Haar-Fisz estimation of evolutionary wavelet spectra

Piotr Fryzlewicz and Guy P. Nason

University of Bristol, UK

[March 21, 2006]

Summary. We propose a new "Haar-Fisz" technique for estimating the time-varying, piecewise constant local variance of a locally stationary Gaussian time series. We apply our technique to the estimation of the spectral structure in the Locally Stationary Wavelet model. Our method combines Haar wavelets and the variance stabilizing Fisz transform. The resulting estimator is mean-square consistent, rapidly computable, easy to implement, and performs well in practice. We also introduce the "Haar-Fisz transform", a device for stabilizing the variance of scaled chi-square data and bringing their distribution close to Gaussianity.

Keywords: heteroscedasticity, log transform, thresholding estimators, wavelet periodogram, wavelet spectrum.

1 Introduction

Time series whose spectral properties vary over time arise in several fields, e.g. finance (Kim (1998), Fryzlewicz (2005)), biomedical statistics (Nason et al. (2000)) or geophysics (Sakiyama (2002)). Estimating the time-varying second-order stucture is essential for understanding the data and forecasting the series.

Models for processes with an evolutionary spectral structure are often modifications of the following classical Cramér representation for stationary processes: all zero-mean discrete-time stationary processes X_t can be represented as

$$X_t = \int_{(-\pi,\pi]} A(\omega) \exp(i\omega t) dZ(\omega), \quad t \in \mathbb{Z}, \tag{1}$$

where $A(\omega)$ is the amplitude, and $Z(\omega)$ is a process with orthonormal increments. Dahlhaus (1996) introduces a class of locally stationary processes which permit a "slow" evolution of the transfer function $A(\omega)$ over time. Other approaches stemming from (1) include Priestley (1965), Battaglia (1979), Mélard and Herteleer-De Schutter (1989), Mallat et al. (1998), Swift (2000) and Ombao et al. (2002).

Being localised both in time and in frequency, wavelets provide a natural alternative to the Fourier-based approach for modelling phenomena whose spectral characteristics evolve over time (see Vidakovic (1999) for an introduction to wavelets and their statistical applications). Nason et al. (2000) introduce the class of locally stationary wavelet (LSW) processes

Address for correspondence: Piotr Fryzlewicz, Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK
Email: p.z.fryzlewicz@bris.ac.uk

which uses non-decimated wavelets, rather than Fourier exponentials, as building blocks. The LSW model enables a time-scale decomposition of the process and permits a rigorous estimation of the *evolutionary wavelet spectrum* and the *local autocovariance*. The LSW class is well-suited for modelling processes believed to have an inherent multiscale structure, such as financial log-returns (see Calvet and Fisher (2001)), and offers the user freedom in choosing the underlying wavelet family. Wavelet-based estimators of the second-order structure of LSW processes are naturally localised and can be computed efficiently.

The estimator of the evolutionary wavelet spectrum proposed in Nason et al. (2000) arises as a linear combination of smoothed *wavelet periodograms*, which can be loosely defined as sequences of squared wavelet coefficients of the process. (Note that Chiann and Morettin (1999) consider a different kind of wavelet periodogram, based on the orthonormal wavelet transform, for stationary processes.) To achieve the smoothing of spatially inhomogeneous wavelet periodograms of Gaussian processes, Nason et al. (2000) recommend using wavelet shrinkage adapted to scaled χ^2 data, or alternatively, applying the variance-stabilizing logarithmic transformation and then proceeding with the smoothing. Neither of these approaches is perfect: for the wavelet shrinkage, a pilot estimate of the local variance is required. On the other hand, the log transform "flattens" the data, often obscuring interesting features, such as peaks or troughs.

The prime objective of this paper is to propose a new technique for estimating the evolutionary wavelet spectrum in the Gaussian LSW model with piecewise constant wavelet spectra. The core of our approach is a new multiscale method for smoothing the wavelet periodogram, powerfully combining Haar wavelets and the variance-stabilizing Fisz transform. To achieve the main objective, we take the following steps:

Section 2. We modify the LSW model by imposing a piecewise constant smoothness constraint on the evolutionary wavelet spectrum. This is a natural requirement as the proposed estimator of the evolutionary wavelet spectrum is based on the piecewise constant Haar wavelets. On the other hand, this approach offers the appealing possibility of modelling processes whose second-order structure evolves over time in a discontinuous fashion.

Section 3. The problem of smoothing the wavelet periodogram in the LSW model can be viewed as the problem of estimating a (piecewise constant) local variance of a zero-mean Gaussian process. With this in mind, we propose a new multiscale technique for estimating the piecewise constant local variance of a zero-mean Gaussian process. The new technique combines Haar wavelet thresholding and the variance-stabilizing Fisz transform; hence we label it the *Haar-Fisz* (HF) method. We prove the mean-square consistency of the HF estimator in this general setting.

Section 4. We propose a new estimator of the evolutionary wavelet spectrum of a Gaussian LSW process, based on the Haar-Fisz-smoothed wavelet periodogram. We prove the mean-square consistency of the proposed estimator.

Section 5. We discuss a data-driven method for selecting the parameter of our estimation procedure, and provide a simulated example.

Additionally, in Section 6, we introduce and discuss the *Haar-Fisz transform*, which brings the distribution of scaled χ^2 data closer to normality and stabilizes their variance. We remark that our estimation procedure introduced in Section 3 can be carried out in three steps, the first and last of which are the Haar-Fisz and the inverse Haar-Fisz transform, respectively. The proofs are deferred to the Appendix.

2 The LSW model

2.1 Definition

We start by defining the LSW model for locally stationary time series.

Definition 2.1 *A triangular stochastic array* $\{X_{t,T}\}_{t=0}^{T-1}$*, for* $T = 1, 2, \ldots$*, is in the class of LSW processes if there exists a mean-square representation*

$$X_{t,T} = \sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_j(k/T)\psi_{j,t-k}\xi_{j,k}, \qquad (2)$$

where $j \in \{-1, -2, \ldots\}$ *and* $k \in \mathbb{Z}$ *are, respectively, scale and location parameters,* $\psi_j = (\psi_{j,0}, \ldots, \psi_{j,\mathcal{L}_j})$ *are discrete, real-valued, compactly supported, non-decimated wavelet vectors, and* $\xi_{j,k}$ *are zero-mean orthonormal identically distributed random variables. Also, for each* $j \le -1$*,* $W_j(z) : [0, 1] \to \mathbb{R}$ *is a real-valued, piecewise constant function with a finite (but unknown) number of jumps. Let* L_j *denote the total magnitude of jumps in* $W_j^2(z)$*. The functions* $W_j(z)$ *satisfy*

- $\sum_{j=-\infty}^{-1} W_j^2(z) < \infty$ *uniformly in* z*,*
- $\sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty$.

In formula (2), the parameters $W_j(k/T)$ can be thought of as a scale- and location-dependent transfer function, while the non-decimated wavelet vectors ψ_j can be thought of as building blocks analogous to the Fourier exponentials in (1). Throughout the paper, we work with Gaussian LSW processes, i.e. our $\xi_{j,k}$ are distributed as $N(0, 1)$.

Haar wavelets are the simplest example of a wavelet system which can be used in formula (2). Denote $\mathbb{I}_A(k) = 1$ when k is in A and zero otherwise. Haar wavelets are defined by

$$\psi_{j,k} = 2^{j/2}\mathbb{I}_{\{0,\ldots,2^{-j-1}-1\}}(k) - 2^{j/2}\mathbb{I}_{\{2^{-j-1},\ldots,2^{-j}-1\}}(k),$$

for $j \in \{-1, -2, \ldots\}$ and $k \in \mathbb{Z}$, where $j = -1$ corresponds to the finest scale. Other Daubechies' compactly supported wavelets (Daubechies (1992)) can also be used.

The main quantity of interest in the LSW framework is the evolutionary wavelet spectrum $S_j(z) := W_j^2(z)$, $j = -1, -2, \ldots$, defined on the rescaled-time interval $z \in [0, 1]$. The main objective of this paper is to propose a new, mean-square consistent estimator of $S_j(z)$. Due to the rescaled time concept, the estimation of $S_j(z)$ is analogous to the estimation of a regression function.

From Definition 2.1, it is immediate that $\mathbb{E}X_{t,T} = 0$ and indeed, throughout the paper, we work with zero-mean processes. Such processes arise, for example, when the trend has been removed from the data, see e.g. von Sachs and MacGibbon (2000) for a wavelet-based technique for detrending locally stationary processes.

Our definition of an LSW process is a modification of the definition from Nason et al. (2000): we have replaced the Lipschitz-continuity constraint on $W_j(z)$ by the piecewise constant constraint, which enables the modelling of processes whose second-order structure evolves over time in a discontinuous (piecewise constant) manner. This regularity constraint is natural given that we base our estimation theory in the LSW model on Haar wavelets which are also piecewise constant. It is unclear to us whether and how the model can be extended to include processes with spectra from other smoothness classes but approximable by piecewise constant functions. The main technical difficulty in doing so stems from the fact that we simultaneously estimate an ensemble of functions $\{S_j(z)\}_j$.

For an extensive discussion of the philosophy and several aspects of LSW modelling the reader is referred to Nason et al. (2000). Estimation in the LSW framework is also considered in Van Bellegem and von Sachs (2003), who propose an adaptive technique for the pointwise estimation of the evolutionary wavelet spectrum, which is postulated to be of bounded total variation as a function of time and to satisfy a number of extra technical assumptions. The estimator is a local average of the “raw" spectrum, where the window length is chosen adaptively via iterative hypothesis testing until the largest interval of near-homogeneity has been found.

2.2 The wavelet periodogram in the LSW model

The basic statistic used by Nason et al. (2000) to estimate the evolutionary wavelet spectrum $S_j(z)$ is the *wavelet periodogram*, which also forms the basis of our estimation theory. The definition follows.

Definition 2.2 *Let $X_{t,T}$ be an LSW process constructed using the wavelet system ψ. The triangular stochastic array*

$$I_{t,T}^{(j)} = \left| \sum_s X_{s,T} \psi_{j,s-t} \right|^2$$

is called the wavelet periodogram of $X_{t,T}$ at scale j.

Throughout the paper, we assume that the reader is familiar with the fast Discrete Wavelet Transform (DWT; see Mallat (1989)), as well as with the fast Non-decimated DWT (NDWT; see Nason and Silverman (1995)). In practice, we only observe a single row of the triangular array $X_{t,T}$. The wavelet periodogram is not computed separately for each scale j but instead, we compute the full NDWT transform of the observed row of $X_{t,T}$ (e.g. with periodic boundary conditions), and then square the wavelet coefficients to obtain $I_{t,T}^{(j)}$ for $t = 0, \ldots, T-1$ and $j = -1, \ldots, -J(T)$, where $J(T) \leq \log_2 T$.

It is convenient to recall two further definitions from Nason et al. (2000) at this point: the *autocorrelation wavelets* $\Psi_j(\tau) = \sum_{k=-\infty}^{\infty} \psi_{j,k}\psi_{j,k+\tau}$ and the *autocorrelation wavelet* *inner product matrix* $A_{i,j} = \sum_\tau \Psi_i(\tau)\Psi_j(\tau)$. We also define the *cross-scale autocorrelation wavelets* as $\Psi_{i,j}(\tau) = \sum_{k=-\infty}^{\infty} \psi_{i,k}\psi_{j,k+\tau}$. Note that as $A_{i,j} = \sum_\tau \Psi_{i,j}^2(\tau)$, it follows that $A_{i,j} > 0$ for all i, j.

In the Lipschitz-continuous setting of Nason et al. (2000), $I_{\lfloor zT \rfloor,T}^{(j)}$ is an asymptotically unbiased estimator of the quantity

$$\beta_j(z) := \sum_{i=-\infty}^{-1} S_i(z)A_{i,j}. \tag{3}$$

In other words, the expectation of the wavelet periodogram, computed at a rescaled-time location z, converges pointwise to a linear combination of wavelet spectra at location z. Naturally enough, in our piecewise-constant setup, no such pointwise convergence occurs at or around the discontinuities; however, we are able to show that the integrated squared bias converges to zero. For this result to hold, it is convenient to introduce the following assumption.

Assumption 2.1 *The set of those locations* z *where (possibly infinitely many) functions* $S_j(z)$ *contain a jump, is finite. In other words, let* $\mathcal{B} := \{z : \exists j \quad \lim_{u\to z^-} S_j(u) \neq \lim_{u\to z^+} S_j(u)\}$. *We assume* $B := \#\mathcal{B} < \infty$.

The result follows.

