha = 1000$. In the numerical studies of [45] and [46], in order to investigate robustness, the 1/f process was not generated via the wavelet model (7), (8). However, as mentioned in [45], this does “not allow us to distinguish between errors inherent in the modeling and errors inherent in the estimation process.” Since the latter are precisely what are under investigation here, the opposite choice is made, and (7) as well as (8) are used to generate 1/f process realizations.

Realizations of length $N = 2^M$ with $M \in [8, 16]$ were used for estimation after being corrupted by white zero-mean Gaussian noise of variance $\sigma_\nu^2 = 0.01$. To avoid border effects, larger realizations than required were generated, and segments of appropriate length were extracted after transients due to nonzero initial conditions had evaporated.

The estimates were found using 500 iterations of the EM algorithm suggested in [45] from starting values of $\hat{\alpha} = 500, \hat{\lambda} = 1.0, \hat{\sigma}_\nu^2 = 0.0001$. For each data length, 100 different noise realizations were obtained with estimation results being averaged over these 100 trials to obtain sample means and variances for the estimates. The results for $\hat{\gamma}$ are shown in Fig. 3. Note that for $\gamma = 0.6$, as predicted by the theoretical analysis of this paper, the estimate is convergent with increasing data length. Note also that for $\gamma = 1.3$ and $\gamma = 1.8$ that (again consistent with the theoretical analysis) the estimates are not convergent to their true values—the variability of the estimates does not continually decrease, and the means drift away from the true values.

$$\mathcal{I}_M(\beta_0) = \frac{N_0}{2^{M+2}} \sum_{m=1}^M \frac{2^m}{(\alpha\lambda^m + \sigma_\nu^2)^2} \begin{bmatrix} \lambda^{2(m-M)} & m/M\lambda^{2(m-M)} & \lambda^{(m-M)} \\ m/M\lambda^{2(m-M)} & m^2/M^2\lambda^{2(m-M)} & m/M\lambda^{(m-M)} \\ \lambda^{(m-M)} & m/M\lambda^{(m-M)} & 1 \end{bmatrix}.$$

