Analysis of Wavelet Based Maximum Likelihood Estimation of $1/f$ Noise.

Brett Ninness*

Abstract

This paper provides a theoretical analysis of the properties of Wavelet based maximum likelihood estimation of the parameters describing $1/f^\gamma$ processes embedded in white noise. This analysis shows that such a 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.

Technical Report EE9351, Department of Electrical and Computer Engineering,
University of Newcastle,AUSTRALIA

1 Introduction

Recently there has been significant interest in 'long memory' [18] and 'multi–scale' stochastic processes [4, 3, 27] and their overlap with work on Fractals and Wavelet Analysis [13], particularly through the study of fractional Brownian motion (fBm) and fractional Gaussian noise (fGn) [15, 14, 33, 38, 41, 5, 11].

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 [24, 2, 28, 37, 42], 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], $|\widehat{x}_N(\omega)|^2$ is of the form $|\widehat{x}_N(\omega)|^2 \approx$

*This work was supported by the Australian Research Council and the Centre for Industrial Control Science, University of Newcastle, Callaghan 2308, Australia. This author is with the Department of Electrical and Computer Engineering, University of Newcastle, Australia and can be contacted at email:brett@tesla.newcastle.edu.au or FAX: +61 49 21 69 93

[1]The absolute value of the N point DFT $\widehat{x}_N(\omega)$ of $\{x_k\}$.

$\sigma_x^2|\omega|^{-\gamma}$ for some finite non–zero σ_x and γ. Such processes (which are sometimes also called 'flicker noise') have been empirically observed in a wide variety of physical processes such as [24, 31] 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 [19].

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 [26, 17] through to approximate and exact maximum likelihood estimation [7, 16, 26, 28] and direct measuring of fractal dimension of observed sample paths [17, 12, 14]. Aside from the maximum likelihood (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 surveyed) are then available.

Unfortunately the ML methods, although being able to cope with measurement noise, are computationally intensive to implement. An exception is the work of Wornell and Oppenheim [46, 45] where the 'whitening filter' property [45, 41, 11] of the Wavelet transform on a $1/f$ processes has been exploited to derive a computationally efficient maximum likelihood 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 [46, 45], 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 [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 [46, 28, 17].

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 estimating γ from noise corrupted observations by any means, maximum likelihood or not.

[2] Throughout 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

2 Modelling of $1/f$ Processes

There are a number of technical difficulties in the modelling 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 [18], although other authors reserve this for the case $\gamma > 2$) the spectrum is non–integrable 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 [34] 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 non–stationary Brownian motion process $B(t)$ and then considering its increments $\mathrm{d}B$ which lead to the required spectra. That is

$$B(t) = \sigma_B^2 \int_0^t \mathrm{d}B(\sigma)$$

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

$$\frac{\mathrm{d}B}{\mathrm{d}t} = \int_{-\infty}^{\infty} e^{j\omega t}\,\mathrm{d}\mu(\omega)$$

where the measure $\mathrm{d}\mu$ satisfies $\mathbf{E}\left\{|\mathrm{d}\mu(\omega)|^2\right\} = \sigma_B^2 \mathrm{d}\omega^2, \sigma_B^2 < \infty$ in which case $B(t)$, since it is the integral of $\dot{B}(t)$, should have a spectrum like $\sigma_B^2\omega^{-2n}$. Following this heuristic line of reasoning, integrating again

$$B_2(t) = \sigma_B^2 \int_0^t \int_0^\xi \mathrm{d}B(\sigma)\,\mathrm{d}\xi$$

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]

$$B_n(t) = \sigma_B^2 \int_0^t \int_0^{t_n} \cdots \int_0^{t_2} \mathrm{d}B(\sigma)\,\mathrm{d}t_2 \cdots \mathrm{d}t_n = \frac{\sigma_B^2}{\Gamma(n)} \int_0^t (t-\xi)^n\,\mathrm{d}B(\xi) \tag{1}$$

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 non–integral spectral exponent γ, then this can be modelled by using a so–called 'fractional integral' of $\mathrm{d}B(t)$ in which the exponent n in (1) is non–integral.

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 [32], by modifying (1) slightly to become

$$B_H(t) = \frac{\sigma_B^2}{\Gamma(H+1/2)} \left\{ \int_0^t |t-\xi|^{H-1/2}\,\mathrm{d}B(\xi) + \int_{-\infty}^0 \left(|t-\xi|^{H-1/2} - |\xi|^{H-1/2}\right)\mathrm{d}B(\xi) \right\}. \tag{2}$$

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 $1/f^\gamma$ spectrums 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 [15]

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

where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. This makes them a particular example of the multi–scale stochastic processes studied recently in [4, 3, 27]. In fact, fBm's are the only Gaussian process to display this self–similar property [32], 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 non–stationary process with covariance function

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

so that for $H > 1/2$ then $\mathbf{E}\left\{B_H(t)[B_H(t+\tau) - B_H(t)]\right\} > 0$ and vice versa for $H < 1/2$. The implication is that sample paths tend to be increasing or decreasing depending on H. This non–stationarity 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–localised transformations such as the Wigner–Ville [14] and Wavelet [14, 33] transforms, others make formal the heuristic motivation for fractional integrals outlined above [32]. All approaches arrive at the relationship between the spectral exponent γ and the Hurst parameter H as $\gamma = 2H + 1$. A particularly appealing interpretation is that of Solo [40] 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 a 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 [18, 9, 22, 23]

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

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

$$R(\tau) = \frac{\Gamma(1-d)\Gamma(\tau+d)}{\Gamma(d)\Gamma(\tau+1-d)}, \quad \phi(\omega) = \frac{\sigma^2}{2\pi(4\sin^2(\omega/2))^d}$$

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 [18] 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 [44, 45] 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.

3 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 realisation $\{x_k\}$ of a $1/f^\gamma$ process. For example, in [28] such an estimate provides a measure of image texture, while in [19] it serves as a cardiac health diagnostic instrument. For detailed surveys on available estimation methods the reader is referred to [26, 17]. The purpose of this section is to provide only a brief overview so that the results to be presented in §5 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 $\widehat{\gamma}$ of γ is then taken as the least–squares estimated slope of $I_N(\omega)$ when plotted on log–log axes; see figure 1 for an illustration of this method. The most fundamental result on the performance of this technique is that [40] it is a mean square consistent estimate if the $1/f$ process can be modelled as a fBm or fGn. More refined information is provided by Leu and Papamorcou [26] who show that the mean square rate of convergence with increasing observation length N is [3]

$$\mathrm{Var}\{\widehat{\gamma}\} = \mathrm{o}\left(\frac{\log^2 N}{N}\right) \qquad \text{as } N \to \infty.$$