Proposition 2.1 *Let* $I_{t,T}^{(j)}$ *be the wavelet periodogram at a fixed scale* j. *We have*

$$\mathbb{E}\, I_{t,T}^{(j)} = \sum_{i=-\infty}^{-1} \sum_{k=-\infty}^{\infty} S_i(k/T)\Psi_{i,j}^2(k-t).$$

Also, if Assumption 2.1 is satisfied, we have

$$T^{-1} \sum_{t=0}^{T-1} \left| \mathbb{E}\, I_{t,T}^{(j)} - \beta_j(t/T) \right|^2 = O(T^{-1}2^{-j}) + b_{j,T}, \tag{4}$$

where $b_{j,T}$ *depends on the sequence* $\{L_j\}_j$. *For example, if* $L_j = O(a^j)$ *for* $a > 2$ *then* $b_{j,T} = O\left(T^{\frac{1}{2\log_2 a - 1} - 1}\right)$, *which implies, in particular, that the rate of convergence in (4) is* $O\left\{T^{-\min\left(1-\frac{1}{2\log_2 a-1}, 1-\varepsilon\right)}\right\}$ *uniformly over* $j = -1, \ldots, -\varepsilon \log_2 T$.

We now briefly study the behaviour of the limiting function $\beta_j(z)$. Before we state the result, we specify two other useful assumptions.

Assumption 2.2 *There exists a positive constant* C_1 *such that for all* j, $S_j(z) \leq C_1 2^j$.

Note that, in particular, Assumption 2.2 is satisfied if $X_{t,T}$ is the standard white noise process, for which $S_j(z) = S_j = 2^j$ (see Fryzlewicz et al. (2003)).

Assumption 2.3 *$\beta_j(z)$ is bounded away from zero for all j.*

By (3), and given that $A_{i,j} > 0$, $\beta_j(z) \equiv 0$ on a subinterval of $[0, 1]$ would imply all $S_i(z) \equiv 0$ on that subinterval; thus the resulting process would be locally deterministic and equal to zero.

The following proposition describes the behaviour of $\beta_j(z)$.

Proposition 2.2 *For each j, the function $\beta_j(z)$ is finite for all z. Further, if Assumption 2.1 holds, then each $\beta_j(z)$ is a piecewise constant function with at most B jumps, each of which occurs in the set $\mathcal{B}$. Additionally, if Assumption 2.2 holds, then for all j, $\beta_j(z) \leq C_1$.*

Proposition 2.1 implies, in particular, the uniqueness of the functions $\beta_j(z)$ in the L_2 sense. Thus, by Proposition 2.2 and in view of the invertibility of the operator A (see Nason et al. (2000)), formula (3) also implies the uniqueness of the functions $S_j(z)$ in the L_2 sense.

Suppose that we knew how to estimate each $\beta_j(z)$. Formula (3) suggests that estimates of $S_j(z)$ could then be constructed by taking an appropriate linear combination of the estimated β_j's. This is in fact exactly the route we follow: loosely speaking, we first obtain estimates $\hat{\beta}_j(z)$ and then estimate $S_j(z)$ by $\hat{S}_j(z) = \sum_i \hat{\beta}_i(z)(A^{-1})_{i,j}$ (see Theorem 4.1 for the exact range of i in the above sum).

Since each wavelet periodogram ordinate is simply a squared wavelet coefficient of a zero-mean Gaussian time series, asymptotically centred (in the sense of Proposition 2.1) on $\beta_j(z)$, its distribution is that of a scaled χ^2_1 variable. This means that it is an inconsistent estimator of $\beta_j(z)$ and needs to be smoothed to achieve consistency. Note that a similar phenomenon also arises in the case of the classical Fourier periodogram for stationary time series, see e.g. Priestley (1981).

Smoothing the wavelet periodogram is by no means an easy task, due to the fact that the variance of the "noise" depends on the level of the signal (as argued directly above), to the low signal-to-noise ratio (we have $\mathbb{E}\, I^{(j)}_{t,T} / \{\text{var}(I^{(j)}_{t,T})\}^{1/2} = 2^{-1/2}$), as well as to the fact that $I^{(j)}_{t,T}$ is typically a correlated sequence, which can be demonstrated using the same techniques as in the proof of Proposition 2.1.

Figure 1 shows the Haar wavelet spectrum of the "concatenated Haar process" also considered in Nason et al. (2000), a sample path of length 1024 generated from it (with Gaussian innovations $\xi_{j,k}$), and the Haar periodograms of the simulated sample path at scales -1 and -2.

To summarise, we are faced with the following statistical problem: our set of observations are $I^{(j)}_{t,T} = (\mathbb{E}\, I^{(j)}_{t,T}) Z_t^2$, where Z_t are correlated $N(0, 1)$ variables; our initial aim is to estimate $\beta_j(t/T)$ using Haar wavelets (which are the natural choice in this setting as the true $\beta_j(z)$'s are piecewise constant). Neumann and von Sachs (1995) used a nonlinear wavelet estimation technique in a setting similar to the above. However, their method involved finding a pre-estimate of the local variance of the observations (in our case: $2(\mathbb{E}\, I^{(j)}_{t,T})^2$). This is an obvious drawback of the estimation procedure and can hamper the practical performance of the method, see e.g. Fryzlewicz (2005).

In order to avoid having to find a pre-estimate of the local variance, an obvious step would be to take logarithms to transform the model from multiplicative to additive and

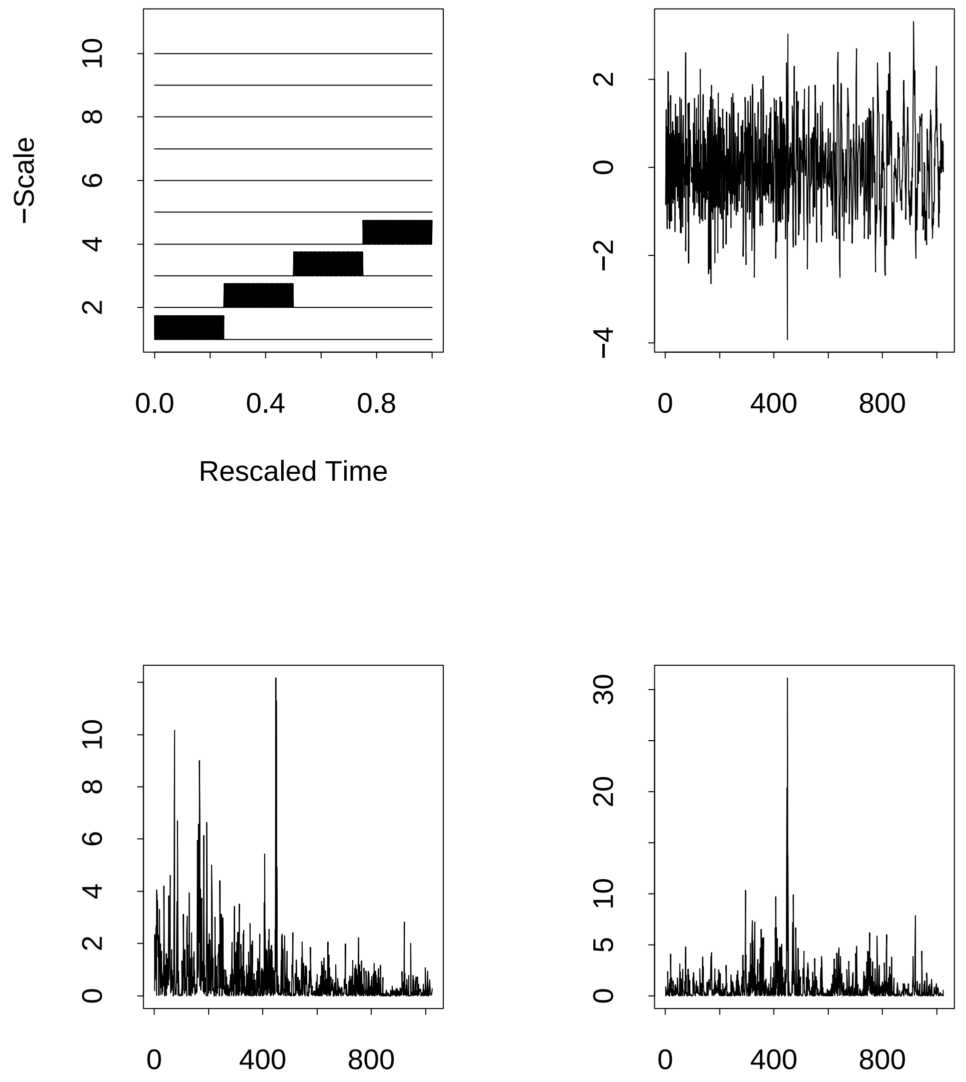

Figure 1: Top left: spectrum of the concatenated Haar process, top right: sample path generated from it, bottom left: wavelet periodogram of the sample path at scale -1, bottom right: wavelet periodogram of the sample path at scale -2.

stabilize the variance. If there was no correlation in Z_t, the resulting model would be similar to the representation of the log-periodogram of a second-order stationary process proposed by Wahba (1980). Several authors proposed wavelet techniques for the estimation of the log-periodogram (see e.g. Moulin (1994), Gao (1997), Walden et al. (1998), Pensky and Vidakovic (2004)), and we conjecture that some of those techniques could be adapted to our framework. However, any wavelet estimator in the logged model would possess two undesirable properties:

- It would be an estimate of $\log \beta_j(t/T)$ (and not $\beta_j(t/T)$ itself). Exponentiating this estimate would yield an estimate of $\beta_j(t/T)$; however, statistical properties of the latter, such as mean-square consistency, would not be easy to establish.
- It would suffer from a bias of order $\mathbb{E} \log Z_t^2$.

Also, the log transform "flattens" the data, thus obscuring potentially interesting features, such as peaks or troughs. In contrast to these unwelcome characteristics, the Haar-Fisz estimation technique which we propose below enjoys the following properties:

- It uses a "multiscale" variance-stablizing step, which eliminates the need for a local variance pre-estimation;
- It yields an asymptotically unbiased, mean-square consistent estimate of $\beta_j(t/T)$ (as opposed to $\log \beta_j(t/T)$), which removes the need for a bias correction factor.

Moreover, as we demonstrate below, it is conceptually simple, fast, easy to code, and performs well in practice. Since our estimator is piecewise constant, we also mention here some other recent estimation techniques for non-stationary time series which yield piecewise constant estimators of their second-order structure (albeit in different models). Apart from the above-mentioned Van Bellegem and von Sachs (2003), we also draw the reader's attention to Davis et al. (2006) and Dahlhaus and Polonik (2006). The former work aims at finding structural breaks in piecewise-stationary time series by use of the Minimum Description Length principle and a computer-intensive genetic algorithm. Although our objectives are different in that we aim at minimising the Mean-Square Error and provide an estimation methodology which is fast, the common denominator is that we also operate in the class of piecewise-stationary processes. Dahlhaus and Polonik (2006) consider, in particular, sieve estimation in the Gaussian autoregressive model with constant coefficients and a monotonically increasing variance function. The resulting variance estimate is piecewise constant.

3 Haar-Fisz estimation of local means of scaled χ_1^2 data

As stated in Section 1, our first objective is to propose a new multiscale technique for smoothing the wavelet periodogram in the LSW model. However, we feel that it is of independent interest to introduce and study our methodology in the broader context of estimating the local means of scaled χ_1^2 data, which do not necessarily need to represent wavelet periodograms. We formulate our generic problem as follows: we observe

$$Y_{t,T}^2 = \sigma_{t,T}^2 Z_{t,T}^2, \quad t = 0, \ldots, T-1, \tag{5}$$

where

- $\sigma_{t,T}^2$ is deterministic and "close" to a piecewise constant function $\sigma^2(z)$ in the sense that $T^{-1}\sum_{t=0}^{T-1} |\sigma_{t,T}^2 - \sigma^2(t/T)|^2 =: a_T = o_T(\log^{-1} T)$; further, $\sigma^2(z)$ is bounded from above and away from zero, with a finite but unknown number of jumps (the number of jumps is denoted by B);
- the vector $\{Z_{t,T}\}_{t=0}^{T-1}$ is multivariate normal with mean zero and variance one, and asymptotically, its autocorrelation sequence is absolutely summable; that is the function $\rho(\tau) := \sup_{i,T} |\mathrm{cor}(Z_{i,T}, Z_{i+\tau,T})|$ satisfies $\rho_\infty^1 < \infty$, where $\rho_\infty^p := \sum_\tau \rho^p(\tau)$. (A simple example of $Z_{t,T}$ satisfying this requirement is a short-memory stationary process with mean zero and variance one, for which $\rho(\tau) = |\mathrm{cor}(Z_{i,T}, Z_{i+\tau,T})|$; the process $Y_{t,T}$ is then a time-modulated stationary process.)

In the above model, our aim is to estimate $\sigma^2(z)$. Later, we will apply our proposed general estimation method to the wavelet periodogram, by taking $Y_{t,T}^2 = I_{t,T}^{(j)}$, $\sigma^2(z) = \beta_j(z)$ and $\sigma_{t,T}^2 = \mathbb{E}\, I_{t,T}^{(j)}$.