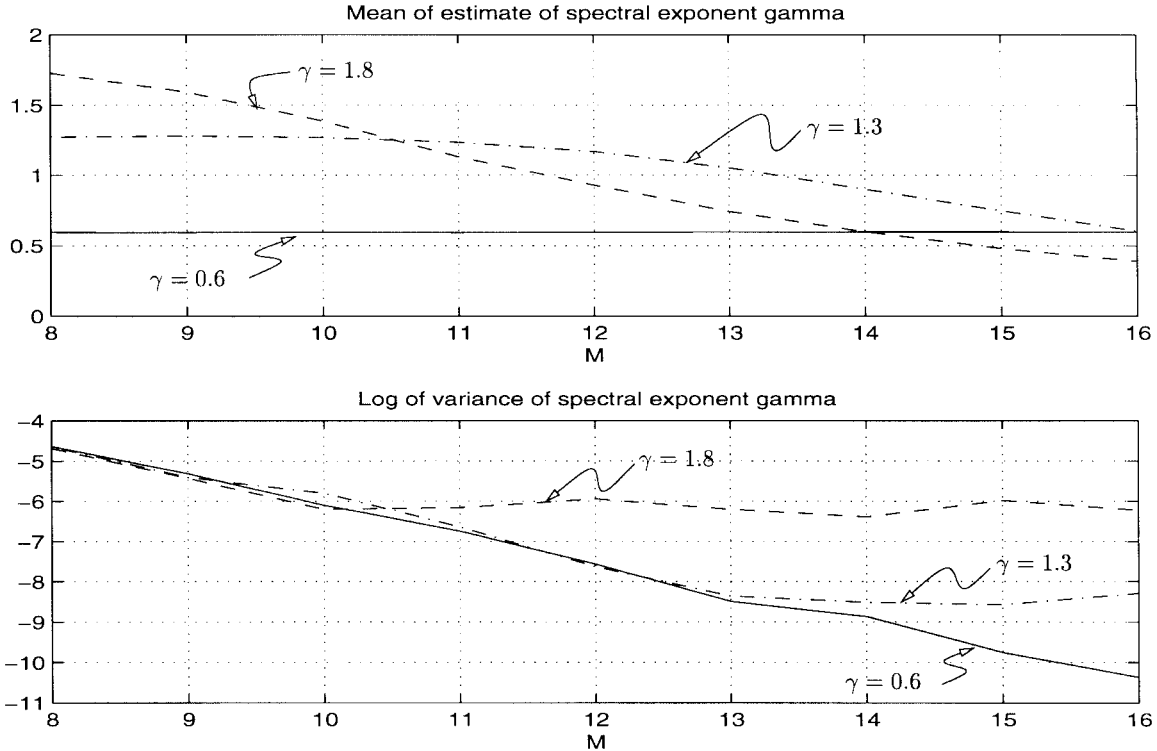


Fig. 3. Illustration of dependence of convergence of the wavelet-based ML method on $\gamma \in (0, 1)$. Solid line is mean and log of variance for $\gamma = 0.6$; estimator is convergent to correct value. Dash-dot and dashed lines are cases of $\gamma = 1.3$ and $\gamma = 1.8$, respectively; estimators are nonconvergent and are wandering around incorrect estimates. Results were obtained via 100 Monte Carlo simulations at each data length 2^M for M between 8 and 16.

At the data length of $N = 2^{16}$, the sensitivity of the estimate to the level of white-noise corruption was examined by the same process of evaluating the expected value of the wavelet-based ML estimate of γ via a sample average over 100 different realizations. The results for a range of white-noise variances from $\sigma_v^2 = 10^{-2}$ (which is the case shown in Fig. 3) down to $\sigma_v^2 = 10^{-8}$ are shown in Fig. 4. Clearly, for $\gamma > 1$ the estimate becomes progressively more biased for increasing white-noise corruption, while is almost invariant for $\gamma < 1$. Balanced against this, Fig. 4 also illustrates that for very small noise corruptions, the wavelet-based ML method can give good estimates for all values of γ .

Note that similar results to those shown in Figs. 3 and 4 were also obtained when the $1/f$ processes were generated not via the wavelet model (7), (8), but by using the “random midpoint displacement method” of simulating fractional Brownian motion [13]. In the interests of brevity, these results are not reproduced here.

VI. CONCLUSION

The use of wavelets in the analysis of fractals and the study of various noise processes has been a topic of recent research interest. More particularly, it has been suggested that wavelets be employed in the estimation of fractal dimension and also in the estimation of various $1/f^\gamma$ processes. The contribution of this paper has been, via a theoretical analysis, to investigate the utility of this approach. The main conclusion arising from this analysis is that even in the presence of external noise corruptions on the observations, the wavelet-based scheme is convergent and statistically efficient, but only for a restricted

range of γ in the range $(0, 1)$. For $\gamma > 1$, a more pessimistic conclusion (that is counter to previous suggestions in the literature) is that the methods do not converge to the correct estimate if the measurements are noise corrupted.

APPENDIX

LEMMAS REQUIRED FOR CONVERGENCE PROOFS

Lemma A.1:

$$\frac{d^2 Q_M(Z|\theta)}{d\theta_j d\theta_k} \text{ a.s.}$$

$$\frac{d^2 \mathbf{E}\{Q_M(Z|\theta)\}}{d\theta_j d\theta_k} \text{ as } M \rightarrow \infty \text{ Uniformly in } \theta \in \Theta.$$

Proof:

$$\begin{aligned} & \frac{dQ_M(Z|\theta)}{d\theta_k} \\ &= \frac{d}{d\theta_k} \left(\frac{1}{2^M} \sum_{m=1}^M N_0 2^{m-1} \log \sigma_m^2(\theta) + \frac{z_m^2}{\sigma_m^2(\theta)} \right) \\ &= \frac{1}{2^M} \sum_{m=1}^M \frac{1}{\sigma_m^2(\theta)} \frac{d\sigma_m^2(\theta)}{d\theta_k} \left(N_0 2^{m-1} - \frac{z_m^2}{\sigma_m^2(\theta)} \right). \\ & \frac{d^2 Q_M(Z|\theta)}{d\theta_j d\theta_k} \\ &= \frac{1}{2^M} \sum_{m=1}^M \frac{1}{\sigma_m^2(\theta)} \frac{d^2 \sigma_m^2(\theta)}{d\theta_j d\theta_k} \left(N_0 2^{m-1} - \frac{z_m^2}{\sigma_m^2(\theta)} \right) \\ & \quad - \frac{1}{[\sigma_m^2(\theta)]^2} \frac{d\sigma_m^2(\theta)}{d\theta_j} \frac{d\sigma_m^2(\theta)}{d\theta_k} \left(N_0 2^{m-1} - 2 \frac{z_m^2}{\sigma_m^2(\theta)} \right). \end{aligned}$$