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

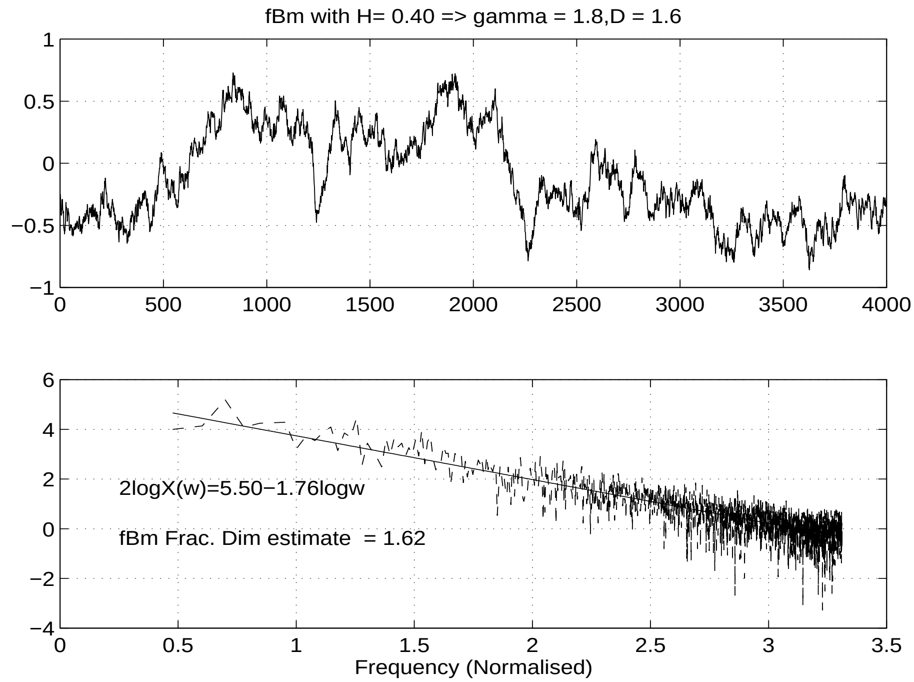

Figure 1: *Illustration of the use of linear regression of a periodogram plot to estimate the spectral exponent. Here the measured sample path is not noise corrupted, the true* $\gamma = 1.8$ *and the estimate is found as* $\widehat{\gamma} = 1.76$.

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 [18, 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 figure 2.

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 $\theta = [\sigma^2, \gamma]$ and $\phi(\omega, \theta) = \sigma^2 |\omega|^{-\gamma}$.

Two approaches have been analysed in the literature. Firstly, the methods of Whittle [43] can be used to approximate the log–likelihood function by using the periodogram so that an approximate Maximum Likelihood estimate $\widehat{\theta}$ is given by

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

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

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

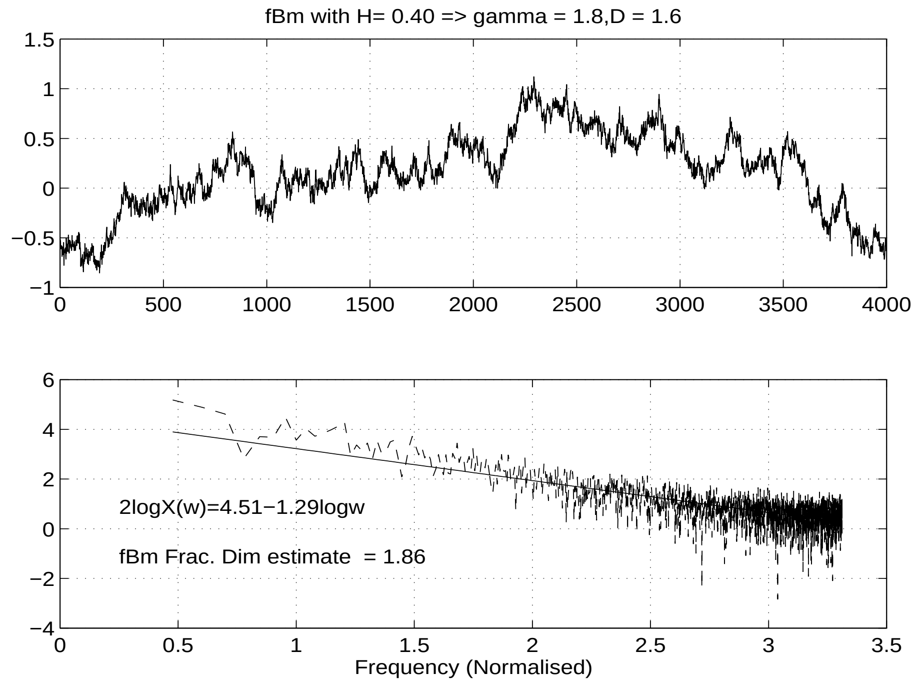

Figure 2: *Illustration of the use of linear regression of a periodogram plot to estimate the spectral exponent. Here the same measured sample path as the previous figure is used, but now it is corrupted by white Gaussian noise of variance* 0.001*. The true value is still* $\gamma = 1.8$ *but now the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

where

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

For $\gamma \in (0,1)$ Fox and Taqqu [16] 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

$$\mathrm{Var}\,\{\widehat{\gamma}\} = \mathrm{o}\left(\frac{1}{N}\right) \qquad \text{as } N \to \infty.$$

For $\gamma \in (0,1)$ Dalhaus [7] 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 [9, 10] that have been developed around the fractional differencing model (3).

In §5, the purpose of this paper will be to analyse 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 [45]). 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 §5 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 §5 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 [45] studied in §5 is a non–stationary one.

Note that other estimation methods based on ideas of 'Allen Variance' and measuring fractal dimension are also available [26, 17], but will not be commented on here.

4 Wavelet Transforms and $1/f$ Processes

Given the multi–scale nature of the fBm model of $1/f$ processes, and the multi–scale motivation of much of Wavelet theory [30, 8], it arises that Wavelets are a powerful analysis tool for studying $1/f$ processes [33, 5, 11, 15, 29].

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 [8]

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

A key feature of this transform [33, 15] is that if $X(t) = B_H(t)$, then even though $X(t)$ is non–stationary, its Wavelet transform $(\mathcal{W}X)(a,t)$ is stationary in t for fixed a and has spectral density $\phi(\omega)$ given by [4]

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

which provides yet another interpretation of the spectrum of a 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 [38].

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[5]

$$\sigma_a^2 \triangleq \mathbf{E}\left\{[(\mathcal{W}B_H)(a,t)]^2\right\} = a\int_{-\infty}^{\infty} \frac{|\widehat{\psi}(a\omega)|^2}{|\omega|^{2H+1}} \, \mathrm{d}\omega \tag{5}$$

so that for $a' \neq a$

$$\sigma_{a'}^2 = a' \int_{-\infty}^{\infty} \frac{|\widehat{\psi}(a'\omega)|^2}{|\omega|^{2H+1}} \, \mathrm{d}\omega = \left(\frac{a'}{a}\right)^{2H+1} \sigma_a^2 \tag{6}$$

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

[5] Modulo a scaling depending on the L_2 norm of $\psi(t)$, and also modulo σ_B^2 which affects the variance of the increments of $B_H(t)$.

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

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

of σ_a^2 can be taken for a range of $a = \{a_1, a_2, \cdots, a_M\}$ 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 $\widehat{\sigma}_{a_k}$ 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 'multi–scale' representation [8] 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 co–efficients $\{c_k^n\}$ may be efficiently calculated using a bank of filters with impulse responses $\{h_k\}$ and $\{g_k\}$ and with sub–sampled outputs

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

This method is commonly initialised 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' [8].

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

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

Wornell [44, 45] 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 co–efficients $\{c_n^m\}$ are uncorrelated across time and scale, are zero mean and have variance

$$\mathbf{E}\left\{(c_n^m)^2\right\} = \frac{\sigma^2}{2^{\gamma m}}, \tag{7}$$

then with $\{\psi(t)\}$ an orthonormal wavelet basis $X(t)$ possesses a 'time averaged' spectrum

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

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,

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

Note that it is known how to design Wavelets with an arbitrarily large number P of vanishing moments [8].

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 by suggested by the above synthesis result. This has been rigorously analysed for the specific case of fBm, where with $a = 2^n$, $\gamma = 2H+1$, (5) gives (7) for the variance of the Wavelet co–efficients and the work in [41] shows that both along and across scales the co–efficients are approximately uncorrelated in that for some $C < \infty$

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

provided that $|2^{-n}k - 2^{-m}\ell| > \max(2^{-n}\xi, 2^{-m}\xi)$. Similar results are reported in [11] so that for $P \gg 1$ the correlation decay along scales is rapid as the figures in [41] 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 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.