Our Haar-Fisz estimator uses the principle of nonlinear (Haar) wavelet shrinkage, thus being potentially well-suited for the estimation of $\sigma^2(z)$ even if it is spatially inhomogeneous, i.e. if its regularity varies from one region to another. Note that Fryzlewicz et al. (2006) propose and study Haar-Fisz estimation in a multiplicative model related to (5), in which the variables $Z_{t,T}$ are independent (but not necessarily Gaussian) and $\sigma_{t,T} = \sigma(t/T)$. These differences mean that distinct proof techniques are needed in our case.

We now outline our Haar-Fisz estimation algorithm. The input to the estimation algorithm is the vector $\{Y_{t,T}^2\}_{t=0}^{T-1}$: here, we assume that T is an integer power of two. To simplify the notation, we drop the subscript T and consider the sequence $Y_t^2 := Y_{t,T}^2$. We denote $M = \log_2 T$. The algorithm proceeds as follows:

1. Compute the Haar decomposition of Y_t^2 using the following algorithm:
   (a) Let $s_{M,n} := Y_n^2$, $n = 0, \ldots, 2^M - 1$.
   (b) For each $m = M-1, M-2, \ldots, 0$, recursively form vectors $\mathbf{s}_m$, $\mathbf{d}_m$ and $\mathbf{f}_m$ with elements:

$$\begin{aligned} s_{m,n} &= \frac{s_{m+1,2n} + s_{m+1,2n+1}}{\sqrt{2}} \\ d_{m,n} &= \frac{s_{m+1,2n} - s_{m+1,2n+1}}{\sqrt{2}} \\ f_{m,n} &= \frac{d_{m,n}}{s_{m,n}}, \end{aligned}$$

   where $n = 0, \ldots, 2^m - 1$.

2. Let $\mu_{m,n}$ denote the true Haar detail coefficient of $\sigma^2(z)$ at scale m, location n (i.e. a quantity computed like the $d_{m,n}$ in point 1.(b) above but with $\sigma^2(t/T)$ as the input vector in point 1.(a)). For most levels m (in a sense to be made precise later), estimate $\mu_{m,n}$ by

$$\hat{\mu}_{m,n} = s_{m,n}\,\text{sgn}(f_{m,n})\,(|f_{m,n}| - t_m)_+ \quad \text{(soft thresholding)}, \tag{6}$$

   where $\mathbb{I}(\cdot)$ and sgn$(\cdot)$ are the indicator and signum functions, respectively, and $(x)_+ = \max(0, x)$. In other words, our estimator returns zero if and only if the corresponding *Haar-Fisz coefficient* $f_{m,n}$ does not exceed (in absolute value) a certain threshold t_m (to be specified later). Note that this is different to classical wavelet thresholding in that the thresholded quantity $s_{m,n}\,\text{sgn}(f_{m,n})$ and the "thresholding statistic" $f_{m,n}$ are *different*.

   As is typically done in classical wavelet thresholding, we leave the coarsest-scale smooth coefficient $s_{0,0}$ intact, i.e. no thresholding is performed on it.

3. Invert the Haar decomposition in the usual way to obtain an estimate of $\sigma^2(t/T)$ at time points $t = 0, \ldots, T-1$. Call the resulting estimate $\hat{\sigma}^2(t/T)$.

Asymptotic Gaussianity and variance stabilization for certain random variables of the form $(U-V)/(U+V)$, where U, V are nonnegative, independent random variables, were studied by Fisz (1955): hence we label $f_{m,n}$ the "Haar-Fisz coefficients", and the division of $d_{m,n}$ by $s_{m,n}$ — the "Fisz transform". The main heuristic idea here is that the variance of $f_{m,n}$ (for most m, n) does not depend on $\sigma^2(z)$. Consider the following example: $\sigma^2_{t,T} = \sigma^2(t/T)$, $m = M-1$, $n = 0$. The Haar-Fisz coefficient $f_{M-1,0}$ has the form:

$$f_{M-1,0} = \frac{Y_0^2 - Y_1^2}{Y_0^2 + Y_1^2} = \frac{\sigma^2(0/T)Z_0^2 - \sigma^2(1/T)Z_1^2}{\sigma^2(0/T)Z_0^2 + \sigma^2(1/T)Z_1^2}.$$

Suppose now that $\sigma^2(0/T) = \sigma^2(1/T)$ (this is likely as $\sigma^2(z)$ is piecewise constant). We then have $f_{M-1,0} = (Z_0^2 - Z_1^2)/(Z_0^2 + Z_1^2)$, and the variance of $f_{M-1,0}$ does not depend on $\sigma^2(z)$. Thus, the thresholds t_m in (6) also do not need to depend on $\sigma^2(z)$, and can therefore be selected more easily.
The following theorem demonstrates the mean-square consistency of $\hat{\sigma}^2(z)$.

Theorem 3.1 *Suppose that $Y_{t,T}$ follows model (5). Construct the estimator $\hat{\sigma}^2(z)$ as follows: fix $\delta \in (0,1)$. For each $T = 2^M$, define the set*

$$\mathcal{I}_T = \{(m,n) : m < M^*\},$$

with $2^{M^*} = O\left(T^{1-\delta}\right)$. Define

$$\hat{\mu}_{m,n} = \begin{cases} s_{m,n}\,\text{sgn}(f_{m,n})\,(|f_{m,n}| - t_m)_+ & (m,n) \in \mathcal{I}_T, \\ 0 & (m,n) \notin \mathcal{I}_T, \end{cases}$$

where

$$t_m = c\,2^{-\frac{M-m-1}{2}}\sqrt{2\log(T)}. \quad (7)$$

Define $a_T = T^{-1}\sum_{t=0}^{T-1}|\sigma^2_{t,T} - \sigma^2(t/T)|^2$ *and assume* $a_T = o_T(\log^{-1} T)$. *We have*

$$T^{-1}\sum_{t=0}^{T-1}\mathbb{E}\left\{\hat{\sigma}^2(t/T) - \sigma^2(t/T)\right\}^2 =$$

$$O\left(a_T c^2 \log T + \sup_z \sigma^4(z)\left\{T^{-\frac{c^2(1-\delta)}{2\rho^2_\infty}}\log T + BT^{-1}c^2\log^2 T\rho^2_\infty + BT^{\delta-1}\right\}\right).$$

Note that we only use non-trivial estimators of $\mu_{m,n}$ in the set $\mathcal{I}_T$, which includes an increasing number of coarsest scales $m \in \{0, \dots, M^* - 1\}$ and excludes an increasing number of finest scales $m \in \{M^*, \dots, M-1\}$. This is done to guarantee the uniform convergence of the thresholds t_m to zero at a certain rate, which occurs over the set $\mathcal{I}_T$, but not over $\mathcal{I}_T^c$. The uniform convergence $t_m \to 0$ is essential for proving the consistency of our estimator: see the proof of Theorem 3.1 and, in particular, the discussion underneath formula (14).

It is impossible to set $\delta = 0$ as then $\mathcal{I}_T$ would include "too many" finest scales m and thus no uniform convergence $t_m \to 0$ at a desired rate would occur. Thus, in view of the term $T^{1-\delta}$ in the mean-square error rate above, δ should be chosen "as small as possible" but positive.

4 Haar-Fisz estimation of the wavelet spectrum

In this section, we apply the result of Section 3 to study consistency properties of our estimator of the evolutionary wavelet spectrum. Because the true $S_l(z)$ can be expressed as $S_l(z) = \sum_j \beta_j(z)(A^{-1})_{l,j}$ (see formula 3), we estimate $S_l(z)$ as $\hat{S}_l(z) = \sum_j \hat{\beta}_j(z)(A^{-1})_{l,j}$, where $\hat{\beta}_j(z)$ are estimates of $\beta_j(z)$ obtained by applying the algorithm of Section 3 to $I^{(j)}_{t,T}$. See Theorem 4.1 for the exact range of j in the above sum. Consistency of $\hat{S}_l$ will follow from the consistency of $\hat{\beta}_j$, the latter being implied by Theorem 3.1.

Note that in this section, we use wavelets at three stages of our modelling+estimation procedure. First, a wavelet family ψ is used to construct the LSW process $X_{t,T}$. Then, the same family ψ is used to compute the wavelet periodogram $I^{(j)}_{t,T}$. Finally, each periodogram sequence is smoothed using Haar wavelets (and the Fisz transform) to obtain the estimates $\hat{\beta}_j(z)$. For the results of this section to hold, two further assumptions are required.

Assumption 4.1 *The constants* L_j *(see Definition 2.1) satisfy* $L_j = O(a^j)$*, for some* $a > 2$.

This is a technical assumption which guarantees that the rate of approximation (in the l_2 norm) of β_j by $I^{(j)}_{t,T}$ is as specified in Proposition 2.1.

Let $e^{(j)}_{t,T}$ denote the wavelet coefficients of the process $X_{t,T}$ at scale j, ie. $e^{(j)}_{t,T} =$

$\sum_s X_{s,T}\psi_{j,s-t}$ (note that $I_{t,T}^{(j)} = (e_{t,T}^{(j)})^2$). Define $\rho_j(\tau) := \sup_{t,T} |\text{cor}(e_{t,T}, e_{t+\tau,T})|$. (Observe that since $\rho_j \leq 1$, we obviously have $\sum_\tau \rho_j^2(\tau) \leq \sum_\tau \rho_j(\tau)$.)

Assumption 4.2 *Functions $\rho_j(\tau)$ satisfy $\sum_\tau \rho_j(\tau) \leq C_3 2^{-j}$.*

The above rate is typical for short-memory LSW processes. As a heuristic motivation, consider a stationary process $X_{t,T}$ with an absolutely summable correlation function $c(l)$. Simple algebra reveals that $\sum_\tau \rho_j(\tau) \leq \sum_l |c(l)| \sum_l |\Psi_j(l)|$, which is in line with the above assumption due to the fact that $\sum_l |\Psi_j(l)| = O(2^{-j})$ for compactly supported wavelets.

Our estimation algorithm proceeds as follows: for each $j = -1, \ldots, -J^*$ (J^* will be specified later), we estimate β_j by applying the estimation algorithm of Section 3 with $Y_{t,T}^2 = I_{t,T}^{(j)}$, $\sigma_{t,T}^2 = \mathbb{E} I_{t,T}^{(j)}$ and $\sigma^2(t/T) = \beta_j(t/T)$. For technical reasons, at each scale j, we use the same parameter $\delta \in (0,1)$ (see Theorem 3.1) and thresholds t_m of the form $\gamma 2^{-j/2} 2^{-\frac{M-m-1}{2}} \sqrt{2 \log T}$ with $\gamma > 0$ (in other words, at each scale j, we set $c = \gamma 2^{-j/2}$). At the remaining scales $j < -J^*$, we simply estimate β_j by zero. Let $\hat{\beta}_j$ denote an estimate obtained in this way. The following proposition prepares the ground for our result on the consistency of $\hat{S}_j$.

Proposition 4.1 *Let J^* be such that $J^* \leq \varepsilon \log_2 T$ for $T \geq T_0$, where $\varepsilon < \delta$. If Assumptions 2.1, 2.2, 2.3, 4.1 and 4.2 hold, then*

$$\frac{2^{2j}}{T} \sum_{t=0}^{T-1} \mathbb{E}(\hat{\beta}_j(t/T) - \beta_j(t/T))^2 = O(T^{-\theta(a,\gamma,C_3,\delta)} \log^2(T))$$

uniformly over $j = -1, \ldots, -J^*$, where $\theta(a,\gamma,C_3,\delta) = \min\left(1 - \frac{1}{2\log_2 a - 1}, \frac{\gamma^2(1-\delta)}{2C_3}, 1 - \delta\right)$.

We are now in a position to demonstrate the consistency of our estimator of the evolutionary wavelet spectrum.

Theorem 4.1 *Let $\hat{S}_l(t/T) = \sum_{j=-J^*}^{-1} \hat{\beta}_j(t/T)(A^{-1})_{l,j}$, where for large T, $J^* = \lfloor \varepsilon \log_2 T \rfloor$ with $\varepsilon = \theta(a,\gamma,C_3,\delta)/3$ and $\theta(a,\gamma,C_3,\delta) < 3\delta$. Further, let the wavelets ψ in Definition 2.1 be Haar wavelets. If Assumptions 2.1, 2.2, 2.3, 4.1 and 4.2 hold, then*

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}(\hat{S}_l(t/T) - S_l(t/T))^2 = O(2^l T^{-\frac{2}{3}\theta(a,\gamma,C_3,\delta)} \log^3(T)).$$

A few remarks are in order.

1. *Other wavelets.* Reasoning as in Nason et al. (2000), proof of Theorem 2.15, we conjecture that the same or faster rate of convergence as in Theorem 4.1 is attained if ψ are compactly supported Daubechies' wavelets other than Haar.