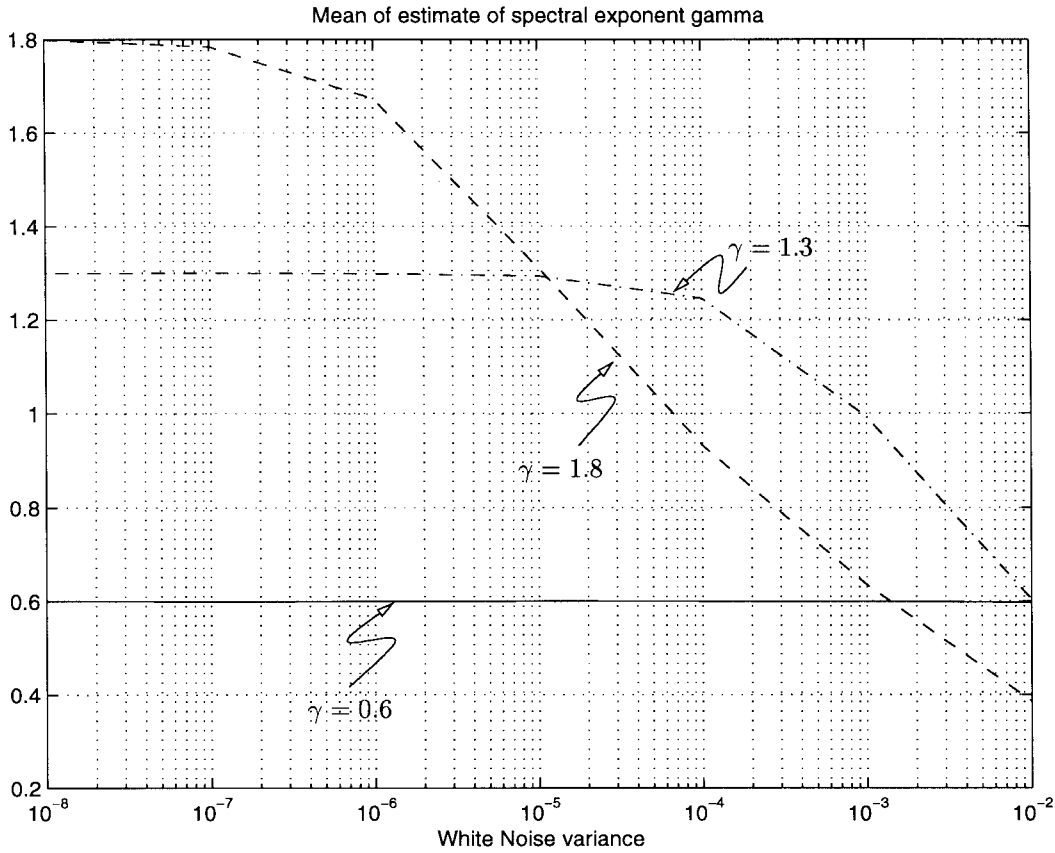


Fig. 4. Effect of estimation accuracy for decreasing signal-to-noise ratio. Solid line is mean of estimate for $\gamma = 0.6$, dash-dot and dashed lines are cases of $\gamma = 1.3$ and $\gamma = 1.8$, respectively. Results were obtained via 100 Monte Carlo simulations for a data length 2^{16} and white-noise variances logarithmically spaced between $\sigma_v^2 = 10^{-8}$ and $\sigma_v^2 = 10^{-2}$.

Therefore, defining

$$\tau_m(\theta) \triangleq \frac{(N_0 2^{2m-1} \sigma_m^2(\theta_0) - z_m^2)}{[\sigma_m^2(\theta)]^2} \cdot \left(\frac{d^2 \sigma_m^2(\theta)}{d\theta_j d\theta_k} - \frac{2}{\sigma_m^2(\theta)} \frac{d\sigma_m^2(\theta)}{d\theta_k} \frac{d\sigma_m^2(\theta)}{d\theta_j} \right)$$

gives on use of (13)

$$\frac{d^2 Q_M(Z|\theta)}{d\theta_j d\theta_k} - \frac{d^2 E\{Q_M(Z|\theta)\}}{d\theta_j d\theta_k} = \frac{1}{2^M} \sum_{m=1}^M \tau_m(\theta). \quad (\text{A.1})$$

