WAVELETS IN STATISTICS

Pedro A. Morettin
Institute of Mathematics and Statistics
University of São Paulo
São Paulo, Brazil

Text for a Tutorial in The Third International Conference on Statistical Data Analysis Based on The L_1-Norm and Related Methods. Neuchâtel, Switzerland, August 11-15, 1997

Contents

Preface

In these notes we present the main uses of wavelets in statistics. Advances in the area are occuring rapidly, with applications in several fields. There are several books about the mathematical aspects of wavelets, like those by Daubechies(1992) and Chui(1992), and others in areas like signal processing and image compression.

The content of these notes borrows heavily from recent works, most of them referenced in the bibliography section. We decided to not include the proofs of the theorems, which the interested reader may find in the original papers.

I would like to thank the organizers of the Third International Conference on Statistical Data Analysis Based on the L_1-Norm and Related Methods for the invitation to give a tutorial on wavelets. This gave me the opportunity to prepare this text. Comments and suggestions are welcome and can be sent to **pam@ime.usp.br**. I would like to thank Chang Chiann for several conversations and David Brillinger for the continuous support.

A version in portuguese was presented as a short course in the Seventh Time Series and Econometrics School, in Canela, RS, Brazil and is part of a book to be published by the University of São Paulo Press.

São Paulo, June 1997.

Pedro A. Morettin

Chapter 1

Preliminaries

1.1 Introduction

The aim of these notes is to give an introduction to wavelets, which we think as contemporary tools with relevant uses in statistics. They are alternatives to other basis functions, like sines and cosines, orthogonal polynomials, Walsh functions, etc.

Meyer(1993) gives a nice hystorical retrospective of the wavelets, from Fourier(1807) to Haar(1910), Calderón(1960), Grossmann and Morlet(1980) and recent developments. The approach is the one of signal processing, potentially more attractive for the applications.

Fourier analysis is today a well established and powerful tool for the analysis of stationary time series. After the influential work of Blackman and Tukey(1959) and the introduction of the fast Fourier transform(FFT) by Cooley and Tukey(1965), the field of spectral analysis experienced a continuous progress. Recent references are Brillinger(1981), Brockwell and Davis(1991) and Percival and Walden(1993).

In the seventies there was an increasing interest in the use of other orthogonal systems(see Harmuth,1969, for example). The Walsh-Fourier analysis(Morettin,1974) became a useful tool for the analysis of categorical time series, or "square waves". The concept of frequency is replaced by that of "sequency", which gives the number of "zero crossings" of the unit interval. Further references are Morettin(1981), Stoffer et al.(1988) and Stoffer(1991).

Another aspect related to the wavelets is that researchers working in distinct areas, eventually not communicating, were responsible for the main developments. The above mentioned book by Meyer is a good reference for the reader, as well as the singular book by Barbara Hubbard(1996), written for non-mathematicians.

The basic fact about wavelets is that they are *localized* in time(or space), contrary to what happens with the trigonomteric functions. This behaviour makes then suitable for the analysis of nonstationary signals, those containing transients and fractal-type structures. Fourier bases are localized in frequency, but not in time: small changes in few observations can change almost all the components of a Fourier expansion.

The idea in Fourier or wavelet analysis is to approximate a function through a linear combination of sines and cosines or wavelets, respectively. For a comparable approximation, functions with different degrees of smoothness(containing peaks and discontinuities, for example) will require less wavelets than sines and cosines in their representations.

In Fourier analysis, every square integrable, periodic function of period 2π, that is, belonging to $L^2(0, 2\pi)$, is generated by a superposition of complex exponentials, $w_n(x) = e^{inx}, n = 0, \pm 1, \ldots$, obtained by dilations of the basic function $w(x) = e^{ix}$: $w_n(x) = w(nx)$. Our objective is to extend this idea to $L^2(\Re)$, that is, to generate this space from a single function, say ψ. This is accomplished by setting

$$\psi_{a,b}(x) = |a|^{-1/2}\psi(\frac{x-b}{a}),\ a > 0, -\infty < b < \infty. \tag{1.1}$$

The function ψ is called the *mother wavelet* and usually we take special values for a and b: $a = 2^{-j}, b = k2^{-j},\ j, k \in Z = \{0, \pm 1, \ldots\}$.

In the field of statistics, wavelets were used in the estimation of densities, nonparametric regression, estimation of the spectrum of stationary processes and of the evolutionary spectrum of non-stationary processes. Besides, they were used in applications in a great variety of situations, ranging from economics to medicine, astronomy, oceanography, etc.

1.2 Fourier Analysis

The key result here is that any periodic function $f(t)$, of period 2π and square integrable, can be written as a linear combination of sinusoidal components, each with a given frequency, namely

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{int} \tag{1.2}$$

where the equality is understood in the quadratic mean sense and the c_n are the *Fourier coefficients*, given by

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int} dt. \tag{1.3}$$

As we already remarked, the functions $w_n(t) = e^{int}$ form an orthonormal basis, generated by dilations of the *sinusoidal wave*

$$w(t) = e^{it} = \cos t + i \sin t, \tag{1.4}$$

which is the known Euler formula. In this context the Parseval relation is given by

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t)|^2 dt = \sum_{n=-\infty}^{\infty} |c_n|^2, \tag{1.5}$$

and this tells us that there exists an isometry between the spaces $L^2(0, 2\pi)$ of the square integrable functions and the space of square summable sequences, that is, $\sum_n |c_n|^2 < \infty$.

In Figure 1.1(taken from Stoffer, 1991) we have some trigonometric and Walsh functions.

The equation (1.2) can also be written as

$$f(t) = \frac{1}{2} a_0 + \sum_{n=1}^{\infty} \{a_n \cos nt + b_n \sin nt\}, \tag{1.6}$$

where the coefficients a_n, b_n are related with the c_n and are given by

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos ntdt, \quad n \geq 0 \tag{1.7}$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin ntdt, \quad n \geq 1. \tag{1.8}$$

The harmonic of (1.6) for $n = 1$ is called *fundamental* and corresponds to a sinusoidal wave of period equals to the period of $f(t)$, for $n = 2$ we have the *first harmonic*, of period equals to half the period of $f(t)$, etc.

The situation described above corresponds to the case of t continuous and (1.2) is the *Fourier series* of $f(t)$. In the case of $f(t)$ non-periodic, we obtain the Fourier integral corresponding to (1.2) and the Fourier transform, in correspondence to (1.3). For a treatment of the four possible situations see Percival and Walden(1993). For the case of Fourier analysis of stochastic processes see Priestley(1981).

We now describe briefly the content of these notes. In Chapter 2 we give the basics on wavelets. The discrete wavelet transform is described in Chapter 3. Chapter 4 introduces the nonparametric setting and the fundamental work of Donoho and co-authors on thresholding techniques. In Chapter 5 we introduce briefly some concepts for stationary processes, but the objective will be the estimation of the spectrum via wavelets. In Chapter 6 we consider nonstationary processes and an approach to estimate the evolutionary spectrum using wavelets.

FIG. 1.1. Trigonometric and Walsh functions.

Chapter 2

Wavelets

2.1 Introduction

In this chapter we present the basics about wavelets, without going into deeper details about the mathematics involved. By analogy with the Fourier analysis, consider the space $L^2(\Re)$ of all square integrable functions on $\Re$. Here the functions $f(t)$ must decay to zero, as $|t| \to \infty$, hence the exponentials (1.4) do not belong to this espace. The functions that generate $L^2(\Re)$ must decay rapidly to zero. The idea is then to consider dilations and translations of a single function ψ in order to cover $\Re$. Therefore, consider the wavelets

$$\psi_{j,k}(t) = \psi(2^j t - k),\ j,k \in Z, \tag{2.1}$$

which shows that $\psi_{j,k}(t)$ is obtained from $\psi(t)$ by a *binary dilation* 2^j and a *dyadic translation* $k2^{-j}$. This is a special case of (1.1). The functions $\{\psi_{j,k}(t), j,k \in Z\}$ form a basis that is not necessarily orthogonal. The advantage of working with orthogonal bases is to allow the perfect reconstruction of a signal from the coefficients of the transform, that is, we have a concise transform and each coefficient is calculated as the scalar product of the signal and the basis function(e^{int} in the case of Fourier analysis). Hence the interest in considering bases that are orthogonal. Meyer proved, in 1985, that is possible to construct such bases.

Consider, then, an orthonormal basis generated by ψ:

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k),\ j,k \in Z, \tag{2.2}$$

such that

$$< \psi_{j,k}, \psi_{\ell,m} >= \delta_{j,\ell}\delta_{k,m},\ j,k,\ell,m \in Z, \tag{2.3}$$

and for any function $f(t)$ of $L^2(\Re)$,

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{j,k}\psi_{j,k}(t). \tag{2.4}$$

The convergence in (2.4) must also be understood in quadratic mean. We say that (2.4) is a *wavelet series* de $f(t)$ and the *wavelet coefficients* are given by

$$c_{j,k} =< f, \psi_{j,k} >= \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)dt. \tag{2.5}$$

The Parseval relation analogous to (1.5) also holds here. Some further properties are:

(P1) $\int_{-\infty}^{\infty} \psi(t)dt = 0$(admissibility).

This condition is equivalent to $\hat{\psi}(0) = 0$(where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$) and this is essentially equivalent to

$$C_\psi = 2\pi \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty.$$

The constant C_ψ appears in the so-called "resolution of the identity" formula. See Chapter 3.

(P2) The first $r-1$ moments of ψ are null, that is,

$$\int_{-\infty}^{\infty} t^j \psi(t)dt = 0,\ j = 0, 1, \ldots, r-1,$$

for some $r \geq 1$ and

$$\int_{-\infty}^{\infty} |t^r \psi(t)|dt < \infty.$$

The value of r is tied to the degree of smoothness of ψ: the bigger r is, the smoother is ψ.

It is convenient to introduce a *scaling function*, or *father wavelet*, ϕ, which is a solution of the equation

$$\phi(t) = \sqrt{2} \sum_k \ell_k \phi(2t - k). \tag{2.6}$$

This function generates an orthonormal family of $L^2(\Re)$,

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k), \; j, k \in Z. \tag{2.7}$$

Then, the mother wavelet ψ can be obtained from ϕ through

$$\psi(t) = \sqrt{2} \sum_k h_k \phi(2t - k), \tag{2.8}$$

where

$$h_k = (-1)^k \ell_{1-k}, \tag{2.9}$$

the so-called "quadrature mirror filter relation". Actually, ℓ_k and h_k are the coefficients of "low-pass" and "high-pass" filters, called "quadrature mirror filters", used for the computation of the discrete wavelet transform. These coefficients are given by

$$\ell_k = \sqrt{2} \int_{-\infty}^{\infty} \phi(t) \phi(2t - k) dt, \tag{2.10}$$

$$h_k = \sqrt{2} \int_{-\infty}^{\infty} \psi(t) \phi(2t - k) dt. \tag{2.11}$$

The equations (2.6) and (2.8) are called *dilation equations*.Typically, ϕ and ψ are bounded and have compact support.

It is convenient to consider the orthonormal system

$$\{\phi_{j,k}(t), \psi_{j,k}(t), j, k \in Z\}_{j \geq j_0, k}, \tag{2.12}$$

in such a way that we can write, for $f(t) \in L^2(\Re)$,

$$f(t) = \sum_k c_{j_0,k} \phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_k d_{j,k} \psi_{j,k}(t), \tag{2.13}$$

where

$$c_{j_0,k} = \int_{-\infty}^{\infty} f(t) \phi_{j_0,k}(t) dt, \tag{2.14}$$

$$d_{j,k} = \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)dt. \tag{2.15}$$

In (2.13), j_0 is the "coarsest scale". Also, in practice, the continuous signal $f(t)$ can be approximated by a finite sum. We will come back to this matter later.

FIG. 2.1. (a) Haar and Daublets (b) Coiflets

FIG. 2.1. (c) Symmlets

FIG. 2.1. (d) Modulated Gaussian, Mexican Hat and Shannon wavelets.

Example 2.1. The oldest and simplest example of a wavelet is the Haar function,

$$\psi^{(H)}(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & \text{otherwise,} \end{cases} \tag{2.16}$$

for which case the scaling function is $\phi(t) = 1, 0 \leq t \leq 1$ and

$$\psi_{j,k}^{(H)}(t) = \begin{cases} 2^{j/2}, & 2^{-j}k \leq t < 2^{-j}(k+1/2) \\ -2^{j/2}, & 2^{-j}(k+1/2) \leq t < 2^{-j}(k+1) \\ 0, & \text{otherwise.} \end{cases} \tag{2.17}$$

For the Haar wavelet, equations (2.6) become

$$\phi(t) = \phi(2t) + \phi(2t-1) = \frac{1}{\sqrt{2}}\sqrt{2}\phi(2t) + \frac{1}{\sqrt{2}}\sqrt{2}\phi(2t-1),$$

and therefore $\ell_0 = \ell_1 = 1/\sqrt{2}$. Analogously, $h_0 = -h_1 = 1/\sqrt{2}$.

Example 2.2. Besides (2.16), the most frequently used orthonormal, compactly supported wavelets, are the "Daublets", "Symmlets" and "Coiflets", introduced by Daubechies. They are all non-symmetric and the "Symmlets" are the "least asymmetric" and used as default in some packages. The term "Coiflet" was coined by Daubechies in honor of R. Coifman.

Example 2.3. Some other wavelets of interest are the following.

(i) Modulated Gaussian(or Morlet) wavelet, given by

$$\psi(t) = e^{i\omega_0 t}e^{-t^2/2}.$$

This is a complex function, for some fixed ω_0.

(ii) The second derivative of the Gaussian(or mexican hat), given by

$$\psi(t) = (1-t^2)e^{-t^2/2}.$$

(iii) The Shannon wavelet, given by

$$\psi(t) = \frac{\sin(\frac{\pi t}{2})}{\frac{\pi t}{2}}\cos(\frac{3\pi t}{2}).$$

In Figure 2.1 we have the plots of some of the wavelets presented in the examples above. The "Daublets"(Daubechies Extremal Phase) and the "Symmlets"(Daubechies Least Asymmetric) were obtained with the use of the software *wavethresh*, developed by Nason and Silverman(1994). The "Coiflets" are from Bruce and Gao(1996), obtained from the software S+WAVELETS. The number N that appears in the figures is related with the support of the wavelet. In the case of the "extremal phase", this support is $[N-1, N]$.

In (2.2) and (2.7), the scale factor 2^{-j} is called the *dilatation factor* and j is the *level* associated with the scale 2^{-j}. The translation $2^{-j}k$ is called *location* and k is the *translation index* associated to $2^{-j}k$.

2.2 Multiresolution Analysis

Formally, a multiresolution analysis(MRA) is an increasing sequence of closed subspaces $\{V_j, j \in Z\}$ approximating $L^2(\Re)$, that is,

(MR1) $\ldots \subset V_{-1} \subset V_0 \subset V_1 \subset \ldots$

(MR2) $L^2(\Re) = \overline{\bigcup_j V_j}$.

(MR3) $\bigcap_j V_j = \{0\}$.

It can be proved that $\{\phi_{j,k}, k \in Z\}$ spans V_j. Moreover,

(MR4) $f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}, \ \forall j$.

V_0 is spanned by $\{\phi(x-k), k \in Z\}$. If V_j is spanned by $\{\phi_{j,k}\}$, then V_{j+1} will be spanned by $\{\phi_{j+1,k}\}$ and hence $\phi_{j+1,k}(t) = \sqrt{2}\phi_{j,k}(2t)$.

Since $V_0 \subset V_1$, any $f \in V_0$ can be written as a linear combination of wavelets $\sqrt{2}\phi(2t-k)$ from V_1. In particular (2.6) holds.

Consider now the subspaces W_j, with $W_j \perp V_j$ and such that

(MR5) $V_{j+1} = V_j \oplus W_j$.

Define $\psi(t)$ by (2.8). Then it can be shown that $\{\psi(2t-k)\}$ is an orthogonal basis for W_1 and, in general, $\{\psi_{j,k}, k \in Z\}$ form an orthogonal basis for W_j. Since $\cup_j V_j = \cup_j W_j$ is dense in $L^2(\Re)$, the functions $\{\psi_{j,k}, j, k \in Z\}$ form an orthogonal basis for $L^2(\Re)$.

Given f of $L^2(\Re)$, there exists J such that $f_J \in V_J$ approximates f. If $g_i \in W_i, f_i \in V_i$, by (MR5),

$$f_J = f_{J-1} + g_{J-1}$$

and repeating the argument

$$f \simeq f_J = g_{J-1} + g_{J-2} + \ldots + g_{J-M} + f_{J-M}. \tag{2.18}$$

We say that (2.18) is a wavelet decomposition of f . Notice that f_{J-M} is a linear combination of the $\phi_{J-M,k}$'s and the g_j's are linear combinations of the $\psi_{j,k}$'s, j=J-M,...,J-1.

2.3 Two-dimensional Wavelets

We will need to analyze, in Chapter 6, the time-dependent spectrum $f(u,\omega)$ of a non-stationary process, which is a function of two variables: u, the normalized time and ω, the frequency. Besides the traditional estimators, we can use estimators based on two-dimensional(briefly, 2-d) wavelets. These are also adequate for the analysis of images and matrices.

There are two possibilities for constructing 2-d wavelet bases:

(a) construct a 2-d basis, with a single scale, from the MRA of two one-dimensional(1-d) bases;

(b) construct a 2-d basis as the tensor product of two 1-d bases with distinct scales for each dimension.

For simplicity, consider $L^2(U)$, where $U=[0,1]$ and suppose that we have a compactly supported, orthonormal basis generating $L^2(U)$. As before let V_j and W_j spanned by the wavelets $\phi_{j,k}$ and $\psi_{j,k}, k\in Z$, respectively. Similarly to condition MR2 of the previous section, we have that

$$L^2(U\times U)=L^2(U^2)=\overline{\bigcup_{j=\ell}^{\infty} V_j\otimes V_j}, \tag{2.19}$$

which shows that we can construct a basis for $L^2(U^2)$ through tensor products of 1-d bases, $\{\phi_{\ell,k},\psi_{j,k}, j\geq \ell, k\}$.

