Some Asymptotics of Wavelet Fits in the Stationary Error Case

By

David R. Brillinger*

Technical Report No. 415
June 1994

*Research partially supported by ONR Grant N00014-94-1-0042

Department of Statistics
University of California
Berkeley, California 94720

$2\pi \neq 1$
$2\pi \neq 1$

SOME ASYMPTOTICS OF WAVELET FITS IN THE STATIONARY ERROR CASE [1]

DAVID R. BRILLINGER

University of California, Berkeley

The model $Y(t) = S(t) + E(t)$, $t = 0, \pm 1, \pm 2, \ldots$ with $S(t)$ a deterministic mean level and $E(t)$ stationary mixing noise is investigated. There is a brief review of traditional methods of estimating $S(.)$, then wavelet techniques of fitting are considered. The large sample distributions of both linear and shrunken wavelet estimates are developed.

1. Introduction. Wavelets are a contemporary tool for function approximation. They are competitors/collaborators with traditional Fourier analysis and other orthogonal function expansions. In particular they are useful for handling localized behavior, discontinuities, and scale and shift transformations. In the time series case they have the useful ability to pick up transient behavior. For example Donoho (1993c) 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

[1] This research was partially supported by the ONR Grant N00014-94-1-0042.
AMS 1991 subject classification. Primary 62M10; secondary 60F05, 60F10.

Keywords and phrases. Asymptotic normality, kernel estimate, large deviations, mean level function, regression estimate, shrinkage estimate, stationary process, wavelet estimate

series falls into this domain. The locations of the jumps could correspond to the times of exogenous events in a practical situation.

Wavelet estimates may be linear in the data available, however a breakthrough occurred when the concept of shrinkage was introduced to wavelet analysis. In it the estimated coefficients of the expansion are moved closer to 0. Quoting from Donoho (1993b): "Traditional methods ... are linear and cannot compete effectively with the wavelet shrinkage method in cases of high spatial variability - either in practice ... or in theory. In estimating functions of bounded variation, linear methods cannot attain the optimal rate ... ; the wavelet shrinkage method ... attains mean-squared error size $(log(n)/n)^{2/3}$ based on n observations, while linear and adaptive linear methods attain only an error size $n^{-1/2}$". There are different ideas re what constitutes high spatial variability. Hall and Patil (1993) are concerned with highly oscillatory behavior near a time point. This paper, motivated by examples in Brillinger (1993, 1994), is concerned with the possible existence of jump discontinuities.

The focus of this paper is the case where the additive error is stationary. The paper begins with some review of existing procedures for the problem of estimating mean level functions of stationary time series. Some of these are regression techniques, others are kernel smoothers. Then wavelet estimates are discussed and the large sample distribution derived for both the linear and shrunken cases. The linear case is studied quite generally. The shrunken case is investigated for a model of the wavelet expansion containing a finite, but unknown, number of terms and with hardlimiters employed in the shrinking. The final estimate studied requires an assumption of normality in the derivation of its asymptotic distribution. The large sample distribution allows the construction of approximate confidence intervals for example.

2. Estimating mean level functions. Consider the model

$$Y(t) = S(t) + \mathrm{E}(t) \tag{2.1}$$

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

$$E\{\, Y(t)\} = S(t \mid \alpha) = \alpha_1 g_1(t) + \cdots + \alpha_J g_J(t) \tag{2.2}$$

with J known and the $g_1(.), \cdots, g_J(.)$ given functions, the large sample distribution of the ordinary and of best linear unbiased least squares estimates were determined long ago (see Grenander and Rosenblatt (1957), Rosenblatt (1959), Hannan (1970), Anderson (1971), Brillinger (1975)). Hannan (1973, 1979) considered the case of $g_j^T(t)$. Results are also available for the case of nonlinear regression, the function $S(t \mid \theta)$ being known up to a finite dimensional parameter θ. Asymptotic distributions may be derived, see Hannan (1971), Robinson (1972), Gallant and Goebel (1976).

In the case that the mean function $S(t)$ is smooth, one can consider its estimation by a running mean or kernel smoother. In one formulation, to handle the discreteness of t, $S(t)$ is written $h(t/T)$ for a function $h(x)$, on [0,1]. The estimate is

$$\hat{h}(x) = \sum_t Y(t)\, w_b(x - \frac{t}{T}) \,/\, \sum_t w_b(x - \frac{t}{T}) \tag{2.3}$$

where the kernel, $w_b(.)$, has binwidth b. In the case of fixed b, the estimate (2.3) is linear in the data so various approximate distribution results may be developed. Härdle and Tuan (1986) present results including robust procedures. The problem of estimating b is considered in Chiu (1989), Hart (1989, 1994), Altman (1990). An optimal b is determined in Truong (1991). Variable binwidth smoothers have been proposed for kernel estimates on occasion, see Müller and Stadtmuller(1987), Staniswalis (1989), Hastie and Tibshirani (1990), Brockman et al. (1993). The wavelet estimates to be presented in Section 3 have a variable character. Further approaches to the estimation of the function $S(.)$ include: orthogonal series expansions (Kronmal and Tarter (1968), Müller (1988)), smoothness priors, (Akaike(1980), Kitagawa (1987), Gersch (1992)), local regression (Cleveland et al. (1991)), penalized likelihood and splines (Silverman (1985), Diggle and Hutchinson (1989)). Müller (1992) and Wu and Chu (1993) investigate the case with a discontinuity in $h(.)$.

3. Wavelets.

3.1 Introduction. Wavelet analyses correspond to particular types of (orthonormal) series expansions. There is a scaling function $\phi(.)$ and a mother wavelet $\psi(.)$ given by

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

for some coefficients c_k. These functions generate families

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

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

such that, for given integer l (which may be $-\infty$),

$$\{ \ \phi_{lk}(x) \quad \text{and} \quad \psi_{jk}(x), \quad j = l+1, \cdots \ k = 0, \pm 1, \pm 2, \cdots \ \}$$

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

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

with

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

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

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 that wavelet approximations are noted for. Also 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 t and this is another of the advantages of the wavelet approach. General references include Daubechies (1992), Walter (1992,1993), Meyer (1993), Strichartz (1993), Benedetto and Frazier (1994).

Two particular examples are:

a) the Haar case where

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

and

$$\psi(x) = \begin{cases} 1 & 0 \le x < 1/2 \\ -1 & 1/2 \le x < 1 \\ 0 & otherwise \end{cases} \tag{3.7}$$

The expansion (3.4) can be anticipated to be particularly appropriate when $h(.)$ is piecewise constant.

b) the Daubechies' case where $\phi(.)$ has support [0,3], is continuous and the coefficients of (3.1) are

$$c_0 = \frac{1+\sqrt{3}}{4} \quad c_1 = \frac{3+\sqrt{3}}{4} \quad c_2 = \frac{3-\sqrt{3}}{4} \quad c_3 = \frac{1-\sqrt{3}}{4}$$

Because of its continuity, the Daubechies' case is finding substantial use in practice. There are variants where $\phi(.)$, $\psi(.)$ have a specified number of derivatives.

