University of Bristol

Statistics Group

Technical Report 08-05

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

Abstract

We suggest a new approach to wavelet threshold estimation of spectral densities of stationary time series. It is well known that choosing appropriate thresholds to smooth the periodogram is difficult because non-parametric spectral estimation suffers from problems similar to curve estimation with a highly heteroscedastic and non-Gaussian error structure. Possible solutions that have been proposed are plug-in estimation of the variance of the empirical wavelet coefficients or the log-transformation of the periodogram.

In this paper we propose an alternative method to address the problem of heteroscedasticity and non-normality. We estimate thresholds for the empirical wavelet coefficients of the (tapered) periodogram as appropriate linear combinations of the periodogram values similar to empirical scaling coefficients. Our solution permits the design of "asymptotically noise-free reconstruction thresholds", paralleling classical wavelet theory for nonparametric regression with Gaussian white noise errors. Our simulation studies show promising results that clearly improve the classical approaches mentioned above. 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 very 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) provides a comprehensive introduction to the spectral analysis of time series. Typically, inference in the spectral domain is based on the periodogram of the data x_t. Often, a data taper is applied prior to computing the periodogram, in order to reduce leakage (Dahlhaus (1983)). It is well known that the (tapered) periodogram is an inconsistent estimator of the spectral density and needs to be smoothed to achieve consistency.

Depending on the theoretical properties of the underlying stationary stochastic process X_t and the associated spectral density $f(\omega)$, various periodogram smoothing techniques have been proposed. For spectral densities with a high degree of regularity, linear smoothing techniques

*Author for correspondence.

(e.g. kernel smoothing) are appropriate. They are covered extensively in the literature: we refer the reader, for example, to the monographs of Brillinger (1981), Koopmans (1995) and Shumway and Stoffer (2000). The application of multiple tapers leads to, both leakage-reduced and smoothed, multitaper spectrum estimators, see e.g. Cristan and Walden (2002) and the references therein.

However, it is well known that linear smoothing methods are incapable of achieving the optimal mean-square rate of convergence in cases where the smoothness of the underlying regression function is distributed in an inhomogeneous fashion over the domain of interest. Thus, in the case of irregular spectral densities, smoothing the periodogram using a nonlinear method might be more suitable. An interesting example of a highly irregular spectral density (which satisfies all our technical assumptions specified later in the paper) is given e.g. in Neumann (1994), Section 8, item 2. Nonlinear methods will also have benefits for spectra with multiple (sharp) peaks and troughs. We note that 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 for many theoreticians and practitioners if the regression function is of low regularity. Thus, the idea of smoothing the periodogram using a nonlinear wavelet method might seem appealing. However, the periodogram approximately follows a multiplicative regression set-up where the variance of the "noise" is not constant over frequencies but is proportional to the level of the underlying spectral density. This represents a hurdle for nonlinear wavelet thresholding, where the variance of the noise needs to be either known or easily estimable. To tackle this problem, two main approaches have been proposed in the literature.

The first approach consists in taking the logarithmic transform of the periodogram to stabilize the variance and transform the model from multiplicative to additive, and only then proceeding with the wavelet smoothing. This idea was first proposed by Gao (1993, 1997). Moulin (1994) derived wavelet thresholds for the logged periodogram using saddle point estimation techniques. Pensky and Vidakovic (2003) derived thresholds for the log-periodogram using the Bayesian paradigm but also demonstrated their frequentist mean-square properties. 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 smoothing technique for the periodogram, where thresholds for the empirical wavelet coefficients are constructed as appropriate local weighted l_1 norms of the periodogram, as opposed to the l_2 norm used in Neumann (1996). As explained in Section 2, 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 the 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.

The paper is organised as follows. In the next section we recall the set-up of nonparametric estimation of spectral densities and give a non-technical motivation for our new approach. Section 3 contains our main theoretical achievements where we show near-optimal rates of convergence of the mean-square risk of our new spectral estimator over a class of spectral densities which also includes those of low regularity. The following section addresses the construction of so-called "noise-free reconstruction thresholds" which are designed to work better in non-asymptotic settings. In a simulation section we compare our new approach with some of the established estimation methods mentioned above. Proofs, and additional theoretical results that complete them, are in the Appendix.

2 Set-up and motivation

In this introductory section, we both establish the technical assumptions for our set-up and give, by a simplified presentation of the spectral estimation problem, the essential motivation for our new approach. It is only in Section 3 that we turn to a more formal asymptotic treatment of

wavelet estimation of spectral densities.

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

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

Throughout the paper, we restrict our interest to processes whose spectral densities satisfy the following assumption.

Assumption 2.1 *The spectral density $f(\omega)$ satisfies*

(i) $f(\omega) \geq \mu > 0$,

(ii) f *is of finite total variation over* $[-\pi, \pi]$,

(iii) f *is continuous.*

Assumption 2.1(i) is essential in obtaining strong asymptotic normality of the empirical wavelet coefficients of the periodogram of X_t. Technically-minded readers may wish to refer to formula (20) at this point. 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*

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

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 asymptotic normality of the (appropriately scaled) local cumulative sums of X_t. The supremum on the left-hand side guarantees that the asymptotic normality is, in a sense, uniform over time 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 will be based on the periodogram of the (possibly tapered) observations

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

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.

Assumption 2.3 *The taper 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.

It is well known that the (tapered) periodogram is an inconsistent estimator of the spectral density and thus needs to be smoothed to achieve consistency. The next section describes our wavelet-Fisz method for smoothing $I_N(\omega)$.

2.2 Spectrum estimation via wavelets: motivation and existing approaches

We base our motivation on the following well known fact of spectral estimation theory (see e.g. Brockwell and Davis, 1987, Section 10.3). 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 $f(\omega_k)$ (except, in most cases, the "edge" frequencies 0 and π, but we shall ignore this fact for the time being).

Motivated by this observation, we choose to demonstrate the basic mechanics of our estimation procedure on the following simplified model:

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

where $\{e_k\}_{k=1}^n$ is a sequence of iid variables distributed as Exp(1). In the model (1) we are now

faced with the problem of estimating the means $f(\omega_k)$ of $J_n(\omega_k)$. The model (1) is considered merely for pedagogical purposes: our rigorous results in Section 3 concern estimation in the full model specified in Section 2.1. The quantity modelled in (1) is labelled as J_n to avoid confusion with I_N.

As mentioned in the Introduction, we base our estimation theory on wavelets. Since the seminal work of Donoho and Johnstone (1994), nonlinear estimation techniques based on wavelets have become a popular and extensively studied tool for non-parametric regression. Many of them combine excellent finite-sample performance, linear computational complexity, and optimal (or near-optimal) asymptotic mean-square error behaviour over a variety of function smoothness classes. A general overview of wavelet methods in statistics can be found, for example, in Vidakovic (1999).

A convenient starting point for wavelet estimation is the formulation of the regression problem at hand in a "function + noise" setting, where the noise has mean zero and its variance is either known or can easily be estimated. Note that the logarithmic transformation transforms the model (1) from multiplicative to additive:

$$\log J_n(\omega_k) = \log f(\omega_k) + \mathbb{E} \log e_k + \varepsilon_k, \tag{2}$$

where $\varepsilon_k = \log e_k - \mathbb{E} \log e_k$ has mean zero and a variance independent of k. Thus, many authors (some references are given in the Introduction) considered wavelet estimation of the log-spectral density in the logged model (2). However, one drawback of using the log transformation is that it flattens out the data, often obscuring interesting features, e.g. spectral peaks which indicate hidden periodicities of the process. Also, the mean-square properties of the resulting exponentiated estimator of the spectral density f are not easy to establish.

