Choice of Thresholds for Wavelet Shrinkage Estimate of the Spectrum

Hong-ye Gao
Statistical Sciences Division
MathSoft, Inc.
1700 Westlake Ave. N., Suite 500
Seattle, WA 98109

Abstract

We study the problem of estimating the log spectrum of a stationary Gaussian time series by thresholding the empirical wavelet coefficients. We propose the use of thresholds $t_{j,n}$ depending on sample size n, wavelet basis ψ and resolution level j. At fine resolution levels ($j = 1, 2, ...$), we propose

$$t_{j,n} = \alpha_j \log n,$$

where $\{\alpha_j\}$ are level-dependent constants and at coarse levels ($j \gg 1$)

$$t_{j,n} = \frac{\pi}{\sqrt{3}} \sqrt{\log n}.$$

The purpose of this thresholding level is to make the reconstructed log-spectrum as nearly noise-free as possible. In addition to being pleasant from a visual point of view, the noise-free character leads to attractive theoretical properties over a wide range of smoothness assumptions. Previous proposals set much smaller thresholds and did not enjoy these properties.

Key Words and Phrases: Log Spectrum Estimation; Orthogonal Wavelet Transformation; Shrinkage Estimator.

Acknowledgements: This work was supported by NASA contract NCA2-488, by NSF DMS-84-51753 and by NSF DMS-92-09130. This is part of the author's Ph.D dissertation, under the supervision of Professor David L. Donoho, whose generous guidance and encouragement are gratefully acknowledged and appreciated.

1 Introduction

Suppose we want to study a time series from an observed segment of the series $X_1, ..., X_N$.

There are two aspects to the study of time series — analysis and modelling. The aim of analysis is to understand the nature of the process. The main reason for modelling a time series is to enable forecasts of future values of the process. No forecasting can be made before we understand the salient features of the process being characterized.

The analysis of a time series can be done either in the time domain or in the frequency domain. In the time domain, attention is focused on the relationship between observations at different points in time, e.g. ARMA models; while in the frequency domain it is cyclical movements which are studied and this can be done by studying the spectral density h. The two forms of analysis are complementary rather than competitive. They give different insights into the nature of the process.

Our goal in this paper is to estimate h from the data $X_1, ..., X_N$. Here

$$h(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \gamma(s) \cos(s\omega).$$

The techniques currently used in spectral analysis are:

- Window methods (also known as kernel methods), §6.2.3 of Priestley (1981). A spectral window (also called weight function) is used to smooth the periodogram, or equivalently a lag window (also known as weight sequences, Priestley (1981), pp. 437, covariance window, Bentkus and Sušinskas (1982)) $W(\cdot)$ is applied to the sample auto-covariances. The resulting estimate is:

  $$\hat{h}(\omega) = \frac{1}{2\pi} \sum_{|s|<N} W(s)\hat{\gamma}(s) \cos(s\omega)$$

  Typical choices of windows include the Féjèr window: $W(s) = I_{[|s|\leq M]}$ for some $1 \ll M \ll N$, for example $M = \sqrt{N}$; Bartlett window: $W(s) = (1 - |s|/M)_+$; Daniell window, $W(s) = \sin(\pi s/M)/(\pi s/M)$; see §6.2.3 of Priestley (1981) for more choices of windows. Bentkus and Sušinskas (1982) has studied optimal L_2 convergence rate for the window method over certain classes of smooth spectra.

- Autoregressive spectral estimation (AR approximation), Parzen (1974). A high order autoregressive model

  $$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

  is fit and the corresponding spectrum, a rational function,

  $$\hat{h}(\omega) = \frac{\hat{\sigma}^2}{2\pi|1 - \hat{\phi}_1 z - \cdots - \hat{\phi}_p z^p|^2}$$

  is taken as an approximation to f.

- Mixtures of the above, e.g. the prewhitening technique, §7.4.1 of Priestley (1981).
- Maximum entropy (ME) methods, introduced by Burg (1967, 1972). This method is essentially equivalent to AR method.
- ARMA approximation: similar idea as AR approximation.

They perform well in many cases and have been used frequently in practice. They have both advantages and disadvantages. For example, window methods have computational advantages but they may perform poorly in cases of high dynamic range; AR approximation is suitable only for relatively smooth spectra, Tukey (1978); ARMA procedures are almost always effective and furnishes attractive interpretations for the derived model; but it is computationally unrealistic to fit high-order ARMA models, and rigorous theory of asymptotic properties is lacking.

In this paper, we will develop a technique, based on wavelet decomposition of the periodogram and reconstruction of the spectrum; our new approach avoids shortcomings of earlier methods. Our new approach is computationally efficient; it can estimate spectra which are nonsmooth at a near-optimal rate; and it can be rigorously analyzed.

It is well known that the asymptotic variances of the empirical wavelet coefficients are functions of the unknown spectrum, and that non-linear Wavelet Shrinkage procedure below heavily depends on the asymptotic variances. Since the asymptotic variances are "proportional" to the coefficients, i.e., the bigger the coefficients at location (j,k), the higher the noise level at this location. This makes estimation more difficult in practice. In order to overcome the heteroscedasticity, we therefore consider estimating the log-spectrum.

Suppose we observe a segment $X_0, X_1, ..., X_{2n-1}$ of a Gaussian time series with mean zero and spectral density h, where for simplicity we assume that $n = 2^m$ is dyadic. In this paper we discuss the following 4-step wavelet shrinkage procedure for estimation of log spectrum, $g = \log h$:

[1] Calculate the log-periodogram

$$z_l = \log I(\omega_l), \qquad l = 0, \ldots, n-1,$$

where $\omega_l = l/2n$ and

$$I(\omega) = \frac{1}{4n\pi} \left| \sum_{t=0}^{2n-1} X_t e^{-it\omega} \right|^2 \tag{1}$$

[2] Take a standard periodic wavelet transform of (z_l) to get the empirical wavelet coefficients $\{y_{j,k}\}_{j=1,2,...m-1, k=0,...,n/2^j-1}$.

