Locally stationary wavelet packet processes: basis selection and model fitting.

Alessandro Cardinali*and Guy P. Nason†

November 21, 2016

Abstract

For nonstationary time series the fixed Fourier basis is no longer canonical. Rather than limit our basis choice to wavelet or Fourier functions, we propose the use of a library of non-decimated wavelet packets from which we select a suitable basis (frame). Non-decimated packets are preferred to decimated basis libraries so as to prevent information "loss" at scales coarser than the finest. This article introduces a new class of locally stationary wavelet packet processes and a method to fit these to time series. We also provide new material on the boundedness of the inverse of the inner product operator of autocorrelation wavelet packet functions. We demonstrate the effectiveness of our modelling and basis selection on simulated series and Standard and Poor's 500 index series.

Keywords: Local stationarity; wavelet packet; locally stationary Fourier process; locally stationary wavelet process.

1 M.B. Priestley: Giant of Time Series

Today, it is clear that Maurice Priestley's fascinating, lucid and encyclopædic body of work was way ahead of its time. Certainly, his path-breaking work on nonstationary time series provides the basis for a great deal of the academic work carried out today: that in the scientific literature, that contained in modern software packages and through into applications. Many were privileged to benefit from scientific interlocution with Professor Priestley. In our case, this communication was about the then new field of wavelets, but also about the 'oscillatory process' idea which was, and is, a key inspiration to all working in nonstationary time series.

This article is focused on a, maybe, little-explored part of Priestley's panoply which can be summarized by the following quote from Priestley (1983) page 822 which refers to representations for nonstationary processes:

> "Parzen (1959) has pointed out that if there exists a representation $X(t) = \int \phi_t(\omega) dZ(\omega)$, then there is a *multitude of different representations of the process*, each representation based on a different family of functions."

and

> "The situation is in some ways similar to the selection of a basis for a vector space."

and

> "However, if the process is non-stationary this choice [*complex exponential family*] of functions is no longer valid."

*University of Plymouth. Email: alessandro.cardinali@plymouth.ac.uk
†University of Bristol. Email: g.p.nason@bristol.ac.uk

The italics are ours. Probably, it is the case that, since those statements were published, apart from Priestley's abstract work on oscillatory processes, there has been little in the literature that diverges from local Fourier representations. This changed for *representations* in the early 2000s by the introduction of the locally stationary wavelet processes in Nason et al. (2000) and an early foray into statistical time series using Smooth Localized complex EXponentials (SLEX) basis *libraries* (the first multitude?) Ombao et al. (2001), Ombao et al. (2002) and Ombao et al. (2005). The representation question might seem abstract but it is, we believe, becoming of increasing practical importance.

Perhaps the best place where the benefits of the 'multitude' can be seen is in the area of stationarity testing. Here too Priestley was a pioneer, constructing one of the first practical tests in the elegant Priestley and Subba Rao (1969). Since then, there have been several excellent tests using Fourier-based quantities but also, more recently, tests based on eliciting nonstationarities using wavelets in Cardinali and Nason (2010), Nason (2013) and Walsh functions in Jin et al. (2015). With 'a multitude of representations' as possible underlying models for nonstationary time series one should be able to benefit from using SLEX libraries, or other libraries such as wavelet packets (WPs), for testing stationarity. For example, Cardinali and Nason (2016) demonstrate the benefits of using wavelet packets for stationarity testing using a mixture of theoretical and empirical arguments. The benefits of the 'multitude' arise because the diversity of basis functions in libraries permit discovery of structure that one basis alone cannot detect.

This article extensively elaborates on a suggestion in Nason et al. (2000), page 276, to use nondecimated wavelet packets (a basis library), and not just wavelets (basis), for *modelling* of nonstationary time series and not only testing for stationarity as in Cardinali and Nason (2016). A basis library is a collection of bases. The article below is deliberately computational: we postulate some potentially useful models, fit them using computational methods, obtain some useful illustrations via simulation and an analysis of the Standard and Poor's 500 Index. Our aim is to raise the profile further of the 'multitude' and stimulate future research in this area, not least in terms of further expanding the mathematical underpinning.

2 Introduction

If a time series is stationary then classical (Fourier) theory provides optimal and well-tested means for its analysis. Indeed, if the series, $X_t, t \in \mathbb{Z}$, is stationary then it is *required* by theory to possess the following well-known decomposition:

$$X_t = \int_{-\pi}^{\pi} A(\omega) \exp(i\omega t) d\xi(\omega), \tag{1}$$

where $d\xi(\omega)$ is a zero-mean orthonormal increments process and $A(\omega)$ is the amplitude function (for a process with absolutely continuous spectral distribution function, see Priestley (1983) §4.11, for example). There are several beautiful proofs that establish that the Fourier representation is the canonical one in the stationary situation. See, for example the nice expositions in Hannan (1960) and Priestley (1983) §4.11. We are interested in the case where X_t might be locally stationary (LS), that is, over short periods of time the series appears to be stationary but it *can* change its statistical properties slowly over (longer periods of) time. The concept of nonstationary time series has been appreciated for many years. The theory of nonstationary processes was significantly advanced by a series of papers by M.B. Priestley and co-authors from the mid 1960s notably the RSS Read Paper: Priestley (1965). A rigorous asymptotic framework for local stationarity modelling was introduced in Dahlhaus (1996a, 1997) within a framework that we call *locally stationary Fourier* processes.

Remark 1 (Rescaled time asymptotics and locally stationary Fourier processes) *The* locally stationary Fourier *model from Dahlhaus (1997) is a (triangular array of) stochastic*

process(es) represented by:

$$X_{t;T} = \int_{-\pi}^{\pi} \exp(i\omega t) A^0_{t,T}(\omega) d\xi(\omega), \tag{2}$$

where $d\xi(\omega)$ is a zero-mean orthonormal increment process. The transfer function $A^0_{t,T}$ satisfied $\sup_{t,\omega} |A^0_{t,T}(\omega) - A(t/T, \omega)| \leq KT^{-1}$, for all T, for some constant K and 2π-periodic function $A : [0,1] \times \mathbb{R} \to \mathbb{C}$ satisfying $A(u, -\omega) = \overline{A(u,\omega)}$ and $A(u,\omega)$ is a continuous function in $u \in [0,1]$. The quantity $u = t/T$ was called rescaled time and $\omega \in (-\pi, \pi)$ the frequency. This definition permits the uniform convergence $A^0_{t,T}(\omega) \to A(z,\omega)$ to be well defined and therefore allows meaningful asymptotics for the locally stationary spectra. When the function $A_{t,T}(\omega)$ is constant with respect to t, then the locally stationary Fourier process becomes stationary. (Dahlhaus's definition is more detailed with more technical conditions that we omit here).

Most locally stationary representations, up to and including Dahlhaus (1997), rely on the Fourier basis to furnish 'oscillation'. One of the key messages that we wish to emphasize is that for nonstationary processes the Fourier basis is no longer canonical. Silverman (1957) remarked on this predominance of "harmonizable processes". However, Priestley (1988) (and others) already explicitly considered the possibility of using oscillatory functions other than Fourier for the purpose of basis representation and this observation constitutes one of the main inspirations for the current article. For example, Nason et al. (2000) address this by introducing locally stationary time series models based on wavelets that they call *locally stationary wavelet* processes. The remainder of this section focuses on process definitions; more explicit definitions of the underlying bases of oscillatory functions is provided in the next section.

Remark 2 (Wavelets and locally stationary wavelet processes) *The* locally stationary wavelet *model from Nason et al. (2000) represents the process (array) by:*

$$X_{t;T} = \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} w_{j,k;T}\, \psi_{j,k}(t)\, \xi_{j,k}, \tag{3}$$

where $\{\xi_{j,k}\}$ is a collection of zero mean uncorrelated random variables, the vectors $\{\psi_{j,k}(t)\}$ is a set of non-decimated discrete Daubechies wavelets (defined below in (6)) and $\{w_{j,k;T}\}$ is a set of amplitudes. The amplitudes have further technical conditions imposed on them, but they are analogues of the quantities in the Fourier representation in (2): $\{\psi_{j,k}(t)\}$ is the analogue of $\{\exp(i\omega t)\}$, $\{\xi_{j,k}\}$ the analogue of $d\xi(\omega)$ and $\{w_{j,k;T}\}$ the analogue of $\{A_{t,T}(\omega)\}$. As with the locally stationary Fourier model the amplitudes $w_{j,k;T}$ are closely related to an amplitude function $W_j(t/T)$ and an underlying rescaled time asymptotic model. The evolutionary wavelet spectrum $S_j(z) = W_j(z)^2$. When $w_{j,k;T}$ is a constant function of k or, equivalently, $S_j(z), W_j(z)$ are constant functions of z, then the associated locally stationary wavelet processes are second-order stationary.

The locally stationary wavelet framework has also been successfully used to model multivariate time series as in Sanderson et al. (2010) and Park et al. (2014) and references therein. However, rather than limit the choice to wavelet or Fourier bases, a further alternative would be to select a basis from an overcomplete set of alternatives that is commonly referred to as *basis library*. The benefits of basis libraries in statistical time series modelling were first realized by Ombao et al. (2001), Ombao et al. (2002) and Ombao et al. (2005) who used the SLEX functions from Wickerhauser (1994) as follows.

Remark 3 (Basis libraries and SLEX processes) *Ombao et al. (2002) introduce the lo-*

cally stationary SLEX processes $\{X_{t;T}\}_{t=1,\ldots,T}$ *by*

$$X_{t;T} = \sum_{i:\cup S_i \sim B_T} M_i^{-1/2} \sum_{k_i=-M_i/2+1}^{M_i/2} \theta_{S_i,k_i,T}\, \overline{\phi_{S_i,\omega_{k_i}}}(t)\, z_{S_i,k_i}, \tag{4}$$

where B_T *is the* SLEX basis*, an adaptive dyadic segmentation of the time interval* $\{0, 1, \ldots, T-1\}$*,* $S_i \in B_T$ *are segments (of time points) from the SLEX basis,* $M_i = |S_i|$ *is the length of the segment* S_i*,* $k_i = -M_i/2+1, \ldots, M_i/2$*,* $\omega_{k_i} = k_i/M_i$ *are radiant frequencies,* $\theta_{S_i,k_i,T}$ *and* z_{S_i,k_i} *are, respectively, amplitude coefficients and random increments for given time-segments and frequencies. For* $\alpha_i = \min(S_i)$*, the SLEX basis functions* $\phi_{S_i,\omega_{k_i}}(t)$ *from Wickerhauser (1994) have the form:*

$$\phi_{S_i,\omega_{k_i}}(t) = \Psi_+\left(\frac{t-\alpha_i}{M_i}\right)\exp\left\{2\pi i\omega_{k_i}(t-\alpha_i)\right\} + \Psi_-\left(\frac{t-\alpha_i}{M_i}\right)\exp\left\{-2\pi i\omega_{k_i}(t-\alpha_i)\right\}, \tag{5}$$

where $\Psi_+(t), \Psi_-(t)$ *are two specially constructed smooth real-valued window functions. The SLEX time (time-block)-varying transfer function can be computed as the inner product of* $X_{t;T}$ *with the respective SLEX basis function.*

Ombao et al. (2001) use this system for adaptive segmentation of a time series into piecewise stationary processes and for spectral smoothing. Ombao et al. (2002) introduces the process, estimation theory and shows asymptotic equivalence to the Dahlhaus locally stationary Fourier model. Ombao et al. (2005) extends the idea to multivariate time series. Another example is Donoho et al. (2003) which is concerned with locally stationary covariance estimation using penalized basis methods. A general review of locally stationary time series models can be found in Nason and von Sachs (1999) and Dahlhaus (2012), see also Cardinali and Nason (2008) for an additional recent set of references. There are many possible models ("the multitude") and not much is known about how the respective process classes of locally stationary Fourier, locally stationary wavelet, SLEX and our model overlap. From a more theoretical standpoint these different models correspond to different tilings of the time-frequency plane and, hence, have different characteristics in analysis mode extracting often very different aspects of information from a time series. This article proposes the use of the overcomplete dictionary of non-decimated wavelet packets from which we select a suitable basis. Non-decimated packets are preferred to decimated basis libraries so as to prevent information "loss" at scales coarser than the finest. Therefore, this article introduces the new class of locally stationary wavelet packet (LSWP) processes and a method to successfully fit these to time series data. We propose a complete framework for process representation and inference for the associated time-frequency spectra and we provide theoretical results concerning the existence of an asymptotically unbiased spectral estimator in this setting.

