2018-05

Practical powerful wavelet packet tests for second-order stationarity

Cardinali, A

10.1016/j.acha.2016.06.006
Applied and Computational Harmonic Analysis
Elsevier BV

Practical Powerful Wavelet Packet Tests for Second-order Stationarity

Alessandro Cardinali and Guy P. Nason*

23rd March 2015

Abstract

Methods designed for second-order stationary time series can be misleading when applied to nonstationary series, often resulting in inaccurate models and poor forecasts. Hence, testing time series stationarity is important especially with the advent of the 'data revolution' and the recent explosion in the number of nonstationary time series analysis tools. Most existing stationarity tests rely on a single basis. We propose new tests that use nondecimated basis libraries which permit discovery of a wider range of nonstationary behaviours, with greater power whilst preserving acceptable statistical size. Our tests work with a wide range of time series including those whose marginal distributions possess heavy tails. We provide freeware R software that implements our tests and a range of graphical tools to identify the location and duration of nonstationarities. Theoretical and simulated power calculations show the superiority of our wavelet packet approach in a number of important situations and, hence, we suggest that the new tests are useful additions to the analyst's toolbox.

Keywords: stationarity test, local stationarity, bootstrap

1 Introduction

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

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

where $d\xi(\omega)$ is an orthonormal increments process and $A(\omega)$ is the amplitude function, see, for example, Hannan (1960) or Priestley (1983).

We are interested in the case where X_t might be locally stationary: a nonstationary process which appears to be stationarity over short periods but *can* change its statistical properties relatively slowly. Nonstationary time series have been studied over many years, see Page (1952) or Silverman (1957), for example, and theory was significantly advanced

*g.p.nason@bristol.ac.uk

by a series of papers by M.B. Priestley and co-authors from the mid 1960s such as Priestley (1965). Dahlhaus (2012) provides a recent review of locally stationary series.

Many nonstationary representations rely on the Fourier basis to provide oscillation: Silverman (1957), Priestley (1965), Dahlhaus (1997), for example. However, for nonstationary processes the Fourier basis and (1) are no longer canonical. For example, Priestley (1983) admits general oscillatory basis functions and stipulates conditions on their form and Nason et al. (2000) introduced models based on nondecimated wavelets called locally stationary wavelet (LSW) processes and also suggested using wavelet packets to provide oscillation.

If nonstationarities in a time series are not detected, then one will proceed as if the series were stationary — with potentially erroneous results for models and forecasts as stationary analyses average out all the interesting nonstationary behaviour. A simple time series plot can often aid stationarity determination although their interpretation can be somewhat subjective and maybe less desirable than objective rigorous statistical tests.

An early hypothesis test for stationarity was proposed by Priestley and Subba Rao (1969) (PSR) that performs an ANOVA analysis on the logarithm of a time-varying spectral estimate at a predefined set of times and frequencies. Software for the PSR test has recently been made publicly available via CRAN in the `fractal` package by Constantine and Percival (2007). Many other tests exist. For example: those that measure correlation between periodogram ordinates such as Hurd and Gerr (1991) and Dwivedi and Subba Rao (2011), those that measure the discrepancy between a time-varying spectral estimate and its 'closest' stationary spectrum such as Gardner and Zivanovic (1991) and Dette et al. (2011), and those that measure constancy of some Fourier spectral functional such as Priestley and Subba Rao (1969), von Sachs and Neumann (2000) and Paparoditis (2010). See also Andrieu and Duvaut (1996) for the specific alternative of a Gaussian cyclostationary process. These tests all work with the Fourier spectrum or closely-related quantities.

Tests that work with other bases include Nason (2013), which utilized a wavelet spectrum, whereas Cardinali and Nason (2010) works with either wavelet or Fourier bases. Another recent alternative is Jin et al. (2015) which makes use of the Walsh basis. All these tests have different operational characteristics but it has become evident that nonstationarity can manifest itself in distinctly non-Walsh, non-Fourier and non-wavelet ways. For example, simulations in Nason (2013) show that there are cases where (i) neither, (ii) both, or (iii) one or the other of a wavelet or Fourier test successfully detect certain kinds of nonstationarity. Section 2 provides additional evidence as to why measuring in both the Fourier and the wavelet directions alone is unlikely to be enough.

The tests mentioned above measure departures from a *single* type of representation: either Walsh, Fourier or wavelet. A key contribution here is a new approach to testing which looks for departures from *multiple* basis representations. Since our test can look in more directions than one basis it should achieve higher power whilst retaining acceptable size characteristics. In fact, we show this by examining theoretical power in section 5 and via a comprehensive simulation study in section 6.

Section 3 introduces our first stationarity test based on wavelet packets using the 'significant Haar wavelet coefficient' method introduced by von Sachs and Neumann (2000) and verifies its theoretical credentials. In practice, this test relies on a hard-to-estimate vari-

ance which sometimes results in poor empirical power. Section 4 circumvents this issue by establishing asymptotic equivalence of the Section 3 wavelet packet statistic to the L_2 test statistic from Cardinali and Nason (2010) and Dette et al. (2011) but assesses significance via a simple bootstrap method.

All our tests are encapsulated in the freeware R software package `BootWPTOS` which also includes graphical tools to identify the location and scale of discovered nonstationarities,

1.1 Background and notation

We assume a basic knowledge of wavelet theory but comprehensive treatments can be found in Daubechies (1992), Chui (1997) or Burrus et al. (1997) for example. For wavelets and signal processing, see Mallat (1998) or Flandrin (1998) and see Vidakovic (1999) or Nason (2008) for their use in statistics.

Informally, the *locally stationary Fourier* (LSF) model from Dahlhaus (1997) represents a (triangular array of) stochastic process(es) defined by:

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

for $t = 1, \ldots, T$. The *locally stationary wavelet* (LSW) 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 $\{\psi_{j,k}(t)\}$ is a set of discrete nondecimated wavelets (defined and extended below). The $\exp(i\omega t)$ in (2) and the $\psi_{j,k}(t)$ in (3) supply oscillation for the models at different frequencies and scales, respectively. The $d\xi(\omega)$ in (2) and $\{\xi_{j,k}\}$ in (3) are uncorrelated zero mean, unit variance random variables and a zero mean orthonormal increments process, respectively (which could be Gaussian to create Gaussian processes). The functions $A_{t,T}(\omega)$ and $w_{j,t;T}$ are time-varying amplitudes that cause the processes to be nonstationary if they are not constant functions of t.

The usual, fundamental (and estimable) parameters underlying models (2) and (3) are the time-varying spectrum $f(t/T, \omega) = |A_t(\omega)|^2$ and the time-scale spectrum $S_j(t/T) \approx w^2_{j,t/T}$.

The Fourier basis in (2) is well-known. The nondecimated discrete wavelets in (3) are defined in Section 2 of Nason et al. (2000) by

Definition 1 (Nondecimated wavelets)**.** *Let $\{h_k, g_k\}$ be the low and high pass quadrature mirror filters used in the construction of the Daubechies (1992) compactly supported continuous time wavelets. The* discrete wavelets $\psi_j = \{\psi_{j,0}, \ldots, \psi_{j,(N_j-1)}\}$ *are compactly supported of length N_j for scale $j \in \mathbb{N}$ obtained by using the formulae:*

$$\psi_{1,n} = \sum_k g_{n-2k} \delta_{0,k} = g_n, \quad \textit{for } n = 0, \ldots, N_1 - 1, \tag{4}$$

and

$$\psi_{j+1,n} = \sum_k h_{n-2k}\psi_{j,k}, \quad \textit{for } n = 0, \ldots, N_j - 1, \tag{5}$$

where $N_j = (2^j - 1)(N_h - 1) + 1$, $\delta_{0,k}$ *is the Kronecker delta and* N_h *is the number of non-zero elements of* $\{h_k\}$.

We include this definition only to extend it below. This section only provides minimal detail on the LSF and LSW processes: for complete technical details see section 3.3 for LSF and in Nason (2013) for LSW processes.

However, our key point is that both LSF and LSW models rely on a single basis for their definition. The next section provides heuristics as to why multiple bases are important for stationarity tests.

2 Choice of Basis and Stationarity Tests

This section justifies why basis choice is crucial for stationarity determination.

2.1 Empirical evidence

Nason (2013) and Cardinali and Nason (2013) report results of separate empirical studies comparing the PSR Fourier-based test with different types of wavelet-based tests. The studies examined the same four nonstationary time series models, P1–P4, defined later in Section 6.3.2. For example, Nason (2013) showed that the PSR test achieves high power (100%) for P2 and P4 and low power (around 40%) for P1 and P3, whereas the wavelet-based test achieves high power ($> 94\%$) for P1 and P4, but low (17% and 1%) for P2 and P3 respectively. Sometimes wavelet-based tests are more powerful, sometimes Fourier are and sometimes they perform similarly. Results in Table 2 in Cardinali and Nason (2010), using the same test statistic and the same bootstrap method of inference, strongly indicate that the best test to use in a given situation depends on the kind of nonstationarity present in a signal.

Overall, though, simulations are not definitive and testing software is also subject to other sources of variation, such as choice of smoothing and other parameters. However, simple theoretical heuristics, demonstrated next, also show that the detection basis is important.

2.2 Theoretical evidence: LSF analysis of LSW

Suppose we subject the LSW process $X_{t,T}$ from (3) to a Fourier spectral analysis. For example, apply the localized periodogram estimator as given in von Sachs and Schneider (1996) with no tapering to obtain

$$I_T(z, \lambda) = \left| \sum_{s=0}^{T-1} X_{[zT-(T/2)+s+1],T} \exp(-2\pi i \lambda s) \right|^2. \tag{6}$$

Then, taking the central component (CC) of (6) applied to the LSW process gives:

$$\begin{aligned} \text{CC} &= \sum_j \sum_k w_{j,k;T} \xi_{j,k} \sum_{s=0}^{T-1} \psi_{j,k}([zT] - T/2 + s + 1) e^{-2\pi i \lambda s} \\ &= e^{-2\pi i \lambda [-zT+T/2-1]} \sum_j \sum_k w_{j,k;T} \xi_{j,k} \sum_{t=[zT]-T/2+1}^{[zT]+T/2} \psi_{j,k}(t) e^{-2\pi i \lambda t}. \end{aligned} \quad (7)$$

Definition 1 shows that $\psi_{jk}(t)$ is non-zero for $0 \leq k-t \leq N_j - 1$. Hence, in (7), for suitably large T, the lower limit on the final sum is less than $k - N_j + 1$ and the upper limit is greater than k, or equivalently $N_j - T/2 \leq k \leq T/2$. Then the final summation is equivalent to $\hat{\psi}_{j,k}(\lambda) = \sum_{-\infty}^{\infty} \psi_{j,k}(t) \exp(-2\pi i \lambda t)$ the discrete Fourier transform of $\psi_{j,k}(\cdot)$ and hence:

$$\text{CC} = e^{-2\pi i \lambda [-zT+T/2-1]} \sum_j \sum_k w_{j,k;T} \xi_{j.k} \hat{\psi}_{j,k}(\lambda) = e^{-2\pi i \lambda [-zT+T/2-1]} \hat{F}(\lambda), \quad (8)$$

for some function $\hat{F}(\lambda)$. To obtain the localized periodogram with no tapering we need to form $|\text{CC}|^2$ which gives $I_T(z, \lambda) = |\hat{F}(\lambda)|^2$ as the exponential part cancels. In summary, under the right conditions, the time-localized *Fourier* periodogram of a time-varying LSW process can effectively be *independent of time* z.

Under which conditions does this phenomenon occur? Examine the inequalities $N_j - T/2 \leq k \leq T/2$ from above. If j is coarse scale then N_j is large and potentially the interval $[N_j - T/2, T/2]$ is very small, or even non-existent. However, if j is small, e.g. the smallest value is $j = 1$, then $N_1 = N_h$ and the interval can be large, especially if T is large. So, the conditions for the 'time-independence' of $I_T(z, \lambda)$ of a LSW process are most likely to occur at the finer scales.

The conclusion is that quantities such as $I_T(z, \lambda)$ will not easily detect nonstationarities if the LSW process is nonstationary at the finer scales. Intuitively, this makes sense because the fine scale components of LSW processes have a large bandwidth, roughly $[\pi/2, \pi)$ for integer-sampled series, that is not efficiently representable by a few Fourier coefficients.