[3] Apply the soft threshold

$$\delta_t(x) = \mathrm{sgn}(x)(|x| - t)_+ \tag{2}$$

to the empirical wavelet coefficients $\{y_{j,k}\}$, with level-dependent thresholds $t = t_{j,n}$ according to formula

$$t_{j,n} = \frac{\pi}{\sqrt{3}}\sqrt{\log n}\bigvee \alpha_j \log n \tag{3}$$

where $(\alpha_1, ..., \alpha_{10}) = (1.29, 1.09, 0.92, 0.77, 0.65, 0.54, 0.46, 0.39, 0.32, 0.27)$ for the commonly used compactly supported orthogonal wavelet bases (*coiflets, daublets* and *symmlets*, Chapters 6 and 8 of Daubechies, 1992) and some selected n.

[4] Invert the wavelet transform, producing an estimate $(\hat{g}_l^*)$ of the log-spectrum at the Fourier frequencies ω_l.

This procedure falls in the general category of wavelet shrinkage estimates for noisy data. Donoho and Johnstone (1992a,b,c), have discussed the application of such thresholded wavelet transforms for recovering curves from noisy data, where the noise is assumed to have a Gaussian distribution, and they have proposed the level-independent threshold $t_n = \sigma\sqrt{2\log(n)}$, where σ is the noise level. Such methods have also been developed in density estimation: Johnstone, Kerkyacharian and Picard (1992) propose the use of level-dependent thresholds $\sqrt{j}$. In general such wavelet shrinkage methods have a number of theoretical advantages, including near-optimal mean-squared error and near-ideal spatial adaptation.

Recently, the author and others have studied the possibility of using wavelet shrinkage to de-noise periodogram data. Application of wavelet shrinkage to the log-periodogram is particularly attractive, Moulin (1993a, b), since the logarithm is the variance stabilizing transformation for the periodogram, Wahba (1980).

It is easy to apply wavelet shrinkage software designed for Gaussian noise to the log periodogram, and in some cases acceptable reconstructions have been obtained. However, this will not always be the case. The noise in the wavelet coefficients of the log-periodogram has a non-Gaussian character; it reaches high levels somewhat more frequently than Gaussian theory would predict. Consequently, thresholds set based on Gaussian theory will not be high enough to completely suppress noise in the coefficients. The thresholding we propose is based on a careful analysis of the non-Gaussian character, and is somewhat larger than Gaussian theory would predict.

As an example of the difference between our method and other proposals, we present in Figure 1 a display for an $AR(24)$ time series: (a) its log spectrum; (b) log periodogram; and wavelet reconstructions based on (c) our proposal and (d) based on Gaussian theory. It is visually evident that the spectral estimate based on Gaussian theory has spurious noise spikes, which are unrelated to the true underlying spectrum. More plots for this example, along with other examples, for different sample sizes and different wavelet bases, are given at the end of the paper.

The pattern visually evident in this example is confirmed by two theorems which we prove in this paper.

Theorem 1 *Under the Wahba approximation for log periodograms (see section 2.2 below), as* $n \to \infty$,

$$P\{\bigcup_j [\sup_k |y_{j,k} - Ey_{j,k}| > t_{j,n}]\} \to 0. \tag{4}$$

In words, the thresholds are set high enough so that the noise does not surpass them.

Theorem 2 *Under the Wahba approximation for log periodograms and a compactly supported wavelet basis, at the finest level* $j = 1$, *as* $n \to \infty$,

$$P\{\sup_k |y_{1,k} - Ey_{1,k}| > \sqrt{2 \log n}\} \to 1. \tag{5}$$

Similar results hold at levels $j = 2, 3, ...$

In words: noise spikes are almost surely to be absent for our proposal and almost surely to be present in a proposal based on Gaussian noise.

The significance of these noise spikes and the need to de-noise is more than cosmetic. In a recent article, Donoho, Johnstone, Kerkyacharian and Picard (1992) study the Gaussian white noise model, and show that by exploiting a noise-free property like (4) one can show that the estimator attains near-optimal reconstruction in a wide variety of norms simultaneously over a wide variety of smoothness assumptions. Their arguments are general and abstract and transfer to the present setting after one has established the de-noising property (4)

If noise-free property like (4) is not maintained, then an inspection of their arguments will show that the resulting estimate does not achieve near optimal reconstruction in norms which measure smoothness of the reconstruction, although it may well be true that the error in ℓ_2 norm, which does not measure smoothness, is still acceptable.

2 The Choice of Thresholds

Our choice of the thresholds $(t_{j,n})$ depends on the tail behavior of the empirical coefficients $\{y_{j,k}\}$. Let

$$p_j(t) = \max_k P\{|y_{j,k} - Ey_{j,k}| > t\} \tag{6}$$

be the tail probability of $y_{j,k}$, where wavelet coefficient $y_{j,k}$ is a standardized linear combination of the log periodogram ordinates. Then we aim to establish that

$$n\, p_j(t_{j,n}) \to 0, \qquad n \to \infty, \tag{7}$$

uniformly for $j = 1, 2, ..., m - 1$ as this implies (4) in Theorem 1 (see (15). In this section, we would like to give two heuristic arguments, upon which our proposed thresholds (3) are based.

