USES OF CUMULANTS IN WAVELET ANALYSIS

DAVID R. BRILLINGER

Department of Statistics, University of California

Berkeley, CA 94720-3860

Cumulants are useful in studying nonlinear phenomena and in developing (approximate) statistical properties of quantities computed from random process data. Wavelet analysis is a powerful tool for the approximation and estimation of curves and surfaces. This work considers both wavelets and cumulants, developing some sampling properties of linear wavelet fits to a signal in the presence of additive stationary noise via the calculus of cumulants. Of some concern is the construction of approximate confidence bounds around a fit. Some extensions to spatial processes, irregularly observed processes and long memory processes are indicated.

KEYWORDS: long memory, point process, spatial process, time series, wavelet estimate.

1. INTRODUCTION

Wavelets are a contemporary tool for function approximation and mean level estimation. They are competitors/collaborators with traditional Fourier analysis, with other orthogonal function expansions, with linear and nonlinear regression estimates and with kernel estimates. In particular they are useful for handling localized behavior, discontinuities, and scale and shift transformations. In the time series case they have the ability to pick up transient behavior. In particular Donoho [18] records,

Mallat's Heuristic: "Bases of smooth wavelets are the best bases for representing objects composed of singularities, when there may be an arbitrary number of singularities, which may be located in all possible spatial positions."

For example the case with piecewise continuous mean level of a time series falls into this domain. This present work was motivated in part by examples in Brillinger [10] concerning the possible existence of jump discontinuities in the mean level function of a time series.

Wavelet estimates may be linear in the data available, however a breakthrough occurred when the concept of shrinkage was introduced, a breakthrough in the sense that asymptotically efficient estimates are realized, see Donoho et al. [22]. In this procedure the estimated coefficients of the expansion are moved closer to 0. Shrinkage estimates are discussed in Section 8, but not investigated in any detail in this present paper.

The focus of the paper is the case where an additive error is stationary and mixing. The work begins with some mention of existing procedures for estimating mean level functions of time series, then presents some pertinent properties of cumulants. The usefulness of the cumulants lies in their ability to elicit basic statistical properties of estimates quite directly. Linear wavelet estimates are indicated and illustrated in practice

---

AMS SUBJECT CLASSIFICATION: Primary 62M10; Secondary 62G99.

This research was partially supported by the ONR Grant N00014-94-1-0042.

with some microtubule movement data. Large sample distributions are developed for the linear case. The large sample distribution presented allows the construction of approximate confidence intervals for example.

Intentions of this paper are to illustrate that pertinent time series results are available to study wavelet estimates and that there are direct extensions to irregularly observed processes, spatial processes, long memory processes and to both continuous and discrete time. It may be remarked that the proofs required are analagous to large sample results for tapered Fourier transforms of stationary processes, see eg. Brillinger [8].

The focus of the work is on the simplest form of wavelet analysis, namely orthonormal expansions based on a single function in the case of the line. Here there are functions $\phi(.)$, $\psi(.)$ such that the functions

$$2^l \phi(2^l x - k), \quad 2^j \psi(2^j x - k)$$

provide a complete orthonormal basis for $L_2(R)$. The function $\phi(.)$ is called the scaling function and $\psi(.)$ the mother function. There are other wavelet analyses based on biorthogonal functions, based on discrete domains and based on functions constructed to handle finite regions, discontinuities and boundaries.

The substantive example presented is based on Haar wavelets and has in mind the detection of step discontinuities. Haar wavelet expansions have the advantage of not being subject to the Gibbs phenomenon at jump discontinuities, see Kelly [30].

1.1 Estimating Mean Level Functions

Consider the model

$$Y(t) = S(t) + \mathrm{E}(t) \tag{1.1}$$

$t = 0, \pm 1, \pm 2, \cdots$ with $S(.)$ a deterministic signal and $\mathrm{E}(.)$ a zero mean stationary noise. Hence $E\{Y(t)\} = S(t)$ is the mean level of the series $Y(.)$ at time t and the problem is to estimate this. Quite a variety of different procedures have been proposed for estimating $S(t)$ given data $Y(t)$, $t = 0, ..., T-1$. These methods can be linear or nonlinear and parametric or nonparametric. One might have a finite parameter linear model, such as

$$E\{\, Y(t)\} = S(t \,|\, \alpha) = \alpha_1 g_1(t) + \cdots + \alpha_J g_J(t) \tag{1.2}$$

with J known and the $g_1(.), ..., g_J(.)$ given functions. In a nonlinear regression formulation one would write $S(t) = S(t \,|\, \theta)$ with θ a finite dimensional parameter to be estimated. In the case that the mean function $S(t)$ is smooth, writing

$$S(t) = h(t/T)$$

suggests that as an alternative one can consider the estimation of $h(x)$ by a kernel smoother such as

$$\hat{h}(x) = \sum_t Y(t)\, w_U(x - \frac{t}{T}) \,/\, \sum_t w_U(x - \frac{t}{T}) \tag{1.3}$$

with $w_U(.)$ a kernel function and U a binwidth parameter. Härdle and Tuan [27] present results including robust procedures. The problem of estimating U is considered in Chiu [11], Hart [28-29], Altman [0]. An optimal U is determined in Truong [45].

As will be seen, wavelets provide another technique for the estimation of mean level functions such as $h(.)$. Their use has been illustrated in Antoniadis [1], Antoniadis et al. [2], Brillinger [10], Donoho [17].

2. CUMULANTS

The multilinear and dependence description properties of cumulants allow analytic derivation of various characteristics of empirical wavelets, particularly sample distributions. Some pertinent properties are now reviewed.

Given an r vector-valued random variable $(Y_1, \cdots, Y_r)$ the multilinear property is described by

$$cum\{a_1 Y_1, \ldots, a_r Y_r\} = a_1 \cdots a_r\, cum\{Y_1, \ldots, Y_r\} \tag{2.1}$$

$$cum\{Y_1 + Z_1, \ldots, Y_r\} = cum\{Y_1, Y_2, \ldots, Y_r\} + cum\{Z_1, Y_2, \ldots, Y_r\} \tag{2.2}$$

with $(Z_1, Y_1, \cdots, Y_r)$ an $r+1$ variate and cum symmetric in its arguments.

The dependence description property is based on the result that if any subset of the Y's is independent of those remaining, then

$$cum\{Y_1, \ldots, Y_r\} = 0 \tag{2.3}$$

This property is useful for formalizing mixing conditions in process cases.

In considering these results it is to be remembered that some of the variates may be identical, eg. $var\{Y\} = cov\{Y, Y\}$. Cumulants have the further property of measuring degree of nonnormality since the normal cumulants of order greater than 2 vanish. Further there are rules for developing joint cumulants of polynomial functions of basic variates. The above properties are discussed in Brillinger [8].

Cumulants provide easy proofs of asymptotic normality. For example, suppose that $Y_1, Y_2, \cdots$ are independent and identically distributed with $E\{Y\} = 0$ and $var\{Y\} = 1$. Suppose all moments of Y exist. Consider

$$S_n = (Y_1 + \cdots + Y_n)/\sqrt{n}$$

then following (2.1) and (2.2)

$$cum_k\{S_n\} = n\ cum_k\{Y\} / n^{k/2}$$

which tends to 0 for $k > 2$ as n tends to infinity. The normal is determined by its moments, in consequence S_n has a limiting normal distribution. There are improved approximations, Edgeworth expansions, based on higher-order cumulants, see Barndorff-Nielsen and Cox [3].