To avoid these problems, it might be beneficial to work with the model (1) directly, without the prior logarithmic transform. The 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. For a large class of functions f, the sequence $\vartheta_{j,k}$ is *sparse*, with most $\vartheta_{j,k}$'s being equal, or close, to zero, which motivates the use of simple thresholding estimators $\hat{\vartheta}_{j,k}$ which estimate $\vartheta_{j,k}$ by zero if and only if the corresponding empirical wavelet coefficient $\tilde{\vartheta}_{j,k}$ falls below certain threshold in absolute value. This ensures that a large proportion of the noise $\tilde{\varepsilon}_{j,k}$ gets removed. The inverse DWT of the thresholded coefficients $\hat{\vartheta}_{j,k}$ then yields an estimate $\hat{f}$ of the original function f.

Drawing inspiration from the "universal" threshold theory first developed by Donoho and Johnstone (1994) in the Gaussian regression case, 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

$$t_{j,k} = \tilde{\sigma}_{j,k}\sqrt{2 \log n} \tag{3}$$

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

$$\tilde{\sigma}_{j,k} = \left\{ \sum_{l=1}^{L_j} \psi_{j,l}^2 f^2(\omega_{l+\tau}) \right\}^{1/2}, \tag{4}$$

where τ is a shift parameter dependent on j and k, ψ_j are discrete wavelet vectors (as described, for example, in Nason et al. (2000)), and L_j are their support lengths. Obviously, f^2 is unknown, and Neumann (1996) overcomes this by pre-estimating $f(\omega_k)$ via kernel smoothing, and using the pre-estimate to obtain estimates of $\tilde{\sigma}_{j,k}$ which are then used in (3). 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 (discontinuous derivatives), sharp peaks or troughs, will be smoothed out. This is then likely to affect the performance of the final estimator $\hat{f}$. Also, as

with any other non-parametric plug-in estimator, the kernel 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).

2.3 Motivation for our wavelet-Fisz approach

To circumvent the drawbacks of the plug-in procedure, it may be advantageous to pre-estimate $f(\omega_k)$ in (4) by the inherently local estimate $J_n(\omega_k)$, for most scales j coarser than the observation scale. To the best of 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 (4) 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, as detailed in Section 4. In addition, as with many 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 wavelet spectrum and locally stationary volatility estimation, respectively. The 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 the remainder, 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, in particular, that they have finite total variation over $[-\pi, \pi]$,

(ii) $\int \phi(x)\,dx \;=\; 1,$

(iii) $\int \psi(x) x^l\,dx \;=\; 0 \quad$ *for* $\quad 0 \le l \le r.$

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

$$\widehat{f}(\omega) \;=\; \sum_k \widetilde{\alpha}_k\; \phi_{0,k}(\omega) \;+\; \sum_{(j,k)\in \mathcal{J}_N} \rho^{(.)}(\widetilde{\alpha}_{j,k}, \lambda_{j,k})\; \psi_{j,k}(\omega)\;, \tag{5}$$

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

As in Neumann (1996), we define $\mathcal{J}_N = \{(j,k) \mid\; 2^j \le C\; N^{1-\delta}\}$ 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, see formula (20) in the Appendix. 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 actually the appropriate rate of convergence to zero of the squared bias due to the encountered approximation error. As will be seen from equation (9) below, the choice of $0 < \delta \le 1/3$ and an arbitrary $C < \infty$ is sufficient, for our special case of $m \ge 1,\; p \ge 1$ as considered in Theorem 3.1.

3.2 Construction of the wavelet-Fisz thresholds

As thoroughly motivated in Section 2.3, we use estimated thresholds which are "local averages" of the periodogram. More specifically, we propose the choice

$$\widehat{\lambda}_{j,k} = \widehat{\theta}_{j,k}\; \sqrt{2 \log(\#\mathcal{J}_N)}\;, \tag{6}$$

where

$$\widehat{\theta}_{j,k} = c_N \int_{-\pi}^{\pi} \kappa_{j,k}(\omega)\ I_N(\omega) d\omega. \tag{7}$$

With this choice, our estimator becomes

$$\widetilde{f}(\omega) \ = \ \sum_k \widetilde{\alpha}_k\ \phi_{0,k}(\omega) \ + \ \sum_{(j,k)\in\mathcal{J}_N} \rho^{(\cdot)}(\widetilde{\alpha}_{j,k}, \widehat{\lambda}_{j,k})\ \psi_{j,k}(\omega)\,. \tag{8}$$

For technical reasons outlined in Section A.1, 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 therefore can be viewed as providing "local averaging" of the periodogram $I_N(\omega)$. More 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 2.3, 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. We note that 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. Note that in practice, the above integrals are replaced by sums, as in formula (12) below.

3.3 Near-optimal rate of mean-square convergence

The near-optimal mean-square convergence rate of our coordinatewise thresholded wavelet spectral estimator $\widetilde{f}$ is formulated in Theorem 3.1 below. As the rate of convergence will be faster the higher the regularity of the target function to estimate, we assume that the spectral density f lies in a ball $\mathcal{F}$ of a quite general function space: a Besov space $B^m_{p,q}$ (with $m, p \geq 1$). This means a slightly more general set-up than a Sobolev regularity for f, in that f is assumed to have m generalised derivatives in L_p, with the parameter q allowing for additional spatial inhomogeneity. For these function balls, the optimal rate of mean-square convergence, the so-called "minimax rate", is $N^{-\frac{2m}{2m+1}}$. More details on Besov spaces in the context of wavelet thresholding can be found, e.g., in 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 $B^m_{p,q}$,

$$\sup_{f\in\mathcal{F}} \sum_{j>J} \sum_{k=0}^{2^j-1} \alpha_{j,k}^2 \;=\; O\left(2^{-2J(m+1/2-1/(\min(p,2))}\right), \tag{9}$$

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 now state our main theorem.

Theorem 3.1 *Suppose that Assumptions 2.1, 2.2, 2.3 and 3.1 hold. Let $B^m_{p,q}(C)$ 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 $0 \leq l \leq m$). We have*

$$\sup_{f\in B^m_{p,q}(C)} E \int_{-\pi}^{\pi} |\widetilde{f}(\omega) - f(\omega)|^2 d\omega \;=\; O\left((\log\ N/N)^{2m/(2m+1)}\right). \tag{10}$$

Theorem 3.1 shows that our data-driven wavelet threshold estimator $\widetilde{f}$ achieves near- minimaxity in the mean-square sense over a large class of function spaces, and hence 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 as stated in Theorem 3.1, in practice it often oversmooths. This is not surprising as the thresholds $\widehat{\lambda}_{j,k}$ contain the same logarithmic term as the universal thresholds in Gaussian regression, and the latter tend to oversmooth. The aim of this section is to propose an alternative wavelet-Fisz thresholding estimator of f which performs better in practice.

The new estimator is constructed to possess the following *noise-free reconstruction property*:

if the true function $f(\omega)$ is a constant function of ω, then, with high probability, our estimate of f 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.

To set the scene, we recall that our estimation procedure described in Section 3 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 particular choice of $t_{j,N}$. Our new noise-free estimator follows the same principle, but uses a different set of $t_{j,N}$'s, which we construct as follows.