2.3 Theoretical evidence: LSW analysis of LSF

Suppose now that we examine the LSF process $X_{t,T} = \sigma(t/T) Z_t$, where Z_t is an iid $N(0, 1)$ sequence and $\sigma(z)$ is a continuous function on $(0, 1)$. The $X_{t,T}$ is LSF by Dahlhaus (1997, page 4, Example (i)) with transfer function $A(z, \lambda) = \sigma(z)\pi^{-1/2}$. Now subject this LSF process to a wavelet analysis: the finest scale Haar wavelet coefficients of $X_{t,T}$ are given by

$$d_{1,k} = 2^{-1/2} \left[\sigma(k/T) Z_t - \sigma\{(k-1)/T\} Z_{t-1} \right].$$

Hence, the expectation of the raw wavelet periodogram at the finest scale $I_{1,k} = d_{1,k}^2$ is given by $\mathbb{E} d_{1,k}^2 = \left\{ \sigma^2 \left(\frac{k}{T}\right) + \sigma^2 \left(\frac{k-1}{T}\right) \right\} /2$. Similarly, at scale two, the expectation is $\mathbb{E} d_{2,k}^2 = 4^{-1} \sum_{r=0}^{3} \sigma^2\{(k-r)/T\}$. Let A be the inner product matrix of the Haar autocorrelation wavelet as given in Nason et al. (2000). Then the evolutionary wavelet spectral

estimate at the finest scale is given by:

$$\hat{S}_1(k/T) = \sum_{j=1}^{\log_2(T)} A_{1,j}^{-1} d_{j,k}^2. \tag{9}$$

Substituting in actual approximate values for $A_{1,j}^{-1}$ we can obtain

$$\mathbb{E}\hat{S}_1(k/T) \approx [\sigma^2(k/T) + \sigma^2\{(k-1)/T\}]/3 - [\sigma^2\{(k-2)/T\} + \sigma^2\{(k-3)/T\}]/11$$

which is not a good estimator of $\sigma(z)$ as it smooths out variability. Hence, $\hat{S}_1$ would not be good for detecting non-constant behaviour in $\sigma(z)$. For this purpose it would be preferable to use a good LSF spectral estimator such as the asymptotically unbiased estimator from Neumann and von Sachs (1997).

2.4 Wavelet, Fourier or something else?

The discussion in sections 2.1–2.3 shows that success in testing for stationarity may depend heavily on the choice of analysis basis. Normally, one does not know in advance which basis offers the best chance of success. In the coming section, we develop a strategy that implicitly exploits multiple bases, and which attempts to mimic the success of the true best basis.

3 A Wavelet Packet Stationarity Test

3.1 Nondecimated wavelet packets and basis libraries

An improved test might be constructed using both wavelet and Fourier bases. However, this seems limited as there exist *basis libraries* containing a wide selection of basis elements. Let $\mathcal{B}$ denote a particular library and $|\mathcal{B}|$ denote the number of bases it contains. Let $b \in \mathcal{B}$ denote a given basis and let $|b|$ define the number of basis functions (packets) in b. The next two sections construct stationarity tests relative to a basis $b \in \mathcal{B}$.

We use the wavelet packet library, see Coifman and Wickerhauser (1992), Wickerhauser (1994), Hess-Nielsen and Wickerhauser (1996), which, of course, includes the wavelet basis as a special case. Each wavelet packet basis library depends on an underlying Daubechies' wavelet and hence there are different libraries for different wavelets and each of those will have different powers for detecting nonstationarities. Wavelet choice could be incorporated into our scheme but, for definiteness, we restrict ourselves to a wavelet packet basis computed with respect to a given particular compactly supported Daubechies' wavelet.

There are many basis library types: wavelet packets are used here, but others could be used resulting in tests with different operating characteristics. We actually use *nondecimated* wavelet packets, defined below. The nondecimation is extremely important for stationarity tests, but also for locally stationary time series models generally, see, e.g. Nason et al. (2000), Fryzlewicz et al. (2003); Van Bellegem and von Sachs (2004); Fryzlewicz

(2005); Park et al. (2005); Borchi et al. (2006); Fryzlewicz and Nason (2006); Khoklov et al. (2006); Woyte et al. (2006, 2007b,a); Van Bellegem and von Sachs (2008); Triantafyllopoulos and Nason (2008); Fryzlewicz and Ombao (2009); Xie et al. (2009); Cardinali and Nason (2010); Eckley et al. (2010); Sanderson et al. (2010); Eckley and Nason (2011); Cho and Fryzlewicz (2012); Gott and Eckley (2013); Nason (2013). Nondecimation is also known by other names such as the maximal overlap discrete wavelet transform see Percival and Walden (2000) and used extensively.

The reason why nondecimation is vital for testing stationarity is so that nonstationarities do not get missed in the gaps left by decimation inherent in regular wavelets or wavelet packets. Decimated wavelet packets are defined on an implicit grid that has gaps which scale dyadically and grow large quickly as we go from fine to coarse scales. Consequently, they possess a preferred set of locations at which changes should take place. The methods we construct, by being nondecimated, have no such preferred set of locations and, we believe, are therefore more suitable for detecting changes at arbitrary and unknown locations. More generally, nondecimated bases often result in more effective estimators, e.g. see Coifman and Donoho (1995) and Nason and Silverman (1995) or Percival (1995), Gabbanini et al. (2004) in conjunction with variance estimation and Cardinali (2008) in conjunction with cross-covariance estimation. Nason et al. (2001); Nason and Sapatinas (2002) used nondecimated wavelet *packets* to construct transfer function models between two time series to enable prediction (or classification) of one time series from an explanatory time series.

We now extend Definition 1 from wavelets to wavelet packets.

Definition 2 (Discrete Non-decimated Wavelet Packets)**.** *The construction of* discrete nondecimated wavelet packets *(DNWP) is identical to* (4) *and* (5) *except that the* $\{g_k\}$ *and* $\{h_k\}$ *can both be replaced by either of* $\{g_k\}$ *or* $\{h_k\}$ *at each scale. The vectors produced by this construction are referred to as* discrete wavelet packets *and denoted by* $\psi_{\ell,k}$*, where* ℓ *refers to a packet and can be written as the doublet* $\ell = (j, i)$ *where* j *refers to scale and* i *to the index within that scale. The* k *index controls shift as in regular wavelets. Define the set of packet indices by* $\mathcal{L} = \{\ell = (j, m) : m = 0 \ldots, 2^j - 1, j = 1, \ldots, J\}$.

The DNWP construction process enables simple packet indexing. Starting from the finest scale either the low-pass filter, h_k, or the high-pass filter, g_k, (as in (4)) can be applied. The resulting coefficients for low-pass can be indexed by 0 and the high-pass by 1. Then, both of these packet sets can then be further processed by low-pass, h_k, or high-pass, g_k and hence the second finest coefficients can be indexed by 00, 01, 10 or 11. More generally, the rth scale packets can be indexed by a binary string of length r. The ordering of wavelet packets in this way is known as the Paley ordering; see Coifman and Wickerhauser (1993).

For regular wavelet packets, which are functions defined on $\mathbb{R}$, rather than vectors as in Definition 2, a basis, $b \in \mathcal{B}$ is a particular set of packets that forms an orthonormal basis for $L^P(\mathbb{R})$ which corresponds to a subtree of the full binary packet tree where the leaves represent the basis packets. For function estimation purposes we would require a complete basis for representation, but for stationarity testing we merely require a selection of packets. The nondecimated discrete wavelet packets will be used to focus on time series features to be tested for stationarity. We now give a few pedagogical examples for those readers not

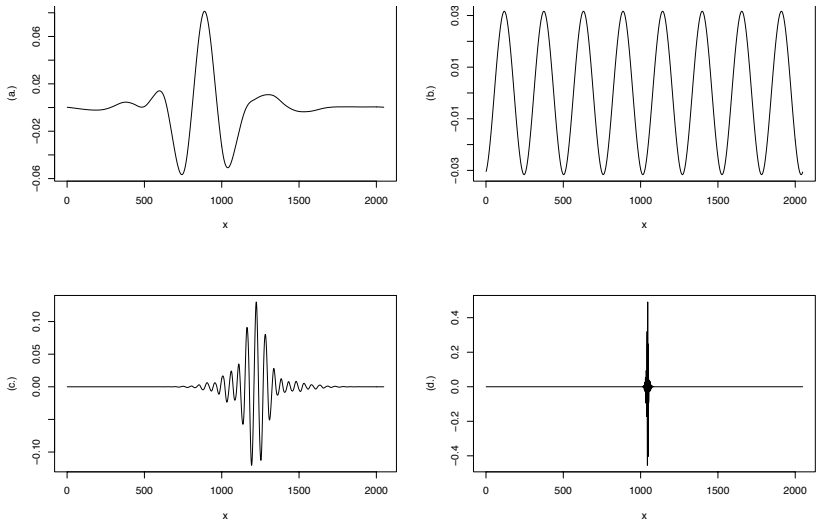

Figure 1: Four wavelet packets derived from Daubechies (1992) least-asymmetric mother wavelet with ten vanishing moments for $T = 2048$: (a.) $(8, 1)$, (b.) $(11, 8)$, (c.) $(7, 6)$, (d.) $(3, 2)$. Drawn by the `drawwp.default()` function in `WaveThresh`.

familiar with wavelet packets.

Example 1. *The scale $j = 1$ and $j = 2$ discrete Haar wavelets are given by $\psi_1 = (1, -1)/\sqrt{2}$ and $\psi_2 = (1, 1, -1, -1)/2$ as in Nason et al. (2000). The two Haar wavelet packets at scale $j = 1$ are $\psi_{(1,1)} = (1, -1)/\sqrt{2} = \psi_1$ and $\psi_{(1,0)} = (1, 1)/\sqrt{2}$. There are four Haar wavelet packets at scale $j = 2$. They are $\psi_{(2,0)} = (1, 1, 1, 1)/2$, $\psi_{(2,1)} = (1, 1, -1, -1)/2 = \psi_2$ and $\psi_{(2,2)} = (1, -1, 1, -1)/2$ and $\psi_{(2,3)} = (1, -1, -1, 1)/2$.*

Example 2. *Figure 1 shows four wavelet packets derived from a Daubechies mother wavelet. The figure gives an idea of the diversity of the time-frequency behaviour compared to the Fourier or wavelet bases.*

Remark 1. *At each scale father wavelets correspond wavelet packets with index $m = 0$ and (regular) wavelets correspond to index $m = 1$. A standard wavelet basis, up to scale J, can be obtained from a wavelet packet library by choosing wavelets at each scale from 1 to J and additionally the father wavelet at the coarsest scale. The regular wavelet basis is described by the following packet indices $\{(1, 1), (2, 1), (3, 1), \ldots, (J, 1), (J, 0)\}$.*

3.2 Underlying Stochastic Model and Assumptions

To enable a rigorous development we now specify our underlying time series model.

Assumption 1 (Dahlhaus (1997))**.** *Our time series are modelled as a sequence of stochastic processes $X_{t,T}$, $t = 1, \ldots, T$, a* locally stationary (Fourier) *process with transfer function*

A^0 with representation

$$X_{t,T} = \int_{-\pi}^{\pi} A^0_{t,T}(\omega) \exp(i\omega t) d\xi(\omega), \tag{10}$$

where

(i) ξ is an orthonormal increments process on $[-\pi, \pi)$ with $\overline{\xi(\omega)} = \xi(-\omega)$, $\mathbb{E}\xi(\omega) = 0$ and $\operatorname{cov}\{d\xi(\omega), d\xi(\nu)\} = \delta(\omega - \nu)d\omega d\nu$ and

$$\operatorname{cum}\{d\xi(\omega_1), \ldots, d\xi(\omega_k)\} = \eta\left(\sum_{j=1}^{k} \omega_j\right) h_k(\omega_1, \ldots, \omega_{k-1}) d\omega_1 \cdots d\omega_k,$$

where $\operatorname{cum}\{\cdot\}$ denotes the cumulant of order k, $|h_k(\omega_1, \ldots, \omega_{k-1})| \leq const_k$ for all k (with $h_1 = 0$, $h_2(\omega) = 1$) and $\eta(\omega) = \sum_{j\in\mathbb{Z}} \delta(\omega + 2\pi j)$ is the Dirac comb.

(ii) Let $\Pi = [-\pi, \pi)$. There exists a positive constant K and smooth function $A(u, \omega)$ on $[0, 1] \times \Pi$ which is 2π-periodic in ω, with $A(u, -\omega) = \bar{A}(u, \omega)$, such that $\forall T$

$$\sup_{t,\omega} \left| A^0_{t,T}(\omega) - A\left(\frac{t}{T}, \omega\right) \right| \leq KT^{-1}, \tag{11}$$

and (A^0, A) are known as a close pair, see Cardinali and Nason (2010).