Error bounds may be given for the degree of approximation of the distribution of a random variable by a normal, through bounds on the cumulants. For example Rudzkis et al. [41] develop the following result. Consider a variate Y with mean 0 and variance 1. Suppose that

$$|cum_k\{Y\}| \le \frac{H(k!)^{1+\nu}}{\Delta^{k-2}}$$

for some $\nu \ge 0$, $H \ge 1$, then in the interval $0 \le u \le \delta/H$

$$\sup_u |Prob\{Y < u\} - \Phi(u)| \le 18H/\delta$$

where

$$\delta = \frac{1}{7}\left(\frac{\sqrt{2}\Delta}{6}\right)^{1/(1+2\nu)}$$

Statulevicius [42] provides the large deviation result that if

$$|cum_k\{Y\}| \le (k-1)!H \,/\, \Delta^{k-2}$$

then

$$Prob\{\,|Y| \ge u\,\} \le 2\exp\{-u^2\Delta / 2(H\Delta + u)\}$$

Such results are useful in deriving the type of results presented in this paper. A particular thing to note is that the results are expressed via conditions on cumulants.

Cumulants, specifically factorial cumulants, are also useful in studying integer-valued random variables and developing Poisson approximations, see Brillinger [9], Statulevicius and Aleskeviciene [43].

3. WAVELETS

3.1 Introduction

Wavelet analyses correspond to particular types of series expansions. In one simple construction, with $x \,\varepsilon\, R$, there is a single scaling function $\phi(.)$ and a mother wavelet $\psi(.)$ given by

$$\psi(x) = \sum_{k=0}^{2N-1} (-1)^k c_{-k+1}\phi(2x-k) \tag{3.1}$$

for some particular coefficients c_k. The integer N relates to the chosen regularity of the wavelets. If $\phi(.)$ has support $[0,2N-1]$, then that of $\psi(.)$ is $[1-N,N]$. This is the construction of Daubechies [13]. Pertinent coefficients, c_k, are listed in Daubechies [14]. The functions

$$\phi_{lk}(x) = 2^{l/2}\phi(2^l x-k) \tag{3.2}$$

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x-k) \tag{3.3}$$

are such that

$$\{\ \phi_{lk}(x) \ \text{ and } \ \psi_{jk}(x), \quad j = l, l+1, \cdots \ \ k = 0, \pm1, \pm2, \ \cdots\ \}$$

provide an orthonormal basis for $L_2(R)$ for any integer l. A square-integrable function $h(x)$ can thus be written as

$$h(x) = \sum_{k=-\infty}^{\infty} \alpha_{lk}\phi_{lk}(x) + \sum_{j=l}^{\infty}\sum_{k=-\infty}^{\infty} \beta_{jk}\psi_{jk}(x) \tag{3.4}$$

for almost all x with

$$\alpha_{lk} = \int \phi_{lk}(x)h(x)dx \tag{3.5}$$

$$\beta_{jk} = \int \psi_{jk}(x)h(x)dx \tag{3.6}$$

and

$$\sum_{k=-\infty}^{\infty} \alpha_{lk}^2 + \sum_{j=l}^{\infty}\sum_{k=-\infty}^{\infty} \beta_{jk}^2 < \infty \tag{3.7}$$

In the case that $\phi(.)$ has compact support, many of the coefficients in (3.4) will be 0. Expression (3.4) is referred to as a scaling expansion. The pair (3.5), (3.6) are called a wavelet transform of $h(.)$. The presence of the 2^j factor in (3.2), (3.3) is what leads to the variable scale character and k is what leads to the translation character, that wavelet approximations are noted for. When $\phi(.)$ and $\psi(.)$ have compact or near compact support, the effects of the individual $\phi_{lk}(.)$, $\psi_{jk}(.)$ terms in (3.3) are localized in x and this is one of the advantages of the wavelet approach. One would pick a scaling function $\phi(.)$ having in mind a desire that the high order coefficients in the expansion (3.4) are small.

One notes that the family $\phi_{lk}^U(x) = \sqrt{U}\,\phi_{lk}(Ux)$, $\psi_{jk}^U(x) = \sqrt{U}\,\psi_{jk}(Ux)$ is also orthonormal and complete. These will be used later in the paper. A parameter like U was introduced by Hall and Patil [24] to facilate the study of large sample properties of wavelet estimates. General references to wavelet analysis are Daubechies [14], Walter [47-49], Meyer [36], Strichartz [44], Benedetto and Frazier [4].

One particular example, and the one employed in the computations of Section 4, is the Haar case where

$$\phi(x) = \begin{cases} 1 & 0 \le x < 1 \\ 0 & otherwise \end{cases}$$

and

$$\psi(x) = \begin{cases} 1 & 0 \le x < 1/2 \\ -1 & 1/2 \le x < 1 \\ 0 & otherwise \end{cases} \tag{3.8}$$

Here N of (3.1) is 1. In this case the expansion (3.4) can be anticipated to be particularly appropriate when $h(.)$ is piecewise constant, as when $E\{Y(t)\}$ is constant at one level for a while and later constant at other levels.

In practice, equivalent expansions like the following, with J a large integer $\ge l+1$ and $l \le J-1$ are employed.

$$\begin{aligned} h_J(x) &= \sum_k \alpha_{Jk}\,\phi_{Jk}(x) \\ &= \sum_{j=-\infty}^{J-1} \sum_k \beta_{jk}\,\psi_{jk}(x) \\ &= \sum_k \alpha_{lk}\,\phi_{lk}(x) + \sum_{j=l}^{J-1} \sum_k \beta_{jk}\,\psi_{jk}(x) \end{aligned}$$

These expansions contain a finite number of items, at each x, in the case that $\phi(.)$ has compact support.

One may deal with convergences other than that of L_2. In what follows, concern will be with the pointwise case. One reference concerning the mathematics of pointwise convergence of wavelet expansions, like those above, as $J \to \infty$ is Kelly et al. [31].

The discusion so far has referred to the cases of R. On occasion a restricted domain, such as [0,1], will be of principal interest. The Haar wavelets actually provide an orthonormal basis for $L_2[0,1]$, but are special in that sense. The ψ_{jk} generated by other ϕ are not generally orthonormal for [0,1]. Researchers have constructed wavelet like orthonormal bases for $L_2[0,1]$, but additional functions have to be introduced and one needs to orthogonalize eg. by Gram-Schmidt, and so loses the simple form, see Daubechies [15], Cohen et al. [12]. The difficulty is with the boundaries. The work of this paper will persist with the functions of the form (3.2), (3.3) in order to better understand this particular situation.

Given data $Y(t), t = 0, \ldots, T-1$ and $T = 2^p$ there are fast discrete wavelet transforms taking the form

$$y_{(j,k)} = \sum_{t=0}^{T-1} w_{(j,k)t}\,Y(t) \tag{3.9}$$

with

$$w_{(j,k)t} \approx \psi_{jk}\left(\frac{t}{T}\right)$$

for $k = 0,...,2^j-1$, $j = 0,...,p-1$, see Donoho et al. [22] .

In practice the spatial case is of great importance, eg. in image processing. Suppose $\mathbf{x} \varepsilon R^p$. Now several scaling functions are involved. Wavelet theory indicates the existence of I functions $\psi^{(i)}(\mathbf{x})$, $i = 1, ..., I$ such that

$$\psi_{j\mathbf{k}}^{(i)}(\mathbf{x}) = 2^{jp/2}\psi^{(i)}(2^j\mathbf{x} - \mathbf{k})$$

for $j \varepsilon Z$, $\mathbf{k} \varepsilon Z^p$, $i = 1, \cdots, I$ provide an orthonormal basis for $L_2(R^p)$. The existence of such functions is discussed in eg. Meyer [36], Daubechies [14], Benedetto and Frazier [4]. Square integrable $h(\mathbf{x})$ is now written as

$$h(\mathbf{x}) = \sum_{i=1}^{I}\sum_{j=-\infty}^{\infty}\sum_{\mathbf{k}} \beta_{j\mathbf{k}}^{(i)}\psi_{j\mathbf{k}}^{(i)}(\mathbf{x}) \qquad (3.10)$$

for $\mathbf{x} \varepsilon R^p$ with

$$\beta_{j\mathbf{k}}^{(i)} = \int \psi_{j\mathbf{k}}^{(i)}(\mathbf{x})h(\mathbf{x})d\mathbf{x}$$

Expression (3.10) is a so-called homogeneous expansion in contrast to the scaling expansion of (3.4). One means to construct such a basis is via the tensor product method. For example when $p = 2$, then $I = 3$ and one sets

$$\psi_{j\mathbf{k}}^{(1)}(\mathbf{x}) = 2^j\,\phi(2^jx_1-k_1)\psi(2^jx_2-k_2)$$

$$\psi_{j\mathbf{k}}^{(2)}(\mathbf{x}) = 2^j\,\psi(2^jx_1-k_1)\phi(2^jx_2-k_2)$$

$$\psi_{j\mathbf{k}}^{(3)}(\mathbf{x}) = 2^j\,\psi(2^jx_1-k_1)\psi(2^jx_2-k_2)$$

with $\phi(.)$ and $\psi(.)$ scaling and wavelet functions on R. The two dimensional Haar system involves functions constant on rectangles.