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. In other words, we require that

$$P\left(\bigcup_{j,k}\left\{\int \psi_{j,k}(\omega) I_N(\omega) > \tilde{t}_{j,N} \int |\psi_{j,k}(\omega)| I_N(\omega)\right\}\right) \to 0 \quad \text{as} \quad N \to \infty. \tag{11}$$

Deriving $\tilde{t}_{j,N}$ from (11) in an exact manner is possible, although computationally inefficient. Below, we describe a set of approximations to (11), which facilitate the computation, although obviously yield a slightly different set of $\tilde{t}_{j,N}$. Simulations described in Section 5.2 demonstrate good practical performance of the approximate noise-free thresholding estimator.

We first note that in practice, the integrals in (11) are replaced by sums, which gives the condition

$$P\left(\bigcup_{j,k}\left[\frac{\sum_{l=1}^{L_j} \psi_{j,l}\ I_N\{2\pi(l/N + k/2^j) - \pi\}}{\sum_{l=1}^{L_j} |\psi_{j,l}|\ I_N\{2\pi(l/N + k/2^j) - \pi\}} > \tilde{t}_{j,N}\right]\right) \to 0 \quad \text{as} \quad N \to \infty, \tag{12}$$

where, as in Section 2.2, ψ_j are discrete wavelet vectors and L_j are their support lengths. By the Bonferroni inequality, (12) is implied by

$$\sum_{j,k} P\left(\frac{\sum_{l=1}^{L_j} \psi_{j,l}\ I_N\{2\pi(l/N + k/2^j) - \pi\}}{\sum_{l=1}^{L_j} |\psi_{j,l}|\ I_N\{2\pi(l/N + k/2^j) - \pi\}} > \tilde{t}_{j,N}\right) \to 0 \quad \text{as} \quad N \to \infty. \tag{13}$$

Denote $Y_{j,k}^+ := \sum_{l:\psi_{j,l}>0} \psi_{j,l}\ I_N(2\pi(l/N + k/2^j) - \pi)$ and $Y_{j,k}^- := -\sum_{l:\psi_{j,l}<0} \psi_{j,l}\ I_N(2\pi(l/N + k/2^j) - \pi)$. 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 exponential variables centred at f. We rewrite (13) as

$$\sum_{j,k} P\left(\frac{Y^+_{j,k}}{Y^-_{j,k}} > \frac{1+\tilde{t}_{j,N}}{1-\tilde{t}_{j,N}}\right) \to 0 \quad \text{as} \quad N \to \infty. \tag{14}$$

Note that by 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 (14) can be approximated as

$$\sum_j 2^j \left(1 - F_{\nu_{1j},\nu_{2j}}\left(\frac{\beta_{2j}(1+\tilde{t}_{j,N})}{\beta_{1j}(1-\tilde{t}_{j,N})}\right)\right) \to 0 \quad \text{as} \quad N \to \infty, \tag{15}$$

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 (15) is $\frac{1}{2}\{\pi \log_2(N/2)\}^{-1/2}$, and the analogues of the probabilities

$$\alpha_j(N) := F_{\nu_{1j},\nu_{2j}}\left(\frac{\beta_{2j}(1+\tilde{t}_{j,N})}{\beta_{1j}(1-\tilde{t}_{j,N})}\right) \tag{16}$$

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 (16).

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 Implementation and simulations

5.1 Implementation

We now outline some implementational details of our wavelet-Fisz spectral density estimator. The algorithm for computing the estimator proceeds as follows.

1. Given the data $\{X_t\}_{t=1}^N$, compute the (tapered) periodogram $I_N(\omega_k)$ at the Fourier fre-

quencies $\omega_k = 2\pi k/N - \pi$, $k = N/2, \ldots, N$. Note that we do not need to compute $I_N(\omega_k)$ at the negative Fourier frequencies as $I_N(\omega) = I_N(-\omega)$.

2. Given a wavelet basis ψ, compute the standard discrete decimated (or non-decimated) wavelet transform of $I_N(\omega_k)$ using a fast algorithm described by Mallat (1989) (or Coifman and Donoho (1995)). Basing the estimator on the non-decimated wavelet transform corresponds to averaging the estimator based on the decimated transform over all possible cyclic shifts of the data, and is well known to reduce the mean-square error and reduce the Gibbs effect, which is why we use the non-decimated (aka: translation-invariant) version of our estimator in the simulations reported below.

3. Compute the thresholds $\widehat{\lambda}_{j,k}$ from Section 3 or $\tilde{t}_{j,N}$ from Section 4. To construct $\widehat{\lambda}_{j,k}$, we need to choose the set $\mathcal{J}_N$. We have empirically found that the choice $C = 1$, $\delta = 0.01$ (as a default "small" value) yields good practical performance of the estimator. Both $\widehat{\lambda}_{j,k}$ and $\tilde{t}_{j,k}$ require the computation of the equivalent of the discrete wavelet transform but computed using the mod-wavelets $\kappa_{j,k}$ instead of the wavelets $\psi_{j,k}$. The latter part of this Section explains how it is done.

4. Threshold the empirical wavelet coefficients of $I_N(\omega_k)$ via hard or soft thresholding, for $(j,k) \in \mathcal{J}_N$ if $\widehat{\lambda}_{j,k}$ are used, or for all j,k if $\tilde{t}_{j,N}$ are used.

5. Invert the wavelet transform to get an estimate of f at the Fourier frequencies.

Computing the mod-wavelet transform. An easy way to implement (7) would be to compute the discrete mod-wavelets, κ_j, one for each scale j by computing ψ_j using existing software and then taking absolute values. One can then filter $I_N(\omega)$ one scale at a time. For the decimated algorithm, the computational complexity would be $O(N \log_2 N)$; for the non-decimated algorithm, it would be $O(N^2 \log_2 N)$.

One of the key advantages of wavelet methods is their efficiency. For decimated transforms we usually 'expect' $O(N)$ algorithms and for non-decimated we would like $O(N \log_2 N)$. Our mod-wavelet transform can achieve these complexities if we adopt an approximation approach similar to the one proposed by Herrick (2000) and subsequently used in Barber et al. (2002) although here it is based on adapting the forward wavelet transform, not the inverse. The idea

is to represent $\kappa_{j,k}$ in terms of scaling functions at a finer scale as follows:

$$\kappa_{j,0}(\omega) \approx \sum_{\ell} e_{j+m_0,\ell}\phi_{j+m_0,\ell}(\omega), \tag{17}$$

where $j+m_0$ is some fixed number of scales finer than j which controls the level of approximation ($m_0 = 3$ is typically very good but higher values are even better). The $e_{j+m_0,\ell}$ sequence is a finite support filter of length $2^{m_0}(b_\psi - a_\psi) + (b_\phi - a_\phi)$ where (a_ψ, b_ψ) is the finite support interval for the wavelet ψ and similarly for ϕ. The $e_{j,\ell}$ sequence can easily be precomputed by the appropriate inner product of κ and ϕ.

Then the mod-wavelet transform of some function $f(\omega)$ can, after some algebra, be approximated by

$$\int f(\omega)\kappa_{j,k}(\omega)\, d\omega \approx \sum_{\ell} e_{j+m_0,\ell-2^{m_0}k} c_{j+m_0,\ell}, \tag{18}$$

where $c_{j,k}$ are the standard discrete father wavelet coefficients produced by the (non-)decimated wavelet transform. Hence the mod-(non-)decimated wavelet coefficients can be produced by first performing the standard (non-)decimated wavelet transform and then computing the finite sum (18) for each coefficient.