2.1 Normal Approximation

For normally distributed wavelet coefficients

$$w_{j,k} = \alpha_{j,k} + \varepsilon_{j,k}$$

where $\varepsilon_{j,k} \sim N(0, \sigma^2)$ are iid, the threshold

$$t_{j,n} \equiv t_n = \sigma\sqrt{2\log n}$$

enjoys a variety of nice theoretical properties, Donoho and Johnstone (1992a, b, c). Under certain regularity conditions, for j close to $m = [\log_2 n]$ (technically $j \to \infty$ as $m \to \infty$), as $n \to \infty$, individual empirical wavelet coefficients will be asymptotically normal:

$$w_{j,k} - Ew_{j,k} \sim N(0,\ \pi^2/6)$$

this follows by applying results of Taniguchi (1979, 1980). Moulin (1993a, b) makes similar argument. So when the sample size n is large, for coarse levels (j close to m), $y_{j,0}, ..., y_{j,2^j-1}$ are approximately normally distributed with the same variance $\pi^2/6$. One may argue that

$$p_j(\pi\sqrt{\log(n)/3}) \le \frac{1+o(1)}{n\sqrt{2\log n}}, \qquad m > j \to \infty. \tag{8}$$

Therefore (7) holds and this leads to the first part of our thresholds.

Here "regularity" refers to the length of memory and the asymptotic normality for the empirical coefficients not only depends on n (large) and j (large), but also the regularity of the time series. In general, the longer the memory, the weaker the asymptotic normality.

2.2 Wahba Approximation

For small j, the normal approximation deteriorates. Under a compactly supported wavelet basis (chapters 6 and 8 of Daubechies, 1992), for fixed j, the coefficients of the finest level $y_{j,k}$'s are linear combinations of *fixed* number of $\{\log I(\omega_l)\}$. For example, with the Haar wavelet, $y_{1,k}$ is just the difference of two adjacent $\log I(\omega_l)$'s.

Consider the following non-Gaussian additive noise model:

$$z_l = g_l + \varepsilon_l \tag{9}$$

$l = 1, 2, ..., n-1$, where (g_l) is the object to be estimated (e.g. log spectrum at frequency ω_l) and

$$\varepsilon_i = \log(\eta_i/2) + \gamma \tag{10}$$

where $\{\eta_i, i = 1, 2, ..., n-1\}$ are independently $\chi^2_2 = \exp(1/2)$ distributed and $\gamma = 0.57721$ is the Euler-Mascheroni constant. It can be shown that $E\varepsilon_i = 0$ and $\mathrm{var}(\varepsilon_i) = \pi^2/6$. Wahba (1980) proposed this as a model for the log periodogram, i.e., $z_l \approx \log I(\omega_l)$ and

$g_l \approx \log f(\omega_l) - \gamma$. Note that $I(\omega_0) \sim \chi^2_1$. Since the influence of this term is negligible for n large, we will ignore this term in our discussion.

For circulant time series, see §4.3 of Harvey (1989), $\log I(\omega_l)$ follows model (9)–(10) exactly. For general stationary processes, the above model is only asymptotically true (see Theorem 5.2.6 of Brillinger, 1981). The exact distribution of log periodogram can be found in Wittwer (1986).

Now we will show that under the model (9)–(10), our advertised thresholds are indeed adequate for the finer levels. Note that under model (9)–(10), each $y_{j,k}$ is a standardized linear combination of z_i's.

Let us use the very finest level, $j = 1$, as an example. Similar results hold for levels $j = 2, 3, ...$ Suppose a compactly supported wavelet basis with L coefficients in the dilation equation. For example, for the Haar wavelet, $L = 2$; D4 wavelet, $L = 4$, etc., more examples can be found in Chapters 6 and 8 of Daubechies (1992). Then

$$y_{1,k} = \sum_{l=1}^{L} a_l z_{l+2k-s}$$

where s is some shift parameter for computational purpose (e.g. $s = L/2$) and $\sum a_l = 0$, $\sum a_l^2 = 1$. Inequalities below leads to the threshold (3).

Let $a = \max_{1 \le l \le L} |a_l|$, from the proofs of the Theorems (see (16) and (22) of Appendix), we can show that for t large,

$$0.25 e^{-2t/a} \le p_1(t) \le (et/aL)^L e^{-t/a}. \tag{11}$$

Hence, for n large, there is a constant A so that

$$\frac{0.25}{n^{2\alpha/a}} \le p_1(\alpha \log n) \le \frac{A(\log n)^L}{n^{\alpha/a}}. \tag{12}$$

Hence thresholds at the finest level must have the form: $\alpha \log n$, for some constant $\alpha > a$, and then

$$n p_1(\alpha \log(2n)) \to 0, \qquad n \to \infty. \tag{13}$$

Note that the wavelet filters obey $\sum_{l=1}^{L} a_l^2 = 1$, which implies $a < 1$. Hence, for the finest level, taking $\alpha = 1$ always guarantees (13) and (7) follows for our proposal $t_{j,n}$.

In general, there are L_j non-zero constants $a_{i,j}$ so that

$$y_{j,k} = \sum_{i=1}^{L_j} a_{i,j} z_{i \oplus k2^j} \tag{14}$$

where $\oplus$ in subscript is interpreted modulo n. Let $a_j = \max_i |a_{i,j}|$, then we have

$$L_j \le 2^j(L-1) - L + 2 \ \text{ and } \ a_j \le \alpha 2^{-(j-1)/2},$$

where L is the length of the filter and $\alpha = \sup_{j \ge 1} 2^{(j-1)/2} a_j < 1.534$. In Tables 1 and 2 of Appendix, we list some L_j's and $2^{(j-1)/2} a_j$'s for the commonly used wavelets (apparently they converge to $\sup_x |\psi(x)|$). This leads to the threshold (3).

3 Numerical Examples

In this section, we study four time series and compare our proposed method with the one based on Gaussian theory.

Figure 1 contains an AR(24) signal, with coefficients:

-2.5216281	4.7715359	-7.9199915	11.9769211	-16.0778828
20.6343346	-25.0531521	28.8738136	-31.8046265	34.0071373
-34.7700272	34.3151321	-32.7861099	30.2861233	-26.7109356
22.8838310	-18.7432098	14.5717688	-10.7177744	7.5322194
-4.7226319	2.6807923	-1.3391306	0.5167125	

Figure 2 contains a white noise siganl, the true log spectral density in this case is $g(\omega) = \log(\sigma^2/2\pi)$, a constant.

