A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series.

(corrected version of *Journal of the Royal Statistical Society*, Series B, 2013, **75**, 879–904)

Guy Nason

28st November 2012 (updated 1st July 2016)

Abstract

Many time series are not second-order stationary and it is not appropriate to analyze them using methods designed for stationary series. This article introduces a new test for second-order stationarity that detects different kinds of departures from stationarity than those based on Fourier methods. The new test is also computationally fast, designed to work with Gaussian and a wide range of non-Gaussian time series, and can locate nonstationarities in time and scale. The test is demonstrated on earthquake, explosion, infant electrocardiogram and simulated time series showing varying degrees of stationarity. The second main contribution develops approximate confidence intervals for time-varying autocovariances for locally stationary series as the usual bands computed for stationary series are not appropriate. Our new bands enable practitioners to statistically assess time-varying autocovariances and are exhibited on localized autocovariances of explosion and simulated time series.

Keywords: stationarity test, local autocovariance, confidence intervals, locally stationary

1 Introduction

Many professionals encounter time series that they suspect are not second-order stationary. Probably, the most common technique for investigating this suspicion is via the regular time series plot, closely followed by applying the

regular autocovariance and spectral estimators to different portions of the series to see whether they differ. Although simple, the use of such 'regular' methods, designed for stationary series, is not to be recommended for nonstationary series. *Ad hoc* use on segments or rolling windows raises all sorts of questions beginning with 'how many observations should a portion contain?' and 'from where should the portion be extracted?', to more advanced inferential questions such as 'do these two, or more, ACF estimates differ statistically?'.

Even if the series is piecewise stationary, detecting breaks is not always an easy proposition. See, for example, in biomedical signal processing: segmentation of extracellular microelectrode recordings in Falkenberg et al. (2003), dynamic spectral analysis of event-related EEG data in Florian and Pfurtscheller (2000); in audio signal processing: automatically determining pitch-side/studio footage from Rugby Union broadcasts in Davies and Bland (2010); in speech processing, see Spanias et al. (1993), and several examples in the statistical literature Adak (1998), Ombao et al. (2001), Ombao et al. (2002), Davis et al. (2006), Draghicescu et al. (2009), Rosen et al. (2009), to name but a few.

This article makes two main contributions. We develop a new test of stationarity which is efficient, works with Gaussian and a wide range of non-Gaussian time series, and identifies the location of nonstationarities. If a series is deemed not to be stationary our second contribution provides a method to compute confidence intervals (CIs) for a localized autocovariance (lacv) estimator. This permits one to investigate how the autocovariance of a nonstationary series changes over time in a statistically rigorous way. Note, our method computes a lacv estimator at a given time point and lag, no associated segment or rolling window is involved.

1.1 Tests of Stationarity

That practitioners often resort to *ad hoc* methods is a somewhat surprising state of affairs as there are several easily-found tests of stationarity that exist in the literature. The idea of series not being second-order stationary has been around for a very long time, certainly since Silverman (1957) and Page (1952). An early test (the PSR test) for stationarity was proposed by Priestley and Subba Rao (1969) which performs an ANOVA analysis on the logarithm of a time-varying spectral estimate at a given set of times and frequencies. The ANOVA can be scrutinized for variation of the spectrum across time (and for effects over frequencies, and interaction effects) to determine, as a *hypothesis test*, whether the spectrum is time-varying and hence the series not stationary. This test has recently been made publicly available via CRAN in the `fractal` package by Constantine and Percival (2007). The `fractal` package's `stationarity` function contains an improved version of the test which uses averages of multitaper estimates which control leakage bias.

Since Priestley and Subba Rao (1969) other tests of stationarity have been developed: von Sachs and Neumann (2000), Paparoditis (2010) and Dwivedi and Subba Rao (2011) are a recent few that we briefly compare and contrast. Some comparative information is shown in Table 1. These works provide a useful survey of the field and they are all representative of certain classes of test. von Sachs and Neumann (2000) delineate the difference between smoothing and non-smoothing tests, the former producing estimates of the spectral density and basing tests on those rather than computing 'lumps' of periodogram. Why do we need another test of stationarity? Different tests pick up on a variety of nonstationarities, and will have differing powers for various alternative hypotheses. Priestley and Subba Rao (1969), von Sachs and Neumann (2000) and Paparoditis (2010) work with local Fourier spectra, and the innovative Dwivedi and Subba Rao (2011) attempts to detect correlation between power at differing Fourier ordinates.