5 Wavelet Based Estimators for $1/f$ Processes

In [46] Wornell and Oppenheim proposed that this 'whitening' filter property of the Wavelet transform be exploited to derive an approximate Maximum–Likelihood estimator for $1/f$ processes that is much more computationally efficient than the ML methods surveyed in §3. Note, that as previously mentioned, this Wavelet based ML approach is also not equivalent to other ML approaches since it employs a non–stationary rather than a stationary model for the underlying $1/f$ process.

In [46], 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 $\mathbf{E}\left\{\nu_k^2\right\} = \sigma_\nu^2$ and $\{c_n^m\}$ are the Wavelet co–efficients 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 $\ell(R_M \mid \theta)$ for the data given the parameter vector θ is (modulo an additive constant):

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

Equivalently, the formulation suggested by Wornell and Oppenheim of

$$\ell(Z_M \mid \theta) \triangleq \sum_{m=1}^{M} N_0 2^{m-1}\log\sigma_m^2(\theta) + \frac{z_m^2}{\sigma_m^2(\theta)} \tag{8}$$

can be used where

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

Now, define the maximum likelihood estimate $\widehat{\theta}_M$ as

$$\widehat{\theta}_M = \arg\min_{\theta\in\Theta}\{Q_M(Z_M \mid \theta)\} \tag{9}$$

where

$$Q_M(Z_M \mid \theta) \triangleq \frac{1}{2^M}\ell(Z_M \mid \theta). \tag{10}$$

Wornell and Oppenheim [46] do not use the normalising factor of $1/2^M$ in (10) and claim in [46] 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 [46] 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 the method is never consistent unless the normalising factor $1/2^M$ is included in the log–likelihood cost as in (10) and that furthermore, 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 maximum likelihood estimators enjoy pleasant properties. As a simple example, consider observations $\{y_k\}$ of a scalar θ that is corrupted by additive uniformly distributed white noise $\{\nu_k\}$

$$y_k = \theta + \nu_k, \qquad \nu_k \sim \mathsf{U}(-\delta, \delta).$$

In this case, the probability density function for data $\{y_0, \cdots, y_{N-1}\}$ conditional upon θ is of the form

$$p(y_0, \cdots, y_{N-1} \mid \theta) = \begin{cases} C > 0 & ; \theta \in \{\theta : |y_k - \theta| \leq \delta \;\forall\; k \in [0, N)\} \\ 0 & ; \theta \notin \{\theta : |y_k - \theta| \leq \delta \;\forall\; k \in [0, N)\} \end{cases}$$

so that the maximum likelihood estimate of θ cannot even be uniquely defined, much less be consistent or distributed in some fashion. Other examples where the ML estimator can be defined, but is still inconsistent my be found in [39].

Following these comments, the convergence of the maximum likelihood estimate (9) needs to be rigorously examined. To begin with, it can be shown that the ML estimator is always convergent at least to a bounded set.

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

$$\lim_{M \to \infty} \widehat{\theta}_M \in \mathcal{S} \triangleq \left\{\theta \in \Theta : \overline{Q}(\theta) \leq \overline{Q}(\beta) \;\forall\; \beta \in \Theta\right\}$$

where

$$\overline{Q}(\theta) \triangleq \lim_{M \to \infty} \mathbf{E}\left\{Q_M(Z|\theta)\right\}$$

Proof. See Appendix A. □□□

It is here that the normalising factor $1/2^M$ included in (10) becomes vital since with it perusal of Appendix A shows that

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

which is obviously convergent to a well defined function $\overline{Q}(\theta)$ on Θ if $0 \notin \Theta$. Without the normalising factor $1/2^M$ then $\mathbf{E}\left\{Q_M(Z \mid \theta)\right\}$ (and hence by (A.5) $Q_M(Z \mid \theta)$ as well) diverges uniformly in θ with increasing M so that the ML estimator is not well defined asymptotically and hence cannot be convergent.

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

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

so that at least $\theta_0 \in \mathcal{S}$. However, for θ_0 to be the only element in the interior of $\mathcal{S}$ it will necessary that in the limit as $M \to \infty$ the Hessian of $\mathbf{E}\left\{Q_M(Z|\theta)\right\}$ be positive definite at θ_0. For finite M the Hessian is easily calculated as follows

Lemma 1

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta \mathrm{d}\theta^T} \mathbf{E}\left\{Q_M(Z_M \mid \theta_0)\right\} = \frac{N_0}{2^{M+2}} \sum_{m=1}^{M} \frac{2^m}{(\alpha\lambda^m + \sigma_\nu^2)^2} \begin{bmatrix} \lambda^{2m} & (\alpha/\lambda) m \lambda^{2m} & \lambda^m \\ (\alpha/\lambda) m \lambda^{2m} & (\alpha/\lambda)^2 m^2 \lambda^{2m} & (\alpha/\lambda) m \lambda^m \\ \lambda^m & (\alpha/\lambda) m \lambda^m & 1 \end{bmatrix}$$

Proof. See Appendix E □□□

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

$$\mathcal{I}_M(\theta_0) \triangleq \mathbf{E}\left\{\frac{\mathrm{d}^2}{\mathrm{d}\theta\mathrm{d}\theta^T}\ell(Z_M \mid \theta_0)\right\} = 2^M \frac{\mathrm{d}^2}{\mathrm{d}\theta\mathrm{d}\theta^T}\mathbf{E}\left\{Q_M(Z_M \mid \theta_0)\right\}.$$

Unfortunately, as the expression in lemma 1 shows, the entries pertaining to α and λ in the Hessian of $\mathbf{E}\{Q_M(Z \mid \theta)\}$ tend to zero as M increases if $2\lambda^2 < 1$ so that $\mathbf{E}\{Q_M(Z \mid \theta)\}$ '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 $[\alpha, \lambda, \sigma_\nu^2]$. Under the conditions that $\theta_0 \in \operatorname{int}\{\Theta\}$, $0 \notin \Theta$ and $\gamma \in (0,1)$ then*

$$\widehat{\theta}_M \xrightarrow{a.s} \theta_0 \quad \textit{as } M \to \infty. \tag{11}$$

If $\gamma > 1$ then $\widehat{\theta}_M$ 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 [46]. This arises since even though the bound applies only to unbiased estimates, many maximum likelihood 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 $1/2^M$ normalisation of the log–likelihood function and only for a narrower range of spectral exponents than originally suggested in [46]

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

$$\sqrt{2^M} P_M^{1/2}\left(\widehat{\theta}_M - \theta_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I) \quad \textit{as } M \to \infty \tag{12}$$

where

$$P_M = \mathbf{E}\left\{\frac{\mathrm{d}^2 Q_M(Z \mid \theta_0)}{\mathrm{d}\theta\mathrm{d}\theta^T}\right\}. \tag{13}$$

Proof. See Appendix B. □□□

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*

$$\lim_{M\to\infty} \sqrt{2^M}\mathsf{E}\left\{(\widehat{\theta}_M - \theta_0)(\widehat{\theta}_M - \theta_0)^T\right\} = \mathcal{I}_M^{-1}(\theta_0)$$

Proof. See Appendix C. □□□

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*

$$\mathrm{Var}\{\widehat{\alpha}\} = \mathsf{O}\left(\frac{1}{N^{1-\gamma}}\right) \quad \textit{as } N \to \infty \tag{14}$$

$$\mathrm{Var}\left\{\widehat{\lambda}\right\} = \mathsf{O}\left(\frac{1}{N^{1-\gamma}\log^2 N}\right) \quad \textit{as } N \to \infty \tag{15}$$

$$\mathrm{Var}\{\widehat{\sigma}_\nu^2\} = \mathsf{O}\left(\frac{1}{N}\right) \quad \textit{as } N \to \infty \tag{16}$$

Proof. See Appendix E. □□□

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 [28, 46, 45, 17] and is also in accordance with intuition.