In practice a finite expansion will be employed, rather than (3.4), and one may be concerned with convergences other than that of L_2. There are results pertinent for the case of general orthonormal systems $\{\nu_n(.), n = 1, 2, \cdots\}$ of $L_2(I)$, I a finite interval. Consider $h(.) \varepsilon L_2(I)$, with

$$\gamma_n = \int_I \nu_n(x) h(x) dx$$

Then $\sum_n \gamma_n^2 < \infty$ and the partial sum

$$h^N(x) = \sum_{n=1}^{N} \gamma_n \nu_n(x) \tag{3.8}$$

converges to $h(x)$ in $L_2(I)$,

$$\int_I |h^N(x) - h(x)|^2 dx \to 0$$

as $N \to \infty$. If, in addition,

$$\sum_n [\log n]^2 \gamma_n^2 < \infty$$

then $h^N \to h$ almost everywhere, see Section 10.21 of Zygmund (1959). Further, if for some positive sequence $\lambda(n)$, that increases to ∞,

$$\sum_n [log\ n]^2 \lambda(n)^2 \gamma_n^2 < \infty$$

then

$$h^N(x) - h(x) = o(1/\lambda(N)) \qquad \textit{almost everywhere in } x$$

as $N \to \infty$, see Tandori (1959). One can insert convergence factors in (3.8) to improve the convergence rate.

In what follows concern will be with the almost everywhere case.

3.2 The statistical setup. Consider the model

$$Y(t) = S(t) + E(t) \tag{3.9}$$

$t = 0, \pm 1, \pm 2, \cdots$ with $S(.)$ deterministic and $E(.)$ zero mean stationary noise. Suppose that the data $Y(t)$, $t = 0, \cdots, T-1$ are available and that one can write

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

with $h(.)$ zero outside [0,1]. Paralleling (3.5), (3.6), (3.4), one can consider the statistics

$$\hat{\alpha}_{lk} = \frac{1}{T}\sum_{t=0}^{T-1} \phi_{lk}(\frac{t}{T})Y(t) \tag{3.10}$$

$$\hat{\beta}_{jk} = \frac{1}{T}\sum_{t=0}^{T-1} \psi_{jk}(\frac{t}{T})Y(t) \tag{3.11}$$

The wavelet transforms of both the signal $S(t/T)$ and the noise $E(t)$ are involved here. The linear estimate of $h(x)$ is

$$\hat{h}_0(x) = \sum_k \hat{\alpha}_{lk}\phi_{lk}(x) + \sum_{j>l}^{J_T}\sum_k \hat{\beta}_{jk}\psi_{jk}(x) \tag{3.12}$$

for some large J_T. Since ϕ has compact support the number of k for which $\phi_{jk}(x) \neq 0$ is bounded, uniformly in j, by $2\,|support\ \phi|$. Similarly for $\psi_{jk}(x)$, so only a finite number of terms are involved in (3.12).

For Haar wavelets the function (3.12) will be piecewise constant. In applications, the times of change of the mean level might correspond to discrete exogenous events. In this case the statistics $\hat{\alpha}$, $\hat{\beta}$ simplify. There is a single $\hat{\alpha}$ in (3.12), say $\hat{\alpha}_0$, and it is the mean of the available Y's. The $\hat{\beta}_{jk}$ are given by

$$\hat{\beta}_{jk} = \frac{2^{j/2}}{T}[\sum_t{}' Y(t) - \sum_t{}'' Y(t)]$$

where Σ' is over $0 \leq 2^j t/T - k < 1/2$ and Σ'' is over $1/2 \leq 2^j t/T - k < 1$. Computing such local means, in either a smoothing or a search for change-points, seems intuitively reasonable. The estimate (3.12) is simply

$$\hat{h}_0(x) = \hat{\alpha}_0 + \sum_{j=0}^{J_T} \sum_{k=0}^{2^j-1} \hat{\beta}_{jk} 2^{j/2} \psi(2^j x - k)$$

with $\psi(.)$ given by (3.7). The estimate, $\hat{h}_0(x)$, is the mean of the data values $Y(t)$ with t/T in the dyadic interval of order J_T containing x.

3.3 Properties of the statistics. The statistics (3.10), (3.11), (3.12) are linear in the Y's, hence certain sampling properties are directly available, eg. large sample variances, cumulants and distributions. Some assumptions and consequent results will be set down.

ASSUMPTION 1. The functions $\phi(.)$ and $\psi(.)$ are of bounded variation and compact support. Further, given l integer, the collection $\{\phi_{lk}(.), \psi_{jk}(.), j = l+1, \cdots, k = 0, \pm 1, \pm 2, \cdots\}$ provides an orthonormal basis for a finite interval containing $[0,1]$.

ASSUMPTION 2. The function $h(.)$ is bounded and of bounded variation on $[0,1]$ and vanishes outside that interval.

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

$$\sum_j \sum_k [\log j]^2 \lambda(j)^2 \beta_{jk}^2 < \infty \quad .$$

for some $\lambda(n)$ increasing to ∞ with n.

Suppose the cumulant functions of the stationary error series E(.) exist and are denoted

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

for $m = 1, 2, \cdots$ and $t, u = 0, \pm 1, \pm 2, \cdots$. The power spectrum at frequency 0 is

$$f_2(0) = \frac{1}{2\pi} \sum_u c_2(u)$$

and will be needed below. Needed as well is

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

$$C_m = \sum_{u_1, \cdots, u_{m-1}} |c_m(u_1, \cdots, u_{m-1})| < \infty \tag{3.13}$$

Also

$$\sum_u |u| |c_2(u)| < \infty$$

and $f_2(0) \neq 0$.