Non-negative spectral estimates. We note here that for ease of exposition, our estimator does not incorporate any safeguards to ensure non-negativity. One such safeguard would simply be to truncate the estimator at zero. Another way to proceed could be to employ the method of information projection, used, for example, by Antoniadis and Bigot (2006). The objective of this type of regularisation is simple: to find a function, which is a member of a suitable "exponential family" and is thus guaranteed to be positive, which best approximates the unregularised (plain) estimate in a given norm. This could be incorporated into our method, with some effort. However, one drawback of this kind of regularisation is that it requires computationally intensive iteration using a Newton-Raphson type algorithm which is described in detail in Section 6 of Antoniadis and Bigot (2006).

5.2 Simulations and comparison with existing estimators

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` (respectively) from the R package `lokern`.

We apply these methods to two "test processes". Process A, the first test process, is the same as in Neumann (1996) and is defined by $X_t^A = Y_t + \frac{1}{2}Z_t$, where

$$Y_t + \frac{1}{5}Y_{t-1} + \frac{9}{10}Y_{t-2} = \varepsilon_t + \varepsilon_{t-2}, \tag{19}$$

and $\{\varepsilon_t\}$, $\{Z_t\}$ are independent Gaussian white noise processes with mean zero and variance one. All our simulations are based on sample paths of size $N = 1024$. Process B is a simple MA(2) process $X_t^B = \varepsilon_t + \varepsilon_{t-1} - \varepsilon_{t-2}$. Neumann's test process A is designed to possess a sharp spectral peak whereas the second test process B's spectrum is much smoother.

Note that 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 an estimate of the spectral density. Both Gao's and our wavelet-Fisz estimator use translation-invariant hard thresholding with all resolution levels thresholded (results for soft thresholding are worse and we do not report them). We use the "noise-free" version of our estimator, described in Section 4. With both Neumann's estimators we set the pilot estimator to be equal to 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 ordinates. Empirically, we have found that the performance of the wavelet-Fisz and Gao's estimators is not much affected by tapering. However, the performance of `glkerns` and `lokerns` deteriorates notably when tapering is introduced. This is due to the fact that the optimal bandwidth selection in both these methods is based on cross-validation which often does not perform well if the data are correlated. Thus, we only report simulation results for the untapered periodogram.

Figure 1 shows the true spectral density for process A and sample reconstructions using 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 given that it is a linear method which is not capable of estimating both smooth and inhomogeneous regions of

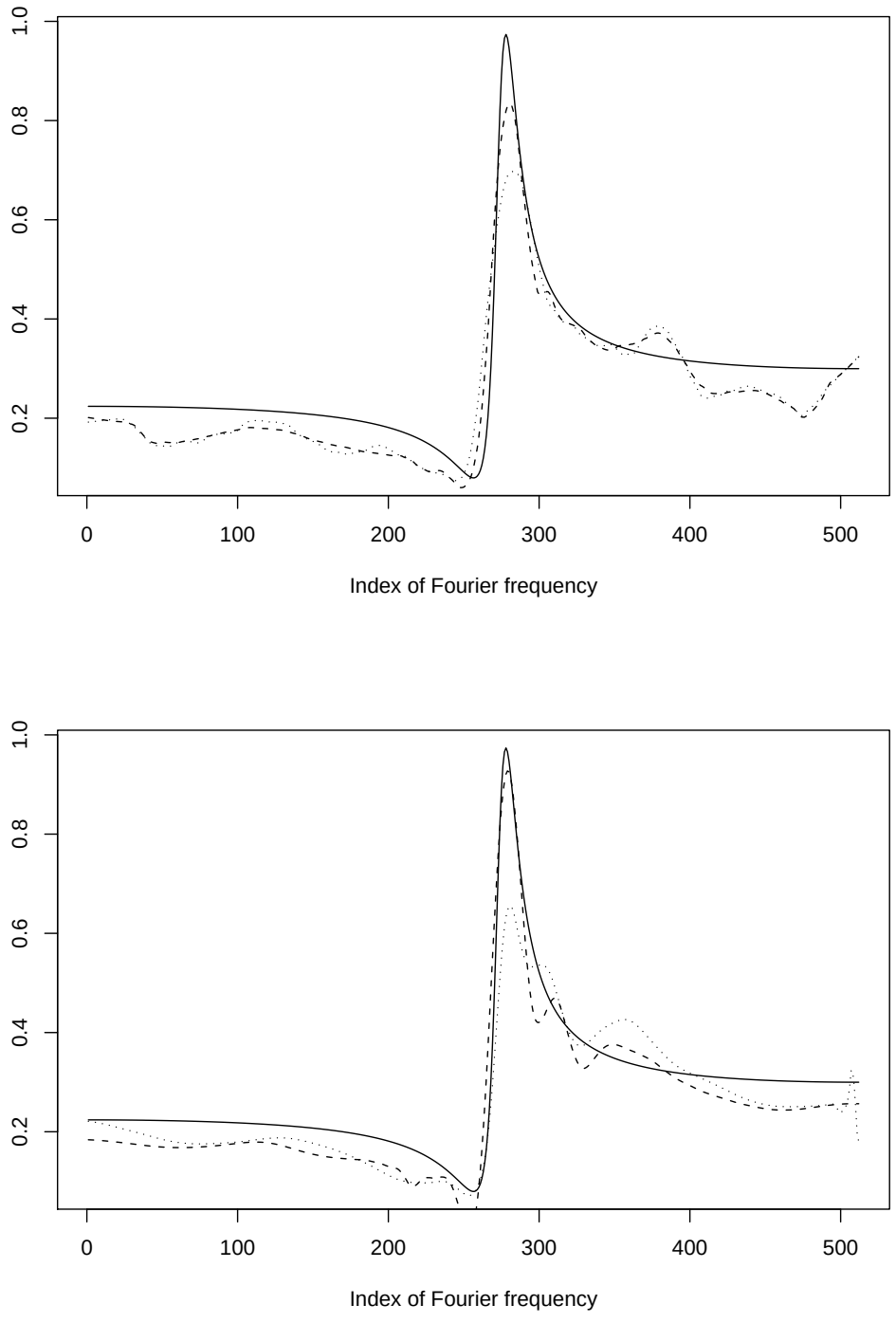

Figure 1: Top: true spectral density for process A (solid); reconstructions using `glkerns` (dotted) and `lokerns` (dashed). Bottom: true spectral density (solid); reconstructions using Gao's method (dotted) and wavelet-Fisz (dashed). See text for details.

the spectral density 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. (Although *lokerns* comes close when wavelet-Fisz uses Daubechies Extremal Phase wavelet with 7 vanishing moments). 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 appears to be 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`. Even though are results seem robust one might be interested in selecting the 'best' wavelet to use. A detailed discussion of this question is beyond the scope of the current paper but, for non-tapered data at least, 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.* Percentage by which the ISE of wavelet-Fisz (averaged over 100 sample paths) is lower than that of the five competitors. (Neumann CS=32 shift cycle-spun 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.* Percentage by which the ISE of wavelet-Fisz (averaged over 100 sample paths) is lower than that of the five competitors. (Neumann CS=32 shift cycle-spun 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

A Proofs

A.1 Auxiliary results

Bias, variance and cumulants of $\widetilde{\alpha}_{j,k}$. We first establish the bias, variance and higher cumulants of $\widetilde{\alpha}_{j,k}$. We recall that $H_p := \int_0^1 h^p(x)\, dx$.