Now put

$$\varphi_{m,j,k}(\theta)$$

$$\triangleq \frac{1}{[\sigma_m^2(\theta)]^2} \left(\frac{d^2 \sigma_m^2(\theta)}{d\theta_j d\theta_k} - \frac{2}{\sigma_m^2(\theta)} \frac{d\sigma_m^2(\theta)}{d\theta_k} \frac{d\sigma_m^2(\theta)}{d\theta_j} \right)$$

$$S_M(\theta)$$

$$\triangleq \sum_{m=1}^M \tau_m(\theta) = \sum_{m=1}^M \varphi_{m,j,k}(\theta) (N_0 2^{2m-1} \sigma_m^2(\theta_0) - z_m^2)$$

$$\mathcal{E}_M(\theta, \varepsilon) \triangleq \{\omega \in \Omega: |S_M(\theta)| > 2^M \varepsilon\}.$$

Then by Chebychev's inequality

$$P\{\mathcal{E}_M(\theta)\} \leq \frac{E\{S_M^2(\theta)\}}{2^{2M} \varepsilon^2}.$$

But

$$E\{S_M^2(\theta)\} = \sum_{m=1}^M E\{\tau_m^2(\theta)\} + 2 \sum_{m=1}^M \sum_{\ell > m}^M E\{\tau_m(\theta) \tau_\ell(\theta)\}.$$

Furthermore, using (13) and (15)

$$E\{\tau_m(\theta) \tau_\ell(\theta)\} = \varphi_{m,j,k}(\theta) \varphi_{\ell,j,k}(\theta) E\{(N_0 2^{2m-1} \sigma_m^2(\theta_0) - z_m^2) \cdot (N_0 2^{2\ell-1} \sigma_\ell^2(\theta_0) - z_\ell^2)\} = 0$$

and by (13) and (14)

$$E\{\tau_m^2(\theta)\} = \varphi_{m,j,k}^2(\theta) E\{(N_0 2^{2m-1} \sigma_m^2(\theta_0) - z_m^2)^2\} = \varphi_{m,j,k}^2(\theta) N_0 2^{2m} \sigma_m^4(\theta_0)$$

so that

$$P\{\mathcal{E}_M(\theta)\} \leq \frac{N_0}{2^{2M} \varepsilon^2} \sum_{m=1}^M 2^{2m} \varphi_{m,j,k}^2(\theta) \sigma_m^4(\theta_0).$$

Now, since $\sigma_m^2(\theta)$ is twice continuously differentiable with respect to θ , and by assumption Θ is compact such that $0 \notin \Theta$, then $\exists \kappa < \infty$ such that $|\varphi_{m,j,k}^2(\theta) \sigma_m^4(\theta_0)| < \kappa \forall \theta \in \Theta$. Therefore,

$$\sup_{\theta \in \Theta} \left\{ \sum_{m=1}^{\infty} P\{\mathcal{E}_M(\theta)\} \right\} \leq \frac{N_0 \kappa}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{1}{2^{2M}} \sum_{m=1}^M 2^{2m} < \infty$$

so that by the Borel-Cantelli Lemma [7]

$$\frac{1}{2^M} \sum_{m=1}^M \tau_m(\theta) \xrightarrow{\text{a.s.}} 0 \text{ as } M \rightarrow \infty \text{ Uniformly in } \theta \in \Theta.$$

Using (A.1) then gives the result. \square

Lemma A.2:

$$\sqrt{2^M} P_M^{-1/2}(\theta_0) \frac{dQ_M(Z|\theta)}{d\theta} \Big|_{\theta=\theta_0} \xrightarrow{\mathcal{D}} \mathcal{N}(0, I) \text{ as } M \rightarrow \infty$$

where

$$P_M(\theta_0) \triangleq \mathbb{E} \left\{ \frac{dQ_M(Z|\theta_0)}{d\theta d\theta^T} \right\}.$$

Proof: Define

$$\begin{aligned} \tau_m &\triangleq \frac{1}{2^M \sigma_m^2(\theta_0)} \left(N_0 2^{m-1} - \frac{z_m^2}{\sigma_m^2(\theta_0)} \right) \\ &\quad \cdot \sum_{k=1}^3 \alpha_k \frac{d\sigma_m^2(\theta_0)}{d\theta_k} \\ X_{m,M} &\triangleq \frac{\tau_m}{\sqrt{\varphi_M}}, \quad \varphi_M \triangleq \sum_{m=1}^M \mathbb{E}\{\tau_m^2\} \end{aligned}$$

where $\{\alpha_1, \alpha_2, \alpha_3\}$ are arbitrary real numbers. Then by (13) and (14)

$$\begin{aligned} \mathbb{E}\{X_{m,M}\} &= 0 \\ \sum_{m=1}^M \mathbb{E}\{X_{m,M}^2\} &= \frac{1}{\varphi_M} \sum_{m=1}^M \mathbb{E}\{\tau_m^2\} = 1. \end{aligned}$$