Our test is different, based on a wavelet process and, hence, likely to detect departures in a different 'direction'. More specifically, we use the principle introduced by von Sachs and Neumann (2000) that assessed constancy of the time-varying Fourier spectrum by examining its Haar wavelet coefficients across time. Our test examines the constancy of a wavelet spectrum by examining its Haar wavelet coefficients but over a *finite* set of (wavelet) scales, rather than a smoothed set of frequencies. Our test inherits the ability of the von Sachs and Neumann (2000) technique to work with a wide range of non-Gaussian data. Indeed, simulation results given in Section 2.4.3 clearly demonstrate our test's better control of statistical size when compared to the Priestley and Subba Rao (1969) method for heavy-tailed time series.

Test speed can be important in many practical situations and ours is designed to be fast. For example, for long series, such as in bioinformatics, see Vannucci and Lio (2001), Lio (2003), or whenever the test has to be repeated multiple times, such as in the determination of costationarity in Cardinali and Nason (2010). We consider a test to be 'fast' whenever its speed achieves $\mathcal{O}(T \log T)$ or better when the length of the series is T: our new test achieves this rate. Although the papers mentioned above do not explicitly mention the computational complexity of their methods, we estimate the order for von Sachs and Neumann (2000) and Paparoditis (2010) to be $\mathcal{O}(T^{3/2} \log T)$. For non-Gaussian data Dwivedi and Subba Rao (2011) requires computation of the tri-spectrum which, even using fast algorithms requires $\mathcal{O}(T^3)$, although the test overall reduces to $\mathcal{O}(T^2)$ for Gaussian data.

In summary, our new test is fast, designed to work with non-Gaussian data and provides sensitivity to different alternatives compared to previous tests.

Table 1: Characteristics of certain tests of stationarity. Type is smoothing or non-smoothing. Non-normal indicates whether theory exists for the non-Gaussian case. `R` indicates whether the software exists in an R package and 'Locates' indicates whether the test directly indicates the location of the nonstationarities.

Test	Type	Non-normal?	Fast	`R`	Locates
PSR69	N	✗	✓	`fractal`	✓
vSN00	S	✓	✗	✗	✓
Pap10	N?	✗	✗	✗	✓
DSR11		✓	✗	✗	✗
ours	S	✓	✓	`locits`	✓

1.2 Autocovariance Confidence Intervals

If a test rejects the null hypothesis of stationarity then a practitioner can proceed to use various tools such as local spectral analysis and local autocovariance estimation. This article introduces CIs for an estimate of the time-varying autocovariance, by deriving the theoretical asymptotic variance of a wavelet-based estimator. Such CIs are useful for determining whether autocovariances are likely to be zero or not, just as in the case for the popular $\pm 2/\sqrt{T}$ bands often used for stationary series, and coded into the `acf()` function in `R`, R Development Core Team (2009).

1.3 Background Model: Locally Stationary Wavelet

Central to the techniques developed in the next sections are the locally stationary wavelet (LSW) processes introduced by Nason et al. (2000). The time series X_t, $t = 1, \ldots, T$, is a LSW process if it has representation:

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

where $w_{j,k;T}$ are the amplitudes of the process, $\{\psi_{j,k}\}_{j,k}$ are a set of nondecimated discrete wavelets and $\xi_{j,k}$ are a set of uncorrelated random variables with mean zero and variance of one. If the $\xi_{j,k}$ are normally distributed then the process is known as a Gaussian LSW process. To control the stochastic evolution of X_t the amplitudes, $w_{j,k}$ are linked to a collection of functions $\{W_j(z)\}_{j=1}^{\infty}$, where $z \in (0,1)$ is rescaled time, by

$$\sup_k |w_{j,k;T} - W_j(k/T)| \leq C_j/T, \tag{2}$$

for $j = 1, 2, \ldots$, where C_j is a sequence of constants satisfying $\sum_{j=1}^{\infty} C_j < \infty$. The rescaled time device was introduced by Dahlhaus (1997) in a seminal