A key conceptual difference between the SLEX model above and our wavelet packet models later is that we use *non-decimated wavelet packet* (NDWP) basis functions. For process representation and spectral estimation of many processes we surmise that probably both work similarly but SLEX, in not being non-decimated will possibly be more computationally efficient for some processes. However, for other processes, especially for finite T, the non-decimation can pick up structure that SLEX might miss. Although widely referred to in the signal processing literature, wavelet packets have not, until now, been extensively used within statistical time series. Exceptions using the non-decimated version are Walden and Contreras Cristan (1998), §6 of Percival and Walden (2000), Nason et al. (2001), Nason and Sapatinas (2002), Gabbanini et al. (2004), Cardinali (2009), Milne et al. (2009), Yang et al. (2009) and Garcia et al. (2013).

There appears to be a misconception about locally stationary processes that use non-decimated transforms e.g. Ombao et al. (2002) p. 173 that claims that it is not straightforward to simulate realizations. On the contrary Cardinali and Nason (2008) mention `LSWsim`, a

fast $\mathcal{O}(T \log T)$ function, that simulates any locally stationary wavelet process. Similar fast functions have been constructed for our current work involving packets with the same order of computation as the fast Fourier transform which SLEX makes use of.

Section 3 provides a quick review of wavelets, wavelet packets and basis libraries and introduces the relevant notation. Section 4 presents our modelling framework and the relevant estimators for a fixed basis eventually selected from an overcomplete basis library. Section 5 illustrate our methodology to select an appropriate basis from a wavelet packet dictionary. Section 6 presents simulations of several LSWP processes for which we assess the finite sample performances of our basis selection method. Section 7 presents an application of our inferential methods on S&P 500 returns and Section 8 concludes outlining directions for future work. Proofs of the main theoretical results are deferred to the appendix.

3 Wavelets, Wavelet Packets and Basis Libraries

Wavelets are locally supported functions that can be used to decompose signals across scales using localized time-scale coefficients. The calculation of such coefficients is often performed by means of the Mallat (1989) Discrete Wavelet Transform, see Daubechies (1992), Percival and Walden (2000) or Nason (2008) for alternative accounts. Wavelets can be used as building blocks for a wide variety of non-smooth signals, in situations where the Fourier functions would not be suited. There are many wavelets one might use. Daubechies (1992) provides an introduction to the mathematical foundation of wavelets, including the *Least Asymmetric* (LA) bases, which were the first compactly supported wavelets designed to be quasi-symmetric. As often emphasized, wavelets have a gender, that is the *father* wavelet is built from a low pass linear filter designed to provide a local signal approximation, whereas the *mother* wavelet is built from a high pass filter identifying the local signal variation. The mother and father wavelets can be dilated and translated in order to form a location-scale family which is used to produce a multiresolution approximation for functions. From the mother wavelet $\psi(t)$ we can form daughters $\psi_{j,k}(t) = 2^{-j/2}\psi\{2^{-j}(t - 2^j\ k)\}$ for translates $k \in \mathbb{Z}$ and scale parameter $j \in \mathbb{Z}$. For suitable choices of mother wavelet the system $\{\psi_{j,k}(t)\}_{j\in\mathbb{Z},k\in\mathbb{Z}}$ can become an orthonormal basis for functions $f \in L^2(\mathbb{R})$ for example. For non-decimated wavelets the $2^j k$ is replaced by k and then we obtain a system of functions able to provide useful representations, but no longer orthogonal. Possibly the simplest example of a mother wavelet is the Haar wavelet defined by $\psi(t) = -2^{-1/2}$ for $t \in (0, 1/2)$, $2^{-1/2}$ for $t \in (1/2, 1)$ and zero elsewhere. However, to build *discrete time* time series we use discretized versions of wavelets as described next. Nason et al. (2000) introduced discrete non-decimated wavelets designed to represent discrete time series $X_t, t \in \mathbb{Z}$ as in (3). These are derived using the same $\{h_k\}_k$ and $\{g_k\}_k$ low and high pass quadrature mirror (finite impulse response) filters that Daubechies (1992) used to build her compactly supported continuous time wavelets. For example, for Haar wavelets $h_0 = h_1 = g_0 = 2^{-1/2}$ and $g_1 = -2^{-1/2}$. At each scale $j \geq 1$ the associated discrete non-decimated wavelets $\psi_j = (\psi_{j,0}, \ldots, \psi_{j,N_j-1})$ are vectors with up to N_j coefficients defined by

$$\begin{aligned} \psi_{1,k} &= \sum_n g_{k-2n}\delta_{0,n} = g_k, \qquad k = 0, \ldots, N_1 - 1, \\ \psi_{j+1,k} &= \sum_n h_{k-2n}\psi_{j,n}, \qquad k = 0, \ldots, N_{j-1} - 1, \end{aligned} \tag{6}$$

where, $\delta_{0,n}$ is the Kronecker delta, $N_j = (2^j - 1)(N - 1) + 1$ and N is the length of the filters $\{h_k\}$. At heart, the discrete wavelet vectors, ψ_j, are oscillatory replacements of the Fourier vectors $\exp(i\omega t)$ both of which satisfy various internal orthogonality conditions. In the locally stationary wavelet process representation (3) the notation $\psi_{j,k}(t)$ actually refers to the basis (vector) element $\psi_{j,k-t}$.

3.1 Wavelet Packets

Wavelet packets are an extension of wavelets whose basis functions, $\psi_{j,i,k}(t)$, depend upon an additional parameter i that measures the number of oscillations of the function. The oscillation parameter i can take values ranging from 0 to $2^j - 1$ for each scale $j = 1, 2, \ldots, J$. See Wickerhauser (1994) or Percival and Walden (2000) for more details. For discrete time series representation we make use of discrete non-decimated wavelet packets, defined next.

Definition 1 (Discrete non-decimated wavelet packets) *Discrete non-decimated wavelet packets (NDWP) are constructed as in* (6) *except that the* $\{g_k\}$ *and* $\{h_k\}$ *can both be replaced by either of* $\{g_k\}$ *or* $\{h_k\}$ *at each scale* j. *At each scale* $j \geq 1$ *and for* $i = 0, \ldots, 2^j - 1$, *the associated discrete wavelet packets* $\psi_{j,i} = (\psi_{j,i,0}, \ldots, \psi_{j,i,N_j-1})$ *are vectors with* N_j *coefficients defined by*

$$\begin{aligned}
\psi_{1,0,k} &= \sum_n h_{k-2n}\delta_{0,n} = h_k, \qquad k = 0, \ldots, N_1 - 1, \\
\psi_{1,1,k} &= \sum_n g_{k-2n}\delta_{0,n} = g_k, \qquad k = 0, \ldots, N_1 - 1, \\
\psi_{j+1,2i,k} &= \sum_n h_{k-2n}\psi_{j,i,n}, \qquad k = 0, \ldots, N_{j-1} - 1, \\
\psi_{j+1,2i+1,k} &= \sum_n g_{k-2n}\psi_{j,i,n}, \qquad k = 0, \ldots, N_{j-1} - 1, \qquad (7)
\end{aligned}$$

where N_j *and* N *are as in* (6). *A wavelet packet* $\psi_{J,i}$ *is also written in short form as* (J, i). *The value of* i *can be obtained by constructing a binary number with 0/1 appearing at position* $j = 1, \ldots, J$ *depending on whether filtering* h_{k-2n} *or* g_{k-2n} *is applied at stage* j *using either the third or fourth equation in* (7).

See Example 1 for an example of the construction. As with wavelet vectors (above) the notation $\psi_{j,i}(t-k)$ actually refers to the element $\psi_{j,i,t-k}$, Figure 1 shows some examples of wavelet packet basis functions derived from two different mother wavelets. The second column in each row corresponds to the wavelet, the other columns correspond to other packets which offer greater oscillatory flexibility compared with just using wavelets alone.

Remark 4 (Frequency coverage) *Hess-Nielsen and Wickerhauser (1996) p. 525 consider*

> *"an abstract two-dimensional signal representation in which time and frequency are indicated along the horizontal and vertical axes, respectively. A waveform is represented by a rectangle in this plane with its sides parallel to the time and frequency axes. ... Let is call such a rectangle an* information cell. *The time and frequency of a cell can be read, for example, from the coordinates of its lower left corner. The uncertainty in time and the uncertainty in frequency are given by the respective dimensions of the cell, it does not matter whether the nominal frequency and time position is taken from the center or from a corner of a rectangle."*

Both wavelets and wavelet packets can be seen to 'cover' certain portions of the time-frequency plane. At each scale $j = 1, 2, \ldots, J$ *wavelets can be associated with the frequency interval (and the vertical axis of the time-frequency plane) of* $\mathcal{I}_j = (2^{-(j+1)}, 2^{-j}]$. *Wavelet packets are associated with the interval* $\mathcal{I}_{j,i} = (2^{-(j+1)}i < \omega \leq 2^{-(j+1)}(i+1)]$ *for* $j = 1, 2, \ldots, J$ *and* $i = 0, 1, \ldots, 2^j - 1$. *Note that* $\mathcal{I}_j = \mathcal{I}_{j,1}$: *that is a wavelet packet with index* $i = 1$ *is equivalent to the wavelet at that scale. Of course, a packet has a time-extent as well and so a wavelet packet basis,* $b \in \mathcal{B}$, *is a disjoint cover of the entire time-frequency plane (see Theorem 3 of Hess-Nielsen and Wickerhauser (1996)).*