Here (3.13) is Assumption 2.6.1 in Brillinger (1975). 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 (3.9) holds with $S(t) = h(t/T)$, then under Assumptions 1, 2, 3

i)

$$E\{2^{-l/2}(\hat{\alpha}_{lk} - \alpha_{lk})\} = O(T^{-1}) \tag{3.14}$$

$$E\{2^{-j/2}(\hat{\beta}_{jk} - \beta_{jk})\} = O(T^{-1}) \tag{3.15}$$

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

ii)

$$var\{2^{-l/2}\hat{\alpha}_{lk}\} = 2\pi f_2(0) 2^{-l} T^{-1} + O(T^{-2})$$

$$var\{2^{-j/2}\hat{\beta}_{jk}\} = 2\pi f_2(0) 2^{-j} T^{-1} + O(T^{-2}) \tag{3.16}$$

$$cov\{2^{-j/2}\hat{\beta}_{jk}, 2^{-j'/2}\hat{\beta}_{j'k'}\} = O(T^{-2}) \tag{3.17}$$

for $(j,k) \neq (j'k')$ with similar results for the remaining

$cov\{\hat{\alpha}, \hat{\beta}\}$, $cov\{\hat{\beta}, \hat{\beta}\}$. The errors terms are uniform in j, j', k, k', l.

iii)

$$|cum\{\hat{\beta}_{j_1k_1}, \cdots, \hat{\beta}_{j_mk_m}\}| \le A^m C_m 2^{(j_1 + \cdots + j_m)(1/2 - 1/m)} T^{-m+1} \tag{3.18}$$

for some finite A.

iv) as $T \to \infty$ finite collections of the $\hat{\alpha}$, $\hat{\beta}$ are asymptotically normal with the indicated first and second order moments.

The proof of the theorem is given in the Appendix. The wavelet transform values are seen to be asymptotically normal. The asymptotic independence of the individual $\hat{\alpha}$'s and $\hat{\beta}$'s follows from the orthogonality of the ϕ's and ψ's. Expression (3.16), with the occurrence of the power spectrum at 0 frequency, makes explicit in a sense that the moderate order wavelet transform is a form of lowpass filtering.

Consideration now turns to the estimate (3.12).

THEOREM II. Under the assumptions of Theorem I

i)

$$E\{\hat{h}_0(x)\} = \sum_k \alpha_{lk}\phi_{lk}(x) + \sum_{j>l}^{J_T} \sum_k \beta_{jk}\psi_{jk}(x) + O(2^{J_T}T^{-1}) \tag{3.19}$$

ii)

$$cov\{\hat{h}_0(x), \hat{h}_0(y)\} =$$

$$\frac{2\pi f_{EE}(0)}{T}\left[\sum_k \phi_{lk}(x)\phi_{lk}(y) + \sum_{j>l}^{J_T}\sum_k \psi_{jk}(x)\psi_{jk}(y)\right] + O(2^{2J_T}T^{-2}) \tag{3.20}$$

iii) joint cumulants of order m are $O(2^{(m-1)J_T}T^{-m+1})$ and

~~iv) if $2^{(m-1)J_T}T^{-m+1}/var\{\hat{h}_0(x)\}^{m/2} \to 0$ as $T \to \infty$~~ for $m = 3, 4, \cdots$,

then $\hat{h}_0(x)$ is asymptotically normal with the indicated first and second order moments.

The proof is given in the Appendix. The division by $var\{\hat{h}_0(x)\}$ in iv) is because the actual order of magnitude of the variance of $\hat{h}_0(x)$ is unclear. It is $O(2^{J_T}T^{-1})$ in any case.

COROLLARY. Under the assumptions of the theorem and Assumption 2' $\hat{h}_0(x)$ is asymptotically unbiased and consistent at almost all x, provided $\lambda(J_T) \to \infty$, $2^{J_T}T^{-1} \to 0$ as $T \to \infty$.

For a proof see the Appendix. The asymptotic distribution of $\sqrt{T}\hat{h}_0(x)$ may be centered at $h(x)$ provided $2^{J_T}T^{-1/2}$, $T^{1/2}\lambda(J_T)^{-1} \to 0$ as $T \to \infty$.

In the Haar case one sees from (3.20) that $var\{\hat{h}_0(x)\} \approx 2^{J_T}2\pi f_2(0)/T$. The condition iv) is immediate and the asymptotic normality is not surprising, since the estimate is the mean of $2^{J_T}T^{-1}$ contiguous values and $2^{J_T}T^{-1} \to \infty$.

To construct a confidence interval for $h(x)$ one will need an estimate of $f_2(0)$. Noting (3.15), (3.16), an estimate could be based on the $\hat{\beta}_{jk}$ for which it is felt that the corresponding $\beta_{jk} = 0$. One estimate has the form

$$\hat{f}_2(0) = \frac{T}{2\pi}\sum_k \hat{\beta}^2_{J_T+1,k} \;/\; K \tag{3.21}$$

where K is the number of k's summed over. In the Appendix the estimate is shown to be consistent when $K_T \to \infty$ appropriately with T. The size of K_T will be order of magnitude 2^{J_T+1} for the present case. One could also base an estimate of the power spectrum on the residuals $Y(t) - \hat{h}_0(t/T)$.

~~4. Shrinkage estimates. In~~ this paper ~~by~~ shrinkage is meant the replacement of coefficients of a statistic by related "smaller" values in an attempt to obtain greater stability at the expense of some increased bias. Shrinkage is basic in statistical work with wavelets, Donoho and Johnstone (1990), Kerkyacharian and Picard (1992), Donoho (1993b) and Hall and Patil (1993). The suggestion to use shrinkers or multipliers to "improve" estimates has been around in statistics for many years, see Lemmer (1988). There is an early spatial harmonic synthesis application in Blow and Crick (1959), concerned with crystal imaging. Thompson (1968) is concerned with improving on a simple sample mean by shrinkage and King (1972) with improving on simple regression.

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

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

where s is an estimate of its standard error and $w(.)$ is a function such that $w(u) \approx 1$ for large $|u|$ and ≈ 0 for small $|u|$. Tukey (1979), for example, proposes

$$w(u) = (1 - 1/u^2)_+$$

It may be noted that this multiplier weights to 0 all terms where $|\hat{\beta}|$ is less than its standard error. There are pertinent connections with the problems of selection of variables and pretest estimates.

4.1 The shrunken wavelet estimate for known variance. In the wavelet case, one can consider the shrinkage estimator

$$\hat{h}(x) = \sum_k \hat{\alpha}_{lk}\phi_{lk}(x) + \sum_{j>l}^{J_T}\sum_k \hat{w}_{jk}\hat{\beta}_{jk}\psi_{jk}(x) \qquad (4.1)$$

where $\hat{w}_{jk}$ is a multiplier depending on $\hat{\beta}_{jk}$. This estimate is nonlinear and as the quote in Section 1 indicates Donoho (1993b) argues that such nonlinearity is necessary to obtain efficient estimates, see also Donoho et al. (1994).

