Piotr Fryzlewicz and Guy P. Nason

Smoothing the wavelet periodogram using the Haar-Fisz transform

Technical report

Technical Report: Smoothing the wavelet periodogram using the Haar-Fisz transform

Piotr Fryzlewicz and Guy P. Nason

University of Bristol, UK

[August 3, 2004]

Summary. The wavelet periodogram is hard to smooth because of the low signal-to-noise ratio and non-stationary covariance structure. This article introduces a method for smoothing a local wavelet periodogram by applying a Haar-Fisz transform which approximately Gaussianizes and approximately stabilizes the variance of the periodogram. Consequently, smoothing the transformed periodogram can take advantage of the wide variety of existing techniques suitable for homogeneous Gaussian data. This article demonstrates the superiority of the new method over existing methods and supplies theory that proves the Gaussianizing, variance stabilizing and decorrelation properties of the Haar-Fisz transform.

Keywords: periodogram smoothing; Gaussianizing transform; variance stabilization; functional central limit theorem.

1 Introduction

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

Address for correspondence: Department of Mathematics, South Kensington Campus, Imperial College, London, SW7 2AZ, England
Email: P.Fryzlewicz@imperial.ac.uk

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 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 & 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 extremely efficiently.

The *wavelet periodogram*, the main quantity of interest in this paper, is a wavelet alternative to the classical periodogram and can be loosely defined as a sequence of squared non-decimated wavelet coefficients of a process. Nason *et al.* (2000) use it as a basic statistic for estimating the evolutionary wavelet spectrum and the local autocovariance in the LSW model. Like the classical periodogram, the wavelet periodogram is not consistent and needs to be smoothed. Nason *et al.* (2000) recommended using wavelet shrinkage with a threshold adapted for χ^2 data or, alternatively, taking the log transform of the wavelet periodogram and then using standard wavelet shrinkage for Gaussian distributed data. Neither of these approaches is perfect: for the non-linear wavelet shrinkage a local pilot estimate of the local variance is required. For the log transform, it stabilizes variance but the log-transformed periodogram is far from Gaussian and then standard denoisers cannot be "automatically" used. Chiann & Morettin (1999) define and analyse some properties of a different kind of wavelet periodogram, based on the orthonormal wavelet transform, for stationary processes.

The prime objective of this paper is to propose a new multiscale transform technique for smoothing the wavelet periodogram. The idea behind our algorithm is the following: we first preprocess the wavelet periodogram using a nonlinear wavelet-based transformation, which we call the Haar-Fisz transformation for χ^2 data. Then we denoise the preprocessed vector as if it were signal plus stationary Gaussian noise. Finally, we apply the inverse Haar-Fisz transform to obtain an estimate of the spectral structure of the original process.

Our Haar-Fisz transform is a new Gaussianizing and variance stabilizing transform, which operates in the wavelet domain, and not in the time domain, like the standard log transformation. Throughout the paper whenever we mention Gaussianization or variance stabilization we mean *approximately*. For a precise technical characterisation of these properties see Section 5.

The advantages of our method are the following.

1. Its performance is good. We often see performance gains in MISE of around 25%. see Section 5.3 below.

2. For a time series of length T the algorithm is of computational order $T \log(T)$ for estimating all scales of the wavelet periodogram.

3. It is simple and easy to code.

4. It is fully automatic (up to any parameters that the Gaussian denoiser requires).

5. It can make use of *any* signal+Gaussian noise denoising technology, an area where a vast amount of research effort has been and is being expended. Hence our method can only get better as we take advantage of newer Gaussian denoisers.

To achieve the main objective, we take the following steps. Section 2 recalls the definitions of an LSW process, the wavelet periodogram, and other preliminaries. We introduce the (multiscale) Haar-Fisz transform for the wavelet periodogram in Section 3 (the Haar-Fisz transform for Poisson data was recently introduced by Fryzlewicz & Nason (2004)). The Haar-Fisz transform works as an approximate Gaussianizer and variance stabilizer and Section 4 proves a functional central limit theorem (FCLT) which is used in Section 5 to formally quantify these Gaussianizing and variance stabilizing assertions. Section 5 also uses simulation to investigate the practical performance of the Haar-Fisz transform and compare it to an exising alternative. Section 6 proposes an algorithm for smoothing the wavelet periodogram, based on our Haar-Fisz transform. This section also contains a simulation study which demonstrates the (usually) superior performance of our algorithm over the existing competitor. Section 7 uses our proposed smoothing methodology to perform a local variance analysis of the Dow Jones index, and concludes that the series can be modelled as Gaussian.

2 Preliminaries

2.1 The LSW model

We start by recalling the definition of an LSW process.

Definition 2.1 (Nason *et al.* (2000)) *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*

$$X_{t,T} = \sum_{j=-J(T)}^{-1} \sum_{k=-\infty}^{\infty} \omega_{j,k;T} \psi_{j,t-k} \xi_{j,k;T}, \tag{2}$$

where $j \in \{-1, -2, \ldots\}$ and $k \in \mathbb{Z}$ are, respectively, scale and location parameters, $\omega_{j,k;T}$ are real constants, ψ_j are discrete non-decimated wavelet vectors, $\{\xi_{j,k;T}\}_{j,k}$ are, for each T, zero-mean orthonormal identically distributed random variables, and for each $j \leq -1$, there exists a Lipschitz function $W_j(z) : [0, 1) \to \mathbb{R}$ such that

- *$\sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty$, uniformly in $z \in (0, 1)$,*
- *the Lipschitz constants L_j satisfy $\sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty$,*
- *there exists a sequence of constants C_j satisfying $\sum_{j=-\infty}^{-1} C_j < \infty$ such that, for each T,*

$$\sup_{k=0,\ldots,T-1} |\omega_{j,k;T} - W_j(k/T)| \leq C_j/T. \tag{3}$$

In formula (2), parameters $\omega_{j,k;T}$ can be thought of as a scale- and location-dependent transfer function, while the nondecimated 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. the $\xi_{j,k;T}$ are distributed as $N(0,1)$.

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

$$\psi_{j,k} = 2^{j/2}\mathbb{I}_{\{0,\ldots,2^{-j-1}-1\}}(k) - 2^{j/2}\mathbb{I}_{\{2^{-j-1},\ldots,2^{-j}-1\}}(k), \tag{4}$$

for $j = -1, -2, \ldots$ and $k \in \mathbb{Z}$, where $j = -1$ corresponds to the finest scale. All results of this paper are true not only for Haar wavelets but also for all other compactly supported wavelets from the Daubechies' families (Daubechies (1992)). For a given wavelet system ψ, we set J in Definition 2.1 to be $J(n) = -\min\{j : \mathcal{L}_j \leq n\}$, where $\mathcal{L}_j$ is the length of support of ψ_j.

Given an LSW process $\{X_{t,T}\}$, we cannot uniquely determine the transfer function $\omega_{j,k;T}$ due to the overcompleteness of the non-decimated wavelet system ψ. However, the asymptotic *evolutionary wavelet spectrum* $S_j(z) := |W_j(z)|^2 = \lim_{T\to\infty} |\omega_{j,\lfloor zT\rfloor;T}|^2$, defined on the rescaled-time interval $z \in [0,1)$, is uniquely defined and can be estimated by means of asymptotically unbiased estimators. 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 & MacGibbon (2000) for a recent wavelet-based technique for detrending locally stationary processes.

The autocovariance function of $\{X_{t,T}\}$ can also be defined in rescaled time: $c_T(z,\tau) = \mathbb{E}(X_{\lfloor zT\rfloor,T}X_{\lfloor zT\rfloor+\tau,T})$. Nason *et al.* (2000) define the corresponding asymptotic *local autocovariance* of $\{X_{t,T}\}$ as

$$c(z,\tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau), \tag{5}$$

where $\Psi_j(\tau) = \sum_{k=-\infty}^{\infty} \psi_{j,k}\psi_{j,k+\tau}$, and show that $|c_T(z,\tau) - c(z,\tau)| = O(T^{-1})$ as $T \to \infty$, uniformly in $\tau \in \mathbb{Z}$ and $z \in (0,1)$. Formula (5) defines the correspondence between the local autocovariance and the evolutionary wavelet spectrum (analogous to the usual Fourier transform link between classical spectrum and autocovariance), so that an estimate of the latter can be used to estimate the former.

2.2 The wavelet periodogram in the LSW model

We are now in a position to recall the definition of the *wavelet periodogram*, the main quantity of interest in this paper.

Definition 2.2 (Nason *et al.* (2000)) *Let $X_{t,T}$ be an LSW process constructed using the*

wavelet system ψ. The triangular stochastic array

$$I_{t,T}^{(j)} = \left| \sum_s X_{s,T} \psi_{j,t-s} \right|^2 \tag{6}$$

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

Throughout the paper, we assume that the reader is familar with the fast Discrete Wavelet Transform (DWT; see Mallat (1989)), as well as with the fast Non-decimated DWT (NDWT; see Nason & 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 $j = -1, -2, \ldots, -J(T)$.

We quote the following result:

Proposition 2.1 (Nason *et al.* (2000)) *Let $X_{t,T}$ satisfy Definition 2.1 and define*

$$A_{i,j} = \sum_\tau \Psi_i(\tau) \Psi_j(\tau).$$

We have

$$\mathbb{E} I_{t,T}^{(j)} = \sum_{i=-\infty}^{-1} S_i \left(\frac{t}{T} \right) A_{i,j} + O(2^{-j}/T). \tag{7}$$

If, in addition, $\xi_{j,k}$ are Gaussian (and hence $X_{t,T}$ is Gaussian), then

$$\mathrm{Var} \left(I_{t,T}^{(j)} \right) = 2 \left(\sum_{i=-\infty}^{-1} S_i \left(\frac{t}{T} \right) A_{i,j} \right)^2 + O(2^{-j}/T). \tag{8}$$

The following proposition shows that the wavelet periodogram at each scale j is typically a correlated sequence.

Proposition 2.2 *Let $X_{t,T}$ be a Gaussian LSW process satisfying $S_j(z) \leq D 2^j$. We have*

$$\mathrm{cov} \left(I_{t,T}^{(j)}, I_{t+s,T}^{(j)} \right) = 2 \left(\sum_{\tau=-\infty}^{\infty} c \left(\frac{t}{T}, \tau \right) \Psi_j(\tau + s) \right)^2 + O(2^{-j}/T). \tag{9}$$

The proof uses exactly the same technique as the proof of (8).

To simplify the notation, we denote

$$\beta_j(z) = \sum_{i=-\infty}^{-1} S_i(z) A_{i,j}. \tag{10}$$

For non-trivial processes we assume that $\beta_j(z)$ is bounded away from zero (note that by (10) $\beta_j(z) \equiv 0$ in a neighbourhood of z_0 would imply all $S_i(z) \equiv 0$ in that neighbourhood and the resulting process would locally be deterministic and exactly zero). The main aim of the paper is to propose a new method for estimating $\beta_j(z)$ for $j = -1, -2, \ldots, -J(T)$, from a single stretch of observations of $X_{t,T}$. By (7) and (8), $I^{(j)}_{t,T}$ is an asymptotically unbiased but inconsistent estimate of $\beta_j(z)$ and needs to be smoothed to obtain consistency. Being able to estimate $\beta_j(z)$ is useful in two ways:

1. An estimate of $\beta_j(z)$ can be used to obtain an estimate of $S_j(z)$ (by (10) and by the invertibility of $(A_{m,n})$, see Nason *et al.* (2000) for details);

2. The estimate of $S_j(z)$ can in turn be used to obtain an estimate of the local autocovariance $c(z, \tau)$ (using (5)).

In short, estimating $\{\beta_j(z)\}_{j=-1}^{-J(T)}$ allows us to make inference about the time-varying second-order structure of $X_{t,T}$.

Example 1. Figure 1 shows an example of a wavelet spectrum, realisation from that spectrum and the wavelet periodogram of that realisation at scale -1. 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 (see formulae (7) and (8)),

- a low signal-to-noise ratio: again by (7) and (8) we obtain, neglecting the remainders, $\mathbb{E}I^{(j)}_{t,T} / \left\{ \mathrm{Var}\left(I^{(j)}_{t,T}\right) \right\}^{1/2} = 2^{-1/2}$,

- the presence of correlation in $I^{(j)}_{t,T}$ (see formula (9)).

Many existing denoising techniques have been designed to handle stationary Gaussian noise and therefore it would be desirable to be able to transform the wavelet periodogram into a signal contaminated with such noise before the denoising is performed. A well-known technique for stabilizing the variance of scaled χ^2_n variables is the log-transform, see e.g. Priestley (1981); however, the resulting variable is still far from Gaussian if, like here, $n = 1$. Nason *et al.* (2000) propose a wavelet-based technique for smoothing the wavelet periodogram without any pre-processing: their method is based on Neumann & von Sachs (1995).

