A wavelet-Fisz approach to spectrum estimation

Piotr Fryzlewicz*, Guy P. Nason
Department of Mathematics,
University of Bristol,
University Walk,
Bristol BS8 1TW,
UK.
{p.z.fryzlewicz,g.p.nason}@bris.ac.uk

and

Rainer von Sachs
Institut de statistique,
Université Catholique de Louvain,
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve,
Belgium.
rvs@uclouvain.be

March 14, 2008

*Author for correspondence.

Abstract

We propose a new approach to wavelet threshold estimation of spectral densities of stationary time series. Our proposal addresses the problem of heteroscedasticity and non-normality of the (tapered) periodogram. We estimate thresholds for the empirical wavelet coefficients of the periodogram as appropriate linear combinations of the periodogram values similar to empirical scaling coefficients. Our solution introduces "asymptotically noise-free reconstruction thresholds" which parallels classical wavelet theory for nonparametric regression with iid Gaussian errors. Our simulations show promising results that clearly improve on existing approaches. In addition, we derive theoretical results on the near-optimal rate of convergence of the minimax mean-square risk for a class of spectral densities, including those of low regularity.
Key words: spectral density estimation, wavelet thresholding, wavelet-Fisz, periodogram, Besov spaces, smoothing.

1 Introduction

The estimation of spectral densities is a fundamental problem in both theoretical and applied time series analysis (Priestley (1981)). Typically, inference in the spectral domain uses the periodogram of the data x_t. A data taper is often applied prior to periodogram computation in order to reduce leakage (Dahlhaus (1983)). The (tapered) periodogram is an inconsistent estimator and needs to be smoothed to achieve consistency. For highly regular spectral densities linear smoothing techniques (e.g. kernel smoothing) are appropriate, see e.g. Brillinger (1981), Koopmans (1995) and Shumway and Stoffer (2000). Multiple tapers lead to, both leakage-reduced and smoothed, multitaper spectrum estimators, see e.g. Cristan and Walden (2002).

However, linear smoothing methods are incapable of achieving the optimal mean-square rate of convergence for spectra whose smoothness is distributed inhomogeneously over the domain of interest, and a need for non-linear methods arises. An interesting example of a highly irregular spectral density (which satisfies all our technical assumptions specified later) is given e.g. in Neumann (1994), Section 8, item 2. Nonlinear methods also have benefits for spectra with multiple (sharp) peaks and troughs. Two nonlinear methods for spectral density estimation were recently proposed by Comte (2001) and Davies and Kovac (2004).

In the "function + iid Gaussian noise" regression model, wavelet thresholding, first proposed by Donoho and Johnstone (1994), has become the nonlinear smoothing method of choice. However, the periodogram approximately follows a multiplicative regression set-up where the "noise" variance is proportional to the level of the underlying spectral density. This is a hurdle for nonlinear wavelet thresholding, where the noise variance needs to be either known or easily estimable. To tackle this problem, two main approaches have been proposed in the literature.

The first approach took the log transform of the periodogram to stabilize the variance, and only then proceeded with the wavelet smoothing, see e.g. Gao (1993, 1997), Moulin (1994) and Pensky and Vidakovic (2003). The "price" for using the log transform is that it flattens out the data, often obscuring interesting features, such as peaks or troughs. Also, the resulting exponentiated estimators of the spectrum are biased, and even after the bias correction, their mean-square properties are not easy to establish.

The second approach (Neumann, 1996) consists in pre-estimating the variance of the periodogram via kernel smoothing, so that it can be supplied to the wavelet estimation procedure. As with other plug-in estimators, the question of the choice of the pre-estimation procedure and its parameters arises. Also, kernel pre-estimation may not be appropriate in cases where the underlying spectral density is of low regularity.

To overcome the drawbacks of the above log-based and plug-in estimators, we propose a new nonlinear wavelet technique for periodogram smoothing, where the empirical wavelet coefficient thresholds are built as appropriate local weighted l_1 norms of the periodogram, as opposed to

the l_2 norm used in Neumann (1996). Section 2 explains that the use of the l_1 norm is motivated by the fact that, asymptotically, the mean of the periodogram is equal to its standard deviation. Also, unlike Neumann (1996), we avoid kernel pre-estimation of the spectral density. Our approach yields a rapidly computable, mean-square consistent estimator which performs well in practice. Also, it permits the construction of *noise-free reconstruction thresholds* which produce visually appealing estimates and offer particularly impressive empirical performance.

2 Set-up and motivation

2.1 Problem set-up

Assume that we observe a sample path $\{X_t\}_{t=1}^N$ of a real-valued, zero-mean, second-order stationary process $\{X_t\}_{t=1}^\infty$. Our aim is to estimate the spectral density

$$f(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \mathrm{cov}(X_t, X_{t+s}) \exp(-i\omega s), \quad \omega \in [-\pi, \pi].$$

We place the following assumption on f.