Figure 3 contains an MA(15001) time series with coefficients

$$a_1 = \pi/4, \quad a_{n+1} = \frac{\sin(\pi n/2)}{n}, \quad n = 1, 2, ..., 15000.$$

AR method is compared in this example.

Figure 4 contains the Sunspots signal, WaveShrink estimate with comparisons with an AR estimate from SPLUS function *spec.ar*. An AR(2) model, suggested by Priestley (1981), pp. 882, is also plotted.

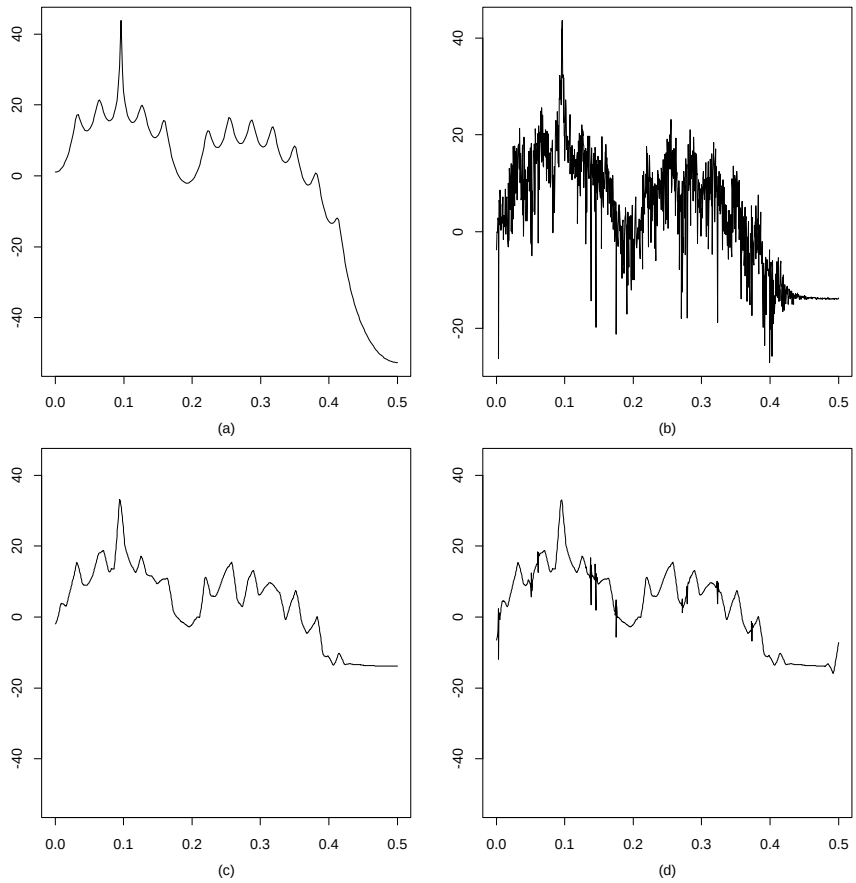

Figure 1: (a). log spectrum of an AR(24) time series. (b). log periodogram, (c). WaveShrink estimate based the new thresholding scheme, and (d). WaveShrink estimate based on Gaussian theory.

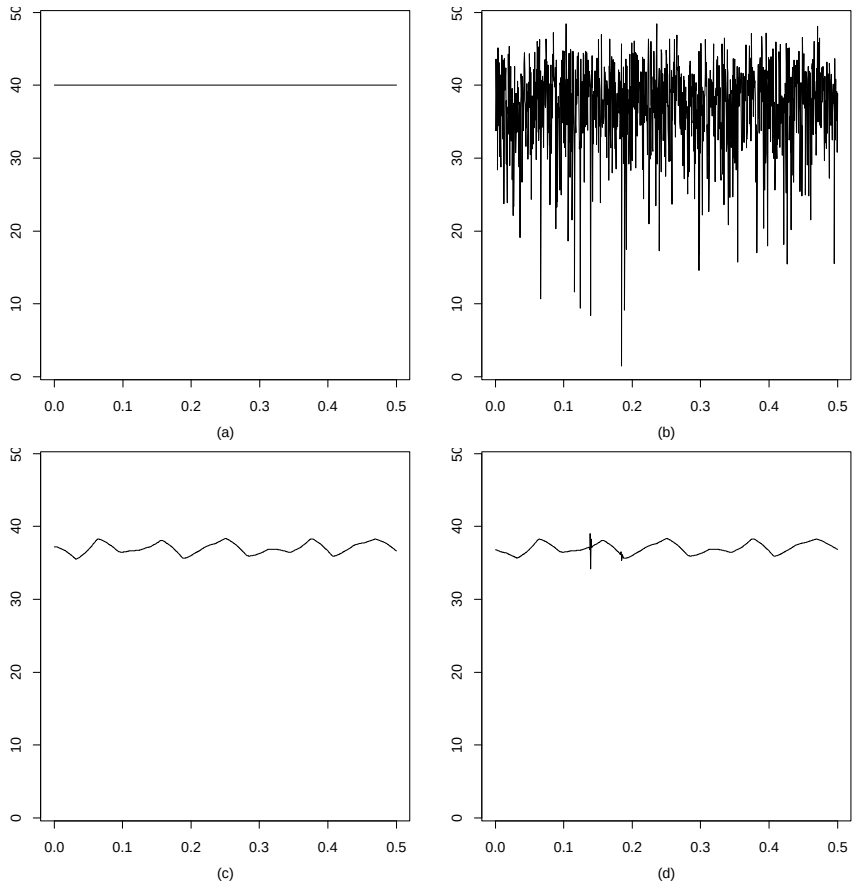

Figure 2: (a). log spectrum of a white noise signal, (b). log periodogram, (c). WaveShrink estimate based the new thresholding scheme, and (d). WaveShrink estimate based on Gaussian theory.