2. *Discussion of the convergence rate.* Theorem 4.1 gives the optimal rate of convergence, which in practice cannot be attained as a and C_3 are unknown so it is not

possible to set $J^* = \lfloor \frac{1}{3}\theta(a, \gamma, C_3, \delta) \log_2 T \rfloor$ (for T large enough). However, it is clear from the proof that our estimator is also consistent (albeit with a slightly slower rate) for $J^* = \lfloor \varepsilon \log_2 T \rfloor$ with any $\varepsilon < \delta$, provided that $\delta < \theta(a, \gamma, C_3, \delta)$.

3. *Choice of γ and δ.* Theorem 4.1 provides some clues as to the choice of γ and δ although they are of a mainly theoretical nature. Consider the requirement that

$$\theta(a, \gamma, C_3, \delta) < 3\delta. \tag{8}$$

If γ was chosen to be "large enough", then condition (8) would simply become $\min\left(1 - \frac{1}{2\log_2 a - 1}, 1 - \delta\right) < 3\delta$. Suppose that a (which is a characteristic of the process and cannot be manipulated by the user) was such that $1 - \frac{1}{2\log_2 a - 1} \geq 1 - \delta$. This would further simplify (8) to $1 - \delta < 3\delta$ or $\delta > 1/4$. On the other hand, recalling the form of θ, the smaller the value of δ, the higher the rate of convergence, which means that in this particular example, δ should be chosen "close" to $1/4$.

In the next section, we propose a practical procedure for selecting J^*, δ (or alternatively M^*) and γ (or alternatively c) from the data.

5 Choice of parameters and a simulated example

The results of the previous section offer some insight into how the parameters J^*, δ (or alternatively M^*) and γ (or alternatively c) should be selected in the asymptotic limit $T \to \infty$. This section complements that theory by describing a practical well-performing procedure for selecting the parameter values in finite samples.

Given each periodogram sequence $I_{t,T}^{(j)}$ for $j = -1, \ldots, -J^*$, we select the constant c in formula (7) using the following observation: note that $I_{t,T}^{(j)}/\beta_j(t/T)$ is (approximately) a sequence of correlated χ_1^2 variables. Therefore, we would expect var$(I_{t,T}^{(j)}/\beta_j(t/T))$ to be close to 2. Thus, we examine a grid of equispaced values of c, and choose the one for which the variance of the empirical residuals $I_{t,T}^{(j)}/\hat{\beta}_j(t/T)$ is the closest to 2. It is easily seen that if $c = 0$, then $\hat{\beta}_j(t/T) = I_{t,T}^{(j)}$ and var$(I_{t,T}^{(j)}/\hat{\beta}_j(t/T)) = 0$. On the other hand, if $c = \infty$, then $\hat{\beta}_j(t/T) =$ sample mean$(I_{t,T}^{(j)})$, and then var$(I_{t,T}^{(j)}/\hat{\beta}_j(t/T)) > 2$ unless $\beta_j(z)$ is constant with respect to z. Empirically, we have found that choosing the (intermediate) value of c which corresponds to var$(I_{t,T}^{(j)}/\hat{\beta}_j(t/T)) \sim 2$ is a reliable way of ensuring the "right" amount of smoothness in $\hat{\beta}_j$. In the example below, c is selected over the grid $1/20, 2/20, \ldots, 1$. The procedure is still fast as our estimation algorithm is of linear computational complexity.

We now describe the choice of M^*. From the definition of $\hat{\mu}_{m,n}$ in Theorem 3.1 it is obvious that $\hat{\mu}_{m,n} = 0$ if $t_m > 1$ (as $|f_{m,n}|$ is always bounded by 1). On the other hand, $\hat{\mu}_{m,n} = 0$ if $m \geq M^*$. Thus, in practice, it is natural to set M^* to be the lowest integer for which $t_m > 1$.

To enable a fair comparison with Nason et al. (2000), we follow them in taking $J^* = \log_2 T$. Please note that this is not in conflict with the theory, which assumes that $J^* \leq \varepsilon \log_2 T$, but only for large T.

We have found the procedure described above to have a very good finite sample performance. To illustrate this, consider again the concatenated Haar process. Let X_t denote its sample path shown in the top right plot of Figure 1. Figure 2 shows reconstructions of the Haar spectrum of X_t obtained using: our method [HF]; the translation-invariant version of our method for comparison with Nason et al. (2000) [HF-TI]; and the translation-invariant method of Nason et al. (2000) based on the log transform, also with Haar wavelets, with a manual choice of the "primary resolution" level (see the paper for details) to guarantee optimal performance [NvsK]. While there is still a slight amount of leakage, HF does remarkably well. The reconstruction HF-TI, although not piecewise constant, has an appearance which is "closer" to piecewise constant than the bumpy reconstruction NvSK. The bumps is NvSK are due to the fact that any, even small, oscillations in the smoothed log-spectrum are significantly magnified after exponentiating the estimate.

6 The Haar-Fisz transform

In this section, we describe a multiscale variance-stabilizing transform which implicitly arises in our generic Haar-Fisz estimation procedure of Section 3. Note that it is possible to decompose our estimation algorithm of Section 3 into the following three steps:

1. *The Haar-Fisz transform.*
   (a) Let $M = \log_2 T$ and let $s_{M,n} := Y_n^2$, $n = 0, \ldots, 2^M - 1$.
   (b) For each $m = M-1, M-2, \ldots, 0$, recursively form vectors $\mathbf{s}_m$, $\mathbf{d}_m$ and $\mathbf{f}_m$ with elements:
   $$\begin{aligned} s_{m,n} &= \frac{s_{m+1,2n} + s_{m+1,2n+1}}{2} \\ f_{m,n} &= \frac{s_{m+1,2n} - s_{m+1,2n+1}}{2 s_{m,n}}, \end{aligned} \quad (9)$$
   where $n = 0, \ldots, 2^m - 1$.
   (c) For each $m = 0, 1, \ldots, M-1$, recursively modify the vectors $\mathbf{s}_{m+1}$:
   $$\begin{aligned} s_{m+1,2n} &= s_{m,n} + f_{m,n} \\ s_{m+1,2n+1} &= s_{m,n} - f_{m,n}, \end{aligned}$$
   where $n = 0, \ldots, 2^m - 1$.
   (d) Let $U_n := s_{M,n}$, $n = 0, \ldots, 2^M - 1$.

   We denote $U_n = \mathcal{F} Y_n^2$. The nonlinear, invertible operator $\mathcal{F}$ is called the *Haar-Fisz operator*, and its action — the *Haar-Fisz transform*.

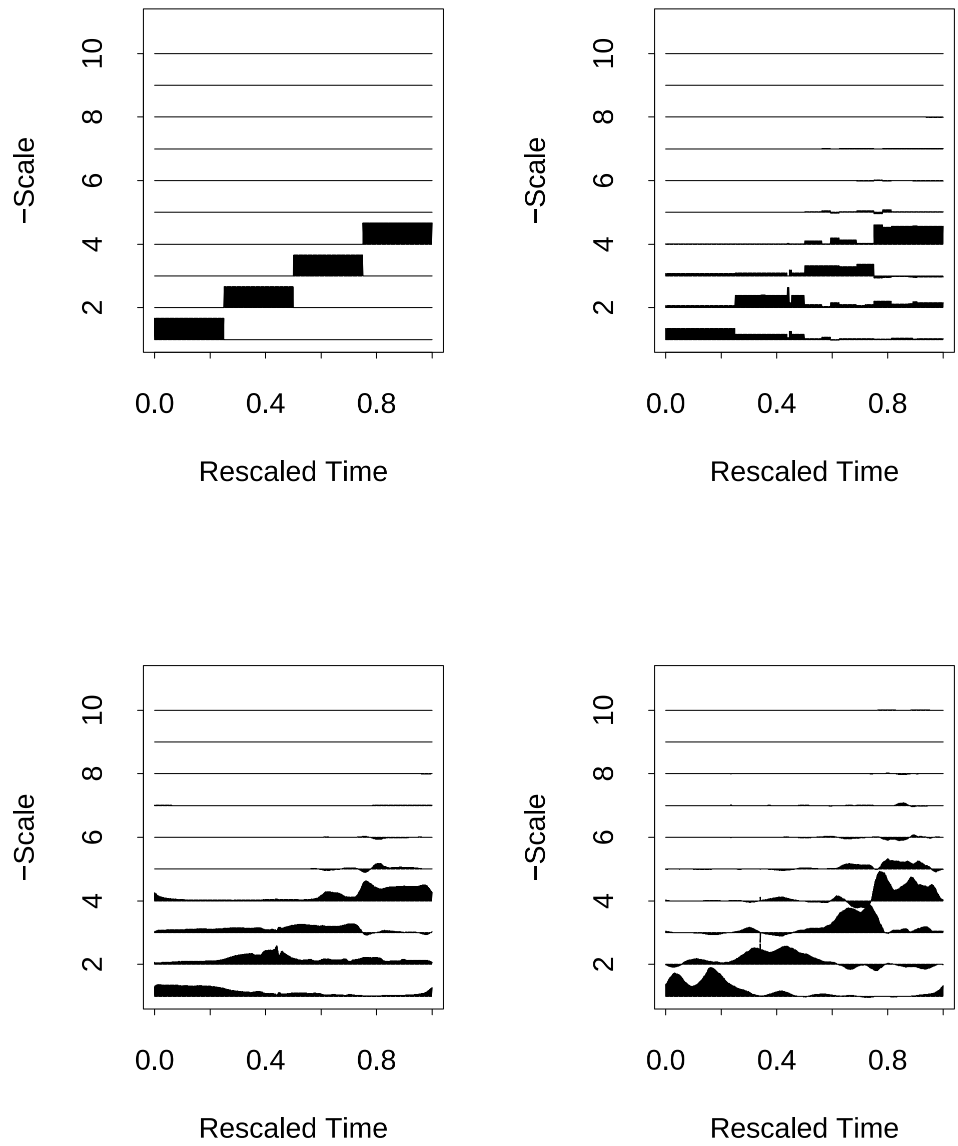

Figure 2: Top left: spectrum of the concatenated Haar process. Reconstruction using: our method (top right); translation-invariant version of our method (bottom left); the method of Nason et al. (2000) (bottom right).

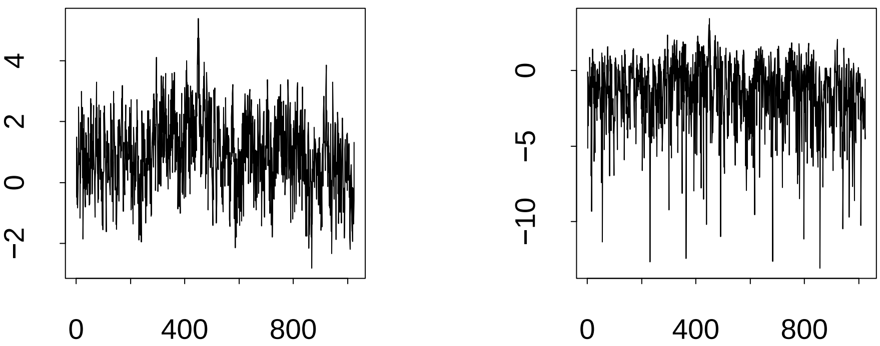

Figure 3: Left: the Haar-Fisz transform of the wavelet periodogram sequence from the bottom right plot of Figure 1; right: its log transform.

2. Smoothing of U_n using a classical wavelet thresholding estimator suitable for homoscedastic Gaussian data, based on Haar wavelets and soft thresholding with a universal-type threshold, see e.g. Donoho and Johnstone (1994). As the focus of this section is on the Haar-Fisz transform (point 1 above), we deliberately avoid giving more details here.

3. The inverse Haar-Fisz transform.

As argued in Section 3, if $\sigma^2_{t,T} = \sigma^2(t/T)$ and $\sigma^2(z)$ is piecewise constant, the division by $s_{m,n}$ in (9) stabilizes the variance of $f_{m,n}$ for "most" m, n. Thus, $\mathcal{F}$ stabilizes the variance of Y_n^2 (this property will be formalised later), but unlike the log transform, operates in the wavelet (Haar) domain instead of the time domain. Figure 3 compares the Haar-Fisz and log transforms of the wavelet periodogram sequence at scale -2 of the concatenated Haar process. Clearly, the noise in the Haar-Fisz-transformed sequence appears more symmetric and closer to homoscedastic Gaussian. Also, the shape of the underlying "signal" seems to be brought out more clearly.