Assumption 2.1 *The spectral density $f(\omega)$ satisfies (i) $f(\omega) \geq \mu > 0$, (ii) f is continuous and of finite total variation over $[-\pi, \pi]$.*

Assumption 2.1(i) is essential in obtaining strong asymptotic normality of the empirical wavelet coefficients of the periodogram of X_t. Assumption 2.1(ii) is a mild smoothness assumption on f. We also place the following technical assumption on the process itself.

Assumption 2.2 *Assume*

$$\sup_{1 \leq t_1 < \infty} \left(\sum_{t_2, \ldots, t_k = 1}^{\infty} |\,\mathrm{cum}(X_{t_1}, \ldots, X_{t_k})| \right) \;\leq\; C^k (k!)^{1+\gamma},$$

for all $k = 2, 3, \ldots$, where C is a generic positive constant and $\gamma \geq 0$.

As in Neumann (1996) Assumption 2.2 implies, uniformly over t, asymptotic normality of the (appropriately scaled) local cumulative sums of X_t. By Remark 3.1 in Neumann (1996), if X_t is α-mixing at an appropriate rate and its marginal distribution is Gaussian, exponential, gamma, or inverse Gaussian, then γ can be set equal to zero. For heavier-tailed distributions, a positive value of γ might be required.

Our nonparametric estimator is based on the periodogram of the (possibly tapered) data

$$I_N(\omega) = (2\pi H_2^{(N)})^{-1} \left| \sum_{s=1}^{N} h\,(s/N)\; X_s \; \exp(-i\omega s) \right|^2 ,$$

where $H_k^{(N)} = \sum_{s=1}^N h^k(s/N)$ and the taper function $h(x) : [0,1] \to \mathbb{R}$ satisfies the following:

Assumption 2.3 *The function h is of bounded variation and satisfies $H := \int_0^1 h^2(x)dx > 0$.*

With this assumption, we obtain, in particular, that $H_2^{(N)} \sim NH$. Note that $h(x) \equiv 1$ yields the non-tapered periodogram. We refer the reader to Dahlhaus (1983) for a discussion of some interesting properties of tapered periodograms. A classical example of a non-trivial taper function is the so-called Hanning window (see e.g. Priestley, 1981, Section 7.4.1). It is well known to reduce leakage effects which occur in spectra with a high dynamic range.

2.2 Motivation for the wavelet-Fisz approach

Ignoring the edge effects, periodogram ordinates $I_N(\omega_k)$, computed at the Fourier frequencies $\omega_k = 2\pi k/N - \pi$, $k = N/2, \ldots, N$, are asymptotically independent and exponentially distributed with means $f(\omega_k)$ (see e.g. Brockwell and Davis, 1987, Section 10.3). With this in mind, we demonstrate the mechanics of our estimation procedure on the following simplified model:

$$J_n(\omega_k) = f(\omega_k)\, e_k, \tag{1}$$

where $\{e_k\}_{k=1}^n$ are iid Exp(1), and the means $f(\omega_k)$ need to be estimated. The model (1) is considered for pedagogical purposes: our rigorous results in Section 3 concern estimation in the full model specified in Section 2.1.

As mentioned in the Introduction, we base our estimation theory on wavelets, without the prior logarithmic transform. Model (1) can be rewritten as $J_n(\omega_k) = f(\omega_k) + \tilde{\varepsilon}_k$, where $\tilde{\varepsilon}_k = f(\omega_k)(e_k - 1)$. Applying the *Discrete Wavelet Transform* (DWT), a multiscale orthonormal linear transform, gives $\tilde{\vartheta}_{j,k} = \vartheta_{j,k} + \tilde{\varepsilon}_{j,k}, \quad j = 0, \ldots, \log_2 n - 1, \quad k = 1, \ldots, 2^j$, and $k = 1$ for $j = -1$, where j and k are (respectively) scale and location parameters, and $\tilde{\vartheta}_{j,k}$, $\vartheta_{j,k}$ and $\tilde{\varepsilon}_{j,k}$ are the wavelet coefficients of $J_n(\omega_k)$, $f(\omega_k)$ and $\tilde{\varepsilon}_k$, respectively. Drawing inspiration from the "universal" threshold theory for Gaussian regression (Donoho and Johnstone (1994)), Neumann (1996) estimates $\vartheta_{j,k}$ by $\hat{\vartheta}_{j,k} = \tilde{\vartheta}_{j,k}\mathbb{I}(|\tilde{\vartheta}_{j,k}| > t_{j,k})$, where the thresholds $t_{j,k}$ are set equal to the *universal* threshold $t_{j,k} = \tilde{\sigma}_{j,k}\sqrt{2\log n}$ with $\tilde{\sigma}_{j,k} = \{\mathrm{var}(\tilde{\varepsilon}_{j,k})\}^{1/2}$. In the simplified model (1), each $\tilde{\sigma}_{j,k}$ is of the form