To begin it will be assumed that $\sigma_{jk}^2 = var\{\hat{\beta}_{jk}\}$ is known and nonzero. The multipliers at level j will be the indicator variables

$$\hat{w}_{jk} = I(|\hat{\beta}_{jk}| \geq \sigma_{jk}\delta_j) \tag{4.2}$$

meant to provide information on whether $\beta_{jk} = 0$ or not. The δ_j will be specified in the theorem. They will slowly increase to ∞ as T increases.

The class of mean level functions to be considered is delineated by the following assumption.

ASSUMPTION 4. The function $h(.)$ satisfies Assumption 2 and in addition only a finite number of the β_{jk} in (3.4) are nonzero.

Provided J_T is large enough, the estimate (4.1) includes all terms with $\beta_{jk} \neq 0$ and is meant to be close to and of the same character as

$$\sum_k \alpha_{lk}\phi_{lk}(x) + \sum_{j>l}^{J_0} \sum_k \beta_{jk}\psi_{jk}(x)$$

where J_0 is the largest j such that $\beta_{jk} \neq 0$. The quantity $\sum \hat{w}_{jk}$ will provide an estimate of the number of terms with $\beta_{jk} \neq 0$.

The following assumption will be needed

ASSUMPTION 5. The cumulant functions of the zero mean stationary series $E(t)$, $t = 0, \pm 1, \pm 2, \cdots$ with

$$\sum_{u_1, \cdots, u_{m-1}} |c_m(u_1, \cdots, u_{m-1})| = C_m$$

satisfy

$$\sum_m C_m z^m / m! < \infty \tag{4.3}$$

for z in a neighborhood of 0.

Assumption 5 is satisfied by a stationary Gaussian series for which $C_2 < \infty$. It is Assumption 2.6.3 in Brillinger (1975) and is needed to obtain large deviation bounds required in the proof of the following theorem. Note that the δ_j depend on T.

THEOREM III. Suppose: Assumptions 1, 4 and 5 hold, b) $\hat{h}(x)$ is given by (4.1) with $\hat{w}_{jk}$ given by (4.2), c) J_T, $2^{-J_T/2}T \to \infty$ as $T \to \infty$, d) the δ_j are such that $2^{j/2}\delta_j = o(T^{1/2})$ for $j = 1, \cdots, J_T$ and

$$\sum_{j=1}^{J_T} 2^{j/2}\exp\{-\delta_j^2/(1+\varepsilon)2\} = o(1) \tag{4.4}$$

for some $\varepsilon > 0$. Then, almost everywhere in x, finite collections of the $\hat{h}(x)$ are asymptotically normal with mean $h(x)$ and covariance function

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

$$\frac{2\pi f_2(0)}{T}\sum_k \phi_{lk}(x)\phi_{lk}(y) + \frac{2\pi f_2(0)}{T}\sum_{j>l}^{J_0}\sum_k w_{jk}^2\psi_{jk}(s)\psi_{jk}(t) \tag{4.5}$$

where $w_{jk} = 1$ if $\beta_{jk} \neq 0$ and equals 0 otherwise.

The proof of the theorem is given in the Appendix. It is notable that the variance here is of order T^{-1} rather than the $O(2^{J_T}T^{-1})$ of Theorem II.

One wants both J_T and δ_j large, but not too large: J_T large to exceed J_0, but not so large that $\hat{\alpha}$, $\hat{\beta}$ become biased. An example of a sequence δ_j satisfying (4.4) and $2^{j/2}\delta_j = o(T^{1/2})$ is given by

$$\delta_j = \sqrt{2\, log(2^{-j}T)}$$

where $2^{J_T(3+\varepsilon)/(1+\varepsilon)2}T^{-1/(1+\varepsilon)}$.

4.2 Unknown variance. In the previous section, the $var\{\hat{\beta}_{jk}\}$ were assumed known. This is unrealistic for practice. In the case they are unknown one needs an estimate of $2\pi f_2(0)/T$. Let $\hat{\sigma}^2$ denote such an estimate, eg. (3.21). The multipliers now employed are

$$\hat{w}_{jk} = I(|\hat{\beta}_{jk}| \geq \hat{\sigma}\delta_j) \tag{4.6}$$

with the δ_j to be specified.

One has

THEOREM IV. Suppose: a) Assumptions 1 and 4 hold, b) the zero mean stationary series, $E(t)$, $t = 0, \pm 1, \cdots$ is Gaussian with $C_2 < \infty$, c) $\hat{h}(x)$ is given by (4.1) with $\hat{w}_{jk}$ given by (4.6), d) $\hat{f}_2(0)$ is given by (3.19) based on K_T coefficients, e) $J_T, 2^{-J_T/2}T \to \infty$ as $T \to \infty$, f) $T^{1/2}K_T^{-1} \to 0$, for some $\varepsilon_0 > 0$, g) $2^{j/2}\delta_j = o(T^{1/2})$ for $j = 1, \cdots, J_T$ and

$$\sum_{j=1}^{J_T} 2^{j/2}\exp\{-\delta_j^2/(1+\varepsilon)2\} = o(1)$$

for some $\varepsilon > 0$. Then almost everywhere in x, finite collections of $\hat{h}(x)$ values are asymptotically normal with mean $h(x)$ and covariance function (4.5).

The proof is given in the Appendix. One wants J_T, K_T, δ_j large, but not too large.

One can estimate $var\{\hat{h}(x)$ by

$$\frac{2\pi\hat{f}_2(0)}{T}\sum_k \phi_{lk}(x)^2 + \frac{2\pi\hat{f}_2(0)}{T}\sum_{j>l}^{J_T}\sum_k \hat{w}_{jk}^2\psi_{jk}(x)^2$$

and thence form approximate confidence intervals.

In the theorem an assumption of normality is employed. Perhaps it can be replaced by the type of assumption employed in Section 7.7 of Brillinger (1975) to obtain almost sure bounds for spectrum estimates.

5. Discusson. Through the inclusion of Section 2 we have sought to contrast regression and kernel estimates with wavelet estimates. The linear

wavelet estimates have important similarities with the kernnel estimates eg. consistency and asymptotic normality, but the shrunken estimates are inherently of different character.

The question of appropriate multipliers to employ in practice is far from settled. Those of Tukey (1979) and Blow and Crick (1959) do not tend to 0 or 1 as $T \to\infty$. whereas those of Donoho and Johnstone do. Donoho and Johnstone also consider another class of multipliers, "soft-thresholders", in contrast to the "hard-thresholders" that have been investigated here. In the case of independent observations they suggest a procedure, based on an identity of Stein (1981), to estimate a threshold level.

The convergence that has been studied in this work is pointwise, because it is felt that this has pertinence to applied work. The class of function 4s considered in Assumption however is quite narrow. Donoho (1993a, b, c), Donoho and Johnstone (1990, 1994), Donoho et al. (1995) consider subtler spaces and types of convergence. Further they are concerned with studying the risk of the estimates relative to the least possible and modifications for discreteness and to handle change. Their focus is on the case of independent observations and consistency type results, whereas the present work has focussed on time series and distributional results.