Further conditions are imposed on the function A by the following.

Assumption 2 (Neumann and von Sachs (1997))**.**

(i) The function $A(u, \omega)$ has bounded total variation on $[0, 1]\times\Pi$ denoted by $\operatorname{TV}_{[0,1]\times\Pi}(A) < \infty$;

(ii) $\sup_u \operatorname{TV}_{\Pi}\{A(u, \cdot)\} < \infty$ and $\sup_\omega \operatorname{TV}_{[0,1]}\{A(\cdot, \omega)\} < \infty$;

(iii) $\sup_{u,\omega} |A(u, \omega)| < \infty$;

(iv) $\inf_{u,\omega} |A(u, \omega)| \geq \kappa$ for some $\kappa > 0$.

Assumption 3. *Let $\hat{A}(u, s) := (2\pi)^{-1} \int_\Pi A(u, \omega) e^{i\omega s} d\omega$, $s \in \mathbb{Z}$, $u \in (0, 1)$. Assume $\hat{A}$ is well-defined and finite. This is a slightly weaker assumption on $\hat{A}$ compared to Assumption 3 in Neumann and von Sachs (1997).*

The equivalent definition for $X_{t,T}$ in von Sachs and Neumann (2000) uses Assumption 1 but also imposes continuity on A. We have chosen to use the less restrictive bounded variational assumptions as found in Neumann and von Sachs (1997). The time-varying spectrum of $X_{t,T}$ is given by $f(u, \omega) = |A(u, \omega)|^2$.

Finally, we reuse the following cumulant assumption.

Assumption 4 (Neumann (1994))**.** $\sup_{1\leq t_1\leq T}\{\sum_{t_2,\ldots,t_k=1}^{T}|\operatorname{cum}(X_{t_1,T},\ldots,X_{t_k,T})|\}\leq C^k(k!)^{1+\gamma}$ *for all* $k=2,3,\ldots,$ *where* $\gamma\geq 0$*.*

Remark 2. von Sachs and Neumann (2000) Remark 3.1 notes that Assumption 4 is satisfied if $\{X_t\}$ is α-mixing with coefficients $\alpha(s)\leq K\exp(-b|s|)$ and

$$\mathbb{E}|X_t|^k\leq C^k(k!)^\rho, \tag{12}$$

for all k. Condition (12) is fulfilled for many distributions, for example, (Gaussian), exponential, gamma and inverse Gamma, for appropriate choice of ρ, see also Neumann (1994).

Remark 3. We could have chosen to model X_t with the locally stationary wavelet (LSW) processes of Nason et al. (2000) and below develop a parallel set of theory and tests.

Remark 4. The focus here is on second-order quantities and hence we assume $\mathbb{E}X_{t,T}=0$ which is certainly satisfied by Assumption 1 and models such as the LSW processes. If a real time series has a non-zero mean then it can be modelled, estimated and removed even in locally stationary noise, see von Sachs and MacGibbon (2000) for example.

3.3 Blueprint of the stationarity tests

Suppose we wish to test X_t, $t=1,\ldots,T$ for stationarity. Let $\mathcal{L}$ be a set of wavelet packet basis functions. Define the *nondecimated wavelet packet coefficients* of $\{X_t\}$, with respect to the discrete wavelet packet ψ_ℓ, $\ell\in\mathcal{L}$, by

$$d_{\ell,k}=\sum_{t=1}^{T}X_t\psi_{\ell,k-t}, \tag{13}$$

for all $k\in\mathbb{Z}$. Define the *raw wavelet packet periodogram* of $\{X_t\}$, with respect to the discrete wavelet packet ψ_ℓ, $\ell\in\mathcal{L}$,by

$$I_{\ell,k}=d_{\ell,k}^2, \tag{14}$$

for $k\in\mathbb{Z}$.

Our stationarity tests all relate to the constancy of the following b-spectrum.

Definition 3 (b-spectrum)**.** *Define the* b-spectrum*,* $\beta_\ell(z)$*, with respect to wavelet packet* ℓ*,* $\ell\in\mathcal{L}$*, and* $z\in(0,1)$ *by*

$$\beta_\ell(k/T)=\mathbb{E}(I_{\ell,k}), \tag{15}$$

for $k=1,\ldots,T$ *and for all* $T\in\mathbb{N}$*.*

Nason et al. (2000) Proposition 4 showed for LSW processes, where ψ_ℓ in (13) is a wavelet, that

$$\mathbb{E}(I_{j,k})=\{\mathcal{A}S(k/T)\}_\ell+\mathcal{O}(T^{-1}), \tag{16}$$

where $\{S_j(z)\}$ is the evolutionary wavelet spectrum (EWS) associated with a locally stationary wavelet process and $\mathcal{A}$ is a blurring matrix/operator. In definition 3 *b-spectrum*

refers to *blurred spectrum* which, when $\ell \in \mathcal{L}$ corresponds to a wavelet, is identical to the quantity in Nason et al. (2000). Fryzlewicz and Nason (2006) denoted $\beta_j(k/T) = \mathbb{E}(I_{j,k})$ for wavelets and it extends here naturally to wavelet packets.

The next proposition establishes the nature of the b-spectrum for our locally stationary series defined in Assumption 1.

Proposition 1. *Let $X_{t,T}$ be a locally stationary Fourier process as in Assumption 1 and let $\beta_\ell(z)$ be its b-spectrum for some wavelet packet $\ell \in \mathcal{L}$. Then:*

$$\beta_\ell(z) = \int_{-\pi}^{\pi} f(z,\omega)|\hat{\psi}_\ell(\omega)|^2 d\omega + \mathcal{O}(T^{-1}), \tag{17}$$

where $\hat{\psi}_\ell(\omega) = \sum_n \psi_{\ell,n} e^{-i\omega n}$ is the Fourier transform of the discrete wavelet packet vector ψ_ℓ.

Proof. See the appendix. A similar result for wavelets was established in von Sachs et al. (1997) formula (35).

The b-spectrum is at the centre of our new stationarity tests. Clearly, $X_{t,T}$ is second-order stationary if and only $f(z,\omega) = f(\omega)$ for all $z \in (0,1)$ if and only if $\beta_\ell(z)$ is a constant function of z for any (and all) $\ell \in \mathcal{L}$.

Example 3 (Example 1 continued)**.** *The squared gains of the scale two Haar wavelet packets are shown in Figure 2 acting as bandpass filters whose union covers the frequency range $[0,\pi)$ and divide the range into approximately equal quadrants. The filters are also orthogonal, i.e. $\sum_k \psi_{j,m,k}\psi_{j,n,k} = \delta_{m,n}$. At scale j the frequency domain is divided into 2^j roughly-equal bandpass filters and, as such, the wavelet packet decomposition provides and organized and coherent decomposition of the time-frequency plane essentially providing exhaustive (within the constraints of a finite discrete sequence) covering of that plane.*

3.4 Haar wavelet coefficient stationarity test

Our first test evaluates the constancy of $\beta_\ell(z)$ by examining its Haar wavelet coefficients, which is inspired by the method of von Sachs and Neumann (2000) who examined the constancy of Fourier spectral estimates.

Formula (17) supplies the clue to the advantage of our technique and a key difference from previous techniques. Rather than look at spectral estimates directly, as in von Sachs and Neumann (2000), we look at wavelet packet filtered versions of the spectrum. The ability of wavelet packets to split up the time-frequency plane permits us to hone in on aspects of nonstationarity that might be obscured by focusing on the entire spectrum. Our new method extends and improves on Nason (2013) also as we allow more diverse, but still coherent, wavelet packet analysis of the spectrum compared to just using wavelets.

The Haar wavelet coefficients of $\beta_\ell(z)$ are given by

$$v_{i,p}^{(\ell)} = \int_0^1 \beta_\ell(z)\psi_{i,p}^H(z)dz, \tag{18}$$

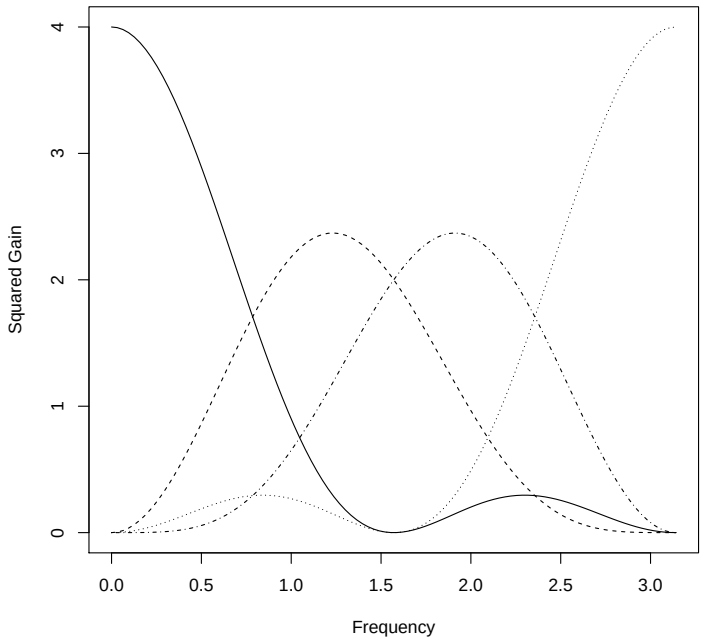

Figure 2: Squared gain functions, $|\hat{\psi}_\ell(\omega)|^2$ for the four scale two Haar wavelet packets: *solid*: $\ell = (2, 0)$; *dashed*: $\ell = (2, 1)$; *dotted*: $\ell = (2, 2)$ and *dot-dashed*: $\ell = (2, 3)$.

for $i = 1, \ldots, J$, $p = 0, \ldots, 2^i - 1$ and $\{\psi^H_{i,p}(t)\}_{i,p}$ are the usual Haar wavelets. Hence $X_{t,T}$ is stationary if and only iff $v^{(\ell)}_{i,p} = 0$ for all i, p. To be precise our null hypothesis is $H_0 : v^{(\ell)}_{i,p} = 0$ for all i, p against the alternative H_A : there exists i^*, p^* such that $v^{(\ell)}_{i^*,p^*} \neq 0$. As in Nason (2013) if $v^{(\ell)}_{i^*,p^*}$ is nonzero then this is evidence for nonstationarities close to $t^* = 2^{-i^*}(p^* + 1/2)$ operating at a length scale of 2^{-i^*}. Hence, not only can an omnibus test of stationarity be constructed, but one can identify its location t^* and extent.

3.5 Haar wavelet coefficient test statistics and their properties

As in von Sachs and Neumann (2000) and Nason (2013) our test statistics consist of empirical versions of $v^{(\ell)}_{i,p}$ given by:

$$\hat{v}^{(\ell)}_{i,p} = 2^{-i/2} \left(\sum_{r=0}^{2^{i-1}-1} I_{\ell,2^i p - r} - \sum_{q=2^{i-1}}^{2^i-1} I_{\ell,2^i p - q} \right), \tag{19}$$

where $I_{j,k}$ was defined in (14) and i, p have the same range as in (18).

As in those previous works, and for the theory below to hold, we need the following assumption which later guides us as to which scales we can use in the (non-bootstrap) Haar wavelet coefficient test.

Assumption 5. *We only use coefficients $v^{(\ell)}_{i,p}$ at coarse enough scales i for which $2^i = \mathcal{O}(T)$.*

This assumption originates from von Sachs and Neumann (2000) who use it to control the wavelet coefficients' asymptotic bias. Another view is that approximate normality of the coefficients only occurs at 'coarse-enough' scales, as it is these coefficients that are sums of 'enough' wavelet packet periodogram values specified by (19).

To utilize the statistics (19) in a test we first establish the following Lemma.

Lemma 1. *Suppose that Assumptions 1–4 are satisfied. Then:*

(i) $\mathbb{E}\hat{v}^{(\ell)}_{i,p} = v^{(\ell)}_{i,p} + o(T^{-1/2})$,

(ii) $\mathrm{var}(\hat{v}^{(\ell)}_{i,p}) = 2T^{-1} \int_0^1 \beta^2_\ell(z)\{\psi^H_{i,p}(z)\}^2 \, dz + \mathcal{O}(T^{-1})$,

(iii) $|\,\mathrm{cum}_n(\hat{v}^{(\ell)}_{i,p})| \leq C^n (n!)^{2+2\gamma} T^{-1} (T^{-1/2})^{n-2}$,

all hold uniformly for ℓ, i, p as above, where $C < \infty$ is an arbitrary, but fixed, constant.

Proof: See the appendix.