For each j^*, we can write $V_{j^*}^{(2)}=V_{j^*}\otimes V_{j^*}$ in two ways:

$$V_{j^*}^{(2)}=(V_\ell\oplus W_\ell\oplus\cdots\oplus W_{j^*-1})\otimes(V_\ell\oplus W_\ell\oplus\cdots\oplus W_{j^*-1})=V_\ell\otimes V_\ell\oplus \left(\sum_{j=\ell}^{j^*-1}(W_j\otimes V_\ell)\right)\oplus\left(\sum_{j=\ell}^{j^*-1}(V_\ell\otimes W_j)\right)\oplus\left(\sum_{j_1,j_2=\ell}^{j^*-1}(W_{j_1}\otimes W_{j_2})\right), \tag{2.20}$$

or, alternatively,

$$V_{j^*}^{(2)}=V_\ell\otimes V_\ell\oplus\bigoplus_{j=\ell}^{j^*-1}[(V_j\otimes W_j)\oplus(W_j\otimes V_j)\oplus(W_j\otimes W_j)]. \tag{2.21}$$

From (2.20) a possible basis for $L^2(U^2)$ is

$$\mathcal{B}_1 = \{\phi_{\ell,k_1}(x)\phi_{\ell,k_2}(y)\}_{k_1,k_2} \cup (\cup_{j_1 \geq \ell}\{\psi_{j_1,k_1}(x)\phi_{\ell,k_2}(y)\}_{k_1,k_2}) \cup$$
$$(\cup_{j_2 \geq \ell}\{\phi_{\ell,k_1}(x)\psi_{j_2,k_2}(y)\}_{k_1,k_2}) \cup (\cup_{j_1,j_2 \geq \ell}\{\psi_{j_1,k_1}(x)\psi_{j_2,k_2}(y)\}_{k_1,k_2}), \tag{2.22}$$

and from (2.21) we have another possible basis,

$$\mathcal{B}_2 = \{\phi_{\ell,k_1}(x)\phi_{\ell,k_2}(y)\}_{k_1,k_2} \cup$$
$$\bigcup_{j \geq \ell} \{\phi_{j,k_1}(x)\psi_{j,k_2}(y), \psi_{j,k_1}(x)\phi_{j,k_2}(y), \psi_{j,k_1}(x)\psi_{j,k_2}(y)\}_{k_1,k_2}. \tag{2.23}$$

Let us consider first the basis $\mathcal{B}_2$, which can be represented as

$$\mathcal{B}_2 = \{\Phi_{\ell,\mathbf{k}}(x,y),\ \mathbf{k} = (k_1,k_2)\}_{\mathbf{k}} \cup \{\Psi^{\mu}_{j,\mathbf{k}}(x,y),\ \mathbf{k} = (k_1,k_2), \mu = h,v,d\}_{j \geq \ell, \mathbf{k}}, \tag{2.24}$$

and we see that we have a father wavelet and three mother wavelets, in the horizontal, vertical and diagonal directions:

$$\begin{aligned}
\Phi_{\ell,\mathbf{k}}(x,y) &= \phi_{\ell,k_1}(x)\phi_{\ell,k_2}(y), \\
\Psi^{h}_{j,\mathbf{k}}(x,y) &= \phi_{j,k_1}(x)\psi_{j,k_2}(y), \\
\Psi^{v}_{j,\mathbf{k}}(x,y) &= \psi_{j,k_1}(x)\phi_{j,k_2}(y), \\
\Psi^{d}_{j,\mathbf{k}}(x,y) &= \psi_{j,k_1}(x)\psi_{j,k_2}(y).
\end{aligned} \tag{2.25}$$

The father wavelet represents the smooth part and the mother wavelets represent the details in the three mentioned directions.

A function f of $L^2(U^2)$ can then be written in the form

$$f(x,y) = \sum_{\mathbf{k}} c_{\ell,\mathbf{k}}\Phi_{\ell,\mathbf{k}}(x,y) + \sum_{j=\ell}^{\infty}\sum_{\mathbf{k}}\sum_{\mu=h,v,d} d^{\mu}_{j,\mathbf{k}}\Psi^{\mu}_{j,\mathbf{k}}(x,y), \tag{2.26}$$

with the wavelet coefficients given by

$$c_{\ell,\mathbf{k}} = \int_{U^2} f(x,y)\Phi_{\ell,\mathbf{k}}(x,y)dxdy, \tag{2.27}$$

$$d^{\mu}_{j,\mathbf{k}} = \int_{U^2} f(x,y)\Psi^{\mu}_{j,\mathbf{k}}(x,y)dxdy. \tag{2.28}$$

Let us turn now to the basis $\mathcal{B}_1$. We need four indices here, since the scale is not the same as in(2.24). With a slight abuse of notation, let $\psi_{\ell-1,k} = \phi_{\ell,k}$. Calling μ_I the basis functions, where $I = (j_1, j_2, k_1, k_2)$, we can write

$$\mathcal{B}_1 = \{\mu_I(x,y),\ I = (j_1, j_2, k_1, k_2)\}. \tag{2.29}$$

Notice that we have generally

$$\mu_I(x,y) = \psi_{j_1,k_1}(x)\psi_{j_2,k_2}(y), \tag{2.30}$$

but some of the wavelets $\psi_{j,k}$ are actually father wavelets $\phi_{j+1,k}$.

Therefore, a function f of $L^2(U^2)$ can be written alternatively as

$$f(x,y) = \sum_I d_I \mu_I(x,y), \tag{2.31}$$

with

$$d_I = \int_{U^2} f(x,y)\mu_I(x,y)dxdy. \tag{2.32}$$

We note that we can use two different bases, one for each direction. In the specific case of the evolutionary spectrum, we will use a 1-d basis in the time direction and a 1-d basis in the frequency direction.

2.4 Comments

The wavelets $\phi_{j,k}$ and $\psi_{j,k}$, given in (2.7) and (2.2), can be defined in a slightly different way, as in Daubechies(1992), for example. Specifically, we have

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \tag{2.33}$$

and

$$\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k). \tag{2.34}$$

Obvious modifications must be made in several places, as in (2.17) and in several formulas of the remaining chapters. In particular, this representation can be more convenient in some situations, as in Chapter 5.

Note that with the representations (2.33) and (2.34), the scale factor becomes 2^j and the translation factor is $k2^j$. With this notation, the lowest resolution levels correspond to the highest ones in the previous notation, and vice versa.

We close this chapter with some comments about the Fourier and wavelet analyses. One analogy is that, given a square integrable function, it can be written as a superposition of sines and cosines or wavelets(centered in a sequency of time points). The difference is that the functions of a wavelet basis are indexed by two parameters, while in the Fourier basis we have a single parameter, the frequency, which has a physical interpretation.

In a wavelet analysis, the parameters j and k represent scale and time localization, respectively. Intuitively, scale can be thought as "inverse frequency", but this connection is tenuous, as the following argument shows(Priestley,1996). Suppose the mother wavelet shows oscillations, as in a Coiflet, for example. As j increases(and hence the scale factor 2^{-j} decreases and there is a shrinking in time) the oscillations increase and the wavelet exhibits "high frequency". On the other hand, if j decreases(the scale increases and there is an expansion in time) the oscillations become slow and a "low frequency" behavior appears. This means that high frequency components are analyzed in short time intervals and low frequency components are analyzed over long time intervals, allowing the analysis of signals with transients and singularities. In other words, the wavelet coefficients characterize the "local" behavior of the signal, while the Fourier coefficients characterize the "global" behavior.

One way to try to achieve time-localization in Fourier analysis is to consider a windowed Fourier transform. Given the function $f(t)$, choose a function $g(t)$ which is concentrated around t and decays to zero at infinity and consider

$$F(t,\omega) = \int f(s)g(s-t)e^{-i\omega s}ds.$$

If $g(s;t,\omega) = g(s-t)e^{-i\omega s}$, this is similar to $\psi_{a,b}(t)$ in formula (3.1) of next chapter, with the difference that $\psi_{a,b}$ has a time-width adapted to its frequency(narrow for high frequencies and broad for low frequencies), while for g fixed, $g(s;t,\omega)$ has a constant width.

Chapter 3

The Wavelet Transform

3.1 Introduction

The wavelet transform appeared in its continuous form with the works of two french researchers, J. Morlet, a geophysicist and A. Grossmann, a theoretical physicist. See Morlet(1981,1983), Grossmann(1988), Grossmann and Morlet(1984,1985).

Y. Meyer, A. Grossmann and I. Daubechies introduced in 1986, a quasi-orthogonal complete discrete set of $L^2(\Re^n)$("wavelet frames"). Actually, Meyer wanted to prove that there was no orthogonal basis constructed from regular wavelets and was surprised when he got an orthogonal basis in 1985(Meyer, 1986,1987,1988). Daubechies(1988,1989) introduced the compactly supported regular wavelets and these are often used in practice.

For f in $L^2(\Re)$, the *continuous wavelet transform*, with respect to ψ, is defined by

$$(W_\psi f)(a,b) = |a|^{-1/2} \int_{-\infty}^{\infty} f(t)\psi(\frac{t-b}{a})dt, \; a,b \in \Re, \; a \neq 0. \tag{3.1}$$

Note that the transform is the inner product $< f, \psi_{a,b} >$ of f and the function $\psi_{a,b}$ given in (1.1). Then f can be reconstructed through

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} < f, \psi_{a,b} > \psi_{a,b}(t) \frac{dadb}{a^2} \tag{3.2}$$

where

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega. \tag{3.3}$$

Formula (3.2) is the "resolution of the identity". Several variations of (3.2) are possible, restricting a to positive values. See Daubechies(1992, chapter 2) for details.

Let us now restrict a and b to discrete values as follows: $a = a_0^m, b = nb_0a_0^m$, with a_0 and b_0 fixed values, $a_0 > 1$ and $b_0 > 0$. Then a formula analogous to (3.2) does not exist. The question is now if still we can recover f from the wavelet transform. It follows that for reasonable wavelets $\psi(t)$(basically, the admissibility condition is satisfied and ψ has a reasonable decay in time and frequency) and suitably chosen a_0 and b_0, there exist $\{\tilde{\psi}_{m,n}(t)\}$ such that

$$f(t) = \sum_m \sum_n \tilde{\psi}_{m,n}(t) < f, \psi_{m,n} >, \tag{3.4}$$

where $\{\tilde{\psi}_{m,n}(t)\}$ is the *dual* frame of the frame $\{\psi_{m,n}\}$ and

$$< f, \psi_{m,n} >= \int_{-\infty}^{\infty} f(t)\psi_{m,n}(t)dt \tag{3.5}$$

is the *discrete parameter wavelet transform*. See Appendix 3A and Daubechies (1992, chapter 3) for more details on frames.

If we choose $a_0 = 1/2, b_0 = 1$, there exists ψ such that $\psi_{m,n}$ form an orthonormal basis for $L^2(\Re)$ and for any function f in this space

$$f(t) = \sum_m \sum_n c_{m,n}\psi_{m,n}(t), \tag{3.6}$$

with

$$c_{m,n} = \int_{-\infty}^{\infty} f(t)\psi_{m,n}(t)dt. \tag{3.7}$$

In the next section we consider the case of a discrete time wavelet transform, when we observe a time series at a finite set on time points.

3.2 The Discrete Wavelet Transform

Suppose we have data $\mathbf{X} = (X_0, X_1, \ldots, X_{T-1})'$, and for the moment asssume that these may be from an i.i.d sample or from a stochastic process. From now on we assume that $T = 2^M, M > 0$ integer.

We define the *discrete wavelet transform* of $\mathbf{X}$, with respect to the mother wavelet ψ, as

$$d_{j,k}^{(\psi)} = \sum_{t=0}^{T-1} X_t \psi_{j,k}(t), \tag{3.8}$$

which we will denote simply $d_{j,k}$, omitting the dependence on ψ. This transform is computed for $j = 0, 1, \ldots, M-1$ and $k = 0, 1, \ldots, 2^j - 1$. Since we have T observations and in the enumeration above we have $T-1$ coefficients, we will denote the remaining one by $d_{-1,0}$(Donoho e Johnstone,1995a).

We can write the transform (3.8) in matrix form

$$\mathbf{d} = \mathbf{W}\mathbf{X}, \tag{3.9}$$

and assuming appropriate boundary conditions, the transform is orthogonal and we can invert it in the form

$$\mathbf{X} = \mathbf{W}'\mathbf{d}, \tag{3.10}$$

where $\mathbf{W}'$ denotes the transpose of $\mathbf{W}$.

The important fact here is that (3.8) is not obtained in practice through the matrix product indicated in (3.9) but through a pyramid algorithm with complexity $O(T)$, consisting of a sequence of low-pass and high-pass filters. See section 3.3. below.

The data vector $\mathbf{X}$ can be associated to a function f on the interval $[0, 1)$, defined by

$$f(t) = \sum_{k=0}^{2^M-1} X_k I_{\{k/2^M \le t < (k+1)/2^M\}}. \tag{3.11}$$

For example, if $T = 8 = 2^3$ and we consider the Haar wavelets, then

$$\begin{pmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \sqrt{2} & 0 & 2 & 0 & 0 & 0 \\ 1 & 1 & \sqrt{2} & 0 & -2 & 0 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 0 & 2 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 0 & -2 & 0 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & 2 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & -2 & 0 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & 2 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} d_{-1,0} \\ d_{0,0} \\ d_{1,0} \\ d_{1,1} \\ d_{2,0} \\ d_{2,1} \\ d_{2,2} \\ d_{2,3} \end{pmatrix}, \tag{3.12}$$

from which we can obtain $\mathbf{d}$. Notice that we can write in this case

$$f(t) = d_{-1,0}\phi(t) + \sum_{j=0}^{2} \sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t), \tag{3.13}$$

where $\phi(t) = 1$ is the scaling function for the Haar wavelet.

Usually we do not consider all the resolution levels, M, but a number J, that corresponds to the coarsest scale, 2^{-J}, or the smooth part of the data and we can write

$$f(t) = \sum_k c_{J,k}\phi_{J,k}(t) + \sum_{j=J}^{M-1} \sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t). \tag{3.14}$$

The relation (3.14) can be written in an alternative form(Donoho and Johnstone,1995a). Let

$$V_J\mathbf{X} = \sum_{j<J} d_{j,k}\psi_{j,k} \tag{3.15}$$

the partial reconstruction of the data from terms with gross structure and

$$W_j\mathbf{X} = \sum_{0\le k<2^j} d_{j,k}\psi_{j,k}, \;\; j \ge J \tag{3.16}$$

the partial reconstruction of the data from terms with resolution level j, or escale 2^{-j}. Then,

$$\mathbf{X} = V_J\mathbf{X} + \sum_{J\le j<M} W_j\mathbf{X}, \tag{3.10}$$

is the *multiresolution decomposition of* $\mathbf{X}$(Mallat,1989).

From (3.14) we can write $\mathbf{d} = (\mathbf{c}_J, \mathbf{d}_J, \mathbf{d}_{J+1}, \ldots, \mathbf{d}_{M-1})'$, where

$$\begin{aligned} \mathbf{c}_J &= (c_{J,0}, \ldots, c_{J,2^J-1})', \\ \mathbf{d}_J &= (d_{J,0}, \ldots, d_{J,2^J-1})', \\ &\vdots \\ \mathbf{d}_{M-1} &= (d_{M-1,0}, \ldots, d_{M-1,2^{M-1}-1})' \end{aligned} \tag{3.17}$$

are called *crystals* or *subbands*.

We remark that the $c_{J,k}$'s capture the low frequency oscillations, while $d_{j,k}$'s capture the high frequency oscillations. The coefficients $d_{M-1,k}$'s represent the fine scale(details) and the $c_{J,k}, d_{J,k}$ represent the coarsest scale(smooth).

FIG. 3.1. The Doppler function.

Example 3.1. In Figure 3.1 we have the plot of the function

$$f(t) = \sqrt{t(1-t)} \sin(2.1\pi/(t+0.05)), \quad 0 \leq t \leq 1, \tag{3.18}$$

known as “Doppler” and computed at T=1024 equally spaced points. In Figure 3.2 the wavelet coefficients are shown, computed with the *wavethresh* package(Nason and Silverman,1994). Here, $T = 2^{10}$, hence $M = 10$. The program computes the coeffcients for $M - 1 = 9$ levels, with level $j = 9$ being the finest one.

FIG. 3.2. Wavelet coefficients of the Doppler function.

3.3 The Pyramid Algorithm

As we already have remarked, the discrete wavelet transform is computed through an algorithm, due to Mallat(1989). Actually there are several pyramid algorithms. See Meyer(1993) for details. There is also an algorithm for the reconstruction of the data(synthesis) from the wavelet coefficients.

The algorithm uses low-pass $\{\ell_k\}$ and high-pass $\{h_k\}$ filters, with coefficients given by (2.10) and (2.11). Figure 3.3 illustrates the procedure. In the figure, L indicates the low-pass filter, H the high-pass filter and $\downarrow 2$ indicates a decimation by 2, that is, we delete every other value of the output.

The procedure starts with the data $c_0 = (c_{0,0}, \ldots, c_{0,T-1})'$, where $c_{0,i} = X_i$, $i = 0, 1, \ldots, T-1$. In the j–th step, the algorithm computes $c_{j,k}$ and $d_{j,k}$ from the smooth coefficients of level $j-1$, $c_{j-1,k}$, through

$$c_{j,k} = \sum_n \ell_{2k-n} c_{j-1,n},$$

$$d_{j,k} = \sum_n h_{2k-n} c_{j-1,n}.$$

If $T_j = T/2^j$, the output of the procedure is the set of detail coefficients $\mathbf{d}_j, j = 1, 2, \ldots, J$ and smooth coefficients $c_J = (c_{J,1}, \ldots, c_{J,T_J})'$.

FIG. 3.3. The pyramid algorithm.

3.4 Software

In this section we describe briefly some of the available computer packages Many others were developed and several are available via ftp or http. One way to get information on wavelets and related topics is to subscribe the Wavelet Digest(send the e-mail add@wavelet.org). The book by Wickerhouser(1994) is another source of programs for several purposes.

1. Wavethresh

This software was developed by Nason(1993) and can be obtained from the

StatLib library via anonymous ftp(ftp lib.stat.cmu.edu/S/wavethresh). See Nason and Silverman(1994) for an overview of the program.

The software operates with S-PLUS, and only a UNIX version is available. A new version, based on the concept of stationary wavelet transform, is under preparation. See Nason and Silverman(1995).

2. S+WAVELETS

This software was developed by Bruce and Gao(1994) to work within S-PLUS. UNIX and WINDOWS versions are available and distributed by MathSoft, Seattle, USA. The book by Bruce and Gao(1996) is a detailed guide of the program, which is quite complete, bringing a variety of routines for the discrete wavelet transform, wavelet packets, several possibilities for thresholds, etc.

3. WaveLab

This is a library of MATLAB routines for wavelet analysis, wavelet packet analysis, etc. It can be obtained via ftp (playfair.stanford.edu/pub/wavelab). There are versions for Macintosh, UNIX and WINDOWS. The package was developed by Buckheit et al.(1995,versão 0.700).