$$\tilde{\sigma}_{j,k} = \left\{ \sum_{l=1}^{L_j} \psi_{j,l}^2 f^2(\omega_{l+\tau}) \right\}^{1/2}, \tag{2}$$

where τ is a shift parameter dependent on j and k, ψ_j are discrete wavelet vectors with support lengths L_j. Obviously, f^2 is unknown, and Neumann (1996) overcomes this by pre-estimating $f(\omega_k)$ via kernel smoothing, and uses this to obtain estimates of $\tilde{\sigma}_{j,k}$ to compute the universal threshold. Although simple and intutive, this approach generates a number of questions.

Firstly, as mentioned in Section 2.1, the spectral density f may not be smooth as apart from continuity (as a consequence of Assumption 2.2) we only impose the total variation constraint on f. In this case, pre-estimating f via kernel smoothing might not be suitable in practice as "irregularities" in f, such as cusps, sharp peaks or troughs, will be smoothed out. This is then likely to affect the performance of the final estimator $\hat{f}$. Also, the kernel pre-estimator would demand a choice of smoothing parameter, which might not be easy to select optimally (most likely by a local bandwidth selection procedure, and hence costly).

To circumvent this, it may be advantageous to pre-estimate $f(\omega_k)$ in (2) by the inherently local estimate $J_n(\omega_k)$. To our knowledge, this approach has not been considered in literature. With this approach, the estimated $\tilde{\sigma}_{j,k}$ is simply a local weighted l_2 norm of $\{J_n(\omega_k)\}_k$.

Our wavelet-Fisz methodology is inspired by this observation, but instead of the local weighted l_2 norm, it estimates thresholds using a local weighted l_1 norm of $\{J_n(\omega_k)\}_k$. The heuristic reason for this is as follows. As the wavelet vectors ψ_j are compactly supported and well localised, and the function $f(\omega)$ is likely to be "mostly smooth" (possibly with occasional irregularities), the hope is that for most values of j and k, the function $f(\omega)$ can be well approximated by a constant over the support of ψ_j. Denoting the approximating constant by $f_{j,k}$, and using the fact that $\|\psi_j^2\|_2 = 1$, formula (2) implies that $\tilde{\sigma}_{j,k}$ is approximately equal to $f_{j,k}$. Thus, estimating $\tilde{\sigma}_{j,k}$ is approximately equivalent to estimating the constant $f_{j,k}$, which is simply the local mean of an iid exponential sample. Because the maximum likelihood estimator of the mean (= standard deviation) of an iid exponential sample is the sample mean, rather than the sample standard deviation, we propose to estimate thresholds as a local weighted l_1 norm of $\{J_n(\omega_k)\}_k$, rather than the local l_2 norm. Another way to think of this is that since

$f_{j,k}$ ($\approx \tilde{\sigma}_{j,k}$) is the value of a spectral density, it is the most naturally pre-estimated by a local average of the periodogram, which is an l_1-type estimate. Because of the above link to maximum likelihood estimation, the hope is that this provides a 'better' estimator. In fact, this argument can be extended to spectral densities which are Lipschitz-continuous. Indeed, it can be shown that for such spectral densities, estimating (squared) thresholds $t_{j,k}^2$ using our l_1-based estimator asymptotically reduces the Mean-Square Error (in comparison to the l_2-estimator) by 20%.

The added benefit of using the local l_1 norm, as opposed to l_2, is that it permits the construction of *noise-free reconstruction* thresholds, see Section 4. In addition, as with other l_1-based estimators, our threshold estimators are less sensitive to outliers than the l_2 estimators, as well as involving lower moments of the periodogram, which makes then potentially less variable.

The idea of using a local weighted l_1 norm also underlies the *Haar-Fisz* estimation theory, introduced by Fryzlewicz and Nason (2006) and Fryzlewicz et al. (2006) for other multiplicative models, namely for evolutionary wavelet spectrum and locally stationary volatility estimation, respectively. A fundamental novelty of our approach is that we use general wavelets, as opposed to the Haar wavelets used in the latter work. This generalisation is essential as it permits us to include some commonly encountered spectral densities (such as those corresponding to AR processes) in our estimation theory, which would not have been possible had we just used Haar wavelets (we remind the readers that Haar wavelet estimators produce piecewise-constant reconstructions, and since spectral densities corresponding to AR processes are continuous, it is "more optimal" to use continuous wavelets to estimate them). On the other hand, more general wavelets require the use of different proof techniques. We also note that Haar-Fisz estimation was first proposed by Fryzlewicz and Nason (2004), albeit in the non-multiplicative context of Poisson intensity estimation.