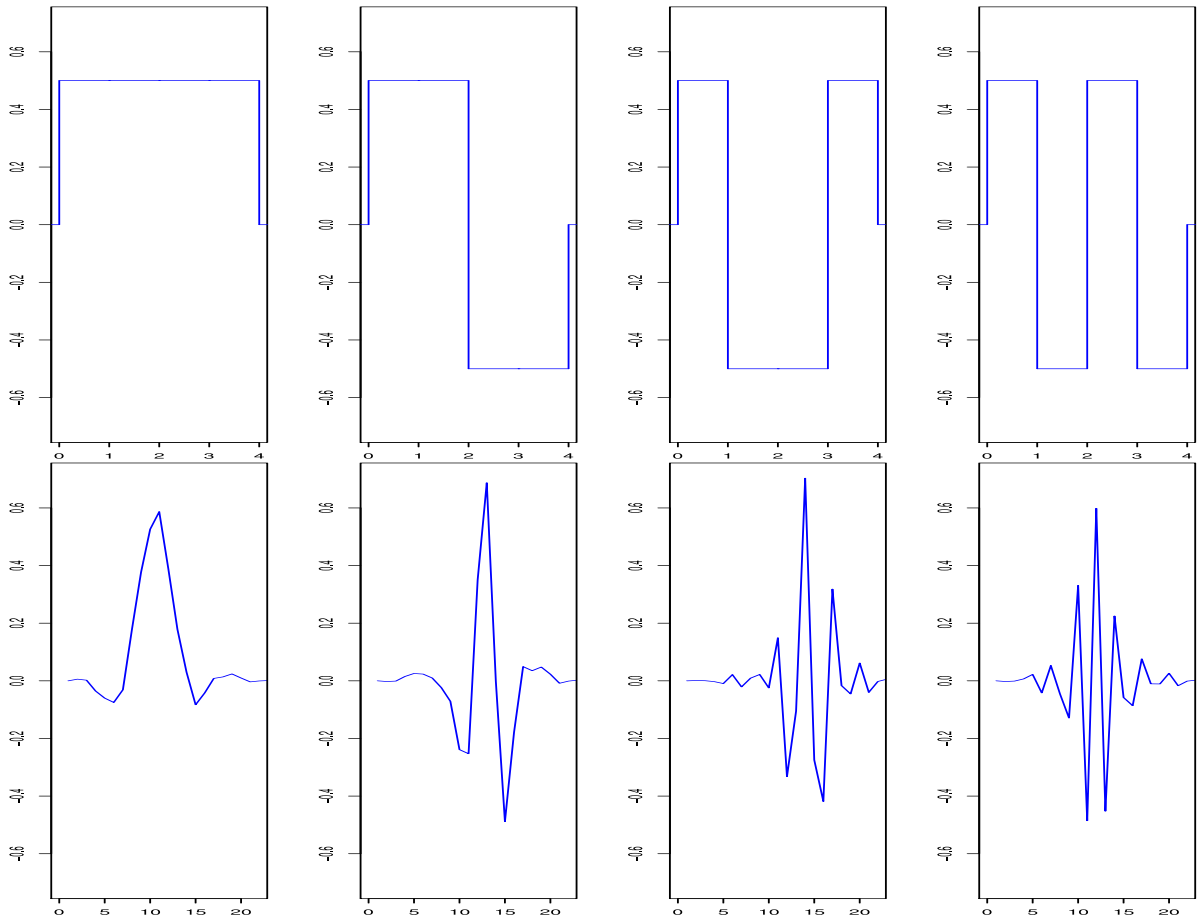

Figure 1: From left: wavelet packets $\psi_{j,i}$ for scale $j = 2$ and oscillations $i = 0, \ldots, 3$ built from Haar wavelets (top row), and from LA(8) filters (bottom row).

3.2 Basis Libraries

A *basis library* is a redundant set of bases from which one can be chosen to represent a data generating process. One example for locally stationary time series representation is provided by the SLEX processes from Section 1. In that example the SLEX library redundancy concerned multiple possible segmentations of the time dimension in the time-frequency plane of those processes. The local cosine bases used in Mallat et al. (1998) provide another example of redundancy in the time dimension. However, other basis libraries exist where redundancy characterizes the frequency domain in the process representation. Typically, basis libraries are redundant (and allow adaptive segmentation) only with respect either the time or frequency domain. For example, local cosine and SLEX libraries allow for adaptive segmentation of the time domain whilst looking at predefined frequencies. Here, we consider libraries of non-decimated wavelet packets. Unlike the aforementioned examples, these libraries offer redundancy (and allow adaptive segmentation) on the frequency interval $(0, 1/2]$, whilst providing fixed resolution in the (rescaled) time domain.

Remark 5 (WP libraries and underlying wavelets) *Any given wavelet packet basis library depends on an underlying Daubechies' mother wavelet. Hence, there are different libraries corresponding to different mother wavelets, and each of those will have different pros and cons. Different mothers could be incorporated into our scheme but, for simplicity of presentation, we restrict ourselves to a wavelet packet basis computed with respect to a single given compactly supported Daubechies' mother wavelet. Our framework will, however, be valid for all wavelet packet libraries built from any of Daubechies' wavelets. Our computational examples consider smooth wavelet packets built from* least asymmetric *filters of length $N = 8$, or $LA(8)$, which are particularly well suited for time series analysis. See Figure 1 for an illustration.*

To establish notation, let $\mathcal{B}$ denote a particular basis library and $|\mathcal{B}|$ denote the number of bases it contains. Let $|b|$ define the number of packets in each basis b. In the following, $\mathcal{B}$ will be the (non-decimated) wavelet packet library built from a given Daubechies' wavelet, see Coifman and Wickerhauser (1992), Wickerhauser (1994), Hess-Nielsen and Wickerhauser (1996) which includes the wavelet basis as a particular entry in the library. Our methodological aim is to devise an approach to identify such a basis that best fits the data with respect to some statistical criterion. This corresponds to a process representation where the time-frequency plane is segmented by a sequence of intervals $\{\mathcal{I}_{j_p,i_p}\}_{p\in b}$, where $\mathcal{I}_{j,i}$ was defined in Remark 4.

Example 1 (Wavelet packet libraries notation) *For $J = 2$ consider the following library of wavelet packet bases $\mathcal{B} = \{b_a, b_b, b_c, b_d\}$ where*

$$\begin{aligned} b_a &= \{(1,0),(1,1)\} \\ b_b &= \{(1,0),(2,2),(2,3)\} \\ b_c &= \{(1,1),(2,0),(2,1)\} \\ b_d &= \{(2,0),(2,1),(2,2),(2,3)\}. \end{aligned}$$

For example, basis b_c contains three packets so $|b_c| = 3$ and we can alternatively use the compact notation $b_c = \{(j_p, i_p)\}_{p=1,2,3}$ so that, for $p = 1$ we have $(j_1, i_1) = (1, 1)$, for $p = 2$ we have $(j_2, i_2) = (2, 0)$ and for $p = 3$ we have $(j_3, i_3) = (2, 1)$. Hence, we can equivalently refer to a wavelet packet either by the doublets (j_p, i_p) or their briefer basis location index $p = 1, \ldots, |b|$. The basis b_c is also the discrete wavelet basis (up to $J = 2$) since this is given in general by packets $\{(j, 1), (J, 0)\}_{j=1,\ldots,J}$. Figure 1 shows the wavelet packets forming the basis b_d with $|b_d| = 4$.

Remark 6 *In general the size of any basis b is finite for finite T. This is because the set of all possible packets is of size $T \times \log_2 T$. We typically refer to bases b of finite size. However,*

whenever appropriate, we specify which results refer to infinite dimensional bases b, which correspond to the limit case $T \to \infty$.

4 Local Stationarity and Wavelet Packet Processes

This section introduces locally stationary processes constructed using a wavelet packet basis using a data-driven basis selection strategy from a library of packet bases. First, for a given fixed basis, b, we introduce the *locally stationary wavelet packet* (LSWP) processes. Elements of the basis b are called packets and denoted by p.

Definition 2 (LSWP process) *Given wavelet packet basis $b \in \mathcal{B}$, the LSWP processes are a sequence of doubly-indexed stochastic processes $\{X_{t;T}\}_{t=0,\ldots,T-1}$, $T = 2^J \geq 1$ having the following representation in the mean-square sense*

$$X_{t;T} = \sum_{p \in b} \sum_{k} w_{j_p,i_p,k;T}\, \psi_{j_p,i_p,k}(t)\, \xi_{j_p,i_p,k}, \tag{8}$$

where $\xi_{j_p,i_p,k}$ and $w_{j_p,i_p,k;T}$ are, respectively, a collection of orthonormal random variables and amplitude coefficients with location index $k = 0, \ldots, T-1$ and packets (j_p, i_p) for $p \in b$. The set $\{\psi_{j_p,i_p,k}(t)\}_{j_p,i_p}$ contains discrete non-decimated wavelet packets that have support length N_{j_p} and are based on a mother wavelet $\psi(t)$ of compact support with length N, as above. Moreover, for $z \in (0,1)$, there exist functions $W_{j_p,i_p}(z)$ that satisfies the following conditions:

i *There exists a sequence of constants C_p such that for each $p \in b$ and T*

$$\sup_k \left| w_{j_p,i_p,k;T} - W_{j_p,i_p}\left(T^{-1}k\right) \right| \leq C_p/T,$$

where C_p fulfils $\sum_{p \in b} C_p < \infty$.

ii *Let $\sum_{p \in b} V_p < \infty$. Then, for $p \in b$, the total variation norm of $W^2_{j_p,i_p}(z)$ is bounded by V_p*

$$||W^2_{j_p,i_p}(z)||_{TV} \;=\; \sup_{\{a_i\}} \left\{ \sum_{d=0}^{D} \left| W^2_{j_p,i_p}(a_d) - W^2_{j_p,i_p}(a_{d-1}) \right| : 0 < a_0 < \cdots < a_D < 1 \right\} \leq V_p,$$

where the sup *is over all partitions $\{a_i\}$ of $(0,1)$.*

For a nontrivial theory we require some further tools and notation. First, we define two operators that generalize the autocorrelation wavelet and associated inner product from §2.3 and §2.4 of Nason et al. (2000).

Definition 3 (Cross-correlation wavelet packets) *For $p, p' \in b$ define the* cross-correlation wavelet packet *by the convolution:*

$$\Psi_{p,p'}(\tau) = \sum_k \psi_{p,k}\ \psi_{p',k-\tau}, \tag{9}$$

where $\psi_{p,k}$ are non-decimated wavelet packets from Definition 1. When the convolution is taken over the same wavelet packet, i.e. when $p' = p$ then $\Psi_{p,p}(\tau) = \Psi_p(\tau)$ is also called autocorrelation wavelet packet. *We also define $A = (A_{p,p'})_{p,p'=1,\ldots,|b|}$ as the inner product operator having entries*

$$A_{p,p'} = \sum_\tau \Psi^2_{p,p'}(\tau) = \sum_\tau \Psi_p(\tau)\ \Psi_{p'}(\tau). \tag{10}$$

The two derivations are equivalent but the latter can be implemented in a more computationally efficient way. For finite samples ($T < \infty$) the operator becomes a square matrix of finite dimensions $|b| \times |b|$. For both finite and infinite dimensional cases we also define the inverse operator $A^{-1} = (A^{-1}_{p,p'})_{p,p'}$. Conditions for the existence of this operator in both finite and infinite dimensional cases will be discussed in the following of this section.