There are many extensions to consider: the continuous time case, the spatial case, series with long range dependence, and borrowing strength across coefficients. The first two extensions are immediate, the same analytic arguments applying. One can consider simultaneous confidence bounds of the estimate and weak convergence in pertinent function spaces. One could consider other types of mixing conditions. The condition employed in the paper has the advantage of having an associated manipulative calculus. It has the disadvantage of requiring the existence of moments, but one could work with truncated variates if that were crucial.

Acknowledgements. I thank Mark Rizzardi for some help with the word processing.

APPENDIX

Throughout the proofs, A with a subscript will denote a finite bound. In particular suppose $|\phi(.)|$, $|\psi(.)| \le A_0$. Sometimes the properties of the $\hat{\alpha}$ will not be developed, but in those cases the argument presented for the $\hat{\beta}$ is applicable.

The following lemma of Polya and Szegö (1925) will be needed.

LEMMA 1. If the function $g(.)$ has finite total variation, V, on $[0,1]$, then

$$|\int_0^1 g(x)dx - T^{-1}\sum_{t=0}^{T-1} g(\frac{t}{T})| \le \frac{V}{T}$$

for integer $T > 0$.

Proof of Theorem I.

Consider (3.15). One has

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

and from Lemma 1

$$|\frac{1}{T}\sum_t \psi_{jk}(\frac{t}{T}) - \int \psi_{jk}(x)h(x)dx| \le V/T$$

where $V = A_1 2^{j/2}$ is the variation of $\psi_{jk}(.)h(.)$. This gives (3.15). The result for $\hat{\alpha}_{jk}$ follows similarly.

~~Consider next (3.16). One has~~

$$var\{\hat{\beta}_{jk}\} = \frac{1}{T^2}\sum_{t_1}\sum_{t_2} \psi_{jk}(\frac{t_1}{T})\psi_{jk}(\frac{t_2}{T})c_2(t_1 - t_2)$$

$$= \frac{1}{T^2} \sum_{u=-T+1}^{T-1} c_2(u) \sum_t \psi_{jk}(\frac{t+u}{T}) \psi_{jk}(\frac{t}{T})$$

where the sum for t is from $max(0,-u)$ to $min(T-1, T-1-u)$. Next, with $u > 0$, (the negative u case follows similarly)

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

$$\leq 2^{j/2} A_0 \sum_t [|\psi_{jk}(\frac{t+u}{T}) - \psi_{jk}(\frac{t+u-1}{T})| + |\psi_{jk}(\frac{t+u-1}{T}) - \psi_{jk}(\frac{t+u-2}{T})|$$
$$+ \cdots + |\psi_{jk}(\frac{t+1}{T}) - \psi_{jk}(\frac{t}{T})|]$$

$$\leq 2^j A_0 |u| V$$

where V denotes the variation of ψ. From Lemma 1

$$|\frac{1}{T} \sum_t \psi_{jk}(\frac{t}{T})^2 - \int \psi_{jk}(x)^2 dx| \leq 2^j A_2 / T$$

and the result (3.16) follows on remembering the definition

$$f_2(0) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} c_2(u)$$

and the assumption

$$\sum_u |u| |c_2(u)| < \infty$$

Next (3.17),

$$cov\{\hat{\beta}_{jk}, \hat{\beta}_{j'k'}\} = \frac{1}{T^2} \sum_{u=-T+1}^{T-1} c_2(u) \sum_t \psi_{jk}(\frac{t+u}{T}) \psi_{j'k'} \psi(\frac{t}{T})$$

as above and further

$$|\sum_t \psi_{jk}(\frac{t+u}{T}) \psi_{j'k'}(\frac{t}{T}) - \sum_t \psi_{jk}(\frac{t}{T}) \psi_{j'k'}(\frac{t}{T})|$$

$$\leq 2^{j/2+j'/2} A_0 V |u|$$

also as above. Finally

$$\left| \frac{1}{T} \sum_t \psi_{jk}(\frac{t}{T}) \psi_{j'k'}(\frac{t}{T}) - \int \psi_{jk}(x) \psi_{j'k'}(x) dx \right| \le 2^{j/2+j'/2} A_3$$

and one gets the desired result, making use of the orthogonality of the $\psi_{jk}(.)$'s .

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

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

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

Abreviating the notation for the moment, from Hölder's inequality

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

and

$$\sum_t |\psi_a(\frac{t}{T})|^m \le 2^{jm/2} A_0^m 2^{-j} T \ |support\ \psi|$$

counting terms. Putting these together one has iii).

The asymptotic normality follows from the fact that the cumulants of $\sqrt{T}\,\hat{\beta}$ tend to those of the normal as $T \to \infty$.

Proof of Theorem II.

At several places the fact that since ψ has compact support, for a given x the number of k for which $\psi_{jk} \ne 0$ is bounded, uniformly in j by $2|support\ \psi|$ will be used.

For (3.19) one uses (3.14,15) and the fact that $|\psi_{jk}(.)| \le 2^{j/2} A_0$. Specifically, consider the expected value of the second term in (3.12). It is

$$\sum_{j>l}^{J_T} \sum_k [\beta_{jk} + O(2^{j/2} T^{-1})] \psi_{jk}(x)$$

and the error term is seen to be

$$O\left(\sum_{j>l}^{J_T} T^{-1} 2^j\right)$$

For (3.20) one uses (3.16), (3.17) and the similar results. Consider the term

$$\sum_{j>l}^{J_T} \sum_{k} \sum_{j'>l}^{J_T} \sum_{k'} cov\{\hat{\beta}_{jk}, \hat{\beta}_{j'k'}\} \psi_{jk}(x) \psi_{j'k'}(y)$$

$$= \frac{2\pi f_2(0)}{T} \sum_{j>l} \sum_{k} \psi_{jk}(x) \psi_{jk}(y) + \frac{1}{T^2} \sum_{j} \sum_{k} \sum_{j'} \sum_{k'} O(2^{j/2 + j'/2}) 2^{j/2} 2^{j'/2}$$

and the second term is $O(2^{2J_T} T^{-2})$.