For the reader's convenience, we now write out the explicit formula for $U = \mathcal{F}Y^2$ in the case $2^M = 8$. For a vector v, let $\overline{v}$ denote its sample mean. We have

$$
\begin{aligned}
U_0 &= \overline{Y^2} + \frac{\sum_{n=0}^3 Y_n^2 - \sum_{n=4}^7 Y_n^2}{\sum_{n=0}^7 Y_n^2} + \frac{Y_0^2 + Y_1^2 - Y_2^2 - Y_3^2}{\sum_{n=0}^3 Y_n^2} + \frac{Y_0^2 - Y_1^2}{Y_0^2 + Y_1^2} \\
U_1 &= \overline{Y^2} + \frac{\sum_{n=0}^3 Y_n^2 - \sum_{n=4}^7 Y_n^2}{\sum_{n=0}^7 Y_n^2} + \frac{Y_0^2 + Y_1^2 - Y_2^2 - Y_3^2}{\sum_{n=0}^3 Y_n^2} - \frac{Y_0^2 - Y_1^2}{Y_0^2 + Y_1^2} \\
U_2 &= \overline{Y^2} + \frac{\sum_{n=0}^3 Y_n^2 - \sum_{n=4}^7 Y_n^2}{\sum_{n=0}^7 Y_n^2} - \frac{Y_0^2 + Y_1^2 - Y_2^2 - Y_3^2}{\sum_{n=0}^3 Y_n^2} + \frac{Y_2^2 - Y_3^2}{Y_2^2 + Y_3^2} \\
U_3 &= \overline{Y^2} + \frac{\sum_{n=0}^3 Y_n^2 - \sum_{n=4}^7 Y_n^2}{\sum_{n=0}^7 Y_n^2} - \frac{Y_0^2 + Y_1^2 - Y_2^2 - Y_3^2}{\sum_{n=0}^3 Y_n^2} - \frac{Y_2^2 - Y_3^2}{Y_2^2 + Y_3^2} \\
U_4 &= \overline{Y^2} - \frac{\sum_{n=0}^3 Y_n^2 - \sum_{n=4}^7 Y_n^2}{\sum_{n=0}^7 Y_n^2} + \frac{Y_4^2 + Y_5^2 - Y_6^2 - Y_7^2}{\sum_{n=4}^7 Y_n^2} + \frac{Y_4^2 - Y_5^2}{Y_4^2 + Y_5^2} \\
U_5 &= \overline{Y^2} - \frac{\sum_{n=0}^3 Y_n^2 - \sum_{n=4}^7 Y_n^2}{\sum_{n=0}^7 Y_n^2} + \frac{Y_4^2 + Y_5^2 - Y_6^2 - Y_7^2}{\sum_{n=4}^7 Y_n^2} - \frac{Y_4^2 - Y_5^2}{Y_4^2 + Y_5^2} \\
U_6 &= \overline{Y^2} - \frac{\sum_{n=0}^3 Y_n^2 - \sum_{n=4}^7 Y_n^2}{\sum_{n=0}^7 Y_n^2} - \frac{Y_4^2 + Y_5^2 - Y_6^2 - Y_7^2}{\sum_{n=4}^7 Y_n^2} + \frac{Y_6^2 - Y_7^2}{Y_6^2 + Y_7^2} \\
U_7 &= \overline{Y^2} - \frac{\sum_{n=0}^3 Y_n^2 - \sum_{n=4}^7 Y_n^2}{\sum_{n=0}^7 Y_n^2} - \frac{Y_4^2 + Y_5^2 - Y_6^2 - Y_7^2}{\sum_{n=4}^7 Y_n^2} - \frac{Y_6^2 - Y_7^2}{Y_6^2 + Y_7^2}.
\end{aligned}
$$

Note that $\overline{U} = \overline{Y^2}$. We also define a "truncated" Haar-Fisz transform $U_n^{(M^*)} = \mathcal{F}^{(M^*)}Y_n^2 := s_{M^*,n}$, where $M^* < M$. As an example, we quote the formula for $U_n^{(M^*)}$ in the case $n = 0$, $M^* = 2$ (note that the length of the vector $U^{(M^*)}$ is $2^{M^*} = 4$).

$$
U_0^{(2)} = \overline{Y^2} + \frac{\sum_{n=0}^3 Y_n^2 - \sum_{n=4}^7 Y_n^2}{\sum_{n=0}^7 Y_n^2} + \frac{Y_0^2 + Y_1^2 - Y_2^2 - Y_3^2}{\sum_{n=0}^3 Y_n^2}.
$$

We now formalise the variance-stabilizing and "Gaussianising" properties of the Haar-Fisz transform, albeit in a simplified setup. Consider the following model:

$$
Y_{t,T}^2 = \sigma^2(t/T) Z_{t,T}^2, \quad t = 0, \ldots, T-1, \tag{10}
$$

where

- $\sigma^2(z)$ is a piecewise constant function, bounded from above and away from zero, with a finite but unknown number of jumps (the number of jumps is denoted by B);
- the variables $\{Z_{t,T}\}_{t=0}^{T-1}$ are i.i.d. normal with mean zero and variance one.

The above model is a special case of (5). For the remainder of this section, let $f_{m,n}^v$ denote the Haar-Fisz coefficient of v at scale m, location n. The following proposition holds.

Proposition 6.1 *Let $Y_{t,T}^2$ follow model (10). The following statements are true:*

1. *(Beta distribution.) $f_{m,n}^{Z^2}$ is distributed as $2\beta(2^{M-m-2}, 2^{M-m-2}) - 1$ and thus $(2^{M-m-1}+1)^{-1/2} f_{m,n}^{Z^2} \to N(0,1)$ as $M \to \infty$, $m \le (1-\delta)M$, $\delta \in (0,1)$.*

2. *(Log-like property of $\mathcal{F}$.) We have*

$$T^{-1} \sum_{t=0}^{T-1} \mathbb{E}\{(\mathcal{F}Y_{t,T}^2 - \overline{Y^2}) - (\mathcal{F}\sigma^2(t/T) - \overline{\sigma^2}) - (\mathcal{F}Z_{t,T}^2 - \overline{Z^2})\}^2 = O(T^{-1}\log_2 T). \tag{11}$$

3. *(Variance stabilization.) We have*

$$\mathit{var}(\mathcal{F}Z_{t,T}^2) = \sum_{l=0}^{M-1}(2^l+1)^{-1} + 2^{1-M} \to \sum_{l=0}^{\infty}(2^l+1)^{-1}$$

   for all t.

4. *(Asymptotic normality.) For any t,*

$$\left(\sum_{l=M-M^*}^{M-1}(2^l+1)^{-1} + 2^{1-M}\right)^{-1/2} \mathcal{F}^{(M^*)} Z_{t,T}^2 \xrightarrow{d} N(0,1),$$

   as $M \to \infty$, $M^* = \lfloor (1-\delta)M \rfloor$, $\delta \in (0,1)$.

5. *(Lack of spurious correlation.) For any t_1, t_2,*

$$\left(\sum_{l=M-M^*}^{M-1}(2^l+1)^{-1} + 2^{1-M}\right)^{-1} \mathit{cov}(\mathcal{F}^{(M^*)} Z_{t_1,T}^2, \mathcal{F}^{(M^*)} Z_{t_2,T}^2) \to 0,$$

   as $M \to \infty$, $M^* = \lfloor (1-\delta)M \rfloor$, $\delta \in (0,1)$.

Note that formula (11) can be interpreted as

$$\mathcal{F}Y_{t,T}^2 - \overline{Y^2} \approx (\mathcal{F}\sigma^2(t/T) - \overline{\sigma^2}) + (\mathcal{F}Z_{t,T}^2 - \overline{Z^2}),$$

which is reminiscent of the log property $\log Y_{t,T}^2 = \log \sigma^2(t/T) + \log Z_{t,T}^2$ (except that we need to subtract the means). This means that the Haar-Fisz transform "approximately" transforms the model from multiplicative to additive: $\mathcal{F}\sigma^2(t/T)$ can be viewed as "signal" and $\mathcal{F}Z_{t,T}^2$ as "noise". The variance stabilization property 3. means that the variance of the "noise" $\mathcal{F}Z_{t,T}^2$ is constant over time. The asymptotic normality property 4. means that as long as we stay away from the finest scales, the "noise" is asymptotically standard normal. The lack of spurious correlation property 5. means that the Haar-Fisz transform asymptotically does not induce any spurious correlation in the "noise" $\mathcal{F}Z_{t,T}^2$ (note that the original noise $Z_{t,T}^2$ was uncorrelated).

We also observe that we do need to stay away from the finest scales (i.e. consider the "truncated" Haar-Fisz transform) to obtain asymptotic normality. Although $\mathcal{F}Z_{t,T}^2$ is a symmetric random variable and, from empirical observation, "close" to Gaussian (certainly closer than $\log Z_{t,T}^2$), it is not true that $\mathcal{F}Z_{t,T}^2 \to$ Gaussian as $M \to \infty$. To see this, note that by the "beta distribution" property 1., $\mathcal{F}Z_{t,T}^2$ is, asymptotically, an infinite sum of random variables whose variances decay approximately geometrically. Thus the form of the distribution of $\mathcal{F}Z_{t,T}^2$ is decided by the summands with the largest variances, which are far from Gaussian (e.g. the finest scale summand, $f_{M-1,n}$, which has a largest variance, has a bimodal distribution).

Extension to correlated innovations $Z_{t,T}^2$. We conjecture that the log-like "separation" formula (11) also holds in the case when $Z_{t,T}^2$ are correlated. As can be seen from the proof of property 2., the key to proving this conjecture is the availability of bounds on moments of ratios of quadratic forms in non-iid normal variables. Exact formulae for them are available, see for example Jones (1987) (note that Ghazal (1994) only treats the iid case), but they involve special functions and are notably complicated so we do not pursue it further here. Once (11) has been established, we conjecture that asymptotic normality of $\mathcal{F}^{(M^*)}Z_{t,T}^2$ can be demonstrated using CLT-type arguments as asymptotically, $\mathcal{F}^{(M^*)}Z_{t,T}^2$ is an infinite sum of random variables, each of which converges to Gaussianity. Obviously, if the correlation structure of $Z_{t,T}^2$ changes over time, then we cannot expect the variance of $\mathcal{F}Z_{t,T}^2$ to be stabilized exactly. However, ample empirical evidence suggests that often the degree of variance stabilization in $\mathcal{F}Z_{t,T}^2$ is remarkable, even in the correlated time-varying case (see eg Figure 3).

We close with a few final remarks:

1. From the computational point of view, the smoothing of the Haar-Fisz transformed vector U_n can be carried out using any smoothing technique suitable for homoscedastic Gaussian data.

2. The representation of our estimator of Section 3 as "the variance-stabilizing Haar-Fisz transform + Gaussian smoothing + the inverse Haar-Fisz transform" follows the same pattern as "the variance-stabilizing log transform + removing homoscedastic noise + exponentiating the estimate". However, note that our estimator is asymptotically unbiased (as it is mean-square consistent by Theorem 3.1). This is in contrast to log-based estimators which require a bias correction involving the Euler-Mascheroni constant.

3. The software implementing our estimators and the Haar-Fisz transform is available on request from the first author.

Acknowledgements

PF is grateful to the Department of Mathematics, Imperial College London, UK (where most of the work for the current version of this paper was done) for the excellent research environment. GPN was partially supported by an EPSRC Advanced Research Fellowship. We thank the Joint Editor, the Associate Editor and two Referees for their helpful comments and suggestions.