3.2 A Statistical Setup

Consider the model

$$Y(t) = S(t) + \mathrm{E}(t) \qquad (3.11)$$

$t = 0, \pm 1, \pm 2, \cdots$ with $S(.)$ deterministic and $\mathrm{E}(.)$ zero mean stationary noise. Suppose that the data $Y(t)$, $t = 0, \pm 1, \cdots, \pm T$ are available, with T large. Suppose further

$$S(t) = h(t/T) \qquad (3.12)$$

for some measurable bounded $h(.)$ on $[-1,1]$.

Suppose that $h(.)$ is expanded using the complete orthonormal family $\{\phi_{lk}^U(x) = \sqrt{U}\,\phi_{lk}(Ux),\ \psi_{jk}^U(x) = \sqrt{U}\,\psi_{jk}(Ux),\ l = l+1,...,\ k = 0,\pm 1,+-2,...\}$. Then

$$h(x) = \sum_k \alpha_{lk}^U\phi_{lk}^U(x) + \sum_{j=l}^{\infty}\sum_k \beta_{jk}^U\psi_{jk}^U(x)$$

almost everywhere with

$$\alpha_{lk}^U = \int \phi_{lk}^U(x)h(x)dx \qquad (3.13)$$

$$\beta_{jk}^U = \int \psi_{jk}^U(x) h(x) dx \tag{3.14}$$

If $U = 2^n$, then $\alpha_{lk}^U = \alpha_{l+n,k}$, $\phi_{lk}^U(x) = \phi_{l+n,k}(x)$ with similar shifts for β and ψ. This identification is useful in simplifying some developments.

Expressions (3.13), (3.14) suggest computing the statistics

$$\hat{\alpha}_{lk}^U = \frac{1}{T} \sum_{t=-T}^{T} \phi_{lk}^U(\frac{t}{T}) Y(t) \tag{3.15}$$

$$\hat{\beta}_{jk}^U = \frac{1}{T} \sum_{t=-T}^{T} \psi_{jk}^U(\frac{t}{T}) Y(t) \tag{3.16}$$

and as an estimate of $h(x)$

$$\hat{h}(x) = \sum_k \hat{\alpha}_{lk}^U \, \phi_{lk}^U(x) + \sum_{j=l}^{J-1} \sum_k \hat{\beta}_{jk}^U \, \psi_{jk}^U(x) \tag{3.17}$$

for some J. The estimate (3.17) may also be written

$$\sum_k \hat{\alpha}_{Jk}^U \phi_{Jk}^U(x) \quad or \quad \sum_{j=-\infty}^{J-1} \sum_k \hat{\beta}_{jk}^U \, \psi_{jk}^U(x) \tag{3.18}$$

In the case $U = 2^n$, these become

$$\sum_k \hat{\alpha}_{J+n} \, \phi_{J+n}(x) \quad \text{and} \quad \sum_{j=-\infty}^{J+n-1} \sum_k \hat{\beta}_{jk} \, \psi_{jk}(x)$$

respectively. For a given x, the number of terms involved in these various expressions is finite, so there are not difficulties with convergence.

Expressions (3.15) and (3.16) constitute an empirical wavelet transform. Supposing that $U = 2^n$ and that the observation domain is $t = 0,...,T-1$ then (3.16) has the approximate form of (3.9), with the coefficients displaced by n. The transform can therefore be (approximately) computed using available discrete wavelet algorithms.

For Haar wavelets the function (3.7) will be piecewise constant, specifically the estimate $\hat{h}(x)$ is the mean of $T/2^J U$ of the $Y(t)$ values around the time point $T([2^J Ux]+.5)/2^J U$, with [.] here indicating integral part.

3.3 Properties of the Statistics

The statistics (3.15), (3.16), (3.17) are linear in the Y's, hence certain sampling properties, eg. large sample variances, cumulants and distributions are directly available. Some assumptions and consequent results follow.

ASSUMPTION 1. The function $h(.)$ is bounded and of bounded variation on $[-1,1]$ and vanishes outside that interval.

The assumption of bounded variation leads to error bounds when certain sums are approximated by integrals.

ASSUMPTION 2. The scaling function $\phi(.)$ is bounded and of bounded variation on $[0,2N-1]$ with N an integer and satisfies (3.1). For given l the collection $\{\phi_{lk}(.), \psi_{jk}(.), j = l, l+1,..., k = 0, \pm 1, \pm 2, \cdots\}$ of (3.1)-(3.3) provides a complete orthonormal basis for $L_2(R)$.

ASSUMPTION 2'. Assumption 1 holds and the coefficients of the expansion (3.4) satisfy

$$\sum_{j=1}^{\infty} \left(\sum_k \beta_{jk}^2 \right) |\log(j+1)|^2 \lambda(j)^2 < \infty \tag{3.19}$$

for some $\lambda(j)$ positive and increasing to ∞ with j.

This last assumption concerns $h(.)$ and is a strengthening of (3.7). Its role is bounding the bias of the estimate of $h(x)$ that will be constructed. In the case that $\psi(.)$ has compact support, the inner sum in (3.19) is over $\le A\,2^j$ terms for some A.

Concerning the time series $Y(t)$, suppose the cumulant functions of the stationary error series E(.) exist and are denoted

$$c_m(u_1, \ldots, u_{m-1}) = cum\{\mathrm{E}(t+u_1), \ldots, \mathrm{E}(t+u_{m-1}), \mathrm{E}(t)\}$$

for $m = 1, 2, \cdots$ and $t, u = 0, \pm 1, \pm 2, \cdots$. In the case $m = 2$, write $c_{\mathrm{EE}}(u)$ for $c_2(u)$. The power spectrum at frequency λ is

$$f_{\mathrm{EE}}(\lambda) = \frac{1}{2\pi} \sum_u e^{-i\lambda u}\, c_{\mathrm{EE}}(u)$$

and will be needed below. Needed as well is

ASSUMPTION 3. The cumulant functions of the zero mean stationary series $\mathrm{E}(t)$, $t = 0, \pm 1, \pm 2, \cdots$ satisfy

$$K_m = \sum_{u_1, \ldots, u_{m-1}} | c_m(u_1, \cdots, u_{m-1}) | < \infty \tag{3.20}$$

for $m = 2, 3, \cdots$. Also

$$\sum_u |u|\, |c_{\mathrm{EE}}(u)| < \infty \tag{3.21}$$

and $f_{\mathrm{EE}}(0) \ne 0$.

Here (3.20) is Assumption 2.6.1 in Brillinger [8]. It is a form of mixing condition and leads to the consistency and asymptotic normality of the estimates to be studied.