Parts iii) and iv) follow likewise from the result iii) of Theorem I. One has

$$cum_m\{\hat{h}_0(x)\} = \sum_{a_1} \cdots \sum_{a_m} cum\{\hat{\beta}_{a_1}, \cdots, \hat{\beta}_{a_m}\} \psi_{a_1}(x) \cdots \psi_{a_m}(x)$$

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

$$= O\left(T^{-m+1} \left[\sum_{a} 2^{j(1 - 1/m)}\right]^m\right)$$

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

Proof of Corollary. This follows from (3.19), Assumption 2' and Tandori's result given at the end of Section 3.1. One counts cases up to and including order j and remembers that the number of k, is, for a given j, bounded.

The proofs of Theorems III and IV follow from a sequence of lemmas.

LEMMA 2. Suppose the assumptions of Theorem III hold. Then for any $\varepsilon > 0$, and T sufficiently large

$$Prob\,\{\,|\hat{\beta}_{jk} - \mu_{jk}| \ge \delta_j\sigma_{jk}\,\} \le 2\exp\{-\delta_j^2/(1+\varepsilon)2\}$$

with $\mu_{jk} = E\{\hat{\beta}_{jk}\}$, $\sigma_{jk}^2 = var\,\{\hat{\beta}_{jk}\}$.

Proof. This result follows as the proofs of Lemmas P4.7 and P4.10, pages 405-407, in Brillinger (1975). Specifically from iii) of Theorem I

$$|cum_m\{\hat{\beta}_{jk}\}| \le A^m C_m 2^{j(m/2-1)} T^{-m+1}$$

and so

$$|cum_m\{2^{-j/2} T\hat{\beta}_{jk}\}| \le A^m C_m 2^{-j} T$$

Therefore

$$|\log E\{\,\exp\{2^{-j/2}T\alpha(\hat{\beta}_{jk} - \mu_{jk})\}\} - 2^{-j}T^2\alpha^2\sigma_{jk}^2/2|$$

$$\le 2^{-j}T\alpha^2 \sum_{m=3}^{\infty} |\alpha|^{m-2} A^m C_m \frac{1}{m!}$$

$$\le 2^{-j}T^2\sigma_{jk}^2\alpha^2\varepsilon/2$$

for α sufficiently small. Here (3.16) and that $f_2(0) \ne 0$ have been used. This gives

$$Prob\,\{\,|\hat{\beta}_{jk} - \mu_{jk}| \ge \delta_j\sigma_{jk}\,\} \le 2\exp\{-2^{-j/2}T\alpha\delta_j\sigma_{jk}\}\exp\{2^{-j}T^2\sigma_{jk}^2\alpha^2(1+\varepsilon)/2\}$$

$$\le 2\exp\{-\delta_j^2/(1+\varepsilon)2\}$$

for the choice $\alpha = \delta_j/2^{-j/2}T\sigma_{jk}(1+\varepsilon)$ which is to be small, and will be so as it has been assumed that $2^{j/2}\delta_j = o(T^{1/2})$.

To continue to develop the proof of Theorem III write

$$\hat{h}(x) - h(x) =$$

$$\sum(\hat{\alpha}_{lk} - \alpha_{lk}) + \sum{}'(\hat{w}_{jk}\hat{\beta}_{jk} - \beta_{jk})\psi_{jk}(x) + \sum{}''\hat{w}_{jk}\hat{\beta}_{jk}\psi_{jk}(x) \qquad \text{(A.1)}$$

with $\sum'$ over the (finite number of) terms with $\hat{\beta}_{jk} \ne 0$ and $\sum''$ over the terms with $\beta_{jk} = 0$.

LEMMA 3. Under the assumptions of Lemma 2

$$E\,|\textstyle\sum''| \le \frac{C}{\sqrt{T}} \sum_{j>J_0}^{J_T} 2^{j/2} \exp\{-\delta_j^2/(1+\bar{\varepsilon})2\} \qquad \text{(A.2)}$$

for some $C > 0$.

Proof. Consider the expected value of the general term in $\sum''$ of (A.1). By Schwarz's inequality it is less or equal to

$$E\{\,|\hat{w}_{jk}\hat{\beta}_{jk}\psi_{jk}(t)|\,\} \le Prob\{\,|\hat{\beta}_{jk}| \ge \delta_j\sigma_{jk}\}\,\sqrt{E\{\hat{\beta}_{jk}^2\}}\;2^{j/2}A_0$$

and one has the result from Lemma 2 and that $\sigma_{jk}^2 \le A_1/T$ for some $A_1 > 0$ following (3.16).

In the theorem, for $\varepsilon = \bar{\varepsilon}$, the righthand side of (A.2) is assumed to be $o(T^{-1/2})$, and one has therefore $T^{1/2}\sum'' = o_p(1)$.

Turning to the $\sum'$ term in (A.1), one has

LEMMA 4. Under assumptions of Theorem III and if $\beta_{jk} \ne 0$

$$\hat{w}_{jk} = O_p(T^{-1})$$

Proof. By definition

$$\hat{w}_{jk} = 1 - I(\,|\hat{\beta}_{jk}| < \delta_j\sigma_{jk})$$

and

$$\begin{aligned} Prob\{\,|\hat{\beta}_{jk}| < \delta_j\sigma_{jk}\} &\le Prob\{\,|\,|\hat{\beta}_{jk} - \mu_{jk}| - |\mu_{jk}|\,| \le \delta_j\sigma_{jk}\} \\ &\le Prob\{\,|\mu_{jk}| - \delta_j\sigma_{jk} \le |\hat{\beta}_{jk} - \mu_{jk}|\,\} \\ &\le \sigma_{jk}^2\,/\,(|\mu_{jk}| - \delta_j\sigma_{jk})^2 \end{aligned}$$

for $\delta_j\sigma_{jk} < |\mu_{jk}|$ via the result of Lemma 2. (That $\delta_j\sigma_{jk} < |\mu_{jk}|$ for T sufficiently large follows from the assumption $2^{j/2}\delta_j = o(T^{1/2})$ and $\sigma_{jk} = O(T^{-1/2})$). This gives the lemma.

Next,

Proof of Theorem III. Using Lemmas 3, 4 and (A.1)

$$\sqrt{T}\,(\hat{h}(x) - h(x)) =$$

$$\sum_k \sqrt{T}\,(\hat{\alpha}_{lk} - \alpha_{lk})\phi_{lk}(x) + \sum_{j>l}^{J_0} \sum_k \sqrt{T}\,(\hat{\beta}_{jk} - \beta_{jk}) w_{jk} \psi_{jk}(x)) + o_p(1) \tag{A.3}$$

The asymptotic normality follows from that of the $\hat{\alpha}$'s and $\hat{\beta}$'s given in Theorem I. Only a finite number of terms are involved in (A.3).

Consideration now turns to the proof of Theorem IV.