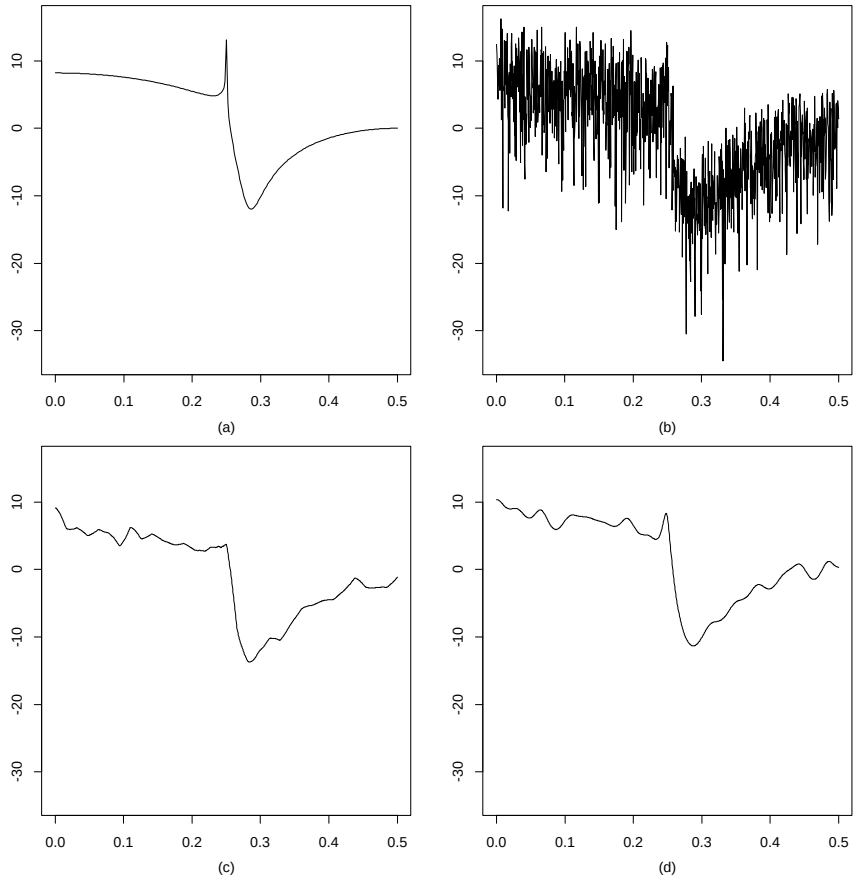

Figure 3: (a). log spectrum of a long MA series, (b). log periodogram, (c). WaveShrink estimate based the new thresholding scheme, and (d). AR estimate from SPLUS function *spec.ar*.

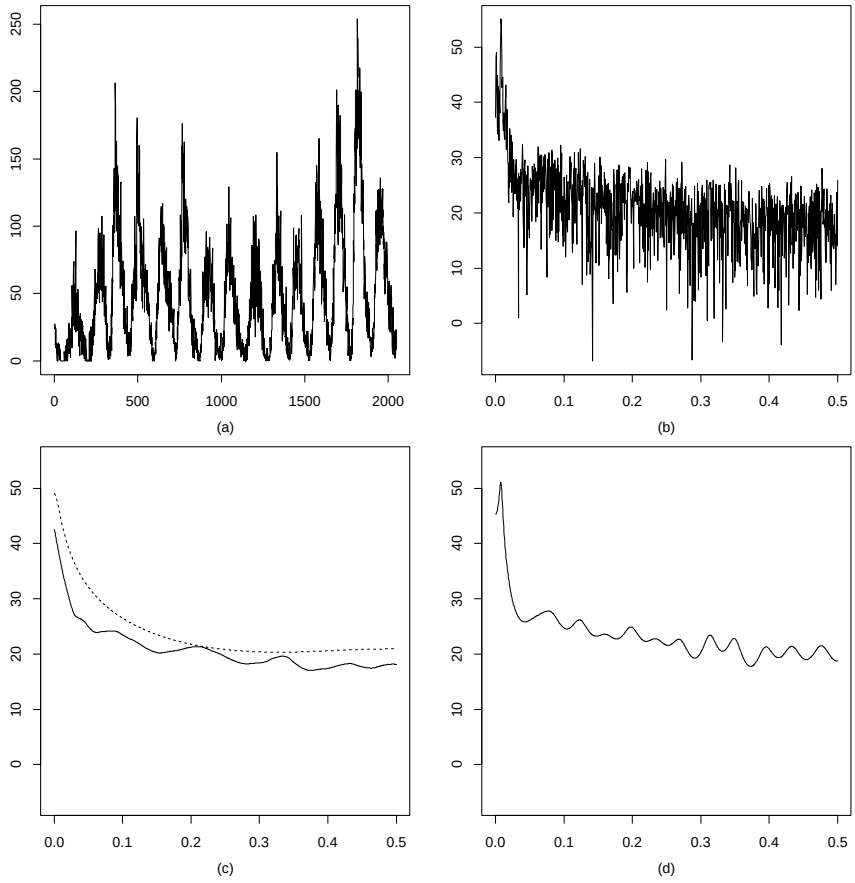

Figure 4: (a). sunspots data, (b). log periodogram, (c). WaveShrink estimate (solid line) and AR(2) estimate (dotted line), and (d). AR (of order 27) estimate.