Remark 7 (Spectra and autocovariances for LSWP processes) *Analogously to the locally stationary wavelet model, we define the* evolutionary wavelet packet spectra *(EWPS) as $S_p(z) = |W_{j_p,i_p}(z)|^2$ and the marginal EWPS as $\bar{S}_p = \int S_p(z)dz$. In what follows we refer to their whole sets of values respectively as $\mathbf{S}(b) = \{S_p(z)\}_{p\in b}$ and $\bar{\mathbf{S}}(b) = \{\bar{S}_p\}_{p\in b}$. The time localized covariance is given by*

$$C(z,\tau) = \sum_{p\in b} S_p(z)\Psi_p(\tau), \tag{11}$$

where $\Psi_p(\tau)$ is the wavelet packet autocorrelation function from Definition 3. In the following we also refer to the whole set of T observations from the model (8) as $\mathbf{X}_T$. As suggested in Fryzlewicz et al. (2003), for $t, s = 0, \ldots, T-1$, we can approximate the entries of $\Sigma = \mathbb{E}(\mathbf{X}_T\ \mathbf{X}_T')$, as $\Sigma(t,s) = C(t/T, t-s) + O(T^{-1})$. Because the only unknown quantities in Σ are the spectral entries we will also refer to it as $\Sigma_{\mathbf{S}(b)}$.

Example 2 (Haar MA packet processes.) *To familiarize the reader with locally stationary wavelet processes Nason et al. (2000) introduced the Haar moving average (MA) processes. Recall that the first order Haar MA process was $X^1_t = 2^{-1/2}(\epsilon_t - \epsilon_{t-1})$ and the second order was $X^2_t = 2^{-1}(\epsilon_t + \epsilon_{t-1} - \epsilon_{t-2} - \epsilon_{t-3})$ for $t \in \mathbb{Z}$, where $\{\epsilon_t\}$ is an i.i.d. zero mean unit variance process. These can be written in the locally stationary wavelet form in (3) by setting (for X^1_t) $S_1(z) = 1$, $\xi_{1,k} = \epsilon_k$ and $\psi_{j,k}(t)$ being non-decimated Haar wavelets, similarly for X^2_t and more generally X^r_t.*

For any given packet, $p \in b$, a similar kind of MA process can be defined. For example, at scale $j = 1$, the process X^1_t above is one process and,, in wavelet packet notation, its scale $j_p = 1$ and index number $i_p = 2$, i.e. $X^{(1,2)}_t$. The other packet process at scale $j_p = 1$ is $X^{(1,1)}_t = 2^{-1/2}(\epsilon_t + \epsilon_{t-1})$. At the second scale there are 4 packets denoted $X^{(2,i)}_t$ for $i = 0, \ldots, 3$ of the same form as X^2_t above but with the signs of each of the coefficients (in the same order) are $(+,+,+,+)$, $(+,+,-,-)$, $(+,-,-,+)$ and $(+,-,+,-)$. The second in this list corresponds to the X^2_t process above.) From this, we can define the MA wavelet packet process *by selecting a particular packet (and underlying wavelet) but using wavelet packets instead of wavelets. For an illustration of these wavelet packets derived from the Haar and least-asymmetric LA(8) wavelets at scale $J = 2$, see Figure 1.*

4.1 Inference for a Fixed Basis

Given a fixed wavelet packet basis, $b \in \mathcal{B}$, we can use results analogous to those in Nason et al. (2000) to derive an estimator for the evolutionary wavelet packet spectra.

Definition 4 (Unbiased wavelet packet periodogram) *For a given packet $p \in b \in \mathcal{B}$ with packet vector ψ_p, define the* wavelet packet process *as the empirical wavelet packet coefficients of $X_{t;T}$:*

$$d_{p,k} = \sum_t X_{t;T}\,\psi_p(t-k). \tag{12}$$

The quantity $d_{p,k}$ is a process *rather than just a set of coefficients because local stationarity of $X_{t;T}$ is conferred onto the* process *$d_{p,k}$ through a time-invariant linear filtering. Also define the (raw)* wavelet packet periodogram *by $I_{p,k} = |d_{p,k}|^2$. As in Nason et al. (2000) the raw wavelet packet periodogram is a biased estimator of the spectra since it can be proved that*

$\mathbb{E}I_{p,k} = \sum_{p' \in b} A_{p,p'} S_{p'}(z) + O(T^{-1})$ *for* $p \in b$*, where* $A_{p,p'}$ *was introduced in Definition 3. However, the estimator can be 'corrected' to make it asymptotically unbiased. Therefore the (asymptotically)* unbiased wavelet packet periodogram *is defined as:*

$$L_{p,k} = \sum_{p' \in b} A^{-1}_{p,p'} I_{p',k}. \tag{13}$$

for $p \in b$*, where* $A^{-1}_{p,p'}$ *was introduced in Definition 3. From these definitions it follows that* $\mathbb{E}L_{p,k} = S_p(z) + O(T^{-1})$.

Obtaining a consistent estimator of $S_p(z)$ can be achieved using similar methods to those described in Nason et al. (2000). The R package `LSWPlib` will be available on CRAN in due course to compute all the quantities defined in this section.

4.2 A Note on the Theory of Locally Stationary Wavelet Packet Processes

A number of theoretical properties of LSWP processes are based on the existence of the (bounded) inverse operator A^{-1} introduced in Definition 3. For example, the existence of an unbiased spectral estimator for LSWP processes directly depends on the existence of such operator as can be appreciated by looking at equation (13). Moreover, an invertible representation between the evolutionary spectra $S_p(z)$ defined in remark 7 and local autocovariances defined in equation (11) is only possible if this (bounded) operator exists since for all $p \in b$ the inverse formula of (11) is

$$S_p(z) = \sum_{p' \in b} A^{-1}_{p,p'} \sum_{\tau} C(z, \tau) \Psi_{p'}(z). \tag{14}$$

The existence of this positive definite operator and its bounded inverse when b is the wavelet basis and when $\psi(\cdot)$ is either the Haar or Shannon wavelet was proved in Theorem 2 of Nason et al. (2000), who also conjectured the existence of a general results for all Daubechies' compactly supported wavelets.

We now show that this result extends not only to all Daubechies' compactly supported wavelets but also to operators A constructed from wavelet packet bases. We use several results from Goodman et al. (1995) and we will refer to specific parts of that paper as GMRS-*page number*. The proof of the following theorem can be found in the appendix.

Theorem 1 (Boundedness of A inverse) *Let* b *be a basis of packets. Let* $A = (A_{p,p'})_{p,p' \in b}$ *for* $|b| \to \infty$*, where* $A_{p,p'}$ *is defined in equation* (10)*. Furthermore, let* b *such that not all packets belong to the same scale, i.e.* $j_p \neq j_{p'}$*, for some* $p, p' \in b$*. Then the inverse of the semi-infinite* A *operator for wavelet packets is bounded.*

5 Locally Stationary Wavelet Packet Processes Basis Selection

In the previous section the 'true' wavelet packet basis is assumed known. However, in practice, the basis is not known and the goal is to find, at least, a good basis. Previous work with adaptive representations in signal processing, e.g. Coifman and Wickerhauser (1992) and time series, e.g. Ombao et al. (2001) or Mallat et al. (1998), for function and process representation has concentrated on using basis libraries, that is, libraries of orthonormal ℓ^2 bases. These studies concentrated on selecting the best basis, $\hat{b} \in \mathcal{B}$ where 'best' can have several different meanings, see Percival and Walden (2000) §6.3 page 221 for a nice example.

Strictly speaking the basis concept is identified with decimated wavelet packets: for non-decimated wavelets the equivalent collection of packets is termed a frame — which, mathematically, has the same representative power but within an overdetermined system and so not

technically a basis. In order to simplify our exposition we will keep using the notion of basis even if we will be referring to nondecimated wavelet packet frames derived from the associated decimated wavelet packet basis. More details on frames can be found in Mallat (2009).

Given an appropriate objective function to be optimized, our goal is to reconstruct the, possibly sparse, *true* representation from a dictionary of ℓ^2 *frames*, i.e. a collection of linearly independent vectors that are almost (but not exactly) orthogonal, see Daubechies (1992) for more details.

This task turns out to be significantly harder than selecting from a dictionary of orthogonal bases. In fact, representations based on ℓ^2 frames account for a significant number of redundant and correlated coefficients, therefore it is crucial to understand how to make good use of these. In our setup the main challenge is therefore the derivation of an appropriate objective function that can ensure good model fitting and the derivation of appropriate cost functionals that can be associated with each packet to ensure successful optimization/basis selection.

5.1 Suggestion: Cost Functionals Based on Profile Likelihood

Inference for locally stationary time series for a fixed (Fourier) basis has been the object of a number of papers such as Dahlhaus (1996b, 1997). However, from the point of view of theoretical inference, the problem of finding an adaptive frequency tiling can be seen as the problem of estimating a number of unknown packet indices $p \in b$, which are the parameters of interest, given the presence of nuisance parameters (the level of the spectra for $p \in b$). Profile likelihood provides a common approach to inference in the presence of nuisance parameters. The use of profile likelihoods for semiparametric models was discussed in Kauermann (2002), where it was established that, as in classical parametric models, profiling leads to systematic bias. For locally stationary processes built upon Gaussian innovations, the negative log-likelihood based on the representation (8) can be written as follows.

Proposition 1 *Let $X_{t;T}$ be defined as in* (8) *and having Gaussian innovations. Then the negative log-likelihood for a basis $b \in \mathcal{B}$ is proportional to*

$$\mathcal{L}_T\{b, \mathbf{S}(b)\} = (2T)^{-1} \sum_{p \in b} \sum_t \left\{ \log S_p(t/T) + \frac{L_{p,t}}{S_p(t/T)} \right\} + O(T^{-1}), \tag{15}$$

where $L_{p,t}$ is the (asymptotically) unbiased wavelet packet periodogram as defined in (13).

For LSWP processes the parameters of interest for selecting a basis are the packet indices $p \in b$, where b is a basis (or, more precisely, a NDWP frame). Here, the nuisance parameters are the vectors $\mathbf{S}(b)$ of the spectra associated with each frame. A profile log-likelihood for b can therefore be derived based on (15) so that we have the following result.

Proposition 2 *Let $X_{t;T}$ be defined as in* (8) *having Gaussian negative log-likelihood proportional to* (15). *Define $\overline{\log L_p} = T^{-1} \sum_t \log L_{p,t}$, where $L_{p,t}$ is defined by*(13). *Furthermore define the negative profile log-likelihood for $b \in \mathcal{B}$ as $\widetilde{\mathcal{L}}_T(b)$. Then we have*

1. $$\widetilde{\mathcal{L}}_T(b) = \frac{1}{2} \left[|b| + \sum_{p \in b} \overline{\log L_p} \right],$$

2. $$\mathbb{E} \left[\widetilde{\mathcal{L}}_T(b) - \mathcal{L}_T\{b, \mathbf{S}(b)\} \right] < 0.$$

These results show that the profile (negative) log-likelihood is a negatively biased estimator for the negative log-likelihood. Moreover Proposition 2 shows that the profile likelihood is characterized by a non-linear relationship with respect the nuisance parameters estimates, which are also strongly correlated at fine scales. Simulation experiments confirm that the basis selection based on the optimization of the profile likelihood lead to large systematic errors in reconstructing known bases. Since the biased profile likelihood and its non-linear dependence with respect nuisance parameter estimates leads to poor basis selection, we consider an alternative approach aimed to improve the basis selection by removing the aforementioned non-linearity. This is achieved by considering an alternative objective function and alternative cost functionals.