LEMMA 5. Under Assumptions of the theorem and particularly that the $\beta_{J_T+1,k} = 0$ for $k = 1, \cdots, K_T$

$$E\{\hat{f}_2(0)\} = f_2(0) + O(2^{J_T} T^{-1})$$

$$var\{\hat{f}_2(0)\} = \frac{2}{K_T} f_2(0)^2 + O(K_T^{-1} 2^{J_T} T^{-1}) + O(2^{2J_T} T^{-2}) \tag{A.4}$$

Proof. The first result follows from (3.17).

For the second,

$$var\{\hat{f}_2(0)\} = var\{\frac{T}{2\pi}\sum_k \hat{\beta}^2_{J_T+1,k}/K_T\}$$

$$= (\frac{T}{2\pi K_T})^2 \sum_k \sum_{k'} cov\{\hat{\beta}_k^2, \hat{\beta}_{k'}^2\}$$

$$= (\frac{T}{2\pi K_T})^2 \sum_k \sum_{k'} 2cov\{\hat{\beta}_k, \hat{\beta}_{k'}\}^2$$

$$= (\frac{T}{2\pi K_T})^2 \left\{ 2\sum_k \left[\frac{2\pi f_2(0)}{T} + O(2^{J_T} T^{-2}) \right]^2 + 2\sum_{k \neq k'} O(2^{J_T} T^{-2})^2 \right\}$$

from (3.16,17). The result then follows.

This estimate of the power spectrum is seen to be consistent provided $2^{J_T}T^{-1}, K_T^{-1} \to 0$ as $T \to \infty$.

LEMMA 6. Under the assumptions of the theorem and given $0 < \omega, \varepsilon_1 < 1$,

$$Prob\{\hat{f}_2(0) \le \omega E\{\hat{f}_2(0)\}$$

$$\le 2\sqrt{1+(1-\omega)(1-\varepsilon_1)\sqrt{K_T}} \exp\{-(1-\omega)(1-\varepsilon_1)\sqrt{K_T}/2\} \qquad \text{(A.5)}$$

for T sufficiently large.

Proof. Ponomarenko (1978) develops the following bound for quadratic forms, Q, in normal variates

$$Prob\{|Q - E\{Q\}| > x\sqrt{var\{Q\}}\} \le 2\sqrt{1+x\sqrt{2}} \exp\{-x/\sqrt{2}\}$$

for $x > 0$. Now

$$Prob\{\hat{f} \le \rho E\{f\}\} \le Prob\{|\hat{f} - E\{\hat{f}\}| \ge (1-\rho)E\{\hat{f}\}\}$$

From Lemma 5, for T sufficiently large

$$\frac{E\{\hat{f}\}}{\sqrt{var\{\hat{f}\}}} \ge (1-\varepsilon_1)\sqrt{K_T/2}$$

and one has the result (A.5) from Ponomareko's inequality.

Next,

LEMMA 7. Under the conditions of the theorem

$$T^{1/2}E\{\Sigma''\} = o(1)$$

Proof.

$$T^{1/2}E\{|\Sigma''|\} \le T^{1/2}\sum_{j>J_0}^{J_T} Prob\{|\hat{\beta}_{jk}| \ge \delta_j\hat{\sigma}\}\frac{A_1}{\sqrt{T}}2^{j/2}$$

for T sufficiently large. Now for $0 < \rho < 1$

$$Prob\{|\hat{\beta}_{jk}| \ge \delta_j\hat{\sigma}\} \le Prob\{|\hat{\beta}_{jk}| \ge \rho\delta_j\sigma_{jk}\} + Prob\{\hat{\sigma} \le \rho\sigma_{jk}\}$$

Following Lemma 2, for any $\bar{\varepsilon} > 0$, the first term on the right here is

$$\leq 2 \exp\{-\rho^2\delta_j^2/(1+\bar{\varepsilon})2\} = 2 \exp\{-\delta_j^2/(1+\varepsilon)2\}$$

with the choice $\bar{\varepsilon} = \varepsilon/2$, $\rho^2 = (1+\bar{\varepsilon})/(1+\varepsilon)$.

Given $\varepsilon_2 > 0$ and T sufficiently large the second term on the right is bounded by $Prob\{\hat{f}_2(0) \leq \rho(1+\varepsilon_2)E\{\hat{f}_2(0)\}$ and one can use (A.5) with the choice $(1-\rho(1+\varepsilon_2))(1-\varepsilon_1) = \varepsilon_0$. The lemma now follows under the conditions of the theorem.

Next one considers $\sum'$. It contains a finite number of terms and one wishes to replace $\hat{w}_{jk}$ by 1 if $\beta_{jk} \neq 0$.

LEMMA 8. Under the assumptions of the theorem and if $\beta_{jk} \neq 0$

$$\hat{w}_{jk} = 1 + o_p(T^{-1}) + o_p(K_T^{-1})$$

Proof. Follows as did Lemma 4. Specifically with $\rho_1 > 1$

$$Prob\{\,|\hat{\beta}_{jk}| < \delta_j\hat{\sigma}_{jk}\}$$

$$\leq Prob\{\,|\hat{\beta}_{jk}| < \rho_1\delta_j\sigma_{jk}\} + Prob\{\hat{\sigma}_{jk}^2 > \rho_1^2\sigma_{jk}^2\}$$

As in the the proof of Lemma 4, the first term on the right is $o_p(T^{-1})$. Now for T sufficiently large

$$Prob\{\hat{\sigma}^2 > \rho_1^2\sigma_{jk}^2\}$$

$$\leq Prob\{\,|\hat{\sigma}^2 - E\{\hat{\sigma}^2\}| > |\rho_1^2\sigma_{jk}^2 - E\{\hat{\sigma}^2\}|\,\}$$

$$\leq var\{\hat{\sigma}^2\} \;/\; |\rho_1^2\sigma_{jk}^2 - E\{\hat{\sigma}^2\}|^2$$

and the result follows from (A.4) of Lemma 5.

And finally,

Proof of Theorem IV. Same as the proof of Theorem III, but using the immediately preceding lemmas.

REFERENCES

ALTMAN, N. S. (1990). Kernel smoothing of data with correlated errors. Journal of American Statistical Association 85 749-759.

ANDERSON, T. W. (1971). The Statistical Analysis of Time Series. Wiley, New York.

BENEDETTO, J.J. and FRAZIER, M. W. Eds. (1994). Wavelets. CRC Press, Boca Raton.

BJERVE, S., DOKSUM, K. A. and YANDELL, B. S. (1985). Uniform confidence bounds for regression based on a simple moving average. Scandinavian Journal of Statistics 12 159-169.

BLOW, D. M. and CRICK, F. H. C. (1959). The treatment of errors in the isomorphous replacement method. Acta Crystalographica 12 794-802.