3 The Haar-Fisz transform

In this section, we propose a multiscale algorithm, called the *Haar-Fisz transform (for χ^2 data)*, for stabilizing the variance of the wavelet periodogram and bringing its distribution closer to normality. The input to the algorithm is:

- A single row of the wavelet periodogram $I^{(j)}_{t,T}$ at a fixed scale j: here, we assume that T is an integer power of two. To simplify the notation in this section, we drop the

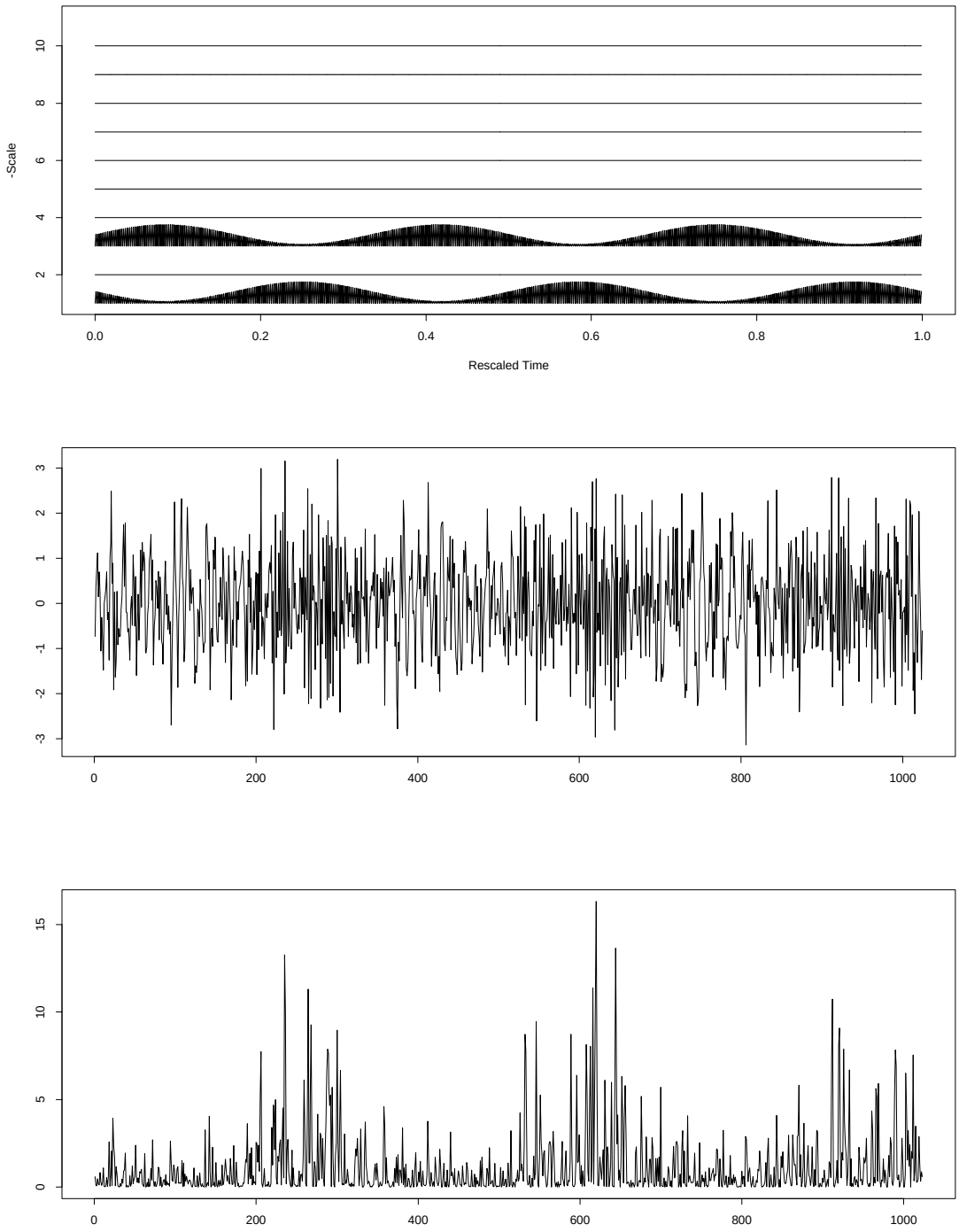

Figure 1: Top: wavelet spectrum example where only $S_{-1}(z)$ and $S_{-3}(z)$ are non-zero; Middle: example sample path from the top spectrum of length $T = 1024$; Bottom: Haar periodogram of the sample path at scale -1.

superscript j and the subscript T and consider the sequence $I_t := I^{(j)}_{t,T}$, or, in vector notation, $\mathbf{I} = (I^{(j)}_{0,T}, \ldots, I^{(j)}_{T-1,T})'$.

- A fixed integer $M \in \{1, 2, \ldots, \log_2(T)\}$; its meaning will become clear later.

The output from the algorithm is:

- The empirical mean of $\mathbf{I}$, denoted by $\bar{\mathbf{I}}$.
- A vector $\mathbf{U}^M$ of length 2^M.

The vector $\mathbf{U}^M$ is constructed as follows:

1. Let $\mathbf{s}^M$ be the vector of local averages of $\mathbf{I}$:

$$s^M_n = \frac{2^M}{T} \sum_{t=nT2^{-M}}^{(n+1)T2^{-M}-1} I_t \quad \text{for} \quad n = 0, 1, \ldots, 2^M - 1. \tag{11}$$

2. For each $m = M-1, M-2, \ldots, 0$, recursively form vectors $\mathbf{s}^m$ and $\mathbf{f}^m$:

$$s^m_n = \frac{1}{2}(s^{m+1}_{2n} + s^{m+1}_{2n+1}) \tag{12}$$

$$f^m_n = \frac{s^{m+1}_{2n} - s^{m+1}_{2n+1}}{2s^m_n}, \tag{13}$$

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

3. For each $m = 0, 1, \ldots, M-1$, recursively modify the vectors $\mathbf{s}^{m+1}$:

$$s^{m+1}_{2n} = s^m_n + f^m_n \tag{14}$$

$$s^{m+1}_{2n+1} = s^m_n - f^m_n, \tag{15}$$

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

4. Set $\mathbf{U}^M := \mathbf{s}^M - \bar{\mathbf{I}}$.

We denote $\mathcal{F}^M\mathbf{I} := \mathbf{U}^M$. The nonlinear operator $\mathcal{F}^M$ is called the *Haar-Fisz transform of* $\mathbf{I}$ *at the resolution level* M.

If $M = \log_2(T)$, then the length of $\mathcal{F}^M\mathbf{I}$ is T and the algorithm is invertible, i.e. $\mathbf{I}$ can be reconstructed from $\mathcal{F}^M\mathbf{I}$ and $\bar{\mathbf{I}}$ by reversing the steps 4.–1. Therefore, the case $M = \log_2(T)$ is the one we are the most interested in in practice. However, the exact asymptotic Gaussianizing properties of the Haar-Fisz transform only hold for M fixed (i.e. independent of T), and this case is investigated theoretically in Section 5.1. Section 5.2 provides some heuristics as to the behaviour of $\mathcal{F}^M\mathbf{I}$ when $M = \log_2(T)$: we still conclude that the distribution of $\mathcal{F}^{\log_2(T)}\mathbf{I}$ is close to Gaussian with a constant variance. To simplify notation, we denote $\mathcal{F} := \mathcal{F}^{\log_2(T)}$.

The reader will note that the steps 2.–4. of the algorithm are similar to the forward and inverse Discrete Haar Transform except the division by s_n^m in formula (13). The division by s_n^m is an application of the theory in Fisz (1955) and hence the name of our algorithm is the Haar-Fisz transform (for the wavelet periodogram). Informally speaking: the division by s_n^m in formula (13) acts as a variance stabilization step and also induces a degree of Gaussianization; formula (11) in step 1. also Gaussianizes (by a CLT type argument); in formulae (14) and (15) of step 3. a further degree of Gaussianization occurs (again by a CLT type argument). Fryzlewicz & Nason (2004) considered a similar Haar-Fisz algorithm for processing an independent Poisson signal. In that case the appropriate normalisation is to divide through by $\sqrt{s_n^m}$.

Example 2. Empirical comparison of log and Haar-Fisz transforms. To give the reader an intuitive feel for the properties of the Haar-Fisz transform when compared to log we describe the following simple example. We performed the log and Haar-Fisz transforms on the sequence of independent variables $\{\Upsilon_t\}_{t=1}^{1024}$ where Υ_t is distributed as $\upsilon_t \chi_1^2$ with $\{\upsilon_t\}_{t=1}^{512} \equiv 1$ and $\{\upsilon_t\}_{t=513}^{1024} \equiv 2$. We simulated 1000 sample paths of Υ_t and computed the mean empirical root signal to noise ratio (SNR=standard deviation of the signal divided by the standard deviation of the noise). The result for $\log \Upsilon_t$ was 0.156 and that for $\mathcal{F}\Upsilon_t$ was 0.297 (to 3 decimal places). For the log transform: the mean variance over the first half of the signal was 4.909 and 4.912 on the second half. For Haar-Fisz the corresponding figures were 1.266 and 1.263. As for Gaussianity, the mean Kolmogorov-Smirnov (K-S) statistic for log transformed values was 0.10 and for Haar-Fisz it was 0.02. For this example the SNR for Haar-Fisz is nearly twice that as achieved by the log transform. Moreover, the Haar-Fisz transformed variables are more Gaussian as demonstrated by their lower K-S statistics. Finally, both transforms stabilize variance extremely well. More extensive comparisons can be found in Section 5.

Example 1. (continued) Figure 2 compares the log and Haar-Fisz transforms of the wavelet periodogram from Figure 1. The Haar-Fisz-transformed wavelet periodogram appears to be much closer to normality.

3.1 Examples of the Haar-Fisz formula

As an example, consider $T = 8$. For $M = 2$, we have:

$$\begin{aligned}
U_0^2 &= \frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} + \frac{I_0 + I_1 - I_2 - I_3}{\sum_{i=0}^{3} I_t} \\
U_1^2 &= \frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} - \frac{I_0 + I_1 - I_2 - I_3}{\sum_{i=0}^{3} I_t} \\
U_2^2 &= -\frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} + \frac{I_4 + I_5 - I_6 - I_7}{\sum_{i=4}^{7} I_t} \\
U_3^2 &= -\frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} - \frac{I_4 + I_5 - I_6 - I_7}{\sum_{i=4}^{7} I_t}.
\end{aligned}$$

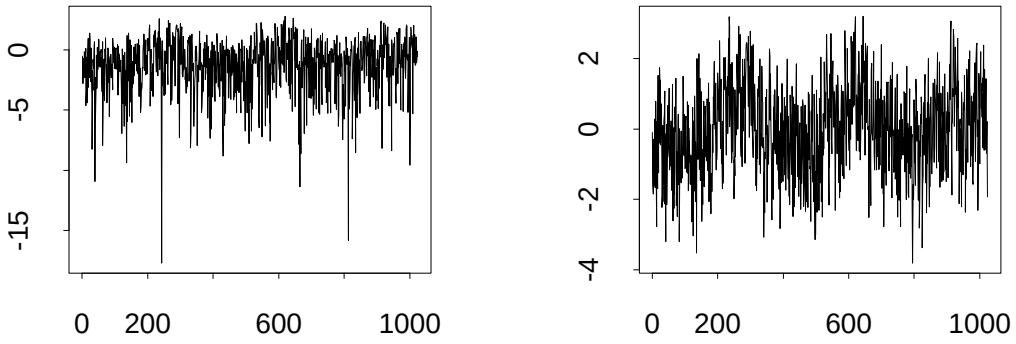

Figure 2: Transforms of the wavelet periodogram from the bottom plot of Figure 1. Left: log transform. Right: Haar-Fisz. Here $T = 1024$ and the full transform is performed, i.e. $M = 10$.

Similarly, for $M = 3$, we have:

$$
\begin{aligned}
U_0^3 &= \frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} + \frac{I_0 + I_1 - I_2 - I_3}{\sum_{i=0}^{3} I_t} + \frac{I_0 - I_1}{I_0 + I_1} \\
U_1^3 &= \frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} + \frac{I_0 + I_1 - I_2 - I_3}{\sum_{i=0}^{3} I_t} - \frac{I_0 - I_1}{I_0 + I_1} \\
U_2^3 &= \frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} - \frac{I_0 + I_1 - I_2 - I_3}{\sum_{i=0}^{3} I_t} + \frac{I_2 - I_3}{I_2 + I_3} \\
U_3^3 &= \frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} - \frac{I_0 + I_1 - I_2 - I_3}{\sum_{i=0}^{3} I_t} - \frac{I_2 - I_3}{I_2 + I_3} \\
U_4^3 &= -\frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} + \frac{I_4 + I_5 - I_6 - I_7}{\sum_{i=4}^{7} I_t} + \frac{I_4 - I_5}{I_4 + I_5} \\
U_5^3 &= -\frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} + \frac{I_4 + I_5 - I_6 - I_7}{\sum_{i=4}^{7} I_t} - \frac{I_4 - I_5}{I_4 + I_5} \\
U_6^3 &= -\frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} - \frac{I_4 + I_5 - I_6 - I_7}{\sum_{i=4}^{7} I_t} + \frac{I_6 - I_7}{I_6 + I_7} \\
U_7^3 &= -\frac{\sum_{t=0}^{3} I_t - \sum_{t=4}^{7} I_t}{\sum_{i=0}^{7} I_t} - \frac{I_4 + I_5 - I_6 - I_7}{\sum_{i=4}^{7} I_t} - \frac{I_6 - I_7}{I_6 + I_7}
\end{aligned}
$$

4 A Functional CLT for the centred wavelet periodogram

This section is concerned with a Functional Central Limit Theorem (FCLT) for the centred wavelet periodogram

$$Z_{t,T}^{(j)} := I_{t,T}^{(j)} - \mathbb{E}I_{t,T}^{(j)} \tag{16}$$

(see Davidson (1994) for more on the stochastic limit theory we use here). Our FCLT demonstrates that the normalised cumulative sum of the centred wavelet periodogram converges in distribution to a transformed Brownian motion. The theory in this section enables us to demonstrate the Gaussianizing, variance stabilizing and decorrelating properties of the Haar-Fisz transform established in Section 5. Before we state the theorem, we introduce some essential notation.