THEOREM I. Suppose the model (3.11) holds with $S(t) = h(t/T)$. As $T \to \infty$, under Assumptions 1, 2, 3

i)

$$E\{(\hat{\alpha}_{lk}^U - \alpha_{lk}^U)\} = O(2^{l/2} U^{1/2} T^{-1}) \tag{3.22}$$

$$E\{(\hat{\beta}_{jk}^U - \beta_{jk}^U)\} = O(2^{j/2} U^{1/2} T^{-1}) \tag{3.23}$$

where the errors terms are uniform in j, k, l, U, T. Also

ii)

$$cov\{\hat{\alpha}_{lk}^U, \hat{\alpha}_{lk'}^U\} = 2\pi f_{\mathrm{EE}}(0) T^{-1} \int_{-U}^{U} \phi_{lk}(u)\phi_{lk'}(u)\,du + O(2^l U T^{-2}) \tag{3.24}$$

$$cov\{\hat{\alpha}_{lk}^U, \hat{\beta}_{j'k'}^U\} = 2\pi f_{\mathrm{EE}}(0) T^{-1} \int_{-U}^{U} \phi_{lk}(u)\psi_{j'k'}(u)\,du + O(2^{(l+j')/2} U T^{-2}) \tag{3.25}$$

$$cov\{\hat{\beta}_{jk}^U, \hat{\beta}_{j'k'}^U\} = 2\pi f_{\mathrm{EE}}(0) T^{-1} \int_{-U}^{U} \psi_{jk}(u)\psi_{j'k'}(u)\,du + O(2^{(j+j')/2} U T^{-2}) \tag{3.26}$$

The errors terms are uniform in j, j', k, k', l, U, T. Next
iii)

$$|\, cum\{\hat\beta^U_{j_1k_1},...,\hat\beta^U_{j_mk_m}\}\,| \le A^m K_m 2^{(j_1+\cdots+j_m)(1/2-1/m)} U^{m/2-1} T^{-m+1} \tag{3.27}$$

for some finite A, with similar expressions for the cumulants involving the $\hat\alpha^U$. Finally
iv) If $U/T \to 0$, $U \ge \varepsilon 0$ as $T \to \infty$, then finite collections of the $\hat\alpha^U$, $\hat\beta^U$ are asymptotically normal with the indicated first and second order moments.

Since $f_{EE}(0) \ne 0$, the variances are actually of order T^{-1} for $U > 0$. If $U \to \infty$ with T, then the integrals in (3.24) - (3.26) vanish in the limit for distinct subscripts and the corresponding statistics are asymptotically independent. If the E(t) are independent with variance σ_{EE} then $2\pi f_{EE}(0)$ is replaced by σ_{EE} in the above expressions.

The proof of the theorem is given in the Appendix. It proceeds by evaluating the joint cumulants of the $\hat\alpha^U$ and $\hat\beta^U$ and seeing that sup U under the indicated assumptions, (3.22)-(3.27) are satisfied. The result might have been anticipated by Theorem 4.4.2 of Brillinger [8] with the $\phi^U_{lk}(.)$ and $\psi^U_{jk}(.)$ in the role of the tapering functions of that theorem.

Expression (3.26) will later suggest an estimate of the power spectrum value $f_{EE}(0)$.

Below it will be seen that

$$var\{\hat h(x)\} \approx \frac{2\pi f_{EE}(0)}{T}\tau(x)$$

with

$$\tau(x) = \sum_k \phi^U_{Jk}(x)^2 = 2^J U \sum_k \phi(2^J Ux - k)^2$$

Because $\phi(.)$ has finite support, $\tau(x) \le A 2^J U$ for some A, but its nearness to 0 for large $2^J U$ is not immediately clear.

THEOREM II. Under the assumptions of Theorem I, for almost all x, y in $[-1,1]$
i)

$$E\{\hat h(x)\} = \sum_k \alpha^U_{Jk}\phi^U_{Jk}(x) + O(2^J UT^{-1}) \tag{3.28}$$

ii) If $U \to \infty$ as $T \to \infty$, then

$$cov\{\hat h(x), \hat h(y)\} \approx$$

$$\frac{2\pi f_{EE}(0)}{T}\sum_k \phi^U_{Jk}(x)\phi^U_{Jk}(y) \tag{3.29}$$

iii) The joint cumulants of order m are $O(2^{(m-1)J}U^{m-1}T^{-m+1})$ as $T \to \infty$ and
iv) If $T/\tau(x_1), ..., T/\tau(x_N) \to \infty$, as $T \to \infty$ then $\hat h(x_1), \cdots, \hat h(x_N)$ are asymptotically jointly normal with the indicated first and second order moments, for N a finite integer.

In the Haar case the asymptotic normality is not surprising, since the estimate is the mean of $T/U2^J$ contiguous values and $T/U2^J \to \infty$. The proof of the Theorem is given in the Appendix. It follows by evaluating the cumulants of $\hat h(x)$, making use of Theorem I.

To study the bias in more detail consider the expression

$$\sum_{j=-\infty}^{J-1} \sum_k \beta_{jk}^U \, \psi_{jk}^U(x) \tag{3.30}$$

which is another way of writing the first term on the right hand side of (3.28). Supposing $U = 2^n$ this expression becomes

$$\sum_{j=-\infty}^{J-1} \sum_k \beta_{j+n,k} \psi_{j+n,k}(x) = \sum_{j=-\infty}^{J+n-1} \sum_k \beta_{jk} \psi_{jk}(x) \tag{3.31}$$

From Theorem 1' of Móricz [37], concerning the degree of approximation of a function by the partial sums of an orthogonal expansion, one has under Assumption 2' that, for almost all x, (3.31) is $h(x) + o(1/\lambda(J+n-1))$ as $n \to \infty$. No smoothness assumptions have been made re ϕ, ψ in that development. Under other assumptions on $\phi(.)$ and $h(.)$ other expressions may be obtained for the degree of approximation of a partial wavelet expansion, see Antoniadis [1], Antoniadis et al. [2], Kon and Rafael [33].

COROLLARY. Under the assumptions of the Theorem and Assumption 2', and if $U = 2^n$, then $\hat{h}(x)$ is asymptotically unbiased and consistent at almost all x in $[-1,1]$, provided $\lambda(J+n-1) \to \infty$, $2^J 2^n T^{-1} \to 0$ as $T \to \infty$.

It may be noted that the asymptotic distribution of $\sqrt{T/2^J U}\,\hat{h}(x)$ can be centered at $h(x)$, provided $2^J U T^{-1}, 2^{-J/2} U^{-1/2} T^{1/2} \lambda(J+n-1)^{-1} \to 0$ as $T \to \infty$.

To construct a confidence interval for $h(x)$, one will need an estimate of $f_{EE}(0)$. Noting (3.23), (3.26), an estimate could be based on the $\hat{\beta}_{jk}^U$ for which it is felt that the corresponding $\beta_{jk} = 0$. Supposing this to be the case for $j = J$, a simple estimate is

$$\hat{f}_{EE}(0) = \frac{T}{2\pi} \sum_k (\hat{\beta}_{Jk}^U)^2 / K \tag{3.32}$$

provided U is large, where K is the number of k's summed over. The estimate is consistent when $K \to \infty$, appropriately with T. Specifically one has

THEOREM III. Under the conditions of Theorem I and if $K, T/K, U \to \infty$ as $T \to \infty$, then $\hat{f}_{EE}(0)$ tends to $f_{EE}(0)$ in probability.

The proof is given in the Appendix. An estimate of $var\{\hat{h}(x)\}$ is now given by

$$\frac{2\pi \hat{f}_{EE}(0)}{T} \sum_k \phi_{Jk}^U(x)^2$$

following (3.29) and one has

COROLLARY. Under the conditions of the Theorem the variate $(\hat{h}(x) - E\{\hat{h}(x)\})/(2\pi \hat{f}_{EE}(0) \sum_k \phi_{Jk}(x)^2/T)^{1/2}$ tends in distribution to the standard normal.

There are similar results for finite collections of such variates.

Antoniadis [2], Antoniadis et al. [1] consider the linear case of a Gasser-Müller type estimator and the E(t) independent and identically distributed. They develop bias, consistency and asymptotic normality results.