A Proofs

Proof of Proposition 2.1

$$\mathbb{E}\left|\sum_s X_{s,T}\psi_{j,s-t}\right|^2 = \mathbb{E}\left|\sum_s \sum_{i=-\infty}^{-1}\sum_{k=-\infty}^{\infty} W_i(k/T)\psi_{i,s-k}\xi_{i,k}\psi_{j,s-t}\right|^2 =$$

$$\mathbb{E}\left|\sum_{i=-\infty}^{-1}\sum_{k=-\infty}^{\infty} W_i(k/T)\Psi_{i,j}(k-t)\xi_{i,k}\right|^2 = \sum_{i=-\infty}^{-1}\sum_{k=-\infty}^{\infty} S_i(k/T)\Psi_{i,j}^2(k-t),$$

using the orthonormality of $\xi_{i,k}$. Also, noting that $A_{i,j} = \sum_\tau \Psi_{i,j}^2(\tau)$,

$$T^{-1}\sum_{t=0}^{T-1}\left|\sum_{i=-\infty}^{-1}\left(\sum_{k=-\infty}^{\infty} S_i(k/T)\Psi_{i,j}^2(k-t) - S_i(t/T)A_{i,j}\right)\right|^2 =$$

$$T^{-1}\sum_{t=0}^{T-1}\left|\sum_{i=-\infty}^{-1}\sum_{k=-\infty}^{\infty}\left(S_i(k/T) - S_i(t/T)\right)\Psi_{i,j}^2(k-t)\right|^2 \leq$$

$$2T^{-1}\sum_{t=0}^{T-1}\left|\sum_{i=-I_T}^{-1}\sum_{k=-\infty}^{\infty}\left(S_i(k/T) - S_i(t/T)\right)\Psi_{i,j}^2(k-t)\right|^2 +$$

$$2T^{-1}\sum_{t=0}^{T-1}\left|\sum_{i=-\infty}^{-I_T}\sum_{k=-\infty}^{\infty}\left(S_i(k/T) - S_i(t/T)\right)\Psi_{i,j}^2(k-t)\right|^2 =: I + II,$$

where the cut-off index $I_T < \log_2 T$ will be specified later. We first consider I. For all $-1 \leq i \leq -I_T$, the length of the support of $\Psi_{i,j}(\cdot)$ can (comfortably) be bounded by $M(2^{-j} + 2^{I_T})$, where M is a constant. Thus, for all t which are not within the distance of $M(2^{-j} + 2^{I_T})$ or less from a breakpoint, we have $|\sum_{i=-I_T}^{-1}\sum_{k=-\infty}^{\infty}(S_i(k/T) -$

$S_i(t/T))\Psi^2_{i,j}(k-t)|^2 = 0$. It follows that we can bound I by

$$4T^{-1}BM(2^{-j}+2^{I_T})\left|\sum_{i=-I_T}^{-1}\sum_{k=-\infty}^{\infty} L_i\Psi^2_{i,j}(k)\right|^2 =$$

$$4T^{-1}BM(2^{-j}+2^{I_T})\left|\sum_{i=-I_T}^{-1} L_i A_{i,j}\right|^2 \leq$$

$$4T^{-1}BM(2^{-j}+2^{I_T})\left|\sum_{i=-I_T}^{-1} L_i 2^{-i}\sum_l 2^l A_{l,j}\right|^2 = O(T^{-1}(2^{-j}+2^{I_T})),$$

using the property $\sum_l 2^l A_{l,j} = 1$ from Fryzlewicz et al. (2003). Similarly, we bound II by

$$2T^{-1}\sum_{t=0}^{T-1}\left|\sum_{i=-\infty}^{-I_T} L_i A_{i,j}\right|^2 \leq 2\left|\sum_{i=-\infty}^{-I_T} L_i 2^{-i}\right|^2.$$

Setting I_T such that $\lim_T I_T = \infty$ and $I_T = o(\log_2 T)$, we get $I + II = o_T(1)$. Further, if $L_i = O(a^i)$ for $a > 2$, then $II = O((2/a)^{2I_T})$. Equating the rates for I and II, $T^{-1}2^{I_T} = (2/a)^{2I_T}$, we obtain the optimal $I_T = \log_2 T/(2\log_2 a - 1)$, which leads to the rate of convergence of $T^{\frac{1}{2\log_2 a - 1}-1}$. □

Proof of Proposition 2.2. Using the same technique as in the proof of Proposition 2.1, we have

$$\beta_j(z) = \sum_i S_i(z)A_{i,j} \leq \sum_i S_i(z)2^{-j},$$

which is uniformly bounded by a multiple of 2^{-j} by Definition 2.1. If Assumption 2.1 holds, then the jumps of any of the functions $\{S_j(z)\}_j$ are in $\mathcal{B}$. It then follows that each $\beta_j(z)$ must be piecewise constant, and its jumps must also be contained in $\mathcal{B}$. Finally, if Assumption 2.2 holds, then

$$\beta_j(z) = \sum_i S_i(z)A_{i,j} \leq C_1\sum_i 2^i A_{i,j} = C_1,$$

again using Fryzlewicz et al. (2003). □

Proof of Theorem 3.1. We introduce the auxiliary process (which in fact we do not observe)

$$\tilde{Y}^2_{t,T} = \sigma^2(t/T)Z^2_{t,T}.$$

Let $\tilde{\hat{\sigma}}^2(t/T)$ denote the soft thresholding estimate of $\sigma^2(t/T)$ constructed from $\tilde{Y}^2_{t,T}$, and let $\tilde{s}_{m,n}$, $\tilde{d}_{m,n}$, $\tilde{f}_{m,n}$ and $\tilde{\hat{\mu}}_{m,n}$ denote quantities constructed from $\tilde{Y}^2_{t,T}$ in the same way as the quantities $s_{m,n}$, $d_{m,n}$, $f_{m,n}$ and $\hat{\mu}_{m,n}$ (respectively) constructed from $Y^2_{t,T}$. We have

$$
\begin{aligned}
T^{-1}\sum_{t=0}^{T-1}\mathbb{E}\left\{\hat{\sigma}^2(t/T)-\sigma^2(t/T)\right\}^2 \quad &\leq \quad 2\,T^{-1}\sum_{t=0}^{T-1}\mathbb{E}\left\{\hat{\sigma}^2(t/T)-\tilde{\hat{\sigma}}^2(t/T)\right\}^2 \\
&+ \quad 2\,T^{-1}\sum_{t=0}^{T-1}\mathbb{E}\left\{\tilde{\hat{\sigma}}^2(t/T)-\sigma^2(t/T)\right\}^2 =: I+II.
\end{aligned}
$$

We first concentrate on I. Using Parseval identity and some elementary properties of the thresholding function $h(x)=\text{sgn}(x)(|x|-t)_+$, we obtain

$$
\begin{aligned}
I &= \frac{2}{T}\mathbb{E}\{s_{0,0}-\tilde{s}_{0,0}\}^2+\frac{2}{T}\sum_{m=0}^{M^*-1}\sum_{n=0}^{2^m-1}\mathbb{E}\{\hat{\mu}_{m,n}-\tilde{\hat{\mu}}_{m,n}\}^2 \leq \frac{2}{T}\mathbb{E}\{s_{0,0}-\tilde{s}_{0,0}\}^2 \\
&\quad +\frac{2}{T}\sum_{m=0}^{M-1}\sum_{n=0}^{2^m-1}\mathbb{E}\{\text{sgn}(d_{m,n})(|d_{m,n}|-s_{m,n}t_m)_+-\text{sgn}(\tilde{d}_{m,n})(|\tilde{d}_{m,n}|-\tilde{s}_{m,n}t_m)_+\}^2 \\
&\quad \leq \frac{2}{T}\mathbb{E}\{s_{0,0}-\tilde{s}_{0,0}\}^2+\frac{4}{T}\sum_{m=0}^{M-1}\sum_{n=0}^{2^m-1}\mathbb{E}\{d_{m,n}-\tilde{d}_{m,n}\}^2+t_m^2\mathbb{E}\{s_{m,n}-\tilde{s}_{m,n}\}^2 \\
&\quad \leq \frac{4}{T}\sum_{t=0}^{T-1}\mathbb{E}\{(\sigma^2_{t,T}-\sigma^2(t/T))Z^2_{t,T}\}^2+\frac{4}{T}\sum_{m=0}^{M-1}\sum_{n=0}^{2^m-1}t_m^2\mathbb{E}\{s_{m,n}-\tilde{s}_{m,n}\}^2 =: III+IV
\end{aligned}
$$

III is of order a_T. To bound IV, we use the explicit form of t_m, $s_{m,n}$ and $\tilde{s}_{m,n}$ as well as the Cauchy-Schwarz inequality:

$$
\begin{aligned}
IV \quad &= \quad \frac{8c^2\log T}{T}\sum_{m=0}^{M-1}\sum_{n=0}^{2^m-1}2^{-M+m+1}2^{m-M}\mathbb{E}\left\{\sum_{s=2^{M-j}n}^{2^{M-n}(n+1)-1}(\sigma^2_{s,T}-\sigma^2(s/T))Z^2_{s,T}\right\}^2 \\
&\leq \quad \frac{8c^2\log T}{T}\sum_{m=0}^{M-1}\sum_{n=0}^{2^m-1}2^{-M+m+1}\mathbb{E}(Z^4_{\cdot,T})\sum_{s=2^{M-j}n}^{2^{M-n}(n+1)-1}(\sigma^2_{s,T}-\sigma^2(s/T))^2 \\
&= \quad \frac{48\,c^2\log T}{T}\sum_{t=0}^{T-1}\mathbb{E}\{\sigma^2_{t,T}-\sigma^2(t/T)\}^2.
\end{aligned}
$$

Thus IV is of order $c^2a_T\log T$. We now turn to II. First consider the MSE of $\tilde{\hat{\mu}}_{m,n}$ for $m<M^*$. For notational clarity, denote $\tilde{d}_{1,m,n}=\tilde{s}_{m+1,2n}/\sqrt{2}$ and $\tilde{d}_{2,m,n}=\tilde{s}_{m+1,2n+1}/\sqrt{2}$, so that $\tilde{d}_{m,n}=\tilde{d}_{1,m,n}-\tilde{d}_{2,m,n}$ and $\tilde{s}_{m,n}=\tilde{d}_{1,m,n}+\tilde{d}_{2,m,n}$. Denote further $\mu_{i,m,n}=\mathbb{E}\,\tilde{d}_{i,m,n}$ for $i=1,2$ (note that $\mu_{m,n}=\mu_{1,m,n}-\mu_{2,m,n}$). Finally denote $w_{i,m,n}=\text{var}\,\tilde{d}_{i,m,n}$ for $i=1,2$, and $w_{m,n}=\text{var}\,\tilde{d}_{m,n}$. We consider two cases.

1. Case $\sigma^2(t/T)=\text{constant}=:\sigma^2$ for $t=2^{M-m}n,\dots,2^{M-m}(n+1)-1$ (so that $\mu_{1,m,n}=\mu_{2,m,n}$). Without loss of generality, suppose $n=0$ to shorten notation. Using a

simple property of $h(x)$, then Cauchy-Schwarz and then Hölder's inequality, we have

$$\mathbb{E}(\tilde{\tilde{\mu}}_{m,0} - \mu_{m,0})^2 = \mathbb{E}\,\tilde{\tilde{\mu}}_{m,0}^2 = \mathbb{E}\left(\tilde{s}_{m,0}\,\mathrm{sgn}(\tilde{f}_{m,0})(|\tilde{f}_{m,0}| - t_m)_+\right)^2 \leq$$
$$\mathbb{E}\left(\tilde{d}_{m,0}\mathbb{I}(|\tilde{f}_{m,0}| > t_m)\right)^2 =$$
$$2^{m-M}\sigma^4\mathbb{E}\left\{\left(\sum_{i=0}^{2^{M-m-1}-1} Z_{i,T}^2 - Z_{i+2^{M-m-1},T}^2\right)^2 \mathbb{I}(|\tilde{f}_{m,0}| > t_m)\right\} \leq$$
$$\sigma^4 \sum_{i=0}^{2^{M-m}-1} \mathbb{E}\left\{Z_{i,T}^4\mathbb{I}(|\tilde{f}_{m,0}| > t_m)\right\} \leq 2^{M-m}\sigma^4\{\mathbb{E}\,Z_{\cdot,T}^{4r}\}^{1/r}\mathbb{P}(|\tilde{f}_{m,0}| > t_m)^{1-1/r} \quad (12)$$

Note that $C_r := \{\mathbb{E}\,Z_{\cdot,T}^{4r}\}^{1/r}$ is a constant depending only on r. We now consider the probability term. We have

$$\mathbb{P}(|\tilde{f}_{m,0}| > t_m) \leq \mathbb{P}(\tilde{f}_{m,0} > t_m) + \mathbb{P}(\tilde{f}_{m,0} < -t_m) =: V + VI.$$

Without loss of generality we may assume that $V \geq VI$ (the proof is identical if $VI \geq V$ and leads to the same bounds) so we bound the above sum of probabilities by $2V$:

$$\begin{aligned} 2V &= 2\mathbb{P}(\tilde{d}_{1,m,0} - \tilde{d}_{2,m,0} > t_m(\tilde{d}_{1,m,0} + \tilde{d}_{2,m,0})) \\ &= 2\mathbb{P}\left(\sum_{i=0}^{2^{M-m-1}-1} Z_{i,T}^2(1-t_m) - \sum_{i=2^{M-m-1}}^{2^{M-m}-1} Z_{i,T}^2(1+t_m) > 0\right), \end{aligned}$$

which, by standard results (see e.g. Johnson and Kotz (1970), p. 151), can be represented as $2\mathbb{P}(\sum_{i=0}^{2^{M-m}-1}\lambda_i U_i^2 > 0)$, where $\{U_i\}_{i=0}^{2^{M-m}-1}$ are i.i.d. standard normal, and λ_i are the eigenvalues of VD, where V is a $2^{M-m} \times 2^{M-m}$ matrix such that $V_{i,l} = \mathrm{cor}(Z_{i,T}, Z_{l,T})$ and D is a $2^{M-m} \times 2^{M-m}$ diagonal matrix for which

$$D_{i,i} = \begin{cases} 1 - t_m & i \leq 2^{M-m-1} - 1 \\ -1 - t_m & i \geq 2^{M-m-1}. \end{cases}$$