To expand on this last point, 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 it's 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 characterise it; observing the spectrum at low frequencies requires observing more data than does observing the spectrum at high frequencies. Corollary 3 also provides further evidence, additional to the results of corollary 1, that the Wavelet based ML scheme is not convergent for $\gamma > 1$.

An illustration of this phenomenon of non–convergence can be provided by reviewing a simulation study in which the ML scheme using Daubechies order 10 wavelets was tested for $1/f$ processes with exponents of $\gamma = 0.6, 1.3$ and 1.8 and with $\alpha = 1000$. Realisations of length $N = 2^M$ with $M \in [8, 15]$ were used for estimation after being corrupted by white zero mean Gaussian noise of variance $\sigma_\nu^2 = 0.01$. The estimates were found using 500 iterations of the EM algorithm suggested in [46] from starting values of $\widehat{\alpha} = 500, \widehat{\lambda} = 1.0, \widehat{\sigma}_\nu^2 = 0.0001$. For each data length, 100 different noise realisations were obtained with estimation results being averaged over these 100 trials to obtain sample means and variances for the estimates. The results for $\widehat{\gamma}$ are shown in figure 3. Note

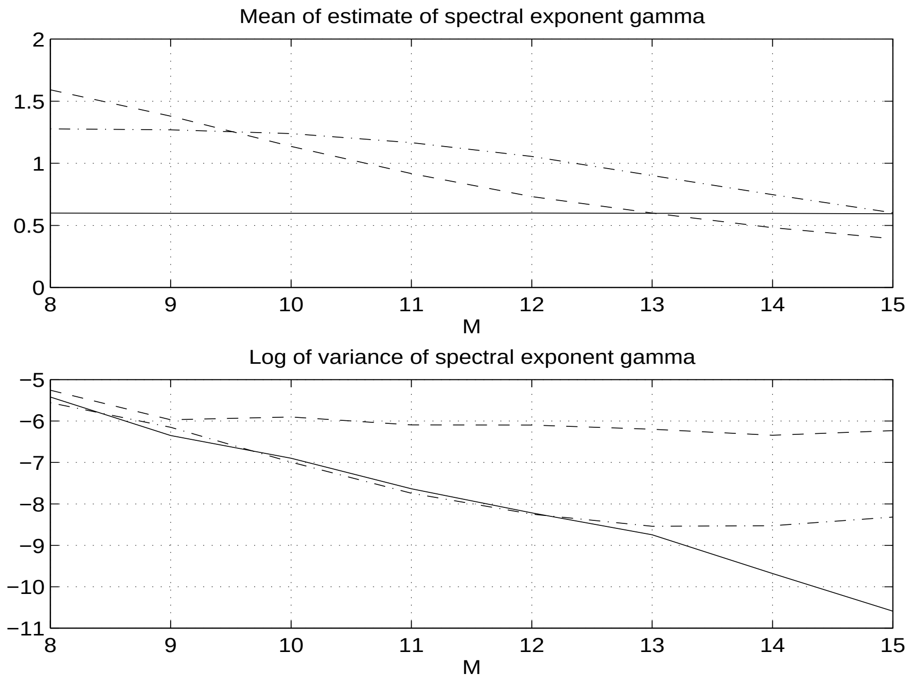

Figure 3: *Illustration of dependence of convergence of the Wavelet Based ML method on* $\gamma \in (0,1)$. *Solid line is mean and log of variance for* $\gamma = 0.6$*; estimator is convergent to correct value. Dash–dot and dashed lines are cases of* $\gamma = 1.3$ *and* $\gamma = 1.8$ *respectively;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

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 their true values.

6 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 A Proof of Strong Consistency

Proof. The idea of the proof is to first show that $Q_M(\theta) \to \mathbf{E}\{Q_M(\theta)\}$ as $M \to \infty$ 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:

$$\begin{aligned}
Q_M(Z \mid \theta) &\triangleq \frac{1}{2^M} \sum_{m=1}^{M} f(m, \theta, z_m), \\
f(m, \theta, z_m) &\triangleq N_0 2^{m-1} \log \sigma_m^2(\theta) + \frac{z_m^2}{\sigma_m^2(\theta)}, \\
\tau_m &\triangleq f(m, \theta, z_m) - \mathbf{E}\{f(m, \theta, z_m)\}, \\
&= \frac{z_m^2 - \mathbf{E}\{z_m^2\}}{\sigma_m^2(\theta)}, \\
S_M &\triangleq \sum_{m=1}^{M} \tau_m, \\
\mathcal{E}_M(\varepsilon) &\triangleq \left\{\omega \in \Omega : |S_M| > 2^M \varepsilon \quad ; \varepsilon > 0\right\}.
\end{aligned}$$

Then, by Chebychev's inequality

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

Now

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

and also

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

where

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

Therefore

$$\frac{1}{\sigma_m^2(\theta_0)} R^T R \sim \chi^2_{N_0 2^{m-1}}, \tag{A.1}$$

so by Lemma D.3

$$\mathbf{E}\left\{z_m^2\right\} = N_0 2^{m-1} \sigma_m^2(\theta_0). \tag{A.2}$$

Also, using Lemma D.4

$$\begin{aligned}
\mathbf{E}\left\{(z_m^2)^2\right\} &= \mathbf{E}\left\{\left(R^T R\right)^2\right\}, \\
&= \left[\sigma_m^2(\theta_0) \operatorname{Tr}\{I\}\right]^2 + 2\sigma_m^4(\theta_0) \operatorname{Tr}\{I\}, \\
&= \sigma_m^4(\theta_0) N_0 2^{m-1} \left[N_0 2^{m-1} + 2\right], \qquad \text{(A.3)}
\end{aligned}$$

so that

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

to give

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

Also

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

and

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

where

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

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

$$\mathbf{E}\left\{z_m^2 z_\ell^2\right\} = \mathrm{Tr}\left\{\sigma_m^2(\theta_0)I\right\}\mathrm{Tr}\left\{\sigma_\ell^2(\theta_0)I\right\} = \sigma_m^2(\theta_0)\sigma_\ell^2(\theta_0)N_0^2 2^{m+\ell-2}. \tag{A.4}$$

But

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

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

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

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 $\mathbf{E}\left\{S_M^2\right\} \leq \kappa' 2^M$ for some $\kappa' < \infty$. Therefore

$$\operatorname*{Sup}_{\theta\in\Theta}\left\{\sum_{M=1}^\infty \mathbf{P}\left\{\mathcal{E}_M(\varepsilon)\right\}\right\} \leq \sum_{M=1}^\infty \frac{1}{2^{2M}\varepsilon^2}\mathbf{E}\left\{S_M^2\right\} \leq \frac{\kappa'}{\varepsilon^2}\sum_{M=1}^\infty \frac{1}{2^M} < \infty,$$

so that by the Borel-Cantelli Lemma [6]

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

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

$$|Q_M(Z \mid \theta) - \mathbf{E}\left\{Q_M(Z \mid \theta)\right\}| < \varepsilon/2 \text{ a.s. } \forall M > M_1, \quad \forall\theta \in \Theta. \tag{A.5}$$

Therefore, for $M > M_1$

$$\mathsf{E}\left\{Q_M(Z \mid \theta)\right\} > Q_M(Z \mid \theta) - \varepsilon/2 \quad \text{a.s} \ \ \forall \theta \in \Theta.$$