3 Wavelet-Fisz spectral density estimation

3.1 Preparing the asymptotic set-up

In order to demonstrate asymptotic mean-square consistency of our proposed estimator, we embed our approach into the appropriate framework for theory, i.e. using orthonormal wavelets defined as continuous square-integrable functions over a unit interval. For what follows we mean our wavelet bases to be $2\pi-$periodic and defined on the $[-\pi, \pi]$ in the (periodised) frequency domain, as both our target function to estimate and our estimators themselves are $2\pi-$periodic. In order to keep our presentation sufficiently simple, we will use a notation with a "classical" wavelet basis; details of how to periodise it can be found in, e.g., Daubechies (1992).

Assumption 3.1 *$\{\phi_{0,k}\}_k \cup \{\psi_{j,k}\}_{j\geq 0;k}$ are chosen to form an orthonormal basis of $L_2[-\pi,\pi]$, where the functions ϕ and ψ satisfy, for any $r > m$ (with $m \geq 1$ given by Theorem 3.1 below),*

(i) ϕ and ψ are in C^r (which implies that they have finite total variation over $[-\pi,\pi]$),
(ii) $\int \phi(x)\,dx = 1$, and (iii) $\int \psi(x)x^l\,dx = 0$ for $0 \leq l \leq r$.

As usual in multiscale wavelet theory, we use the notation $g_{j,k}(x) := 2^{j/2}\ g(2^j\ x - k)$, where $g = \psi, \phi$, so that the scaled and shifted functions $\psi_{j,k}$ and $\phi_{0,k}$ are all normalised to square-integrate to one. The indices j and k are "scale" and "location" parameters, respectively.

To set up the notation for our wavelet threshold spectral estimator, let

$$\widehat{f}(\omega) = \sum_k \widetilde{\alpha}_k\ \phi_{0,k}(\omega) + \sum_{(j,k)\in\partial_N} \rho^{(\cdot)}(\widetilde{\alpha}_{j,k}, \lambda_{j,k})\ \psi_{j,k}(\omega)\ , \tag{3}$$

where $\rho^{(\cdot)}(\alpha,\lambda)$ is the hard or soft threshold rule applied to the coefficient α using threshold λ, and the empirical and true wavelet coefficients are defined as $\widetilde{\alpha}_{j,k} = \int \psi_{j,k}(\omega) I_N(\omega) d\omega$, $\widetilde{\alpha}_k = \int \phi_{0,k} I_N(\omega) d\omega$, $\alpha_{j,k} = \int \psi_{j,k}(\omega) f(\omega) d\omega$ and $\alpha_k = \int \phi_{0,k} f(\omega) d\omega$, respectively.

As in Neumann (1996), we define $\mathcal{J}_N = \{(j,k) \mid 2^j \leq C\ N^{1-\delta}\}$ for some $\delta > 0$; thus, the estimator $\widehat{f}(\omega)$ includes a growing number of coarsest scales j, and excludes a growing number of finest scales. This implies that the number of wavelet coefficients used for thresholding is of order $N^{1-\delta}$, which is less than $O(N)$, and is done to ensure that a certain uniform asymptotic normality effect holds. The parameter δ is chosen so that $\sum_{(j,k)\notin\mathcal{J}_N} \alpha_{j,k}^2 = O(N^{-2m/(2m+1)})$, which is the appropriate rate of convergence to zero of the squared bias due to the encountered approximation error. As will be seen from equation (4) below, the choice of $0 < \delta \leq 1/3$ and an arbitrary $C < \infty$ is sufficient, for our special case of $m \geq 1,\ p \geq 1$ as considered in Theorem 3.1.

3.2 Construction of the wavelet-Fisz thresholds

As thoroughly motivated in Section 2.2, we use estimated thresholds which are "local averages" of the periodogram. More specifically, we propose the choice $\widehat{\lambda}_{j,k} = \widehat{\theta}_{j,k}\ \sqrt{2\log(\#\mathcal{J}_N)}$, where $\widehat{\theta}_{j,k} = c_N \int_{-\pi}^{\pi} \kappa_{j,k}(\omega)\ I_N(\omega) d\omega$. With this choice, our estimator becomes

$$\widetilde{f}(\omega) = \sum_k \widetilde{\alpha}_k\ \phi_{0,k}(\omega) + \sum_{(j,k)\in\mathcal{J}_N} \rho^{(\cdot)}(\widetilde{\alpha}_{j,k}, \widehat{\lambda}_{j,k})\ \psi_{j,k}(\omega)\,.$$

For technical reasons, we take $c_N = \sqrt{2\pi/N\ \ H_4/H_2^2} =: c\,N^{-1/2}$. The $\kappa_{j,k}$ are nonnegative functions on $L_2[-\pi,\pi]$ which are normalised to integrate to one, and hence can be viewed as "local averaging" the periodogram $I_N(\omega)$. Specifically, we set $\kappa_{j,k}(\omega)$ to be the "mod-wavelets", defined as follows: $\kappa_{j,k}(\cdot) := 2^j\ \kappa(2^j \cdot -k)$, where $\kappa(\omega) := |\psi(\omega)| / \int |\psi(\omega)| d\omega$. This choice of $\kappa_{j,k}$ is motivated by the derivation of the "noise-free reconstruction thresholds", see Section 4.