Definition 4.1 (transformed Brownian motion) *Let η be an increasing homeomorphism on $[0,1]$ with $\eta(0) = 0$ and $\eta(1) = 1$. A transformed Brownian motion B_η is defined as*

$$B_\eta(z) \stackrel{D}{=} B(\eta(z)), \quad z \in [0,1],$$

where B is a Brownian motion.

Definition 4.2 (cross-scale autocorrelation wavelets) *Let ψ be a fixed wavelet system. Vectors $\Psi_{i,j}$, for $i, j \in \{-1, -2, \ldots\}$, defined by*

$$\Psi_{i,j}(\tau) = \sum_{s=-\infty}^{\infty} \psi_{i,s+\tau}\psi_{j,s} \tag{17}$$

are called the cross-scale autocorrelation wavelets.

Denote

$$\begin{aligned} \bar{S}_j &= \max_z S_j(z) \\ \delta_k &= \max_{j=-1,\ldots,-k} \frac{C_j}{\bar{S}_j}, \end{aligned}$$

with the convention $0/0 = 0$. Denote further

$$\begin{aligned} A_{i,j}^\tau &= \sum_n \Psi_{i,j}(n)\Psi_{i,j}(n+\tau) \\ \beta_j^\tau(z) &= \sum_i S_i(z)A_{i,j}^\tau \end{aligned} \tag{18}$$

We now state the Functional Central Limit Theorem for the centred wavelet periodogram.

Theorem 4.1 *Let $X_{t,T}$ be a Gaussian LSW process, and let $Z_{t,T}^{(j)}$ be its centred wavelet periodogram at scale j. Define*

$$\begin{aligned} b_T^2 &= \mathbb{E}\left(\sum_{t=0}^{T-1} Z_{t,T}^{(j)}\right)^2 \qquad (19) \\ R_T(z) &= \frac{\sum_{t=0}^{\lfloor zT \rfloor - 1} Z_{t,T}^{(j)}}{b_T} \quad \text{for} \quad z \in [0,1]. \end{aligned}$$

If

$$\begin{aligned} \textit{there exists } \varepsilon > 0 \textit{ such that } \left(\sum_{i<0}\sum_{l \geq m+1} \Psi_{i,j}^2(l)\bar{S}_i\right)^{1/2} &= O(m^{-1/2-\varepsilon}), \qquad (20) \\ \delta_{J(T)}/T &\in l_\infty, \qquad (21) \\ \sup_z \sum_n |c(z,n)| &< \infty, \qquad (22) \\ \textit{there exists } D \textit{ such that } \bar{S}_j 2^{-j} &\leq D \quad \textit{for all } j, \qquad (23) \end{aligned}$$

and if $\beta_j(z)$ is bounded away from zero, then $R_T \xrightarrow{D} B_\eta$, where

$$\eta(z) = \frac{\int_0^z \sum_{\tau=-\infty}^{\infty} \left(\sum_i S_i(u) A_{i,j}^\tau\right)^2 du}{\int_0^1 \sum_{\tau=-\infty}^{\infty} \left(\sum_i S_i(u) A_{i,j}^\tau\right)^2 du}.$$

The proof appears in the Appendix. As is clear from the proof, the left-hand expression in condition (20) is a measure of dependence in the sequence $Z_{t,T}^{(j)}$ at lag m. Condition (21) places an additional restriction on the finite-sample wavelet spectrum $\{\omega_{j,k;T}^2\}_{j,k}$ in relation to the asymptotic spectrum $\{S_j(z)\}_j$. Condition (22) is a short-memory assumption for $X_{t,T}$, and condition (23) requires that the wavelet spectrum should decay at a certain speed as $j \to -\infty$.

We now give an example of a periodogram sequence which satisfies the technical condition (20). Let $X_{t,T}$ be a Gaussian LSW process constructed with Haar wavelets and such that $S_i(z) = S_i = 2^i$. Asymptotically, $X_{t,T}$ is a white noise process (see Fryzlewicz *et al.* (2003) for a proof of this fact). Consider the Haar periodogram of $X_{t,T}$ at scale $j = -1$. Using (4) and (17), simple algebra yields

$$\sum_{i<0} 2^i \sum_{l \geq m+1} \Psi_{i,-1}^2(l) = O(m^{-2}),$$

and (20) is satisfied. Also note that any process with $S_j(z) \equiv 0$ for all but a finite number of scales j satisfies (20).

In particular, Theorem 4.1 implies that $\mathbb{E}(R_T(z)^2) \to \eta(z)$ as $T \to \infty$, and that the increments of $R_T(z)$ are asymptotically independent. Theorem 4.1 is fundamental for the theoretical results of the next section.

5 Properties of the Haar-Fisz transform

5.1 Properties of the Haar-Fisz transform for M fixed

Theorem 5.1 *Let $X_{t,T}$ satisfy the assumptions of Theorem 4.1, and let $I^{(j)}_{t,T}$ be the wavelet periodogram of $X_{t,T}$ at scale j. Let the corresponding functions $\beta_j(z)$ and $\sum_\tau(\beta^\tau_j(z))^2$ (see formulae (10) and (18)) be continuous with bounded left and right derivatives. For M fixed, $\mathbf{U}^M = \mathcal{F}^M I^{(j)}_{t,T}$ admits the following decomposition:*

$$\mathbf{U}^M = \mathbf{V}^M + \mathbf{Y}^M,$$

where

1. *$\mathbf{V}^M$ has an almost-sure deterministic limit as $T \to \infty$;*
2. *$\sqrt{T}\mathbf{Y}^M \xrightarrow{D} N(0, \Sigma)$ as $T \to \infty$, with*

$$\left(2^{M+1}-2\right) \inf_{z\in[0,1]} \frac{\sum_\tau(\beta^\tau_j(z))^2}{(\beta_j(z))^2} - O(M) \le \Sigma_{n,n} \le \left(2^{M+1}-2\right) \sup_{z\in[0,1]} \frac{\sum_\tau(\beta^\tau_j(z))^2}{(\beta_j(z))^2} + O(M) \tag{24}$$

and

$$\Sigma_{n_1,n_2} = O(M) \quad \text{for} \quad n_1 \neq n_2. \tag{25}$$

Theorem 5.1 shows that the Haar-Fisz transformed periodogram, $\mathbf{U}^M$, admits an asymptotic "signal plus noise" type representation where explicit formulae for the "signal", $\mathbf{V}^M$, and "noise", $\mathbf{Y}^M$ appear in the proof. Property 2. above is called the *Gaussianization property* of the Haar-Fisz transform and, theoretically, this is shown by using the "convergence to Brownian motion" result in the FCLT of Theorem 4.1. Formulae (24) and (25) define, respectively, the *variance stabilization* and the *decorrelation properties* of the Haar-Fisz transform. The decorrelation property stems from the independent increments inherent in the Brownian motion that $\mathbf{Y}^M$ converges to, by Theorem 4.1.

As an interesting special case suppose $X_{t,T}$ is a time-modulated stationary process ($X_{t,T} = \sigma(t/T)Y_{t,T}$ where $Y_{t,T}$ is stationary). Then the quantity $\sum_\tau(\beta^\tau_j(z))^2/(\beta_j(z))^2$ is the asymptotic sum of the autocorrelations of $I^{(j)}_{t,T}$ and in this case is not dependent on z. This means that the main terms in formula (24) are identical which implies that $\Sigma_{n,n} = \text{const} + O(M)$ for all n (i.e. the variance is almost exactly stabilized). Note that exact variance stabilization is not possible unless the log transform is used. However, as exemplified by Example 2, in practice Haar-Fisz variance stabilization is very good and one also benefits from Gaussianization and a better SNR.

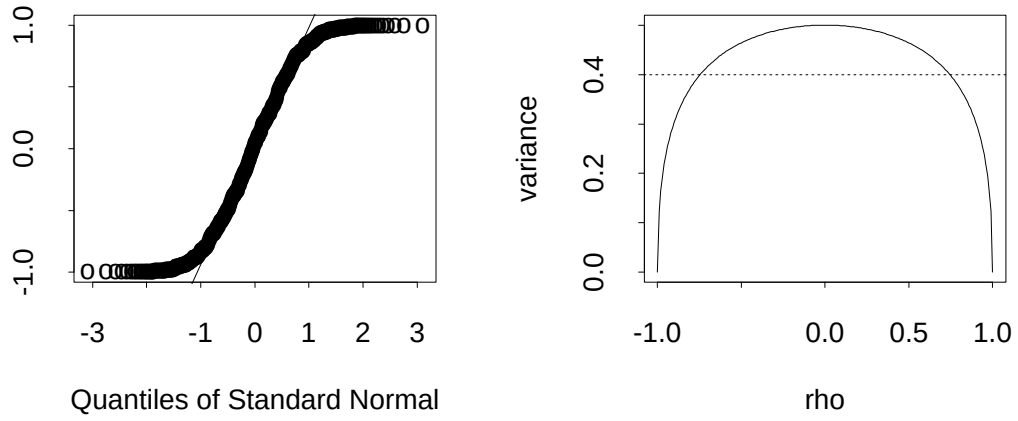

Figure 3: Left plot: the q-q plot of $\mathbf{f}^9$ arising from the Haar periodogram of a pure white noise process at scale $j = -1$ (against the normal quantiles). Right plot: solid line — the variance of $f_n^{\log_2(T)-1}$ against the correlation of the Gaussian variables involved; short-dashed line — variance $= 0.4$ (see text for further description).

Note the different asymptotic regimes for $\mathbf{V}^M$ and $\mathbf{Y}^M$: the multiplication of $\mathbf{Y}^M$ by $\sqrt{T}$ is needed because $\mathrm{Var}(Y_n^M) = O(2^M/T)$; we remind the reader that M is fixed. However, for the invertible case (see the discussion in Section 3), we require $M = \log_2(T)$. For a theoretical analysis of this case we need to be able to determine the distribution of the Haar-Fisz summands f_n^m on the finest scales. We have not been able to determine the general form of this distribution; however, we cast some light on the behaviour of $\mathcal{F}^{\log_2(T)}$ in Section 5.2.

5.2 Heuristics of the Haar-Fisz transform for $M = \log_2(T)$

In the asymptotic framework set out in Section 5.1, we assume that M is fixed, and therefore the length of the Haar-Fisz-transformed vector $\mathcal{F}^M\mathbf{I}$ is always constant and equal to 2^M, even though $T \to \infty$. This ensures the asymptotic Gaussianity of $\mathcal{F}^M\mathbf{I}$, in the sense specified by Theorem 5.1. However, to obtain an invertible operator, we need to set $M = \log_2(T)$. Simulations suggest that the asymptotic distribution of $\mathcal{F}^{\log_2(T)}$ is not exactly Gaussian, which is not surprising given the fact that the distribution of

$$f_n^{\log_2(T)-1} = \frac{I_{2n,T}^{(j)} - I_{2n+1,T}^{(j)}}{I_{2n,T}^{(j)} + I_{2n+1,T}^{(j)}}$$

(see the second example in Section 3.1) is far from Gaussian. To illustrate this statement, let us consider the Haar periodogram sequence $I^{(-1)}_{t,1024}$ of a pure white noise process ($T = 1024, J = 10$). The left plot in Figure 3 shows the q-q plot of the corresponding sequence $\mathbf{f}^9$ against the normal quantiles: its distribution strongly deviates from Gaussian in the tails. Other extensive simulations have shown that, for a wide range of processes, the distribution of $\mathbf{f}^M$ gets closer to Gaussianity as M decreases (as expected, see proof of Theorem 5.1).

However, even though taking $M = \log_2(T)$ (instead of keeping it fixed) spoils the exact asymptotic Gaussianity of $\mathcal{F}^M\mathbf{I}$, it does not seem to upset the other important theoretical property of $\mathcal{F}^M\mathbf{I}$: the variance stabilization. To illustrate this point, we consider the variance of the summand $f_n^{\log_2(T)-1}$ (for definition see formula (13)). Note that $f_n^{\log_2(T)-1}$ is always of the form $f_n^{\log_2(T)-1} = (\zeta_1^2-\zeta_2^2)/(\zeta_1^2+\zeta_2^2)$, where (ζ_1, ζ_2) is bivariate normal with mean $(0, 0)$. For simplicity, we assume that $\mathrm{Var}(\zeta_1) = \mathrm{Var}(\zeta_2)$, which is not a restrictive assumption: due to the local stationarity property, the two variances tend to the same limit as $T \to \infty$. Let $\rho = \mathrm{corr}(\zeta_1, \zeta_2)$. It can be shown that

$$\mathrm{Var}\left(f_n^{\log_2(T)-1}\right) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\left(\left(1-u^2\right)^2 - \left(\left(1-u^2\right)\rho + 2u\sqrt{1-\rho^2}\right)^2\right)^2}{(1+u^2)\left(\left(1-u^2\right)^2 + \left(\left(1-u^2\right)\rho + 2u\sqrt{1-\rho^2}\right)^2\right)^2} du.$$

The right plot in Figure 3 shows the graph of $\mathrm{Var}\left(f_n^{\log_2(T)-1}\right)$ against ρ. It can be seen that $\mathrm{Var}\left(f_n^{\log_2(T)-1}\right)$ is "stable" for a wide range of correlation values: indeed, the variance is between 0.4 and 0.5 for $|\rho| \leq 0.74$. This implies that while incorporating $\mathbf{f}^{\log_2(T)-1}$ spoils the asymptotic Gaussianity property of the Haar-Fisz transform, it helps achieve its variance stabilization property. Note, however, *in practice*, even the full invertible case $M = \log_2 T$ yields a distribution which is very close to variance stabilized Gaussian as Example 2 and simulations in Section 5.3 show.