Appendix 3A

Frames

We have seen in (1.2) and (2.4) the representation of a function f in terms of functions of two orthonormal bases. The theory of frames provides the representation of a given function in terms of functions of a basis that is not necessarily orthonormal(not even linearly independent). An excellent account on frames is given by Daubechies(1992, chapter 3), on which we base what follows.

We say that the set of functions $\{\varphi_j, j \in J\}$ of a Hilbert space $\mathcal{H}$ constitutes a *frame* if there exist $A > 0, B < \infty$ such that, for any $f \in \mathcal{H}$,

$$A\|f\|^2 \leq \sum_{j \in J} | < f, \varphi_j > |^2 \leq B\|f\|^2. \tag{A.3.1}$$

A and B are called *frame bounds*. We say that the frame is *tight* if $A = B$ and in this case $\sum_j | < f, \varphi_j > |^2 = A\|f\|^2$ and we can recover f from $< f, \varphi_j >$, that is,

$$f = A^{-1} \sum_j < f, \varphi_j > \varphi_j. \tag{A.3.2}$$

If the frame is tight, $A = 1$ and $\|\varphi_j\| = 1, \forall j$, then $\{\varphi_j, j \in J\}$ is an orthonormal basis. In general, given a frame $\{\varphi_j\}$, there is a *dual* frame $\{\tilde{\varphi}_j\}$ such that

$$f = \sum_j < f, \varphi_j > \tilde{\varphi}_j, \tag{A.3.3}$$

for any $f \in \mathcal{H}$, which means that we have a reconstruction formula for f from $< f, \varphi_j >$. In this case, $f = \sum_j < f, \tilde{\varphi}_j > \varphi_j$.

If ψ is a mother wavelet and $\psi_{j,k}(t) = a_0^{-j/2}\psi(a_0^{-j}t - kb_0), j, k \in Z$, in order to recosntruct f from $< f, \psi_{j,k} >$ we require that $\psi_{j,k}$ form a frame. Not all choices of a_0, b_0 and ψ lead to wavelet frames, but for reasonable ψ and appropriate a_0, b_0 this will be the case.

Boundary Conditions

The discrete wavelet transform matrix $\mathbf{W}$ in (3.9) is applied to a finite set of data $\mathbf{X}$. This corresponds to a truncation of a theoretical wavelet transform operator applied to an infinite signal. This truncation makes the wavelet coefficients different for the finite and infinite transforms at the boundaries.

There are several rules to deal with this problem. See Bruce and Gao(1996) for details. The most used ones are the *periodic* and *reflection* rules.

In the periodic rule, the data are assumed to be T-periodic: $X_{t+Tk} = X_t$, $t = 0, 1, \ldots, T-1$, $k \in Z$.

In the reflection rule the original data are reflected at the boundaries and then periodically extended. Algorithms are available for this. An inconvenient of this rule is that it applies to biorthogonal wavelets, in order to get perfect reconstruction.

The periodic rule preserves orthogonality, besides being fast and numerically stable.

Chapter 4

Nonparametric Regression

4.1 Introduction

In this chapter we will introduce one of the most important applications of wavelets to statistics. We have here the main contributions of D. Donoho and co-workers, that were afterwards extended to other areas, as in the estimation of the spectrum of a stationary process.

The methodology is based in the principle of wavelet shrinkage, which aims to reduce(and even remove) the noise present in a signal, reducing the magnitude of the wavelet coefficients. The main works in this area are the fundamental articles of Donoho and Johnstone(1994,1995a, 1995b), Donoho et al.(1995,1996a) and Johstone and Silverman(1994).

Consider the model

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, 2, \ldots, T, \tag{4.1}$$

where $\epsilon_i \sim$ iid$N(0, \sigma^2)$ and $t_i = (i-1)/T$.

In the usual case of parametric regression, we assume that f follows a particular model, for example a linear function of p parameters $\theta_1, \ldots, \theta_p$. The estimation of the parameters is done through ordinary least squares. In the situation of model (4.1), we assume that f belongs to some class of functions, satisfying certain regularity(smoothness) properties. The traditional approach here uses estimators based on kernels, splines and Fourier expansions. An excellent review on nonparametric methods in time series is given by Härdle et al.(1997).

If $f_i = f(t_i)$ the purpose is to estimate $\mathbf{f} = (f_1, \ldots, f_T)'$, with the smallest mean square error. For a given class of functions $\mathcal{F}$, we want $\hat{\mathbf{f}} = \hat{\mathbf{f}}(y_1, \ldots, y_T)$ that reaches the minimax risk

$$R(T, \mathcal{F}) = \inf_{\hat{\mathbf{f}}} \sup_{\mathbf{f}} \mathcal{R}(\hat{\mathbf{f}}, \mathbf{f}) \tag{4.2}$$

where $\mathcal{R}(\hat{\mathbf{f}}, \mathbf{f}) = (1/T)E\|\hat{\mathbf{f}} - \mathbf{f}\|_2^2$ and $\|f\|_2^2 = \sum_{i=1}^T f_i^2$.

The shrinkage procedure consists of the three following steps:

[1] Take the discrete wavelet transform of the data $y_1, \ldots, y_T$, leading to the T wavelet coefficients $y_{j,k}$, which are contaminated by noise.

[2] Use thresholds to reduce the coefficients, making null those coefficients below a certain value. Several choices here are possible and we will discuss some of them in the next section. We obtain, in this stage, the coefficients without noise.
[3] Take the inverse wavelet transform of the coefficients in stage [2] to get the estimates $\hat{f}_i$.

The transform (3.3) applied to model (4.1) produces

$$\mathbf{Wy} = \mathbf{Wf} + \mathbf{W}\epsilon \tag{4.3}$$

and since W is orthogonal, white noise is transformed into white noise, that is, if $w_{j,k}$ are the wavelet coefficients of $f(t_i)$, we may write

$$y_{j,k} = w_{j,k} + \sigma z_{j,k} \tag{4.4}$$

where $z_{j,k} \sim$ iid $N(0, 1)$.

Therefore, (4.4) tells us that the wavelet coefficients of a noisy sample may be written as the noiseless coefficients plus white noise.

As in the case of Fourier transform, the wavelet transform establishes an isometry between two spaces, preserving risks. In other words, if $\hat{w}_{j,k}$ are estimates of $w_{j,k}$, there exists an estimate $\hat{\mathbf{f}}$ such that $\hat{\mathbf{f}} = \mathbf{W}'\hat{\mathbf{w}}$ and the Parseval relation implies that $\|\hat{\mathbf{w}} - \mathbf{w}\|_2 = \|\hat{\mathbf{f}} - \mathbf{f}\|_2$. In the other direction, if $\hat{\mathbf{f}}$ is an estimate of $\mathbf{f}$, then $\hat{\mathbf{w}} = \mathbf{W}\hat{\mathbf{f}}$ defines an estimate of $\mathbf{w}$ with isometric risk.

4.2 Choice of the Threshold

Here we have two problems: one is the choice of the threshold scheme(or policy) and the other is the choice of the parameters that govern this scheme.

4.2.1 Choice of the Scheme

Two schemes traditionally used are the following:

(i) Hard Threshlod, defined by

$$\delta_\lambda^H(x) = \begin{cases} 0, & \text{if } |x| \le \lambda \\ x, & \text{if } |x| > \lambda. \end{cases} \tag{4.5}$$

(ii) Soft Threshold, defined by

$$\delta_\lambda^S(x) = \begin{cases} 0, & \text{if } |x| \le \lambda \\ sign(x)(|x| - \lambda), & \text{if } |x| > \lambda. \end{cases} \tag{4.6}$$

FIG. 4.1. Hard and soft thresholds applied to $f(x) = x$.

Figure 4.1 shows how these two schemes act over the function $f(x) = x$. Notice that (i) is of the type "kill" or "preserve", while (ii) is of the type "kill" or "shrink"(by the amount λ).

The smooth policy may present large biases, while the hard produces small biases but larger variance. See Bruce and Gao(1996).

4.2.2 Choice of the Parameters

Besides the parameter λ, which appears in (4.5) and (4.6), we will have to estimate the parameter σ, the noise standard deviation. In general these parameters depend of level j, that is, for each scale we will have a threshold λ_j and a σ_j.

We make now a brief digression on the main proposals.

(P1) **Universal**

Donoho and Johnstone(1994) consider the model (4.1) and propose a minimax estimate for $\mathbf{f}$, called "RiskShrink", based on hard or soft thersholds and on a parameter λ_T obtained through the minimization of a theoretical upper limit of the asymptotic risk. They also suggest to use an alternate, called "VisuShrink", which has similar properties and is good visually. For this estimate the threshold parameter used is

$$\lambda_T = \lambda_{j,T} = \sigma\sqrt{2\log T}, \tag{4.7}$$

which *does not depend of the scale.* This threshold is called *universal* and comes as a default in many available software. The noise level σ is estimated as described below. See also Donoho et al.(1995) and section 4.3.

This threshold leads to estimates that underestimate f, since there is a tendency to eliminate or reduce too many coefficients, specially at high resolution levels.

(P2) **Adaptive**("SureShrink")

Suppose that the vector $\mathbf{x} = (x_1, \ldots, x_d)'$ is such that $x_i \sim N(\mu_i, 1)$ and $\mu = (\mu_1, \ldots, \mu_d)'$. Stein(1981) propposed to estimate the loss $\|\hat{\mu} - \mu\|^2$ in an unbiased way, even if $\hat{\mu}$ is an arbitrary, non-linear and biased estimator.

Using this idea, Donoho and Johstone(1995a) suggest a procedure, denominated "SureShrink", for which in each resolution level an unbiased estimate of Stein's risk is minimized("Sure= Stein unbiased risk estimator"). If in level j we have n_j coefficients, define the threshold

$$\lambda_j = \text{argmin}_{0\le t\le\sqrt{2\log(n_j)}} \text{SURE}(\mathbf{y}_j, t) \tag{4.8}$$

where

$$\text{SURE}(\mathbf{y}_j, t) = n_j - 2\sum_{k=1}^{n_j} I_{\{|y_{j,k}|\le t\sigma_j\}} + \sum_{k=1}^{n_j}\{\frac{y_{j,k}}{\sigma_j^2} \wedge t^2\}. \tag{4.9}$$

In (4.9), σ_j is the noise level for scale j.

(P3) **Cross-Validation**

Cross-validation is a technique for the estimation of the prediction error for a fitted model to the data. The prediction error measures the goodness of a model when it predicts a future observation. In regression models the prediction error is defined as

$$PE = E(\hat{y} - y)^2,$$

where y represents a future observation and e $\hat{y}$ is the prediction, using a particular regression model.

A possibility is to use the mean square error of the residuals as an estimate of the prediction error(PE), but this tends to be optmistic, or it underestimates the true PE, the reason being that we are using the same data to fit and assess the model.

Cross-validation uses part of the data to estimate the model and the remaining part to evaluate if the model is adequate or not. Usually, the cross-validation(CV) algorithm is as follows(Efron and Tibshirani,1993):

[1] If we have n observations $y_1, \ldots, y_n$, for each observation y_i, fit the model, leaving this observation out and compute the predicted value for this i-th data, denoted $\hat{y}_i^{-i}$.

[2] Compute

$$CV = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i^{-i})^2$$

as an estimate of the PE. CV requires that we fit the complete model n times.

An alternative is to split the data into k parts, $k \neq n$. This will be done below, since we will have in general $n = 2^M$ and if we remove one observation we get $2^M - 1$, which may not be a power of two.

Nason(1994) suggests to use CV to find a proper threshold, that minimizes the mean integrated error, defined by

$$M(t) = E \int_{-\infty}^{\infty} [\hat{f}_t(x) - f(x)]^2 dx, \tag{4.10}$$

where $\hat{f}_t$ denotes an estimator of f in model (4.1), if we use the threshold t. Since f is unknown we use an estimator of $M(t)$.

Consider the observations $y_1, \ldots, y_n$. The suggested algorithm is:

[1] The observations y_i with *odd* indexes are removed from the set, leaving 2^{M-1} observations with even indexes, and these are reindexed from $j =$

$1, \ldots, 2^{M-1}$.

[2] An estimator $\hat{f}_t^{even}$ is then constructed, using a particular threshold, from the reindexed y_j.

[3] Using the removed data with odd indexes, form an interpolated version of the odd-indexed data:

$$\overline{y}_j^{odd} = \begin{cases} 1/2(y_{2j-1} + y_{2j+1}), & \text{for } j = 1, \ldots, \frac{n}{2} - 1 \\ 1/2(y_1 + y_{n-1}), & \text{for } j = n/2. \end{cases}$$

Similar calculations are done for $\hat{f}_t^{odd}$ and $\overline{y}_j^{even}$. An estimate of $M(t)$ is then given by

$$\hat{M}(t) = \sum_j \left[\left(\hat{f}_{t,j}^{even} - \overline{y}_j^{odd} \right)^2 + \left(\hat{f}_{t,j}^{odd} - \overline{y}_j^{even} \right)^2 \right]. \tag{4.11}$$

We notice that the estimator of $M(t)$ is based in estimates of f_t computed with $n/2$ values. The universal threshold t_n has to be modified due to this fact. After the minimization of $M(t)$, this correction is applied. See Nason(1994) for details and simulation results comparing this procedure with the universal and a modification of SureShrink.

(P4) Donoho et al.(1995) use a variant of the universal rule, namely

$$\lambda_T = \sigma \sqrt{\frac{2 \log(T)}{T}}, \tag{4.12}$$

with similar properties.

(P5) Ogden and Parzen(1996) suggest to carry on multiple hypotheses tests: for each wavelet coefficient we test if it is zero or not.

Let us turn now to the problem of estimation of the parameter σ. Three possibilities may be considered.

(i) Estimate σ from the coefficients $y_{j,k}$ of the finest scale.

(ii) Consider all the scales and an estimator of σ based in all wavelet coefficients.

(iii) Estimate a σ_j for each scale.

In any case, we can use the variance of the coefficients or the estimator proposed by Donoho et al.(1995),

$$\hat{\sigma} = \text{median}\{|y_{M-1,k}| : 0 \leq k < 2^M\}/0.6745, \tag{4.13}$$

where $M - 1$ is the finest scale(in the case of using (i) above) and the factor 0.6745 is connected with the fact that $0.6745 < \Phi(1) - \Phi(-1)$.

FIG. 4.2. (a) The Doppler function plus Gaussian noise, $\sigma = 1$

Summarizing the shrinkage procedure [1]-[3], if $\mathbf{w} = \mathbf{W}\mathbf{f}$, $\mathbf{z} = \mathbf{W}\epsilon$, define

$$\hat{w}_{j,k} = \delta^S_{\lambda_T}(w_{j,k}) \tag{4.14}$$

as estimators of the coefficients $w_{j,k}$ and, therefore, the estimator of $\mathbf{f}$ will be $\hat{\mathbf{f}} = \mathbf{W}'\hat{\mathbf{w}}$.

Example 4.1. In Figure 4.2(a) we have the function obtained after adding Gaussian white noise(with zero mean and variance one) to the Doppler function of Figure 3.3. The reconstruction using the universal threshold is shown in Figure 4.2(b) and the reconstruction with SureShrink is shown in Figure 4.2(c).

FIG. 4.2. (b) Reconstruction with soft, universal; (c)SureShrink.

4.3 Minimax Estimators

In the classical context of estimation, we assume that we have a known class $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$ of probability distributions, indexed by a parameter $\theta \in \Omega$. For a given loss function $L(\theta, \delta)$, where δ is an estimator of θ, consider the risk

$$R(\theta, \delta) = E_\theta\{L(\theta, \delta(\mathbf{X}))\}, \tag{4.15}$$

for observations $\mathbf{X}$ of P_θ. We would like to find δ that minimizes R, for all possible values of θ. As we know, this problem has no solution, unless θ is a constant. One way of proceeding is to restrict the class of estimators, requiring properties of unbiasedness or equivariance, for example.

We can give up these requirements, but then we will have to accept optimal properties weaker than uniform risk. One possibility is to minimize the meam weighted risk, for some nonnegative weight function. This procedure leads us to Bayes estimators. Other possibility is to minimize the maximum risk, that is, to find $\hat{\theta}$ such that

$$R(\hat{\theta},\theta) = \inf_{\tilde{\theta}} \sup_{\theta} R(\tilde{\theta},\theta). \tag{4.16}$$

Such an estimator is called *minimax*. Usually it is not easy to find minimax estimators. See Lehmann(1983) for details.

An important case is when we have a sample $X_1,\ldots,X_n$ of i.i.d $N(\mu,\sigma^2)$ variables. Then(Wolfowitz,1950) $\overline{X}$ is the minimax estimator of μ.

In the nonparametric situation, of which (4.1) is an example, there are several results in the literature, but as in the parametric situation, it is not possible to obtain minimax estimators for any function f.

Three are the basic ingredients in the search for minimax estimators(Donoho et al., 1995):

[1] Suppose that f belongs to some specific functional space, $\mathcal{F}$, as for example Hölder, Sobolev, Besov, etc.

[2] We assume that we have a specific measure for the risk of the estimator, as for example

$$R_n(\hat{f},f) = E\|\hat{f}-f\|^2. \tag{4.17}$$

[3] We try to find an estimator that is *minimax* for $\mathcal{F}$ and R_n,

$$\sup_{\mathcal{F}}\{R_n(\hat{f},f)\} = \inf_{f} \sup_{\mathcal{F}} R_n(\hat{f},f), \tag{4.18}$$

or *assintotically minimax*, with equality in (4.18) replaced by $\sim$, for $n\to\infty$, or that *it attains the minimax convergence rate*, replacing equality in (4.18) by $\approx$, for $n\to\infty$, where $a_n \approx b_n$ means that $\text{liminf}_n \mid \frac{a_n}{b_n} \mid > 0$ and $\text{limsup}_n \mid \frac{a_n}{b_n} \mid < \infty$.

Several combinations of these ingredients were considered in the literature. See Donoho et al.(1995) for further information.

Let us consider now $X_i,\ i=1,\ldots,n$, such that

$$X_i = \theta_i + Z_i, \tag{4.19}$$

where Z_i are independent $N(0, \sigma^2)$, σ^2 known and we want to estimate θ_i. Here, the X_i's make the role of the coefficients $y_{j,k}$ in model (4.4). Let $\theta = (\theta_1, \ldots, \theta_n)'$ and we are interested in the situation that few of the θ_i are non null.