In Section 2.2, emphasis was laid on the fact that, heuristically, the standard deviation of the wavelet coefficients (which thresholds should be proportional to) is "at the level" of the spectral density at the corresponding location, and is thus best estimated via local averaging of the periodogram. Our estimated thresholds $\widehat{\lambda}_{j,k}$ do precisely that: since $\int \kappa_{j,k}(\omega)d\omega = 1$, the parameters $\widehat{\theta}_{j,k}$ can be interpreted as local weighted averages of the periodogram at scale j and location k. In practice, the above integrals are replaced by sums, as in (7) below.

3.3 Near-optimal rate of mean-square convergence

Assume that the spectral density f lies in a ball $\mathcal{F}$ of a Besov space $B_{p,q}^m$ (with $m, p \geq 1$). For details on Besov spaces in the wavelet thresholding context see e.g. Donoho et al. (1995). In this context, it is known that for the wavelet coefficients $\alpha_{j,k}$ of f in any ball $\mathcal{F}$ of $B_{p,q}^m$,

$$\sup_{f\in\mathcal{F}} \sum_{j>J} \sum_{k=0}^{2^j-1} \alpha_{j,k}^2 = O\left(2^{-2J(m+1/2-1/(\min(p,2))}\right), \tag{4}$$

which is of order $O(N^{-\frac{2m}{2m+1}})$ if $2^{-J} = O(N^{-\frac{2}{3}})$ (see Theorem 8 in, again, Donoho et al. (1995)), for the range of parameters $m \geq 1,\ p \geq 1$ considered in Theorem 3.1. This suggests choosing $0 < \delta \leq 1/3$ in the definition of $\mathcal{J}_N = \{(j,k) \mid 2^j \leq C\ N^{1-\delta}\}$. Theoretically, a large δ would be possible in the case of higher Besov regularity ($m > 1$) but, as in practice the parameter m is unknown, we prefer to select a smaller δ to prevent "too much truncation", which, in our experience, hampers practical performance of the method. (We have found $C = 1$ and $\delta = 0.1$ to be a good practical choice.) We now state our main theorem.

Theorem 3.1 *Suppose that Assumptions 2.1, 2.2, 2.3 and 3.1 hold. Let $B_{p,q}^m(C)$ be a Besov ball of radius $C < \infty$ with $m, p \geq 1$ (recall that by Assumption 3.1(iii), we have, in particular,*

$\int \psi(u)u^l du = 0$ *for* $0 \le l \le m$*). We have*

$$\sup_{f \in B^m_{p,q}(C)} E \int_{-\pi}^{\pi} |\widetilde{f}(\omega) - f(\omega)|^2 d\omega = O\left((\log N/N)^{2m/(2m+1)}\right). \quad (5)$$

The proof, additional theoretical results and implementation details, can be found in the accompanying technical report, available as Technical Report 08:05, 2008, The University of Bristol Statistics Group Research Reports series at URL `http://www.stats.bris.ac.uk` and also via

`http://www.maths.bris.ac.uk/~mapzf/wfspec/wfspec.pdf`

The methodology of the proof is not dissimilar to that used by Neumann (1996).

Theorem 3.1 shows that our estimator $\widetilde{f}$ enjoys the same optimality properties as the analogous estimators in the "function + i.i.d. Gaussian noise" setting. We mention that it is well-known that linear estimators cannot achieve the optimal rate of MSE convergence (not even up to a logarithmic term) when $p < 2$.

4 Noise-free reconstruction thresholds

Although our wavelet-Fisz estimator $\widetilde{f}$ enjoys good theoretical properties (by Theorem 3.1) it often oversmooths in practice. This is not surprising as the thresholds $\widehat{\lambda}_{j,k}$ contain the same logarithmic term as the universal thresholds in Gaussian regression which also tend to oversmooth. Below we propose an alternative wavelet-Fisz thresholding estimator which performs better in practice. The $\widetilde{f}$ estimator essentially consists in "testing" whether each empirical wavelet coefficient $\int \psi_{j,k}(\omega) I_N(\omega)$ exceeds the quantity $t_{j,N} \int |\psi_{j,k}(\omega)| I_N(\omega)$ for a given choice of $t_{j,N}$. Our new estimator follows the same principle, but constructs different $t_{j,N}$ as follows.