A similar variance stabilization phenomenon occurs for $\mathbf{f}^M$ for $M < \log_2(T) - 1$.

5.3 Simulation

As an illustration of the Gaussianization and the variance stabilization properties of the Haar-Fisz transform, consider the process $X_{t,T} = \sigma(t/T)Y_{t,T}$, where $Y_{t,T} = \rho(t/T)Y_{t-1,T} + \varepsilon_t$ with $|\rho(z)| < 1$ and $\varepsilon_t \sim N(0, 1)$ i.i.d. It can easily be shown that the local autocovariance function for $X_{t,T}$ has the form

$$c(z, \tau) = \sigma^2(z)\frac{\rho(z)^\tau}{1-\rho(z)^2}$$

and, for $\beta_j(z)$ arising from the Haar periodogram, we have

$$\beta_j(z) = \sigma^2(z)\frac{1-\rho(z)^2 + 2^{j+3}\rho(z)^{2^{-j-1}+1} - 6\,2^j\rho(z) - 2^{j+1}\rho(z)^{2^{-j}+1}}{(1-\rho(z)^2)(1-\rho(z))^2}.$$

We consider the following two cases:

TVAR Model $\sigma^2(z) = 1$ and $\rho(z) = 1.8z - 0.9$, so that $X_{t,T}$ is a time-varying AR(1) process;

TMWN Model $\sigma^2(z)$ is a scaled Donoho & Johnstone (1995) `bumps` function with (min, max) values of (1/8, 8), and $\rho(z) = 0$, so that $X_{t,T}$ is a time-modulated white noise process.

In both of these models, we simulate 100 sample paths for both $T = 256$ and $T = 1024$. For each of the simulated sample paths, we compute the wavelet periodogram at scales $j = -1, \ldots, -\log_2(T)$. For each of the periodogram sequences $I^{(j)}_{t,T}$ obtained in this way, we compute the residuals $\mathcal{F}^M I^{(j)}_{t,T} - \mathcal{F}^M \beta_j(t/T)$ for $M = \log_2(T) - 2, \log_2(T) - 1, \log_2(T)$. We assess the Gaussianity of each sequence of residuals by looking at the p-value of the Kolmogorov-Smirnov statistic, returned by the S-Plus function `ks.gof`. For comparison, we also consider the residuals from the log transform: $\log(I^{(j)}_{t,T}) - \log(\beta_j(t/T))$.

Figure 4 shows that for $M = \log_2(T) - 2$, the proportion of p-values exceeding 5% is close to 95% for $j = -1, \ldots, -5$, so that residual sequences at these scales can be regarded as approximately Gaussian. However, even for $M = \log_2(T)$ the proportion of p-values exceeding 5% is incomparably larger than the same proportion computed for the log transform. Indeed, for $T = 1024$, no p-value exceeded the 5% threshold for the log transform.

The above experiment demonstrates that even for $M = \log_2(T)$ (the invertible case), the Haar-Fisz transform is a far better Gaussianizer than the log transform.

In practice, we often observe a degree of correlation in $\mathcal{F}^M \mathbf{I}^{(j)}$, particularly at coarser scales, i.e. for large negative j. This has to be taken into account when denoising Haar-Fisz transformed sequences by using suitable methods for correlated data as described later in Section 6.1.

6 Smoothing the wavelet periodogram

In this section, we first outline our general methodology for smoothing the wavelet periodogram of a Gaussian LSW process $X_{t,T}$, basing on a single stretch of observations. Then, we provide simulation results which demonstrate the effectiveness of our technique.

The generic algorithm consists of the following steps.

1. For each $j = -1, \ldots, -J(T)$, compute the raw wavelet periodogram $I^{(j)}_{t,T}$. In practice, this is done by taking the non-decimated wavelet transform of $X_{t,T}$ down to the level $-J(T)$, and then squaring the result. For computational convenience, we use periodic boundary treatment; another option would be to use e.g. symmetric boundary treatment.

2. For each $j = -1, \ldots, -J(T)$, take the Haar-Fisz transform of $I^{(j)}_{t,T}$ at a fixed resolution level $M \leq \log_2(T)$.

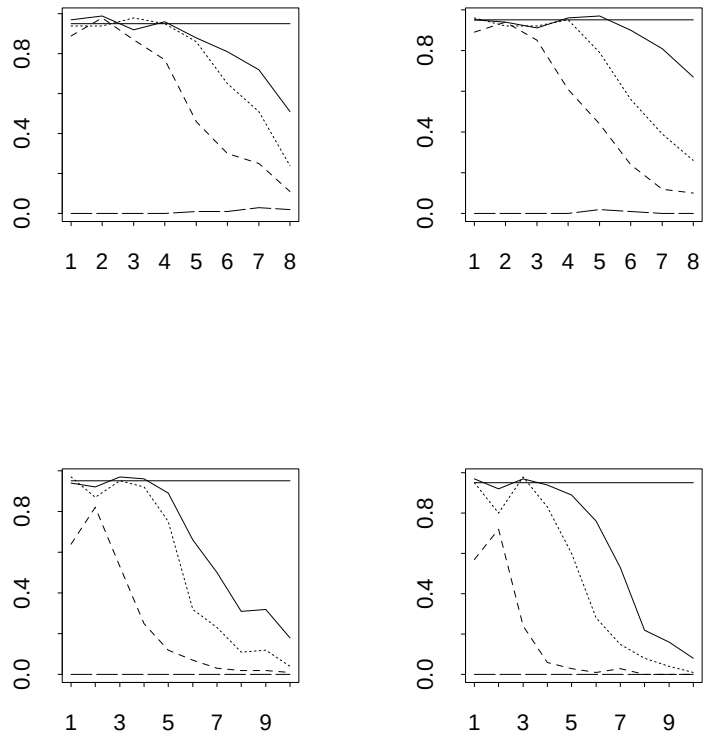

Figure 4: Proportion of p-values exceeding or equal to 5% against $-j$. Left column: results for model TVAR, right column: results for model TMWN. Top row: $T = 256$, Bottom row: $T = 1024$. Solid line: $M = \log_2(T)$, short-dashed line: $M = \log_2(T) - 1$, dashed line: $M = \log_2(T) - 2$, long-dashed line: the log transform. Horizontal solid line: 0.95.

3. For each j, denoise the Haar-Fisz transformed periodogram sequence using any wavelet denoising technique suitable for correlated Gaussian noise with constant variance. The wavelet denoising procedure employed at this stage may be of a translation-invariant (TI) type: we refer to TI-denoising at this stage as "internal" cycle-spinning (CS).

4. For each j, take the inverse Haar-Fisz transform of the denoised data.

5. If $M < \log_2(T)$, then for each j interpolate the estimates obtained in this way to the grid $\{t/T\}_{t=0}^{T-1}$ (so that they are of length T and not $2^M < T$). In our empirical investigation, we used simple linear interpolation. For each j, take the result to be an estimate of $\beta_j(z)$.

6. For a fixed integer S, let $s = 1, \ldots, S-1$. For each j, shift $I_{t,T}^{(j)}$ cyclically by s, smooth the shifted version using steps 2 – 5 of this algorithm, and shift back by s to obtain an estimate of $\beta_j(z)$. The CS at this stage is referred to as "external" cycle-spinning.

7. For each j, the *final estimate* of $\beta_j(z)$ is obtained by averaging over the estimates obtained through the S shifts.

A few remarks are in order.

Consistency. For considerations of consistency we have to let M grow with T: here we examine the case $M = \log_2(T)$. It is easy to see that our whole algorithm is consistent if in step 3 above we use linear Haar wavelet shrinkage. This means taking the Haar wavelet transform of the Haar-Fisz transformed **I**, annihilating the K finest scales and then inverting the Haar transform. Here $K \to \infty$ and $K = o(\log(T))$. The reason we have consistency in this simple case is that using linear Haar wavelet shrinkage causes the overall procedure to be *exactly* the same as: smooth **I** using the linear Haar scheme, i.e. perform a Haar decomposition of **I**, remove the K finest detail levels, invert the Haar transform.

In other words, because of the structure of the Haar and Haar-Fisz transforms, the non-linearities in our algorithm cancel out and it reduces to a linear procedure.We feel that it is worth mentioning that in this special case consistency occurs without needing a bias correction factor which is required in the $\log$ case.

However, if we replace the Haar wavelets in the wavelet shrinkage in step 3 by a different wavelet family, we immediately lose the correspondence of the two procedures and our algorithm becomes non-linear. This also happens if we replace the linear scheme in step 3 by non-linear wavelet shrinkage. In both cases theoretical verification of consistency is not straightforward. However, extensive computer experiments with a variety of Gaussian denoisers in step 3 systematically indicate MISE of the overall procedure decreasing to zero with increasing sample size.

Computational complexity. Steps 1–5 of the algorithm are each of computational order $O(TJ(T))$, provided that the wavelet denoising method used in step 3 has complexity $O(T)$. Therefore, the whole algorithm 1–7 is of computational order $O(STJ(T))$. In practice, the software is fast.

Use of wavelets. It is worth recalling here that, effectively, we use wavelets at four different stages of the smoothing procedure: First of all, a non-decimated wavelet system ψ is used in the construction of the LSW process $X_{t,T}$. Then the same system ψ is used to compute the wavelet periodogram $I^{(j)}_{t,T}$ in step 1 of the smoothing algorithm. Then the (inverse) Haar-Fisz transform in step 2 (4) relies on the Haar transform: thus, wavelets are used for the third time. Finally, we use wavelets (possibly a different family, say $\tilde{\psi}$) to denoise the Haar-Fisz transformed periodogram in step 3.

Cycle-spinning. Let $\mathcal{S}$ be the shift-by-one-operator from Nason & Silverman (1995). The Haar-Fisz transform is not translation-equivariant since $\mathcal{S}\mathcal{F}^M \neq \mathcal{F}^M\mathcal{S}$. Therefore, it is potentially beneficial to apply the external CS of step 6 even if step 3 uses internal CS. Another reason is that cycle-spinning can mitigate minor artefacts induced by the Haar part of the Haar-Fisz transform and improve visual quality. Cycle-spinning is a kind of basis averaging which arose in wavelet shrinkage in Coifman & Donoho (1995).

We now move on to describe our particular simulation setup.

6.1 Simulation

In this section, we describe the details of our simulation study which compares the performance of our Haar-Fisz smoothing algorithm with the original technique of Nason *et al.* (2000).

The “test processes" used in this section are the same as those in Section 5.3 again with 100 simulations: TVAR and TMWN. We consider the Haar periodogram of TVAR and TMWN, for sample paths of length 256 and 1024. In step 3 of the Haar-Fisz smoothing algorithm, we use non-TI level-dependent universal hard thresholding, appropriate for correlated Gaussian data as described in Johnstone & Silverman (1997). At this stage, we use Daubechies' Least Asymmetric wavelets with 4 vanishing moments, in both our algorithm and that of Nason *et al.* (2000).

Computational experiments suggest that for correlated noise, the choice of primary resolution (PR) is of utmost importance. (PR is a common concept in wavelet shrinkage and is defined as the resolution level j_0 such that wavelet coefficients at levels j_0 and coarser than j_0 are not denoised.) We do not choose the PR automatically (actually, we are unaware of any existing technique for performing automatic PR selection when the noise is correlated), but instead, we subjectively choose the PR for which the method of Nason *et al.* (2000) gives the most visually appealing results for the wavelet periodogram at the finest scale, i.e. $j = -1$. We also use the same PR in our algorithm. The particular values of the PR are: 7 for TMWN 1024, 6 for TMWN 256, 4 for TVAR 1024 and 3 for TVAR 256.

We use $S = 10$ external cycle-shifts. Using more shifts is likely to be beneficial in terms of MISE but is also more burdensome computationally and sometimes has the tendency to oversmooth. We only report results for $M = \log_2(T)$ (i.e. for the full invertible Haar-Fisz transform).

Figure 5 shows estimates of the local variance constructed from the estimates of the periodogram obtained using the two methods described above, for particular sample paths of TMWN 1024 and TVAR 1024. For both sample paths, our method achieves lower ISE.

Figure 6 shows, for each j, the differences between the logarithm of the ISE in estimating $\beta_j(z)$ for the method of Nason *et al.* (2000), and for our Haar-Fisz algorithm. The results are averaged over 100 simulated sample paths. Our algorithm is superior in most of the cases, except for the 4 finest scales in TVAR 256, and the 3 coarsest scales in TMWN 1024. A similar pattern has been obtained for other values of the PR.

We have also performed additional simulations for $M = \log_2(T) - 1$ and $M = \log_2(T) - 2$. It turned out that as long as the PR remained fixed, the choice of M had very little influence upon the estimates.

On a final note, it must be mentioned that other denoising methods can also be used in step 3, and our algorithm can only benefit from this flexibility. Some of the techniques for correlated data are reviewed in Opsomer *et al.* (2001). We have also experimented with the eBayes method of Johnstone & Silverman (2003) and Barber & Nason (2004) and obtained encouraging results.

7 Real data example: the Dow Jones index

In this section, we perform a local variance analysis of the Dow Jones Industrial Average series $D_{t,T}$, plotted in the top plot of Figure 7 ($T = 1024$) and obtained from:

`http://bossa.pl/notowania/daneatech/metastock/`

We used the following four methods to compute the local variance of $D_{t,T}$:

1. Our Haar-Fisz method of Section 6, based on the Haar periodogram, with the following parameters: $M = 10$, $S = 10$, step 3 applied non-TI level-dependent hard universal thresholding using Daubechies' Least Asymmetric wavelet with 4 vanishing moments. PR = 7.