Suppose the ideal situation in which an oracle tells us which of the θ_i are null and we look for estimators under this assumption. Consider estimators of the form(diagonal projections-DP)

$$\delta_i = I_{\{|\theta_i| > \sigma\}} \tag{4.20}$$

for which the corresponding risk is

$$R(PD, \theta) = \sum_{i=1}^{n} (\theta_i^2 \wedge \sigma^2). \tag{4.21}$$

Clearly we cannot use (4.20) for practical purposes and (4.21) is unattainable, but we can use this risk as a benchmark to judge estimators constructed from the data. Let $\lambda_n = \sigma\sqrt{2\log n}$ and define the estimators

$$\hat{\theta}_i = \delta^S_{\lambda_n}(X_i) \tag{4.22}$$

based in (4.6). Then we have the following result.

Theorem 4.1.(Donoho and Johnstone,1994). For any $\theta \in \Re^n$,

$$E\|\hat{\theta} - \theta\|^2 \leq (1 + 2\log n)\{\sigma^2 + \sum_i (\theta_i^2 \wedge \sigma^2)\}. \tag{4.23}$$

that is, (4.23) tells us that the estimator $\hat{\theta}$ attains the benchmark $\sigma^2 + R(PD, \theta)$, up to a factor $O(\log n)$. The following result shows that this behavior cannot be essentially improved.

Theorem 4.2.(Donoho and Johnstone,1994). As $n \to \infty$,

$$\frac{1}{2\log n} \inf_{\tilde{\theta}} \sup_{\theta} \frac{E\|\tilde{\theta} - \theta\|^2}{\sigma^2 + \sum_i (\theta_i^2 \wedge \sigma^2)} \to 1. \tag{4.24}$$

These results can be applied to the models (4.1) and (4.4). Consider the estimator $\hat{\mathbf{f}}$ given below (4.14).

Corollary 4.1. For all $\mathbf{f}$ and all n,

$$R(\hat{\mathbf{f}},\mathbf{f}) \leq (1+2\log n)\{\sigma^2 + \sum_{j,k}(w_{j,k}^2 \wedge \sigma^2)\}. \tag{4.25}$$

Corollary 4.2. As $n \to \infty$,

$$\frac{1}{2\log n} \inf_{\tilde{f}} \sup_{f} \frac{E\|\tilde{\mathbf{f}} - \mathbf{f}\|^2}{\sigma^2 + \sum_{j,k}(w_{j,k}^2 \wedge \sigma^2)} \to 1. \tag{4.26}$$

4.4 Model with Stationary Noise

Suppose now that in model (4.1), the errors ϵ_i constitute a Gaussian stationary process, with zero mean and $\boldsymbol{\epsilon}=(\epsilon_1,\ldots,\epsilon_n)' \sim N(\mathbf{0},\boldsymbol{\Gamma}_n)$. Assume that the covariance matrix $\boldsymbol{\Gamma}_n$ is circulant with elements $\gamma_{|s-t|}^{(n)}$. If we call $\mathbf{V}_n$ the covariance matrix of $\mathbf{z}=\mathbf{W}\boldsymbol{\epsilon}$, then

$$\mathbf{V}_n = \mathbf{W}\boldsymbol{\Gamma}_n\mathbf{W}'. \tag{4.27}$$

By the stationarity of the errors, the variance of the coefficients $z_{j,k}$ will depend only on j. Call $s_j^2 = \mathrm{Var}(z_{j,k})$, for all j.

We consider now thresholds for the coefficients $y_{j,k}$ that are level-dependent, of the form

$$\lambda_j = s_j\sqrt{2\log n}, \tag{4.28}$$

and define the estimator

$$\hat{w}_{j,k} = \delta_{\lambda_j}^S(y_{j,k}), \tag{4.29}$$

assuming that the s_j are known. Finally, estimate $\mathbf{f}$ by

$$\hat{\mathbf{f}} = \mathbf{W}'\hat{\mathbf{w}}. \tag{4.30}$$

Johnstone and Silverman(1994) prove results that are similar to (4.23) and (4.24). Suppose that

$$(\gamma_0^{(n)})^{-2}\sum_{j=0}^{n-1}(\gamma_j^{(n)})^2 \leq c_1 < \infty \tag{4.31}$$

and

$$k_n = \frac{\gamma_0}{n}\sum_{j=0}^{n-1}(\tilde{\gamma}_j)^{-1} \leq c_2 < \infty \tag{4.32}$$

where $\tilde{\gamma}_j$ is the Fourier transform of γ_j. Then,

Theorem 4.3.(Johnstone and Silverman,1994) For all $\mathbf{f} \in \Re^n$,

$$E\|\hat{\mathbf{f}} - \mathbf{f}\|^2 \leq (1 + 2\log n)\{\gamma_0^{(n)} + \sum_{j,k}(w_{j,k}^2 \wedge s_j^2)\}. \tag{4.33}$$

Theorem 4.4.(Johstone and Silverman,1994). The estimator $\hat{\mathbf{f}}$ is such that

$$\liminf_{n\to\infty}(2\log n)^{-1}k_n \inf_{\hat{\mathbf{f}}}\sup_{\mathbf{f}}\frac{E\|\hat{\mathbf{f}} - \mathbf{f}\|^2}{\gamma_0^{(n)} + \sum_{j,k}(w_{j,k}^2 \wedge s_j^2)} \geq 1. \tag{4.34}$$

The meaning is the same as in the case of white noise: the estimator is nearly minimax up to a factor $O(\log n)$ and cannot be improved. Since s_j is usually not know it has to be estimated as before.

The case of estimation of functions in a model with fractional Gaussian noise was considered by Wang(1996). This situation contemplates the case of regression with long memory.

Chapter 5

Stationary Processes

5.1 Introduction

In this chapter we will consider the main applications of wavelets to the theory of stationary processes. We will deal basically with the estimation of the spectrum of such a process, using thechniques involving thresholds. The principal results are due to Gao(1993) and Moulin(1994) for Gaussian processes and to Neumann(1996) for general stationary processes.

Eventhough the wavelets have a bigger potential of applications in the non-stationary case, we will consider also a time-scale approach to stationary processes. This may be a starting situation before embarking in the non-stationary case.

We will assume that the reader knows the basic concepts of stationary processes. The references here are Priestley(1981), Brilinger(1981) and Brockwell and Davis(1991).

Let $\{X_t, t \in Z\}$ be a discrete stationary process, with zero mean, autocovariance function $\gamma(\tau)$, supposed to be absolutely summable, and spectrum(or spectral density function)

$$f(\omega) = \frac{1}{2\pi} \sum_{\tau} \gamma(\tau) e^{-i\omega\tau}, \quad -\pi \leq \omega \leq \pi. \tag{5.1}$$

Then $f(\omega)$ is real, non-negative, periodic, with period 2π and uniformly continuous. An initial estimator of the spectrum, from which other estimators are obtained by smoothing, is the *periodogram*

$$I(\omega) = |d(\omega)|^2, \tag{5.2}$$

where $d(\omega)$ is the *discrete Fourier transform*,

$$d(\omega) = \frac{1}{\sqrt{2\pi T}} \sum_{t=0}^{T-1} X_t e^{-i\omega t}, \tag{5.3}$$

computed at the Fourier frequencies $\omega_j = 2\pi j/T,\ j = 0, 1, \ldots, [T/2]$, from the data $\mathbf{X} = (X_0, \ldots, X_{T-1})'$. The transform (5.3) is usually computed via the FFT algorithm, which has complexity $O(T\ell n\, T)$, if T is an integer power of two.

Under regularity conditions it can be shown that $d(\omega)$ has asymptotically a complex normal distribution, with asymptotic variance $E|d(\omega)|^2 \sim f(\omega)$, and this shows that $I(\omega)$ is an asymptotically unbiased estimator of $f(\omega)$, though not consistent(its variance is of the order $f(\omega)^2$). Moreover, ordinates of this transform calculated at a finite number of Fourier frequencies are asymptotically independent, which is the key fact to construct smoothed estimators.

5.2 The Wavelet Spectrum

Consider an orthonormal wavelet basis

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k), \quad j, k \in Z, \tag{5.4}$$

obtained from the mother wavelet ψ, which we suppose fixed in what follows. The analysis will depend on such a function, contrary to the analysis based in (5.3). This section is based in Chiann and Morettin(1996).

The discrete wavelet transform is written as

$$d_{j,k} = \sum_{t=0}^{T-1} X_t \psi_{j,k}(t), \tag{5.5}$$

with $T = 2^M$ observations $\mathbf{X} = (X_0, X_1, \ldots, X_{T-1})'$ obtained from the stationary process X_t. Now (5.5) will be computed for $j = 0, 1, \ldots, M-1$ and $k = 0, 1, \ldots, 2^{M-j} - 1$, giving $T - 1$ coefficients, plus c_{00} corresponding to the scaling function. Notice the change of notation in (5.4).

It follows that

$$E\{d_{j,k}\} = 0$$

and

$$Var\{d_{j,k}\} = \sum_{t=0}^{T-1}\sum_{s=0}^{T-1} \gamma(t-s)\psi_{j,k}(t)\psi_{j,k}(s)$$

$$= \sum_{u=-(T-1)}^{T-1} \gamma(u) \sum_{t=0}^{T-1-|u|} \psi_{j,k}(t)\psi_{j,k}(t+ \mid u \mid). \quad (5.6)$$

Define

$$\Psi_{j,k}(u) = \sum_{t=0}^{\infty} \psi_{j,k}(t)\psi_{j,k}(t+ \mid u \mid) \quad (5.7)$$

as the *wavelet autocorrelation function* at (j,k).

Consider the following assumptions, that will be used in the sequel.

(A1) $\sum_u |\gamma(u)| < \infty$.

(A2) $\sum_u (1+|u|)|\gamma(u)| < \infty$.

(A3) $\sum_{u_1} \cdots \sum_{u_{k-1}} |u_j||C_k(u_1,\ldots,u_{k-1})| < \infty$,

where $C_k(u_1,\ldots,u_{k-1}) = \mathrm{Cum}\{X_{t+u_1},\ldots,X_{t+u_{k-1}},X_t\}$ is the cumulant of order k of X_t, $u_1,\ldots,u_{k-1} = 0,\pm1,\ldots,\ k \geq 2$.

Under (A2), the *wavelet spectrum of the process* $\{X_t, t \in Z\}$, *with respect to* ψ, is defined by

$$\eta_{j,k}^{(\psi)} = \sum_{u=-\infty}^{\infty} \gamma(u)\Psi_{j,k}(u) \quad j,k \in Z. \quad (5.8)$$

From now on we will omitt the upper index ψ. It follows that if (A2) is satisfied, then

$$Var\{d_{j,k}\} \to \eta_{j,k}, \quad \text{as} \quad T \to \infty. \quad (5.9)$$

Under the assumption (A2), $\eta_{j,k}$ is bounded and nonnegative.

Example 5.1. In Figure 5.1 we have the three-dimensional plots of the wavelet spectrum for: (a) a moving average process of order one, MA(1), given by

$X(t) = \epsilon(t) - 0.7\epsilon(t-1)$, where $\epsilon(t)$ is a normal white noise series with mean zero and variance one; (b) an autoregressive process of order one, AR(1), given by $X(t) = 0.8X(t-1) + \epsilon(t)$, with $\epsilon(t)$ as in (a). In both cases the Haar wavelet was used.

FIG. 5.1. Wavelet spectrum: (a) MA(1) (b) AR(1)

For the Haar wavelet, $d_{j,k}^{(H)}$ is given by

$$d_{j,k}^{(H)} = 2^{-j/2}[\sum_{t=k2^j}^{2^j(k+1/2)-1} X_t - \sum_{t=2^j(k+1/2)}^{2^j(k+1)-1} X_t],$$

which are re-scaled differences of local means of $X(t)$.

Let us turn now to investigate the covariance structure of the discrete wavelet transform. Define

$$\eta_{(j,j'),(k,k')} = \sum_{u=-\infty}^{\infty}\sum_{t=0}^{\infty}\gamma(u)\psi_{j,k}(t+ \mid u \mid I_{\{u>0\}})\psi_{j',k'}(t+ \mid u \mid I_{\{u\leq 0\}}), \quad (5.10)$$

for $j, j', k, k' \in Z$, as the *(asymptotic)covariance of the wavelet transform.* Then, we have the following result.

Theorem 5.1. Under (A1) and (A2),

(i) $E\{d_{j,k}d_{j',k'}\} \rightarrow \eta_{(j,j'),(k,k')}$, as $T \rightarrow \infty$;

(ii) If $j = j'$, $k = k'$, then $\eta_{(j,j'),(k,k')} = \eta_{j,k}^{(\psi)}$.

(iii) $E\{d_{j,k}d_{j',k'}\} = O(1)$, as $T \rightarrow \infty$, for all (j,k) and (j', k').

This theorem shows that the asymptotic covariance of the discrete wavelet transform at two distinct pairs (j,k) and (j',k') is not necessarily zero, which contrasts with the case of the discrete Fourier transform. In some sitiuations we may have independence. Also, under regularity conditions, which include mixing conditions on the process, a central limit theorem for the wavelet transform can be proved. See Chiann and Morettin(1996) for details.

The relation (5.9) suggests that $\eta_{j,k}$ can be estimated by the statistic

$$I_{j,k} = (d_{j,k})^2 = [\sum_{t=0}^{T-1} X_t \psi_{j,k}(t)]^2. \tag{5.11}$$

which is called the *wavelet periodogram* of the values $X_0, \ldots, X_{T-1}$. The following result is valid.

Theorem 5.2. Assume (A2) satisfied. Then:

$$E[I_{j,k}] = \eta_{j,k} + O(T^{-1}), \quad \text{as} \quad T \to \infty. \tag{5.12}$$

If (A3) is also satisfied for $k = 4$, then

$$Cov\{I_{j,k}, I_{j',k'}\} = 2\{\eta_{(j,j'),(k,k')}\}^2 + O(1), \quad \text{as} \quad T \to \infty. \tag{5.13}$$

In particular, if $j = j'$, $k = k'$, we get the variance

$$Var[I_{j,k}] = 2\{\eta_{j,k}\}^2 + O(1). \tag{5.14}$$

Theorem 5.2 shows that the periodogram is not consistent. The periodogram gives the "energy" of the process for each (j, k). We can be interested in the distribution of this energy for each scale. For this, consider the *scalegram in scale* j as being

$$S(j) = \sum_{k=0}^{2^{(M-j)}-1} [d_{j,k}]^2 = \sum_{k=0}^{2^{(M-j)}-1} I_{j,k}, \quad j = 0, \ldots, M-1. \tag{5.15}$$

See Scargle(1993) and Arino and Vidakovic(1995) for applications of this concept. Under the assumptions of the theorem 5.2 we can prove that

$$E\{S(j)\} = \eta_{j,\cdot} + O(2^{M-j}T^{-1}), \tag{5.16}$$

$$Cov\{S(j),S(j^{'})\} = 2\sum_{k=0}^{2^{(M-j)}-1}\sum_{k^{'}=0}^{2^{(M-j^{'})}-1}\{\eta_{(j,j^{'}),(k,k^{'})}\}^2 + O(1), \tag{5.17}$$

where

$$\eta_{j,\cdot} = \sum_{k=0}^{2^{(M-j)}-1}\eta_{j,k}.$$

The quantity $\eta_{j,\cdot}$ can be thought as the wavelet energy associated to scale j and $S(j)$ its estimator.

We now reproduce two examples: the first shows the local character of the wavelet transform as compared with the Fourier transform; the second, analyzes a real example.

Example 5.2.(Percival, 1993). Figure 5.2 illustrates two time series $\{X_t^{(i)}, t = 1,2,\ldots,16\}, i = 1,2$, which differ only in two points: $X_{14}^{(2)} = -X_{14}^{(1)}$. In the same figure we have the periodogram (5.2), in this case calculated for $\omega_j,\ j = 0,1,\ldots,8$. Both series have the same "energy", and by the Parseval identity we have

$$\sum_{t=1}^{16} X_t^2 = \sum_{j=0}^{8} I(\omega_j).$$

FIG. 5.2. (a) Two time series series, $X_t^{(1)}$ and $X_t^{(2)}$.
(b) Fourier Periodograms : $X_t^{(1)}$(thick line), $X_t^{(2)}$(thin line).

In Figure 5.3 we have the Fourier synthesis, that is , the reconstruction based in the Fourier coefficients $d_j = d(\omega_j)$. Therefore, this figure shows $d_j e^{-i\omega_j t},\ t = 1,\ldots,16$. We see that the change in a single value implies the change in all, except three, of the Fourier coefficients, while all the ordinates of the periodogram were changed. This shows that the Fourier coefficients are not well "localized" in time, or that each d_j is determined by values of the series that are well dispersed in time.

FIG. 5.3. Fourier synthesis.

FIG. 5.4. Wavelet synthesis, using Haar wavelets.

Figure 5.4 gives the wavelet synthesis, using the Haar wavelet. We see, now, that only one coefficient in each scale is changed, showing that the wavelet basis is localized in time. Figure 5.5 brings a time $\times$ scale plot, with the values of $I_{j,k}$ represented in a grey scale: the darker, the larger the importance of a given scale. In the example, only the grey tones above $t = 14$ are different from plot to plot.

FIG. 5.5. Time $\times$ scale plots: (a) $X_t^{(1)}$ e (b) $X_t^{(2)}$.

Example 5.3.(Chiann and Morettin,1996) Stoffer et al.(1988) analyzed EEG records of infants born from mothers who abstained from alcohol during pregnancy and mothers who used moderate amounts of alcohol. Data were collected for a study of the effects of moderate maternal alcohol consumption on neonatal EEG sleep patterns. A detailed description of the study design can be found in the above paper. The analysis was performed using the Haar wavelet and the software *wavethresh*.

Figure 5.6(a) shows the record from an infant(during120 minutes), whose mother was not exposed to alcohol("unexposed"). The EEG for an infant of an "exposed" mother is shown in Figure 5.6(b). The sleep state is categorized into six states: 1: quiet sleep–trace alternant; 2: quiet sleep–high voltage; 3:

indeterminate sleep; 4: active sleep–low-voltage; 5: active sleep– mixed; 6: awake.

The wavelet transforms of *unexposed* and *exposed* are presented in Figures 5.7(a) and 5.7(b), respectively. Figures 5.8(a) and 5.8(b) show their wavelet periodograms. In the figures, Resolution Level corresponds to level $(M - j)$; in this case, $M = 7$. After applying the universal threshold to the wavelet coefficients we obtain the threshoded periodograms in Figure 5.9. Images of these periodograms are presented in Figures 5.10. These periodograms tell us how the energy in a time series is decomposed on a scale by scale basis at different times. The figures show, for example, that the periodograms of higher scales contribute more to energy in our time series than lower scales. Moreover, the smallest values of the periodograms are present at scale 1, which indicate the presence of noise in time series. On the other hand, in comparing the two periodograms, note that there are larger values of the periodograms at higher scales in Figure 5.9(b) than in Figure 5.9(a), and the distribution of the energy for each level j in Figure 5.9(b) is more uniform than in Figure 5.9(a). Moreover, at higher levels, the energy in Figure 5.9(a) is concentrated at lower values of k, while in Figure 5.9(b) is about uniform for each k.