5.2 Cost Functionals for Penalized Least Squares

The alternative approach that we adopt to overcome the difficulties of working in this highly irregular and non-linear setting is based on the use of an objective function which is still biased with respect the log-likelihood but is now linearly related to the nuisance parameter estimates. Among several possible alternatives we will consider the objective function

$$\widetilde{\mathcal{L}}_{2,T}(b) = \frac{1}{2}\left[2^J + \sum_{p\in b} \alpha_p \bar{L}_p\right], \tag{16}$$

where, for $p = (j_p, i_p)$, $J = \max_p\{j_p\}_p$, $\bar{L}_p = T^{-1}\sum_t L_{p,t}$ and $D_{j_p} = \sum_{i_p}^{2^{j_p}} \bar{L}_p$, the weights are defined by

$$\alpha_p = \frac{a^{j_p}}{J-1}\left(1 - \frac{D_{j_p}}{\sum_{j_p}^{J} D_{j_p}}\right), \tag{17}$$

for some $a \in (0,1)$ and $b \in \mathcal{B}$. The functional form of (17) is based on two main arguments. First we note that $\sum_{j_p}^{J} \alpha_p\, a^{j_p} = 1$, so one component of the weights is self-normalizing. This component is then multiplied by a^{j_p} which compensates for the increasing dependence among $\bar{L}_p$ for different packets at finer scales. It is worth noting that for $\alpha_p = 1$ we have that $\mathbb{E}\widetilde{\mathcal{L}}_{2,T}(b) = 0.5(2^J + \bar{\sigma}_T^2)$ where $\bar{\sigma}_T^2 = \int C(z,0)dz$ is the *marginal variance* for the LSWP process. Since 2^J is constant for all $b \in \mathcal{B}$, the maximization of (16) coincides with the maximization of the fitted marginal variance, and therefore this approach to basis selection can be interpreted as using penalized marginal least squares. This is because by maximizing the variance of the fitted model we minimize the residual variance not captured by the selected basis and the corresponding spectral estimates. Conditionally on the weights α_p, $\widetilde{\mathcal{L}}_{2,T}$ is a consistent estimator for $\mathcal{L}_{2,T}\{b, \mathbf{S}(b)\} = 0.5(2^J + \sum_p \alpha_p \bar{S}_p)$. This is a consequence of the consistency for the marginal periodogram $\{L_p\}_p$ as estimator of the marginal spectra $\{S_p\}_p$. Under the assumptions of Definition 2, this latter proof is a direct consequence of Theorem 3 from Cardinali and Nason (2010) that proved the same result for locally stationary wavelet processes.

5.3 Basis Selection

The optimization of (16) necessary to implement basis selection of LSWP needs to be carried out over non-overlapping tiling of the frequency interval $(0, 1/2]$. This can still be achieved by using the best basis algorithm where, for each packet, we will consider the cost functionals $-\alpha_p \bar{L}_p$. Here negative signs are used because the best basis algorithm will instead minimize $-\widetilde{\mathcal{L}}_{2,T}(b)$ over $b \in \mathcal{B}$. Our algorithm for basis selection is very fast and is based on the following steps:

1. Calculate $\bar{L}_{(j_p,i_p)}$ for $j_p = 1$ and $i_p = 1, 2, \ldots, 2^{j_p}$. This is done by calculating a bias

matrix A for the scale $j_p = 1$ and then calculating the unbiased periodogram;

2. Repeat step 1 for scales $j_p = 2, 3, \ldots, J$, where J is the maximum level of the NDWP transform that we consider in the analysis;

3. Calculate the whole set of weights $\alpha_{(j_p,i_p)}$ from (17);

4. Apply the best basis algorithm to the set of cost functionals as $-\alpha_{(j_p,i_p)}\bar{L}_{(j_p,i_p)}$;

5. Select $\hat{b} = \underset{b\in\mathcal{B}}{\operatorname{argmin}} \left\{-\widetilde{\mathcal{L}}_{2,T}(b)\right\}$, where $\widetilde{\mathcal{L}}_{2,T}(b)$ is defined in (16).

5.4 Practical Advice for Parameter Settings

Our basis selection method based on penalized least squares requires three parameters to be set. The first choice to be made is which (mother) wavelet filter to use in order to build the wavelet packets library. We mainly refer to Daubechies filters here and have used both $LA(8)$ and Haar wavelets in our simulations. Generally $LA(8)$ filters performed better than Haar filters in our experiments, therefore we recommend their use. Further work would be required to investigate the performance of other wavelets and provide recommendations for their use in specific situations.

The second parameter that needs to be set is the the depth of the wavelet packet library, represented by the maximum scale J. This parameter should be selected by taking into account the wavelet filter used to build the library. Filters of greater length allow smaller values of J to be set. If N represents the filter length then a recommended choice for the library depth is to set $J = [\log_2 T] - [\log_2 N] - g$, where $[x]$ is the integer part of x and $g = 3$ is an integer that reduces the computational boundary of J further. This is in order to avoid overfitting due to large positive correlation of wavelet packet coefficients at large scales. In our examples we used this setting but even for $g = 2$ we obtained good results.

The third parameter to be set is the (penalty) rate of a^j from equation (17). Penalties are parameterised by a geometric progression which allows us to interpret the penalty in terms of compensation for increasing positive correlation of the wavelet packet periodogram coefficients at large scales j. We mainly used a penalty of rate $a = 0.98$, but our experiments suggest that values larger than 0.95 perform similarly. It should be also noted that the penalty rate should be set in regard to the wavelet filter used and for Haar filters setting $a = 0.95$ provides better results than larger values. Our experiments show that this rate should be proportional to the filter length N: the larger is the latter, the larger should be the rate. This implies that cost functionals at large scales should be penalized more for shorter filters. The need of larger penalization for Haar filters seems also due to the frequency leakage that characterize filters with shorter length.

6 Simulation Examples

This section simulates several LSWP processes and empirically evaluates our basis selection methodology. We simulate processes using representation (8) with $w_{j_p,i_p,k;T} = S_{p,k}^{1/2}$ for $k = 0, \ldots, T-1$, and independent draws $\xi_{j_p,i_p,k}$ from the standard Gaussian distribution. We consider both stationary and locally stationary processes with fixed bases and energy distributions, and compare the estimated bases with the truth.

To evaluate our fits, we derive a measure of the chance of correctly selecting increasing proportions of true packets in the estimated basis. We aim to construct a measure that accounts for the different portions of the spectra that are represented by each packet within a given basis, so we define $|\mathcal{I}_p|$ as the length of the frequency intervals associated with a generic

packet p and defined by $\mathcal{I}_{j,i}$ from Remark 4. Therefore, the 'portion' of the true spectrum b that is correctly fitted by $\hat{b}$ can be expressed as

$$|\mathcal{I}_{\hat{b},b}| = 2 \sum_{\hat{p}\in\hat{b}} \mathbb{I}(\hat{p} \in b)|\mathcal{I}_{\hat{p}}|, \tag{18}$$

where $\mathbb{I}(A)$ is the usual indicator function. Hence, $\mathbb{I}(\{\hat{p} \in b\})$ is one if $\hat{p}$ is contained within the true basis b and zero otherwise. Hence, the quantity $|\mathcal{I}_{\hat{b},b}|$ is larger when more of the estimated packets are contained within the true basis. Indeed, if the estimated basis is equal to the true basis then the complete frequency interval $(0, 1/2]$ is covered and the portion is one. (Since all the separate $|\mathcal{I}_{\hat{p}}|$ add up to $1/2$ and then multiplying by two gives a portion of 1).

When we run M separate simulations, indexed by $m = 1, 2, \ldots, M$, the portion for simulation m will be written as $|\mathcal{I}_{\hat{b}_m,b}| \in [0, 1]$. Define

$$R(q) = R_{M,b}(q) = M^{-1} \sum_{m=1}^{M} \mathbb{I}\left(|\mathcal{I}_{\hat{b}_m,b}| \geq q\right), \tag{19}$$

to be the empirical proportion of bases that correctly fit at least $100q\%$ of the true spectra for $q \in (0, 1)$, $b \in \mathcal{B}$ and M some positive integer. For clear axes labels in figures, below, we suppress the dependence of b, M on the $R_{M,b}(q)$ below, but it should be remembered that $R(q)$ depends on M and b.

We next exhibit our simulation results on six different process types labelled LSWP_1 to LSWP_6. Here $\text{LSWP}_1, \text{LSWP}_2, \text{LSWP}_3$ and LSWP_5 are stationary and LSWP_4 and LSWP_6 are locally stationary. In all cases $M = 1000$.

6.1 Simulating Stationary LSWP_1 Processes

Example LSWP_1 uses the basis

$$b_a = \{(1, 0), (4, 8), (4, 9), (4, 10), (4, 11), (4, 12), (4, 13), (4, 14), (4, 15)\}. \tag{20}$$

The frequency design implied by this basis gives high resolution to the highest frequencies and minimum resolution to the first half of the spectra. The evolutionary wavelet packet spectrum (EWPS) for our simulation is

$$\text{LSWP}_1 \to S_p = 2^{-(j_p-1)/2}\, \mathbb{I}(p \in b_a). \tag{21}$$

This spectrum does not depend on z, rescaled time, and hence the process specified by (8) is stationary and the amplitudes $w_{j_p,i_p,k} \approx S_p^{1/2}$. Figure 2 shows a single realization and the marginal spectra for this process. For the $(1, 0)$ basis element we have $S_{(1,0)} = 2^{-(1-1)/2} = 1$ and this element's tile occupies $(0, 0.25)$ of the marginal spectrum. For the remaining basis elements, that cover the upper half of the spectrum we have, e.g. $S_{(4,8)} = 2^{-(4-1)/2} \approx 0.353$ as indicated in the marginal spectrum for each of the scale 4 basis packets. Figure 3 shows the performance of our basis fitting estimator assessed by the metric $R(q)$ over $M = 1000$ simulations for different sample sizes. So, for example, for $T = 1024$ the bottom-right graph of Figure 3 approximately 80% of the simulations achieved just under 80% of the true basis packets. The figures indicate statistical consistency as the area under the curve increases with sample size.

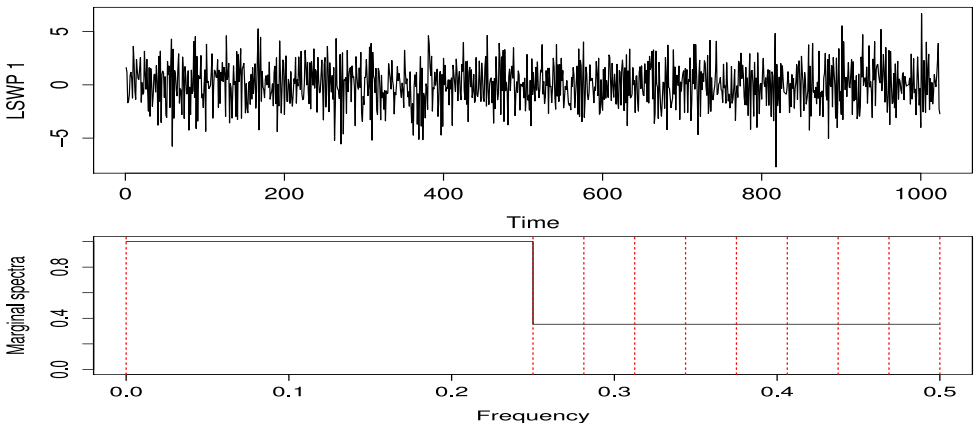

Figure 2: Top: realization of LSWP_1 process. Bottom: marginal spectra of process. Vertical lines show the tiling $\mathcal{I}_p$, for $p \in b_a$.