development of locally stationary Fourier processes. Stochastic evolution of X_t is controlled by constraints on $\{W_j(z)\}$: the precise constraints we employ are determined by Assumption 2 for the tests of stationarity and in the statement of Theorem 1 for CIs of localized autocovariance, below. However, if suffices to say here that if $W_j(z)$ is a constant function of z for all $j \in \mathbb{N}$, then the process X_t is stationary. If $W_j(z)$ changes slowly as a function of z (for some $j \in \mathbb{N}$) then the process is 'near stationary', and the faster the oscillation of $W_j(z)$ means that the process X_t deviates further from stationarity. Nason et al. (2000, Section 2.2) define the evolutionary wavelet spectrum (EWS), $\{S_j(z)\}$, by $S_j(z) = |W_j(z)|^2$ for all $z \in (0, 1)$ and $j \in \mathbb{N}$. As in stationary theory the (EW) spectrum is a key quantity. However, unlike stationary theory the EWS is a time-varying spectrum, $S_j(t/T)$, which quantifies the contribution to variance in the series at scale $j \in \mathbb{N}$ and over time t.

Nason et al. (2000, Def. 6 and 7) estimate $\{S_j(z)\}$ from $X_1, \ldots, X_T$ by first computing the discrete nondecimated wavelet coefficients of the series: $d_{j,k} = \sum_{t=1}^{T} X_t \psi_{j,k-t}$ and then the wavelet periodogram $I_{j,k} = d_{j,k}^2$. Proposition 4 of Nason et al. (2000) shows that $\mathbb{E}(I_{j,k}) = \beta_j(k/T)$ where

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

where the matrix $A_{i,j} = \sum_{\tau} \Psi_i(\tau)\Psi_j(\tau)$, and the autocorrelation wavelets $\Psi_j(\tau) = \sum_k \psi_{j,k}\psi_{j,k-\tau}$ from Definitions 3 and 5 from Nason et al. (2000). An asymptotically unbiased estimator of $\{S_j(z)\}$ can be obtained by premultiplying $\mathbf{I}_k = (I_{1,k}, \ldots, I_{J,k})^T$ by A^{-1} where $z = k/T$ and $J = \log_2 T$ a finite number of scales from a finite series T. The quantity $\beta_j(z)$ was introduced by Fryzlewicz and Nason (2006) and is often easier to work with theoretically than the spectrum $S_j(z)$.

1.4 Article Structure

Section 2 describes our new test of stationarity and introduces: the theoretical basis in Section 2.1; the empirical Haar wavelet coefficients of the wavelet periodogram in Section 2.2 and demonstrates their asymptotic normality; the test statistics in Section 2.3 and the role of multiple hypothesis testing. Section 2.4 exhibits the new test on real data and compares it to the PSR on simulated data.

Section 3 introduces our CIs for a localized autocovariance estimator: Section 3.1 rehearses the method of calculation of the localized autocovariance from Nason et al. (2000); Section 3.2 shows how *ad hoc* methods can be misleading; Section 3.3 explains our method for computing approximate $100(1 - \alpha)\%$ CIs; Section 3.4 exhibits our CIs on real and simulated data.

Section 4 concludes and suggests topics for further examination. Further simulations, comparisons and investigations into both the new test of station-

arity and the CIs can be found in Nason (2012) along with instructions on how to reproduce the figures in this paper. An `R` package, `locits`, containing software to implement the methods can be obtained from the author.

2 A Wavelet-based Test of Stationarity

2.1 Theoretical Basis for the Test

Our test of stationarity is based on examination of the wavelet periodogram $I_{\ell,\lfloor zT \rfloor;T}$, or $I_{\ell,t}$ for short, where $z = t/T$. As first shown in Nason et al. (2000, Proposition 4) $I_{\ell,k}$ is an asymptotically unbiased estimator of the quantity $\beta_\ell(z)$ defined above in (3). In principle, $\ell \in \mathbb{N}$, but, in practice, since T is finite, we have access to a finite set of scales $\ell = 1, \ldots, J = \log_2(T)$.

The series X_t is stationary if and only for all $\ell \in \mathbb{N}$ the quantity $\beta_\ell(z)$ is constant function of $z \in (0, 1)$. As proposed by von Sachs and Neumann (2000) for Fourier spectra we evaluate the constancy of of $\beta_\ell(z)$ by examining its Haar wavelet coefficients:

$$v_{i,p}^{(\ell)} = \int_0^1 \beta_\ell(z) \psi_{i,p}^H(z)\, dz, \tag{4}$$

for $i = 1, 2, \ldots, J$, $p = 1, \ldots, 2^i - 1$, and $\{\psi_{i,p}^H(t)\}_{i,p}$ are the usual Haar wavelets. von Sachs and Neumann (2000) note that the Haar wavelets are perfect for stationarity testing for detecting differences in $\beta_\ell(z)$. This is in contrast to estimation of $\beta_\ell(z)$ itself, where smoother wavelets can offer advantages in terms of better rates of convergence.