FIG. 5.6. Sleep states of infants: (a) Unexposed; (b) Exposed.

FIG. 5.7. Wavelet coefficients(Haar): (a) Unexposed; (b) Exposed.

In Figure 5.11 we have the reconstructed sleep states from the thresholded wavelet coefficients. The scalegrams are presented in Figure 5.12. We note that there is a strong peak in the neighborhood of a high level ($j = 5$) in these two scalegrams, indicating that a large period is present in both series. However, the peak presented in unexposed is much stronger than in exposed, indicating that the energy of this series is concentrated at level 5 and lower values of k, according to the analysis above. We also note the different behavior of the scalegrams at low and high scales.

The Walsh-Fourier analysis of this data is given in Stoffer et al (1988). To detect whether alcohol affects the infant′s sleep we suggest to use a wavelet analysis of variance to compare the two groups of infants. This research will be presented elsewhere.

FIG. 5.8. Wavelet periodograms: (a) Unexposed, (b) Exposed.

FIG. 5.9. Thresholded periodograms: (a) Unexposed, (b) Exposed

FIG. 5.10. Time-scale plots: (a) Unexposed, (b) Exposed

FIG. 5.11. Reconstructed sleep states

FIG 5.12. Scalegrams of Unexposed and Exposed

5.3 Estimation of the Fourier Spectrum

The purpose here is to estimate the spectrum of a real stationary process, of zero mean, given by (5.1), using non-linear estimation techniques based on the thresholding of empirical wavelet coefficients.

In the spectral estimation through Fourier methods we search for a compromise between resolution(bias) and stability(variance). The smoothed estimators based on spectral windows are appropriate for the estimation of functions with a homogeneous degree of smoothness. The presence of peaks, corners and transients causes the known problems in the classical spectral analysis.

From a theoretical point of view, cosine bases are optimal for L^2–Sobolev spaces, while wavelet bases(also optimal for these spaces) are optimal for a wide range of spaces, including Besov spaces, which contain functions that are not smooth in a homogeneous way, that is, can be smooth in part of the domain but less smooth in other parts.

The two main requirements to be satisfied by estimators based on orthogonal expansions in order to have high spatial adaptation are(Neumann, 1996):

(a) the basis should be well localized;

(b) the estimator should discriminate noise from variance.

Fourier bases do not have property (a) for example. The idea is to use this technique with high power of local adaptation to the problem of spectrum estimation.

We consider first the case of a Gaussian stationary process. We follow Gao(1993) and Moulin(1994), and mention also Wahba(1980). The purpose is to estimate the log-spectrum $g(\omega) = \log f(\omega)$.

Suppose that we have the observations $X_0, X_1, \ldots, X_{2T-1}$, with $T = 2^M$, $M > 0$ integer. Wahba proposed the following model:

$$\log I(\omega_j) = \log f(\omega_j) + \gamma + \epsilon_j, \quad j = 1, \ldots, T-1, \tag{5.18}$$

where $\epsilon_j = \log(\eta_j/2) - E\{\log(\eta_j/2)\}$, $\eta_j \sim \chi_2^2$. It can be proved that $Var(\epsilon_j) = \pi^2/6$.

The proposed non-linear procedure to estimate $g(\omega)$ is the following:

[1] Compute the log-periodogram

$$g_\ell = \log I(\omega_\ell), \quad \ell = 0, 1, \ldots, T-1,$$

for frequencies $\omega_\ell = \frac{2\pi\ell}{2T}$.

[2] Consider the wavelet transform of g_ℓ, and get the empirical wavelet coefficients $\{y_{j,k}\}, j = 0, 1, \ldots, M-1, k = 0, 1, \ldots, 2^j - 1$.

[3] Apply the soft threshold

$$\delta_\lambda^S(x) = \text{sign}(x)(|x| - \lambda)_+, \tag{5.19}$$

to the coefficients $\{y_{j,k}\}$, with threshold parameters dependent on the level j, $\lambda = \lambda_{j,T}$, calculated by

$$\lambda_{j,T} = \frac{\alpha \log T}{2^{(M-j-1)/4}}, \tag{5.20}$$

for fine resolution levels(j=M-1,M-2,...), where α is a constant(basically the supremum of the mother wavelet ψ) and

$$\lambda_{j,T} = \pi\sqrt{\frac{\log T}{3}}, \tag{5.21}$$

for the coarse resolution levels($j << M-1$). Notice that we are using again the notation (2.2).

[4] Take the inverse wavelet transform, producing an estimator $\hat{g}_\ell^*$ of the log-spectrum in frequency ω_ℓ.

Let us comment briefly the rational for using (5.20) and (5.21).

For normal wavelet coefficients, the universal threshold

$$\lambda_{j,T} = \lambda_T = \sigma\sqrt{2\log T}, \tag{5.22}$$

where σ^2 is the noise variance, satisfies certain interesting properties, as already mentioned. Under regularity conditions, for resolution levels $j << M-1$, the coefficients $\{y_{j,k}, k = 0, 1, \ldots, 2^j - 1\}$ are approximately normal, with variance $\pi^2/6$. In this case, by (5.22), we get

$$\lambda_{j,T} = \sqrt{\frac{\pi^2}{6}}\sqrt{2\log T} = \pi\sqrt{\frac{\log T}{3}}.$$

For j near $M-1$, the normal approximation is not good, and in this case Gao(1993) shows that we have to use (5.20). The argument can be found in the proof of the following result.

Theorem 5.3.(Gao, 1993) (a) Under the model (5.18), as $T \to \infty$,

$$P\{\bigcup_j [sup_k |y_{j,k} - E(y_{j,k})| > \lambda_{j,T}]\} \to 0. \tag{5.23}$$

(b) Under the model (5.18) and for a wavelet basis with compact support, for the finest resolution level $M-1$, as $T \to \infty$,

$$P\{sup_k|y_{M-1,k} - E(y_{M-1,k})| > \sqrt{2\log T}\} \to 1. \tag{5.24}$$

Similar results hold for the levels $M-2, M-3, \ldots$.

Result (5.23) shows that the thresholds are set sufficiently high so that the noise does not exceed it. On the other hand, (5.24) shows that noise peaks will be present with probability one in a proposal based in the assumption of Gaussian noise. Therefore, given the non-Gaussian character of the wavelet coefficients of the log-periodogram, thresholds based in the Gaussian theory will not be big enough to suppress completely the noise in these coefficients.

We now turn to the non-Gaussian case, considered by Neumann(1996). Consider, initially, the tapered periodogram

$$I_T(\omega) = \frac{1}{2\pi H_2^{(T)}} | \sum_{t=0}^{T-1} h_t X_t e^{-it\omega} |^2, \tag{5.25}$$

where h_t is a taper with $h_t = h(t/T)$ and $H_k^{(T)} = \sum_{t=0}^{T-1} h_t^k$.

As the usual periodogram($h_t \equiv 1$), $I_T(\omega)$ is asymptotically unbiased, non-consistent, chi-squared distributed and for $\omega_1 \neq \omega_2$, $I_T(\omega_1)$ and $I_T(\omega_2)$ are asymptotically non-correlated.

From the wavelets ϕ and ψ, consider the orthonormal basis in $L^2(\Re)$,

$$\{\tilde{\phi}_{\ell,k}\}_{k\in Z} \cup \{\tilde{\psi}_{j,k}\}_{j\geq \ell, k\in Z}, \tag{5.26}$$

generated in the usual way. Since we want an orthonormal basis in $L^2(\Pi), \Pi = [-\pi, \pi]$, from (5.26) we generate

$$\{\phi_{\ell,k}\}_{k\in I_\ell} \cup \{\psi_{j,k}\}_{j\geq \ell, k\in I_j}, \tag{5.27}$$

with $I_j = \{1, 2, \ldots, 2^j\}$,

$$\phi_{\ell,k}(t) = \sum_{n\in Z} (2\pi)^{-1/2} \tilde{\phi}_{\ell,k}((2\pi)^{-1}t + n)$$

and

$$\psi_{j,k}(t) = \sum_{n\in Z} (2\pi)^{-1/2} \tilde{\psi}_{j,k}((2\pi)^{-1}t + n).$$

Here, $L^2(\Pi)$ is the space of all square integrable functions with period 2π over Π.

For any function f of this space,

$$f = \sum_{k \in I_\ell} \alpha_\ell \phi_{\ell,k} + \sum_{j \geq \ell} \sum_{k \in I_j} \alpha_{j,k} \psi_{j,k}, \tag{5.28}$$

where

$$\alpha_k = \int f(t) \phi_{\ell,k}(t) dt, \tag{5.29}$$

and

$$\alpha_{j,k} = \int f(t) \psi_{j,k}(t) dt \tag{5.30}$$

are the wavelet coefficients.

Let us define the estimator of $\alpha_{j,k}$ by

$$\tilde{\alpha}_{j,k} = \int \psi_{j,k}(\omega) I_T(\omega) d\omega, \tag{5.31}$$

and an analogous definition for $\tilde{\alpha}_k$.

The purpose is to consider an estimator of the spectral density function (5.1) based in (5.31). Assume that the following conditions are valid.

(A1) $H = \int_0^1 h^2(t)dt > 0$, and h is a function of bounded variation.

(A2) $\sup_{1 \leq t < \infty} \{ \sum_{t_2,\ldots,t_k=1}^{\infty} \mid \mathrm{cum}(X_{t_1}, \ldots, X_{t_k}) \mid \} \leq C_1^k (k!)^{1+\gamma}$, for all $k = 2, 3, \ldots, \gamma \geq 0$.

(A3) $TV(f) \leq C_2$, $\| f \|_\infty \leq C_3$, where $TV(f)$ denotes the total variation of f.

(A4) For any $r > m$ suppose that:

(i) ϕ and ψ belong to C^r,

(ii) $\int \phi(t) dt = 1$,

(iii) $\int \psi(t) t^k dt = 0,\ 0 \leq k \leq r$.

We have denoted by m the degree of smoothness of the spectrum. Hence, in order that the estimator attain the optimal convergence rate, ϕ and ψ should be smoother than f.

The next result gives the main properties of the estimator (5.31).

Theorem 5.4.(Neumann,1996) Assuming (A1)-(A4) we have:

(i) $E(\tilde{\alpha}_{j,k}) = \alpha_{j,k} + O(2^{j/2}T^{-1}\log T)$.

(ii) $Var(\tilde{\alpha}_{j,k}) = 2\pi(H_4^{(T)}/(H_2^{(T)})^2)\int_{\Pi}\psi_{j,k}(\alpha)[\psi_{j,k}(\alpha)+\psi_{j,k}(-\alpha)]|f(\alpha)|^2 d\alpha + o(T^{-1}) + O(T^{-1}2^{-j})$.

(iii) $\mid \mathrm{cum}_n(\tilde{\alpha}_{j,k}) \mid \leq C^n(n!)^{2+2\gamma}T^{-1}(2^{j/2}T^{-1}\log(T))^{n-2}$ holds uniformly in $k \in I_j$ and $2^j \leq CT^{1-\alpha}$, where $C < \infty$ and $\alpha > 0$ are fixed but arbitrary constants.

To prove the asymptotic normality of the wavelet coefficients it is necessary to restrict the domain of the indices (j,k) . Let

$$\mathcal{I} = \{(j,k) \,:\, 2^j \leq CT^{1-\alpha}, k \in I_j\},$$

where $C < \infty$ and $0 < \alpha \leq 1/3$ are fixed constants. Let $\sigma_{j,k}^2$ be the variance of $\tilde{\alpha}_{j,k}$.

From (ii) above, $\sup_{(j,k)\in\mathcal{I}}\{\sigma_{j,k}\} = O(T^{-1/2})$. Consider, then, for some fixed $C_0 > 0$ the set

$$\mathcal{I}^0 = \{(j,k) \in \mathcal{I} \,:\, \sigma_{j,k} \geq C_0 T^{-1/2}\}.$$

Using (iii) of Theorem 5.4 we get that

$$\mid \mathrm{cum}_n(\tilde{\alpha}_{j,k}/\sigma_{j,k}) \mid \leq (n!)^{2+2\gamma}C^n(T^{-\alpha/2}\log(T))^{n-2}$$

holds uniformly for $(j,k) \in \mathcal{I}^0$. This result is used to prove the theorem that follows.

Theorem 5.5.(Neumann, 1996) Suppose that assumptions (A1)–(A4) are valid. Then,

$$\frac{P(\pm((\tilde{\alpha}_{j,k} - \alpha_{j,k})/\sigma_{j,k} \geq x)}{1 - \Phi(x)} \to 1$$

holds uniformly for $(j,k) \in \mathcal{I}^0$, $-\infty < x \leq \Delta_\gamma$, where $\Delta_\gamma = o(\Delta^{1/(3+4\gamma)})$ and $\Delta = T^{\alpha/2}(\log T)^{-1}$.

For those coefficients that have standard deviation below $C_0 T^{-1/2}$ define

$$\sigma_T = \max\{\max_{(j,k)\in\mathcal{I}}\{\sigma_{j,k}\}, C_0 T^{-1/2}\},$$

and let $\theta_{j,k} \sim N(0, \sigma_T^2 - \sigma_{j,k}^2)$ independent of $\tilde{\alpha}_{j,k}$. It follows that the variable $\tilde{\alpha}_{j,k} + \theta_{j,k}$ has the same mean and the same cumulants of order n, for $n \geq 3$, as $\tilde{\alpha}_{j,k}$, while its variance is $\sigma_T^2 \simeq T$. We then get the result corresponding to Theorem 5.5.

Theorem 5.6.(Neumann, 1996) Under (A1)-(A4) we have

$$\frac{P(\pm((\tilde{\alpha}_{j,k} + \theta_{j,k}) - \alpha_{j,k})/\sigma_T \geq x)}{1 - \Phi(x)} \to 1$$

uniformly for $(j,k) \in \mathcal{I}$, $-\infty < x \leq \Delta_\gamma$.

Let us consider now the estimation of the spectrum $f(\omega)$. The idea is to use (5.28) with some type of threshold for the wavelet coefficients. We can use the hard or soft schemes and choose $\lambda = \lambda_j(\sigma_{j,k}, \mathcal{F})$ depending on the scale, on the standard deviation of $\tilde{\alpha}_{j,k}$ and on the class $\mathcal{F}$ of functions to which $f(\omega)$ belongs.

Neumann(1996) propposes two types of thresholds that satisfy certain desirable conditions, namely

$$\lambda_{j,k} = (\sigma_{j,k} \vee \sigma_T)\sqrt{2\log(\#\mathcal{I})} \tag{5.32}$$

and

$$\lambda_{j,k} = (\sigma_{j,k} \vee \sigma_T)\sqrt{2\log((\#\mathcal{I})/2^\ell)}. \tag{5.33}$$

If $\mathcal{F}$ is a space with norm essentially equivalent to that of a Besov space, then the estimator

$$\hat{f} = \sum_{k\in I_\ell} \tilde{\alpha}_{\ell,k}\phi_{\ell,k} + \sum_{(j,k)\in\mathcal{I}} \delta^{(\cdot)}(\tilde{\alpha}_{j,k}, \lambda_{j,k})\psi_{j,k}, \tag{5.34}$$

leads to the adequate convergence rate for the estimator risk, specifically

$$\sup_{f \in \mathcal{F}} \{E\|\hat{f} - f\|^2_{L^2(\Pi)}\} = O((\log T/T)^{2m/(2m+1)}), \tag{5.35}$$

assuming (A1)–A(4) and additional conditions for the $\lambda_{j,k}$'s, which are satisfied by (5.32) and (5.33). In (5.34), $\delta^{(\cdot)}$ may be a hard or soft policy.

It is known that $T^{-2m/(2m+1)}$ is the optimal convergence rate for classes with degree of smoothness m, if X_t is stationary and Gaussian. Hence the above estimator is nearly minimax up to a factor of order of $\log T$. Neumann(1996) gives an estimator that attains this rate and further suggestions to improve these estimators.

In practice we need to find rules from the data to construct the thresholds, in particular to estimate $\sigma_{j,k}$ in (5.32) and (5.33). One possibility is to use

$$\hat{\sigma}^2_{j,k} = 2\pi(H_4^{(T)}/(H_2^{(T)})^2) \int_\Pi \psi_{j,k}(\alpha)[\psi_{j,k}(\alpha) + \psi_{j,k}(-\alpha)]|\tilde{f}(\alpha)|^2 d\alpha \tag{5.36}$$

as an estimator of the variance of the empirical wavelet coefficient, suggested by Theorem 5.4, where $\tilde{f}$ is any consistent estimator of the spectrum. From this we get estimators $\hat{\lambda}_{j,k}$ using (5.32) and (5.33).

Chapter 6

Non-stationary Processes

6.1 Introduction

The methods used in time series analysis are heavily based on the concepts of stationarity and linearity. Linear models as autoregressive(AR), moving average(MA) and mixed(ARMA) models are often used. These are special cases of the general discrete linear model, which is a linear combination of present and past values of a white noise process.

But there are fields where non-stationary and non-linear models are necessary, as in economics, oceanography, engineering, medicine, etc. A wide variety of non-linear models has been considered in the literature, as the bilinear models, threshold models, ARCH models, etc. We refer the reader to the books by Tong(1990), Subba Rao and Gabr(1984), Priestley(1988) and Granger and Terasvirta(1993) for further information. A classical treatment, in terms of Volterra type expansions is given in Wiener(1958).

Concerning non-stationary processes, several attempts have been made to treat special forms of non-stationarity in the frequency domain, defining what is called a *time-dependent spectrum*. This means that we will have a spectrum $f(t, \omega)$, in the non-stationary case, depending of time and frequency.

The first attempts considered processes that locally are stationary. Page (1952) introduced the definition of instantaneous power spectrum and Silverman(1957) considered *locally stationary processes*, for which the covariance function $\gamma(t_1, t_2) = Cov\{X(t_1), X(t_2)\}$ can be written in the form

$$\gamma(t_1, t_2) = m(\frac{t_1 + t_2}{2})\gamma_1(t_1 - t_2), \tag{6.1}$$

where $m(t) \geq 0$ and $\gamma_1(t)$ is a non-negative definite function(stationary covariance function). For $m(t)$ constant, we have a stationary process. Writing (6.1) in the form

$$\gamma(t+\frac{\tau}{2}, t-\frac{\tau}{2}) = m(t)\gamma_1(\tau), \tag{6.2}$$

the local stationary behavior becomes evident.