Also, since by assumption all the $\{z_k\}$ are not point masses, then $\varphi_M > 0$ to give that $\exists \kappa < \infty$ such that

$$\mathbb{E}\{|X_{m,M}|^3\} \leq \kappa \mathbb{E}\{|\tau_m|^3\} \leq \kappa \sqrt{\mathbb{E}\{\tau_m^2\}} \sqrt{\mathbb{E}\{\tau_m^4\}}.$$

But by (13), (14), and the definition of $\tau_m(\theta)$

$$\mathbb{E}\{\tau_m^2\} = \frac{N_0 2^m}{2^{2M} [\sigma_m^2(\theta_0)]^2} \left| \sum_{k=1}^3 \alpha_k \frac{d\sigma_m^2(\theta_0)}{d\theta_k} \right|^2.$$

Also,

$$\begin{aligned} \mathbb{E}\{\tau_m^4\} &= \frac{1}{2^{4M} [\sigma_m^2(\theta_0)]^4} \left| \sum_{k=1}^3 \alpha_k \frac{d\sigma_m^2(\theta_0)}{d\theta_k} \right|^4 \sum_{\ell=0}^4 \binom{4}{\ell} \\ &\quad \cdot (-1)^\ell N_0^\ell 2^{m\ell - \ell} \mathbb{E} \left\{ \left(\frac{z_m^2}{\sigma_m^2(\theta_0)} \right)^{4-\ell} \right\}. \quad (\text{A.2}) \end{aligned}$$

But by (12), $[\sigma_m^2(\theta_0)]^{-1} z_m^2 \sim \chi_{N_0 2^{m-1}}^2$ which has moment-generating function $\phi(s)$ given by [35]

$$\phi(s) = (1 - 2s)^{-N_0 2^{m-2}}$$

so that

$$\mathbb{E}\{(z_m^2)^\ell\} = [\sigma_m^2(\theta_0)]^\ell \left. \frac{d^\ell \phi(s)}{ds^\ell} \right|_{s=0}.$$

Using this in (A.2) then gives

$$\mathbb{E}\{\tau_m^4\} = \frac{3N_0 2^m (N_0 2^m + 8)}{2^{4M} [\sigma_m^2(\theta_0)]^4} \left| \sum_{k=1}^3 \alpha_k \frac{d\sigma_m^2(\theta_0)}{d\theta_k} \right|^4 \quad (\text{A.3})$$

so that

$$\begin{aligned} \mathbb{E}\{|X_{m,M}|^3\} &\leq \kappa \sqrt{3} \frac{N_0 2^m \sqrt{N_0 2^m + 8}}{2^{3M} [\sigma_m^2(\theta_0)]^3} \\ &\quad \cdot \left| \sum_{k=1}^3 \alpha_k \frac{d\sigma_m^2(\theta_0)}{d\theta_k} \right|^3. \end{aligned}$$

Therefore, since $\sigma_m^2(\theta)$ is C^1 on compact Θ

$$\begin{aligned} \mathbb{E}\{|X_{m,M}|^3\} &< \infty \\ \lim_{M \rightarrow \infty} \sum_{m=1}^M \mathbb{E}\{|X_{m,M}|^3\} &= 0 \end{aligned}$$

so that by the definition of $X_{m,M}$, use of [7, Theorem 7.1.2] leads to (note that by Craig's Theorem, the $\{z_k\}$ are independent)

$$\sum_{m=1}^M X_{m,M} = \frac{1}{\sqrt{\varphi_M}} \sum_{k=1}^3 \alpha_k \frac{dQ_M(Z|\theta)}{d\theta_k} \Big|_{\theta=\theta_0} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } M \rightarrow \infty.$$

Also, for $m \neq \ell$ by (13) and (15)

$$\mathbb{E}\{\tau_m \tau_\ell\} = 0$$

so that

$$\begin{aligned} \varphi_M &= \mathbb{E} \left\{ \left(\sum_{m=1}^M \tau_m \right)^2 \right\} \\ &= \mathbb{E} \left\{ \left(\sum_{k=1}^3 \alpha_k \frac{dQ_M(Z|\theta_0)}{d\theta_k} \right)^2 \right\} \\ &= 2^{-M} \alpha^T P_M(\theta_0) \alpha \end{aligned}$$

where use has been made of the notation $\alpha^T \triangleq [\alpha_1, \alpha_2, \alpha_3]$ so that

$$\sqrt{\frac{2^M}{\alpha^T P_M(\theta_0) \alpha}} \alpha^T \frac{dQ_M(Z|\theta)}{d\theta} \Big|_{\theta=\theta_0} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } M \rightarrow \infty.$$