Remark 5. The statement of Lemma 1 is identical to that in Nason (2013) and both are slight adaptions of the statement of Lemma 3.2 from Neumann and von Sachs (1997). However, the underlying quantities are completely different and the details of the proof are different because (i) our Lemma deals with locally stationary (Fourier) processes with spectra of bounded variation and (ii) the quantities underlying $\beta_\ell(z)$ are wavelet packets.

Asymptotic normality of the $\hat{v}_{i,p}^{(\ell)}$ now follows.

Proposition 2. *Suppose that Assumptions 1–4 hold. Let* $\Delta_T = K(\log T)^{1/2}$ *for any fixed* $K < \infty$. *Then:*

$$\mathbb{P}\{\pm(\hat{v}_{i,p}^{(\ell)} - v_{i,p}^{(\ell)})/\sigma_{i,p}^{(\ell)} \geq x\} = \{1 - \Phi(x)\}\{1 + o(1)\}, \tag{20}$$

holds uniformly in $-\infty \leq x \leq \Delta_T$ *with* i, p *in the range given in* (18)*, where* $\Phi(x)$ *is the standard normal CDF and* $\sigma_{i,p}^{(\ell)2} = \mathrm{var}(\hat{v}_{i,p}^{(\ell)})$.

Proof: Proposition 4 directly parallels Proposition 3.1 from von Sachs and Neumann (2000) and Proposition 1 from Nason (2013) and both rely on Proposition 3.1 from Neumann and von Sachs (1997) and the properties established in Lemma 1. □

We could now follow the established route of von Sachs and Neumann (2000) which takes X_t, forms its raw wavelet packet periodogram, $I_{\ell,k}$, constructs empirical Haar wavelet coefficients, $\hat{v}_{i,p}^{(\ell)}$ and performs significance tests using the asymptotic normality result furnished in Proposition 4. As in von Sachs and Neumann (2000) or Nason (2013) we have to use some form of multiple test size control such as Bonferroni or false discovery rate due to Benjamini and Hochberg (1995). Up to this point the innovation would be the use of wavelet packets instead of wavelets and hence a potential power gain due to the extensive range of wavelet packet basis functions (of which wavelets are a subset).

Other benefits of the Bonferroni approach are that it does not require knowledge of the joint distribution of the wavelet coefficients and, related to this, results in a rather conservative test. In our context conservatism is a good thing as we would not want to force analysts to use complicated nonstationary methods unless we absolutely have to. A further benefit of the Bonferroni testing is that it is relatively simple to elicit the theoretical power properties of the test, see Section 5.

Another possibility would be to use some combination of the $\hat{v}_{i,p}^{(\ell)}$ coefficients to obtain some kind of omnibus test. For example, using $\max_{\ell,i,p} |\hat{v}_{i,p}^{(\ell)}/\sigma_{i,p}^{(\ell)}|$ as the basis of a test statistic and examining its asymptotic distribution to obtain critical values. Unfortunately, all of these quantities require estimation of $\sigma_{i,p}^{(\ell)}$ which is known to be problematic. Hence, the next section adopts a different tack by using the bootstrap which additionally permits us to make use of the important finer scale coefficients.

4 A Bootstrap Test

Another recent test of stationarity (for wavelet or Fourier) was introduced by Cardinali and Nason (2010). The underlying test statistic was also suggested by Dette et al. (2011) who also established the statistic's asymptotic normality to provide inference.

The wavelet version of the Cardinali and Nason (2010) test used the metric:

$$T_{\mathrm{C}}\{S_j(z)\} = J^{-1} \sum_{j=1}^{J} \int_0^1 \{S_j(z) - \bar{S}_j\}^2 \, dz, \tag{21}$$

where $\bar{S}_j = \int_0^1 S_j(z)\,dz$ to measure departures of $S_j(z)$ from constancy and the notation T_{C} indicates a test statistic that looks for departures from constancy of the spectrum. A test statistic was formed by estimating $S_j(z)$ by an asymptotically unbiased estimator obtained by applying the matrix $\mathcal{A}^{-1}$ to the wavelet periodogram in (16) and inserting it into (21).

There is a useful theoretical equivalence between the tests in Cardinali and Nason (2010) and Nason (2013).

Proposition 3. *The time series $\{X_{t,T}\}$ is stationary iff the Haar wavelet coefficients of $\{\beta_j(z)\}_{j=1}^{\infty}$ are all zero iff $T_{\mathrm{C}}\{S_j(z)\} = 0$ for $j \in \mathbb{N}$.*

Proof: See the appendix.

Although there is a theoretical equivalence, Cardinali and Nason (2013) indicate that the T_{C} bootstrap test has generally better empirical size and power characteristics compared to the multiple Haar wavelet coefficient test using Bonferroni correction from Nason (2013). There appear to be two reasons for this:

1. The empirical tests in von Sachs and Neumann (2000) and Nason (2013) both make use of Assumption 5 to provide asymptotic normality for inference. This means that only medium- or coarse-scale Haar wavelet coefficients of $\beta_j(z)$ from Nason (2013), or $f(u,\omega)$ for von Sachs and Neumann (2000), are used in the stationarity test. The test statistic, T_{C}, in (21) implicitly used *all* scales including the finest.

2. von Sachs and Neumann (2000) and Nason (2013) both estimate each wavelet coefficients' variance using a global quantity. This variance is $\hat{\sigma}_I^2$ in (3.6) from von Sachs and Neumann (2000) and $\hat{\sigma}_{i,p}^{(\ell)}$ in (9) in Nason (2013). These estimators enable theoretical establishment of the size of the test under H_0, but they result in excessively conservative tests. The reason being that nonstationarity, as well as inducing large Haar wavelet coefficients of β_ℓ or f, also can cause spuriously large values for the estimators $\hat{\sigma}_I^2$ and $\hat{\sigma}_{i,p}^{(\ell)}$. Under the alternative hypothesis what is required is the *local* variance of a particular wavelet coefficient and not the whole-series time-averaged variance estimate appropriate for the stationary case (which is what the existing literature prescribes).

Taking these points together we can construct our second new test of stationarity as follows.

Definition 4. *Let $\mathcal{L}_1 = \{\ell_1, \ldots, \ell_L\}$ be a fixed set of wavelet packets where $\ell_p \in \mathcal{L}$ for $p = 1, \ldots, L$ for some $L \in \mathbb{N}$.*

(a) The wavelet packet measure *is given by*

$$T_{\mathrm{C}}\left[\{\beta_\ell(z)\}_{\ell\in\mathcal{L}_1}\right] = L^{-1}\sum_{\ell\in\mathcal{L}_1}\int_0^1 \{\beta_\ell(z) - \bar{\beta}_\ell\}^2 dz, \tag{22}$$

where $\bar{\beta}_\ell = \int_0^1 \beta_\ell(z)\,dz$.

(b) The empirical wavelet packet test statistic $\hat{T}_{\mathrm{vS}}$ is obtained by substituting $I_{\ell,k}$ for its expectation $\beta_\ell(k/T)$ in (22) *and replacing the integral by the sum over all possible time values, k.*

The next result quantifies the asymptotic properties of the test statistic under the stationary null hypothesis.

Proposition 4. *Let $\{X_{t,T}\}$ satisfy Assumptions 1–4 and, in addition, be second-order stationary with associated b-spectrum of $(\beta_{\ell_1}, \ldots, \beta_{\ell_L})$ for some fixed set of $\ell_p \in \mathcal{L}$ for $p = 1, \ldots, L$. Let $R_{\ell,k} = I_{\ell,k} - \bar{I}_\ell$ and $S_{\ell,k} = R^2_{\ell,k}$. Then $\mu_\ell := \mathbb{E}(S_{\ell,k}) < \infty$ and $\sigma^2_\ell := \mathrm{var}(S_{\ell,k}) < \infty$ and define $\mu = (TL)^{-1} \sum_{i=1}^{L} \mu_{\ell_i}$ and $\sigma^2 = L^{-2} \sum_{i=1}^{L} \sigma^2_\ell$. Then $\hat{T}_{\mathrm{vS}}$ is asymptotically $N(\mu, \sigma^2/T)$.*

Proof: See the appendix.

One can obtain a similar result using a modified test statistic by replacing $\beta_\ell(z), \hat{\beta}_\ell$ by their projections onto a Haar wavelet basis of scale greater than i to satisfy Assumption 5. Then Proposition 4 can be used to show asymptotic normality for the β estimates, χ^2 for the $S_{\ell,k}$ in Theorem 1 and hence overall asymptotic normality.

With the asymptotic normality result we could, in principle, use this to construct an appropriate critical region for a relatively standard hypothesis test. However, accurate estimation of μ_ℓ and σ^2_ℓ is challenging for reasons similar to why estimating $\hat{\sigma}^2_I$ and $\hat{\sigma}^{(\ell)}_{i,p}$ above is challenging. It is in this situation that a bootstrap test can produce good results with the bootstrap samples providing insurance by exploring the variability around estimates of μ_ℓ and σ^2_ℓ. We adopt the bootstrap procedure from Cardinali and Nason (2010) which uses the surrogate sampling techniques of Davies and Harte (1987) and Percival and Constantine (2006) and the `fractal` package Constantine and Percival (2007). Our bootstrap test algorithm is as follows.

`BootWPTOS`:

1. The user specifies a set $\mathcal{L}_1$ of wavelet packet indices, and a number B of bootstrap simulations.
2. The test statistic in (22) is computed on the data using $\mathcal{L}_1$ and its valued recorded as $\hat{T}_{\mathrm{vS},1}$.
3. For $b = 2, \ldots, B$ we generate a surrogate series using the `"phase"` option of the `surrogate()` function from the `fractal` package. We compute the test statistic in (22) using $\mathcal{L}_1$ on the surrogate and record its value as $\hat{T}_{\mathrm{vS},b}$.
4. The p-value of the test is given by $p = \#_{b=2,\ldots,B}\{i : \hat{T}_{\mathrm{vS},b} \geq \hat{T}_{\mathrm{vS},1}\}/B$.

The `"phase"` option of the `surrogate` function implements the phase randomization technique due to Theiler et al. (1992) which produces a simulated time series with the same estimated Fourier spectral structure as the original series. Theiler et al. (1992) is more appropriate than the LSW simulation technique used in Nason (2013) since, under the null hypothesis of stationarity, the process has a Fourier/Cramer representation.

5 Theoretical Power

A hypothesis test's theoretical power can be useful to compare tests, or elicit quantities such as the minimum observable amount of nonstationarity for a given sample size, or what sample size is necessary to detect certain kinds of nonstationarity. This section derives expressions for the theoretical power for three sample alternative hypotheses for the stationarity tests introduced in Section 3.5.

Example: TVMA(1). The spectrum for a time-varying moving average model of order one, X_t, with mean zero, unit variance innovations, is given by $f(z,\omega) = \pi^{-1}|1+a(z)e^{i\omega}|^2$. Here, the function $a : (0,1) \to \mathbb{R}$ controls the degree of nonstationarity over rescaled time, z, and X_t is stationary if $a(z)$ is a constant function of $z \in (0,1)$. For this example we will let $a(z) = c - mz$ where $c, m \in \mathbb{R}$. Clearly, if $m = 0$ then X_t is stationary. We will consider Haar wavelet packets and, for example, the Fourier transform for the finest scale father wavelet is given by

$$\hat{\psi}_{(1,0)}(\omega) = 2^{-1/2}(1 + e^{-i\omega}), \tag{23}$$

and one of second-finest scale wavelet packets that is not the mother or father wavelet is given by

$$\hat{\psi}_{(2,2)}(\omega) = 2^{-1}(1 - e^{-i\omega} + e^{-2i\omega} - e^{-3i\omega}), \tag{24}$$

the other wavelet packets similarly defined. Using formula (17) the $\beta_\ell(z)$ functions for these two packets and the spectrum $f(z,\omega)$ are — ignoring remainder terms —

$$\beta_{(1,0)}(z) = 2(1 + c + c^2 - mz - 2cmz + m^2z^2), \tag{25}$$

and

$$\beta_{(2,2)}(z) = 2c^2 - c(4mz+3) + 2m^2z^2 + 3mz + 2, \tag{26}$$

respectively.

Next we calculate the Haar wavelet coefficients using formula (18). The Haar wavelet coefficients of (25) are given by

$$v_{i,p}^{(1,0)} = 2^{-\frac{5i}{2}-1} m \left\{ c2^{i+1} + 2^i - (2p+1)m \right\}, \tag{27}$$

and of (26) are given by

$$v_{i,p}^{(2,2)} = 2^{-\frac{5i}{2}-2} m \left\{ c2^{i+2} - 3\,2^i - 2(2p+1)m \right\}, \tag{28}$$

and, as expected, $v_{i,p}^{(1,0)} = v_{i,p}^{(2,2)} = 0$ for $m = 0$ and X_t stationary.