However, by the definition of $\widehat{\theta}_M$, $Q_M(Z \mid \theta) \geq Q_M(Z \mid \widehat{\theta}_M) \ \ \forall \theta \in \Theta$. Therefore:

$$\mathsf{E}\left\{Q_M(Z \mid \theta)\right\} > Q_M(Z \mid \widehat{\theta}_M) - \varepsilon/2 \quad \text{a.s} \ \ \forall \theta \in \Theta. \tag{A.6}$$

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

$$\left|Q_M(Z \mid \widehat{\theta}_{M_n}) - Q_M(Z \mid \theta^\star)\right| < \varepsilon/4 \ \ \forall \theta \in \mathcal{D}.$$

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

$$\left|Q_{M_n}(Z \mid \theta^\star) - \mathsf{E}\left\{Q_{M_n}(Z \mid \theta^\star)\right\}\right| < \varepsilon/4.$$

Combining the previous two bounds with (A.6) and defining M_2 such that $M_n > M_2 \Rightarrow \widehat{\theta}_{M_n} \in \mathcal{D}$ means that $\forall M_n > \max(M_1, M_2, M_3)$

$$\mathsf{E}\left\{Q_{M_n}(Z \mid \theta)\right\} > \mathsf{E}\left\{Q_{M_n}(Z \mid \theta^\star)\right\} - \varepsilon.$$

Since ε is arbitrary this means that

$$\lim_{M_n \to \infty} \widehat{\theta}_{M_n} = \theta^\star \in \mathcal{S} \triangleq \left\{\theta \in \Theta : \overline{Q}(\theta) \leq \overline{Q}(\beta) \ \forall \ \beta \in \Theta\right\}$$

where

$$\overline{Q}(\theta) \triangleq \lim_{M \to \infty} \mathsf{E}\left\{Q_M(Z|\theta)\right\}$$

Noting that this applies for any convergent subsequence of $\widehat{\theta}_M$ then completes the proof. □□□

Appendix B Proof of Asymptotic Normality

Proof. By Theorem 1, for large enough M, $\widehat{\theta}_M$ 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, $Q_M(Z \mid \theta))$ is minimized at a $\widehat{\theta}_M \in \operatorname{int}\{\Theta\}$ so $\widehat{\theta}_M$ satisfies the necessary condition for minimisation of a differentiable function on an open domain:

$$\left.\frac{\mathrm{d}Q_M(Z \mid \theta)}{\mathrm{d}\theta}\right|_{\theta = \widehat{\theta}_M} = 0 \quad \text{a.s.}$$

Therefore, using the Mean Value Theorem, for large enough M $\exists \mu \in [0, 1]$ such that

$$\frac{\mathrm{d}Q_M(Z \mid \theta_0)}{\mathrm{d}\theta} = R_M(\beta)(\theta_0 - \widehat{\theta}_M) \quad \text{a.s.}$$

where

$$R_M(\beta) \triangleq \left.\frac{\mathrm{d}^2 Q_M(Z \mid \theta)}{\mathrm{d}\theta \mathrm{d}\theta^T}\right|_{\theta=\beta},$$
$$\beta \triangleq \mu\widehat{\theta}_M + (1-\mu)\theta_0.$$

Therefore, using Lemma D.2

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

But, by Lemma D.1

$$R_M(\beta) \xrightarrow{a.s.} P_M(\beta) \quad \text{Uniformly in } \beta \in \Theta, \tag{B.2}$$

and by Theorem 1

$$\widehat{\theta}_M \xrightarrow{a.s.} \theta_0 \quad \text{as } M \to \infty \tag{B.3}$$

so that $\beta \xrightarrow{a.s.} \theta_0$. Combining (B.1),(B.2) and (B.3) then gives

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

□□□

Appendix C Proof of Mean Square Convergence

Proof. From the proof of Theorem 1 in Appendix 1, for M large enough

$$\theta_0 - \widehat{\theta}_M = R_M^{-1}(\beta)\frac{\mathrm{d}Q_M(Z \mid \theta_0)}{\mathrm{d}\theta}$$

where $\beta = \mu\widehat{\theta}_M + (1-\mu)\theta_0$ 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 $\mathbf{E}\{R_M(\theta_0)\}$ 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:

$$\Omega_M^1 \triangleq \left\{\omega \in \Omega : \|\widehat{\theta}_M - \theta_0\| < \delta\right\},$$
$$\Omega_M^2 \triangleq \{\omega : \|R_M(\beta)\| \geq \varepsilon/2 \;\forall \beta \text{ such that } \|\beta - \theta_0\| < \delta\}$$

and put $\Omega_M = \Omega_M^1 \cap \Omega_M^2$. In this case

$$\mathbf{E}\left\{\|\sqrt{2^M}(\widehat{\theta}_M - \theta_0)\|^4\right\} \leq 4^M \mathbf{E}\left\{\|R_M^{-1}(\beta)\|^4 \left\|\frac{\mathrm{d}Q_M(Z \mid \theta_0)}{\mathrm{d}\theta}\right\|^4\right\}$$
$$\leq \frac{4^{M+2}}{\varepsilon^4}\int_{\Omega_M} \left\|\frac{\mathrm{d}Q_M(Z \mid \theta_0)}{\mathrm{d}\theta}\right\|^4 \mathrm{d}\mathbf{P} + 4^M \int_{\overline{\Omega}_M} \|\widehat{\theta}_M - \theta_0\|^4 \,\mathrm{d}\mathbf{P}.$$

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

$$\mathbf{E}\left\{\|\sqrt{2^M}(\widehat{\theta}_M - \theta_0)\|^4\right\} \leq C4^M \mathbf{E}\left\{\left\|\frac{\mathrm{d}Q_M(Z \mid \theta_0)}{\mathrm{d}\theta}\right\|^4\right\} + C4^M \mathbf{P}\{\overline{\Omega}_M\}. \tag{C.4}$$

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

$$\overline{\Omega}^2_M \subset \left\{\omega : \operatorname*{Sup}_{\theta\in\Theta} \|R_M(\theta) - \mathbf{E}\{R_M(\theta)\}\| > \varepsilon/2\right\}$$

and so by Lemma 2 and Chebychev's inequality $\exists K < \infty$ such that

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

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

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

Therefore

$$\overline{\Omega}^1_M \subset \left\{\omega : \operatorname*{Sup}_{\theta\in\Theta} |Q_M(\theta) - \mathbf{E}\{Q_M(\theta)\}| > \rho/2\right\}$$

so that again by using Lemma 2 and Chebychev's inequality

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

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

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

Substituting this into (C.4) together with the use of Lemma 2 gives that

$$\mathbf{E}\left\{\|\sqrt{2^M}(\widehat{\theta}_M - \theta_0)\|^4\right\} \leq C4^M \frac{K}{4^M} + C4^M \frac{K}{4^M} < \infty$$

so that combining Theorem 2 with Theorem 4.5.2 of [6] gives the result. □□□

Appendix D Lemmata required for Convergence Proofs

Lemma D.1

$$\frac{\mathrm{d}^2 Q_M(Z \mid \theta)}{\mathrm{d}\theta_j \mathrm{d}\theta_k} \xrightarrow{a.s.} \frac{\mathrm{d}^2 \mathbf{E}\{Q_M(Z \mid \theta)\}}{\mathrm{d}\theta_j \mathrm{d}\theta_k} \quad \textit{as } M \to \infty \text{ Uniformly in } \theta \in \Theta$$

Proof.