In some circulstances one can employ cumulant techniques to construct simultaneous confidence bounds. This could proceed for example by showing that the distribution of the number of crossings of a high level is asymptotically Poisson.

4. A SUBSTANTIVE EXAMPLE

As an illustration of wavelet analysis, consider searching for jumps in records of microtubule movement. Microtubules are linear polymers basic to cell motility. A concern is whether their movement is smooth, or rather via a series of jumps. See Malik et al. [35] for a discussion of the experiments involved.

Suppose $Y(t)$ denotes the distance a microtubule has travelled at time t. A model to consider is

$$Y(t) = \gamma + \delta t + h(t/T) + \mathrm{E}(t) \qquad (4.1)$$

with δ a parameter related to diffusion motion and $h(.)$ a step function corresponding to jumps. Re the parameter J involved in the analysis, it will be had in mind that there are but a moderate number of jumps.

The Figure presents a Haar wavelet plus linear trend fit to the data on the location of a microtubule as a function of time in one experiment. The top panel graphs the data itself. The y-units in the panels are nanometers and the x-units in milleseconds. The middle panel provides a linear trend plus Haar wavelet fit. The bottom panel provides a corresponding shrunken wavelet fit bringing in the next level of detail. (The steps of such a fit are described in Section 8.) The dashed lines provide approximate ±2 standard error limits about the fits. There is minimal evidence for jumps in this particular trajectory and not much difference between the linear and shrunken estimates. The length of the series here was 1024, $l = 3$ and $J = 4$.

The calculations here were done by two stage regression: a) fit wavelets to the data, b) fit wavelets to t, c) fit the result of b) to the result of a).

The results of similar computations, but for stream flow data, are provided in Brillinger [10].

5. IRREGULARLY OBSERVED PROCESSES

Suppose

$$Y(t) = S(t) + \mathrm{E}(t)$$

for $t \varepsilon R$, but that the values are observed at time points τ_n of a stationary stochastic point process. Denote the data by $Y(\tau_n)$, $-T \le \tau_n < T$. Define

$$N(t) = \#\{\tau_n \ in \ (-1,t]\}$$

and let the rate of the process $N(.)$ be given by $E\{dN(t)\} = p_N\, dt$.

Suppose $S(t) = h(t/T)$, with support [−1,1] as before and for some l consider the wavelet expansion,

$$h(x) = \sum_{k=-\infty}^{\infty} \alpha_{lk}^U \phi_{lk}^U(x) + \sum_{j=l}^{\infty} \sum_{k=-\infty}^{\infty} \beta_{jk}^U \psi_{jk}^U(x)$$

As estimates of the coefficients take

$$\hat{\alpha}_{lk}^U = \frac{1}{N(T)} \sum_n \phi_{lk}^U(\tau_n/T)\; Y(\tau_n)$$

and

$$\hat{\beta}_{jk}^U = \frac{1}{N(T)} \sum_n \psi_{jk}^U(\tau_n/T)\; Y(\tau_n)$$

To study these estimates consider the bivariate process $\mathbf{X}(.)$ defined by

$$d\mathbf{X}(t) = [dN(t),\ \mathrm{E}(t)dN(t)] \qquad (5.1)$$

Suppose that $\mathbf{X}$ has stationary increments and cumulant measures defined for $a_n = 1, 2,\ n = 1,...,m\ \ m = 2, 3, ..$ by

$$cum\{dX_{a_1}(t+u_1), \ldots, dX_{a_{m-1}}(t+u_{m-1}), dX_{a_m}(t)\} = dC_{a_1,\ldots,a_m}(u_1, \ldots, u_{m-1})dt$$

with $C_{a_1,\ldots,a_m}(u_1, \ldots, u_{m-1})$ of bounded variation in finite intervals. Such an assumption is employed in Brillinger [6]. The power spectral density matrix of $\mathbf{X}(.)$ at frequency λ has entry

$$f_{cd}(\lambda) = \frac{1}{2\pi}\int e^{-i\lambda u}\, dC_{cd}(u)$$

in row c and column d. In the case that E(.) and $N(.)$ above are stationary and independent, the matrix is diagonal with entries $f_{11}(\lambda) = f_{NN}(\lambda)$ and

$$f_{22}(\lambda) = p_N^2\, f_{\mathrm{EE}}(\lambda) + \int f_{\mathrm{EE}}(\lambda-\gamma) f_{NN}(\gamma) d\gamma$$

where $f_{\mathrm{EE}}(\lambda)$ is the power spectrum of E(.) and $f_{NN}(.)$ that of $N(.)$.

Paralleling Assumption 3, it will be required that

ASSUMPTION 3'. The process $\mathbf{X}(.)$ of (5.1) has stationary increments and

$$\int |dC_{a_1,\ldots,a_m}(u_1,\ldots,u_{m-1})| = K_m < \infty$$

for $a_1, \cdots, a_m = 1,2$. Also

$$\int |u|\, |dC_{cd}(u)| < \infty$$

for $c, d = 1,2$.

Now $N(T)\hat{\beta}_{jk}^U$ is given by

$$\int_{-T}^{T} \psi_{jk}^U(t/T) h(t/T) dN(t) + \int_{-T}^{T} \psi_{jk}^U(t/T) dX(t) \qquad (5.2)$$

and its expected value is

$$p_N T \int \psi_{jk}^U(x) h(x) dx = p_N T \beta_{jk}^U$$

Its variance is approximately

$$2\pi T f_{NN}(0) \int_{-1}^{1} \psi_{jk}^U(x)^2 h(x)^2 dx + 2\pi T f_{22}(0) \int_{-1}^{1} \psi_{jk}^U(x)^2 dx \qquad (5.3)$$

which may be estimated as at (3.32) following an assumption that $\beta_{jk}^U = 0$.

The cumulants of order m are in absolute value are $\le A^m K_m 2^{(j_1+\ldots+j_m)(1/2-1/m)} T$ for some finite A as before. The asymptotic normality of the $\hat{\alpha}_{lk}^U$, $\hat{\beta}_{jk}^U$ follows as Theorems 4.1 and 4.2 of Brillinger (1972) with the $\phi_{lk}^U(.)$, $\psi_{jk}^U(.)$ again playing the role of tapers.

One defines the estimate $\hat{h}(x)$ by (3.12) once again. Because of the presence of $h(.)$ in the first term of (5.3) however the variance estimate is more complicated since the terms with different jk are no longer asymptotically independent.

The τ_n could be assumed fixed as in Brillinger [7] and alternate results developed.

6. SPATIAL PROCESSES

Consider $\mathbf{t} \,\varepsilon\, R^p$, so a change has been made to a spatial and continuous domain for the basic process. In the case of $p = 2$ one might be studying an image. Consider the model

$$Y(\mathbf{t}) = S(\mathbf{t}) + \mathrm{E}(\mathbf{t})$$

for $\mathbf{t} \,\varepsilon\, R^p$ and suppose one wishes to estimate $h(\mathbf{x})$ where $S(\mathbf{t}) = h(\mathbf{t}/T)$ where $h(.)$ has support $[-1,1]^p$ and E(.) is stationary spatial noise. Suppose that $h(.)$ has the homogeneous wavelet expansion (3.10).

The analogs of Assumptions 1 to 3 above are immediate.