Similarly, using Lemma 1(ii) we can show that

$$\begin{aligned} \mathrm{var}(\hat{v}_{i,p}^{(1,0)}) = T^{-1} \Big(& -5(2c+1)\left(c^2+c+1\right) 2^{3-i} m(2p+1) + 40\left(c^2+c+1\right)^2 \\ & - 5(2c+1)2^{2-3i} m^3 [2p\{p(2p+3)+2\}+1] \\ & + 5\{2c(c+1)+1\}2^{3-2i} m^2 \{3p(p+1)+1\} \\ & + 2^{3-4i} m^4 \left\{ 5p(p+1)\left(p^2+p+1\right)+1 \right\} \Big) /5, \end{aligned} \tag{29}$$

and

$$\begin{aligned} \mathrm{var}(\hat{v}_{i,p}^{(2,2)}) = 2T^{-1} \big(& -15(4c-3)8^{-i} m^3 [2p\{p(2p+3)+2\}+1] \\ & + 5\{12c(2c-3)+17\}4^{-i} m^2 \{3p(p+1)+1\} \\ & - 15(4c-3)\{c(2c-3)+2\}2^{-i} m(2p+1) + 15\{c(2c-3)+2\}^2 \\ & + 3\,4^{1-2i} m^4 \left\{ 5p(p+1)\left(p^2+p+1\right)+1 \right\} \big) /15. \end{aligned} \tag{30}$$

Similar formulae can be derived for other packets.

For a single Haar wavelet coefficient for the wavelet packet $(1,0)$, e.g., and using the asymptotic normality result of Proposition 4 the approximate power, for large T, of the size α test is given by:

$$\rho_{(1,0)}(\alpha, T, i, p) = 1 + \Phi\left\{ -C_\alpha - T^{1/2} v_{i,p}^{(1,0)} / s_{i,p}^{(1,0)} \right\} - \Phi\left\{ C_\alpha - T^{1/2} v_{i,p}^{(1,0)} / s_{i,p}^{(1,0)} \right\}, \tag{31}$$

where C_α is the usual critical value $C_\alpha = \Phi^{-1}(1-\alpha/2)$ and $s_{i,p}^{(\ell)2} = T\,\mathrm{var}(\hat{v}_{i,p}^{(\ell)})$.

Using the Bonferroni method of multiple hypothesis test size control, the overall power of the test of stationarity using a collection of Haar wavelet coefficients associated with a single packet, e.g. $(1,0)$ is given by

$$\mathrm{P}_{(1,0)}(\alpha, T) = 1 - \prod_{(i,p)\in\mathcal{I}} \{1 - \rho_{(1,0)}(\alpha/|\mathcal{I}|, T, i, p)\}, \tag{32}$$

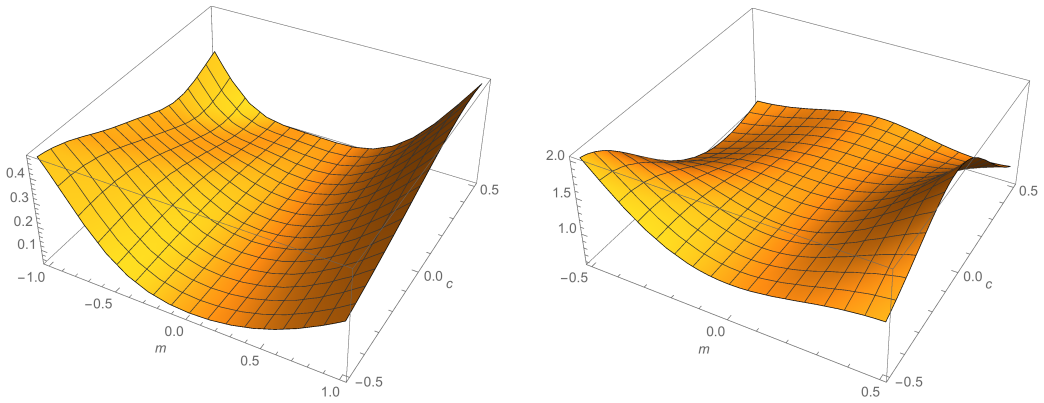

Figure 3: *Left:* Power function $P_{(2,2)}(\alpha, T)$. *Right:* Ratio of power functions $P_{(1,1)}(\alpha, T)/P_{(2,1)}(\alpha, T)$. Here $\alpha = 5\%$, $T = 512$ and number of tests $|\mathcal{I}| = 6$ in both cases.

where $\mathcal{I}$ is the set of Haar coefficients used in the test and $|\mathcal{I}|$ is the number of coefficients used. Given the complicated form of quantities like (27) and (29) it does not seem likely that a simple immediately interpretable formula for the overall power, $P_{(1,0)}$ exists although it can be calculated exactly.

For example, Figure 3 shows a plot of a power function and the ratio of two power functions. In the left plot of Figure 3 the power reaches high values of 40–50% when $(c, m) = (-0.5, -1)$ and $(c, m) = (0.5, 1)$ respectively. Further, the power is 5%, equivalent to the size of the test, when $m = 0$ and the underlying process is stationary as one would expect. The right plot in Figure 3 shows the ratio of power functions for packets $(1, 1)$ and $(2, 1)$ and shows that the former test is about twice as powerful as the latter for the parameter values of $(c, m) = (-0.5, -0.5)$, e.g.

Example: Piecewise stationary with single jump discontinuity. For this example, suppose the time-varying spectrum is given by

$$f(z, \omega) = \begin{cases} f_1(\omega) & \text{for } 0 \leq z < \frac{1}{2}, \\ f_2(\omega) & \text{for } \frac{1}{2} \leq z < 1. \end{cases} \tag{33}$$

Then using (17) and ignoring remainders means that $\beta_\ell(z) = K_1^{(\ell)}$ for $0 \leq z < 1/2$ and $\beta_\ell(z) = K_2^{(\ell)}$ for $1/2 \leq z < 1$ where $K_i^{(\ell)} = \int_{-\pi}^{\pi} f_i(\omega)|\hat{\psi}_\ell(\omega)|^2 \, d\omega$ for $i = 1, 2$. As the breakpoint has been deliberately set at $z = 1/2$ the only non-zero wavelet coefficient is $v_{0,0} = \{K_1^{(\ell)} - K_2^{(\ell)}\}/2$. The variance of the $(0, 0)$ Haar wavelet coefficient is given by Lemma 1(ii) as

$$\text{var}(\hat{v}_{0,0}^{(\ell)}) = 2T^{-1} \int_0^1 \beta_\ell^2(z)\phi(z)dz = 2T^{-1} \int_0^1 \beta_\ell^2(z) \, dz = 2T^{-1}(K_1^{(\ell)2} + K_2^{(\ell)2}). \tag{34}$$

Now define $\Delta = \frac{1}{2}(K_1^{(\ell)} - K_2^{(\ell)})\{K_1^{(\ell)2} + K_2^{(\ell)2}\}^{-1/2}$, which is a normalised measure of the jump between the first half of the spectrum and the second. Then, it can be shown that the

power function with N_T Haar wavelet coefficient tests, of overall size α, with Bonferroni correction, sample size T is given to a high degree of approximation by

$$P_{(\ell)}(\alpha, T) = 1 - \left(1 - \frac{\alpha}{N_T}\right)^{N_T - 1} \left\{\Phi(C_{\alpha/N_T} - T^{1/2}\Delta) - \Phi(-C_{\alpha/N_T} - T^{1/2}\Delta)\right\}, \tag{35}$$

where the last term in the product corresponds to the power associated with the single non-zero Haar wavelet coefficient.

Equation (35) shows that, for a general nonstationary alternative $f(z, \omega)$, every extra non-zero wavelet coefficient causes terms in the product to change from $1 - \alpha/N_T$ to $\Phi(C_{\alpha/N_T} - T^{1/2}\Delta_{i,p}) - \Phi(-C_{\alpha/N_T} - T^{1/2}\Delta_{i,p})$. This causes the asymptotic power to be equal to one in each of these cases. Hence, this kind of test is statistically consistent.

Theoretical power comparison with Dette et al. (2011). Here we compare the theoretical power of our packet test with the D^2 test from Dette et al. (2011). Their Example 1 considers the TVMA(2) process $X_{t,T} = \cos(2\pi t/T)Z_t - (t/T)^2 Z_{t-1}$, where $\mathrm{var}(Z_t) = \sigma^2$ is a stationary white noise process. The nonstationary spectrum associated with this alternative hypothesis is given by:

$$f(z, \omega) = \{\cos(2\pi z)^2 - 2z^2 \cos(2\pi z)\cos(\omega) + z^4\}. \tag{36}$$

Their test criterion is the minimal distance $D^2 = \min_g \int_{-\pi}^{\pi} \int_0^1 \{f(z, \omega) - g(\omega)\}^2 \, dz d\omega$ and its minimizing function is shown to be $g(\omega) = 7/(20\pi) - \cos(\omega)/(2\pi^3)$. Working through the asymptotic theory in Dette et al. (2011) Section 3 we can use their asymptotic normality result for their test statistic, $\hat{D}_T^2$, from Theorem 2 to show that for $T = 128$, $\alpha = 0.05$ that the theoretical power is 72.3%. For $T = 512$ this increases to 91.9%. This assumes that we choose the 'number of blocks' parameter $M = 8$ as suggested in their Section 4.1.1.

For our test, we can derive $\beta_\ell(z)$ exactly for the alternative hypothesis spectrum (36) for each packet ℓ, e.g., $\beta_{(2,2)}(z) = \left\{\cos(4\pi z) + 3z^2 \cos(2\pi z) + 2z^4 + 1\right\}/2$. We can also derive $v_{i,p}^{(\ell)}$ and $\mathrm{var}(\hat{v}_{i,p}^{(\ell)})$ analytically although the expressions are long and complicated. Using these we can form exact expressions for the analytical power of any combination of Haar wavelet coefficients as we did earlier in this section. For the test in Section 3.4 using the two finest scales of Haar wavelet coefficients with $\alpha = 0.05$ and $T = 128$ the power for each individual packet test is $P_{(1,0)}(0.05, 128) = 62.2\%$, $P_{(1,1)}(0.05, 128) = 80.0\%$, $P_{(2,0)}(0.05, 128) = 75.6\%$, $P_{(2,1)}(0.05, 128) = 61.1\%$, $P_{(2,2)}(0.05, 128) = 86.2\%$ and $P_{(2,3)}(0.05, 128) = 71.9\%$. For $T = 512$ the theoretical power for each packet is greater than 99.93%. So, for $T = 128$, four out of the six packets in our test have greater power than the D^2 test and the remaining two packets have power within 12%. For $T = 512$ <u>all</u> packets have greater power. Hence, a packet-based test can produce superior results when compared to the D^2-based statistic.

Importantly, the theoretical power of wavelet packets that are not wavelets (e.g. $\ell = (2, 2)$) is greater than that of the wavelets alone. Hence, there is an advantage in using a wavelet packet based test compared to one based on wavelets alone. This theoretical fact is further supported by simulations in the next section.

6 Example and Simulations

6.1 Infant Electrocardiogram Data

The Infant Electrocardiogram data was originally analyzed in Nason et al. (2000, Section 4.2) and their first differences analysed in Nason (2008, Section 5.3.7). From the time series plot one might think that the series is obviously non-stationary but it is not impossible for stationary time series models to mimic much of the behaviour of such series and hence the need for an objective test. The first differences were tested for stationarity in Nason (2013) and a plot indicating the location and extent of the nonstationarities arising from the significant Haar coefficients was presented in Figure 3 of that paper.

The plots produced here are similar except that we produce one plot per wavelet packet selected. Figure 4 shows the differenced infant electrocardiogram time series (in grey) with significant Haar wavelet coefficients superimposed as double-headed arrows. We first executed the `BootWPTOS` 'omnibus' test of stationarity from Section 4 on all packets $\mathcal{L}_1 = \{\ell = (j, m) : j = 1, \ldots, 3; m = 1, \ldots, 2^j - 1\}$, i.e. all 11 (non-father wavelet) wavelet packets at scales $j = 1, \ldots, 3$ which resulted in a very small p-value much less than 0.01. We then selected four particular packets: $(1, 1)$, $(2, 2)$, $(2, 3)$ and $(3, 5)$ and applied the Haar wavelet coefficient test from Section 3.4 to each one. So, for example, the top-right plot in Figure 4 corresponds to packet $(2, 2)$: wavelet packet two at the second-finest scale (which is not a wavelet). For this plot there are three significant Haar wavelet coefficients, $\hat{v}^{(2,2)}_{5,22}$ at resolution level 5 and $\hat{v}^{(2,2)}_{6,43}$, $\hat{v}^{(2,2)}_{6,45}$ at resolution level 6. The other three plots correspond to three other highly significant packets.