$$\begin{aligned}\frac{\mathrm{d}Q_M(Z \mid \theta)}{\mathrm{d}\theta_k} &= \frac{\mathrm{d}}{\mathrm{d}\theta_k}\left(\frac{1}{2^M}\sum_{m=1}^{M} N_0 2^{m-1}\log\sigma_m^2(\theta) + \frac{z_m^2}{\sigma_m^2(\theta)}\right), \\ &= \frac{1}{2^M}\sum_{m=1}^{M}\frac{1}{\sigma_m^2(\theta)}\frac{\mathrm{d}\sigma_m^2(\theta)}{\mathrm{d}\theta_k}\left(N_0 2^{m-1} - \frac{z_m^2}{\sigma_m^2(\theta)}\right).\end{aligned}$$

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

Therefore, defining

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

gives on use of (A.2)

$$\frac{\mathrm{d}^2Q_M(Z \mid \theta)}{\mathrm{d}\theta_j\mathrm{d}\theta_k} - \frac{\mathrm{d}^2\mathbf{E}\left\{Q_M(Z \mid \theta)\right\}}{\mathrm{d}\theta_j\mathrm{d}\theta_k} = \frac{1}{2^M}\sum_{m=1}^{M}\tau_m(\theta). \tag{D.1}$$

Now put

$$\begin{aligned}\varphi_{m,j,k}(\theta) &\triangleq \frac{1}{[\sigma_m^2(\theta)]^2}\left(\frac{\mathrm{d}^2\sigma_m^2(\theta)}{\mathrm{d}\theta_j\mathrm{d}\theta_k} - \frac{2}{\sigma_m^2(\theta)}\frac{\mathrm{d}\sigma_m^2(\theta)}{\mathrm{d}\theta_k}\frac{\mathrm{d}\sigma_m^2(\theta)}{\mathrm{d}\theta_j}\right), \\ S_M(\theta) &\triangleq \sum_{m=1}^{M}\tau_m(\theta) = \sum_{m=1}^{M}\varphi_{m,j,k}(\theta)\left(N_0 2^{m-1}\sigma_m^2(\theta_0) - z_m^2\right), \\ \mathcal{E}_m(\theta,\varepsilon) &\triangleq \left\{\omega \in \Omega : |S_M(\theta)| > 2^M\varepsilon\right\}.\end{aligned}$$

Then by Chebychev's inequality

$$\mathbf{P}\left\{\mathcal{E}_M(\theta)\right\} \le \frac{\mathbf{E}\left\{S_M^2(\theta)\right\}}{2^{2M}\varepsilon^2}.$$

But

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

Furthermore, using (A.2) and (A.4)

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

and by (A.2),(A.3)

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

so that

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

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 $|\varphi_{m,j,k}^2(\theta)\sigma_m^4(\theta)| < \kappa \quad \forall \theta \in \Theta$. Therefore

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

so that by the Borel-Cantelli Lemma [6]

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

Using (D.1) then gives the result. □□□

Lemma D.2

$$\sqrt{2^M} P_M^{-1/2}(\theta_0)\left.\frac{\mathrm{d}Q_M(Z \mid \theta)}{\mathrm{d}\theta}\right|_{\theta=\theta_0} \xrightarrow{\mathcal{D}} \mathcal{N}(0, I) \quad \textit{as } M \to \infty,$$

where

$$P_M(\theta_0) \triangleq \mathbf{E}\left\{\frac{\mathrm{d}Q_M(Z \mid \theta_0)}{\mathrm{d}\theta\mathrm{d}\theta^T}\right\}.$$

Proof. Define

$$\begin{aligned}
\tau_m &\triangleq \sum_{k=1}^{3}\alpha_k \left.\frac{\mathrm{d}Q_M(Z \mid \theta)}{\mathrm{d}\theta_k}\right|_{\theta=\theta_0} = \frac{1}{2^M\sigma_m^2(\theta_0)}\left(N_0 2^{m-1} - \frac{z_m^2}{\sigma_m^2(\theta_0)}\right)\sum_{k=1}^{3}\alpha_k\frac{\mathrm{d}\sigma_m^2(\theta_0)}{\mathrm{d}\theta_k}, \\
X_{m,M} &\triangleq \frac{\tau_m}{\sqrt{\varphi_M}}, \\
\varphi_M &\triangleq \sum_{m=1}^{M}\mathbf{E}\left\{\tau_m^2\right\},
\end{aligned}$$

where $\{\alpha_1, \alpha_2, \alpha_3\}$ are arbitrary real numbers. Then by (A.2),(A.3)

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

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

$$\mathbf{E}\left\{|X_{m,M}|^3\right\} \le \kappa \mathbf{E}\left\{|\tau_m|^3\right\} \le \kappa\sqrt{\mathbf{E}\left\{\tau_m^2\right\}}\sqrt{\mathbf{E}\left\{\tau_m^4\right\}}.$$

But by (A.2),(A.3) and the definition of $\tau_m(\theta)$

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

Also

$$\mathbf{E}\left\{\tau_m^4\right\} = \frac{1}{2^{4M}[\sigma_m^2(\theta_0)]^4}\left|\sum_{k=1}^{3}\alpha_k \frac{\mathrm{d}\sigma_m^2(\theta_0)}{\mathrm{d}\theta_k}\right|^4 \sum_{\ell=0}^{4}\left(\begin{array}{c}4\\ \ell\end{array}\right)(-1)^\ell N_0^\ell 2^{m\ell-\ell}\mathbf{E}\left\{\left(\frac{z_m^2}{\sigma_m^2(\theta_0)}\right)^{4-\ell}\right\}, \tag{D.2}$$

but by (A.1), $[\sigma_m^2(\theta_0)]^{-1}z_m^2 \sim \chi^2_{N_0 2^{m-1}}$ which has moment generating function $\phi(s)$ given by [36]

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

so that

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

Using this in (D.2) then gives

$$\mathbf{E}\left\{\tau_m^4\right\} = \frac{3N_0 2^m(N_0 2^m + 8)}{2^{4M}[\sigma_m^2(\theta_0)]^4}\left|\sum_{k=1}^{3}\alpha_k \frac{\mathrm{d}\sigma_m^2(\theta_0)}{\mathrm{d}\theta_k}\right|^4, \tag{D.3}$$

so that

$$\mathbf{E}\left\{|X_{m,M}|^3\right\} \le \kappa\sqrt{3}\frac{N_0 2^m\sqrt{N_0 2^m+8}}{2^{3M}[\sigma_m^2(\theta_0)]^3}\left|\sum_{k=1}^{3}\alpha_k \frac{\mathrm{d}\sigma_m^2(\theta_0)}{\mathrm{d}\theta_k}\right|^3.$$

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

$$\begin{aligned}\mathbf{E}\left\{|X_{m,M}|^3\right\} &< \infty,\\ \lim_{M\to\infty}\sum_{m=1}^{M}\mathbf{E}\left\{|X_{m,M}|^3\right\} &= 0\end{aligned}$$

so that by the definition of $X_{m,M}$,

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

Also, for $m \neq \ell$ by (A.2),(A.4)

$$\mathbf{E}\left\{\tau_m\tau_\ell\right\} = 0$$

so that

$$\varphi_M = \mathbf{E}\left\{\left(\sum_{m=1}^{M}\tau_m\right)^2\right\} = \mathbf{E}\left\{\left(\sum_{k=1}^{3}\alpha_k\frac{\mathrm{d}Q_M(Z \mid \theta_0)}{\mathrm{d}\theta_k}\right)^2\right\} = 2^{-M}\alpha^T P_M(\theta_0)\alpha$$

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

$$\sqrt{\frac{2^M}{\alpha^T P_M(\theta_0)\alpha}}\alpha^T \left.\frac{\mathrm{d}Q_M(Z \mid \theta)}{\mathrm{d}\theta}\right|_{\theta=\theta_0} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{as } M \to \infty.$$