Our new estimator is constructed to possess the following *noise-free reconstruction property*: if the truth $f(\omega)$ is a constant function of ω, then, with high probability, our estimate is also constant and equal to the empirical mean of $\{I_N(\omega_k)\}_{k=1}^N$. The noise-free reconstruction property guarantees that asymptotically, no noise survives the estimation procedure and thus the resulting estimate is visually appealing and does not display spurious spikes.

For the noise-free reconstruction property to hold, we require that, for a constant spectral density f, all empirical wavelet coefficients fall below their corresponding thresholds with a high probability. I.e. we require that

$$P\left(\bigcup_{j,k}\left\{\int \psi_{j,k}(\omega) I_N(\omega) > \tilde{t}_{j,N} \int |\psi_{j,k}(\omega)| I_N(\omega)\right\}\right) \to 0 \quad \text{as} \quad N \to \infty. \quad (6)$$

Below, we describe a set of approximations to (6), which facilitate the computation. Define $\xi_{j,l,k} = \psi_{j,l} I_N(2\pi(l/N + k/2^j) - \pi)$ and $\Xi_{j,k} = \sum_{l=1}^{L_j} \xi_{j,l,k} / \sum_{l=1}^{L_j} |\xi_{j,l,k}|$. We first note that in practice, the integrals in (6) are replaced by sums, which gives the condition

$$P\left(\cup_{j,k}\left\{\Xi_{j,k} > \tilde{t}_{j,N}\right\}\right) \to 0 \quad \text{as} \quad N \to \infty, \quad (7)$$

where, as in Section 2.2, ψ_j are discrete wavelet vectors and L_j are their support lengths. By the Bonferroni inequality (7) is implied by

$$\sum_{j,k} P\left(\Xi_{j,k} > \tilde{t}_{j,N}\right) \to 0 \quad \text{as} \quad N \to \infty. \quad (8)$$

Denote $Y^+_{j,k} := \sum_{l:\psi_{j,l}>0} \xi_{j,l}$ and $Y^-_{j,k} := -\sum_{l:\psi_{j,l}<0} \xi_{j,l}$. Note that the distribution of each $Y^+_{j,k}$ (similarly, each $Y^-_{j,k}$) is asymptotically the same, since $I_N(\omega_k)$ is asymptotically a sequence of independent exp. variables centred at f. We rewrite (8) as

$$\sum_{j,k} P\left(Y^+_{j,k}/Y^-_{j,k} > (1+\tilde{t}_{j,N})/(1-\tilde{t}_{j,N})\right) \to 0 \quad \text{as} \quad N \to \infty. \tag{9}$$

By the Satterthwaite approximation (see, e.g. Johnson and Kotz (1975)), $Y^+_{j,k}$ and $Y^-_{j,k}$ are approximately independently distributed as $\beta_{1j}\chi^2_{\nu_{1j}}$ and $\beta_{2j}\chi^2_{\nu_{2j}}$, for appropriate values of the constants $\beta_{\cdot j}$ and $\nu_{\cdot j}$. Thus (9) can be approximated as

$$\sum_j 2^j \left(1 - F_{\nu_{1j},\nu_{2j}}\left(\frac{\beta_{2j}(1+\tilde{t}_{j,N})}{\beta_{1j}(1-\tilde{t}_{j,N})}\right)\right) \to 0 \quad \text{as} \quad N \to \infty, \tag{10}$$

where $F_{a,b}(\cdot)$ is the cdf of the F distribution with a, b degrees of freedom. To derive $\tilde{t}_{j,N}$, we now mimick standard Gaussian universal thresholding, where the speed of convergence of the quantity analogous to (10) is $\frac{1}{2}\{\pi \log_2(N/2)\}^{-1/2}$, and the analogues of the probabilities

$$\alpha_j(N) := F_{\nu_{1j},\nu_{2j}}\left(\frac{\beta_{2j}(1+\tilde{t}_{j,N})}{\beta_{1j}(1-\tilde{t}_{j,N})}\right) \tag{11}$$

are constant across scales; that is $\alpha_j(N) = \alpha(N)$. To find $\tilde{t}_{j,N}$, we first solve $\sum_j 2^j(1-\alpha(N)) = \frac{1}{2}\{\pi \log_2(N/2)\}^{-1/2}$ for $\alpha(N)$, and then obtain each $\tilde{t}_{j,N}$ numerically from (11).

This method requires computation of the 'mod wavelet' coefficients of I_N. We compute these coefficient (and a TI-version of them) using a fast $\mathcal{O}(N \log N)$ algorithm building on the approximative ideas for powers of wavelets in Barber et al. (2002), see technical report.

We mention that the idea of noise-free reconstruction in the context of volatility estimation via Haar wavelet thresholding was considered in Section 3.2 of Fryzlewicz et al. (2006).