All of the plots identify nonstationarities around $t = 1400$; these were identified also by the 'wavelet-only' plot in Figure 3 of Nason (2013). The bottom-right plot of Figure 4 also identifies nonstationarities around $t = 500$, and these too were identified by Figure 3 of Nason (2013). However, the wavelet packet test here also *additionally* identifies some plausible nonstationarities at about $t = 1600$ (bottom two plots) and $t = 1750$ (top left, bottom right).

6.2 Size Simulations and Comparisons

Nason (2012, 2013) presented a comprehensive range of simulations to evaluate the empirical size and power characteristics of stationarity tests for time series with sample sizes of $2^9 = 512$. In particular, they compare the Priestley-Subba Rao (PSR) test with the new test in Nason (2013). Here, we repeat these simulations for our new testing procedure for two cases (A) $\mathcal{L}_A = \{(1, 1), (2, 1), (3, 1)\}$, wavelets at the first three finest scales and (B)

$$\mathcal{L}_B = \{(1, 1), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (3, 7)\},$$

all wavelet packets at the first three finest scales apart from the father wavelet coefficients.

Nason (2012, 2013) simulated data from a number of stationary models using Gaussian innovations and assessed how often their tests rejected the null hypothesis. The models are:

S1 iid standard normal;

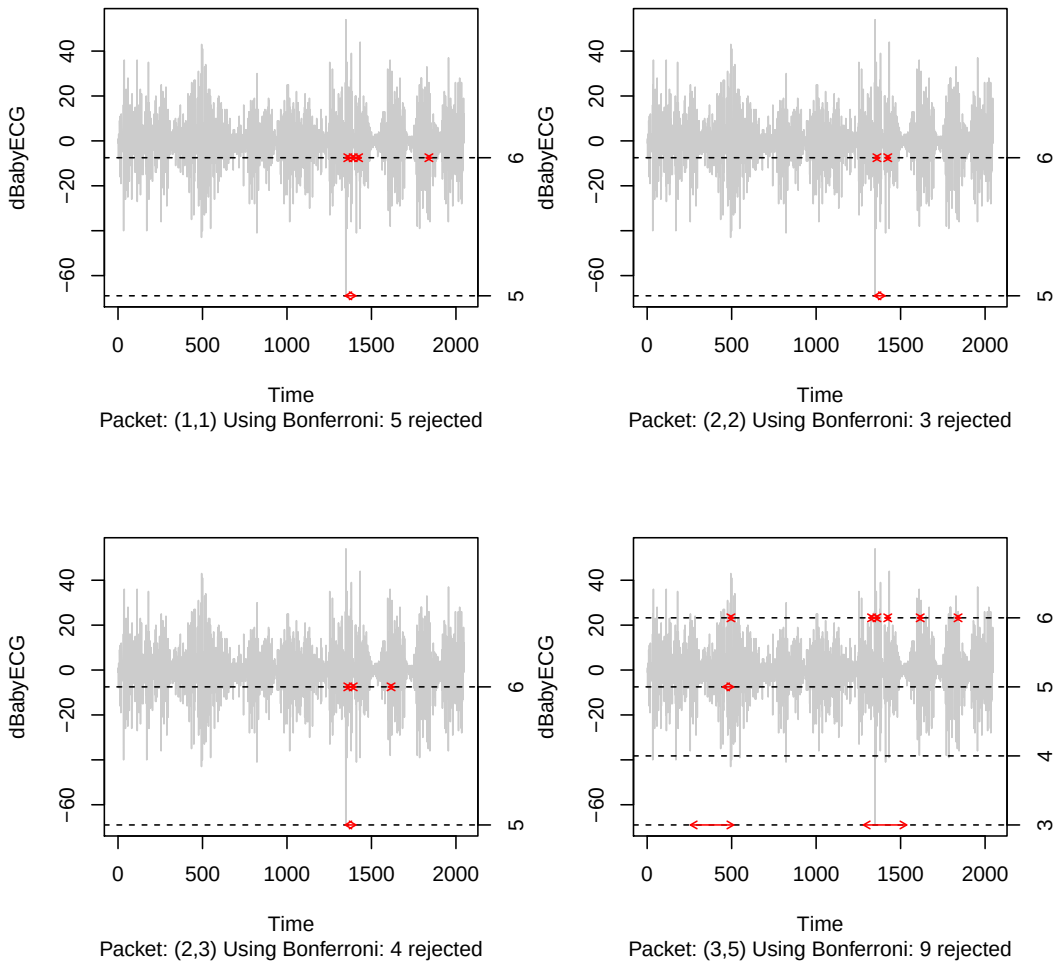

Figure 4: Plots identifying significant Haar wavelet coefficients, $\hat{v}_{i,p}^{(\ell)}$, that reject the null hypothesis of stationarity for four different wavelet packets, ℓ. Grey time series are first differences of infant electrocardiogram data (`dBabyECG`). Right-hand axis indicates scale of discovered Haar wavelet coefficients, i.

Table 1: Simulated size estimates (%) for stationary Gaussian models $T = 512$ based on 1000 simulations and $B = 1000$ with 5% nominal power.

	From Nason (2013)		BootWPTOS	
Model	PSR	HWTOS	$\mathcal{L}_A$	$\mathcal{L}_B$
S1	5.6	4.3	5.3	5.7
S2	12.4	4.0	10.6	11.5
S3	6.2	20.3	5.3	5.6
S4	6.0	3.4	4.5	4.9
S5	6.5	0.7	5.2	5.7
S6	7.5	0.1	5.3	6.6
S7	23.9	7.3	5.3	5.3

S2 AR(1) model with AR parameter of 0.9 with standard normal innovations;

S3 As S2 but with AR parameter of -0.9;

S4 MA(1) model with parameter of 0.8;

S5 As S4 but with parameter of -0.8.

S6 ARMA$(1, 0, 2)$ with AR parameter of -0.4, and MA parameters of $(-0.8, 0.4)$.

S7 AR(2) with AR parameters of $\alpha_1 = 1.385929$ and $\alpha_2 - 0.9604$. The roots associated with the auxiliary equation, see Chatfield (2003, page 44), are $\beta_1 = \bar{\beta_2} = 0.98e^{i\pi/4}$. This process is stationary, but close to the 'unit root': a 'rough' stochastic process with spectral peak near $\pi/4$.

The empirical size values are shown in Table 1 and demonstrates that our new BootWPTOS tests have superior size characteristics compared to both the PSR and HWTOS tests. The ranges of PSR and HWTOS are $(5.6, 23.9)$ and $(0.1, 20.3)$ respectively whereas the ranges for the BootWPTOS tests, $(4.5, 10.6)$ and $(4.9, 11.5)$, are much smaller and closer to 5%. A notable outlier is the poor calibration of the wavelet packet and PSR tests in response to S2. We speculate that this is due to the dominant low-frequency behaviour of the S2 process which might mask as trend, which none of these tests are designed to handle without trend removal, as mentioned in Remark 4 in Section 3.3. The empirical sizes of the BootWPTOS tests are more uniformly closer to the nominal size of 5%.

Table 2: Simulated power estimates (%) for models (4.6)–(4.10) with a nominal size of 5% based on 1000 simulations each with $B = 200$ bootstrap simulations.

		Model				
	Method	(4.6)	(4.7)	(4.8)	(4.9)	(4.10)
$T = 256$	DPV11	82.8	68.1	89.9	20.7	74.0
	HWTOS	37.6	28.6	31.7	37.4	37.8
	BootWPTOS	96.4	93.2	96.0	52.4	99.7
$T = 512$	DPV11	98.6	94.3	99.7	27.7	95.1
	HWTOS	73.5	66.5	85.7	78.9	99.6
	BootWPTOS	100	100	100	78.2	100
$T = 1024$	DPV11	100	99.9	100	38.6	99.0
	HWTOS	99.9	99.7	100	100	100
	BootWPTOS	100	100	100	94.4	100

6.3 Simulated Power Comparisons

6.3.1 Comparisons to Dette et al. (2011) and Nason (2013)

Some nonstationary alternatives specified by Dette et al. (2011) are given by

$$
\begin{aligned}
(4.6) \quad & X_{t,T} = 1.1\cos\{1.5 - \cos(4\pi t/T)\}Z_{t-1} + Z_t, \\
(4.7) \quad & X_{t,T} = 0.6\sin(4\pi t/T)X_{t-1} + Z_t, \\
(4.8) \quad & X_{t,T} = (0.5X_{t-1} + Z_t)I_{\{[1,T/4]\cup[3T/4,T]\}}(t) \\
& \qquad + (-0.5X_{t-1} + Z_t)I_{[T/4+1,3T/4]}(t), \\
(4.9) \quad & X_{t,T} = (-0.5X_{t-1} + Z_t)I_{\{[1,T/2]\cup[T/2+T/64+1,T]\}}(t) \\
& \qquad + 4Z_t I_{[T/2+1,T/2+T/64]}(t),
\end{aligned}
$$

where $I_A(t)$ denotes the indicator function of a set A and the equation labels are those specified in Dette et al. (2011). The final model we will consider is the following locally stationary wavelet process:

$$
(4.10) \quad X_{t,T} = \sum_{k=0}^{T-1} w_1(k/T)\psi_{1,k-t}Z_{1,k} + \sum_{k=0}^{T-1} w_2(k/T)\psi_{2,k-t}Z_{2,k},
$$

where $w_1(x) = 0.6\cos(4\pi x)$, $w_2(x) = 0.8x^2$ and $\psi_{1,\cdot}, \psi_{2,\cdot}$ are the finest and second-finest scale nondecimated Haar discrete wavelets from Nason et al. (2000). In all of these models $Z_t, Z_{1,k}, Z_{2,k}$ are Gaussian white noise with mean zero and variance one. Table 2 shows simulations where the method from Dette et al. (2011) beats the single basis wavelet test (`HWTOS`) for all models apart from (4.9). However, the new wavelet packet test, `BootWPTOS`, dominates both `HWTOS` and the test from Dette et al. (2011). Here the `BootWPTOS` tests uses the four, five and six finest packet scales with all packets for sample sizes of $T = 256, 512$ and 1024 respectively.

Table 3: Simulated power estimates (%) for models P1–P4 with nominal size of 5% based on 1000 simulations and $B = 1000$.

	From Nason (2013)		BootWPTOS	
Model	PSR	HWTOS	$\mathcal{L}_A$	$\mathcal{L}_B$
P1	37.2	99.7	99.3	99.7
P2	100	17.3	74.9	83.5
P3	44.3	1.3	17.2	22.5
P4	100	94.8	99.1	100

6.3.2 Comparisons to PSR and Nason (2013)

The models are:

P1 Time-varying AR model $X_t = \alpha_t X_{t-1} + \epsilon_t$ with iid standard normal innovations and the AR parameter evolving linearly from 0.9 to -0.9 over the 512 observations.

P2 A LSW process based on Haar wavelets with spectrum $S_j(z) = 0$ for $j > 1$ and $S_1(z) = \frac{1}{4} - (z - \frac{1}{2})^2$ for $z \in (0, 1)$. This process is, of course, a time-varying moving average process.

P3 A LSW process based on Haar wavelets with spectrum $S_j(z) = 0$ for $j > 2$ and $S_1(z)$ as for P2 and $S_2(z) = S_1(z + \frac{1}{2})$ using periodic boundaries (for the construction of the spectrum only).

P4 A LSW process based on Haar wavelets with spectrum $S_j(z) = 0$ for $j = 2, j > 4$ and $S_1(z) = \exp\{-4(z - \frac{1}{2})^2\}$, $S_3(z) = S_1(z - \frac{1}{4})$, $S_4(z) = S_1(z + \frac{1}{4})$ again assuming periodic boundaries.

Spectra and single realisations for these processes are displayed in Nason (2012). Table 3 shows that the BootWPTOS test based on wavelet packets $\mathcal{L}_B$ is more powerful than the BootWPTOS test based on wavelets alone ($\mathcal{L}_A$). Secondly, our new BootWPTOS test dominates HWTOS for these simulations. So, we compare PSR to BootWPTOS $\mathcal{L}_B$: (P1) BootWPTOS $\mathcal{L}_B$ dominates; (P2) PSR dominates, but the BootWPTOS is not far behind with 83.5%, considerably better than 17.3% from HWTOS; (P3): None of the three tests is really impressive; PSR performs better than BootWPTOS $\mathcal{L}_B$, which is itself much better than the 1.3% from HWTOS; (P4) both perform extremely well.