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

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

$$\begin{aligned}
\mathbf{E}\left\{\|R_M(\theta) - \mathbf{E}\left\{R_M(\theta)\right\}\|^4\right\} &\leq \frac{K}{4^M}, \\
\mathbf{E}\left\{|Q_M(\omega) - \mathbf{E}\left\{Q_M(\omega)\right\}|^4\right\} &\leq \frac{K}{4^M}, \\
\mathbf{E}\left\{\left\|\frac{\mathrm{d}Q_M(Z \mid \theta_0)}{\mathrm{d}\theta}\right\|^4\right\} &\leq \frac{K}{4^M}.
\end{aligned}$$

Proof. From the proof of Lemma D.1

$$[R_M(\theta) - \mathbf{E}\left\{R_M(\theta)\right\}]_{j,k} = \frac{1}{2^M}\sum_{m=1}^{M}[N_0 2^{m-1}\sigma_m^2(\theta_0) - z_m^2]\varphi_{m,j,k}(\theta).$$

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

$$\|\Lambda_m\|^2 \leq \mathrm{Tr}\left\{\Lambda_m\right\} \leq Cm^2\lambda^{2m}$$

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

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

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

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

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

$$\left\|\frac{\mathrm{d}Q_M(Z \mid \theta_0)}{\mathrm{d}\theta}\right\|^2 = \frac{1}{4^M}\sum_{m=1}^{M}\sum_{\ell=1}^{M}\frac{1}{\sigma_m^2(\theta_0)\sigma_\ell^2(\theta_0)}\frac{\mathrm{d}\sigma_\ell^2(\theta)}{\mathrm{d}\theta^T}\frac{\mathrm{d}\sigma_m^2(\theta)}{\mathrm{d}\theta}\left(N_0 2^{m-1} - \frac{z_m^2}{\sigma_m^2(\theta_0)}\right)\left(N_0 2^{\ell-1} - \frac{z_\ell^2}{\sigma_\ell^2(\theta_0)}\right)$$

so that

$$\begin{aligned}
\mathbf{E}\left\{\left\|\frac{\mathrm{d}Q_M(Z \mid \theta_0)}{\mathrm{d}\theta}\right\|^4\right\} &\le \frac{1}{16^M}\sum_{m=1}^{M}\frac{1}{[\sigma_m^2(\theta_0)]^4}\left\|\frac{\mathrm{d}\sigma_m^2(\theta)}{\mathrm{d}\theta}\right\|^4\mathbf{E}\left\{\left(N_0 2^{m-1} - \frac{z_m^2}{\sigma_m^2(\theta_0)}\right)^4\right\} \\
&= \frac{3}{16^M}\sum_{m=1}^{M}\frac{1}{[\sigma_m^2(\theta_0)]^4}\left\|\frac{\mathrm{d}\sigma_m^2(\theta)}{\mathrm{d}\theta}\right\|^4 N_0 2^m(N_0 2^m + 8) \le \frac{K}{4^M}.
\end{aligned}$$

□□□

Lemma D.4 *If $X \sim N_m(\mu, \Sigma)$ where Σ is non-singular and $m = dim\{X\}$, then*

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

where $\chi_m^2(\delta)$ is a non-central chi squared distribution with non-centrality parameter

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

Proof. See [35] □□□

Lemma D.5 *Distribution of Quadratic Forms Suppose $Z \sim \mathcal{N}(0, P)$, $W \sim \mathcal{N}(0, Q)$ with $\mathbf{E}\{ZW^T\} = R$ and $A \in \mathbf{R}^{n\times n}$,$B \in \mathbf{R}^{m\times m}$ are symmetric where $n = \dim Z$, $m = \dim W$. Then:*

$$\mathbf{E}\left\{\left(Z^TAZ\right)\left(W^TBW\right)\right\} = (\mathrm{Tr}\{AP\})(\mathrm{Tr}\{BQ\}) + 2\,\mathrm{Tr}\{RBRA\} \qquad \text{(D.4)}$$

Proof.

$$\mathbf{E}\left\{\left(Z^TAZ\right)\left(W^TBW\right)\right\} = \mathrm{Tr}\left\{\mathbf{E}\left\{PZW^TQWZ^T\right\}\right\}$$

and

$$\left(PZW^T\right)_{k,j} = \sum_r\sum_n A_{k,n}z_n w_r\sum_m B_{r,m}w_m z_j$$

So using the formula for the 4th moments of a Multivariate Gaussian Distribution [20]

$$\begin{aligned}
\mathsf{E}\left\{\left(Z^T A Z\right)\left(W^T B W\right)\right\} &= \sum_k \sum_r \sum_n \sum_m A_{k,n} B_{r,m} \mathsf{E}\left\{z_n z_k w_r w_m\right\} \\
&= \sum_k \sum_r \sum_n \sum_m A_{k,n} B_{r,m} \left(R_{k,r} R_{n,m} + P_{n,k} Q_{r,m} + R_{n,r} R_{k,m}\right) \\
&= \sum_k \sum_r R_{k,r} \sum_n A_{k,n} \sum_m B_{r,m} R_{n,m} + \\
&\quad \sum_k \sum_r \sum_n A_{k,n} P_{n,k} \sum_m B_{r,m} Q_{r,m} + \\
&\quad \sum_k \sum_r \sum_n A_{k,n} R_{n,r} \sum_m B_{r,m} R_{k,m} \\
&= \mathrm{Tr}\left\{RBRA\right\} + \left(\mathrm{Tr}\left\{AP\right\}\right)\left(\mathrm{Tr}\left\{BQ\right\}\right) + \mathrm{Tr}\left\{ARBR\right\}
\end{aligned}$$

which is (D.4). □□□

Appendix E Cramér-Rao Bound Related Calculations

E.1 Proof of Cramér-Rao Bound.

Proof. From (8)

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

so that

$$\frac{\mathrm{d}}{\mathrm{d}\theta_k} \ell(Z_M \mid \theta) = -\frac{1}{2} \sum_{m=1}^{M} \frac{1}{\sigma_m^2(\theta)} \frac{\mathrm{d}\sigma_m^2(\theta)}{\mathrm{d}\theta_k} \left(N_0 2^{m-1} - \frac{z_m^2}{\sigma_m^2(\theta)} \right).$$

Therefore, using (A.2),(A.3) and (A.4)

$$\mathsf{E}\left\{ \frac{\mathrm{d}}{\mathrm{d}\theta_k} \ell(Z_M \mid \theta_0) \frac{\mathrm{d}}{\mathrm{d}\theta_j} \ell(Z_M \mid \theta_0) \right\} = \frac{N_0}{4} \sum_{m=1}^{M} \frac{2^m}{[\sigma_m^2(\theta_0)]^2} \frac{\mathrm{d}\sigma_m^2(\theta_0)}{\mathrm{d}\theta_k} \frac{\mathrm{d}\sigma_m^2(\theta_0)}{\mathrm{d}\theta_j}.$$

□□□

E.2 Proof of Asymptotic Variance Expressions

Proof. Define

$$\widehat{\beta}_M^T \triangleq \sqrt{2^M} \left[\lambda^M \widehat{\alpha}, M \lambda^M \left(\frac{\widehat{\lambda}}{\alpha} \right), \sigma_\nu^2 \right]. \tag{E.5}$$

Then by Theorem 2

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

and by the definition (E.5) $\widehat{\beta}_M = \mathcal{T}\widehat{\theta}_M$ where

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

so that use of Slutsky's Theorem gives (E.6) and also

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

Using the previous Lemma this information matrix can be evaluated as

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

Therefore, $\gamma \in (0,1)$ and hence $2\lambda^2 > 1$

$$\widehat{\beta}_M - \beta_0 \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}^{-1}(\beta_0)) \quad \text{as } M \to \infty \tag{E.6}$$

where $0 < \mathcal{I}(\beta_0) < \infty$ and

$$\mathcal{I}(\beta_0) = \lim_{M\to\infty}\mathcal{I}_M(\beta_0). \tag{E.7}$$

and hence using corollary 2

$$\lim_{M\to\infty}\mathbf{E}\left\{(\widehat{\beta}_M-\beta_0)(\widehat{\beta}_M-\beta_0)^T\right\} = \mathcal{I}_M^{-1}(\beta_0)$$

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. □□□

Appendix References

[1] J. Barnes and D. Allan, *A statistical model of flicker noise*, Proceedings of the IEEE, 54 (1966), pp. 176–178.