5 Simulations

In this section, we compare the empirical performance of our wavelet-Fisz estimator to both the unshifted and 32-shift cycle shifted method of Neumann (1996), to the method of Gao (1993), as well as to the kernel smoothing of $I_N(\omega)$ with optimally chosen global, and locally varying, bandwidth. (The 32-shift cycle-spun estimate from Neumann (1996) is equivalent to full cycle-spinning since the primary resolution $l = 3$ as noted there.) For the latter two methods, we have used the routines `glkerns` and `lokerns` from the R package `lokern`. Further implementational issues are discussed more fully in our associated technical report.

We apply these methods to two "test processes". Process A is from Neumann (1996) and defined by $X^A_t = Y_t + \frac{1}{2}Z_t$, where $Y_t + \frac{1}{5}Y_{t-1} + \frac{9}{10}Y_{t-2} = \varepsilon_t + \varepsilon_{t-2}$, and $\{\varepsilon_t\}$, $\{Z_t\}$ are iid $N(0,1)$ variables. Process B is the MA(2) process $X^B_t = \varepsilon_t + \varepsilon_{t-1} - \varepsilon_{t-2}$. For both $N = 1024$. Neumann's process A has a sharp spectral peak whereas process B's spectrum is much smoother.

Gao (1993) estimates the *log*-spectral density. In this simulation study, we correct the bias of his estimator using the Euler-gamma constant, and then exponentiate it to obtain the spectral density estimate. Both Gao's and our wavelet-Fisz estimator use translation-invariant (TI) hard thresholding with all resolution levels thresholded. We use the "noise-free" version of our estimator. With both Neumann's estimators we set the pilot estimator to be the true spectrum and so that results reported are somewhat better than those that could be obtained in practice.

Tapering often introduces substantial correlation in the periodogram. Empirically, we found that the wavelet-Fisz and Gao's estimators' performance is not much affected by tapering. However, the `glkerns` and `lokerns` performance deteriorates notably when tapering is introduced.

This is because the optimal bandwidth choice for these methods is based on cross-validation which often does not work well for correlated data. Thus, we only report the untapered results.

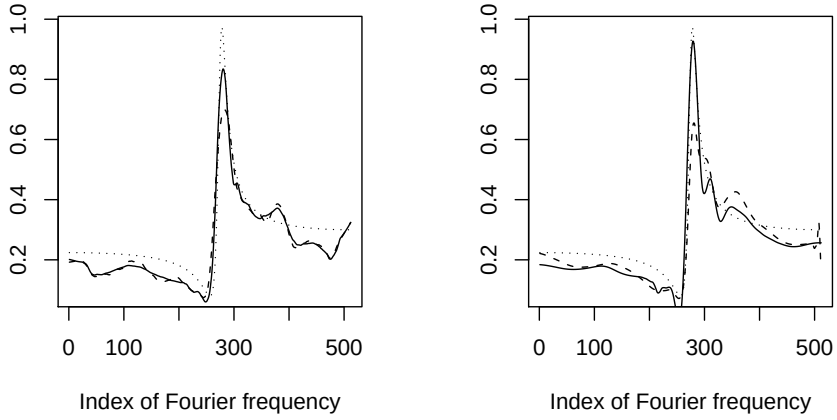

Figure 1: Both: True spectral density for process A (dotted). Left: `glkerns` (dashed) and `lokerns` (solid) estimate. Right: Gao's method (dashed) and wavelet-Fisz (solid).

Figure 1 shows the true spectral density for process A and sample reconstructions using the four methods tested. Gao's method and wavelet-Fisz use Daubechies' "Least Asymmetric 5" wavelet. The `glkerns` method oversmooths the peak, which is not surprising because it is a linear method which is not capable of estimating both smooth and inhomogeneous regions well. The method of Gao also oversmooths the peak, which is due to the magnification of the slight oversmooth of the peak of the log-spectral density after exponentiating the estimate.

The `lokerns` and wavelet-Fisz methods perform the best. Wavelet-Fisz reconstructs the peak better than `lokerns` and it also flatter in the tails. On the other hand, `lokerns` does a better job at reconstructing the region immediately to the right of the peak. The ISE of wavelet-Fisz is around 12% lower than that of `lokerns` for this sample. It is also 39% lower than the ISE of Gao's method and 41% lower than that of `glkerns`.

We have also computed the ISE averaged over 100 sampled paths for all 6 methods (and a selection of wavelet bases for Gao's, Neumann's and wavelet-Fisz) for processes A and B. Table 1 shows percentage improvements in ISE of the wavelet-Fisz method over its competitors for process A. Wavelet-Fisz is always the best option for this 'peaked' spectral density, sometimes substantially so. Note that, in practice, Neumann's estimator would be worse since here we set the pilot estimate equal to the actual spectrum (although with an estimated pilot estimator the results are not dramatically worse).

A similar table of results was produced for process B in Table 2. For the smooth spectrum the wavelet-Fisz method outperforms Gao's by about 8%, and Neumann's methods by 10–25%. The wavelet-Fisz method is not more than 6% worse than the `glkerns` and `lokerns` methods and, for Haar wavelets, is nearly 10% better.