But α is arbitrary, so that use of the Cramér-Wold device completes the proof. \square

Lemma A.3: If $\gamma \in (0, 1)$ then for some $K < \infty$

$$\begin{aligned} \sup_{\theta \in \Theta} \mathbb{E}\{\|R_M(\theta) - \mathbb{E}\{R_M(\theta)\}\|^4\} &\leq \frac{K}{4^M} \\ \sup_{\theta \in \Theta} \mathbb{E}\{|Q_M(\theta) - \mathbb{E}\{Q_M(\theta)\}\|^4\} &\leq \frac{K}{4^M} \\ \sup_{\theta \in \Theta} \mathbb{E}\left\{ \left\| \frac{dQ_M(Z|\theta)}{d\theta} \right\|^4 \right\} &\leq \frac{K}{4^M}. \end{aligned}$$

Proof: From the proof of Lemma A.1

$$\begin{aligned} [R_M(\theta) - \mathbb{E}\{R_M(\theta)\}]_{j,k} &= \frac{1}{2^M} \sum_{m=1}^M [N_0 2^{m-1} \sigma_m^2(\theta_0) - z_m^2] \varphi_{m,j,k}(\theta). \end{aligned}$$

Lengthy but straightforward arithmetic shows that with the definition $[\Lambda_m]_{j,k} \triangleq \varphi_{m,j,k}(\theta)$

$$\|\Lambda_m\|^2 \leq \text{Trace}\{\Lambda_m\} \leq C m^2 \lambda^{2m}$$

where $C < \infty$. Therefore, using the Cauchy-Schwarz inequality, the fact that $\gamma \in (0, 1) \Rightarrow \lambda < 1$ and the fact that

uncorrelated Gaussian random variables are also independent gives

$$\begin{aligned} & \mathbb{E}\{\|R_M(\theta) - \mathbb{E}\{R_M(\theta)\}\|^4\} \\ & \leq \frac{C}{16^M} \sum_{m=1}^M \sum_{\ell=1}^M \mathbb{E}\{[N_0 2^{2m-1} \sigma_m^2(\theta_0) - z_m^2]^2 \\ & \quad \cdot [N_0 2^{\ell-1} \sigma_\ell^2(\theta_0) - z_\ell^2]^2\} \\ & = \frac{C}{16^M} \left(\sum_{m=1}^M \mathbb{E}\{[N_0 2^{2m-1} \sigma_m^2(\theta_0) - z_m^2]^2\} \right)^2 \\ & = \frac{C}{16^M} \left(\sum_{m=1}^M N_0 2^m \sigma_m^4(\theta_0) \right)^2 \leq \frac{K}{4^M}. \end{aligned}$$

Similarly, from the proof of Theorem 1 and using (A.3)

$$\begin{aligned} & \mathbb{E}\{|Q_M(\theta) - \mathbb{E}\{Q_M(\theta)\}\|^4\} \\ & = \frac{N_0}{16^M} \sum_{m=1}^M \mathbb{E}\left\{ \left(\frac{N_0 2^{2m-1} \sigma_m^2(\theta_0) - z_m^2}{\sigma_m^2(\theta)} \right)^4 \right\} \\ & = \frac{3N_0^2}{16^M} \sum_{m=1}^M 2^m (N_0 2^m + 8) \left(\frac{\sigma_m^2(\theta_0)}{\sigma_m^2(\theta)} \right)^4 \leq \frac{K}{4^M}. \end{aligned}$$

Finally, using the same methods as for the previous two cases

$$\begin{aligned} & \left\| \frac{dQ_M(Z\theta_0)}{d\theta} \right\|^2 \\ & = \frac{1}{4^M} \sum_{m=1}^M \sum_{\ell=1}^M \frac{1}{\sigma_m^2(\theta_0) \sigma_\ell^2(\theta_0)} \frac{d\sigma_\ell^2(\theta)}{d\theta^T} \frac{d\sigma_m^2(\theta)}{d\theta} \\ & \quad \cdot \left(N_0 2^{2m-1} - \frac{z_m^2}{\sigma_m^2(\theta_0)} \right) \left(N_0 2^{2\ell-1} - \frac{z_\ell^2}{\sigma_\ell^2(\theta_0)} \right) \end{aligned}$$