Our overall conclusion from the size and power simulations is that BootWPTOS $\mathcal{L}_B$ appears to be more stable with empirical size values closer to the nominal. However, at the same time, BootWPTOS $\mathcal{L}_B$ exhibits excellent power properties in many, but not all, situations.

7 Conclusions and Further Work

This article introduces two new tests of second-order stationarity based on assessing the constancy of a wavelet packet periodogram (the b-spectrum). Two forms of assessment

are proposed: one via Haar wavelet coefficients of the periodogram and one using an L_2 norm. Theoretical and empirical investigation shows superior power characteristics of the new wavelet packet based tests whilst still retaining good control of statistical size. We also present an R software package, `BootWPTOS`, which carries out the tests and provides graphical depiction of the location and extent of the nonstationarities. We intend that `BootWPTOS` be made freely available on the `CRAN` R software archive in due course.

The current work assumes the use of a fixed mother wavelet (underlying wavelet and wavelet packet methods) and also a fixed choice of wavelet packet basis functions. An interesting question for future study is whether it is possible to construct a method that adaptively chooses the best wavelet packet basis functions to use to test for stationarity and to develop theory for increasing numbers of packets as T increases. Another interesting possibility is whether it is possible to use the bootstrap and the asymptotic normality results in a hybrid method: the bootstrap is slower than the methods of inference using the asymptotic normality. Might one use asymptotic normality tests to quickly identify packets of interest, followed by a bootstrap to validate and, perhaps, refine the test? Of course, such a more complex two-stage procedure would require careful and rigorous examination of its size and power characteristics.

8 Acknowledgements

GPN was supported by the Research Councils UK Energy Programme grant EP/I01697X/1 and EPSRC grant EP/K020951/1. The Energy Programme is an RCUK cross-council initiative led by EPSRC and contributed to by ESRC, NERC, BBSRC and STFC. We acknowledge and are grateful to Dr. Philip Preuß for code and advice on the specification of the models and operation of his simulations.

A Appendix

A.1 Proof of Proposition 1

The proof proceeds as follows with $z = k/T$ for $z \in (0,1)$ and some $k \in \mathbb{Z}$:

$$\begin{aligned}
\beta_\ell(z) &= \mathbb{E}[I_{j,k}] = \mathbb{E}[|d_{j,k}|^2] && (37)\\
&= \sum_t \sum_s \psi_{\ell,k-t}\psi_{\ell,k-s}\mathbb{E}[X_t\overline{X_s}] && (38)\\
&= \sum_{t,s} \psi_{\ell,k-t}\psi_{\ell,k-s}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} A^0_{t,T}(\omega)\overline{A^0_{s,T}(\nu)}e^{i(\omega t-\nu s)}\mathbb{E}[d\xi(\omega)d\xi(\nu)]\\
&= \sum_{t,s} \psi_{\ell,k-t}\psi_{\ell,k-s}\int_{-\pi}^{\pi} A^0_{t,T}(\omega)\overline{A^0_{s,T}(\omega)}e^{i\omega(t-s)}d\omega, && (39)
\end{aligned}$$

as $\{\xi(\omega)\}$ is an orthonormal increments process. Using the fact that A^0 is in a close pair with A and the continuity of A we obtain:

$$\begin{aligned}
\beta_\ell(z) &= \sum_{t,s} \psi_{\ell,k-t}\psi_{\ell,k-s} \int_{-\pi}^{\pi} \left\{ A\left(\frac{k}{T},\omega\right) + \mathcal{O}(T^{-1})\,\mathrm{TV}_{[0,1]\times\Pi}(A) \right\} && (40)\\
&\times \left\{ \overline{A\left(\frac{k}{T},\omega\right)} + \mathcal{O}(T^{-1})\,\mathrm{TV}_{[0,1]\times\Pi}(A) \right\} e^{i\omega(t-s)} d\omega && (41)\\
&= U + V + W, && (42)
\end{aligned}$$

using similar arguments to the bounding of $|R_T^{(1,1)}|$ on page 66 of Neumann and von Sachs (1997), where

$$\begin{aligned}
U &= \int_{-\pi}^{\pi} \left| A\left(\frac{k}{T},\omega\right) \right|^2 \sum_t \psi_{\ell,k-t} e^{i\omega t} \sum_s \psi_{\ell,k-s} e^{-i\omega s} d\omega && (43)\\
&= \int_{-\pi}^{\pi} f(z,\omega) |\hat{\psi}_\ell(\omega)|^2 d\omega, && (44)
\end{aligned}$$

as $\sum_s \psi_{\ell,k-s} e^{-i\omega s} = \sum_n \psi_{\ell,n} e^{-i\omega(k-n)} = e^{-i\omega k} \overline{\sum_n \bar{\psi}_{\ell,n} e^{-i\omega n}} = e^{-i\omega k} \overline{\hat{\psi}_\ell(\omega)}$ since the discrete wavelet packets are real-valued (in this case). Similarly, $\sum_t \psi_{\ell,k-t} e^{i\omega t} = e^{i\omega k} \sum_n \psi_{\ell,n} e^{-i\omega n} = e^{i\omega k} \hat{\psi}_\ell(\omega)$. In forming the product of these two terms the phase terms cancel leaving $|\hat{\psi}_\ell(\omega)|^2$ as required.

The order of magnitude of the V term can be established by looking at the product of $A(k/T,\omega)$ from (40) and the $\mathcal{O}(1/T)$ term from (41) (and vice versa, hence the 2 in the next expression):

$$\begin{aligned}
|V| &\leq 2 \left| \sum_{t,s} \psi_{\ell,k-s}\psi_{\ell,k-t} \mathcal{O}(T^{-1})\,\mathrm{TV}_{[0,1]\times\Pi}(A) \int_{-\pi}^{\pi} A\left(\frac{k}{T},\omega\right) e^{i\omega(t-s)} d\omega \right|\\
&= \mathcal{O}(T^{-1}), && (45)
\end{aligned}$$

by Assumption 1, that $\psi_{\ell,n}$ are compactly supported sequences and $\ell \in \mathcal{L}$ is fixed. The order of magnitude in the cross-term, W, from the remainders in (40) and (41) can be established as follows:

$$W = \sum_{s,t} \psi_{\ell,k-s}\psi_{\ell,k-t} \mathcal{O}(T^{-2})\,\mathrm{TV}_{[0,1]\times\Pi}(A)^2 \int_{-\pi}^{\pi} e^{i\omega(t-s)} d\omega, \qquad (46)$$

and using $\int_{-\pi}^{\pi} e^{i\omega t} d\omega = 2\sin(\pi t)/t = 2\pi \mathrm{sinc}(\pi t)$ gives:

$$|W| = 2\pi\,\mathrm{TV}_{[0,1]\times\Pi}(A)^2 \mathcal{O}(T^{-2}) \sum_{s,t} \psi_{\ell,k-s}\psi_{\ell,k-t} |t-k||s-k|\,\mathrm{sinc}\{\pi(t-s)\} = \mathcal{O}(T^{-2}), \qquad (47)$$

again due to the compact support of the discrete wavelet packets. Putting together U, V and W together establishes the result. □

A.2 Proof of Lemma 1

The proof of (i) follows the same pattern of reasoning as that for Lemma 1(i) in Nason (2013), except we need now additionally to verify that $\beta_\ell(z)$ is a function of bounded variation for every wavelet packet $\ell \in \mathcal{L}$. Note that $\beta_\ell(z)$ has a finite and bounded first derivative because, by Assumption 2(iii), A is bounded; moreover, $\hat{\psi}_\ell(\omega)$ can be seen to be bounded by invoking Lemma 2.28 of Nielsen (1999) and using standard Fourier estimates.

For part (ii) we can write $\hat{v}_{i,p}^{(\ell)}$ as

$$\begin{aligned}
\hat{v}_{i,p}^{(\ell)} &= \sum_r \psi_{i,r}^H I_{\ell,2^i p-r} = \sum_r \psi_{i,r}^H d_{\ell,2^i p-r}^2 && (48)\\
&= \sum_r \psi_{i,r}^H \left(\sum_t X_t \psi_{\ell,2^i p-r-t} \sum_s \psi_{\ell,2^i p-r-s} \right) && (49)\\
&= \sum_t \sum_s X_t X_s \sum_r \psi_{i,r}^H \psi_{\ell,2^i p-r-t} \psi_{\ell,2^i p-r-s} && (50)\\
&= \sum_t \sum_s X_t X_s m_{t,s}, && (51)
\end{aligned}$$

where the ψ^H are Haar wavelets and

$$m_{t,s} = \sum_r \psi_{i,r}^H \psi_{\ell,2^i p-r-t} \psi_{\ell,2^i p-r-s}, \tag{52}$$

which depends on ℓ, p and i also. It is immediate that $M = (m_{t,s})_{t,s}$ is a symmetric operator.

Hence, $\hat{v}_{i,p}^{(\ell)}$ satisfies the conditions for Lemma 3.1 of Neumann and von Sachs (1997) which states that if X_t satisfies Assumptions 1–5 then $\hat{v}_{i,p}^{(\ell)} = \eta_T = X^T M X$ with M symmetric, and if we let $\xi_T = Y^T M Y$ where $Y \sim N(0, \mathrm{cov}(X))$ then:

$$\mathrm{cum}_n(\eta_T) = \mathrm{cum}_n(\xi_T) + R_n, \tag{53}$$

where

$$R_n \leq 2^{n-2} C^{2n} \{(2n)!\}^{1+\gamma} \max_{s,t}\{|M_{s,t}|\} \tilde{M} ||M||_\infty^{n-2}, \tag{54}$$

and $\mathrm{cum}_n |\xi_T|$ is bounded, see Neumann and von Sachs (1997). So far, the development for the proof of (ii) parallels symbolically that in Nason (2013). However, the situation is somewhat different here because in this article the ψ_ℓ are wavelet packets and not just wavelets. So, the quantities in (52) and (54) need to be verified for wavelet packets.

The matrix M is banded in that $m_{t,s} = 0$ if $s, t < 2^i(p-1) + 2 - N_\ell$ or $s, t > 2^i p$ where N_ℓ is the number of nonzero entries in the discrete wavelet packet vector ψ_ℓ (the packet vectors are compactly supported). The remaining quantities satisfy $\max_{s,t}\{M_{s,t}\} = \mathcal{O}(T^{-1/2})$, $||M||_2 \leq ||M||_\infty = \mathcal{O}(T^{-1/2})$ and $\tilde{M} = \mathcal{O}(T^{-1/2})$. This can be seen by replacing wavelets by wavelet packets in (a), (b) and (c) of the proof of Lemma 1(ii) from

Nason (2013) and nothing that the wavelet packets are also bounded and compactly supported.