If any $v_{i,p}^{(\ell)}$ are non-zero then $\exists \ell$ such that $\beta_\ell(z)$ is not constant and hence the associated LSW process, X_t, is not stationary. In reality we do not know $\beta_\ell(z)$ and we replace it by its asymptotically unbiased estimator $I_{\ell,k}$ to obtain estimates $\hat{v}_{i,p}^{(\ell)}$ of $v_{i,p}^{(\ell)}$ and perform multiple hypothesis testing on the whole collection of hypotheses: $H_0 : v_{i,p}^{(\ell)} = 0$ for all ℓ, i, p against the alternative that there exists ℓ^*, i^*, p^* such that $v_{i^*,p^*}^{(\ell^*)} \neq 0$. The hypothesis tests rely on the establishment of asymptotic normality of the $\hat{v}_{i,p}^{(\ell)}$ under mild assumptions.

In what follows, the one-dimensional wavelet coefficient analysis parallels that for the two-dimensional situation in von Sachs and Neumann (2000). The assumptions we make are:

Assumption 1. *For control of bias asymptotically we consider only wavelet coefficients $v_{i,p}$ at coarse enough scales, i.e. $2^i = \mathcal{O}(T)$. As in von Sachs and Neumann (2000) there is no additional segmentation bias as we average the wavelet periodogram over pre-defined dyadic intervals.*

Assumption 2. *(a)* $\sup_z |W_\ell(z)| < \infty$*, for all* $\ell \in \mathbb{N}$.

(b) $\inf_z |W_\ell(z)| \geq \kappa$ *for some* $\kappa > 0$*, for all* $\ell \in \mathbb{N}$.

(c) $W_\ell(z)$ *has uniformly bounded total variation with respect to* z*, again for all* $\ell \in \mathbb{N}$*. In other words* $\sup_z TV\{W_\ell(z)\} < \infty$ *for* $\ell \in \mathbb{N}$.

The next assumption is made on the cumulants of the process X_t:

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

Remark 1. *Assumption 2 is satisfied if* $W_\ell(z)$ *is differentiable (with uniformly bounded partial derivatives), but as with Van Bellegem and von Sachs (2008) the bounded variation assumption permits jumps in the spectrum, which can be important in practice.*

Remark 2. *von Sachs and Neumann (2000) note that Neumann (1994) showed that Assumption 3 is satisfied if* $\{X_t\}$ *is* α*-mixing with coefficients* $\alpha(s) \leq K\exp(-b|s|)$ *and*

$$\mathbb{E}|X_t|^k \leq C^k(k!)^\rho, \tag{5}$$

for all k*. Condition* (5) *is fulfilled for many distributions, for example, exponential, gamma and inverse Gaussian distribution for appropriate choice of* ρ *(and also for Gaussian, see Neumann (1994, Remark 3.1(ii))).*

2.2 The Test Statistic(s)

In practice, we use the empirical wavelet periodogram values $I_{\ell,k} = I_\ell(k/T)$ for $k = 1, \ldots, T$ and $z = k/T$ is rescaled time. The Haar wavelet coefficient estimates, for scale ℓ of the wavelet periodogram, are given by

$$\hat{v}_{i,p}^{(\ell)} = 2^{-i/2}\left(\sum_{r=0}^{2^{i-1}-1} I_{\ell,2^ip-r} - \sum_{q=2^{i-1}}^{2^i-1} I_{\ell,2^ip-q}\right), \tag{6}$$

for the same range of i, p as in (4).

We now derive the asymptotic distribution of these empirical coefficients under the general assumptions listed in Section 2.1.

Lemma 1. *Suppose that Assumptions 1–3 hold. Then:*

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

(ii) $\operatorname{var}(\hat{v}_{i,p}^{(\ell)}) = 2T^{-1}\int_0^1 \beta_\ell^2(z)\{\psi_{i,p}^H(z)\}^2\,dz + \mathcal{O}(T^{-1})$.