so that

$$\begin{aligned} & \mathbb{E}\left\{ \left\| \frac{dQ_M(Z\theta_0)}{d\theta} \right\|^4 \right\} \leq \frac{1}{16^M} \sum_{m=1}^M \frac{1}{[\sigma_m^2(\theta_0)]^4} \left\| \frac{d\sigma_m^2(\theta)}{d\theta} \right\|^4 \\ & \quad \cdot \mathbb{E}\left\{ \left(N_0 2^{2m-1} - \frac{z_m^2}{\sigma_m^2(\theta_0)} \right)^4 \right\} \\ & = \frac{3}{16^M} \sum_{m=1}^M \frac{1}{[\sigma_m^2(\theta_0)]^4} \left\| \frac{d\sigma_m^2(\theta)}{d\theta} \right\|^4 \\ & \quad \cdot N_0 2^m (N_0 2^m + 8) \leq \frac{K}{4^M}. \quad \square \end{aligned}$$

Lemma A.4: If $X \sim N_m(\mu, \Sigma)$ where Σ is nonsingular and $m = \dim\{X\}$, then

$$X^T \Sigma^{-1} X \sim \chi_m^2(\delta)$$

where $\chi_m^2(\delta)$ is a noncentral chi-squared distribution with noncentrality parameter

$$\delta = \mu^T \Sigma^{-1} \mu.$$

Proof: See [34].

Lemma A.5:

Distribution of Quadratic Forms: Suppose $Z \sim \mathcal{N}(0, P)$, $W \sim \mathcal{N}(0, Q)$ with $\mathbb{E}\{ZW^T\} = R$, and $A \in \mathbb{R}^{n \times n}$ $B \in$

$\mathbb{R}^{m \times m}$ are symmetric where $n = \dim Z$, $m = \dim W$. Then

$$\begin{aligned} \mathbb{E}\{(Z^T A Z)(W^T B W)\} & = (\text{Trace}\{AP\})(\text{Trace}\{BQ\}) \\ & \quad + 2 \text{Trace}\{RBRA\}. \end{aligned} \quad (\text{A.4})$$

Proof: See [34]. \square

REFERENCES

- [1] J. Barnes and D. Allan, "A statistical model of flicker noise," *Proc. IEEE*, vol. 54, pp. 176–178, 1966.
- [2] R. J. Barton and H. V. Poor, "Signal detection in fractional Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 34, pp. 943–959, 1988.
- [3] M. Basseville, A. Benveniste, K. Chou, S. Golden, R. Nikoukah, and A. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Inform. Theory*, vol. 38, pp. 766–784, 1992.
- [4] A. Benveniste, R. Nikoukah, and A. Willsky, "Multiscale system theory," *IEEE Trans. Circuits Syst. I*, vol. 41, pp. 2–15, 1994.
- [5] L. L. Cam, "Maximum likelihood: An introduction," *Int. Statist. Rev.*, vol. 58, pp. 153–171, 1990.
- [6] S. Cambanis, "Wavelet approximation of deterministic and random signals: Convergence properties and rates," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1013–1029, 1994.
- [7] K. Chung, *A Course in Probability Theory*. New York: Harcourt, Brace and World, 1968.
- [8] R. Dahlhaus, "Efficient parameter estimation for self similar processes," *Ann. Statist.*, vol. 17, pp. 1749–1766, 1989.
- [9] I. Daubechies, "Ten lectures on wavelets," in *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia, PA: Soc. Indust. Appl. Math., 1992.
- [10] M. Deriche and A. Tewfik, "Signal modeling with filtered discrete fractional noise processes," *IEEE Trans. Signal Processing*, vol. 41, pp. 2839–2849, 1993.
- [11] ———, "Maximum likelihood estimation of the parameters of discrete fractionally differenced Gaussian noise process," *IEEE Trans. Signal Processing*, vol. 41, pp. 2977–2989, 1993.
- [12] R. Dijkerman and R. Mazumdar, "On the correlation structure of the wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1609–1612, 1994.
- [13] K. Falconer, *Fractal Geometry*. New York: Wiley, 1987.
- [14] M. Farge, J. Hunt, and J. Vassilicos, "Wavelets, fractals and Fourier transforms," in *The Institute of Mathematics and its Applications Conference Series*. Oxford, U.K.: Clarendon, 1993.
- [15] P. Flandrin, "On the spectrum of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 35, pp. 197–199, 1989.
- [16] ———, "Wavelet analysis and synthesis of fractional Brownian motions," *IEEE Trans. Inform. Theory*, vol. 38, pp. 910–917, 1992.
- [17] R. Fox and M. S. Taqqu, "Large sample properties of parameter estimates for strongly dependent stationary time series," *Ann. Statist.*, vol. 14, pp. 517–532, 1986.
- [18] N. Gache, P. Flandrin, and D. Garreau, "Fractal dimension estimators for fractional Brownian motions," in *Proc. IEEE Int. Conf. on Acoustics Speech and Signal Processing*, 1991, pp. 3557–3560.
- [19] J. Geweke and S. Porter-Hudak, "The estimation and application of long memory time series models," *J. Time Series Anal.*, vol. 4, pp. 221–238, 1983.
- [20] A. L. Goldberger and B. West, "Fractals in physiology and medicine," *Yale J. Medicine and Biology*, vol. 60, pp. 421–435, 1987.
- [21] H. Hurst, "Long term storage capacity in reservoirs," *Trans. Amer. Soc. Civil Eng.*, vol. 116, pp. 770–799, 1951.
- [22] G. Jumarie, "Stochastic differential equations with fractional Brownian motion input," *Int. J. Syst. Sci.*, vol. 24, pp. 1113–1131, 1993.
- [23] R. Kashyap, "Estimation in long memory time series model," *J. Time Series Anal.*, vol. 9, pp. 35–41, 1988.
- [24] M. Keshner, " $1/f$ noise," *Proc. IEEE*, vol. 70, pp. 212–218, 1982.
- [25] E. Lehmann, *Theory of Point Estimation*. New York: Wiley, 1983.
- [26] J.-S. Leu and A. Papamarcou, "On estimating the spectral exponent of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 41, pp. 233–244, 1995.
- [27] M. Luetttgen and A. Willsky, "Multiscale smoothing error models," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 173–175, 1995.
- [28] T. Lundahl, W. Ohley, S. Kay, and R. Siffert, "Fractional Brownian motion: A maximum likelihood estimator and its application to image texture," *IEEE Trans. Med. Imag.*, vol. MI-5, pp. 152–161, 1986.
- [29] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet transform," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, 1989.