BRILLINGER, D. R. (1975). Time Series: Data Analysis and Theory. Holt, New York.

BRILLINGER, D. R. (1993). Some examples of the use of the Fourier transform in the analysis of scientific data. Tech. Report 407, Statistics Dept., Univ. of Calif. Berkeley.

BRILLINGER, D. R. (1994). Some river wavelets. Environmetrics. To appear.

BROCKMAN, M., GASSER, T. and HERRMANN, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. J. Amer. Statist. Assoc. 88 1302-1309.

CHIU, S. T. (1989). Bandwidth selection for kernel estimate with correlated noise. Statistics Probability Letters 8 347-354.

CLEVELAND, W. S. , GROSSE, E. and SHYU, W.M. (1991). Local regression models. Pp. 309-376 in Statistical Models in S (eds. J. M. Chambers and T. J. Hastie). Wadsworth, Pacific Grove.

DAUBECHIES, I. (1992). Ten Lectures on Wavelets. SIAM, Philadelphia.

DIGGLE, P. J. and HUTCHINSON, M. F. (1989), On spline smoothing with autocorrelated errors. Australian Journal of Statistics 83 86-101.

DONOHO, D. (1993a). On minimum entropy segmentation. Preprint, Stanford University .

DONOHO, D. (1993b). Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. Proceedings of Symposia in Applied Mathematics, American Mathematical Society.

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

DONOHO, D. L. and JOHNSTONE, I. M. (1990). Wavelets and optimal nonlinear function estimates. Technical Report 281, Statistics Department, University of California, Berkeley.

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Adapting to unknown smoothness via wavelet shrinkage, J. Amer. Statist. Assoc. To appear.

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: asymptopia? J. Roy. Statist. Soc. B 57. To appear.

EUBANK, R. L. and SPECKMAN, P. L. (1993). Confidence bounds in nonparametric regression. Journal of American Statistical Association 88 1287-1301.

GALLANT, A. R. and GOEBEL, J. J. (1976). Nonlinear regression with autocorrelated errors. Journal of American Statistical Association 76 961-967.

GERSCH, W. (1992). Smoothness priors. Pp. 113-146 in New Directions in Time Series, Part II (eds. D. R. Brillinger et al.). Springer, New York.

GRENANDER, U. and ROSENBLATT, M. (1957). Statistical Analysis of Stationary Time Series. Wiley, New York.

HALL P., and PATIL, P. (1993). On wavelet methods for estimating smooth functions. Preprint, Australian National University.

HANNAN, E. J. (1970). Multiple Time Series. Wiley, New York.

HANNAN, E. J. (1971). Non-linear time series regression. Journal of Applied Probability 8 767-780.

HANNAN, E. J. (1973). Central limit theorems for time series regression. Zeitshrift für Wahrscheinlichkeitstheorie und Gebiete 26 157-170.

HANNAN, E. J. (1979). The central limit theorem for time series regression. Journal of Stochastic Proccesses Applications 9 281-289.

HÄRDLE, W. and PHAM, D. T. (1986). Some theory of M-smoothing of time series. Journal of Time Series Analysis 7 191-204.

HÄRDLE, W. and VIEU, P. (1992). Kernel regression smoothing of time series. J. Time Series Anal. 13 209-232.

HART, J. D. (1989). Kernel regression with time series errors. Journal of Royal Statistical Society B 53 173-187.

HART, J. D. (1994). Automated kernel smoothing of dependent data by using time series cross-validation. J. R. Statist. Soc. B 56 529-542.

HASTIE, T. J. and Tibshirani, R. J. (1990). Generalized Additive Models. Chapman and Hall, London.

KERKYACHARIAN, G. and PICARD, D. (1992). Density estimation in Besov spaces. Statistics and Probability Letters 13 15-24.

KING, N. (1972). An alternative for the linear regression equation when the predictor variable is uncontrolled and the sample size is small, J. Amer. Statist. Assoc. 67 217-219.

KITAGAWA, G. (1987). Non-Gaussian state space modelling of nonstationary time series. Journal of American Statistical Association 82 1032-1063.

KRONMAL, R. A. and TARTER (1968) The estimation of probability densities and cumulatives by Fourier series methods. Journal of American Statistical Association 63 925-952.

MEYER, Y. (1993). Wavelets: Algorithms and Applications. SIAM, Philadelphia.

MULLER, H.-G. (1988). Nonparametric Regression Analysis of Longitudinal Data. Springer, New York.

MULLER, H.-G. (1992). Change-points in nonparametric regression analysis. Ann. Statist. 20 737-761.

MULLER, H.-G. and STADTMULLER, U. (1987). Variable bandwidth kernel estimators of regression curves. Ann. Statist. 15 182-201.

POLYA, G. and SZEGO, G. (1925). Aufgaben and Lehrsatz aus der Analysis I. Springer, Berlin.

PONOMARENKO, L. S. (1978). Inequalities for distributions of quadratic forms in normal random variables. Theory Prob. Appl. 23 652-656.

RICE, J. and ROSENBLATT, M. (1983). Smoothing splines: regression, derivatives and deconvolution. Annals of Statistics 11 141-156.

ROBINSON, P. M. (1972). Non-linear regression for multiple time-series. Journal of Applied Probability 9 758-768.

ROSENBLATT, M. (1959). Statistical analysis of stochastic processes with stationary residuals. pp. 246-275 in Probability and Statistics (ed. U. Grenander). Wiley, New York.

SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting. Journal of Royal Statistical Society B 47 1-21.

STANISWALIS, J. (1989). Local bandwidth selection for kernel estimates. J. Amer. Statist. Assoc. 84 284-287.

STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9 1135-1151.

STRICHARTZ, R. S. (1993). How to make wavelets. American Mathematical Monthly 100 539-556.

TANDORI, V. K. (1959). Uber die orthogonalen Funktionen. VII (Approximationssätze), Acta Sci Math (Szegeg) 20 19-24.

THOMPSON, J. R. (1968). Some shrinkage techniques for estimating the mean. J. Amer. Statist. Assoc. 113-122.

TRUONG, Y. K. (1991). Nonparametric curve estimation with time series errors. Journal of Statistical Planning Inference 28 167-183.

TUKEY, J. W. (1979). Introduction to the dilemmas and difficulties of regression. Preprint, Princeton University.

WALTER, W. W. (1992). Approximation of the delta function by wavelets. Journal of Approximation Theory 71 329-343.

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

WU, J. S. and CHU, C. K. (1993). Kernel-type estimators of jump points and values of a regression function. Ann. Statist. 21 1545-1567.

ZYGMUND, A. (1959). Trigonometric Series Vol. II, Cambridge press, London.