Lemma A.1 *Suppose that Assumptions 2.1, 2.2, 2.3 and 3.1 hold. The following hold uniformly over* $(j,k) \in \mathcal{J}_N$.

(a) $E\ \widetilde{\alpha}_{j,k} = \alpha_{j,k} + O(2^{j/2} \log N/N)$

(b) $\sigma_{j,k}^2 := var(\widetilde{\alpha}_{j,k}) = 2\pi/N\ \ H_4/H_2^2 \int_\pi^\pi \{f(\omega)\}^2\, \psi_{j,k}(\omega)\, [\psi_{j,k}(\omega)+\psi_{j,k}(-\omega)]\, d\omega + o(N^{-1}) + O(2^{-j}N^{-1})$

(c) $|\mathrm{cum}_p(\widetilde{\alpha}_{j,k}/\sigma_{j,k})| \le C^p\ (p!)^{2+2\gamma}\, N^{-1}\ (2^{j/2} \log N/N)^{-(p-2)}$ *for* $p \ge 3$.

For the proof we refer the reader to Neumann (1996), Proposition 3.1.

Strong asymptotic normality of $\widetilde{\alpha}_{j,k}$. The technique for proving the near-optimal rate of mean-square convergence of ($\widehat{f}$ and) $\widetilde{f}$ is based on the following strong form of asymptotic normality

of the empirical wavelet coefficients $\widetilde{\alpha}_{j,k}$:

$$\frac{P(\pm(\widetilde{\alpha}_{j,k} - \alpha_{j,k})/\sigma_{j,k} \geq x)}{1 - \Phi(x)} \to 1, \tag{20}$$

uniformly over $(j,k) \in \mathcal{J}_N$, $x \leq \Delta_\gamma$, with $\Delta_\gamma = o(\Delta^{1/(3+4\gamma)})$ and $\Delta = O(N^{\delta/2}/\log N)$, where δ is as described in the last paragraph of Section 3.1. Note that $\Phi(x)$ is the cdf of the standard normal. The proof of (20) relies, amongst others, on the asymptotic behaviour of the cumulants of $\widetilde{\alpha}_{j,k}$ of order two and higher, as specified in Lemma A.1. For details, we refer again to Neumann (1996).

A.2 'Theoretical' estimator $\widehat{f}$

The proof of the near-optimal rate of convergence of the estimator $\widetilde{f}$ proceeds via proving, first, a similar result for the 'theoretical' estimator $\widehat{f}$ defined in formula (5), with a particular choice of $\lambda_{j,k}$. In fact, we define $\lambda_{j,k}$ to be 'theoretical counterparts' of the estimated thresholds $\widehat{\lambda}_{j,k}$ of formula (6):

$$\lambda_{j,k} = \theta_{j,k}\sqrt{2\log(\#\mathcal{J}_N)}, \tag{21}$$

with

$$\theta_{j,k} = c_N \int_{-\pi}^{\pi} \kappa_{j,k}(\omega) f(\omega) d\omega. \tag{22}$$

Note that since $\int \kappa_{j,k}(\omega)d\omega = 1$, the parameters $\theta_{j,k}$ can be interpreted as local weighted l_1 norms of f at scale j and location k.

Mean-square convergence of $\widehat{f}$. We now state a theorem on the mean-square convergence of our 'bona fide' estimator $\widehat{f}$ defined in (5).

Assumption A.1 *Let $\sigma_{j,k}^2 = var(\widetilde{\alpha}_{j,k})$ (see Lemma A.1 for the exact formula for $\sigma_{j,k}^2$), and let m be as in the statement of Theorem A.1. Thresholds $\lambda_{j,k}$ are such that*

$$\sum_{(j,k)\in\mathcal{J}_N} \left(\frac{\lambda_{j,k}}{\sigma_{j,k}} + 1\right) \phi(\lambda_{j,k}/\sigma_{j,k}) = O(N^{1/(2m+1)}) \tag{23}$$

$$\sup_{(j,k)\in\mathcal{J}_N} \lambda_{j,k} = O(N^{-1/2}\sqrt{\log N}), \tag{24}$$

where ϕ is the standard normal density.

Note that by Lemma A.2 (with $\alpha_N = 1$), our thresholds $\lambda_{j,k}$, defined in formula (21), satisfy Assumption A.1. Heuristically speaking, Assumption A.1 controls the distance between the thresholds $\lambda_{j,k}$ and the benchmark "universal" thresholds $\lambda^N_{j,k} = \sigma_{j,k}\sqrt{2\log(\#\mathcal{J}_N)}$, used, for example, by Neumann (1996). This is needed as the latter are motivated by the universal threshold theory in the Gaussian regression case, and our thresholds $\lambda_{j,k}$ also rely on asymptotic Gaussianity arguments.

Theorem A.1 *Suppose that Assumptions 2.1, 2.2, 2.3 and 3.1 hold. Let $\lambda_{j,k}$ be any thresholds satisfying Assumption A.1; for example, those defined in formula (21). Let $B^m_{p,q}(C)$ be a Besov ball of radius $C < \infty$ with $m, p \geq 1$. We have*

$$\sup_{f \in B^m_{p,q}(C)} E \int_{-\pi}^{\pi} |\widetilde{f}(\omega) - f(\omega)|^2 d\omega = O\left((\log N/N)^{2m/(2m+1)}\right). \tag{25}$$

For the proof of this theorem we refer to Neumann (1996), as it is parallel to the proof of Theorem 5.1 therein.

Choice of c_N. Finally, we motivate the choice $c_N = \sqrt{2\pi/N \;\; H_4/H_2^2}$ in formulae (7) and (22). A desired property of $\theta_{j,k}$, required so that (23) may hold, is that $\theta_{j,k}$ should be "almost equal" to $\sigma_{j,k}$ for fine-scale coefficients. From Lemma A.1(b), we observe that

$$\sigma_{j,k} \leq \sqrt{\frac{2\pi H_4}{N\ H_2^2}}\ \overline{f}_{j,k}\ \sqrt{1 + \int_{-\pi}^{\pi} |\psi_{j,k}(\omega)\psi_{j,k}(-\omega)| d\omega} + o(N^{-1}) + O(2^{-j}N^{-1})\ , \tag{26}$$

where $\overline{f}_{j,k} = \sup\{f(\omega) : \omega \in \mathrm{supp}(\psi_{j,k})\}$ (for later purposes, we also denote $\underline{f}_{j,k} = \inf\{f(\omega) : \omega \in \mathrm{supp}(\psi_{j,k})\}$). Further we note that $\int_{-\pi}^{\pi} |\psi_{j,k}(\omega)\psi_{j,k}(-\omega)|\ d\omega = 0$ except for a finite (and the same) number of k at each scale j for which $0 \in \mathrm{supp}\{\psi_{j,k}\}$. This motivates setting c_N to be equal to $c_N = \sqrt{2\pi/N \;\; H_4/H_2^2} =: c\, N^{-1/2}$.

A.3 Proof of Theorem 3.1

Instead of Theorem 3.1, we prove the more general Theorem A.2 below.