Consider the statistic

$$\hat{\beta}_{j\mathbf{k}}^{U(i)} = T^{-p} \int \psi_{j\mathbf{k}}^{U(i)}(\mathbf{t}/T) Y(\mathbf{t}) d\mathbf{t}$$

with $\psi_{j\mathbf{k}}^{U(i)}(x) = U^p \psi_{jk}(U\mathbf{x})$. This statistic is linear in the data so one can evaluate cumulants directly. As an estimate $\hat{\beta}_{j\mathbf{k}}^{U(i)}$ is unbiased. Arguing as in Brillinger [5], (6.1) has variance

$$\left(\frac{2\pi}{T}\right)^p f_{\mathrm{EE}}(0) \int_{-U}^{U} \cdots \int_{-U}^{U} \psi_{j\mathbf{k}}^{(i)}(\mathbf{u})^2 d\,\mathbf{u}$$

and the various $\hat{\beta}$'s are asymptotically normal. If $U \to \infty$, they will be asymptotically independent. The asymptotic normality is Theorem 3.2 of Brillinger (1970) with $\psi_{j\mathbf{k}}^{U(i)}(.)$ in the role of taper. One can follow (3.32) and consider

$$\hat{f}_{\mathrm{EE}}(0) = \left(\frac{T}{2\pi}\right)^p \sum_{i=1}^{I} \sum_{k} (\hat{\beta}_{Jk}^{U(i)})^2 / K$$

as an estimate of $f_{\mathrm{EE}}(0)$ where K is the number of $\hat{\beta}_{j\mathbf{k}}^{U(i)}$ involved.

A homogeneous linear wavelet estimate is provided by

$$\hat{h}(\mathbf{x}) = \sum_{i=1}^{I} \sum_{j=-\infty}^{J-1} \sum_{\mathbf{k}} \hat{\beta}_{j\mathbf{k}}^{U(i)} \psi_{j\mathbf{k}}^{U(i)}(\mathbf{x})$$

for some J. As a variance estimate for $\hat{h}(\mathbf{x})$ one can consider

$$\left(\frac{2\pi}{T}\right)^p \hat{f}_{\mathrm{EE}}(0) \sum_{i=1}^{I} \sum_{j=-\infty}^{J-1} \sum_{\mathbf{k}} \psi_{j\mathbf{k}}^{U(i)}(\mathbf{x})^2$$

and use this to set approximate confidence intervals.

Once again these properties may be derived by arguments involving cumulants.

7. LONG MEMORY PROCESSES

The processes considered up to now have been mixing, for example it has been assumed that

$$\sum_{u} |c_{\mathrm{EE}}(u)| < \infty \tag{7.1}$$

In the cases of a long memory process this sum may be ∞. Consider for example a discrete time circumstance where the moments of the process E(.) exist, but its power spectrum has the form

$$f_{\mathrm{EE}}(\lambda) = |1 - e^{-i\lambda}|^{2d} f^*(\lambda) \tag{7.2}$$

where $f^*(\lambda)$ is strictly positive, continuous and has bounded variation, $0 < d < 1/2$. Then

$$c_{\mathrm{EE}}(u) \approx |u|^{2d-1} 2 f^*(0)\, \Gamma(1-2d) \sin(\pi d)$$

as $|u| \to \infty$, see Künsch [34] and so (7.1) does not hold.

Under some regularity conditions including (7.2), Yajima [52] considers the large sample distribution of statistics, like (3.15-16) with $U = 1$ and $t = 0, \ldots, T-1$. He finds they are asymptotically normal, but for example

$$var\{\hat{\beta}_{jk}\} \approx 2T^{-1+2d} f^*(0)\Gamma(1-2d)\sin(\pi d) \int_0^1\int_0^1 \psi_{jk}(x)\psi_{jk}(y)|x-y|^{2d-1} dxdy$$

His argument also gives

$$cov\{\hat{\beta}_{jk}, \hat{\beta}_{j'k'}\} \approx 2T^{-1+2d} f^*(0)\Gamma(1-2d)\sin(\pi d) \int_0^1\int_0^1 \psi_{jk}(x)\psi_{j'k'}(y)|x-y|^{2d-1} dxdy$$

expressions which may be usefully compared with (3.26) with the difference that asymptotic independence can not be anticipated. As before, one could use these last to develop estimates of $f^*(0)$ and $var\{\hat{h}(x)\}$.

Rosenblatt [40] also develops such large sample distributions in the long memory case. Robinson [39] reviews the kernel estimation of mean level functions in the presence of long memory noise. Wang [51] considers efficiency results. Percival and Guttorp [38] employ the Haar transform to study the Hurst effect. Gao [23] considers the L_2 norm under a long memory assumption.

8. SHRINKAGE ESTIMATES

By shrinkage is here meant the replacement of the sample coefficients of a statistic by related "smaller" values in an attempt to obtain greater stability at the expense of some increased bias, particularly when it is felt that the actual coefficients are small. Shrinkage is basic in statistical work with wavelets, Donoho and Johnstone [19], Kerkyacharian and Picard [32], Donoho [8] and Hall and Patil [24].

There are a variety of forms of shrinkage estimate. In a regression setup coefficients $\hat{\beta}$ are multiplied by factors between 0 and 1 depending on their individual uncertainty. For example $\hat{\beta}$ may be shrunk to

$$w(\hat{\beta}/\hat{\sigma})\,\hat{\beta}$$

where $\hat{\sigma}$ is an estimate of the standard error, of $\hat{\beta}$ and $w(.)$ is a sigmoidal function such that $w(u) \approx 1$ for large $|u|$ and ≈ 0 for small $|u|$. Tukey [46], for example, proposes

$$w(u) = (1 - 1/u^2)_+ \tag{8.1}$$

while Donoho and Johnstone [19] emphasize the cases of hard and soft limiters. It may be noted that these multipliers generally weight to 0 all terms where $|\hat{\beta}|$ is less than its standard error, as is intuitively plausible.

In the wavelet case, one can consider the shrinkage estimator

$$\hat{h}(x) = \sum_k \hat{\alpha}_{lk}^U \phi_{lk}^U(x) + \sum_{j=l}^{J-1} \sum_k \hat{w}_{jk}^U \hat{\beta}_{jk}^U \psi_{jk}^U(x) \tag{8.2}$$

where $\hat{w}_{jk}$ is a multiplier depending on $\hat{\beta}_{jk}$. This type of estimate has been proposed by Donoho and Johnstone [19-21] and studied by them and by Hall and Patil [25] amongst others. The estimate is nonlinear and such nonlinearity can be necessary to obtain efficient estimates, see Donoho et al. [22].

In practice the multipliers could have the form

$$\hat{w}_{jk} = w(\hat{\beta}_{jk}/\hat{\sigma}_{jk}\delta) \tag{8.3}$$

for a sigmoidal function $w(.)$ where $\hat{\sigma}_{jk}$ is an estimate of the standard error of $\hat{\beta}_{jk}$ where the constants $1/\delta$ and J tend to ∞ at a some rate. Through choice of J and δ one can affect the location and spread characteristics of the estimate. The bias can be anticipated to be less for large J and small δ, while the variance would be less for small J and large δ. Some discussion of the situation is provided in Hall and Patil [25].

An estimate based on (8.1-8.2) was employed in the Figure of the example of Section 4. The properties of the estimate will be developed in a second paper.

9. DISCUSSION

The usefulness of the cumulants in this work is that they provide an algebraic-analytic calculus allowing routine derivation of probability bounds and approximate distributions. Various of the results are pertinent to other series estimators.

Other approaches to picking up discontinuities via wavelet techniques are indicated in: Donoho [16], Wang [50], and Hall et al. [26].

Acknowledgements

The referee made some basic comments that improved the substance of the paper. Anestis Anoniadis, Peter Bühlmann and Jonathan Raz also made helpful comments. I thank them all.

References

[0] N. S. Altman, Kernel smoothing of data with correlated errors, J. Amer. Statist. Assoc. 85 (1990), 749-759.

[1] A. Antoniadis, Wavelet methods for smoothing noisy data, Wavelets, Images and Surface Fitting (eds. P. J. Laurent, A. Le Méhauté and L. I. Schumaker), Peters, Wellesley (1994), 21-28.

[2] A. Antoniadis, G. Gregoire and I. W. McKeague, Wavelet methods for curve estimation, J. Amer. Statist. Assoc. 89 (1994), 1340-1353.