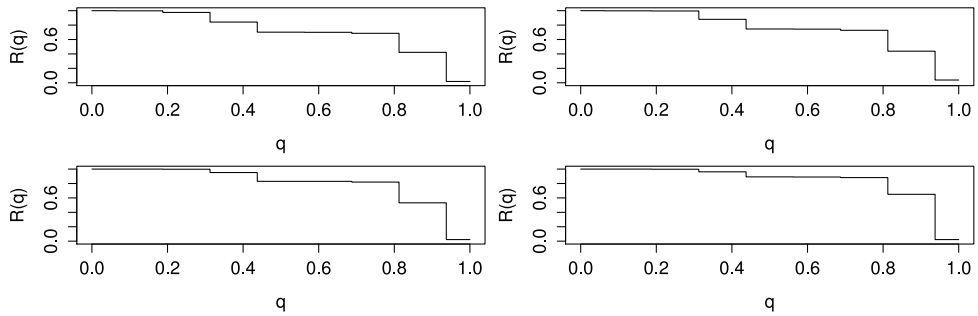

Figure 3: Survival probabilities, $R(q)$, for proportion of true spectra correctly fitted for processes LSWP_1 over 1000 realizations. Clockwise from top-left: $T = 128, 256, 512, 1024$.

6.2 Simulating Stationary LSWP_2 Processes

Example LSWP_2 uses the basis

$$b_b = \{(4,0),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),(4,7),(1,1)\}, \tag{22}$$

with spectrum given by

$$\mathrm{LSWP}_2 \quad \rightarrow \quad S_p = 2^{-(j_p-1)/2}\, \mathbb{I}(p \in b_b), \tag{23}$$

The frequency tiling implied by b_b is the opposite to that of the process LSWP_1, as it gives highest resolution to the lower frequencies and minimum resolution to the second half of the spectra. A single realization and the marginal spectra for this process are illustrated in Figure 4. Figure 5 displays the survival probabilities for the LSWP_2 model and the conclusions are

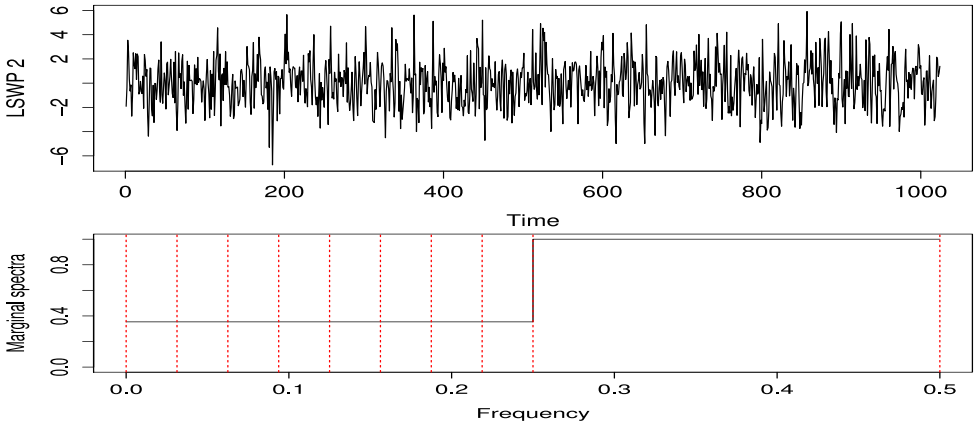

Figure 4: Top: one realization of LSWP_2 process. Bottom: marginal spectra for the process LSWP_2. The vertical lines show the tiling $\mathcal{I}_p$, for $p \in b_b$.

broadly the same as for the LSWP_1 process above.

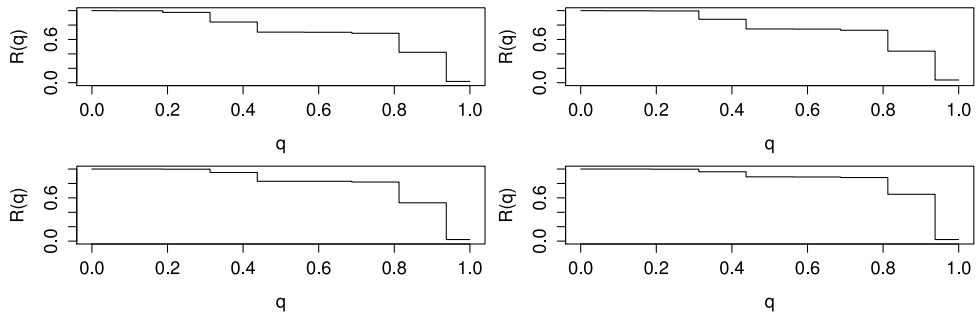

Figure 5: Survival probabilities, $R(q)$, for proportion of true spectra correctly fitted for processes LSWP_2 over 1000 realizations. Clockwise from top-left: $T = 128, 256, 512, 1024$.

6.3 Simulating LSWP_3 and LSWP_4 Processes

Examples LSWP_3 and LSWP_4 both use the basis

$$b_c = \{(1,1),(2,1),(3,1),(4,0),(4,1)\}, \tag{24}$$

which corresponds to a wavelet basis and hence the process will be a locally stationary wavelet process introduced by Nason et al. (2000). The frequency resolution is low at high frequencies and progressively better for lower frequencies. Unlike the previous two examples, we will

now consider both stationary and time-varying energy distributions. The two distributions correspond to different processes that we will refer to as

$$\begin{aligned} \text{LSWP}_3 &\rightarrow S_p = 2^{-j_p/2}\, \mathbb{I}(p \in b_c), \\ \text{LSWP}_4 &\rightarrow S_p(z) = 2^{-(j_p-2)/2} \cos^2(4\pi z)\, \mathbb{I}(p \in b_c) \qquad z \in (0,1). \end{aligned}$$

The processes are chosen to have identical marginal spectra which is motivated by our desire to see how the time-varying nature affects the fit (irrespective of the marginal spectrum). Single realizations of the two processes and their true marginal spectra are illustrated in Figure 6, where it is shown the typical discrete wavelet transform spectral resolution, which increases dyadically moving towards lower frequencies. This property is commonly known as *adaptivity*

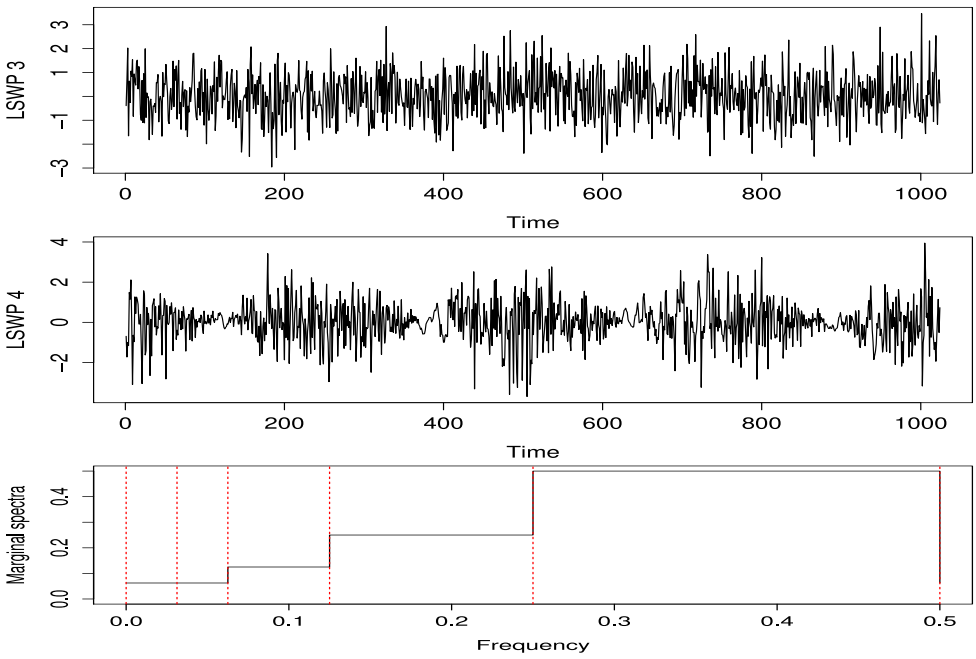

Figure 6: Top: realization of LSWP_3 process. Center: realization of LSWP_4 process. Bottom: marginal spectra for both processes. Vertical lines show the tiling $\mathcal{I}_p$, for $p \in b_c$.

of the wavelet transform, and implies that noisier signal components are averaged over a wider frequency band, and those bands decrease for less noisy components. Figures 7 and 8 show the survival probabilities for different sample sizes. The figures provide empirical evidence

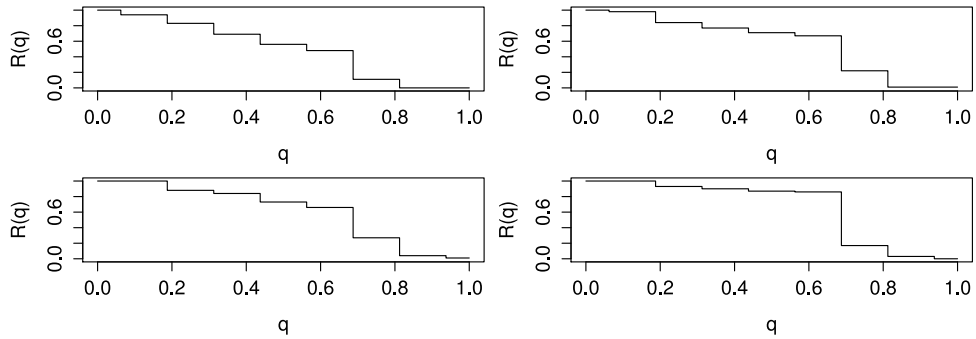

Figure 7: Survival probabilities, $R(q)$, for proportion of true spectra correctly fitted for processes LSWP_3 over 1000 realizations. Clockwise from top-left: $T = 128, 256, 512, 1024$.

that our procedure is consistent, and is also invariant to the specification of the time-varying energy distribution since the shape of the curve $R(q)$ is very similar for the two processes.

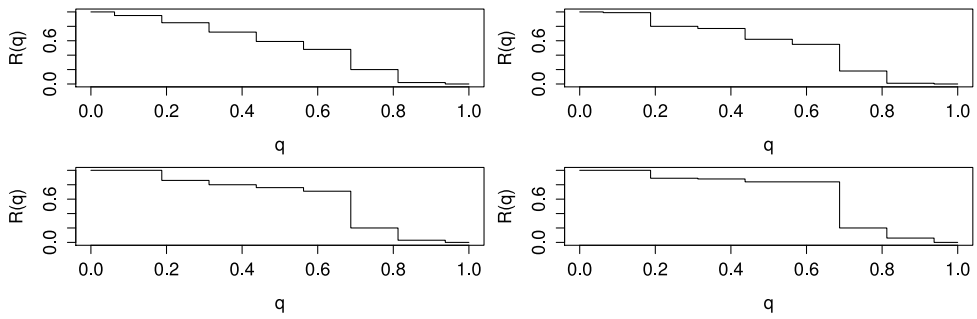

Figure 8: Survival probabilities, $R(q)$, for proportion of true spectra correctly fitted for processes LSWP_4 over 1000 realizations. Clockwise from top-left: $T = 128, 256, 512, 1024$.