2. Our Haar-Fisz method of Section 6, based on the Haar periodogram, with the following parameters: $M = 10$, $S = 10$, step 3 used the S-Plus spline smoothing routine `smooth.spline` with default parameters.

3. A modification of our Haar-Fisz method: instead of the sequences of the wavelet periodogram of $D_{t,T}$, the input to the Haar-Fisz algorithm was $D^2_{t,T}$. We took the smoothed version of $D^2_{t,T}$ to be an estimate of the local variance. The parameters of the Haar-Fisz algorithm were: $M = 10$, $S = 10$, step 3 used the S-Plus spline smoothing routine `smooth.spline` with default parameters.

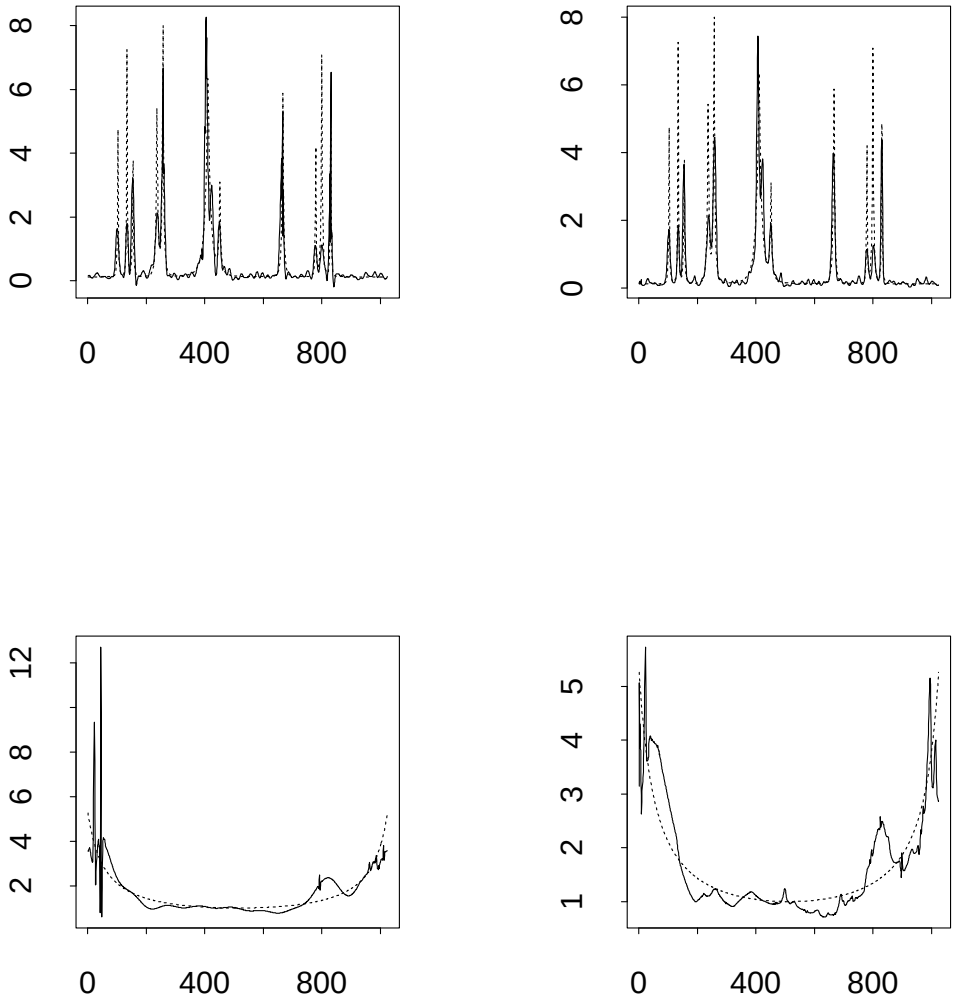

Figure 5: Solid lines: estimates of the local variances for $T = 1024$ in the TMWN model (top row), and the TVAR model (bottom row), using the method of Nason *et al.* (2000) (left column) and the Haar-Fisz algorithm (right column) as described in the text. Note that the vertical scales in the two bottom plots are different. Short-dashed lines: true local variances.

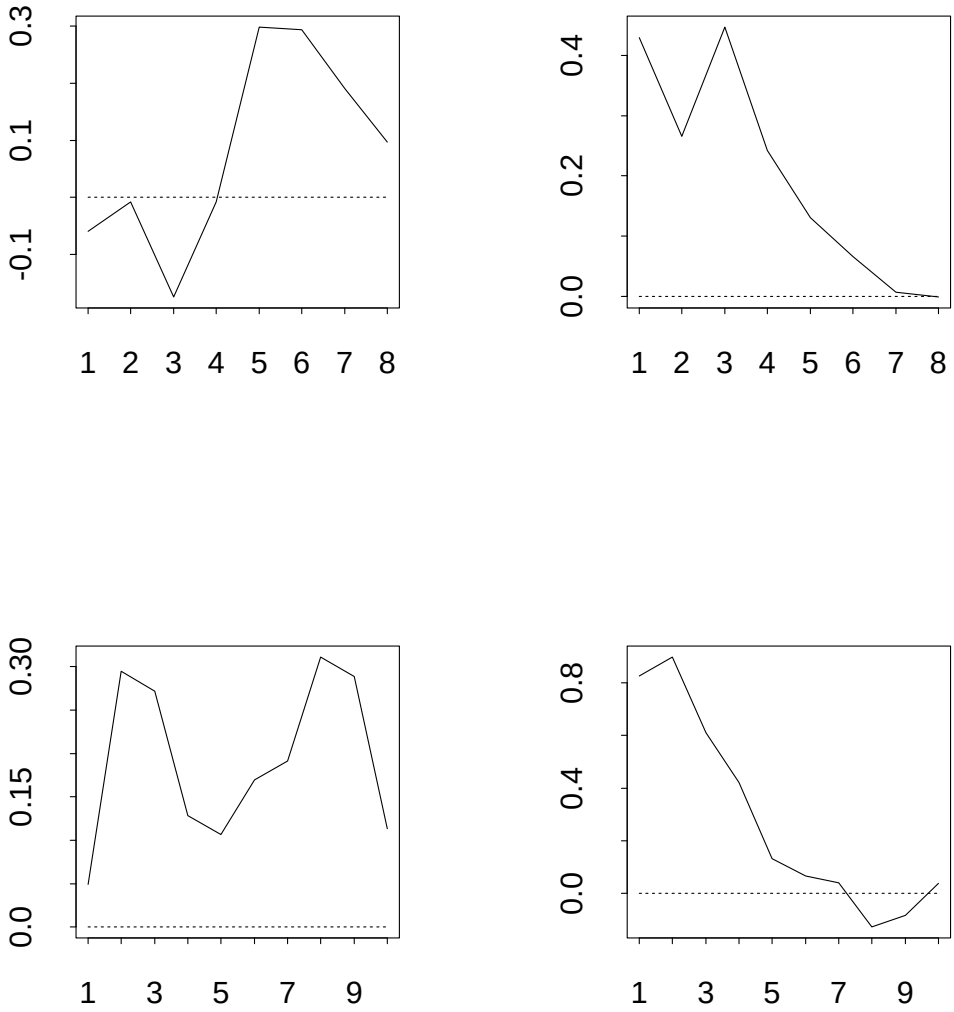

Figure 6: Solid line: difference between logged MISE for Nason *et al.* (2000) and for our Haar-Fisz algorithm (x-axis shows negative scale $-j$). Positive value means our algorithm does better. Left column: results for TVAR, right column: results for TMWN. Top row: $T = 256$, bottom row: $T = 1024$. Short-dashed line: zero.

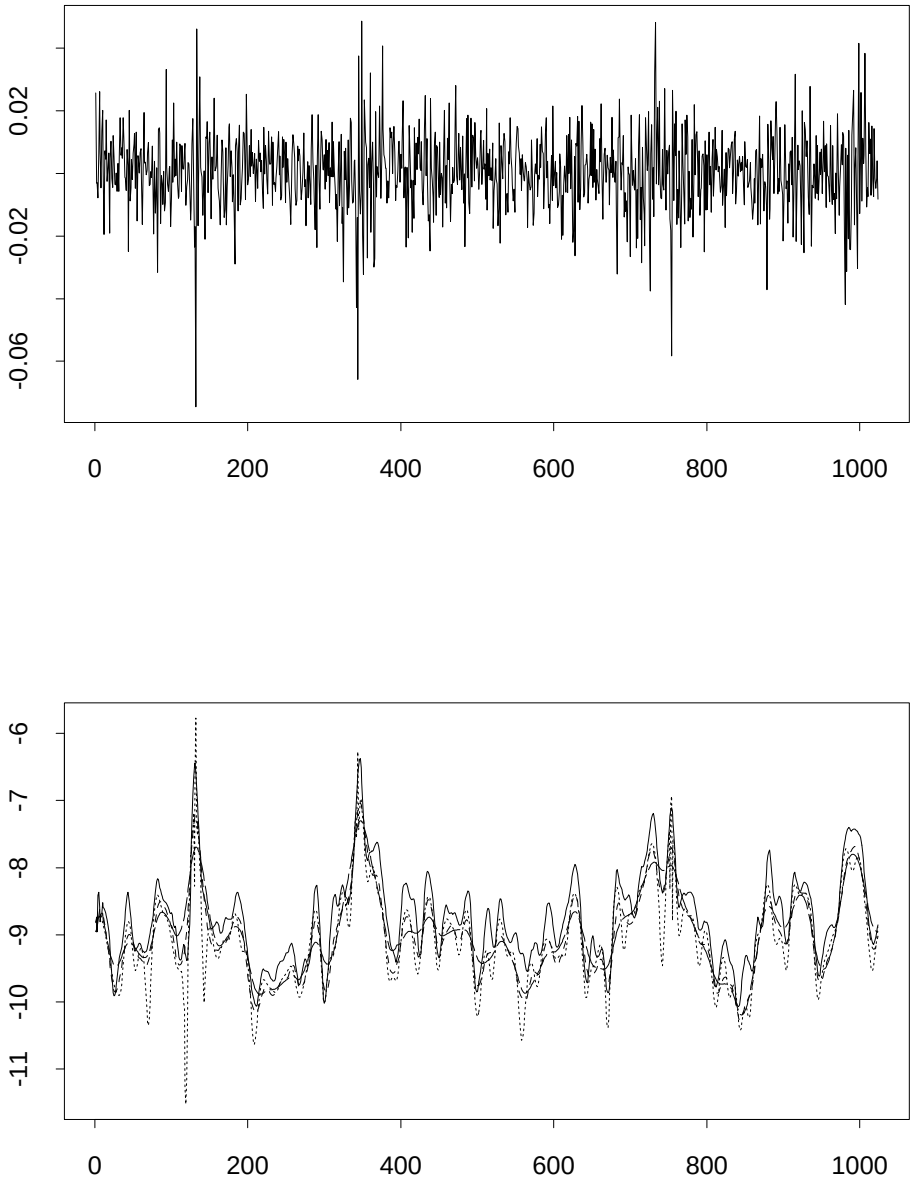

Figure 7: Top plot: log-returns on daily closing values of the Dow Jones Industrial Average. 1024 observations; the last one corresponds to 10/11 May 2001. Bottom plot: four estimates of the local variance of $D_{t,T}$ on a log scale. Solid line: method 1. Dashed line: method 2. Long-dashed line: method 3. Short-dashed line: method 4. See text for further description.

4. The method of Nason *et al.* (2000) with the following parameters: TI level-dependent universal hard thresholding using Daubechies' Least Asymmetric wavelet with 4 vanishing moments, PR $= 7$. The `smooth.dev` parameter in the `ewspec` routine (Nason (1998)) was set to `var`.

The results for PR $\neq 7$ were less convincing. The bottom plot in Figure 7 shows all four estimates plotted on a log scale. The two estimates based on spline smoothing show the least variability, the estimate 4 is the most variable, and the estimate 1 — the second most variable. Moreover, 1 estimates the variance at a slightly higher level than the other three.

One interesting question which can be asked is whether or not $D_{t,T}$ can be modelled as Gaussian. This can be examined, for example, by dividing $D_{t,T}$ by the square root of the estimates of the local variance, and looking at the distribution of the residuals. Figure 8 shows the `qqnorm` plot of the empirical quantiles of the residuals against the quantiles of the standard normal, for the four methods described above. The surprising observation is that all four plots consistently indicate that the upper tail is slightly platykurtic. However, there is no consistency in the assessment of the behaviour of the lower tail: here, 3 plots indicate platykurtosis, but the result of method 3 suggests slight leptokurtosis.

However, the p-values of the Kolmogorov-Smirnov test (returned by the S-Plus routine `ks.gof`) are large for each of the 4 sequences of residuals. In this sense, it can be concluded that the departure of $D_{t,T}$ from Gaussianity is insignificant.

This is in stark contrast to stationary nonlinear modelling (e.g. (G)ARCH or Stochastic Volatility), where, typically, the marginal distribution of financial log-returns is modelled as heavily leptokurtic.

8 Conclusions and further work

In this paper, we have introduced a Haar-Fisz variance-stabilizing transform for the wavelet periodogram (WP) of a Gaussian LSW process. The transform, performed in the wavelet domain by dividing the Haar detail coefficients of the WP by the corresponding smooth coefficients (an instance of the so-called Fisz transform), brings the distribution of the WP closer to normality, as well as stabilizing its variance. This makes the WP more amenable to standard denoising techniques which require stationary Gaussian noise. The computational complexity of the Haar-Fisz transform is linear in the number of data points, which is required to be a power of two.