[3] O. E. Barndorff-Nielsen and D. R. Cox, Asymptotic Techniques for Use in Statistics, Chapman and Hall, London, 1989.

[4] J. J. Benedetto and M. W. Frazier Eds., Wavelets. CRC Press, Boca Raton, 1994.

[5] D. R. Brillinger, The frequency analysis of relations between stationary spatial series. Proc. Twelfth Biennial Seminar Canadian Math. Congress (ed. R. Pyke). Canadian Math. Congress, Montreal, 1970, 39-81.

[6] D. R. Brillinger, The spectral analysis of stationary interval functions, Proc. Sixth Berkeley Symp. Vol. I. University of California Press, Berkeley, 1972, 483-513.

[7] D. R. Brillinger, Estimation of the mean of a stationary time series by sampling. J. Appl. Prob. 10 (1973), 419-431.

[8] D. R. Brillinger, Time Series: Data Analysis and Theory, Holt, New York, 1975.

[9] D. R. Brillinger, Moments, cumulants and some applications to stationary random processes, Moments and Signal Processing (eds. P. Purdue and H. Solomon) Naval Postgraduate School, Monterey NPS-OR-92-012 (1992), 108-126.

[10] D. R. Brillinger, Some river wavelets, Environmetrics 5 (1994), 211-220.

[11] S-T. Chiu, Bandwidth selection for kernel estimate with correlated noise, Stat. Prob Letters 8 (1989), 347-354.

[12] A. Cohen, I. Daubechies and P. Vial, Wavelets on the interval and fast wavelet transforms. Appl. Comp. Harmon. Anal. 1 (1993), 54-81.

[13] I. Daubechies, Orthonormal bases of compactly supported wavelets, Comm. Pure Appl. Math. 41 (1988), 909-996.

[14] I. Daubechies, Ten Lectures on Wavelets. SIAM, Philadelphia, 1992.

[15] I. Daubechies, Two recent results on wavelets: wavelet bases for the interval and biorthogonal wavelets diagonalizing the derivative operator, Recent Advances in Wavelet Analysis (eds. L. L. Schumaker and G. Webb), Academic, 1994, 237-258.

[16] D. Donoho, On minimum entropy segmentation, preprint, Stanford University, 1993.

[17] D. Donoho, Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. Proceedings of Symposia in Applied Mathematics, American Mathematical Society 1993, 173-205.

[18] D. L. Donoho, Unconditional bases are optimal bases for data compression and for statistical estimation, Appl. Computational Harmonic Analysis 1 (1993), 100-115.

[19] D. L. Donoho and I. M. Johnstone, Wavelets and optimal nonlinear function estimates. Technical Report 281, Statistics Department, University of California, Berkeley, 1990.

[20] D. L. Donoho and I. M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, J. Amer. Statist. Assoc. 1995, to appear.

[21] D. L. Donoho and I. M. Johnstone, Ideal spatial adaptation via wavelet shrinkage, Biometrika, 1995, to appear.

[22] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian and D. Picard. Wavelet shrinkage: asymptopia? J. Roy. Statist. Soc. B 57 (1995), 301-369.

[23] H. Gao, On the rates of convergence for nonparametric curve estimation with dependent data, preprint, University of California, Berkeley (1992).

[24] P. Hall and P. Patil, On wavelet methods for estimating smooth functions. SMS 59-93, Australian National University, 1993.

[25] P. Hall and P. Patil, Effect of threshold rules on performance of wavelet-based curve estimators. Report SRR 013-94, Centre for Mathematics and its Applications, Australian National University, Canberra, 1994.

[26] P. Hall, P. Patil and Y. K. Truong, On estimating possibly discontinuous regression involving stationary time series, Statistics Research Report No. SRR 035-94, Australian National University, 1994.

[27] W. Härdle and P-D. Tuan, Some theory of M-smoothing of time series, J. Times Series Anal. 7 (1986), 191-204.

[28] J. D. Hart, Kernel regression estimation with time series errors, J. R. Statist. Soc. B (1991), 173-187.

[29] J. D. Hart, Automated kernel smoothing of dependent data by using time series cross-validation, J. R. Statist. Soc. B (1994), 529-542.

[30] S. E. Kelly, Gibbs phenomenon for wavelets, preprint, University of Wisconsin-La Crosse, 1995.

[31] S. E. Kelly, M. A. Kon and L. A. Raphael, Local convergence for wavelet expansions, J. Functional Analysis 126 (1994), 102-138.

[32] G. Kerkyacharian and D. Picard, Density estimation in Besov spaces, Statistics and Probability Letters 13 (1992), 15-24.

[33] M. Kon and L. Raphael, Convergence rates of wavelet expansions, preprint, Boston University, 1994.

[34] H. R. Künsch, Dependence among observations: consequences and methods to deal with it. Directions in Robust Statistics and Diagnostics (eds. W. Stahel and S. Weisberg). Springer, New York, 1991 131-140.

[35] F. Malik, D. Brillinger and R. D. Vale, High-resolution tracking of microtubule motility driven by a single kenesin motor, Proc. Natl. Acad. Sci. USA 91, 4584-4588.

[36] Y. Meyer, Wavelets: Algorithms and Applications. SIAM, Philadelphia, 1993.

[37] F. Móricz, Approximation theorems for double orthogonal series. J. Approx. Theory 42(1984), 107-137.

[38] D. B. Percival and P. Guttorp, Long-memory processes, the Allan variance and wavelets, Wavelets in Geophysics (eds. E. Foufoula-Georgiou and P. Kumar), Academic, 1994, 325-344.

[38a] G. Polya and G. Szegö, Aufgaben und Lehrsätze aus Analysis I, Springer, Berlin (1925).

[39] P. M. Robinson, Nonparametric function estimation for long memory time series. Nonparametric and Semiparametric Methods in Econometrics and Statistics (eds. W. A. Barnett, J. Powell and G. E. Tauchen). Cambridge University Press, Cambridge, 1991, 437-457.

[40] M. Rosenblatt, Fractional integrals of stationary processes and the central limit theorem. J. Applied Prob. 13 (1976), 723-732.

[41] R. Rudzkis, L. Saulis and V. Statulevicius, A general lemma on probabilities of large deviations. Liet. Mat. Rink. 18 (1978), 99-116.

[42] V. A. Statulevicius, Applications of semi-invariants to asymptotic analysis of distributions of random processes. Mulitvariate Analysis - IV (ed. P. R. Krsihnaiah). North-Holland, Amsterdam, 1977, 325-337.

[43] V. Statulevicius and A. Aleskeviciene, On large deviations in the Poisson approximation, Theory Prob. Appl. 38 (1993), 385-393.

[44] R. S. Strichartz, How to make wavelets. American Math. Monthly 100 (1993), 539-556.

[45] Y. K. Truong, Nonparametric curve estimation with time series errors, J. Statist. Plan. Inf. 28 (1991), 167-183.

[46] J. W. Tukey, Introduction to the dilemmas and difficulties of regression, preprint, Princeton University, 1979.

[47] G. G. Walter, Approximation of the delta function by wavelets. Journal of Approximation Theory 71 (1992), 329-343.

[48] G. G. Walter, Wavelets: a new tool in applied mathematics. Journal of Undergraduate Mathematics Applications 14 (1993), 155-178.

[49] G. G. Walter, Wavelets and Other Orthogonal Systems With Applications, CRC Press, Boca Raton, 1994.

[50] Y. Wang, Jump and sharp cusp detection by wavelets - one dimensional case, preprint, University of Missouri-Columbia, 1994.

[51] Y. Wang, Function estimation via wavelets for data with long-range dependence, preprint, University of Missouri-Columbia, 1994.

[52] Y. Yajima, A central limit theorem of Fourier transforms of strongly dependent stationary processes. J. Time Series Analysis 10 (1989), 375-383.