[2] R. J. Barton and H. V. Poor, *Signal detection in fractional Gaussian noise*, IEEE Transactions on Information Theory, IT-34 (1988), pp. 943–959.

[3] M. Basseville, A. Benveniste, K. Chou, S. Golden, R. Nikoukah, and A. Willsky, *Modeling and estimation of multiresolution stochastic processes*, IEEE Transactions on Information Theory, 38 (1992), pp. 766–784.

[4] A. Benveniste, R. Nikoukah, and A. Willsky, *Multiscale system theory*, IEEE Transactions on Circuits and Systems I, 41 (1994), pp. 2–15.

[5] S. Cambanis, *Wavelet approximation of deterministic and random signals: Convergence properties and rates*, IEEE Transactions on Information Theory, IT-40 (1994), pp. 1013–1029.

[6] K. Chung, *A Course in Probability Theory*, Harcourt, Brace and World Inc., 1968.

[7] R. Dahlhaus, *Efficient parameter estimation for self similar processes*, The Annals of Statistics, 17 (1989), pp. 1749–1766.

[8] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[9] M. deriche and A. Tewfik, *Signal modeling with filtered discrete fractional noise processes*, IEEE Transactions on Signal Processing, 41 (1993), pp. 2839–2849.

[10] M. Deriche and A. H. Tewfik, *Maximum likelihood estimation of the parameters of discrete fractionally differenced gaussian noise process*, IEEE Transactions on Signal Processing, 41 (1993), pp. 2977–2989.

[11] R. Dijkerman and R. Mazumdar, *On the correlation structure of the wavelet coefficients of fractional brownian motion*, IEEE Transactions on Information Theory, IT-40 (1994), pp. 1609–1612.

[12] K. Falconer, *Fractal Geometry*, John Wiley and Sons, 1987.

[13] M. Farge, J. Hunt, and J. Vassilicos, *Wavelets, Fractals and Fourier Transforms*, The Institute of Mathematics and its Applications Conference Series, Clarendon Press, Oxford, 1993.

[14] P. Flandrin, *On the spectrum of fractional Brownian motion*, IEEE Transactions on Informaton Theory, IT-35 (1989), pp. 197–199.

[15] ——, *Wavelet analysis and synthesis of fractional Brownian motions*, IEEE Transactions on Informaton Theory, IT-38 (1992), pp. 910–917.

[16] R. Fox and M. S. Taqqu, *Large sample properties of parameter estimates for strongly dependent stationary time series*, The Annals of Statistics, 14 (1986), pp. 517–532.

[17] N. Gache, P. Flandrin, and D. Garreau, *Fractal dimension estimators for fractional brownian motions*, in Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, 1991, pp. 3557–3560.

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

[19] A. L. Goldberger and B. West, *Fractals in physiology and medicine*, Yale Journal of Medicine and Biology, 60 (1987), pp. 421–435.

[20] G. Goodwin and R. Payne, *Dynamic System Identification*, Academic Press, 1977.

[21] H. Hurst, *Long term storage capacity in reservoirs*, Transactions of the American Society of Covil Engineers, 116 (1951), pp. 770–799.

[22] G. Jumarie, *Stochastic differential equations with fractional Brownian motion input*, International Journal of Systems Science, 24 (1993), pp. 1113–1131.

[23] R. Kashyap, *Estimation in long memory time series model*, Journal of Time Series Analysis, 9 (1988), pp. 35–41.

[24] M. Keshner, *$1/f$ noise*, Proceedings of the IEEE, 70 (1982), pp. 212–218.

[25] E. Lehmann, *Theory of Point Estimation*, John Wiley & Sons, 1983.

[26] J.-S. Leu and A. Papamarcou, *On estimating the spectral exponent of fractional brownian motion*, IEEE Transactions on Information Theory, 41 (1995), pp. 233–244.

[27] M. Luettgen and A. Willsky, *Multiscale smoothing error models*, IEEE Transactions on Automatic Control, 40 (1995), pp. 173–175.

[28] T. Lundahl, W. Ohley, S. Kay, and R. Siffert, *Fractional Brownian motion:A maximum likelihood estimator and its application to image texture*, IEEE Transactions on Medical Imaging, MI-5 (1986), pp. 152–161.

[29] S. Mallat, *A theory for multiresolution signal decomposition:The Wavelet transform*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (1989), pp. 674–693.

[30] S. G. Mallat, *A theory for multiresolution signal decomposition:The wavelet representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (1989), pp. 674–693.

[31] B. Mandelbrot and R. Voss, *Noises in Physical Systems and $1/f$ Noise*, Elsevier Science, 1983, ch. Why is Nature Fractal and when should noises by scaling?, pp. 31–39.

[32] B. B. Mandelbrot and J. W. V. Ness, *Fractional Brownian motions, fractional noises and applications*, SIAM Review, 10 (1968), pp. 422–437.

[33] E. Masry, *The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion*, IEEE Transactions on Information Theory, IT-39 (1993), pp. 260–264.

[34] A. McBride, *Fractional calculus and Integral transforms of generalised functions*, vol. 31 of Research Notes in Mathematics, Pitman, San Fransisco, 1979.

[35] R. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley and Sons, 1982.

[36] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, London, 1965.

[37] V. Paxson, *Fast approximation of self-similar network traffic*, Tech. Rep. LBL-36750, Lawrence Berkeley Laboratory, April 1995.

[38] J. Ramanathan and O. Zeitouni, *On the wavelet transform of fractional brownian motion*, IEEE Transactions on Information Theory, 37 (1991), pp. 1156–1158.

[39] M. J. Schervish, *Theory of Statistics*, Springer Series in Statistics, Springer Verlag, 1995.

[40] V. Solo, *Intrinsic random functions and the paradox of $1/f$ noise*, Siam Journal of Applied Mathematics, 52 (1992), pp. 270–291.

[41] A. Tewfik and M. Kim, *Correlation structure of the discrete wavelet coefficients of fractional Brownian motion*, IEEE Transactions on Information Theory, 38 (1992), pp. 904–909.

[42] R. F. Voss and J. Clarke, *$1/f$ Noise in music and speech*, Nature, 258 (1975), pp. 317–318.

[43] P. Whittle, *Estimation and information in stationary time series*, Arkiv För Matematik, 2 (1953), pp. 423–434.

[44] G. Wornell, *A Karhunen-Loève-like expansion for $1/f$ processes*, IEEE Transactions on Information Theory, IT-36 (1990), pp. 859–861.

[45] G. Wornell, *Wavelet-based representations for the $1/f$ family of fractal processes*, Proceedings of the IEEE, 81 (1993), pp. 1428–1450.

[46] G. Wornell and A. Oppenheim, *Estimation of fractal signals from noise measurements using wavelets*, IEEE Transactions on Signal Processing, 40 (1992), pp. 611–623.