In order to analyse theoretical properties of the Haar-Fisz transform in a certain asymptotic setting, we have formulated and proved a functional central limit theorem (FCLT) for the centred WP. Next, we have applied our FCLT to demonstrate the Gaussianizing, variance-stabilizing and decorrelating properties of the Haar-Fisz transform in the case where the length of the output vector remains constant as the length of the input vector goes to infinity.

Exact asymptotic Gaussianity does not hold if the length of the output vector of the Haar-Fisz transform matches the length of the input vector (which is the more interesting case in practice). However, we have provided some numerical evidence that the limiting

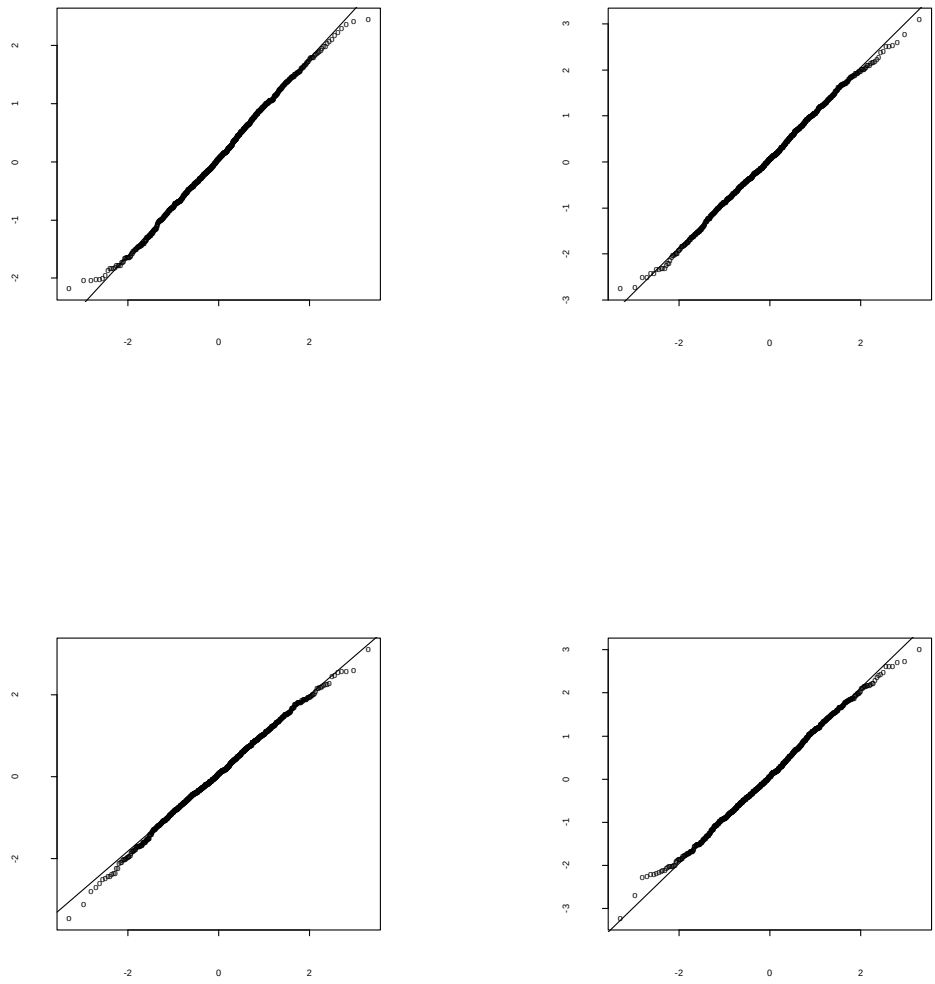

Figure 8: Empirical quantiles of the residuals of $D_{t,T}$ against the quantiles of the standard normal. Top left: method 1. Top right: method 2. Bottom left: method 3. Bottom right: method 4. See text for further description.

distribution is still not far from Gaussian, and that its variance is well stabilized. Extensive simulations have shown that even in this case, the Haar-Fisz transform is a far more effective Gaussianizer than the usual log transform.

Next, we have considered a smoothing algorithm for the WP, based on the Haar-Fisz transform. Theory has shown that the new algorithm is computationally fast, and simulation — that its MISE performance is better than that of the existing competitor.

Finally, several variants of the algorithm have been used to compute the local variance of the time series of daily log returns on the Dow Jones index. All of them consistently demonstrated that the series can be modelled as Gaussian.

Further work. An interesting avenue for future work would be to explore the possibility and utility of using the Haar-Fisz methodology for smoothing of the classical periodogram from stationary time series theory. Of course, it remains to be seen whether such a method could compete with the wide variety of excellent existing techniques, for example, Walden *et al.* (1998).

Another idea, suggested by a referee, would be whether other transforms could be used instead of the Haar wavelet transform within the Haar-Fisz algorithm.

It would also be interesting to show consistency for the more general cases, not just a Haar wavelet transform and a linear smoother as we have shown.

Free software and documentation to carry out the analyses can be found at

`http://www.ma.ic.ac.uk/~pzf/wavper/wavper.html`

8.1 Acknowledgment

Fryzlewicz gratefully acknowledges support by the ORS scheme and the University of Bristol. Nason gratefully acknowledges support from EPSRC Advanced Fellowship Grant GR/A01664/01. The authors would like to acknowledge helpful discussions with Anestis Antoniadis and Stuart Barber.

A Proofs

For this section we again refer the reader to Davidson (1994) for further technical background.

We first recall the definition of L_2-Near Epoch Dependence (L_2-NED). For a random variable X define $\|X\|_r = (\mathbb{E}|X|^r)^{1/r}$.

Definition A.1 *For a stochastic array $\{\{V_{t,T}\}_{t=-\infty}^{\infty}\}_{T=1}^{\infty}$, possibly vector-valued, on a probability space $(\Omega, \mathcal{G}, P)$, let $\mathcal{G}_{t-m,T}^{t+m} = \sigma(V_{t-m,T}, \ldots, V_{t+m,T})$. If an integrable array $\{\{X_{t,T}\}_{t=-\infty}^{\infty}\}_{T=1}^{\infty}$ satisfies*

$$\|X_{t,T} - \mathbb{E}(X_{t,T}|\mathcal{G}_{t-m,T}^{t+m})\|_2 \leq h_{t,T}\nu_m,$$

where $\nu_m \to 0$ as $m \to \infty$, and $\{h_{t,T}\}$ is an array of positive constants, it is said to be L_2-NED on $\{V_{t,T}\}$ with constants $\{h_{t,T}\}$. Further, if there exists $\varepsilon > 0$ such that

$\nu_m = O(m^{\lambda-\varepsilon})$, *then* $\{\{X_{t,T}\}_{t=-\infty}^{\infty}\}_{T=1}^{\infty}$ *is said to be* L_2*-NED of size* λ *on* $\{V_{t,T}\}$.

Lemma A.1 *Define* $Z_{t,T}^{(j)}$ *and* b_T *as in (16) and (19). Define* $\underline{\xi}_{t,T} = (\xi_{-1,t}, \ldots, \xi_{-J(T),t})'$. *If there exists*

$$\varepsilon > 0 \text{ such that } \left(\sum_{i<0} \sum_{l \geq m+1} \Psi_{i,j}^2(l) \bar{S}_i \right)^{1/2} = O(m^{-1/2-\varepsilon}),$$

then $Z_{t,T}^{(j)}/b_T$ *is* L_2*-NED of size* $-1/2$ *on* $\{\underline{\xi}_{t,T}\}$. *If in addition* $\delta_{J(T)}/T \in l_\infty$, *then the NED constants can be set to* $1/b_T$.

Proof. It suffices to examine the L_2-Near Epoch Dependence for $Z_{t,T}^{(j)}$. Define

$$\mathcal{G}_{t-m,T}^{t+m} = \sigma(\underline{\xi}_{t-m,T}, \ldots, \underline{\xi}_{t+m,T}).$$

Recall the definition of $I_{t,T}^{(j)}$ and $X_{t,T}$ from (6) and (2). We have

$$\begin{aligned}
Z_{t,T}^{(j)} - \mathbb{E}(Z_{t,T}^{(j)}|\mathcal{G}_{t-m,T}^{t+m}) &= I_{t,T}^{(j)} - \mathbb{E}(I_{t,T}^{(j)}|\mathcal{G}_{t-m,T}^{t+m}) \\
&= \left| \sum_{i=-1}^{-J(T)} \sum_{k} \omega_{i,k;T} \Psi_{i,j}(t-k) \xi_{i,k} \right|^2 \\
&- \left| \sum_{i=-1}^{-J(T)} \sum_{|k-t| \leqslant m} \omega_{i,k;T} \Psi_{i,j}(t-k) \xi_{i,k} \right|^2 \\
&- \sum_{i=-1}^{-J(T)} \sum_{|k-t|>m} \omega_{i,k;T}^2 \Psi_{i,j}^2(t-k) \\
&= Y_1^2 - Y_2^2 - K_1^2 = (Y_1 - Y_2)(Y_1 + Y_2) - K_1^2,
\end{aligned}$$

where Y_n^2 are random and K_1 is deterministic. Note that $Y_1 - Y_2$ and $Y_1 + Y_2$ are Gaussian and that $\mathbb{E}(Y_1 - Y_2)^2 = \mathbb{E}((Y_1 - Y_2)(Y_1 + Y_2)) = K_1^2$. Simple algebra yields

$$\mathbb{E}\left((Y_1 - Y_2)(Y_1 + Y_2) - K_1^2\right)^2 = 2K_1^2 \mathbb{E}(Y_1^2 + Y_2^2) \leq 4K_1^2 \mathbb{E}Y_1^2.$$

Noting that

$$\begin{aligned}
K_1^2 &\leq 2\left(1 + \frac{\delta_{J(T)}}{T}\right) \sum_{i=-1}^{-\infty} \sum_{l \geq m+1} \Psi_{i,j}^2(l) \bar{S}_i \\
\mathbb{E}Y_1^2 &\leq \left(1 + \frac{\delta_{J(T)}}{T}\right) \sum_{i=-1}^{-\infty} \sum_{l} \Psi_{i,j}^2(l) \bar{S}_i,
\end{aligned}$$

and recalling that $\delta_{J(T)}/T \in l_\infty$, the assertion of the Lemma follows with $v_m^2 := 4K_1^2\mathbb{E}Y_1^2$ and $h_{t,T}^2 = b_T^{-2}$, see Definition A.1. □

Lemma A.2 *If*

$$\sup_{z\in[0,1]} \sum_\tau |c(z,\tau)| \quad < \quad \infty \tag{26}$$

$$\textit{there exists } D \textit{ such that } \bar{S}_i 2^{-i} \quad \leq \quad D \quad \textit{for all } i, \tag{27}$$

then for fixed j

$$\frac{b_T^2}{T} \to 2\int_0^1 \sum_{\tau=-\infty}^{\infty} \left(\sum_i S_i(z) A_{i,j}^\tau\right)^2 dz$$

as $T \to \infty$.

Proof. All summations $\sum_i$ mean $\sum_{i=-1}^{-\infty}$. Using Gaussianity, we have

$$\begin{aligned}
b_T^2 &= 2\sum_{t=0}^{T-1}\sum_{\tau=-t}^{T-1-t}\left(\sum_{i=-1}^{-J(T)}\sum_k \omega_{i,k;t}^2 \Psi_{i,j}(t-k)\Psi_{i,j}(t+\tau-k)\right)^2 \\
&= 2\sum_{t=0}^{T-1}\sum_{\tau=-t}^{T-1-t}\left(\sum_{i=-1}^{-J(T)}\sum_k \left\{S_i\left(\frac{t}{T}\right) + O\left(\frac{C_i + L_i(t-k)}{T}\right)\right\}\right. \\
&\times \left.\Psi_{i,j}(t-k)\Psi_{i,j}(t+\tau-k)\right)^2 \\
&= 2\sum_{t=0}^{T-1}\sum_{\tau=-t}^{T-1-t}\left(\sum_i S_i\left(\frac{t}{T}\right) A_{i,j}^\tau\right)^2 + \text{Rest}_T,
\end{aligned}$$

where

$$\begin{aligned}
\text{Rest}_T &= 2\sum_{t=0}^{T-1}\sum_{\tau=-t}^{T-1-t}\left(\sum_{i=-1}^{-J(T)}\sum_k O\left(\frac{C_i + L_i(t-k)}{T}\right)\Psi_{i,j}(t-k)\Psi_{i,j}(t+\tau-k)\right. \\
&- \left.\sum_{i=-J(T)-1}^{-\infty} S_i\left(\frac{t}{T}\right) A_{i,j}^\tau\right) \\
&\times \left(\sum_{i=-1}^{-J(T)}\sum_k \left\{2S_i\left(\frac{t}{T}\right) + O\left(\frac{C_i + L_i(t-k)}{T}\right)\right\}\right. \\
&\times \left.\Psi_{i,j}(t-k)\Psi_{i,j}(t+\tau-k) + \sum_{i=-J(T)-1}^{-\infty} S_i\left(\frac{t}{T}\right) A_{i,j}^\tau\right).
\end{aligned} \tag{28}$$

Let us first show two simple auxiliary results.