Given $\{X(t), t \in \Re\}$, a zero mean stationary process, a key result is the spectral representation

$$X(t) = \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega), \tag{6.3}$$

such that

$$\int_{-\infty}^{\infty} e^{i(\omega_1-\omega_2)t} dt = \delta(\omega_1 - \omega_2) \tag{6.4}$$

and where $\{Z(\omega), \omega \in \Re\}$ is a process with orthogonal increments, in the sense that

$$E\{dZ(\omega_1)\overline{dZ(\omega_2)}\} = \delta(\omega_1 - \omega_2) dF(\omega_1) d\omega_2. \tag{6.5}$$

Moreover, the autocovariance function of $X(t)$ can be written

$$\gamma(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} dF(\omega), \tag{6.6}$$

where F is the spectral distribution function of the process. If this is absolutely continuous(with respect to Lebesgue measure), then $dF(\omega) = f(\omega)d\omega$ and f is the spectral density function.

For non-stationary processes, the relations (6.3)-(6.6) are not valid. We will have to relax (6.4) or (6.5).

In this chapter we will be interested mainly in the estimation of the evolutionary spectrum via wavelets. Two important families of waveforms which are useful for time-frequency representations of many types of processes are the wavelet packets and cosine packets. We will not consider them here and the interested reader may refer to Bruce and Gao(1996) and Coifman et al.(1992) for details.

6.2 Time-Dependent Spectra

As a basic requirement we would like that the concept of frequency be preserved, when we define the concept of time-dependent spectrum, $f(t, \omega)$, say. Some conditions can be imposed on $f(t, \omega)$, as for example that it be real, non-negative and reduces to $f(\omega)$ in the stationary case. It can be shown, however, that it is not possible to obtain a definition of evolutionary spectrum that is unique.

There are two possible approaches(Flandrin,1989):

(1) To preserve (6.4)-(6.5), but abandon sines and cossines and loose the concept of frequency.

(2) To preserve the classical(stationary) concept of frequency and then accept some correlation in (6.5).

We present below brief comments on both cases. For further information see Flandrin(1989) and Loynes(1968).

6.2.1 Solutions Preserving Orthogonality

In this approach we can consider Karhunen decompositions, Priestley evolutionary spectra, evolutionary spectrum of Tjøstheim and Mélard, rational evolutionary spectra of Grenier, among other suggestions. We will limit ourselves to present the approach by Priestley(1965, 1981), who considered the representation of $X(t)$ in the form

$$X(t) = \int_{-\infty}^{\infty} A(t, \omega) e^{i\omega t} dZ(\omega), \tag{6.7}$$

where $Z(\omega)$ is an orthogonal process such that $E\{|dZ(\omega)|^2\} = d\mu(\omega)$, for some measure μ. Processes that admit the representation (6.7) are called *oscillatory*. The function $A(t, \omega)$ is supposed to vary slowly in a neighborhood of t and to have the representation

$$A(t, \omega) = \int_{-\infty}^{\infty} e^{it\theta} dK_{\omega}(\theta), \tag{6.8}$$

assuming that $|K_{\omega}(\theta)|$ has an absolute maximum at $\theta = 0$.

The covariance function can be written in the form

$$\gamma(s,t) = \int_{-\infty}^{\infty} \overline{A(s,\omega)} A(t,\omega) e^{i\omega(t-s)} d\mu(\omega) \tag{6.9}$$

from which we obtain

$$\mathrm{Var}\{X(t)\} = \int_{-\infty}^{\infty} |A(t,\omega)|^2 d\mu(\omega). \tag{6.10}$$

Similarly to the stationary case, since the variance is a measure of the total power in the series in instant t, we define the *evolutionary spectrum in time* t *and frequency* ω as

$$dH(t,\omega) = |A(t,\omega)|^2 d\mu(\omega). \tag{6.11}$$

In the case of absolute continuity relative to Lebesgue measure, that is, $dH(t,\omega) = f(t,\omega)d\omega$ and $d\mu(\omega) = f(\omega)d\omega$, we get

$$f(t,\omega) = |A(t,\omega)|^2 f(\omega). \tag{6.12}$$

We notice that the definition (6.11) depends on the choice of the family of oscillatory functions $\mathcal{F} = \{A(t,\omega)e^{i\omega t}\}$. For stationary processes, $A(t,\omega) = 1$.

The estimation of the evolutionary spectrum and other aspects were considered by Priestley(1965). One difficulty is to know if a given process belongs or not to the class $\mathcal{F}$.

6.2.2 Solutions Preserving Frequency

Here (6.5) is replaced by

$$E\{dZ(\omega_1)\overline{dZ(\omega_2)}\} = \Phi(\omega_1,\omega_2)d\omega_1 d\omega_2, \tag{6.13}$$

that is, we have a two-dimensional distribution function that is not concentrated in the diagonal $\omega_1 = \omega_2$, as in the stationary case.

A non-stationary process is said to be *harmonizable* if

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\Phi(\omega_1,\omega_2)| d\omega_1 d\omega_2 < \infty. \tag{6.14}$$

In this case,

$$\gamma(s,t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(\omega_1 s - \omega_2 t)} \Phi(\omega_1,\omega_2) d\omega_1 d\omega_2, \tag{6.15}$$

which is the analogous of (6.6).

For deterministic signals, Wigner(1932) introduced in quantum mechanics the function

$$W(t,\omega) = \int_{-\infty}^{\infty} x(t+\tau/2)\overline{x}(t-\tau/2)e^{-i\omega\tau}d\tau, \tag{6.16}$$

called the *Wigner-Ville distribution*, due to the fact that Ville(1948) also used it in signal theory.

In the case of a continuous stochastic process, the *Wigner-Ville spectrum* is defined as

$$W(t,\omega) = \int_{-\infty}^{\infty} \gamma(t-\tau/2, t+\tau/2)e^{-i\omega\tau}d\tau = \int_{-\infty}^{\infty} \Phi(\omega-\theta/2, \omega+\theta/2)e^{-it\theta}d\theta. \tag{6.17}$$

Notice that (6.17) reduces to the classical definition of spectrum in the stationary case, but it lacks an adequate physical interpretation, and also can assume negative values.

In the case of a discrete non-stationary process $\{X_t, t \in Z\}$, the Wigner-Ville is defined by

$$W(t,\omega) = \frac{1}{2\pi}\sum_{\tau} \gamma(t-\tau/2, t+\tau/2)e^{-i\omega\tau},\ t \in Z,\ |\omega| < \pi. \tag{6.18}$$

6.3 Locally Stationary Processes

The approaches of the previous section present a difficulty inherent to the study of non-stationary processes: an adequate asymptotic theory cannot be established, which allows us to find biases, variances and asymptotic distributions of statistics, since finite-samples properties are often difficult to derive.

In the case of stationary processes, when the sample size increases, we get more information of the same kind about the process, since the probabilistic structure does not change through translations of time. But if the process is non-stationary, observed for $t = 1, \ldots, T$, for $T \to \infty$ we will not get information on the process over the initial interval.

Let us consider the following example, due to Dahlhaus(1996). Let

$$X_t = g(t)X_{t-1} + \epsilon_t, \quad t = 1, \ldots, T$$

where $\epsilon_t \sim$ $i.i.d$ $N(0, \sigma^2)$ and $g(t) = a + bt + ct^2$. Then we may have, for example, $|g(t)| < 1$ on $[1, T]$, but $g(t) \to \infty$, as $T \to \infty$.

This difficulty led Dahlhaus(1997) to introduce the class of locally stationary processes. The idea is to consider an asymptotic theory such that $T \to \infty$ does not mean to "look into the future", but we "observe " $g(t)$ on a finer grid, but in the same interval, that is, we consider

$$X_{t,T} = g(\frac{t}{T})X_{t-1,T} + \epsilon_t, \quad t = 1, \dots, T,$$

such that $u = \frac{t}{T}$ belongs to the interval $[0, 1]$. Hence, for T increasing, we have more and more information in the sample $X_{1,T}, \dots, X_{T,T}$ to estimate the local structure of g at each time point.

Definition 6.1. A sequence of stochastic processes $\{X_{t,T}, t = 1, \dots, T\}$ is called *locally stationary*, with transfer function A^o and trend μ if there exists a representation of the form

$$X_{t,T} = \mu(\frac{t}{T}) + \int_{-\pi}^{\pi} A^o_{t,T}(\omega)e^{i\omega t}d\xi(\omega), \tag{6.19}$$

where:

(i) $\xi(\omega)$ is a stochastic process on $[-\pi, \pi]$, with $\overline{\xi(\omega)} = \xi(-\omega)$, $E\{\xi(\omega)\} = 0$ and with orthogonal increments, that is, $\mathrm{Cov}\{d\xi(\omega), d\xi(\omega^{'})\} = \delta(\omega - \omega^{'})d\omega$;

(ii) there exists a constant $K > 0$ and a smooth function $A(u, \lambda)$ on $[0, 1] \times [-\pi, \pi]$ with period 2π in ω, $A(u, -\omega) = \overline{A(u, \omega)}$ and such that for all T,

$$\sup_{t,\omega} \mid A^o_{t,T}(\omega) - A(\frac{t}{T}, \omega) \mid \leq KT^{-1}. \tag{6.20}$$

The functions $A(u, \omega)$ and $\mu(u)$are supposed to be continuous in u. The smoothness of the function $A(u, \omega)$ in u controls the local variation of $A^o_{t,T}(\omega)$ as a function of t, giving the local stationary caracter of $X_{t,T}$.

The above definition can be simplified, without loss of much generality, replacing $A^o_{t,T}$ by $A(u, \omega)$ in (6.19). We remark again that we are re-scaling time to the interaval interval $[0, 1]$ through $u = t/T$.

Definition 6.2. The *evolutionary spectrum* of the locally stationary process $X_{t,T}$ is defined by

$$f(u, \omega) = |A(u, \omega)|^2. \tag{6.21}$$

It can be shown(Neumann and von Sachs, 1997) that $f(u,\omega)$ is the limit in quadratic mean of

$$f_T(u,\omega) = \frac{1}{2\pi}\sum_s \mathrm{Cov}\{X_{[uT-s/2],T}, X_{[ut+s/2],T}\}e^{-i\omega s},$$

which is similar to the Wigner-Ville spectrum, defined in(6.18).

Let us consider now some examples of locally stationary processes(LSP).

Example 6.1. (i) Let Y_t be a stationary process with spectral density $f_Y(\omega)$ and μ, σ real functions, defined on $[0,1]$. Consider the modulated process

$$X_{t,T} = \mu(t/T) + \sigma(t/T)Y_t. \tag{6.22}$$

Then, $X_{t,T}$ is a LSP with $A^o_{t,T}(\omega) = A(t/T,\omega)$ and $f(u,\omega) = \sigma^2(u)f_Y(\omega)$.

(ii) Consider $\epsilon_t \sim \mathrm{WN}(0,\sigma^2)$ and

$$X_{t,T} = \sum_{j=0}^{\infty} a_j(t/T)\epsilon_{t-j}, \quad a_0(u) = 1. \tag{6.23}$$

It follows that this is a LSP with

$$A^o_{t,T}(\omega) = A(t/T,\omega) = \{\sum_{j=0}^{\infty} a_j(t/T)e^{-i\omega j}\}\frac{\sigma(t/T)}{\sqrt{2\pi}}$$

and $f(u,\omega) = |A(u,\omega)|^2$.

A special case of this general linear model with time-varying coefficients is the moving average model, assuming that $a_j(u) = 0,\ j > q$. A system similar to (6.23), with input a locally stationary process and a white noise additive term is considered by Chiann and Morettin(1997).

(iii) The autoregressive process

$$\sum_{j=0}^{p} b_j(t/T)X_{t-j,T} = \sigma(t/T)\epsilon_t, \quad b_0(u) = 1, \tag{6.24}$$

with $\epsilon_t \sim$ WN(0,1), is also a LSP with transfer function $A(u,\omega) = \frac{\sigma(u)}{\sqrt{2\pi}}(1 + \sum_{j=0}^{p} b_j(u)e^{-i\omega j})^{-1}$. See Dahlhaus, Neuman and Von Sachs(1996) for details.

The *local covariance of lag k and time u* is defined by

$$c(u,k) = \int_{-\pi}^{\pi} f(u,\omega)e^{i\omega k}d\omega. \tag{6.25}$$

This covariance may be estimated by

$$\hat{c}_T(u,k) = \frac{1}{b_T T}\sum_t K\left(\frac{u-(t+k/2)T}{b_T}\right) X_{t,T}X_{t+k,T}. \tag{6.26}$$

where $K : \Re \to [0,\infty)$ is a kernel and b_T a time domain bandwidth.

6.4 Estimation of the Evolutionary Spectrum

An estimator for the evolutionary spectrum is

$$\hat{f}(u,\omega) = \frac{1}{b_f}\int K_f\left(\frac{\omega-\alpha}{b_f}\right) I_N(u,\alpha)d\alpha, \tag{6.27}$$

where

$$I_N(u,\omega) = \frac{1}{2\pi H_N} \mid \sum_{s=0}^{N-1} h(s/N)X_{[uT-N/2+s+1],T}e^{-i\omega s} \mid^2 \tag{6.28}$$

is the peridogram over the segment $\{[uT]-N/2+1, [uT]+N/2\}$, $h : [0,1] \to \Re$ is a taper, $H_N = \sum_{s=0}^{N-1} h^2(s/N)$, with $H_N \sim N$, $K_f : \Re \to [0,\infty)$ is a kernel and b_f is the frequency domain bandwidth. Properties of this nonparametric estimator are studied by Dahlhaus(1996).

Let us see now how to use wavelets in the estimation of $f(u,\omega)$. Since this is a function of two parameters, we will have to use two-dimensional wavelet bases.

The procedure consists in applying non-linear thresholds to the empirical wavelet coefficients of the evolutionary spectrum. As mentioned before, this procedure is adaptative, in the sense that it is optimal relative to some error criterium, as the L^2 risk.

The two possibilities mentioned in section 2.4 were treated by von Sachs and Schneider(1996) and Neumann and von Sachs(1997) in the problem of the estimation of the evolutionary spectrum.

Initially, consider estimators using the basis $\mathcal{B}_2$, defined in (2.29) and the segmented periodogram (6.28), with $N = 2^J$. Suppose, also, that in the representation (2.26) there exists a coarsest resolution scale $j = 0$. Consider the projection of $f(u, \omega)$ over the 2^{2J}–dimesional space $V_J \subset L^2(U \times V)$, denoted $f_J(u, \omega)$, and its decomposition in wavelets in the form

$$f_J(u,\omega) = c_{0,\mathbf{0}} + \sum_{j=0}^{J} \sum_{\mathbf{k}=\mathbf{0}}^{2^j-1} \sum_{\mu=h,v,d} d^{\mu}_{j,\mathbf{k}} \Psi^{\mu}_{j,\mathbf{k}}(u,\omega), \tag{6.29}$$

sampled in the grid (u_i, ω_i), $0 \leq i, n \leq N - 1$. We are, therefore, restricting ourselves to the finest scale J. The coefficients of (6.29) are given by

$$c_{0,\mathbf{0}} = \int_0^1 \int_{-1/2}^{1/2} f(u,\omega) du d\omega, \tag{6.30}$$

$$d_{j,\mathbf{k}} = \int_0^1 \int_{-1/2}^{1/2} f(u,\omega) \Psi^{\mu}_{j,\mathbf{k}}(u,\omega) du d\omega. \tag{6.31}$$

Notice that $U \times V = [0,1] \times [-1/2, 1/2]$ and ω is frequency in cycles by unit time. Since we do not know the true spectrum, this is replaced in (6.30) and (6.31) by the periodogram (6.28), and this is computed for overlapping segments of $X_{t,T}$, all with length N. Let S be the translation from segment to segment, $1 \leq S \leq N$. It follows that the periodogram is computed for

$$u_i = \frac{t_i}{T},\ t_i = Si + \frac{N}{2},\ 0 \leq i \leq N-1,$$

with $T = SN$, and for the frequencies $\omega_n = 2\pi n/N$, $0 \leq n \leq N - 1$.

The resulting empirical coefficients are

$$\hat{c}_{0,\mathbf{0}} = \tfrac{1}{N} \textstyle\sum_{i=0}^{N-1} \int_{-1/2}^{1/2} I_N(u_i,\omega) d\omega, \tag{6.32}$$

$$\hat{d}_{j,\mathbf{k}} = \tfrac{1}{N} \textstyle\sum_{i=0}^{N-1} \int_{-1/2}^{1/2} I_N(u_i,\omega) \Psi^{\mu}_{j,\mathbf{k}}(u_i,\omega) d\omega. \tag{6.33}$$

In practice the coefficients are not computed as in (6.32)-(6.33), but through a wavelet transform that uses the Lagrange cardinal function of the espace V_j.

See von Sachs and Schneider(1996, section 2.2) for details.

The following assumptions are necessary in order to obtain the properties of the empirical coefficients.

(A1) Assume that $A(u, \omega)$ and $\Psi^{\mu}_{j,\mathbf{k}}(u, \omega)$ are differentiable in u and ω, with uniformly bounded first partial derivatives.

(A2) The parameters N, S and T satisfy

$$T^{1/4} << N << T^{1/2}/\ell n\, T \text{ and } S = N \text{ or } S/N \to 0,\ T \to \infty.$$

(A3) $h(x)$ is continuous on $[0, 1]$ and twice differentiable for all $x \notin P$, where P is a finite set and $\sup_{x \notin P} |h''(x)| < \infty$.

The following theorem gives the asymptotic properties of the wavelet coefficients.

Theorem 6.1.(von Sachs and Schneider, 1996). If the assumptions (A1)-(A3) are satisfied, then for $T \to \infty$, uniformly in $j, \mathbf{k}$, with $2^j = o(N)$,

(i)$E(\hat{d}^{\mu}_{j,\mathbf{k}} - d^{\mu}_{j,\mathbf{k}}) = O(2^{-j} N^{-1}) = o(T^{-1/2})$, for all μ.

(ii) $Var(\hat{d}^{\mu}_{j,\mathbf{k}}) = A^{\mu}_{j,\mathbf{k}}/T + O(2^j N/T^2) + O(2^{-j} T^{-1})$, for all μ.

(iii) $T^{L/2}\mathrm{cum}_L\{\hat{d}^{\mu}_{j,\mathbf{k}}\} = o(1),\ \forall L \geq 3$.