Noting that $\sum_{i=0}^{2^{M-m}-1}\lambda_i = \mathrm{tr}(VD) = -t_m 2^{M-m}$, we rewrite $2V$ as $2\mathbb{P}(\sum_{i=0}^{2^{M-m}-1}\lambda_i(U_i^2 - 1) > t_m 2^{M-m})$. We bound this tail probability using Bernstein's inequality (Bosq (1998), p. 24). The conditions for the latter are satisfied as, by simple properties of normal variables, there exists a C_2 such that, for all $i = 0, \ldots, 2^{M-m} - 1$ and $k \geq 3$, we have $\mathbb{E}|\lambda_i(U_i^2 - 1)|^k \leq (C_2 \max_l |\lambda_l|)^{k-2} k! \mathbb{E}|\lambda_i(U_i^2 - 1)|^2$. Thus, by Bernstein's inequality, we have the bound

$$2V \leq 4\exp\left(-\frac{t_m^2 2^{2M-2m}}{8\sum_{i=0}^{2^{M-m}-1}\lambda_i^2 + 2\max_i|\lambda_i| C_2 t_m 2^{M-m}}\right). \quad (13)$$

It remains to assess $\sum_{i=0}^{2^{M-m}-1}\lambda_i^2$ and $\max_i|\lambda_i|$. By standard results, we have $\sum_{i=0}^{2^{M-m}-1}\lambda_i^2 = \mathrm{tr}(VD)^2$, which by direct verification can be bounded from above by $2^{M-m}(1+t_m^2)\rho_\infty^2$. To bound the largest eigenvalue, let $\|\cdot\|$ denote the spectral norm of a matrix. By standard results, we have $\max_i|\lambda_i| \le (1+t_m)\|V\|$. Using that V is nonnegative definite (being a correlation matrix), it is easily seen that $\|V\| \le \rho_\infty^1$. Using these results and the explicit form of t_m, we can bound (13) from above by

$$4\exp\left(-\frac{2c^2\log T}{4(1+t_m^2)\rho_\infty^2+\rho_\infty^1 C_2 t_m(1+t_m)}\right) = \\ 4T^{-\frac{2c^2}{4(1+t_m^2)\rho_\infty^2+\rho_\infty^1 C_2 t_m(1+t_m)}} = O(T^{-\frac{c^2}{2\rho_\infty^2}}), \qquad (14)$$

where the last equality follows from the fact that $t_m < t_{M^*} = O(T^{-\delta/2}\sqrt{\log T})$ uniformly on $\mathcal{I}_T$. Plugging it into (12), we obtain the final bound for the risk as

$$2^{M-m}\sup_z \sigma^4(z)\tilde{C}_r T^{-\frac{c^2(1-1/r)}{2\rho_\infty^2}}. \qquad (15)$$

2. Case $\sigma^2(t/T) \neq$ constant for $t = 2^{M-m}n,\dots,2^{M-m}(n+1)-1$ (so that possibly $\mu_{1,m,n} \neq \mu_{2,m,n}$). Again suppose $n=0$ to shorten notation. Denote $\tilde{\tilde{\mu}}_{m,0}^{(h)} = \tilde{d}_{m,0}\mathbb{I}(|\tilde{f}_{m,0}| > t_m)$. We have

$$\mathbb{E}(\tilde{\tilde{\mu}}_{m,0}-\mu_{m,0})^2 \le 2\mathbb{E}(\tilde{\tilde{\mu}}_{m,0}-\tilde{\tilde{\mu}}_{m,0}^{(h)})^2 + 2\mathbb{E}(\tilde{\tilde{\mu}}_{m,0}^{(h)}-\mu_{m,0})^2 =: VII + VIII.$$

Using the representation $\tilde{\tilde{\mu}}_{m,0}^{(h)} = \tilde{s}_{m,0}\tilde{f}_{m,0}\mathbb{I}(|\tilde{f}_{m,0}| > t_m)$, it is easily seen that $|\tilde{\tilde{\mu}}_{m,0} - \tilde{\tilde{\mu}}_{m,0}^{(h)}| \le \tilde{s}_{m,0}t_m$, which leads to $VII \le 2t_m^2\mathbb{E}\,\tilde{s}_{m,0}^2 = 2t_m^2(\mathrm{var}\,\tilde{s}_{m,0} + (\mathbb{E}\,\tilde{s}_{m,0})^2)$. Using an explicit formula for $\tilde{s}_{m,0}$, we bound $\mathrm{var}\,\tilde{s}_{m,0} \le 2\sup_z\sigma^4(z)\rho_\infty^2$ and $\mathbb{E}\,\tilde{s}_{m,0} \le 2^{\frac{M-m}{2}}\sup_z\sigma^2(z)$, which, using the explicit form of t_m, finally leads to

$$VII \le 8c^2\sup_z\sigma^4(z)\log T(2\rho_\infty^2+1).$$

We bound $VIII$ as follows:

$$\begin{aligned} VIII &= 2\mathbb{E}(\tilde{d}_{m,0}\mathbb{I}(|\tilde{f}_{m,0}|>t_m)-\mu_{m,0})^2 \le 4\mathbb{E}((\tilde{d}_{m,0}-\mu_{m,0})\mathbb{I}(|\tilde{f}_{m,0}|>t_m))^2 \\ &+ 4\mu_{m,0}^2\mathbb{P}(|\tilde{f}_{m,0}|<t_m) \le 4w_{m,0} + 4\mu_{m,0}^2\mathbb{P}(|\tilde{f}_{m,0}|<t_m). \end{aligned} \qquad (16)$$

Note $w_{m,0} \le 2\sup_z\sigma^4(z)\rho_\infty^2$. If $\mu_{m,0}=0$, then the second summand disappears. Assume w.l.o.g. that $\mu_{m,0} > 0$. Using Markov's inequality and the fact that $(A+B+C)^2 \le$

$3(A^2+B^2+C^2)$, we bound (16) by

$$8\sup_z \sigma^4(z)\rho^2_\infty + 4\mu^2_{m,0}\mathbb{P}(\tilde{f}_{m,0} < t_m) \leq$$
$$8\sup_z \sigma^4(z)\rho^2_\infty + 4\mu^2_{m,0}\mathbb{P}((\tilde{d}_{1,m,0}-\mu_{1,m,0})(t_m-1)+(\tilde{d}_{2,m,0}-\mu_{2,m,0})(t_m+1)+$$
$$+2\mu_{1,m,0}t_m > (1+t_m)\mu_{m,0}) \leq \quad \text{[Markov's inequality]}$$
$$8\sup_z \sigma^4(z)\rho^2_\infty + 12(1+t_m)^{-2}((1-t_m)^2 w_{1,m,0} + (1+t_m)^2 w_{2,m,0} + 4\mu^2_{1,m,0}t_m^2) \leq$$
$$16\sup_z \sigma^4(z)(2\rho^2_\infty + 3c^2\log T).$$

We are now in a position to bound II. Using Parseval's equality, we have

$$II \leq 2T^{-1}\mathbb{E}\left(\tilde{s}_{0,0} - T^{-1/2}\sum_{t=0}^{T-1}\sigma^2(t/T)\right)^2 + 2T^{-1}\sum_{m=0}^{M-1}\sum_{n=0}^{2^m-1}\mathbb{E}(\tilde{\hat{\mu}}_{m,n}-\mu_{m,n})^2.$$

Using the Gaussianity of $Z_{t,T}$, the first term above, labelled II_1, is bounded as follows.

$$II_1 = 2T^{-2}\mathbb{E}\left(\sum_{t=0}^{T-1}\sigma^2(t/T)(Z^2_{t,T}-1)\right)^2 \leq 2T^{-2}\sup_z\sigma^4(z)\sum_{t=0}^{T-1}\sum_{s=0}^{T-1}\text{cov}(Z^2_{t,T}, Z^2_{s,T})$$
$$= 4T^{-2}\sup_z\sigma^4(z)\sum_{t=0}^{T-1}\sum_{s=0}^{T-1}(\text{cor}(Z_{t,T}, Z_{s,T}))^2 \leq 4T^{-1}\sup_z\sigma^4(z)\rho^2_\infty.$$

We now bound the second term. Denote $\mathcal{N}_T = \{(m,n) : \sigma^2(t/T) = \text{const, for } t = 2^{M-m}n, \ldots, 2^{M-m}(n+1)-1\}$. At each scale m, at most B indices (m,n) are in $\mathcal{N}_T^c$. We have

$$2T^{-1}\sum_{m=0}^{M-1}\sum_{n=0}^{2^m-1}\mathbb{E}(\tilde{\hat{\mu}}_{m,n}-\mu_{m,n})^2 =$$
$$2T^{-1}\left\{\sum_{(m,n)\in\mathcal{I}_T\cap\mathcal{N}_T}\mathbb{E}\,\tilde{\hat{\mu}}^2_{m,n} + \sum_{(m,n)\in\mathcal{I}_T\cap\mathcal{N}_T^c}\mathbb{E}(\tilde{\hat{\mu}}_{m,n}-\mu_{m,n})^2 + \sum_{(m,n)\in\mathcal{I}_T^c}\mu^2_{m,n}\right\} \leq$$
$$2T^{-1}\sum_{(m,n)\in\mathcal{I}_T\cap\mathcal{N}_T}2^{M-m}\sup_z\sigma^4(z)\tilde{C}_r T^{-\frac{c^2(1-1/r)}{2\rho^2_\infty}} +$$
$$+16BM^*T^{-1}\sup_z\sigma^4(z)(2\rho^2_\infty(c^2\log T+2)+7c^2\log T)+$$
$$+2BT^{-1}\sum_{m=M^*}^{M-1}2^{M-m-2}\sup_z\sigma^4(z) =$$
$$O\left(\sup_z\sigma^4(z)\left\{T^{-\frac{c^2(1-1/r)}{2\rho^2_\infty}}\log T + BT^{-1}c^2\log^2 T\rho^2_\infty + BT^{\delta-1}\right\}\right).$$

Upon setting $r = 1/\delta$ and combining I and II, the result follows. $\square$

Proof of Proposition 4.1.

1. Assumptions 2.1 and 4.1 together with the specific form of the constant c, imply that the term I in the proof of Theorem 3.1 is of order $(T^{-1}2^{-j} + T^{\frac{1}{2\log_2 a - 1} - 1})\gamma^2 2^{-j} \log T$.

2. Thresholds t_m are of order $2^{-j/2} 2^{-\frac{M-m}{2}} \log^{1/2} T \leq T^{(\varepsilon-\delta)/2} \log^{1/2} T$ (as $J^* \leq \varepsilon \log_2 T$), and therefore $t_m \to 0$ uniformly over $j = -1, \ldots, -J^*$ (as $\varepsilon < \delta$). Thus the equality (14) holds uniformly over j. By Proposition 2.2 and Assumption 4.2, the risk bound in formula (15) is at most $2^{M-m} C_1^2 \tilde{C}_r T^{-\frac{\gamma^2(1-\delta)}{2C_3}}$ uniformly over $j = -1, \ldots, J^*$. This implies that the term II in the proof of Theorem 3.1 is of order $C_1^2 (T^{-\frac{\gamma^2(1-\delta)}{2C_3}} \log T + BT^{-1}\gamma^2 2^{-2j} C_3 \log^2 T + BT^{\delta-1})$.

Combining I and II, the risk is easily found to be at most of the order $2^{-2j} T^{-\theta(a,\gamma,C_3,\delta)} \log^2 T$, where

$$\theta(a, \gamma, C_3, \delta) = \min\left(1 - \frac{1}{2\log_2 a - 1}, \frac{\gamma^2(1-\delta)}{2C_3}, 1 - \delta\right),$$

which leads to the result. $\square$

Proof of Theorem 4.1. We have

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}(\hat{S}_l(t/T) - S_l(t/T))^2 \leq$$
$$\frac{2}{T}\sum_{t=0}^{T-1} \mathbb{E}\left(\sum_{j=-J^*}^{-1} (\hat{\beta}_j(t/T) - \beta_j(t/T))(A^{-1})_{l,j}\right)^2 + \frac{2}{T}\sum_{t=0}^{T-1}\left(\sum_{j=-\infty}^{-J^*-1} \beta_j(t/T)(A^{-1})_{l,j}\right)^2$$
$$=: I + II.$$

From the proof of Theorem 2.15 in Nason et al. (2000), we know that for Haar wavelets, $(A^{-1})_{l,j} \leq C_4 2^{l/2} 2^{j/2}$. Using this, applying the Cauchy-Schwarz inequality to the sum over j, and then using Proposition 4.1, we obtain

$$I \leq \frac{2}{T}\sum_{t=0}^{T-1} \log T \sum_{j=-J^*}^{-1} (A^{-1})^2_{l,j} \mathbb{E}(\hat{\beta}_j(t/T) - \beta_j(t/T))^2 \leq 4C_4^2 2^l T^{-\theta(a,\gamma,C_3,\delta)+\varepsilon} \log^3 T.$$

Using Proposition 2.2 and the above-mentioned property of A^{-1}, II can easily be shown to be bounded by $2^{2l+1} C_1^2 C_4^2 T^{-2\varepsilon}$. Equating the powers of T in the bounds for I and II, we obtain the "optimal" ε as $\varepsilon = \theta(a, \gamma, C_3, \delta)/3$. Note that Proposition 4.1 requires that $\varepsilon < \delta$: this is satisfied as by assumption, $\theta(a, \gamma, C_3, \delta)/3 < \delta$. The result follows. $\square$

Proof of Proposition 6.1.