1. Summability of constants C_i and L_i. We use the properties of A from Fryzlewicz *et al.* (2003) Lemma 8.

$$\begin{aligned}\sum_i (C_i + L_i(2^{-i} + 2^{-j}))A_{i,j} &= \sum_i (C_i + L_i 2^{-i}) 2^j 2^{-j} A_{i,j} \\ &+ \sum_i L_i 2^{-j} 2^i 2^{-i} A_{i,j} \\ &\leq 2^{-j} \sum_i (C_i + L_i 2^{-i}) \sum_k 2^k A_{i,k} \\ &+ 2^{-j} \sum_i L_i 2^{-i} \sum_k 2^k A_{k,j} \\ &= O(2^{-j}). \end{aligned} \tag{29}$$

2. Summability of covariance of wavelet coefficients.

$$\begin{aligned}\sum_\tau \left| \sum_i S_i(z) A_{i,j}^\tau \right| &= \sum_\tau \left| \sum_i S_i(z) \sum_n \Psi_i(n) \Psi_j(n+\tau) \right| \\ &= \sum_\tau \left| \sum_n c(z,n) \Psi_j(n+\tau) \right| \\ &\leq \sum_n |c(z,n)| \sum_\tau |\Psi_j(n+\tau)| \\ &\leq K_2 2^{-j} \sum_n |c(z,n)| \\ &= O(2^{-j}), \end{aligned} \tag{30}$$

by assumption (26) where K_2 is a constant. By formula (29) and assumption (27), we have

$$\begin{aligned}&\max_{t,\tau} \left| \sum_{i=-1}^{-J(T)} \sum_k O\left(\frac{C_i + L_i(t-k)}{T} \right) \Psi_{i,j}(t-k) \Psi_{i,j}(t+\tau-k) - \sum_{i=-J(T)-1}^{-\infty} S_i\left(\frac{t}{T}\right) A_{i,j}^\tau \right| \leq \\ &O(T^{-1}) \max_{t,\tau} \sum_i C_i + K_3 L_i (2^{-i} + 2^{-j}) \sum_k |\Psi_{i,j}(t-k) \Psi_{i,j}(t+\tau-k)| + \\ &+ \max_{t,\tau} \sum_k \sum_{i=-J(T)-1}^{-\infty} \bar{S}_i |\Psi_j(k)| \leq \\ &O(T^{-1}) \sum_i \left(C_i + K_3 L_i (2^{-i} + 2^{-j}) \right) A_{i,j} + O(2^{-j} T^{-1}) = O(2^{-j} T^{-1}), \end{aligned} \tag{31}$$

where K_3 is a constant.

Using first (31), and then (30) and (29), we bound (28) as follows

$$\begin{aligned}
\mathrm{Rest}_T &\leq O(2^{-j}T^{-1})\sum_{t=0}^{T-1}\sum_{\tau=-t}^{T-1-t}\Bigg|\sum_{i,k}\left\{3S_i\left(\frac{t}{T}\right)+O\left(\frac{C_i+L_i(t-k)}{T}\right)\right\} \\
&\times \Psi_{i,j}(t-k)\Psi_{i,j}(t+\tau-k)\Bigg| \\
&\leq O(2^{-j}T^{-1})\sum_{t=0}^{T-1}\sum_{\tau=-t}^{T-1-t}\left|\sum_i S_i\left(\frac{t}{T}\right)A_{i,j}^{\tau}\right| \\
&+ O(2^{-j}T^{-2})\sum_{t=0}^{T-1}\sum_{\tau=-t}^{T-1-t}\sum_i (C_i+L_i(2^{-i}+2^{-j}))A_{i,j} \\
&= O(2^{-2j})+O(2^{-2j}),
\end{aligned}$$

which yields the result. □

Proof of Theorem 4.1. We apply Theorem 29.14 from Davidson (1994), with

$$U_{t,T} = Z_{t,T}^{(j)}/b_T \tag{32}$$
$$c_{t,T} = 1/b_T \tag{33}$$
$$K_T(z) = \lfloor zT \rfloor, \tag{34}$$

where the LHS's of (32) – (34) use the notation from Davidson (1994), and the RHS's of these formulae use the notation from the article. We now check conditions (a) — (f) from Davidson (1994).

(a) Clearly, $\mathbb{E}Z_{t,T}^{(j)} = 0$.

(b) For Gaussian LSW processes, we have $\sup_{t,T}\|Z_{t,T}^{(j)}\|_r < \infty$ for $r > 2$.

(c) Satisfied by Lemma A.1 as $\{\underline{\xi}_{t,T}\}$ independent.

(d) Satisfied by Lemma A.2 as

$$\limsup_{T\to\infty}\frac{\sum_{t=\lfloor zT\rfloor}^{\lfloor(z+\omega)T\rfloor-1}\frac{1}{b_T^2}}{\omega} = \limsup_{T\to\infty}\frac{T}{b_T^2}.$$

(e) We clearly have $1/b_T = O(T^{1/2-1}) = O(T^{-1/2})$.

(f) Again by Lemma A.2, we have

$$\mathbb{E}R_T^2(z) = \frac{b_{\lfloor zT\rfloor}^2}{b_T^2} \to \eta(z).$$

This completes the proof. □

Proof of Theorem 5.1. Denote $Z_t = I_t - \mathbb{E}I_t$ and $\beta_j^\tau(z) = \sum_i S_i(z)A_{i,j}^\tau$. Note that $\beta_j^0(z) = \beta_j(z)$. Consider a single Haar-Fisz summand f_n^m, for $m \in \{0, 1, \ldots, M-1\}$ and $n \in \{0, 1, \ldots, 2^m - 1\}$. In what follows, $\gamma_{m,n}$ are appropriate integers and $\alpha_{m,n} \in \{0, 1\}$. We have

$$\begin{aligned}
(-1)^{\alpha_{m,n}} f_n^m &= (-1)^{\alpha_{m,n}} \frac{\sum_{t=\gamma_{m,n}T2^{-(m+1)}}^{(\gamma_{m,n}+1)T2^{-(m+1)}-1} I_t - \sum_{t=(\gamma_{m,n}+1)T2^{-(m+1)}}^{(\gamma_{m,n}+2)T2^{-(m+1)}-1} I_t}{\sum_{t=\gamma_{m,n}T2^{-(m+1)}}^{(\gamma_{m,n}+2)T2^{-(m+1)}-1} I_t} \\
&= (-1)^{\alpha_{m,n}} \frac{\sum_{t=\gamma_{m,n}T2^{-(m+1)}}^{(\gamma_{m,n}+1)T2^{-(m+1)}-1} Z_t - \sum_{t=(\gamma_{m,n}+1)T2^{-(m+1)}}^{(\gamma_{m,n}+2)T2^{-(m+1)}-1} Z_t}{\sum_{t=\gamma_{m,n}T2^{-(m+1)}}^{(\gamma_{m,n}+2)T2^{-(m+1)}-1} Z_t + \sum_{t=\gamma_{m,n}T2^{-(m+1)}}^{(\gamma_{m,n}+2)T2^{-(m+1)}-1} \mathbb{E}I_t} \\
&+ (-1)^{\alpha_{m,n}} \frac{\sum_{t=\gamma_{m,n}T2^{-(m+1)}}^{(\gamma_{m,n}+1)T2^{-(m+1)}-1} \mathbb{E}I_t - \sum_{t=(\gamma_{m,n}+1)T2^{-(m+1)}}^{(\gamma_{m,n}+2)T2^{-(m+1)}-1} \mathbb{E}I_t}{\sum_{t=\gamma_{m,n}T2^{-(m+1)}}^{(\gamma_{m,n}+2)T2^{-(m+1)}-1} Z_t + \sum_{t=\gamma_{m,n}T2^{-(m+1)}}^{(\gamma_{m,n}+2)T2^{-(m+1)}-1} \mathbb{E}I_t} \\
&= y_n^m + v_n^m.
\end{aligned}$$

Note that $\left\{\mathcal{F}^M \mathbf{I}\right\}_n = \sum_{m=0}^{M-1} (-1)^{\alpha_{m,n}} f_n^m = \sum_{m=0}^{M-1} y_n^m + v_n^m$. By Theorem 4.1 and Cramér's theorem (Davidson (1994), Theorem 22.14), we have

$$\begin{aligned}
\sqrt{T} y_n^m &= \frac{\sum_{t=\gamma_{m,n}T2^{-(m+1)}}^{(\gamma_{m,n}+1)T2^{-(m+1)}-1} Z_t - \sum_{t=(\gamma_{m,n}+1)T2^{-(m+1)}}^{(\gamma_{m,n}+2)T2^{-(m+1)}-1} Z_t}{b_T} \\
&\times (-1)^{\alpha_{m,n}} \frac{b_T\sqrt{T}}{\sum_{t=\gamma_{m,n}T2^{-(m+1)}}^{(\gamma_{m,n}+2)T2^{-(m+1)}-1} Z_t + \sum_{t=\gamma_{m,n}T2^{-(m+1)}}^{(\gamma_{m,n}+2)T2^{-(m+1)}-1} \mathbb{E}I_t} \\
&\xrightarrow{D} \left\{B_\eta((\gamma_{m,n}+2)2^{-m+1}) - 2B_\eta((\gamma_{m,n}+1)2^{-m+1}) + B_\eta(\gamma_{m,n}2^{-m+1})\right\} \\
&\times \frac{(-1)^{\alpha_{m,n}} 2^{1/2} \left(\sum_{\tau=-\infty}^{\infty} \int_0^1 (\beta_j^\tau(z))^2 dz\right)^{1/2}}{\int_{\gamma_{m,n}2^{-m+1}}^{(\gamma_{m,n}+2)2^{-m+1}} \beta_j(z) dz},
\end{aligned}$$

as $T \to \infty$. Denote the distributional limit by $\tilde{y}_n^m$. Set $Y_n^M = \sum_{m=0}^{M-1} y_n^m$ and $\tilde{Y}_n^M = \sum_{m=0}^{M-1} \tilde{y}_n^m$. Denote further $c_{(1)} = 2^{1/2} \left(\sum_{\tau=-\infty}^{\infty} \int_0^1 (\beta_j^\tau(z))^2 dz\right)^{1/2}$. We have

$$\begin{aligned}
&\sqrt{T} Y_n^M \xrightarrow{D} \tilde{Y}_n^M = \\
&\quad c_{(1)} \sum_{m=0}^{M-1} (-1)^{\alpha_{m,n}} \frac{B_\eta((\gamma_{m,n}+2)2^{-m+1}) - 2B_\eta((\gamma_{m,n}+1)2^{-m+1}) + B_\eta(\gamma_{m,n}2^{-m+1})}{\int_{\gamma_{m,n}2^{-m+1}}^{(\gamma_{m,n}+2)2^{-m+1}} \beta_j(z) dz},
\end{aligned}$$

as $T \to \infty$. It is immediate that $\mathbb{E}\tilde{Y}_n^M = 0$. We now look at the variance-covariance matrix

of $\tilde{\mathbf{Y}}^M$. We have

$$\mathrm{Var}(\tilde{Y}_n^M) =$$
$$c_{(1)}^2 \left(\sum_{m=0}^{M-1} \frac{\eta((\gamma_{m,n}+2)2^{-(m+1)}) - \eta(\gamma_{m,n}2^{-(m+1)})}{\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz\right)^2} + 2\sum_{m=0}^{M-1}\sum_{m'=m+1}^{M-1} (-1)^{\alpha_{m,n}+\alpha_{m',n}} \times \right.$$
$$\left\{ \frac{-2\eta(\rho_{m,n} \wedge \gamma_{m',n}2^{-(m'+1)}) + 4\eta(\rho_{m,n} \wedge (\gamma_{m',n}+1)2^{-(m'+1)})}{\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz \int_{\gamma_{m',n}2^{-(m'+1)}}^{(\gamma_{m',n}+2)2^{-(m'+1)}} \beta_j(z)dz} + \right.$$
$$\frac{-2\eta(\rho_{m,n} \wedge (\gamma_{m',n}+2)2^{-(m'+1)}) + \eta(\gamma_{m',n}2^{-(m'+1)}) - 2\eta((\gamma_{m',n}+1)2^{-(m'+1)})}{\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz \int_{\gamma_{m',n}2^{-(m'+1)}}^{(\gamma_{m',n}+2)2^{-(m'+1)}} \beta_j(z)dz} +$$
$$\left. \left. \frac{\eta((\gamma_{m',n}+2)2^{-(m'+1)})}{\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz \int_{\gamma_{m',n}2^{-(m'+1)}}^{(\gamma_{m',n}+2)2^{-(m'+1)}} \beta_j(z)dz} \right\} \right),$$

where $\rho_{m,n} = (\gamma_{m,n}+1)2^{-(m+1)}$.