4 Appendix

4.1 Tables

	L1	H1	L2	H2	L3	H3	L4	H4	L5	H5	L6	H6
C6	0.85	0.85	0.95	1.10	1.02	1.28	1.07	1.40	1.10	1.48	1.13	1.53
C12	0.81	0.81	0.85	1.01	0.86	1.08	0.87	1.10	0.87	1.11	0.87	1.12
C18	0.79	0.79	0.81	0.96	0.81	0.99	0.81	1.00	0.81	1.00	0.81	1.00
C24	0.78	0.78	0.79	0.93	0.79	0.94	0.79	0.95	0.79	0.95	0.79	0.95
C30	0.77	0.77	0.78	0.91	0.78	0.91	0.78	0.92	0.78	0.92	0.78	0.92
D2	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71
D4	0.84	0.84	0.95	1.08	0.99	1.23	1.00	1.28	1.00	1.29	0.99	1.28
D6	0.81	0.81	0.83	0.98	0.86	1.03	0.89	1.11	0.90	1.15	0.90	1.18
D8	0.71	0.71	0.79	0.77	0.78	0.95	0.79	0.96	0.79	0.94	0.79	0.96
D10	0.72	0.72	0.81	0.94	0.74	0.89	0.75	0.83	0.76	0.84	0.76	0.84
D12	0.75	0.75	0.75	0.92	0.74	0.80	0.73	0.80	0.73	0.80	0.73	0.8
D14	0.73	0.73	0.74	0.79	0.73	0.82	0.73	0.79	0.73	0.79	0.73	0.79
D16	0.68	0.68	0.74	0.82	0.71	0.82	0.71	0.79	0.71	0.78	0.71	0.78
D18	0.66	0.66	0.72	0.87	0.70	0.74	0.70	0.75	0.70	0.76	0.70	0.76
D20	0.69	0.69	0.69	0.82	0.68	0.75	0.68	0.73	0.69	0.72	0.69	0.73
S8	0.80	0.80	0.85	0.99	0.85	1.05	0.85	1.05	0.85	1.06	0.86	1.07
S10	0.72	0.72	0.77	0.77	0.78	0.89	0.79	0.93	0.79	0.94	0.79	0.95
S12	0.79	0.79	0.80	0.95	0.80	0.97	0.81	0.97	0.81	1.00	0.81	1.00
S14	0.77	0.77	0.79	0.92	0.79	0.95	0.77	0.92	0.77	0.89	0.77	0.89
S16	0.78	0.78	0.78	0.91	0.78	0.92	0.79	0.93	0.79	0.95	0.79	0.95
S18	0.72	0.72	0.76	0.76	0.77	0.89	0.77	0.89	0.77	0.89	0.77	0.89
S20	0.77	0.77	0.77	0.89	0.77	0.89	0.78	0.91	0.78	0.92	0.78	0.92

Table 1: Maximum Coefficients, Lj low-pass cascade filter at level j, Hj high-pass cascade filter at level j. C for Coiflets; D for Daublets, Daubechies' original wavelets; S for Symmlet, Daubechies' near-symmetric wavelets.

	L=2	L=4	L=6	L=8	L=10	L=12	L=14	L=16	L=18	L=20
j=1	2	4	6	8	10	12	14	16	18	20
j=2	4	10	16	22	28	34	40	46	52	58
j=3	8	22	36	50	64	78	92	106	120	134
j=4	16	46	76	106	136	166	196	226	256	286
j=5	32	94	156	218	280	342	404	466	528	590
j=6	64	190	316	442	568	694	820	946	1072	1198

Table 2: Number of Non-zero Coefficients, L is the length of the filter and j = level.

4.2 Proofs of the Theorems

Proof of Theorem 1: First of all,

$$\begin{aligned} P\{\bigcup_j [\sup_k |y_{j,k} - Ey_{j,k}| > t_{j,n}]\} &\leq \sum_{j=1}^{m-1} P\{\sup_k |y_{j,k} - Ey_{j,k}| > t_{j,n}\} \\ &\leq \sum_{j=1}^{m-1} 2^j p_j(t_{j,n}) \leq n \max_{1\leq j<m} p_j(t_{j,n}) \end{aligned} \tag{15}$$

Under model (9) and (10), from (14), coefficients in level j can be written as

$$y_{j,k} - Ey_{j,k} = \sum_{i=1}^{L_j} a_{i,j}\varepsilon_{i\oplus k2^j}$$

with $\sum_i a_{i,j} = 0$ and $\sum_i a_{i,j}^2 = 1$. Let $a_j = \max_i |a_{i,j}|$.

The moment generating function for $\varepsilon \sim \log(\chi_2^2/2)$ is

$$M(u) \equiv Ee^{u\varepsilon} = \Gamma(u+1)$$

where Γ is Euler's Gamma function. Let

$$\Lambda_j^*(t) = \inf_s\{-st + \sum_{i=1}^{L_j}(\log\Gamma(1+a_{i,j}s) + \log\Gamma(1-a_{i,j}s))\}$$

It can be shown that for any random variable X with $EX = 0$,

$$P(X > a) \leq e^{-at}Ee^{tX}$$

for all $a > 0$ and t. Then

$$p_j(t) \leq e^{\Lambda_j^*(t)}$$

and therefore,

$$\max_{1\leq j<m} p_j(t) \leq \sup_{1\leq j<m} e^{\Lambda_j^*(t)} = \exp(\sup_{1\leq j<m} \Lambda_j^*(t)) \tag{16}$$

Let $\ell_n = (\log n)^{7/6}$ and consider

$$t_j^* = \begin{cases} \frac{\alpha \log n}{L_j^{1/4}} & L_j < \ell_n \\ \beta\sqrt{\log n} & L_j \geq \ell_n \end{cases} . \tag{17}$$

For the thresholds

$$t_{j,n} = \frac{\alpha \log n}{L_j^{1/4}} \bigvee \beta\sqrt{\log n}$$

we have $t_{j,n} \geq t_j^*$. Therefore,

$$\sup_{1\leq j<m} p_j(t_{j,n}) \leq \sup_{1\leq j<m} p_j(t_j^*) \leq \exp(\sup_{1\leq j<m} \Lambda_j^*(t_j^*))$$

and

$$\begin{aligned} \sup_{1\leq j<m} \Lambda_j^*(t_j^*) &\leq \sup_{1\leq L_j<\ell_n} \Lambda_j^*(t_j^*) \bigvee \sup_{\ell_n\leq L_j\leq n} \Lambda_j^*(t_j^*) \\ &= \sup_{1\leq L_j<\ell_n} \Lambda_j^*(\frac{\alpha \log n}{L_j^{1/4}}) \bigvee \sup_{\ell_n\leq L_j\leq n} \Lambda_j^*(\beta\sqrt{\log n}) \end{aligned} \tag{18}$$