- [30] B. Mandelbrot and R. Voss, "Why is nature fractal and when should noises by scaling?," in *Noises in Physical Systems and 1/f Noise*. Amsterdam, The Netherlands: Elsevier, 1983, pp. 31–39.
- [31] B. B. Mandelbrot and J. W. V. Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Rev.*, vol. 10, pp. 422–437, 1968.
- [32] E. Masry, "The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 39, pp. 260–264, 1993.
- [33] A. McBride, "Fractional calculus and Integral transforms of generalised functions," in *Research Notes in Mathematics*, vol. 31. San Francisco: Pitman, 1979.
- [34] R. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
- [35] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. London, U.K.: McGraw-Hill, 1991.
- [36] V. Paxson, "Fast approximation of self-similar network traffic," Tech. Rep. LBL-36750, Lawrence Berkeley Lab., Apr. 1995.
- [37] J. Ramanathan and O. Zeitouni, "On the wavelet transform of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1156–1158, 1991.
- [38] M. J. Schervish, "Theory of statistics," in *Springer Series in Statistics*. Berlin, Germany: Springer-Verlag, 1995.
- [39] V. Solo, "Intrinsic random functions and the paradox of $1/f$ noise," *SIAM J. Appl. Math.*, vol. 52, pp. 270–291, 1992.
- [40] A. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 904–909, 1992.
- [41] R. F. Voss and J. Clarke, " $1/f$ noise in music and speech," *Nature*, vol. 258, pp. 317–318, 1975.
- [42] P. Whittle, "Estimation and information in stationary time series," *Arkiv För Matematik*, vol. 2, pp. 423–434, 1953.
- [43] G. Wornell, "A Karhunen–Loève-like expansion for $1/f$ processes," *IEEE Trans. Inform. Theory*, vol. 36, pp. 859–861, 1990.
- [44] ———, "Wavelet-based representations for the $1/f$ family of fractal processes," *Proc. IEEE*, vol. 81, pp. 1428–1450, 1993.
- [45] G. Wornell and A. Oppenheim, "Estimation of fractal signals from noise measurements using wavelets," *IEEE Trans. Signal Processing*, vol. 40, pp. 611–623, 1992.
- [46] G. W. Wornell, "Signal processing with fractals: A wavelet-based approach," in *Signal Processing Series*. Englewood Cliffs, NJ: Prentice-Hall, 1996.