Theorem A.2 *Suppose that Assumptions 2.1, 2.2, 2.3 and 3.1 hold. Let $\widehat{\lambda}_{j,k}$ be any thresholds satisfying Assumption A.2; for example, those defined in formula (6). Let $B^m_{p,q}(C)$ 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*

$$\sup_{f \in B^m_{p,q}(C)} \boldsymbol{E} \int_{-\pi}^{\pi} |\widetilde{f}(\omega) - f(\omega)|^2 d\omega \;=\; O\left((\log\, N/N)^{2m/(2m+1)}\right). \tag{27}$$

The "logic" of the proof of Theorem A.2 is like the proof of Theorem 6.1 of Neumann (1996). It works via showing the analogous result for our 'bona fide' estimator $\widehat{f}$ defined in (5), based on thresholds $\lambda_{j,k}$ which are required to satisfy Assumption A.1.

In the remainder of this section, we demonstrate that if $\lambda_{j,k}$ are as defined in (21), then our random thresholds $\widehat{\lambda}_{j,k}$, defined in formula (6), satisfy Assumption A.2.

Assumption A.2 *Let* $\lambda_{j,k}$ *be any thresholds satisfying Assumption A.1, and let* m *be as in the statement of Theorem A.2. Thresholds* $\widehat{\lambda}_{j,k}$ *are such that*

(i) There exists a $\nu < 1/(2m+1)$ *and a positive sequence* α_N *approaching 1 from below as* $N \to \infty$ *such that*

$$\sum_{(j,k)\in\mathcal{J}_N} \mathbb{P}(\hat{\lambda}_{j,k} < \alpha_N\ \lambda_{j,k}) \;= O(N^{\nu});$$

(ii) There exists a constant $\tilde{C} < \infty$ *such that*

$$\sum_{(j,k)\in\mathcal{J}_N} \mathbb{P}(\hat{\lambda}_{j,k} > \tilde{C}\ N^{-1/2}\sqrt{\log\, N}) \;=\; O(N^{-\frac{2m}{2m+1}}).$$

Essentially, Assumption A.2 requires that the random thresholds $\widehat{\lambda}_{j,k}$ are not "too far off", in an appropriate sense, from the theoretical thresholds $\lambda_{j,k}$.

To show that our data-driven wavelet-Fisz thresholds (6) fulfil Assumption A.2, we specify what we call, respectively, "low" and "high" deterministic thresholds. We propose

$$\begin{aligned} \lambda^{(l)}_{j,k} &= \alpha_N \theta_{j,k}\ \sqrt{2\log(\#\mathcal{J}_N)} \\ \lambda^{(u)}_{j,k} &= \tilde{C}\ N^{-1/2}\ \sqrt{\log\, N}, \end{aligned}$$

with an $\alpha_N \to 1_-$, and with an appropriate constant $\tilde{C}$ specified below. Both $\lambda^{(l)}_{j,k}$ and $\lambda^{(u)}_{j,k}$ need to satisfy Assumption A.1.

Lemma A.2 *Suppose that Assumptions 2.1 and 2.3 hold. If* $\tilde{C} \geq c\sqrt{2}\sup_\omega f(\omega)$, *then* $\forall\, j,k \quad \lambda_{j,k}^{(u)} \geq \sup_{j,k} \lambda_{j,k}^{(l)}$, *and both* $\lambda_{j,k}^{(l)}$ *and* $\lambda_{j,k}^{(u)}$ *satisfy Assumption A.1.*

Proof. It is easy to check that if $\tilde{C} \geq c\sqrt{2}\sup_\omega f(\omega)$, then $\forall\, j,k \quad \lambda_{j,k}^{(u)} \geq \sup_{j,k} \lambda_{j,k}^{(l)}$. We now check (23) for $\lambda_{j,k}^{(l)}$. The factor $\lambda_{j,k}^{(l)}/\sigma_{j,k} + 1$ only contributes a logarithmic term so we skip it. Denote $\mathcal{Z}_N = \mathcal{J}_N \cap \{(j,k) : 0 \in \mathrm{supp}(\psi_{j,k})\}$. We have

$$\left(\sum_{\mathcal{Z}_N} + \sum_{\mathcal{J}_N \setminus \mathcal{Z}_N}\right) \phi(\lambda_{j,k}^{(l)}/\sigma_{j,k}) = O(M \log\, N) + \sum_{\mathcal{J}_N \setminus \mathcal{Z}_N} \phi(\lambda_{j,k}^{(l)}/\sigma_{j,k}),$$

where M is a constant. It remains to investigate the sum on the RHS. Recalling that $H_p^{(N)} := \sum_{t=1}^N h^p(t/N)$ and using (26), we bound $\lambda_{j,k}^{(l)}/\sigma_{j,k}$ from below as follows:

$$\frac{\lambda_{j,k}^{(l)}}{\sigma_{j,k}} \geq \frac{\alpha_N c N^{-1/2} \underline{f}_{j,k} \sqrt{2\log(\#\mathcal{J}_N)}}{\sqrt{\frac{2\pi H_4^{(N)}}{(H_2^{(N)})^2}}\, \overline{f}_{j,k} + o(N^{-1}) + O(2^{-j}N^{-1})}.$$

We introduce $\beta_N = cN^{-1/2}/\sqrt{2\pi H_4^{(N)}/(H_2^{(N)})^2}$, noting that $\beta_N \to 1$. The above bound can be rewritten as

$$\begin{aligned}
&\frac{\alpha_N \beta_N \underline{f}_{j,k} \sqrt{2\log(\#\mathcal{J}_N)}}{\overline{f}_{j,k} + o(N^{-1/2}) + O(2^{-j}N^{-1/2})} \geq \\
&\quad \frac{\alpha_N \beta_N \underline{f}_{j,k} \sqrt{2\log(\#\mathcal{J}_N)}}{\overline{f}_{j,k}} + o(N^{-1/2}) + O(2^{-j}N^{-1/2}) = \\
&\quad \frac{\alpha_N \beta_N \underline{f}_{j,k} \sqrt{2\log(\#\mathcal{J}_N)}}{\overline{f}_{j,k}} \left(1 + o((N\log\, N)^{-1/2}) + O(2^{-j}(N\log\, N)^{-1/2})\right) =: \\
&\quad \frac{\alpha_N \tilde{\beta}_N \underline{f}_{j,k} \sqrt{2\log(\#\mathcal{J}_N)}}{\overline{f}_{j,k}},
\end{aligned}$$

where the first inequality holds by the convexity of $u(x) = 1/(a^2 + x)$ for small x; note that $\tilde{\beta}_N \to 1$. Denoting $\mu = \inf f(\omega)$, we have

$$\frac{\underline{f}_{j,k}}{\overline{f}_{j,k}} \geq \frac{\underline{f}_{j,k}}{\underline{f}_{j,k} + \mathrm{TV}(f)|_{\mathrm{supp}(\psi_{j,k})}} \geq \frac{\mu}{\mu + \mathrm{TV}(f)|_{\mathrm{supp}(\psi_{j,k})}},$$

where the last inequality follows from the fact that $v(x) = x/(x+a^2)$ is increasing on $[0,\infty)$. As in Neumann (1996), the proof of Lemma 6.1 (ii), we have $\sum_k \mathrm{TV}(f)|_{\mathrm{supp}(\psi_{j,k})} \leq O(1)\mathrm{TV}(f)$ and thus for a sequence $d_N \to 0$, at each scale j we have

$$\#\{k : \mathrm{TV}(f)|_{\mathrm{supp}(\psi_{j,k})} > d_N\} = O(d_N^{-1}). \tag{28}$$