When $L_j < \ell_n = (\log n)^{7/6}$ and n large, we have

$$\frac{\alpha \log n}{L_j^{1/4}} > \alpha\sqrt{L_j} \geq a_j L_j$$

Consider the following non-positive function H, defined on R_+ by

$$H(t) = -t + 1 + \log t$$

Lemma 1 *For $t \geq a_j L_j$,*

$$\Lambda_j^*(t) \leq L_j H(\frac{t}{a_j L_j})$$

and

$$L\ H(\frac{\alpha \log n}{a L^{3/4}})\ \downarrow\ \textit{in } L \textit{ for large } n \textit{ and } 1 \leq L \leq \ell_n.$$

The proof will be given later. Combining (16) and Lemma 1 yields the second part of (11).

From the Lemma,

$$\begin{aligned} \sup_{1\leq L_j<\ell_n} \Lambda_j^*(\frac{\alpha \log n}{L_j^{1/4}}) &\leq \sup_{1\leq L_j<\ell_n} L_j H(\frac{\alpha \log n}{a L_j^{3/4}}) \\ &\leq H(\frac{\alpha}{a} \log n) = -\frac{\alpha}{a} \log n\ (1 + o(1)) \end{aligned} \tag{19}$$

When $L_j \geq \ell_n$, we have

$$\beta\sqrt{\log n} \leq \beta L_j^{3/7}$$

and

Lemma 2 *For $\delta_j \in (0,1)$ and $0 \le t \le \pi^2\delta_j/6a_j$,*

$$\Lambda_j^*(t) \le -\frac{t^2}{\pi^2/3 + 6\delta_j}.$$

In particular, for n large and $\delta_j \equiv 6\beta a_j L_j^{3/7}/\pi^2 \asymp L_j^{-1/14} \le (\log n)^{-1/12}$,

$$\Lambda_j^*(t) \le -\frac{3t^2}{\pi^2}(1+o(1)) \qquad t \le \beta L_j^{3/7}.$$

The proof will be given later.

From the Lemma,

$$\sup_{\ell_n \le L_j \le n} \Lambda_j^*(\beta\sqrt{\log n}) \le -\frac{3\beta^2 \log n}{\pi^2}(1+o(1)). \tag{20}$$

Combining (18), (19) and (20), we have

$$\sup_{1\le j<m} \Lambda_j^*(t_j^*) \le -(\frac{\alpha}{a} \bigwedge \frac{3\beta}{\pi^2}) \log n \ (1+o(1)).$$

Therefore, for any $\alpha > a$ and $\beta > \pi/\sqrt{3}$,

$$n \sup_{1\le j<m} p_j(t_{j,n}) \le n \exp(\sup_{1\le j<m} \Lambda_j^*(t_{j,n})) = o(1).$$

and this completes the proof of Theorem 1.

Proof of Lemma 1: First of all, it can be shown that for $0 \le x < 1$, $\Gamma(1+x) \le \Gamma(1-x)$ and $\Gamma(x) \le 1/x$. (In fact, $\Gamma(x) = x^{-1}\prod_{k=1}^{\infty} \frac{(1+1/k)^x}{1+x/k}$ and $(1+1/k)^x \le 1 + x/k$ when $0 \le x < 1$.) Therefore,

$$\begin{aligned} \Lambda_j^*(t) &\le \inf_{|s|<1/a_j} \{-st + \sum_{i=1}^{L_j} \log\Gamma(1+a_{i,j}s)\} \\ &\le L_j \inf_{0\le s<1} \{-\frac{st}{a_j L_j} - \log(1-s)\} \end{aligned}$$

and the first part of the Lemma can be easily solved from the last expression.

For n large, such that $C_n \equiv \alpha a^{-1}\log n > 11$ and $\ell_n = (\log n)^{7/6} \le (C_n/11)^{4/3}$. Let $h(x) = xH(C_n x^{-3/4})$, then for $1 \le x < \ell_n$, $h'(x)$ is increasing in x and $h'(\ell_n) < 0$. Therefore, $h'(x) < 0$ for $1 \le x < \ell_n$ and this means $h(x)$ is decreasing in x. The proof is completed.

Proof of Lemma 2: Let $K(x) = \log\Gamma(1+x)$, then for $x > -1$, by Taylor expansion,

$$K(x) = K(0) + K'(0)x + x^2K''(\theta x)/2 = -\gamma x + x^2K''(\theta x)/2, \tag{21}$$

where $\gamma = 0.5772....$ is the Euler constant and $\theta = \theta(x) \in (0,1)$. Note that $K'(x-1)$ is the famous Psi function. By the properties of Psi function (Davis (1935), pp. 11), $K''(x) = \sum_{k=1}^{\infty}(k+x)^{-2}$ and $K''(0) = \pi^2/6$. Hence, for any $\delta \in (0,1)$, we have $K''(-\delta) \leq K''(0) + 3\delta \equiv K_\delta$ and for $x \geq -\delta$,

$$K(x) \leq -\gamma x + x^2 K_\delta/2.$$

Recall that $\sum_i a_{i,j} = 0$ and $\sum_i a_{i,j}^2 = 1$,

$$\begin{aligned}
\Lambda_j^*(t) &= \inf_s\{-st + \sum_i K(a_{i,j}s)\} \\
&\leq \inf_{|s|<\delta_j/a_j}\{-st + \sum_i(-\gamma a_{i,j}s + a_{i,j}^2 s^2 K_\delta/2)\} \\
&= \inf_{|s|<\delta_j/a_j}\{-st + s^2 K_\delta/2\}
\end{aligned}$$

and Lemma 2 follows immediately.