APPENDIX

Throughout the proofs, A will denote a finite bound. Sometimes the properties of the $\hat{\alpha}^U$ will not be developed, but in those cases the argument presented for the $\hat{\beta}^U$ is applicable.

The following lemma of Polya and Szegö [38a] will be needed.

LEMMA 1. If the function g has finite total variation, V, on $[-1,1]$, then

$$\left| \int_{-1}^{1} g(x)dx - T^{-1} \sum_{t=-T}^{T} g\left(\frac{t}{T}\right) \right| \le \frac{V}{T}$$

for integer $T > 0$.

PROOF OF THEOREM I.

Consider (3.23). One has

$$E\{\hat{\beta}_{jk}^U\} = \frac{1}{T}\sum_t \psi_{jk}^U(\frac{t}{T})h(\frac{t}{T})$$

and from Lemma 1

$$\left| \frac{1}{T}\sum \psi_{jk}^U(\frac{t}{T})h(\frac{t}{T}) - \int \psi_{jk}^U(x)h(x)dx \right| \le variation\{\psi_{jk}^U\, h\}/T$$

and that variation is $A\,2^{j/2}U^{1/2}$. This gives (3.23). The result for $\hat{\alpha}_{jk}^U$ follows similarly.

Consider next (3.26) in the case $(j,k) = (j',k')$. One has

$$var\{\hat{\beta}_{jk}^U\} = \frac{1}{T^2}\sum_{t_1}\sum_{t_2} \psi_{jk}^U(\frac{t_1}{T})\psi_{jk}^U(\frac{t_2}{T})c_{\mathrm{EE}}(t_1 - t_2)$$

$$= \frac{1}{T^2}\sum_{u=-2T}^{2T} c_{\mathrm{EE}}(u) \sum_t \psi_{jk}^U(\frac{t+u}{T})\psi_{jk}^U(\frac{t}{T})$$

where the sum for t is from $max(-T,-T-u)$ to $min(T,T-u)$. Next, with $u > 0$, (the negative u case follows similarly)

$$\left| \sum_t \psi_{jk}^U(\frac{t+u}{T})\psi_{jk}^U(\frac{t}{T}) - \sum_t \psi_{jk}^U(\frac{t}{T})^2 \right|$$

$$\le A\,2^{j/2}U^{1/2} \sum_t \Big[\left|\psi_{jk}^U(\frac{t+u}{T}) - \psi_{jk}^U(\frac{t+u-1}{T})\right| + \left|\psi_{jk}^U(\frac{t+u-1}{T}) - \psi_{jk}^U(\frac{t+u-2}{T})\right|$$
$$+ \cdots + \left|\psi_{jk}^U(\frac{t+1}{T}) - \psi_{jk}^U(\frac{t}{T})\right| \Big]$$

$$\le 2^{j/2}U^{1/2}A\ |u|\ variation\{\psi_{jk}^U\} \le 2^j\, UA\, |u|$$

Again from Lemma 1

$$\left| \frac{1}{T}\sum_t \psi_{jk}^U(\frac{t}{T})^2 - \int \psi_{jk}^U(x)^2 dx \right| \le 2^j\, UA/T$$

and the result (3.26) follows on remembering the definition

$$f_{\mathrm{EE}}(0) = \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} c_{\mathrm{EE}}(u)$$

and the assumption (3.21). The covariance result follows similarly.

For part iii), writing a for a subscript pair jk

$$cum\{\hat{\beta}_{a_1}^U, \cdots, \hat{\beta}_{a_m}^U\} =$$

$$\frac{1}{T^m}\sum_{u_1=-2T}^{2T} \cdots \sum_{u_{m-1}=-2T}^{2T} c_{m-1}(u_1, \cdots, u_{m-1}) \sum_t \psi_{a_1}^U(\frac{t+u_1}{T}) \cdots \psi_{a_{m-1}}^U(\frac{t+u_{m-1}}{T})\psi_{a_m}^U(\frac{t}{T})$$

Abreviating the notation, from Hölder's inequality

$$\left| \sum_t \psi_{a_1}^U \cdots \psi_{a_m}^U \right| \le (\sum_t |\psi_{a_1}|^m)^{1/m} \cdots (\sum_t |\psi_{a_m}|^m)^{1/m}$$

and

$$\sum_t |\psi_a^U(\frac{t}{T})|^m \le 2^{jm/2} A^m 2^{-j} U^{-1} T \ \ |support\ \psi|$$

counting terms. Putting these together one has iii).

The asymptotic normality, part iv), follows from the fact that the cumulants of $\sqrt{T}\,\hat{\beta}$ tend to those of the normal as $T \to \infty$ and Lemma P4.5, page 403 in Brillinger (1975).

PROOF OF THEOREM II.

The proof will use the first representation (3.18) The number of k with nonzero terms is bouned by $2N$. Using (3.22) the expected value is

$$\sum_k E\{\hat{\alpha}_{Jk}^U\}\phi_{Jk}^U(x) = \sum_k (\alpha_{Jk}^U + O(2^{J/2}U^{1/2}T^{-1}))\phi_{Jk}^U(x) = \sum_k \alpha_{Jk}^U \phi_{Jk}^U(x) + O(2^J U T^{-1})$$

as indicated.

For (3.29), one uses (3.24). Consider

$$\sum_k \sum_{k'} cov\{\hat{\alpha}_{Jk}^U, \hat{\alpha}_{Jk'}^U\}\phi_{Jk}^U(x)\phi_{Jk'}^U(y)$$

$$\approx \frac{2\pi f_{\mathrm{EE}}(0)}{T} \sum_k \phi_{Jk}^U(x)\phi_{Jk}^U(y)$$

as desired for U sufficiently large.

Parts iii) and iv) follow likewise from the result iii) of Theorem I with $a = (J,k)$. One has

$$cum_m\{\hat{h}(x)\} = \sum_{a_1} \cdots \sum_{a_m} cum\{\hat{\alpha}_{a_1}^U, \cdots, \hat{\alpha}_{a_m}^U\}\phi_{a_1}^U(x) \cdots \phi_{a_m}^U(x)$$

$$= O(\ T^{-m+1} \sum_{a_1} \cdots \sum_{a_m} 2^{(j_1 + \cdots + j_m)(1/2 - 1/m)} 2^{j_1/2} \cdots 2^{j_m/2}\ )$$

$$= O(\ T^{-m+1} [\sum_a 2^{j(1-1/m)}]^m\ )$$

The convergence of the standardized cumulants to those of the normal gives the asymptotic normality.

PROOF OF THEOREM III.

From (3.26) and the fact that $E\{\hat{\beta}\}$ is 0

$$E\{\hat{f}_{\mathrm{EE}}(0)\} = \frac{2\pi}{T} f_{\mathrm{EE}}(0)\ (\sum_k \int_{-U}^{U} \psi_{Jk}(u)^2 du)/K + O(2^J U/T)$$

From (3.27) the variance of $\hat{\beta}^U$ is $O(T^{-1})$ uniformly and the fourth cumulants are $O(2^J U T^{-3})$ uniformly. It follows that the covariances of the $(\hat{\beta}^U)^2$'s are $O(T^{-2})$+$O(2^J U T^{-1})$ and so $var\{\hat{f}_{\mathrm{EE}}(0)\}$ is $O(K^2 T^{-2})$. Thus the estimate is mean square consistent.

PROOF OF COROLLARY.

One uses the result that if as $T \to \infty$, a sequence $(Y_T - \mu_T)/\sigma_T$ approaches the standard normal in distribution and if $\hat{\sigma}_T/\sigma_T$ tends to 1 in probability, then $(Y_T - \mu_T)/\hat{\sigma}_T$ also tends to the standard normal in distribution.

Figure legend

The top panel provides one of the traces of movement of a microtubule. The second panel provides a fit of the model (4.1) employing Haar wavelets and approximate 95% confidence limits around the fitted line. The third panel provides the shrinkage estimate (8.2) employing the Tukey multiplier (8.1).