Denote $\mathcal{D}_N = \mathcal{J}_N \cap \mathcal{Z}_N^c \cap \{(j,k) : \mathrm{TV}(f)|_{\mathrm{supp}(\psi_{j,k})} > d_N\}$. Note that by (28), at each scale j at most $O(d_N^{-1})$ coefficients are in $\mathcal{D}_N$. Denote further $\mathcal{E}_N = \mathcal{J}_N \cap \mathcal{Z}_N^c \cap \mathcal{D}_N^c$. Let $J^* - 1$ be the finest scale in $\mathcal{J}_N$. We have

$$\begin{aligned}
&\sum_{\mathcal{J}_N \setminus \mathcal{Z}_N} \phi(\lambda_{j,k}^{(l)}/\sigma_{j,k}) = \left(\sum_{\mathcal{D}_N} + \sum_{\mathcal{E}_N}\right) \phi(\lambda_{j,k}^{(l)}/\sigma_{j,k}) = O(d_N^{-1}\log\, N) + \\
&\quad \sum_{j=0}^{J^*-1}\sum_{k=1}^{2^j} \phi\left(\frac{\alpha_N \tilde{\beta}_N \mu \sqrt{2\log(\#\mathcal{J}_N)}}{\mu + d_N}\right) \leq O(d_N^{-1}\log\, N) + (2\pi)^{-1/2} \sum_{j=0}^{J^*-1} 2^{j - J^* \alpha_N^2 \tilde{\beta}_N^2 \left(\frac{\mu}{\mu+d_N}\right)^2} \\
&\quad = O(d_N^{-1}\log\, N) + O((\#\mathcal{J}_N)^{1-\alpha_N^2\tilde{\beta}_N^2\left(\frac{\mu}{\mu+d_N}\right)^2}) = O(d_N^{-1}\log\, N) + o(N^{1/(2m+1)}),
\end{aligned}$$

for any $m > 0$. The last equality follows from the fact that $1 - \alpha_N^2\tilde{\beta}_N^2\left(\frac{\mu}{\mu+d_N}\right)^2 \to 0$. Choosing $d_N = \log^{-1}\, N$ (say), we have that (23) is satisfied irrespective of the smoothness parameter m. Because the thresholds $\lambda_{j,k}^{(u)}$ are higher than $\lambda_{j,k}^{(l)}$, (23) also holds for $\lambda_{j,k}^{(u)}$. Obviously, (24) holds for $\lambda_{j,k}^{(u)}$, which implies that it also holds for $\lambda_{j,k}^{(l)}$, since $\lambda_{j,k}^{(l)}$ are lower than $\lambda_{j,k}^{(u)}$. □

To continue the proof of Theorem A.2, we show that our random thresholds (6) satisfy Assumption A.2. In order to do so, we show that our "estimators" $\widehat{\theta}_{j,k}$ fulfil a strong form of asymptotic normality, paralleling the one in (20) for $\widetilde{\alpha}_{j,k}$.

Proposition A.1 *Suppose that Assumptions 2.1, 2.2, 2.3 and 3.1 hold. Denote* $s_{j,k}^2 = var(\hat{\theta}_{j,k})$. *We have*

$$\frac{P(\pm(\widehat{\theta}_{j,k} - \theta_{j,k})/s_{j,k} \geq x)}{1 - \Phi(x)} \to 1,$$

uniformly over $(j,k) \in \mathcal{J}_N$, $x \leq \Delta_\gamma$, *where* $\Delta_\gamma = o(\Delta^{1/(3+4\gamma)})$ *and* $\Delta = O(N^{\delta/2}/\log\, N)$.

Proof. We start with some clarifying remarks as to the orders of magnitude of $\hat{\theta}_{j,k}$, $\theta_{j,k}$ and $s_{j,k}$. As each $\kappa_{j,k}$ integrates to one in L_1, we have $\theta_{j,k} = O(N^{-1/2})$ uniformly over j,k.

The bias in estimating $\theta_{j,k}$ by $\hat{\theta}_{j,k}$ is of order $O(2^j \ \log N/N^{3/2})$. Its derivation completely parallels the proof of Lemma A.1(a), which is Proposition 3.1(i) in Neumann (1996): $E\hat{\theta}_{j,k} = \theta_{j,k} + O(c_N 2^j \log N/N)$.

The variance $s_{j,k}^2$ can be derived using Lemma 6 in Dahlhaus (1983), which is analogous to the proof of Lemma A.1(b):

$$s_{j,k}^2 \ = \ c_N^2 \ [c_N^2 \int_{-\pi}^{\pi} f^2(\omega)\kappa_{j,k}(\omega)(\kappa_{j,k}(\omega) + \kappa_{j,k}(-\omega))d\omega + o(2^j \ N^{-1}) + O(N^{-1})] \ .$$

With the normalisation $\int \kappa_{j,k}(\omega) \ d\omega = 1$, we remark that the first term in brackets is of order $O(2^j/N)$, which is easy to see since $\kappa_{j,k}(\omega) = 2^j \ \kappa(2^j\omega - k)$.

With this observation, one gets the following auxiliary results, useful for the remainder: $\int \kappa_{j,k}^2(\omega) \ d\omega = 2^j$ and $\sup_k \sup_\omega \kappa_{j,k}(\omega) = O(2^j)$. Thus, we note that the overall order of the leading term of $s_{j,k}$ is $2^{j/2}N^{-1}$. By Assumption 2.1(i), we have a uniform lower bound on $s_{j,k}$, which ensures that $c_N^{-1} \ s_{j,k} \geq \tilde{c} \ (2^j/N)^{1/2}$.

However, as we are going to study the ratios of $\hat{\theta}_{j,k}/s_{j,k}$ and of the bias divided by $s_{j,k}$, the normalisation of each of the quantities $\hat{\theta}_{j,k}$, $\theta_{j,k}$ and $s_{j,k}$ by the factor c_N (which is of order $N^{-1/2}$) will cancel out. For example, for the bias treatment considered in Theorem A.1, we get

$$E\{(\hat{\theta}_{j,k} - \theta_{j,k})/s_{j,k}\} = \ O\left(\frac{2^j \log N/N}{(2^j/N)^{1/2}}\right) = O((2^j/N)^{1/2} \log N) \ . \tag{29}$$

Note that since $2^j \leq C \ N^{1-\delta}$, the rescaled bias converges to zero.

The key property to make the strong asymptotic normality work is the treatment of all higher-order cumulants of $\hat{\theta}_{j,k}/s_{j,k}$. We first parallel Lemma A.1(c):

$$\mathrm{cum}_p(\hat{\theta}_{j,k}) \ = \ O(c_N^p \ (p!)^{2+2\gamma} \ 2^j/N \ (2^j/N \ \log N)^{p-2}) \ , \tag{30}$$

for all $p \geq 3$. This immediately entails the result for the cumulants rescaled by the standard deviation $s_{j,k}$:

$$\mathrm{cum}_p(\hat{\theta}_{j,k}/s_{j,k}) \ \leq \ C(p,\gamma) \ (N^{-\delta/2} \ \log N)^{p-2} \ ,$$

uniformly over $(j,k) \in \mathcal{J}_N$ for all $p \geq 3$, with a bounded constant $C(p,\gamma)$. The proof of this last result is straightforward if we observe that the normalisation by the inverse of $s_{j,k}$ leads

to the division of the order of equation (30) by the factor of $c_N^p\ (2^j/N)^{p/2}$. (Compare also the order of the leading term of $s_{j,k}$ shown to be $2^{j/2}N^{-1}$). Elementary calculations give the order for the normalised cumulant as $((2^j/N)^{1/2}\ \log\ N)^{p-2}$, which, however, is clearly of order $(N^{-\delta/2}\ \log\ N)^{p-2}$.