(iv) $\sqrt{T}(\hat{d}^{\mu}_{j,\mathbf{k}} - d^{\mu}_{j,\mathbf{k}})$ is asymptotically normal with zero mean and variance $A^{\mu}_{j,\mathbf{k}}$,

where in sum (6.33) N is replaced by M, such that

$$T = S(M-1) + N,\ 1 \leq S \leq N,\ u_i = t_i/T,\ t_i = S_i + N/2,\ 0 \leq i \leq M-1,$$

and

$$A^{\mu}_{j,\mathbf{k}} = 2C_h \int_{U \times V} \{f(u, \omega) \Psi^{\mu}_{j,\mathbf{k}}(u, \omega)\}^2 du d\omega, \tag{6.34}$$

with $C_h = \int_0^1 h^4(x)dx / (\int_0^1 h^2(x)dx)^2$, for $S = N$, and $C_h = 1$ if $S/N \to 0$.

The introduction of the additional parameter M, which controls the smoothness in the u direction, is motivated by the assumption (A2). See von Sachs and Schneider(1996) for additional comments .

As an estimator of the spectrum $f(u,\omega)$ we take

$$\hat{f}_T(u,\omega) = \hat{c}_{0,\mathbf{0}} + \sum_{j=0}^{J-1} \sum_{\mathbf{k}=\mathbf{0}}^{2^j-1} \sum_{\mu=h,v,d} \hat{d}^S_{j,\mathbf{k}} \Psi^\mu_{j,\mathbf{k}}(u,\omega), \tag{6.35}$$

with soft threshold $\hat{d}^S_{j,\mathbf{k}} = \delta^S_{\lambda_T}(\hat{d}^\mu_{j,\mathbf{k}})$ and parameter

$$\lambda_T = \lambda_{T;j,\mathbf{k}} = \frac{K \log T}{T^{1/2}},$$

hence the threshold is the same for all $j, \mathbf{k}$ and μ.

An estimate for K can be taken as

$$\hat{K} = 2C_h \max_{0<u<1} \max_{-1/2<\omega<1/2} |\tilde{f}(u,\omega)|,$$

with C_h given in (6.34) and $\tilde{f}$ any consistent estimator of f.

Such an estimator nearly minimizes the risk $R_T(\hat{f}, \mathcal{F}) = \sup_{f\in\mathcal{F}} E\|\hat{f} - f\|^2$ over a wide class of functions $\mathcal{F}$.

Let us turn now to estimators using the basis $\mathcal{B}_1$. The idea(see section 2.4) is to consider a 1-d basis with different scale parameter for each direction(time and frequency). If we have different smoothness degrees in both directions, this basis is superior to that used above. The estimators so obtained will have the property of attaining the optimal rate of convergence in anisotropic smoothness classes

In the previous approach it was necessary to introduce a parameter, N, in order to achieve smoothness, that is, we used local periodograms of length N. Let us suppose now that the function $A(u,\omega)$ is of bounded variation in $U \times \Pi = [0,1] \times [-\pi,\pi]$. Define the *total variation* of any function f in $U \times \Pi$ by

$$TV(f) = \sup \sum_i \sum_j \mid f(u_i,\omega_j) - f(u_i,\omega_{j-1}) - f(u_{i-1},\omega_j) + f(u_{i-1}, w_{j-1}) \mid,$$

where the supremum is taken over all partitions of $U \times \Pi$.

Assume the following conditions satisfied:

(S1) $A(u,\omega)$ has bounded total variation in $U \times \Pi$. Moreover,
(i) $\sup_u TV_\Pi(A(u,\cdot)) < \infty$,
(ii) $\sup_\omega TV_U(A(\cdot,\omega)) < \infty$,
(iii) $\sup_{u,\omega} |A(u,\omega)| < \infty$,
(iv) $\inf_{u,\omega} |A(u,\omega)| \geq K$, for some $K > 0$.

(S2) If $\hat{A}(u,s) = \frac{1}{2\pi}\int_\Pi A(u,\omega)e^{-i\omega s}dw,\ s \in Z, u \in [0,1]$, then $\sup_u \sum_s |\hat{A}(u,s)| < \infty$.

(S3) The wavelets in both directions are functions of bounded variation and the Fourier transforms of the wavelets in the frequency direction are absolutely summable.

(S4) $\sup_{1\leq t\leq T} \{\sum_{t_1,\ldots,t_{k-1}=1}^T |\mathrm{Cum}(X_{t_1,T},\ldots,X_{t_{k-1},T})| \leq C^k(k!)^{1+\gamma}$, $\forall k \geq 2, \gamma \geq 0$.

We assume that f has different degrees of smoothness in both directions, which is equivalent to say that $A(u,\omega)$ belongs to the class of functions $\mathcal{F}_{p_1,p_2}^{m_1,m_2}(C) = \{f : \sum_{i=1}^2(\|f\|_{p_i} + \|\frac{\partial^{m_i}}{\partial x_i^{m_i}} f\|_{p_i}) \leq C\}$, for all $C > 0,\ m_i \geq 1, p_i \geq 1,\ m_i > \frac{1}{p_i}$.

Define the periodogram

$$I_{t,T}(\omega) = \frac{1}{2\pi}\sum_{|s|\leq((t-1)\wedge(T-t))} X_{[t-s/2],T}X_{[t+s/2],T}e^{-i\omega s}. \tag{6.36}$$

This is a very irregular estimator and may assume also negative values. Consider two orthonornal bases, of compact supports, with the usual properties, one in the time direction, denoted $\{\phi_{\ell,k}\}_k \cup \{\psi_{j,k}\}_{j\geq\ell,k}$, and the other in the frequency direction, denoted $\{\tilde{\phi}_{\ell,k}\}_k \cup \{\tilde{\psi}_{j,k}\}_{j\geq\ell,k}$, periodic, as seen in section 5.3(see also Daubechies, 1992, section 9.3).

The wavelet coefficients in the expansion of $f(u,\omega)$ in terms of functions of the basis $\mathcal{B}_1$ are

$$\begin{aligned} d_I &= \int_{U\times\Pi} f(u,\omega)\mu_I(u,\omega)dud\omega \\ &= \int_{U\times\Pi} f(u,\omega)\psi_{j_1,k_1}(u)\tilde{\psi}_{j_2,k_2}(\omega)dud\omega, \end{aligned} \tag{6.37}$$

hence the empirical wavelet coeffcients will be given by

$$\tilde{d}_I = \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \psi_{j_1,k_1}(u)du \int_{-\pi}^{\pi} \tilde{\psi}_{j_2,k_2}(\omega) I_{t,T}(\omega) d\omega. \tag{6.38}$$

Restrict to those coefficients $\tilde{d}_I$ with indices in $\mathcal{I}_T = \{I : \ 2^{j_1+j_2} \leq T^{1-\delta}\}$, for some $\delta > 0$. Let $\sigma_I^2 = \mathrm{Var}(\tilde{d}_I)$.

Theroem 6.2.(Newmann and von Sachs, 1997). If the assumptions (S1)-(S4) are valid, then:

(i)$E(\tilde{d}_I) = d_I + o(T^{-1/2})$,

(ii) $\mathrm{Var}(\tilde{d}_I) = 2\pi T^{-1} \int_{U\times\Pi} \{f(u,\omega)\psi_{j_1,k_1}(u)\}^2 du \tilde{\psi}_{j_2,k_2}(\omega)$
$\times[\tilde{\psi}_{j_2,k_2}(\omega) + \tilde{\psi}_{j_2,k_2}(-\omega)]d\omega + o(T^{-1}) + O(2^{-j_2}T^{-1})$,

(iii) $|\mathrm{Cum}_n(\tilde{d}_I)| \leq (n!)^{2+2\gamma} C^n T^{-1} (T^{-1} 2^{(j_1+j_2)/2} \log(T))^{n-2}$, for $n \geq 3$ and $C > 0$, uniformly in $I \in \mathcal{I}_T$.

(iv) Let $\Delta_T = (\log(T))^{\lambda}$, for all fixed λ, $0 < \lambda < \infty$. Then,

$$P\left(\frac{\pm(\tilde{d}_I - d_I)}{\sigma_I} \geq x\right) = (1 - \Phi(x))(1 + o(1)). \tag{6.39}$$

holds uniformly for $-\infty \leq x \leq \Delta_T$, $I \in \mathcal{I}_T \cap \{I : \ 2^{j_2} \geq T^{\rho}\}$, $\rho > 0$(arbitrarily small).

Consider, now, the estimation of the evolutionary spectrum. Define

$$\hat{f}(u,\omega) = \sum_I \delta^{(\cdot)}(\tilde{d}_I, \lambda_{I,T}) \mu_I(u,\omega), \tag{6.40}$$

where the sum excludes the indices $(j_1, j_2) = (\ell - 1, \ell - 1)$, that is, we do not apply, as usual, the threshold to the coarsest scale. Call $\mathcal{I}_T^o$ the set of all indices of $\mathcal{I}_T$ except the excluded ones. Let

$$\vartheta(m_1, m_2) = \frac{2m_1m_2}{2m_1m_2 + m_1 + m_2}.$$

As in section 5.3 we can use a consistent estimator in (ii) of Theorem 6.2 to get an estimator of σ_I^2, $\hat{\sigma}_I^2$. Let $\hat{\lambda}_I$ a stochastic threshold and $\tilde{f}$ the resulting estimator. It can be shown that

$$E\{\|\tilde{f} - f\|^2_{L_2(U\times\Pi)} = O(T^{-\vartheta(m_1,m_2)}\log(T)), \tag{6.41}$$

which means that the estimator is nearly simultaneously optimal over the class $\mathcal{F}(C)$, which includes sufficiently wide smoothness classes. One practical rule is to use the threshold

$$\hat{\lambda}_I = \hat{\sigma}_I\sqrt{2\log(\#\mathcal{I}_T^o)}.$$

6.5 Wavelet Processes

Based on the definition of LSP given in the previous section, von Sachs, Nason and Kroisandt(1996a) developed a time-scale spectral representation for a particular class of non-stationary processes.

Definition 6.3. The doubly indexed sequence $\{X_{t,T}, t = 1,\ldots,T,\ T \geq 1\}$ is said to be a *locally stationary wavelet process*, with respect to the wavelet basis $\{\psi_{j,k}(t), j,k \in Z\}$, if

$$X_{t,T} = \sum_{j=-J}^{-1}\sum_k w^o_{j,k;T}\psi_{j,k}(t)\xi_{j,k}, \tag{6.42}$$

where $T = 2^J$, the $\psi_{j,k(t)}$ are given by (2.2), with $j = -1,-2,\ldots,-J(T) = -\log_2(T), k \in Z$, and with the properties:

(i) $E\{\xi_{j,k}\} = 0$, for all j,k. Hence, $E\{X_{t,T}\} = 0$, for all t,T.

(ii) $\mathrm{Cov}(\xi_{j,k},\xi_{\ell,m}) = \delta_{j,\ell}\delta_{k,m}$.

(iii) For each $j \leq -1$ there exists a function $W_j(z)$ on $(0,1)$, Lipschitz-continuous, such that

$$\sup_k \mid w^o_{j,k;T} - W_j(\frac{k2^{-j}}{T}) \mid = O(T^{-1}), \quad T \to \infty,$$

where for each j, the sup is taken over $k = 1,\ldots,2^jT$.

(iv) The basis $\{\psi_{j,k}\}$ is orthonormal and the wavelets have compact support.

(v) $\sum_{j=-\infty}^{-1} \mid W_j(z) \mid^2 < \infty$, for all $z \in (0,1)$.

We remarh that the notation for the wavelets here is different: -1 corresponds to the finest resolution scale and $-J$ to the coarsest.

Definition 6.4. Consider

$$S_j(z) = \mid W_j(z) \mid^2, \; z \in (0,1). \tag{6.43}$$

The sequence $\{S_j(z), \; j \leq -1\}$ is called the *evolutionary wavelet spectrum* of the sequence $\{X_{t,T}\}$ with respect to the basis $\{\psi_{j,k}\}$.

From the assumption (ii) of Definition 6.3 it follows that

$$S_j(z) = \lim_{T\to\infty} \mid w^o_{j,2^j[zT];T} \mid^2, \; \forall z \in (0,1). \tag{6.44}$$

The estimation of the spectrum given by (6.43) is based on the discrete wavelet transform (5.5) applied to the locally stationary wavelet process $X_{t,T}$.

Definition 6.5. Let

$$d_{j,k;T} = \sum_{t=1}^{T} X_{t,T}\psi_{j,k}(t) \tag{6.45}$$

be the discrete wavelet transform of $X_{1,T}, \ldots, X_{T,T}$.

The basic statistic, similar to the usual periodogram, is defined now.

Definition 6.6. The *wavelet periodogram* of $X_{1,T}, \ldots, X_{t,T}$ is defined by

$$I_{j,k;T} = \mid d_{j,k;T} \mid^2 . \tag{6.46}$$

Suppose that the variables $\xi_{j,k}$ are normal, hence the process $X_{t,T}$ is also Gaussian. The following theorem gives the properties of the statistic (6.46).

Theorem 6.3. For the estimator (6.46) we have:

(a) $E\{I_{j,2^j[zT];T}\} = S_j(z) + O(T^{-1}), \; \forall z \in (0,1)$.

(b) $\mathrm{Var}\{I_{j,2^j[zT];T}\} = 2S_j(z)^2 + O(T^{-1}), \; \forall z \in (0,1)$.

(c) $\mathrm{Cov}\{I_{j,k;T}, I_{\ell,m;T}\} = 2\delta_{j\ell}\delta_{km}\left\{S_j\left(\frac{k}{2^jT}\right)^2 + O(T^{-1})\right\}$.

Part (b) of Theorem 6.3 shows that the wavelet periodogram is not consistent. To obtain better estimators we will have to smooth this periodogram and for this purpose we can use non-linear threshold techniques. For details on proofs and other properties of $I_{j,k;T}$ and on smoothed estimators see von Sachs, Nason and Kroisandt(1996a,b).

Chapter 7

Further Topics

7.1 Bayesian Analysis

We saw that there are basically three approaches for the determination of thresholds: (a) minimax; (b) cross-validation; (c) multiple hyphoteses testing.

Given that a Bayes rule is usually a shrinker, it is natural to use some Bayesian procedure in wavelet shrinkage. There is though one difference: a shrinkage rule decreases the coefficients in absolute value, without changing its sign. A thresholding rule shrinks the coefficients, but also effectively removes those coefficients which are in absolute value smaller than a certain value.

The Bayesian approach has been used mainly in the thresholding procedure, but there are also applications as in density estimation, for example.

Let us consider again the nonparametric regression model (4.1). Applying the wavelet transform we get the model in (4.4), hence estimating the function f is equivalent to estimate the coefficients $w_{j,k}$, assuming that

$$y_{j,k} \sim N(w_{j,k}, \sigma^2).$$

Assume that $y_{j,k}|w_{j,k}, \sigma^2 \sim N(w_{j,k}, \sigma^2)$, σ^2 unknown and set a prior for σ^2, for example an exponential with parameter μ: $\sigma^2 \sim E(\mu)$. It follows that the marginal likelihood is double exponential: $y_{j,k}|w_{j,k} \sim ED(w_{j,k}, 1/\sqrt{2\mu})$.

Admitting a prior $t_n(0, \tau)$, say, for $w_{j,k}$, we can get a Bayes rule under a quadratic loss function. Observe that $E(\sigma^2) = 1/\mu$, and hence the hyperparameter μ estimates the precision $1/\sigma^2$. The hyperparameter τ controls the shrinking. See Vidakovic(1994) for details and examples.

Abramovich et al.(1997) suggest a weighted combination of L_1 losses based in the wavelet coefficients. Then the Bayes rule is the posterior median. Specifically, consider the model

$$w_{j,k} \sim \pi_j N(0, \tau_j) + (1 - \pi_j)\delta(0), \; j = 0, \ldots, J, k = 0, \ldots, 2^j - 1, \tag{7.1}$$

where $0 \leq \pi_j \leq 1$, $\delta(0)$ is the point mass at the origin and we assume that the $w_{j,k}$'s are independent. Notice that we take the same hyperparameters π_j and τ_j for all the coefficients in level j.

Model (7.1) tells us that π_j is the proportion of non-null coefficients in level j and τ_j^2 is the measure of their magnitudes. The posterior $w_{j,k}|y_{j,k}$ can then be computed. If F denotes the distribution function of this posterior, then $F(w_{j,k}|y_{j,k}) = 0,5$ implies that the median of the posterior distribution is

$$\text{med}(w_{j,k}|y_{j,k}) = \text{sign}(y_{j,k})\max(0, \xi_{j,k}), \tag{7.2}$$

where

$$\xi_{j,k} = \frac{\tau_j^2}{\sigma^2 + \tau_j^2}|y_{j,k}| - \frac{\sigma\tau_j}{\sqrt{\sigma^2 + \tau_j^2}}\Phi^{-1}\left(\frac{1 + (w_{j,k} \wedge 1)}{2}\right). \tag{7.3}$$

It is easy to see that (7.2) is a level-dependent threshold with parameter λ_j, and $[-\lambda_j, \lambda_j]$ is an interval where (7.3) is negative, for any $y_{j,k}$. For a detailed analysis of this model, suggestion of the choice of hyperparameters and applications see Abramovich et al.(1997).

Müller and Vidakovic(1995) consider the estimation of a density using a Bayesian approach. Further references are Vidakovic and Müller(1995) and Vannucci and Corradi(1997).

7.2 Model with Stationary Errors

A commonly used model in time series is

$$X(t) = S(t) + E(t), \tag{7.4}$$

where $S(t)$ is a deterministic function and $E(t)$ is a stationary process, with zero mean and spectrum $f_{EE}(\omega)$. It follows that $S(t)$ is the mean level of the series in time t and the interest is to estimate it. There are several ways to do it, see Brillinger(1996). Using a wavelet expansion we have

$$h(x) = \sum_k \alpha_{\ell,k}\phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty}\sum_k \beta_{j,k}\psi_{j,k}(x), \tag{7.5}$$

where we set $S(t) = h(t/T),\ t = 0, 1, \ldots, T-1$ and assume that $h(x) \in L_2(\Re)$. The coefficients are given by

$$\beta_{j,k} = \int \psi_{j,k}(x)h(x)dx, \tag{7.6}$$

and a similar relation for $\alpha_{\ell,k}$. Hall and Patil(1993) and Brillinger(1996) introduce an additional parameter U, such that

$$\psi^U_{j,k}(x) = \sqrt{U}\psi_{j,k}(Ux),$$

and the analogous for $\alpha_{\ell,k}(x)$, in such a way that the resulting system is orthonormal and complete. For example, we may take $U = 2^n$. Consider, then, the expansion (7.5) above and the modified coefficients.