6.4 Simulating LSWP_5 and LSWP_6 Processes

Examples LSWP_5 and LSWP_6 both use the basis

$$b_d = \{(1,0), (2,2), (3,6), (4,14), (4,15)\}. \tag{25}$$

The frequency tiling implied by this basis is still an example of smooth change in frequency resolution, but corresponds to the opposite design in comparison to that of the discrete wavelets above. In fact, now the frequency resolution is low at low frequencies and then increases for higher frequencies. Analogously to the previous section, we will now consider both examples of stationary and time-varying energy distributions with spectra given by

$$\begin{aligned} \text{LSWP}_5 &\rightarrow S_p = 2^{2j_p - 8}\, \mathbb{I}(p \in b_d), \\ \text{LSWP}_6 &\rightarrow S_p = 2^{2j_p - 7} \cos^2(4\pi z)\, \mathbb{I}(p \in b_d). \end{aligned} \tag{26}$$

Single realizations from these processes and their (identical) marginal spectra are shown in Figure 9. Performance of our estimator is again summarized by the plots in Figures 10 and 11.

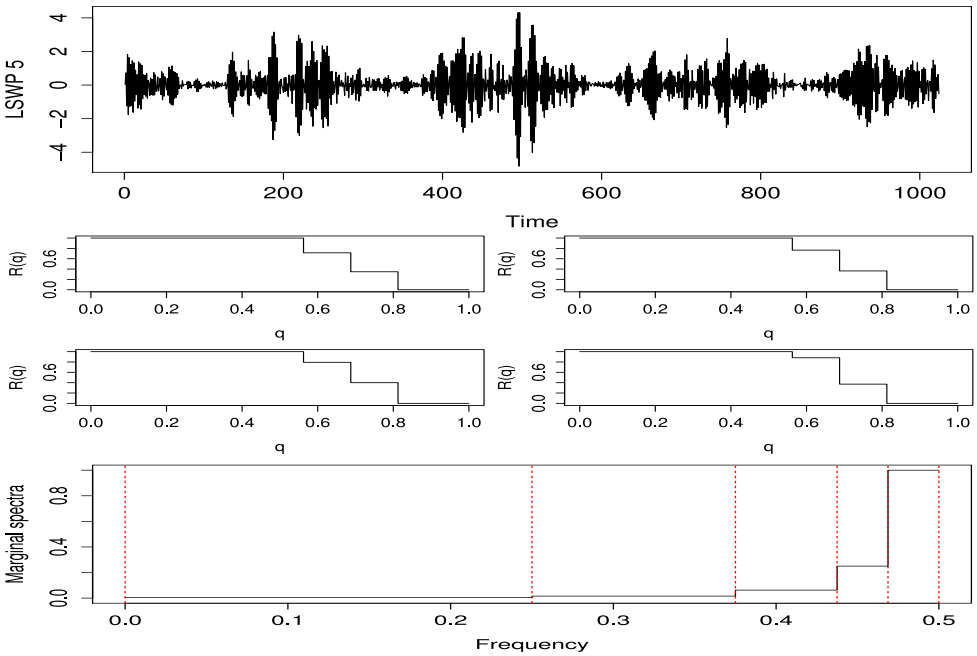

Figure 9: Top: one realization of LSWP_5 process. Center: one realization of LSWP_6 process. Bottom: marginal spectra for the processes LSWP_5 and LSWP_6. The vertical lines show the tiling $\mathcal{I}_p$, for $p \in b_d$.

The results indicate that our procedure is consistent also for these examples and is also

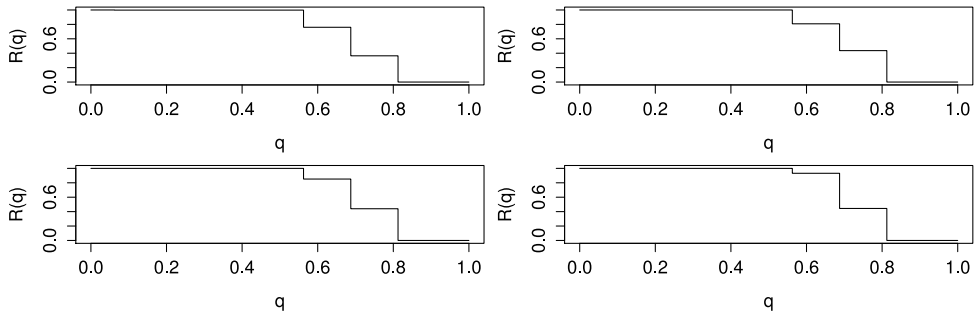

Figure 10: Survival probabilities, $R(q)$, for proportion of true spectra correctly fitted for processes LSWP_5 over 1000 realizations. Clockwise from top-left: $T = 128, 256, 512, 1024$.

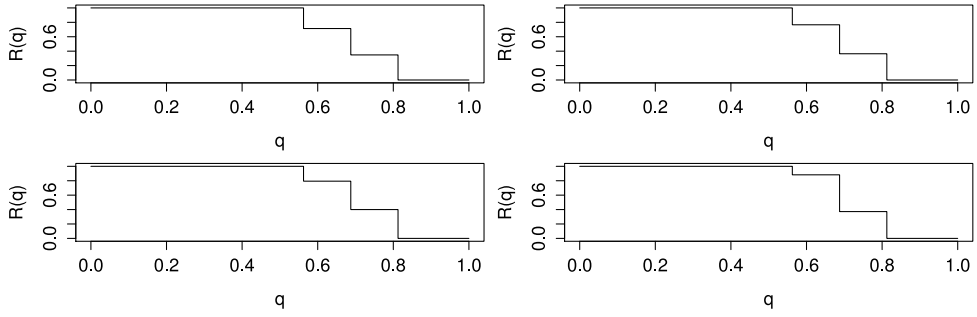

Figure 11: Survival probabilities, $R(q)$, for proportion of true spectra correctly fitted for processes LSWP_6 over 1000 realizations. Clockwise from top-left: $T = 128, 256, 512, 1024$.

invariant to the specification of the time-varying energy distribution since the shape of the curve $R(q)$ is basically identical for the two processes. Overall, we can estimate about 70% of the true basis with high probability, and just below 50% chance to achieve about 80% fit.

7 Time-frequency Analysis of S&P 500 Log-returns

Figure 12 shows a series of 1024 S&P500 log-returns from the period November 1999 to July 2002. These data have been widely analyzed in many different ways and GARCH models

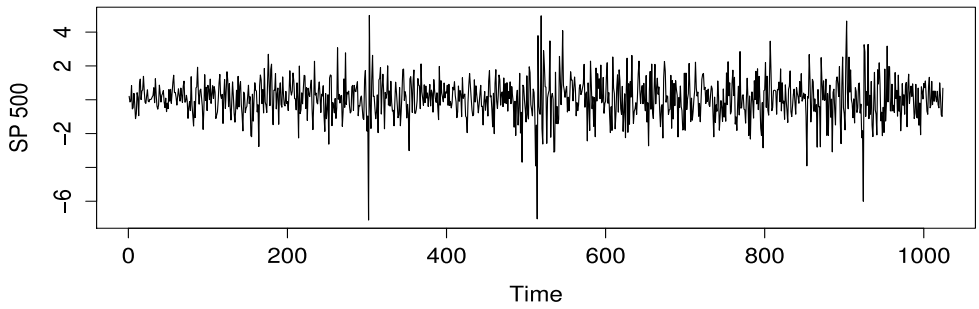

Figure 12: Log-returns for S&P 500 index.

have been proposed by several studies. Our simulations above indicated that LSWP models can *reproduce* heteroscedasticity even from a stationary specification. In particular, LSWP_5 realizations show that this can occur when the estimated marginal spectra accounts for high frequency resolution at highest frequencies. In these situations the *intensity* of heteroscedasticity is positively correlated with the energy level of the marginal spectra. In a similar setup,

simulations of the LSWP_6 process have also shown that time-varying energy can only *reinforce or mitigate* the intensity of heteroscedasticity.

We applied our basis selection methodology to the S&P 500 returns, setting $J = 4$, and estimated the basis

$$\hat{b}_{SP} = \{(1,0),(3,6),(3,7),(4,8),(4,9),(4,10),(4,11)\}. \tag{27}$$

It is exciting that this basis is not a wavelet nor close to a Fourier basis. So, using wavelet packets has really made a difference here. In particular, the basis selection indicates that it is probably wasteful to use a traditional Fourier spectral analysis which has too fine frequency resolution at the lower frequencies.

Our fitted basis accounts for higher resolution and (relatively) low energy at higher frequencies, and this seems the characteristic time-frequency scenario for financial returns in an efficient market. The estimated marginal spectra is illustrated in Figure 13. We have evidence

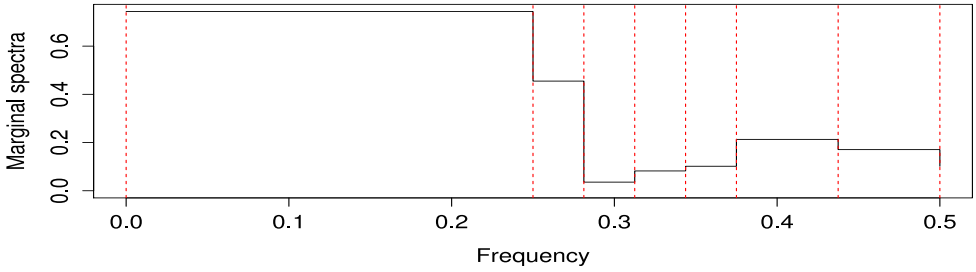

Figure 13: Estimated marginal spectra for S&P 500 log-returns. Vertical red dotted lines indicate frequency division of associated best basis.

that the lower frequencies account for higher energy than higher frequencies. We are therefore in an intermediate situation with respect the simulated scenarios.

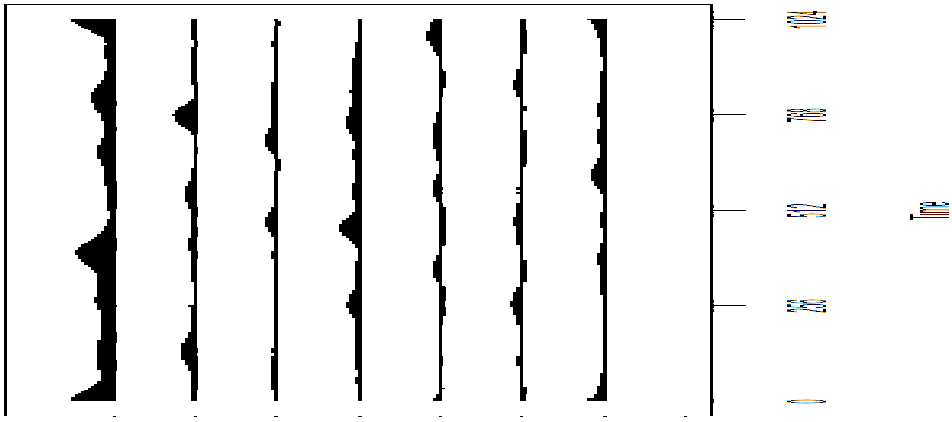

Figure 14: Wavelet packet periodogram for S&P 500 log-returns. From top $L_{p,t}$, for $p \in \hat{b}_{SP}$ and $t = 0, 1, .., 1023$. Within the green vertical lines coefficients are free of boundary effects.