1. *(Beta distribution.)* Proof easy once one notes that $f^{Z^2}_{m,n} = (U - V)/(U + V)$ where U, V are independent $\chi^2_{2^{M-m-1}}$ variables; the desired density is then obtained by computing $P(f^{Z^2}_{m,n} < t)$ in terms of the cdf of an appropriate F distribution and then differentiating with respect to t.

2. *(Log-like property of $\mathcal{F}$.)* Let $T = 2^M$ and let $d^{Y^2}_{m,n}$ ($d^{\sigma^2}_{m,n}$, $d^{Z^2}_{m,n}$) denote the Haar coefficients of $\mathcal{F}Y^2_{t,T} - \overline{Y^2}$ ($\mathcal{F}\sigma^2(t/T) - \overline{\sigma^2}$, $\mathcal{F}Z^2_{t,T} - \overline{Z^2}$), for $m = 0, \ldots, M-1$ and $n = 0, \ldots, 2^m - 1$. Further, let $f^{Y^2}_{m,n}$ ($f^{\sigma^2}_{m,n}$, $f^{Z^2}_{m,n}$) denote the Haar-Fisz coefficients of Y^2 (σ^2, Z^2). Note that $d^{\cdot}_{m,n} = 2^{\frac{M-m}{2}} f^{\cdot}_{m,n}$. To prove (11), note that the l_2 distance between two vectors in the time domain is equal to their l_2 distance in the Haar domain, due to Parseval identity. Thus

$$T^{-1}\sum_{t=0}^{T-1}\mathbb{E}\{(\mathcal{F}Y^2_{t,T} - \overline{Y^2}) - (\mathcal{F}\sigma^2(t/T) - \overline{\sigma^2}) - (\mathcal{F}Z^2_{t,T} - \overline{Z^2})\}^2 =$$
$$T^{-1}\sum_{m=0}^{M-1}\sum_{n=0}^{2^m-1}\mathbb{E}\{d^{Y^2}_{m,n} - d^{\sigma^2}_{m,n} - d^{Z^2}_{m,n}\}^2 =$$
$$T^{-1}\sum_{m=0}^{M-1}\sum_{n=0}^{2^m-1} 2^{M-m}\mathbb{E}\{f^{Y^2}_{m,n} - f^{\sigma^2}_{m,n} - f^{Z^2}_{m,n}\}^2. \quad (17)$$

We now consider two cases:

(A) Case $\sigma^2(t/T) = \text{constant} =: \sigma^2$ for $t = 2^{M-m}n, \ldots, 2^{M-m}(n+1) - 1$. We then have $f^{\sigma^2}_{m,n} = 0$ and $f^{Y^2}_{m,n} = f^{Z^2}_{m,n}$ so that $2^{M-m}\mathbb{E}\{f^{Y^2}_{m,n} - f^{\sigma^2}_{m,n} - f^{Z^2}_{m,n}\}^2 = 0$.

(B) Case $\sigma^2(t/T) \neq \text{constant}$ for $t = 2^{M-m}n, \ldots, 2^{M-m}(n+1) - 1$. Suppose $n = 0$ to shorten notation; bounds for $n \neq 0$ are identical. We have

$$f^{Y^2}_{m,n} - f^{\sigma^2}_{m,n} - f^{Z^2}_{m,n} = \frac{\sum_{i=0}^{2^{M-m-1}-1}\sigma^2(i/T)Z^2_{i,T} - \sum_{i=2^{M-m-1}}^{2^{M-m}-1}\sigma^2(i/T)Z^2_{i,T}}{\sum_{i=0}^{2^{M-m-1}-1}\sigma^2(i/T)Z^2_{i,T} + \sum_{i=2^{M-m-1}}^{2^{M-m}-1}\sigma^2(i/T)Z^2_{i,T}} -$$
$$\frac{\sum_{i=0}^{2^{M-m-1}-1}\sigma^2(i/T) - \sum_{i=2^{M-m-1}}^{2^{M-m}-1}\sigma^2(i/T)}{\sum_{i=0}^{2^{M-m-1}-1}\sigma^2(i/T) + \sum_{i=2^{M-m-1}}^{2^{M-m}-1}\sigma^2(i/T)} -$$
$$\frac{\sum_{i=0}^{2^{M-m-1}-1}Z^2_{i,T} - \sum_{i=2^{M-m-1}}^{2^{M-m}-1}Z^2_{i,T}}{\sum_{i=0}^{2^{M-m-1}-1}Z^2_{i,T} + \sum_{i=2^{M-m-1}}^{2^{M-m}-1}Z^2_{i,T}} =: \frac{A_1 - A_2}{A_1 + A_2} - \frac{B_1 - B_2}{B_1 + B_2} - \frac{C_1 - C_2}{C_1 + C_2} =$$
$$\frac{C_1 - C_2}{C_1 + C_2}\left\{\frac{(A_1 - A_2)(B_1 - B_2) - 2(A_1B_1 + A_2B_2)}{(A_1 + A_2)(B_1 + B_2)}\right\} +$$
$$2\frac{A_1B_2 - A_2B_1}{(A_1 + A_2)(B_1 + B_2)} \leq 3\frac{|C_1 - C_2|}{C_1 + C_2} + 2\frac{|A_1B_2 - A_2B_1|}{(A_1 + A_2)(B_1 + B_2)}. \quad (18)$$

By the "beta distribution" property, $(C_1 - C_2)/(C_1 + C_2)$ is distributed as $2\beta(2^{M-m-2}, 2^{M-m-2}) -$

1 and thus

$$\mathbb{E}\{|C_1 - C_2|/(C_1 + C_2)\}^2 = (2^{M-m-1} + 1)^{-1}. \tag{19}$$

Further, note that

$$\frac{|A_1 B_2 - A_2 B_1|}{(A_1 + A_2)(B_1 + B_2)} \le \frac{|A_1 B_2 - A_2 B_1|}{\min_z \sigma^2(z)(C_1 + C_2)(B_1 + B_2)},$$

and the ratio of quadratic forms in normal variables $(A_1 B_2 - A_2 B_1)/(C_1 + C_2)$ satisfies the assumptions of Corollary 1 in Ghazal (1994), and so

$$\mathbb{E}\{(A_1 B_2 - A_2 B_1)/(C_1 + C_2)\}^2 = \frac{B_2^2 \sum_{i=0}^{2^{M-m-1}-1} \sigma^4(i/T) + B_1^2 \sum_{i=2^{M-m-1}}^{2^{M-m}-1} \sigma^4(i/T)}{2^{M-m-1}(2^{M-m} + 2)} \le \frac{\max_z \sigma^4(z)(B_1^2 + B_2^2)}{2^{M-m} + 2} \tag{20}$$

Using (19) and (20) in (18), we obtain

$$\mathbb{E}(f^{Y^2}_{m,n} - f^{\sigma^2}_{m,n} - f^{Z^2}_{m,n})^2 \le (2^{M-m-1} + 1)^{-1} \left\{ 18 + \frac{4 \max_z \sigma^4(z)}{\min_z \sigma^4(z)} \right\} \tag{21}$$

Observe that at each scale m, at most B indices n fall into category (B) above. Therefore, using (21), we bound (17) from above by

$$BT^{-1} \sum_{m=0}^{M-1} 2^{M-m}(2^{M-m-1}+1)^{-1} \left\{ 18 + \frac{4 \max_z \sigma^4(z)}{\min_z \sigma^4(z)} \right\} \le \frac{\log_2 T}{T} 2B \left\{ 18 + \frac{4 \max_z \sigma^4(z)}{\min_z \sigma^4(z)} \right\},$$

which completes the proof.

Techniques for proving the following properties: 3. *(variance stabilization)*, 4. *(asymptotic normality)* and 5. *(lack of spurious correlation)* are the same as those used in the proofs of Propositions 1 and 2 in Fryzlewicz and Nason (2004). We omit the details.

References

F. Battaglia. Some extensions in the evolutionary spectral analysis of a stochastic process. *Boll. Un. Mat. Ital. B (5)*, 16:1154–1166, 1979.

D. Bosq. *Nonparametric Statistics for Stochastic Processes*. Springer-Verlag, New York, 2 edition, 1998.

L. Calvet and A. Fisher. Forecasting multifractal volatility. *J. Econometrics*, 105:27–58, 2001.

C. Chiann and P. Morettin. A wavelet analysis for time series. *J. Nonparametr. Statist.*, 10: 1–46, 1999.

R. Dahlhaus. On the Kullback-Leibler information divergence of locally stationary processes. *Stoch. Proc. Appl.*, 62:139–168, 1996.

R. Dahlhaus and W. Polonik. Nonparametric quasi maximum likelihood estimation for Gaussian locally stationary processes. *Ann. Statist.*, to appear, 2006.

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

R.A. Davis, T.C.M. Lee, and G.A. Rodriguez-Yam. Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.*, 101:223–239, 2006.

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

P. Fryzlewicz. Modelling and forecasting financial log-returns as locally stationary wavelet processes. *Journal of Applied Statistics*, 32:503–528, 2005.

P. Fryzlewicz and G. P. Nason. A Haar-Fisz algorithm for Poisson intensity estimation. *J. Comput. Graph. Stat.*, 13:621–638, 2004. Technical Report version: `http://www.stats.bris.ac.uk/research/stats/pub/ResRept/2003/03.pdf`.

P. Fryzlewicz, S. Van Bellegem, and R. von Sachs. Forecasting non-stationary time series by wavelet process modelling. *Ann. Inst. Statist. Math.*, 55:737–764, 2003.

P. Fryzlewicz, T. Sapatinas, and S. Subba Rao. A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika*, to appear, 2006.

H-Y. Gao. Choice of thresholds for wavelet shrinkage estimate of the spectrum. *J. Time Ser. Anal.*, 18:231–252, 1997.

G.A. Ghazal. Moments of the ratio of two dependent quadratic forms. *Statistics & Probability Letters*, 20:313–319, 1994.

N. Johnson and S. Kotz. *Distributions in Statistics: Continuous Univariate Distributions-1*. Houghton Mifflin Company, Boston, 1970.

M.C. Jones. On moments of ratios of quadratic forms in normal variables. *Statistics & Probability Letters*, 6:129–136, 1987.

W. Kim. Econometric analysis of locally stationary time series models. *Manuscript, Yale University*, 1998.

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

S. Mallat, G. Papanicolaou, and Z. Zhang. Adaptive covariance estimation of locally stationary processes. *Ann. Stat.*, 26:1–47, 1998.

G. Mélard and A. Herteleer-De Schutter. Contributions to evolutionary spectral theory. *J. Time Ser. Anal.*, 10:41–63, 1989.

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

G. P. Nason and B. W. Silverman. The stationary wavelet transform and some statistical applications. *Lect. Notes Statist.*, 103:281–300, 1995.

G. P. Nason, R. von Sachs, and G. Kroisandt. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. Roy. Statist. Soc. B*, 62:271–292, 2000.

M.H. Neumann and R. von Sachs. Wavelet thresholding: beyond the Gaussian iid situation. *Lect. Notes Statist.*, 103:301–329, 1995.

H. Ombao, J. Raz, R. von Sachs, and W. Guo. The SLEX model of a non-stationary random process. *Ann. Inst. Statist. Math*, 54:171–200, 2002.

M. Pensky and B. Vidakovic. Bayesian decision theoretic scale adaptive estimation of a log-spectral density. *Preprint*, 2004.

M. Priestley. Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc. B*, 27: 204–237, 1965.

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

K. Sakiyama. Some statistical applications for locally stationary processes. *Sci. Math. Jpn.*, 56:231–250, 2002.

R. Swift. The evolutionary spectra of a harmonizable process. *J. Appl. Statist. Sci.*, 9: 265–275, 2000.

S. Van Bellegem and R. von Sachs. Locally adaptive estimation of evolutionary wavelet spectra. *Preprint, available from* `http://www.stat.ucl.ac.be/ISpersonnel/vanbelle/pub/VBvS.pdf`, 2003.

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

R. von Sachs and B. MacGibbon. Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scand. J. Statist.*, 27:475–499, 2000.

G. Wahba. Automatic smoothing of the log periodogram. *J. Ameri. Statist. Assoc.*, 75: 122–132, 1980.

A.T. Walden, D.B. Percival, and E.J. McCoy. Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Trans. Sig. Proc.*, 46:3153–3165, 1998.