An estimator of $h(x)$ is obtained estimating the theoretical wavelet coefficients (7.6) by the empirical coefficients

$$\hat{\alpha}^U_{\ell,k} = \frac{1}{T}\sum_{-T}^{T}\phi^U_{\ell,k}(t/T)X(t), \tag{7.7}$$

$$\hat{\beta}^U_{j,k} = \frac{1}{T}\sum_{-T}^{T}\psi^U_{j,k}(t/T)X(t), \tag{7.8}$$

and defining

$$\hat{h}(x) = \sum_k \hat{\alpha}^U_{\ell,k}\phi^U_{\ell,k}(x) + \sum_{j=1}^{J-1}\sum_k \hat{\beta}^U_{j,k}(x)\psi^U_{j,k}(x), \tag{7.9}$$

for some J.

Set down the following assumptions:

(1) $h(\cdot)$ is a function of bounded variation on $[-1, 1]$ and it is zero outside this interval.

(2) $\phi(\cdot)$ is of bounded variation on $[0, 2N-1]$, N integer and the functions

$\phi_{\ell,k}$ and $\psi_{j,k}$ form a complete orthonormal system.

(3) If $c_m(u_1, \ldots, u_{m-1})$ indicates the cumulant of order m of $E(t)$, then

$$K_m = \sum_{u_1,\ldots,u_{m-1}} |c_m(u_1, \ldots, u_{m-1})| < \infty, \; m \geq 2.$$

Moreover, $\sum_u |u||c_{EE}(u)| < \infty, \; f_{EE}(0) \neq 0$.

Theorem 7.1.(Brillinger,1996). Under the assumptions (1)-(3), we have:

(i) $E(\hat{\beta}^U_{j,k} - \beta^U_{j,k}) = O(2^{j/2}U^{1/2}T^{-1})$, the error being uniform in j, k, U, T. An analogous expression holds for $\hat{\alpha}_{\ell,k}$.

(ii) $\mathrm{Cov}(\hat{\beta}^U_{j,k}, \hat{\beta}^U_{j'k'}) = 2\pi f_{EE}(0)T^{-1} \int_{-U}^{U} \psi_{j,k}(u)\psi_{j'k'}(u)du + O(2^{(j+j')/2}UT^{-2})$, with uniform error in j, j', k, k', U and T. Similar expressions hold for the covariances of the $\hat{\phi}$'s and between the $\hat{\phi}$'s and $\hat{\psi}$'s.

(iv) If $U/T \to 0, \; U \geq \epsilon > 0$, as $T \to \infty$, then finite collections of $\hat{\alpha}^U, \hat{\beta}^U$ are asymptotically normal, with the indicated first and second order moments.

The asymptotic properties of $\hat{h}(x)$, including normality, are similarly obtained. See Brillinger(1996, Theorem2). Extensions for spatial processes, processes with long memory and processes irregularly observed are easily derived. Threshold estimators and applications to hydrologic data were considered by Brillinger(1994).

7.3 Estimation of Densities

The interest here lies in the estimation of a probability density function $f(x)$ based in observations $X_1, \ldots, X_n$ forming a sample of f. For a review of the usual methods see Silverman(1986).

Consider the expansion

$$f(x) = \sum_k \alpha_{\ell,k}\phi_{\ell,k}(x) + \sum_{j\geq\ell}\sum_k \beta_{j,k}\psi_{j,k}(x), \tag{7.10}$$

where the coefficients are estimated by

$$\hat{\alpha}_{\ell,k} = n^{-1} \sum_{k=1}^{n} \phi_{\ell,k}(X_i), \tag{7.11}$$

$$\hat{\beta}_{j,k} = n^{-1} \sum_{k=1}^{n} \psi_{j,k}(X_i). \tag{7.12}$$

To estimate f, we use thresholds

$$\tilde{\beta}_{j,k} = \delta^{(\cdot)}(\hat{\beta}_{j,k}, \lambda_j), \tag{7.13}$$

where $\delta^{(\cdot)}$ represents a hard or soft policy and λ_j is a level-dependent parameter. In general, we do not apply thresholds to the coefficients $\hat{\alpha}_{\ell,k}$.

As in the nonparametric regression case, we are interested in global error measures to assess the performance of the estimators over a wide range of smoothness classes. The estimator of f, replacing the $\tilde{\beta}_{j,k}$ in (7.10), attains a nearly optimal performance, in terms of convergence rates.

Donoho et al.(1996b) use thresholds given by

$$\tilde{\beta}_{j,k} = \begin{cases} \hat{\beta}_{j,k}, & \text{if } |\hat{\beta}_{j,k}| > KC(j)\sqrt{n} \\ 0, & \text{otherwise,} \end{cases} \tag{7.14}$$

with $C(j) = \sqrt{j}$ and K constant.

Delyon and Juditsky(1993) use the thresholds of (7.14) with $C(j) = \sqrt{j - \ell}$. Johnstone et al.(1992) consider $\lambda_j = A\sqrt{j}$, with A constant. See also Hall and Patil(1993).

Bibliography

Abramovich, F., Sapatinas, T. and Silverman, B.W.(1997). Wavelet thresholding via a Bayesian approach. Technical Report, University of Bristol, U.K.

Adak, S.(1995). Time-dependent spectral analysis of nonstationary time series. Technical Report 478, Department of Statistics, Stanford University.

Blackman, R.B. and Tukey, J.W.(1958). *The Measurement of Power Spectra.* New York: Dover.

Brillinger, D.R.(1981). *Time Series: Data Analysis and Theory.* San Francisco: Holden-Day.

Brillinger, D.R.(1994). Some river wavelets. *Environmetrics,* **5**, 211–220.

Brillinger, D.R.(1996). Some uses of cumulant in wavelet analysis. *Nonparametric Statistics*, **6**, 93–114.

Brockwell, P.J and Davis, R.A.(1991). *Time Series: Theory and Methods.* Second Edition. New York: Springer Verlag.

Bruce, A.G. and Gao, Y.-Ye(1994). *S+ WAVELETS Users Manual.* StatSci Division of MathSoft Inc., 1700 Westlake Ave N, Seattle, WA 98109-9891, USA.

Bruce, A.G. and Gao, H.-Ye(1995). Understanding waveshrink: Variance and bias estimation. *Biometrika*, **83**, 727–745.

Bruce, A.G. and Gao, H.-Ye(1996). *Applied Wavelet Analysis with S-PLUS*. New York: Springer.

Buckheit, J., Chen, S., Donoho, D.L., Johstone, I. and Scargle, J.(1995). About WaveLab. Available from ftp playfair.stanford.edu/pub/wavelab.

Chiann, C. and Morettin, P.A.(1996). A wavelet analysis for stationary processes. Under revision, *J. Nonparametric Statistics*.

Chiann, C. and Morettin, P.A.(1997). Estimation of time-varying linear systems using wavelets. Preprint.

Chui, C.K.(1992).*An Introduction to Wavelets.* San Diego: Academic Press.

Coifman, R., Meyer, Y. and Wickerhauser, V.(1992). Wavelet analysis and signal processing. In *Wavelets and Their Applications*(M.B. Ruskai et al., editors), 153–178. Jones and Bartlett Publishers, Boston.

Cooley, J.W. and Tukey, J.W.(1965). An algorithm for the machine calculation of complex Fourier series.*Math. Comp.*,**19**, 297–301.

Dahlhaus, R.(1996). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In *Athens Conference on Applied Probability and Time Series Analysis, Vol. II*(P.M. Robinson and M. Rosenblatt, eds.). New York: Springer-Verlag.

Dahlhaus, R., Neumann, M.H. and von Sachs, R.(1996). Nonlinear wavelet estimation of time-varying autoregressive processes. Preprint, Universitat Heidelberg.

Dahlhaus, R.(1997). Fitting time series models to nonstationary processes. To appear, *Ann. Statist.*

Daubechies, I.(1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, **41**, 909–996.

Daubechies, I.(1989) Orthonormal bases of wavelets with finite support- Connection with discrete filters. In *Wavelets, Time-Frequency Methods and Phase Space. First Int. Wavelets Conf.*(J.M. Combes, A. Grossmann, Tchamitchian,P., editors), Marseille, Dec. 1987, 38–66. Springer-Verlag.

Daubechies, I.(1992).*Ten Lectures on Wavelets.*Philadelphia: SIAM.

Delyon, B. and Juditsky, A.(1993). Wavelet estimator, global error measures: revisited. TR 782, IRISA, Campus de Beaulieu.

Donoho, D.L.(1993). Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. Proceedings of Symposia in Applied Mathematics,Vol. 47, 173–205,AMS.

Donoho, D.L. and Johnstone, I.M.(1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

Donoho, D.L. and Johnstone,I.M.(1995a). Adapting to unknown smoothing via wavelet shrinkage. *J. Amer. Statist. Association*, **90**, 1200–1224.

Donoho, D.L. and Johnstone, I.M.(1995b). Minimax estimation via wavelet shrinkage. Submitted to the *Annals of Statistics*.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D.(1995). Wavelet shrinkage: Asymptopia?(with discussion). *J. Royal Statist. Society*, Series B, **57**, 301–369

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996a). Universal near minimaxity of wavelet shrinkage. In *Festschrift for Lucien Le Cam*(D. Pollard and G. Yang, eds.). Springer-Verlag.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D.(1996b). Density estimation by wavelet thresholding. *The Ann. Statist.*, **24**, 508–539.

Efron, B. and Tibshirani, R.J.(1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Flandrin, P.(1989). Time dependent spectra for nonstationary stochastic processes. In *Time and Frequency Representations of Signals and Systems*(G. Longo and B. Picinbono, eds.), 69–124. New York: Springer-Verlag.

Gao, H.-Y.(1993). Choice of thresholds for wavelet estimation of the log spectrum. TR 438, Dept. Statistics, Stanford University.

Granger, C.W.J. and Terasvirta, T.(1993). *Modelling Nonlinear Economic Relationships.*. Oxford University Press.

Grossmann, A. and Morlet, J.(1984). Decomposition of Hardy functions into square integrable wavelets of constant shape.*SIAM J. Math.*,**15**, 723–736.

Grossmann, A. and Morlet, J.(1985). Decomposition of functions into wavelets of constant shape, and related transforms. In *Mathematics + Physics, Lectures on Recent Results*(L. Streit,editor), 135–165. Singapore: World Scientific.

Grossmann, A.(1988). Wavelet transforms and edge detection. *Stochastic processes in Physics and Engineering*(P. Blanchard et al, editors). Boston:

Reidel.

Haar, A.(1910). Zur theorie der orthogonalen Funktionensysteme. *Mathematiche Annalen*,**69**, 331–371.

Hall, P. and Patil, P.(1993). On wavelet methods for estimating smoothing functions. TR 12-93, CMA, Australian National University

Härdle, W., Lütkepohl, H. and Chen, R.(1997). A review of nonparametric time series analysis. *International Statistical Review*, **65**, 49-72.

Harmuth, H.F.(1969). *The Transmission of Information by Ortogonal Functions.* New York: Springer-Verlag.

Hubbard, B.B.(1996).*The World According to Wavelets.* Wellesley: A.K. Peters.

Johnstone, I.M., Kerkyacharian, G. and Picard, D.(1992). Estimation d'une densité de probabilité par méthode d'ondelettes. *Compt. Rend. Acad. Sci. Paris A,* **315**, 211–216.

Johnstone, I.M. and Silverman, B.(1994). Wavelet threshold estimates for data with correlated noise. Technical Report, Stanford University.

Lehmann, E.L.(1983). *Theory of Point Estimation.* New York: John Wiley and Sons.

Loynes, R.M.(1968). On the concept of the spectrum for nonstationary processes. *J. Royal Statist. Society,* Series B, **30**, 1–20.

Mallat, S.(1989). A theory for multiresolution signal decomposition: Wavelet representation.*IEEE Trans. Pattern Anal. Mach. Intellig.*, **11**, 674–693.

Meyer, Y.(1986). Ondelettes et fonctions splines. *Séminaire Equations aux Derivées Partielles*. Cent. Ecole Polytech., Palaiseau, France.

Meyer, Y.(1987). Wavelets and operators. *Rapport* CEREMADE, No. 8704, Univ. Paris-Dauphine, Paris, France.

Meyer, Y.(1988). Orthonormal wavelets. *Congr. Int. Phys. Math.*, Swansea, July.

Meyer, Y.(1990).*Ondelettes.* Paris: Hermann.

Meyer, Y.(1993). *Wavelets: Algoritms and Applications.* Philadelphia: SIAM.

Morettin, P.A.(1974). Walsh function analysis of a certain class of time series. *J. Stochastic Processes and Their Applications*, **2**, 183–193.

Morettin, P.A.(1981). Walsh spectral analysis. *SIAM Review*, **23**, 279–291.

Morettin, P.A.(1996). From Fourier to wavelet analysis of time series. In *Proceedings in Computational Statistics*(A. Prat, editor),111–122. Physica–Verlag.

Morettin, P.A.(1997).*Waves and Wavelets: From Fourier to Wavelet Analysis of Time Series*. To appear, University of São Paulo Press.

Morlet, J.(1981). Sampling theory and wave propagation. *Proc. 51st Annu. Meet. Soc. Explor. Geophys., Los Angeles.*

Morlet, J.(1983). Sampling theory and wave propagation. *NATO ASI, Vol. F1, Issues on Acoustic Signal/Image Processing and Recognition*(C.H. Chen editor). Springer-Verlag.

Moulin, P.(1994). Wavelet thresholding techniques for power spectrum estimation. *IEEE Trans. Signal Processing,* **42**, 3126–3136.

Müller,P. and Vidakovic, B.(1995). Bayesian inference with wavelets: density estimation. *Discussion Paper* 95-33, ISDS, Duke University.

Nason, G.P.(1993). *The* wavethresh *Package; Wavelet Transform and Thresholding Software for S.* Available from the StatLib archive, statlib@lib.stat.cmu.edu.

Nason, G.P.(1994). Wavelet regression by cross-validation. TR 447, Dept. Statistics, Stanford University.

Nason, G.P.(1995). Wavelet function estimation using cross-validation. In *Wavelets and Statistics*(A. Antoniadis and G. Oppenheim, eds.), 261-280. Lecture Notes in Statistics, Springer-Verlag.

Nason, G.P. and Siverman,B.W.(1994). The discrete wavelet transform in S.

J. Computational and Graphical Statistics, **3**, 163–191.

Nason, G.P. and Silverman, B.(1995). The stationary wavelet transform and some statistical applications. In *Wavelets and Statistics*(A. Antoniadis and G. Oppenheim,eds), 281-300. Lecture Notes in Statistics, Springer-Verlag.

Neumann, M.H.(1996). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *J. Time Series Analysis,* **17**, 601–633.

Neumann, M.H. and von Sachs, R.(1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra.*Ann. Statist,* **25**, Number 1.

Ogden, T. and Parzen, E.(1996). Data dependent wavelet thresholding in nonparametric regression with change points applications. *Computational Statistics and Data Analysis*, **22**, 53–70.

Page, C.H.(1952). Instantaneous power spectra. *J. Applied Physics,* **23**, 103–106.

Percival, D.B.(1993). An introduction to spectral analysis and wavelets. In *Proceedings of the workshop on Advanced Mathematical Tools in Metrology,* Torino, Italy, October,1993.

Percival, D.B. and Walden,A.T.(1993).*Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate techniques.* Cambridge: Cambridge University Press.

Priestley, M.B.(1965). Evolutionary spectra and nonstationary processes. *J. Roy. Statist. Soc.*, Series B, **27**, 204–237.

Priestley, M.B.(1981). *Spectral Analysis and Time Series.* London: Academic Press.

Priestley, M.B.(1988). *Nonlinear and non-stationary time series analysis.* London: Academic Press.

Priestley, M.B.(1996). Wavelets and time-dependent spectral analysis. *J. Time Series Analysis*, **17**, 85–103.

Scargle, J.D.(1993). Wavelet methods in astronomical time series. In *Appli-*

cations of Time Series in Astronomy and Meteorology(ed. T. Subba Rao and O. Lessie). London: Chapman and Hall.

Silverman, B.W.(1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Silverman, R.A.(1957) Locally stationary random processes.*IRE Trans. Information Theory,* **IT-3**, 182–187.

Stein, C.(1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.,* **9**, 1135–1151.

Subba Rao, T. and Gabr, M.M.(1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models.*. Lecture Notes in Statistics, **24**. New York: Springer-Verlag.

Tong, H.(1990). *Non-linear Time Series. A Dynamical System Approach.* New York: Oxford University Press.

Vannucci, M. and Corradi, F.(1997). Some findings on the covariance structure of wavelet coefficients: theory and models in a Bayesian perspective. Report University of Kent at Canterbury, 97/05.

Vidakovic, B.(1994). Nonlinear wavelet shrinkage via Bayes rules and Bayes factors. *Discussion Paper* 94-24, ISDS, Duke University.

Vidakovic, B. and Müller, P.(1995). Wavelet shrinkage with affine Bayes rules with applications. *Discussion Paper* 95-34, ISDS, Duke University.

Ville, J.(1948). Théorie et applications de la notion de signal analytique. *Cables et Transm., 2ème A*, 61–74.

von Sachs, R.(1996). Modelling and estimation of time-varying structure of nonstationary time series. *Brazilian J. Probab. and Statistics*, **10**, 181–204.

von Sachs, R., Nason, G.P. and Kroisandt, G.(1996a). Spectral representation and estimation for locally stationary wavelet processes. In *Proc. Workshop on Spline Functions and Theory of Wavelets.* Montréal, Canada, 1996.

von Sachs, R., Nason, G.P. and Kroisandt, G.(1996b). Adaptive estimation of the evolutionary wavelet spectrum. Technical Report, Department of Statistics, Stanford University.

von Sachs, R. and Schneider, K.(1996). Wavelet smoothing of evolutionary spectra by nonlinear thresholding. *Appl. Comput. Harmonic Analysis,* **3**, 268–283.

Wahba, G.(1980). Automatic smoothing of the log periodogram.*J. Amer. Statist. Association*, **75**, 122–132.

Wang, Y.(1996). Function estimation via wavelet shrinkage for long-memory data. *The Ann. Statist.*, **24**, 466–484.

Wickerhauser, M.(1994). *Adapted Wavelet Analysis from Theory to Software.* Wellesley: A. K. Peters.

Wiener, N.(1958). *Non-linear Problems in Random Theory.* Cambridge, Mass.: M.I.T Press.

Wigner, E.P.(1932). On the quantum correction for thermodynamic equilibrium. *Phys. Rev.*, **40** 749–759.

Wolfowitz, J.(1950). Minimax estimation of the mean of a normal distribution with known variance. *Ann. Math. Statist.,* **21**, 218–230.