Remark 8 (Robustness of basis selection) *An important question is 'how robust is our basis selection to variations in the parameter settings discussed in Section 5.4?' We conducted our analysis using wavelet packets built from Daubechies $LA(8)$ filters since these performed better than Haar filters in our simulations.*

However, the results presented for the S&P data have also been obtained using the other recommended settings listed in Section 5.4. We have repeated our analysis using different values for both the library depth J and the penalty rate a. When considering $J = 5$ only one packet from $\hat{b}_{SP}$ was omitted and was replaced by two child *packets from scale $j = 5$. The missing*

packet corresponds to the high-frequency end of the spectra, therefore to the more noisy data component. For $J = 6$ we have the same basis than that selected for $J = 5$. For all those values of J we were able to select the packet corresponding to $j_p = 1$ which in our basis corresponds to (low) frequencies in the interval $(0, 1/4]$. As for the penalty rate a from equation (17) we have repeated our analysis for rates $0.95 < a < 0.98$ and have observed the same results in each of the cases $J = 4, 5, 6$.

Remark 9 (Interpreting evidence from time-frequency analysis) *A wavelet packets basis represents the oscillatory components accounting for the most energy in a time series. The associated frequency intervals provide twofold information both on a range of frequencies/periods characterizing each component, and on their length (which is inversely proportional to the precision of measuring the evolution of the (time-varying) amplitude and variance). Therefore, wide frequency intervals correspond to more precise measurement of time-varying features and should be of particular interest in applications focusing on precise measurements in the time domain such as change-point analysis.*

Figure 14 shows the asymptotically unbiased wavelet packet periodogram from (13). Each spectral line in Figure 14 (top to bottom) corresponds to the elements (left to right) in the selected wavelet packet basis given in (27). The estimated basis $\hat{b}_{SP}$ is characterized by a wide packet at low frequencies $\mathcal{I}_{1,0} = (0, 1/4)$ corresponding to oscillatory dynamics with period larger than 4 days (since we analysed daily data). The relatively high energy associated with this wavelet packet suggests that this asset could be of interest to investors looking for weekly (or longer) investment horizons. The time-frequency periodogram does not show evidence of strong time-varying effects apart from two packets that correspond to lower and medium frequencies. Those packets do not belong to the classical discrete wavelet transform tiling, therefore the identification of the time-varying effects could be compromised if a non-adaptive tiling (such as that from the discrete wavelet transform) was imposed. The overall limited presence of time-varying effects (apart the aforementioned episodes around February 2001) seems to reflect the prevalent non-transitory nature of the heteroscedasticity that characterizes this particular dataset. The two changes identified at the beginning of 2001 seems to mark the U.S. recession (and U.S. stock market drop) that started in that trimester.

8 Conclusions and Further Work

This article introduces locally stationary wavelet packet processes (LSWP) and shows how they can be used for modelling and analysis of locally stationary time series. Unlike other locally stationary models based on orthonormal ℓ^2 bases the LSWP model includes finite sample stationary processes as particular cases. Furthermore, the LSWP family provides a very flexible framework for analyzing locally stationary time series by allowing the most important periodic components to be selected by a data-driven criterion. However, basis selection for LSWP is a difficult problem: the ℓ^2 frame design of NDWP potentially introduces over-parametrization and correlation in a standard best-basis selection. We are able to derive a modified best-basis selection that uses functionals of the unbiased periodogram.

We also use a penalty to ensure that a sparse basis is selected, and to compensate for the leakage that affects non-decimated wavelet packet designs, especially at fine scales. We show with a number of prototype simulation examples that LSWP can represent a large variety of empirical features and provide a novel framework for analyzing heteroscedastic time series. Our simulations show that, by using different designs for energy distribution and frequency resolution, we can represent many heteroscedastic features by using little or no time-varying spectral coefficients. Our empirical analysis based on the S&P 500 returns confirms this finding, and, in particular, that heteroscedastic time series are better represented by a frequency tiling which is different from the one implied by the classical discrete wavelet transform. These

findings can lead to significant improvements in the accuracy of in-sample and out-of-sample analysis of locally stationary time series.

Acknowledgements

AC is grateful to GPN for having introduced him to the study of locally stationary time series. GPN was honoured to have been invited to speak at, and participate in, the Professor Maurice Priestley Commemoration Day on 18th December 2013 at the School of Mathematics in Manchester and would like to thank deeply the organizers of that meeting. GPN is grateful to the EPSRC for grant EP/K020951/1 which partly funded this work.

References

Cardinali A. 2009. A generalized multiscale analysis of the predictive content of Eurodollar implied volatilities. *International Journal of Theoretical and Applied Finance* **12**: 1–18.

Cardinali A, Nason. GP. 2008. Costationarity and stationarity tests for stock index returns. *Technical Report* 08:08, Statistics Group, University of Bristol.

Cardinali A, Nason, GP. 2010. Costationarity of locally stationary time series. *Journal of Time Series Econometrics* **2** Issue 2.

Cardinali A, Nason, GP. 2016. Practical powerful wavelet packet tests for second-order stationarity. (submitted for publication).

Coifman R, Wickerhauser M. 1992. Entropy-based algorithms for best-basis selection. *IEEE Transactions on Information Theory* **38**: 713–718.

Dahlhaus R. 1996a. Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In *Athens Conference on Applied Probability and Time Series Analysis. Lecture Notes in Statistics, Vol. 115*. New York: Springer Verlag.

Dahlhaus R. 1996b. Maximum likelihood estimation and model selection for locally stationary processes. *Journal of Nonparametric Statistics* **6**: 171–191.

Dahlhaus R. 1997. Fitting time series models to nonstationary processes. *Annals of Statistics* **25**: 1–37.

Dahlhaus R. 2012. Locally stationary processes. *Handbook of Statistics, Time Series Analysis: Methods and Applications* **30**: 351–413.

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

Donoho DS, Mallat S, von Sachs R, Samuelides, Y. 2003. Locally stationary covariance and signal estimation with macrotiles. *IEEE Transactions on Signal Processing* **51**: 614–627.

Fryzlewicz P, Van Bellegem S, von Sachs R. 2003. Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics* **55**: 737–764.

Gabbanini F, Vannucci M, Bartoli G, Moro A. 2004. Wavelet packet methods for the analysis of variance of time series with applications to crack widths on the Brunelleschi dome. *Journal of Computational and Graphical Statistics* **13**: 639–658.

Garcia CA, Otero A, Vila X, Márquez DG. 2013. A new algorithm for wavelet-based heart rate variability analysis. *Biomedical Signal Processing and Control* **8**: Issue 6.

Goodman T, Micchelli C, Rodriguez G, Seatzu S. 1995. On the Cholesky factorization of the Gram matrix of locally supported functions. *BIT Numerical Mathematics* **35**: 233–257.

Hannan E. 1960. *Time series analysis*. Chapman and Hall.

Hess-Nielsen N, Wickerhauser MV, 1996. Wavelets and time-frequency analysis. *Proceedings of the IEEE* **84**: 523–540.

Jin L, Wang S, Wang H. 2015. A new non-parametric stationarity test of time series in the time domain. *Journal of the Royal Statistical Society* B **77**: 893–922.

Kauermann G. 2002. On a small sample adjustment for the profile score function in semiparametric smoothing models. *Journal of Multivariate Analysis* **82**: 471–485.

Mallat S. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**: 674–693.

Mallat, S. 2009. *A Wavelet Tour of Signal Processing*. 3rd edn. Paris: Academic Press.

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

Milne AE, Macleod CJA, Haygarth PM, Hawkins JMB, Lark RM. 2009. The wavelet packet transform: A technique for investigating temporal variation of river water solutes. *Journal of Hydrology* **379**: 1–19.

Nason GP. 2008. *Wavelet Methods in Statistics with R*. Springer: New York.

Nason GP 2013. A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *Journal of the Royal Statistical Society* B **75**: 879–904.

Nason GP, Sapatinas T. 2002. Wavelet packet transfer function modelling of nonstationary time series. *Statistics and Computing* **12**: 19–56.

Nason, GP, Sapatinas T, Sawczenko A. 2001. Wavelet packet modelling of infant sleep state using heart rate data. *Sankhya* B **63**: 199–217.

Nason GP, von Sachs R. 1999. Wavelets in time series analysis. *Philosophical Transactions of the Royal Society of London* A **357**: 2511–2526.

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

Ombao H, Raz J, von Sachs R, Guo W. 2002. The SLEX model of non-stationary random processes. *Annals of the Institute of Statistical Mathematics* **54**: 171–200.

Ombao H, Raz J, von Sachs R, Malow B. 2001. Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association* **96**: 543–560.

Ombao H, von Sachs R, Guo W. 2005. SLEX analysis of multivariate nonstationary time series. *Journal of the American Statistical Association* **100**: 519–531.

Parzen E. 1959. Statistical Inference on Time Series by Hilbert Space Methods, I. *Technical Report* 23, Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California.

Park TA, Eckley I, Ombao H. 2014. Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Transactions on Signal Processing*. **62:20**: 5240–5250.

Percival DB, Walden AT. 2000. *Wavelet Methods for Time Series Analysis*. Cambridge University Press.

Percival DP. 1995. On estimation of the wavelet variance. *Biometrika* **82**: 619–631.

Priestley MB. 1965. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society* B **27**: 204–237.

Priestley MB. 1983. *Spectral Analysis and Time Series*. Academic Press.

Priestley MB. 1988. *Non-linear and Non-stationary Time Series Analysis*. Academic Press.

Priestley MB, Subba Rao T. 1969. A test for non-stationarity of time-series. *Journal of the Royal Statistical Society* B **31**: 140–149.

Sanderson J, Fryzlewicz PZ, Jones M. 2001. Measuring dependence between non-stationary time series using the locally stationary wavelet model. *Biometrika* **97**: 435–446.

Silverman R. 1957. Locally stationary random processes. *IRE Transactions on Information Theory* **IT-3**: 182–187.

Von Sachs R, Schneider K. 1996. Wavelet smoothing of evolutionary spectra by nonlinear thresholding. *Applied Computational Harmonic Analysis* **3**: 268–282.

Walden AT, Contreras Cristan AC. 1998. The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events. *Proceedings of the Royal Society of London, Series A* **454**: 2243–2266.

Wickerhauser M. 1994. *Adapted Wavelet Analysis from Theory to Software*. Wellesley, MA: A.K. Peters.

Yang Y, He Y, Cheng J, Dejie Y. 2009. A gear fault diagnosis using Hilbert spectrum based on MODWPT and a comparison with EMD approach. *Measurement* **42**: 542–551.