For part (iii) we use Lemma 3.1 from Neumann and von Sachs (1997) together with Assumption 4 to verify:

$$\lambda_{\max}(M)\lambda_{\max}\{\mathrm{cov}(X)\} = \mathcal{O}(T^{-1/2}) \sum_{1\le t\le T} \{\mathrm{cov}(X_s, X_t)\} = \mathcal{O}(T^{-1/2}), \tag{55}$$

where $\lambda_{\max}(M)$ is the maximum eigenvalue of M using the estimates from the norm-quantities of M given in the proof to part (ii) above and Neumann (1994). □

A.3 Proof of Proposition 3

It is the case that:

$$\begin{aligned} T_{\mathrm{C}}\{S_j(z)\} &= J^{-1}\sum_{t=1}^{T}(\mathbf{S}_t - \bar{\mathbf{S}})^T(\mathbf{S}_t - \bar{\mathbf{S}}) && (56)\\ &= J^{-1}\sum_{t=1}^{T}(\boldsymbol{\beta}_t - \bar{\boldsymbol{\beta}})^T(A^{-1})^T A^{-1}(\boldsymbol{\beta}_t - \bar{\boldsymbol{\beta}}), && (57)\end{aligned}$$

where $\mathbf{S}_t$ is the vector $\{S_1(t/T), \ldots, S_J(t/T)\}^T$ and $\bar{\mathbf{S}} = T^{-1}\sum_{t=1}^{T}\mathbf{S}_t$, and similarly for $\boldsymbol{\beta}_t$ and $\bar{\boldsymbol{\beta}}$. Nason et al. (2000) Theorem 1 established that, for all Daubechies' compactly supported wavelets, all the eigenvalues of A are positive. Hence, the norms induced by $T_{\mathrm{C}}\{S_j(z)\}$ and $T_{\mathrm{C}}\{\beta_j(z)\}$ are equivalent. Hence, X_t is stationary iff the wavelet coefficients of $\{\beta_j(z)\}_{j=1}^{J}$ are all zero iff $T_{\mathrm{C}}\{S_j(z)\} = 0$. □

A.4 Proof of Proposition 4

Without loss of generality we will consider a single $\ell \in \mathcal{L}$ the result being a simply extended for the average over a finite collection of $\ell \in \mathcal{L}$ via Lindeberg's theorem, see Billingsley (1995) page 359. For the asymptotic normality we merely need to check Lindeberg's condition in a similar fashion to the proof of Proposition 10.3.2 from Brockwell and Davis (1991). The mean and variance of S_ℓ are finite due to Assumption 4 and the fact that the discrete wavelet packets are compactly supported. Define $s_T^2 = \sum_{i=1}^{T}\sigma_\ell^2 = T\sigma_\ell^2$. For any $\epsilon > 0$

$$\begin{aligned} &\lim_{T\to\infty}(T\sigma^2)^{-1}\sum_{i=1}^{T}\mathbb{E}\left\{(S_{\ell,i}-\mu_\ell)^2\mathbb{I}(|S_{\ell,i}-\mu_\ell| > \epsilon s_n)\right\} && (58)\\ = \;& \sigma^{-2}\lim_{T\to\infty}\mathbb{E}\left\{(S_{\ell,1}-\mu_\ell)^2\mathbb{I}(|S_{\ell,1}-\mu_\ell| > \epsilon T^{1/2}\sigma)\right\} = 0. && (59)\end{aligned}$$

□

References

Andrieu, C. and Duvaut, P. (1996) Measure of cyclostationarity for Gaussian processes based on the likelihood ratio test., in *Proceedings of the 8th IEEE Workshop on Statistical Signal and Array Processing.*, pp. 416–419.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Statist. Soc.* B, **57**, 289–300.

Billingsley, P. (1995) *Probability and Measure*, Wiley, New York.

Borchi, F., Naveau, P., Keckhut, P., and Hauchecome, A. (2006) Detecting variability changes in arctic total ozone column, *J. Atmos. Sol-Terr. Phy.*, **68**, 1383–1395.

Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*, Springer, New York.

Burrus, C. S., Gopinath, R. A., and Guo, H. (1997) *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice Hall, Upper Saddle River.

Cardinali, A. (2008) A generalized multiscale analysis of the predictive content of Eurodollar implied volatilities., *Int. J. Theor. Appl. Fin.*, **12**, 1–18.

Cardinali, A. and Nason, G. P. (2010) Costationarity of locally stationary time series, *J. Time Ser. Econom.*, **2**, Article 1.

Cardinali, A. and Nason, G. P. (2013) Costationarity of locally stationary time series using costat, *J. Statist. Soft.*, **55**, 1–22.

Chatfield, C. (2003) *The Analysis of Time Series: An Introduction*, Chapman and Hall/CRC, London, sixth edition.

Cho, H. and Fryzlewicz, P. (2012) Multiscale and multilevel technique for consistent segmentation of nonstationary time series, *Statistica Sinica*, **22**, 207–229.

Chui, C. K. (1997) *Wavelets: a Mathematical Tool for Signal Analysis*, SIAM, Philadelphia.

Coifman, R. R. and Donoho, D. L. (1995) Translation-invariant de-noising, in A. Antoniadis and G. Oppenheim, eds., *Wavelets and Statistics*, volume 103 of *Lecture Notes in Statistics*, pp. 125–150, Springer-Verlag, New York.

Coifman, R. R. and Wickerhauser, M. V. (1992) Entropy-based algorithms for best-basis selection, *IEEE Trans. Inf. Th.*, **38**, 713–718.

Coifman, R. R. and Wickerhauser, M. V. (1993) Wavelets and adapted waveform analysis toolkit for signal processing and numerical analysis, in I. Daubechies, ed., *Different Perspectives on Wavelets*, pp. 119–154, American Mathematical Society.

Constantine, W. L. B. and Percival, D. B. (2007) fractal: Insightful fractal time series modeling and analysis, *R package version 1.0-3*, uRL: `http://CRAN.R-project.org/package=fractal`.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes, *Ann. Statist.*, **25**, 1–37.

Dahlhaus, R. (2012) Locally stationary processes, in T. Subba Rao, S. Subba Rao, and C. Rao, eds., *Handbook of Statistics*, volume 30, pp. 351–413, Elsevier.

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

Davies, R. and Harte, D. (1987) Tests for the Hurst effect, *Biometrika*, **74**, 95–102.

Dette, H., Preuss, P., and Vetter, M. (2011) A measure of stationarity in locally stationary processes with applications to testing, *J. Am. Statist. Ass.*, **106**, 1113–1124.

Dwivedi, Y. and Subba Rao, S. (2011) A test for second-order stationarity of a time series based on the discrete Fourier transform, *J. Time Ser. Anal.*, **32**, 68–91.

Eckley, I. and Nason, G. (2011) `LS2W`: Implementing the locally stationary 2-D wavelet process approach in R, *J. Statist. Soft.*, **43**, issue 3.

Eckley, I., Nason, G., and Treloar, R. L. (2010) Locally stationary wavelet fields with application to the modelling and analysis of image texture, *J. R. Statist. Soc.* C, **59**, 595–616.

Flandrin, P. (1998) *Time-Frequency/Time-Scale Analysis*, Academic Press, San Diego.

Fryzlewicz, P. (2005) Modelling and forecasting financial log-returns as locally stationary wavelet processes, *J. Appl. Stat.*, **32**, 503–528.

Fryzlewicz, P. and Nason, G. P. (2006) Haar–Fisz estimation of evolutionary wavelet spectra, *J. R. Statist. Soc.* B, **68**, 611–634.

Fryzlewicz, P. and Ombao, H. (2009) Consistent classification of nonstationary time series using stochastic wavelet representations, *J. Am. Statist. Ass.*, **104**, 299–312.

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

Gabbanini, F., Vannucci, M., Bartoli, G., and Moro, A. (2004) Wavelet packet methods for the analysis of variance of time series with applications to crack widths on the Brunelleschi dome, *J. Comp. Graph. Stat.*, **13**, 639–658.

Gardner, W. and Zivanovic, G. (1991) Degrees of cyclostationarity and their application to signal detection and estimation, *Sig. Proc.*, **22**, 287–297.

Gott, A. and Eckley, I. (2013) A note on the effect of wavelet choice on the estimation of the evolutionary wavelet spectrum, *Comm. Stat. - Simulation and Computation*, **42**, 393–406.

Hannan, E. J. (1960) *Time Series Analysis*, Chapman and Hall, London.

Hess-Nielsen, N. and Wickerhauser, M. V. (1996) Wavelets and time-frequency analysis, *Proceedings of the IEEE*, **84**, 523–540.

Hurd, H. and Gerr, N. (1991) Graphical methods for determining the presence of periodic correlation in time series, *IEEE Trans. Inf. Th.*, **12**, 337–350.

Jin, L., Wang, S., and Wang, H. (2015) A new non-parametric stationarity test of time series in the time domain, *J. R. Statist. Soc.* B, **77**, 893–922.

Khoklov, V., Glushkov, A., and Loboda, N. (2006) On the nonlinear interaction between global teleconnection patterns, *Quart. J. Roy. Met. Soc.*, **132**, 447–465.

Mallat, S. G. (1998) *A Wavelet Tour of Signal Processing*, Academic Press, San Diego.

Nason, G. P. (2008) *Wavelet Methods in Statistics with R*, Springer, New York.

Nason, G. P. (2012) Simulation study comparing two tests of second-order stationarity and confidence intervals for localized autocovariance, Technical Report 12:02, Statistics Group, University of Bristol.

Nason, G. P. (2013) A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series., *J. R. Statist. Soc.* B, **75**, 879–904.

Nason, G. P. and Sapatinas, T. (2002) Wavelet packet transfer function modelling of non-stationary time series, *Statistics and Computing*, **12**, 45–56.

Nason, G. P. and Silverman, B. W. (1995) The stationary wavelet transform and some statistical applications, in A. Antoniadis and G. Oppenheim, eds., *Wavelets and Statistics*, volume 103 of *Lecture Notes in Statistics*, pp. 281–229, Springer-Verlag, New York.

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

Nason, G. P., Sapatinas, T., and Sawczenko, A. (2001) Wavelet packet modelling of infant sleep state using heart rate data, *Sankhyā* B, **63**, 199–217.

Neumann, M. (1994) Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series, Statistics Research Report SRR 028-94, Australian National University, Canberra, Australia.

Neumann, M. and von Sachs, R. (1997) Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra, *Ann. Statist.*, **25**, 38–76.

Nielsen, M. (1999) *Size Properties of Wavelet Packets*, Ph.D. thesis, Department of Mathematics, Washington University, Saint Louis, MI, USA.

Page, C. H. (1952) Instantaneous power spectra, *J. Appl. Phys.*, **23**, 103–106.

Paparoditis, E. (2010) Validating stationarity assumptions in time series analysis by rolling local periodograms, *J. Am. Statist. Ass.*, **105**, 839–851.

Park, J., Lee, J., Lee, H., and Lee, J. (2005) Accuracy enhancements of the ARX model by introducing LSW theory in ozone peak prediction, in C. Brebbia, ed., *Air Pollution III*, volume 82 of *WIT Transactions on Ecology and the Environment*, pp. 477–486, WIT Press, Southampton.

Percival, D. B. (1995) On estimation of the wavelet variance, *Biometrika*, **82**, 619–631.

Percival, D. B. and Constantine, W. L. B. (2006) Exact simulation of Gaussian time series from nonparametric spectral estimates with application to bootstrapping, *Statistics and Computing*, **16**, 25–35.

Percival, D. B. and Walden, A. T. (2000) *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Cambridge.

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

Priestley, M. B. (1983) *Spectral Analysis and Time Series*, Academic Press, London.

Priestley, M. B. and Subba Rao, T. (1969) A test for non-stationarity of time-series, *J. R. Statist. Soc.* B, **31**, 140–149.

Sanderson, J., Fryzlewicz, P., and Jones, M. (2010) Measuring dependence between non-stationary time series using the locally stationary wavelet model, *Biometrika*, **97**, 435–446.

Silverman, R. A. (1957) Locally stationary random processes, *IRE Trans. Information Theory*, **IT-3**, 182–187.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. D. (1992) Testing for nonlinearity in time series: the method of surrogate data, *Physica D: Nonlinear Phenomena*, **58**, 77–94.

Triantafyllopoulos, K. and Nason, G. P. (2008) A note on state-space representations of locally stationary wavelet time series, *Stat. Prob. Lett.*, **79**, 50–54.

Van Bellegem, S. and von Sachs, R. (2004) On adaptive estimation for locally stationary wavelet processes and its applications, *Int. J. of Wavelets, Multiresolution and Information Processing*, **2**, 545–565.

Van Bellegem, S. and von Sachs, R. (2008) Locally adaptive estimation of evolutionary wavelet spectra, *Ann. Statist.*, **36**, 1879–1924.

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

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

von Sachs, R. and Neumann, M. H. (2000) A wavelet-based test for stationarity, *J. Time Ser. Anal.*, **21**, 597–613.

von Sachs, R. and Schneider, K. (1996) Wavelet smoothing of evolutionary spectra by non-linear thresholding, *App. Comp. Harm. Anal.*, **3**, 268–282.

von Sachs, R., Nason, G., and Kroisandt, G. (1997) Adaptive estimation of the evolutionary wavelet spectrum, Technical Report 516, Statistics Department, Stanford University, Stanford, CA, USA.

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

Woyte, A., Van Thong, V., Belmans, R., and Nijs, J. (2006) Voltage fluctuations on distribution level introduced by photovoltaic systems, *IEEE Trans. Energ. Conv.*, **21**, 202–209.

Woyte, A., Belmans, R., and Nijs, J. (2007a) Fluctuations in instantaneous clearness index: analysis and statistics, *Sol. Energy*, **81**, 195–206.

Woyte, A., Belmans, R., and Nijs, J. (2007b) Localized spectral analysis of fluctuating power generation from solar energy systems, *Eurasip J. Adv. Sig. Pr.*, 080919.

Xie, Y., Yu, J., and Ranneby, B. (2009) Forecasting using locally stationary wavelet processes, *J. Statist. Comp. Simul.*, **79**, 1067–1082.