Diagonal contribution. Let us first consider the diagonal contribution to $\mathrm{Var}(\tilde{Y}_n^M)$. We have

$$\frac{c_{(1)}^2}{2} \frac{\eta((\gamma_{m,n}+2)2^{-(m+1)}) - \eta(\gamma_{m,n}2^{-(m+1)})}{\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz\right)^2} = \frac{\sum_{\tau=-\infty}^{\infty} \int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} (\beta_j^\tau(z))^2 dz}{\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz\right)^2}. \tag{35}$$

By Cauchy inequality and the extended mean-value theorem, we have

$$\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz\right)^2 \le 2^{-m} \int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} (\beta_j(z))^2 dz =$$
$$2^{-m} \int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \sum_\tau (\beta_j^\tau(z))^2 dz \frac{\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} (\beta_j(z))^2 dz}{\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \sum_\tau (\beta_j^\tau(z))^2 dz} =$$
$$2^{-m} \int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \sum_\tau (\beta_j^\tau(z))^2 dz \frac{(\beta_j(\omega))^2}{\sum_\tau (\beta_j^\tau(\omega))^2},$$

where $\omega \in [\gamma_{m,n}2^{-(m+1)}, (\gamma_{m,n}+2)2^{-(m+1)}]$. This, combined with (35), gives

$$2^{m+1} \inf_{\omega\in[0,1]} \frac{\sum_\tau (\beta_j^\tau(\omega))^2}{(\beta_j(\omega))^2} \le c_{(1)}^2 \frac{\eta((\gamma_{m,n}+2)2^{-(m+1)}) - \eta(\gamma_{m,n}2^{-(m+1)})}{\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz\right)^2}. \tag{36}$$

To obtain the upper bound, note that there exists $\omega_1, \omega_2 \in [\gamma_{m,n}2^{-(m+1)}, (\gamma_{m,n}+2)2^{-(m+1)}]$ such that

$$\sum_{\tau=-\infty}^{\infty} \int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} (\beta_j^\tau(z))^2 dz \leq$$

$$2^m \left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz \right)^2 \frac{2^{-m} \int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} (\beta_j(z))^2 dz}{\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz \right)^2} \sup_{\omega\in[0,1]} \frac{\sum_\tau (\beta_j^\tau(\omega))^2}{(\beta_j(\omega))^2} =$$

$$2^m \left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz \right)^2 \frac{\beta_j(\omega_1)}{\beta_j(\omega_2)} \sup_{\omega\in[0,1]} \frac{\sum_\tau (\beta_j^\tau(\omega))^2}{(\beta_j(\omega))^2} \leq$$

$$2^m \left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz \right)^2 \left(1 + \frac{2^{-m} \sup_{\omega\in[0,1]} |\beta_j'(\omega)|}{\inf_{\omega\in[0,1]} \beta_j(\omega)} \right) \sup_{\omega\in[0,1]} \frac{\sum_\tau (\beta_j^\tau(\omega))^2}{(\beta_j(\omega))^2},$$

where β_j' is the one-sided derivative of β_j. The above, combined with (35), yields

$$\begin{aligned} & c_{(1)}^2 \frac{\eta((\gamma_{m,n}+2)2^{-(m+1)}) - \eta(\gamma_{m,n}2^{-(m+1)})}{\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz \right)^2} \\ \leq \; & 2^{m+1} \sup_{\omega\in[0,1]} \frac{\sum_\tau (\beta_j^\tau(\omega))^2}{(\beta_j(\omega))^2} + \frac{2\sup_{\omega\in[0,1]} |\beta_j'(\omega)|}{\inf_{\omega\in[0,1]} \beta_j(\omega)} \sup_{\omega\in[0,1]} \frac{\sum_\tau (\beta_j^\tau(\omega))^2}{(\beta_j(\omega))^2} \\ = \; & 2^{m+1} \sup_{\omega\in[0,1]} \frac{\sum_\tau (\beta_j^\tau(\omega))^2}{(\beta_j(\omega))^2} + O(1). \end{aligned} \tag{37}$$

Off-diagonal contribution. Two cases are possible: either $\rho_{m,n} \geq (\gamma_{m',n}+2)2^{-(m'+1)}$ or $\rho_{m,n} \leq \gamma_{m',n}2^{-(m'+1)}$. In either of the two cases, we have

$$\begin{aligned} & \Big| -2\eta\left(\rho_{m,n} \wedge \gamma_{m',n}2^{-(m'+1)}\right) + 4\eta\left(\rho_{m,n} \wedge (\gamma_{m',n}+1)2^{-(m'+1)}\right) + \\ & \quad -2\eta\left(\rho_{m,n} \wedge (\gamma_{m',n}+2)2^{-(m'+1)}\right) + \eta\left(\gamma_{m',n}2^{-(m'+1)}\right) - 2\eta\left((\gamma_{m',n}+1)2^{-(m'+1)}\right) + \\ & \quad \eta\left((\gamma_{m',n}+2)2^{-(m'+1)}\right) \Big| = \Big| \eta\left(\gamma_{m',n}2^{-(m'+1)}\right) - 2\eta\left((\gamma_{m',n}+1)2^{-(m'+1)}\right) + \\ & \quad \eta\left((\gamma_{m',n}+2)2^{-(m'+1)}\right) \Big| \leq 2^{-2m'-1} \sup_{\omega\in[0,1]} |\eta''(\omega)|, \end{aligned}$$

where the last inequality follows by the mean-value theorem and η'' denotes the one-sided derivative of η'. Using the above, and, again, the mean-value theorem, we bound the off-

diagonal contribution by

$$2c_{(1)}^2 \sum_{m=0}^{M-1} \frac{1}{\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz} \sum_{m'=m+1}^{M-1} 2^{-(m'+1)} \frac{2^{-m'} \sup_{\omega\in[0,1]} |\eta''(\omega)|}{\int_{\gamma_{m',n}2^{-(m'+1)}}^{(\gamma_{m',n}+2)2^{-(m'+1)}} \beta_j(z)dz} \leq$$

$$\frac{2c_{(1)}^2 \sup_{\omega\in[0,1]} |\eta''(\omega)|}{\inf_{\omega\in[0,1]} \beta_j(\omega)} \sum_{m=0}^{M-1} \frac{1}{\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz} \sum_{m'=m+1}^{M-1} 2^{-(m'+1)} \leq$$

$$\frac{2c_{(1)}^2 \sup_{\omega\in[0,1]} |\eta''(\omega)|}{\inf_{\omega\in[0,1]} \beta_j(\omega)} \sum_{m=0}^{M-1} \frac{2^{-(m+1)}}{\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz} \leq$$

$$\frac{c_{(1)}^2 \sup_{\omega\in[0,1]} |\eta''(\omega)|}{\inf_{\omega\in[0,1]} (\beta_j(\omega))^2} M = O(M). \tag{38}$$

Putting together (36), (37) and (38), we finally arrive at

$$\left(2^{M+1}-2\right) \inf_{\omega\in[0,1]} \frac{\sum_\tau (\beta_j^\tau(\omega))^2}{(\beta_j(\omega))^2} - O(M) \leq \mathrm{Var}(\tilde{Y}_n^M) \leq \left(2^{M+1}-2\right) \sup_{\omega\in[0,1]} \frac{\sum_\tau (\beta_j^\tau(\omega))^2}{(\beta_j(\omega))^2} + O(M). \tag{39}$$

Let us now consider $\mathrm{Cov}(\tilde{Y}_{n_1}^M, \tilde{Y}_{n_2}^M)$ for $n_1 \neq n_2$. Let $M' = \#\{m : \tilde{y}_{n_1}^m = \tilde{y}_{n_2}^m\}$. Let us first look at the case $M' > 0$. It is straightforward to show that

$$\mathrm{Cov}(\tilde{Y}_{n_1}^M, \tilde{Y}_{n_2}^M) =$$

$$\mathrm{Cov}\left(\sum_{m=0}^{M'-1} \tilde{y}_{n_1}^m + \tilde{y}_{n_1}^{M'} + \sum_{m=M'+1}^{M-1} \tilde{y}_{n_1}^m, \sum_{m=0}^{M'-1} \tilde{y}_{n_1}^m - \tilde{y}_{n_1}^{M'} + \sum_{m=M'+1}^{M-1} \tilde{y}_{n_2}^m\right) =$$

$$\mathrm{Var}\left(\sum_{m=0}^{M'-1} \tilde{y}_{n_1}^m\right) - \mathrm{Var}\left(\tilde{y}_{n_1}^{M'}\right) + \mathbb{E}\left(\sum_{m=0}^{M'-1} \tilde{y}_{n_1}^m \left(\sum_{m=M'+1}^{M-1} \tilde{y}_{n_1}^m + \sum_{m=M'+1}^{M-1} \tilde{y}_{n_2}^m\right) + \right.$$

$$\left. \tilde{y}_{n_1}^{M'} \left(\sum_{m=M'+1}^{M-1} \tilde{y}_{n_1}^m - \sum_{m=M'+1}^{M-1} \tilde{y}_{n_2}^m\right)\right).$$

The expectation can be shown to be $O(M)$ using the same methodology as for bounding the off-diagonal component of $\mathrm{Var}(\tilde{Y}_n^M)$. We will now show that $\mathrm{Var}(\sum_{m=0}^{M'-1} \tilde{y}_{n_1}^m) - \mathrm{Var}(\tilde{y}_{n_1}^{M'}) = O(M)$. We first quote two simple facts: let g be a continuous function with a bounded one-sided derivative over $[0,1]$ and let $[c,d] \subset [a,b] \subset [0,1]$. We have

$$\left| \int_a^b g(z)dz - \frac{b-a}{d-c} \int_c^d g(z)dz \right| \leq (b-a)^2 \sup_z |g'(z)| \tag{40}$$

$$\left| \left(\int_c^d g(z)dz\right)^2 - \left(\frac{d-c}{b-a}\right)^2 \left(\int_a^b g(z)dz\right)^2 \right| \leq (d-c)^2(b-a) \sup_z |(g^2(z))'| \tag{41}$$

For simplicity, denote $n = n_1$. Using again the same method as for bounding the off-diagonal component of the variance, we obtain

$$\mathrm{Var}(\sum_{m=0}^{M'-1}\tilde{y}_n^m) - \mathrm{Var}(\tilde{y}_n^{M'}) = O(M) + 2\sum_{m=0}^{M'-1}\frac{\sum_{\tau=-\infty}^{\infty}\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}}(\beta_j^\tau(z))^2dz}{\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}}\beta_j(z)dz\right)^2} -$$

$$\frac{2\sum_{\tau=-\infty}^{\infty}\int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}}(\beta_j^\tau(z))^2dz}{\left(\int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}}\beta_j(z)dz\right)^2} = O(M) + 2\times$$

$$\sum_{m=0}^{M'-1}\left(\frac{\sum_{\tau=-\infty}^{\infty}\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}}(\beta_j^\tau(z))^2dz}{\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}}\beta_j(z)dz\right)^2} - \frac{\sum_{\tau=-\infty}^{\infty}\int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}}(\beta_j^\tau(z))^2dz}{2^{M'-m}\left(\int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}}\beta_j(z)dz\right)^2}\right) +$$

$$\frac{2^{-M'+1}\sum_{\tau=-\infty}^{\infty}\int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}}(\beta_j^\tau(z))^2dz}{\left(\int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}}\beta_j(z)dz\right)^2}.$$

Consider a single component of the sum over m: it is a difference of two ratios which we denote here by $I - II$ to shorten the notation. We have $|I - II| \leq |I - III| + |III - II|$, where

$$III = \frac{2^{M'-m}\sum_{\tau=-\infty}^{\infty}\int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}}(\beta_j^\tau(z))^2dz}{\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}}\beta_j(z)dz\right)^2}.$$

Using (40), we get

$$|I - III| \leq \frac{2^{-2m}\sup_{\omega\in[0,1]}\left|\left(\sum_\tau(\beta_j^\tau(\omega))^2\right)'\right|}{\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}}\beta_j(z)dz\right)^2} \leq \frac{\sup_{\omega\in[0,1]}\left|\left(\sum_\tau(\beta_j^\tau(\omega))^2\right)'\right|}{\left(\inf_{\omega\in[0,1]}\beta_j(\omega)\right)^2}.$$

On the other hand, using (41) we have

$$|III - II| =$$

$$2^{M'-m} \left| \frac{\left(\int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}} \beta_j(z)dz \right)^2 - 2^{2(m-M')} \left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz \right)^2}{\left(\int_{\gamma_{m,n}2^{-(m+1)}}^{(\gamma_{m,n}+2)2^{-(m+1)}} \beta_j(z)dz \right)^2} \right| \times$$

$$\frac{\sum_{\tau=-\infty}^{\infty} \int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}} (\beta_j^\tau(z))^2 dz}{\left(\int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}} \beta_j(z)dz \right)^2} \leq$$

$$\frac{2^{-M'} \sum_{\tau=-\infty}^{\infty} \int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}} (\beta_j^\tau(z))^2 dz}{\left(\int_{\gamma_{M',n}2^{-(M'+1)}}^{(\gamma_{M',n}+2)2^{-(M'+1)}} \beta_j(z)dz \right)^2} \frac{\sup_{\omega\in[0,1]}((\beta_j(\omega))^2)'}{\left(\inf_{\omega\in[0,1]} \beta_j(\omega)\right)^2},$$

which is bounded by (39). This proves the assertion that $\mathrm{Var}(\sum_{m=0}^{M'-1} \tilde{y}_{n_1}^m) - \mathrm{Var}(\tilde{y}_{n_1}^{M'}) = O(M)$. Setting $V_n^M = \sum_{m=0}^{M-1} v_n^m$ completes the proof of the theorem. □

References

Barber, S., & Nason, G.P. 2004. Real nonparametric regression using complex wavelets. *J. Roy. Statist. Soc. B*, **66**, (to appear).

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

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

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

Coifman, R. R., & Donoho, D. L. 1995. Translation-invariant de-noising. *Lect. Notes Statist.*, **103**, 125–150.

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

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

Davidson, J. 1994. *Stochastic Limit Theory*. Oxford University Press.

Donoho, D. L., & Johnstone, I. M. 1995. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, **90**, 1200–1224.

Fisz, M. 1955. The limiting distribution of a function of two independent random variables and its statistical application. *Colloquium Mathematicum*, **3**, 138–146.

Fryzlewicz, P. 2002. Modelling and forecasting financial log-returns as locally stationary wavelet processes. *Technical Report 02:14, Department of Mathematics, University of Bristol*.

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

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

Johnstone, I. M., & Silverman, B. W. 1997. Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B*, **59**, 319–351.

Johnstone, I. M., & Silverman, B. W. 2003. Empirical Bayes selection of wavelet thresholds. *Submitted*.

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

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

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

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

Nason, G. P. 1998. *WaveThresh3 Software*. Available from `http://www.stats.bris.ac.uk/~wavethresh/`.

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

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

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

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

Opsomer, J.D., Wang, Y., & Yang, Y. 2001. Nonparametric regression with correlated errors. *Stat. Sci.*, **16**, 134–153.

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

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

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

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

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

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

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