Overall, the performance of our wavelet-Fisz estimator is extremely good on the 'peaky' spectrum. For a smooth spectrum our estimator is better than Gao's and Neumann's and comparable to `glkerns` and `lokerns`. To choose the 'best' wavelet to use, one might consider using a cross-validatory approach for wavelet selection such as that proposed by Nason (2002).

Table 1: *Peaked process A spectrum simulation.* % by which the wavelet-Fisz ISE (mean over 100 sample paths) is less than that of the five competitors. (Neumann CS=32 shift estimate).

	Spectral estimation method				
Wavelet	`glkerns`	`lokerns`	Gao	Neumann	Neumann CS
Haar	26	20	18	40	7
DaubExPhase 7	8	0.4	28	48	43
DaubLeAsymm 5	22	15	21	56	50

Table 2: *Smooth process B spectrum simulation.* % by which the wavelet-Fisz ISE (mean over 100 sample paths) is less than that of the five competitors. (Neumann CS=32 shift estimate).

	Spectral estimation method				
Wavelet	`glkerns`	`lokerns`	Gao	Neumann	Neumann CS
Haar	9	9	9	13	11
DaubExPhase 7	-4	-4	9	25	23
DaubLeAsymm 5	-6	-6	7	23	21

Acknowledgement

Supported by EPSRC Small Grant GR/T22278/01 for a visit by RvS to Bristol in 2004. PF thanks RvS for financial support & hospitality for visits to Louvain-la-Neuve & Kaiserslautern and is also grateful to the Maths Dept., Imperial College London, where a large part of this work was done, for an excellent research environment. GPN was supported by EPSRC ARF Grant GR/A01664/01 and by Responsive Mode Grant EP/D005221/1. RvS was supported by IAP research network grant P5/24 of the Belgian Government (Belgian Science Policy).

References

S. Barber, G.P. Nason, and B.W. Silverman. Posterior probability intervals for wavelet thresholding. *J. Roy. Stat. Soc. B*, 64:189–205, 2002.

D.R. Brillinger. *Time Series: Data Analysis and Theory.* McGraw-Hill Inc., New York, 1981.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer, 1987.

F. Comte. Adaptive estimation of the spectrum of a stationary Gaussian sequence. *Bernoulli*, 7:267–298, 2001.

A. C. Cristan and A. T. Walden. Multitaper power spectrum estimation and thresholding: Wavelet packets versus wavelets. *IEEE Trans. Sig. Proc.*, 50:2976–2986, 2002.

R. Dahlhaus. Spectral analysis with tapered data. *J. Time Ser. Anal.*, 4:163–175, 1983.

I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, Pa., 1992.

P.L. Davies and A. Kovac. Densities, spectral densities and modality. *Ann. Stat.*, 32:1093–1136, 2004.

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

D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? (with discussion). *J. Roy. Stat. Soc. B*, 57:301–369, 1995.

P. Fryzlewicz and G.P. Nason. A Haar-Fisz algorithm for Poisson intensity estimation. *J. Comput. Graph. Stat.*, 13: 621–638, 2004.

P. Fryzlewicz and G.P. Nason. Haar-Fisz estimation of evolutionary wavelet spectra. *J. Roy. Stat. Soc. B*, 68:611–634, 2006.

P. Fryzlewicz, T. Sapatinas, and S. Subba Rao. A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika*, 93:687–704, 2006.

H-Y. Gao. *Wavelet Estimation of Spectral Densities in Time Series Analysis*. PhD thesis, University of California Berkeley, 1993.

H-Y. Gao. Choice of thresholds for wavelet shrinkage estimate of the spectrum. *J. Time Ser. Anal.*, 18:231–251, 1997.

N. Johnson and S. Kotz. *Distributions in Statistics: Continuous Univariate Distributions*. Wiley, 1975.

L.H. Koopmans. *The Spectral Analysis of Time Series*. Academic Press, San Diego, 1995.

P. Moulin. Wavelet thresholding techniques for power spectrum estimation. *IEEE Trans. Sig. Proc.*, 42:3126–3136, 1994.

G. P. Nason. Choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage. *Stat. Comput.*, 12: 219–227, 2002.

M. H. Neumann. Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *Statistics Research Report No. SRR 028–94, CMA, Australian National University, Canberra, Australia*, 1994.

M. H. Neumann. Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *J. Time Ser. Anal.*, 17:601–633, 1996.

M. Pensky and B. Vidakovic. Bayesian decision theoretic scale-adaptive estimation of a log-spectral density. *Preprint*, 2003.

M. B. Priestley. *Spectral Analysis and Time Series*. Academic Press, 1981.

R.H. Shumway and D.S. Stoffer. *Time Series Analysis and Its Applications*. Springer-Verlag, New York, 2000.