Proof of Theorem 2: To prove Theorem 2, we need a lower bound for $p_j(t)$. It is easy to show that if X, Y are independent, then for any $s, t > 0$,

$$P\{|X+Y| \geq t\} \geq P\{|X| < s\}P\{|Y| \geq t+s\}.$$

Suppose $|a_1| = \max|a_l| = a$. There exists $t_0 > 0$ such that for $t \geq t_0$,

$$P\{|\sum_{l=2}^{L} a_l \log(\eta_l/2)| < t\} \geq \frac{1}{2}.$$

Also notice that for $0 \leq x \leq 1$, $1 - e^{-x} \geq x/2$. Then

$$\begin{aligned}
p_1(t) &= P\{|\sum_{l=1}^{L} a_l \log(\eta_l/2)| \geq t\} \\
&\geq P\{|a_1 \log(\eta_1/2)| \geq 2t\}P\{|\sum_{l=2}^{L} a_l \log(\eta_l/2)| < t\} \\
&\geq \frac{1}{2}P\{|\log(\eta_1/2)| \geq \frac{2t}{a}\} \\
&\geq \frac{1}{2}P\{\log(\eta_1/2) \leq -\frac{2t}{a}\} \\
&= \frac{1}{2}\{1 - \exp(-e^{-2t/a})\} \geq \frac{1}{4}e^{-2t/a}. \qquad (22)
\end{aligned}$$

and this completes the proof of Lemma 1.

From (12) we can see that for any constant $C < a/2$,

$$np_1(C\log n) \to \infty, \qquad n \to \infty.$$

In particular,

$$np_1(\sqrt{2\log n}) = np_1(\sqrt{\frac{2}{\log n}}\log n) \to \infty, \qquad n \to \infty.$$

There exists $K > 0$ such that $\{y_{1,iK}\}$ are independent (linear combinations over disjoint intervals). Therefore,

$$\begin{aligned} P\{\sup_k |y_{1,k} - Ey_{1,k}| > \sqrt{2\log n}\} &\geq P\{\sup_i |y_{1,iK} - Ey_{1,iK}| > \sqrt{2\log n}\} \\ &= 1 - \prod_{i=0}^{n/2K} (1 - p_1(\sqrt{2\log n})) \to 1, \end{aligned}$$

and this implies the Theorem 2.

References

[1] Bentkus, R. Yu. and Sušinskas, Ju. V. (1982) On Optimal Statistical Estimators of the Spectral Density. *Soviet Math. Dokl.*, Vol. **25**, No. **2**, 415-419.

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

[3] Burg, J. P. (1967). Maximum Entropy Spectral Analysis. Paper Presented at the 37th Annual international S.E.G. Meeting, Oklahoma City, Oklahoma.

[4] Burg, J. P. (1972). The Relationship between Maximum Entropy Spectra and Maximum Likelihood Spectra. *Geophysice*, **37**, 375-376.

[5] Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM. Philadelphia.

[6] Davis, H. T. (1935). *Tables of The Mathematical Functions*, Vol II, Trinity University.

[7] Donoho, D. (1992). De-Noising By Soft-Thresholding. Technical Report, Department of Statistics, Stanford University.

[8] Donoho, D. and Johnstone, I (1992a). Minimax Estimation via Wavelet Shrinkage. Technical Report, Department of Statistics, Stanford University.

[9] Donoho, D. and Johnstone, I (1992b). Adapting to Unknown Smoothness via Wavelet Shrinkage. Technical Report, Department of Statistics, Stanford University.

[10] Donoho, D. and Johnstone, I (1992c). Ideal Spatial Adaptation by Wavelet Shrinkage. Technical Report, Department of Statistics, Stanford University.

[11] Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1992). Wavelet Shrinkage: Asymptopia? Technical Report, Department of Statistics, Stanford University.

[12] Gao, Hong-Ye (1992). Spectral Density Estimation by Using Wavelets. Technical Report, Department of Statistics, Stanford University.

[13] Harvey, A. C. (1981). *Time Series Models*, Philip Allan.

[14] Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

[15] Johnstone, I., Kerkyacharian, G. and Picard, D. (1992). Estimation d'une densité de probabilité par méthode d'ondelettes. *Comptes Rendus Acad. Sciences Paris* (A), **315** 211-216.

[16] Moulin, P. (1993a) A Wavelet Regularization Method for Diffuse Radar-Target Imaging and Speckle-Noise Reduction. *Journal of Math. Imaging and Vision*, Special Issue on Wavelet, Vol. 3, 123-134.

[17] Moulin, P. (1993b) Wavelet Thresholding Techniques for Power Spectrum Estimation. Submitted to *IEEE Trans. on Signal Processing*.

[18] Parzen, E. (1974). Some Recent Advances in Time Series Modelling. *IEEE Trans. Automatic Control*, **AC-19**, 723-729.

[19] Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Academic Press, London.

[20] Taniguchi, M. (1979). On Estimation of Parameters of Gaussian Stationary Processes. *J. Appl. Probab.*, **16**, 575-591.

[21] Taniguchi, M. (1980). On Estimation of the Integrals of Certain Functions of Spectral Density. *J. Appl. Probab.*, **17**, 73-83.

[22] Tukey, J. W. (1978). When Should Which Spectrum Approach Be Used? Paper Presented at the Institute of Statisticians Conference, King's College, Cambridge, July, 1978.

[23] Wahba, G. (1980). Automatic Smoothing of the Log Periodogram. *Journal of the American Statistical Association*, **75**, 122-132.

[24] Wittwer, G. (1986) On the Distribution of the Periodogram for Stationary Random Sequences (with Discussion). *Math. Opera. und Stat.*, **17**, 201-219.