We close by noting that one can choose $\Delta = N^{\delta/2}/\log\ N$ to control the bias in formula (29) above. This completes the proof of Proposition A.1, which is similar to the proof of Neumann (1996), Theorem 4.1. □

With this strong form of asymptotic normality, we now verify that our random thresholds (6) satisfy Assumption A.2.

Verifying Assumption A.2(i):

Define $\Delta_\delta = (N^{\delta/2}/\log\ N)^{1/(3+4\gamma)}/\log\ N$ (note that it satisfies the requirement for Δ_γ of Proposition A.1) and let Z be standard normal. All C_i's are positive constants. Assumption A.2(i) writes as follows:

$$\sum_{(j,k)\in\mathcal{J}_N} P(\hat{\lambda}_{j,k} < \lambda_{j,k}^{(l)}) = \sum_{(j,k)\in\mathcal{J}_N} P((\theta_{j,k} - \hat{\theta}_{j,k})/s_{j,k} > \theta_{j,k}(1-\alpha_N)/s_{j,k}) \leq$$
$$\sum_{(j,k)\in\mathcal{J}_N} P((\theta_{j,k} - \hat{\theta}_{j,k})/s_{j,k} > \min(\theta_{j,k}(1-\alpha_N)/s_{j,k}, \Delta_\delta)) \leq$$
$$C_1 \sum_{(j,k)\in\mathcal{J}_N} P(Z > \min(\theta_{j,k}(1-\alpha_N)/s_{j,k}, \Delta_\delta)) \leq$$
$$C_1 \sum_{(j,k)\in\mathcal{J}_N} P(Z > \theta_{j,k}(1-\alpha_N)/s_{j,k}) + C_1 \sum_{(j,k)\in\mathcal{J}_N} P(Z > \Delta_\delta) =: I + II.$$

Recalling the orders of $\theta_{j,k}$ and $s_{j,k}$, and denoting $b = \exp(C_2^2/2)$, we bound I as follows:

$$C_1 \sum_{(j,k)\in\mathcal{J}_N} P(Z > C_2 2^{(J-j)/2}(1-\alpha_N) \leq C_3 \sum_{(j,k)\in\mathcal{J}_N} \exp(-C_2^2 2^{J-j}(1-\alpha_N)^2/2) =$$
$$C_3 \sum_{j=0}^{J^*} 2^j b^{-(1-\alpha_N)^2 2^{J-j}} \leq C^3 b^{-(1-\alpha_N)^2 2^{J-J^*}} \sum_{j=0}^{J^*} 2^j \leq C_4 2^{J^*} b^{-(1-\alpha_N)^2 2^{J-J^*}}.$$

Recalling that $2^{J^*} = N^{1-\delta}$, and logging the above bound we obtain $(1-\delta)\log_b N - (1-\alpha_N)^2 N^\delta$. To ensure that it is bounded in N (say $\leq C_5$), we need to take

$$\alpha_N \leq 1 - \sqrt{\frac{(1-\delta)\log_b N - C_5}{N^\delta}}.$$

Turning now to II, we have $2^{J^*} P(Z > \Delta_\delta) \leq N^{1-\delta} \exp(-\Delta_\delta^2/2)$. Logging this, we get $(1-\delta)\log N - N^{\delta/(3+4\gamma)}/(2\log^{2+2/(3+4\gamma)} N)$, which tends to $-\infty$, which means that $II \to 0$.

Verifying Assumption A.2(ii):

$$\begin{aligned}
&\sum_{(j,k)\in\mathcal{J}_N} P(\hat{\lambda}_{j,k} > \lambda_{j,k}^{(u)}) \leq \sum_{(j,k)\in\mathcal{J}_N} P(\hat{\theta}_{j,k} > C_6) = \\
&\quad \sum_{(j,k)\in\mathcal{J}_N} P((\hat{\theta}_{j,k} - \theta_{j,k})/s_{j,k} > (C_6 - \theta_{j,k})/s_{j,k}) \leq \\
&\quad \sum_{(j,k)\in\mathcal{J}_N} P((\hat{\theta}_{j,k} - \theta_{j,k})/s_{j,k} > C_7 N^{1/2} 2^{(J-j)/2}) \leq \\
&\quad \sum_{(j,k)\in\mathcal{J}_N} P((\hat{\theta}_{j,k} - \theta_{j,k})/s_{j,k} > \Delta_\delta) \leq C_8 \sum_{(j,k)\in\mathcal{J}_N} P(Z > \Delta_\delta) \leq C_8 N^{1-\delta} \exp(-\Delta_\delta^2/2).
\end{aligned}$$

This implies that we need to show $N^{1-\delta}\exp(-N^{\delta/(3+4\gamma)}/(2\log^{2+2/(3+4\gamma)} N)) \leq N^{-2m/(2m+1)}$. But this is asymptotically true, which becomes obvious once both sides are logged.

The remainder of the proof of Theorem A.2 is like the proof of Theorem 6.1 of Neumann (1996).

Acknowledgement

This work was supported by EPSRC Small Grant GR/T22278/01 which funded a visit by RvS to Bristol in Autumn 2004. PF thanks RvS for financial support and hospitality during visits to Louvain-la-Neuve and Kaiserslautern. He is also grateful to the Department of Mathematics, Imperial College London, where a large part of this work was done, for an excellent research environment. GPN was also supported by EPSRC Advanced Research Fellowship Grant GR/A01664/01 and by Responsive Mode Grant EP/D005221/1. The research of RvS was also supported by IAP research network grant P 5/24 of the Belgian government (Belgian Science

Policy).

References

A. Antoniadis and J. Bigot. Poisson inverse problems. *Annals of Statistics*, 34:2132–2158, 2006.

S. Barber, G.P. Nason, and B.W. Silverman. Posterior probability intervals for wavelet thresholding. *Journal of the Royal Statistical Society Series 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.

R.R. Coifman and D.L. Donoho. Translation-invariant de-noising. In Anestis Antoniadis and Georges Oppenheim, editors, *Wavelets and Statistics*, volume 103 of *Lecture Notes in Statistics*, pages 125–150. Springer-Verlag, New York, 1995.

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. *Journal of Time Series Analysis*, 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). *Journal of the Royal Statistical Society Series B*, 57:301–369, 1995.

P. Fryzlewicz and G.P. Nason. A Haar-Fisz algorithm for Poisson intensity estimation. *Journal of Computational and Graphical Statistics*, 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, 1997.

D. Herrick. *Wavelet Methods for Curve Estimation*. PhD thesis, University of Bristol, Bristol, U.K., 2000.

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, 2 edition, 1995.

S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattn Anal. Mach. Intell.*, 11:674–693, 1989.

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. *Statistics and Computing*, 12:219–227, 2002.

G. P. Nason, R. von Sachs, and G. Kroisandt. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society. Series B*, 62:271–292, 2000.

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. *Journal of Time Series Analysis*, 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.

B. Vidakovic. *Statistical Modeling by Wavelets*. Wiley, New York, 1999.