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

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

The proof of this Lemma is in the appendix and is heavily modified version of Lemma 3.2 from Neumann and von Sachs (1997). We can demonstrate the asymptotic normality of the $\hat{v}_{i,p}$ using the following Proposition:

Proposition 1. *Suppose that Assumptions 1–3 hold. Let* $\Delta_T = K(\log T)^{1/2}$ *for any fixed* $K < \infty$. *Then:*

$$\mathbb{P}\{\pm(\hat{v}_{i,p}^{(\ell)} - v_{i,p}^{(\ell)})/\sigma_{i,p}^{(\ell)} \geq x\} = \{1 - \Phi(x)\}\{1 + o(1)\}, \tag{7}$$

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

Proposition 1 parallels Proposition 3.1 in von Sachs and Neumann (2000) for the local Fourier case, and both rely on the asymptotic normality proof of Proposition 3.1 found in Neumann and von Sachs (1997) and the properties established in Lemma 1.

2.3 Testing

Under H_0 the quantity $\beta_\ell(z)$ is a constant function of z and hence $v_{i,p}^{(\ell)} = 0$ for all ℓ, i, p from (4) as $\int \psi(z)\, dz = 0$ is a defining wavelet property. The test statistic we form is:

$$T_{i,p}^{(\ell)} = \hat{v}_{i,p}^{(\ell)} \hat{\sigma}_{i,p}^{(\ell)-1}, \tag{8}$$

To estimate $\sigma_{i,p}^{(\ell)2}$ note that for stationary series we have $\beta_\ell = \int_{-\pi}^{\pi} f(\omega)|\hat{\psi}_\ell(\omega)|^2\, d\omega$ where the classical stationary spectrum $f(\omega)$ could be estimated by the regular periodogram. In this case, von Sachs and Neumann (2000, Theorem 3.1) allows us to obtain convergence rates that guarantee that $P_{H_0}\{|\hat{v}_{i,p}^{(\ell)}| > \hat{\sigma}_{i,p}^{(\ell)}\Phi^{-1}(1 - \alpha/2)\} \leq \alpha + o(1)$, for parameters ℓ, i, p as above and size α. Another estimate of $\sigma_{i,p}^{(\ell)2}$ can be obtained by replacing $\beta_\ell^2(z)$ in Lemma 1 (ii) by

$$\hat{\sigma}_{i,p}^{(\ell)2} = 2T^{-1} I_{\ell,<1,T>}^2 \int_0^1 \{\psi_{i,p}^H(z)\}^2\, dz = 2T^{-1} I_{\ell,<1,T>}^2, \tag{9}$$

where $I_{\ell,<1,T>}^2 = T^{-1}\sum_t I_{\ell,t}^2$ and because Haar wavelets have unit norm for all i, p. Each test statistic is compared to a critical value derived from the normal distribution in the normal way.

For stationarity testing we wish to test many hypotheses $H_0 : v_{i,p}^{(\ell)}$ for several values of ℓ, i, p. Hence, we are in the world of multiple hypothesis testing. We propose to use Bonferroni correction and, for a less conservative procedure, we propose using the false discovery rate (FDR) procedure introduced by Benjamini and Hochberg (1995). Our simulations, below, show that these methods work well. However, the multiple hypothesis testing methods

we use do not take into account dependence of the $\hat{v}_{i,p}^{(\ell)}$, this is an interesting avenue for further work.

2.4 Data: Real Examples and Simulations

2.4.1 The Earthquake and Explosion Data

Figures 1 and 2 show our test being applied to the Earthquake P and the Explosion P data from Shumway and Stoffer (2006) respectively. Each plot also uses a horizontal double-headed arrow to indicate the location of a nonstationarity, significant Haar coefficient, picked up by the multiple hypothesis testing, and are produced automatically in the associated `R` package, `locits`.

The vertical position of the horizontal arrow encodes the scale ℓ of the periodogram that was tested, and i the scale of the Haar wavelet transform of that periodogram. The right-hand axis indicates ℓ and the wavelet scales, i, are arranged in ascending order of scale within that. The span of the arrow corresponds to the support of the underlying Haar wavelet and hence the arrow indicates the time period over which the nonstationarity has been detected. The plot is only meant to give an overall impression, identification of i is hard from the plot. However, the software provides detailed information on both Bonferroni and FDR nonstationarities which can either can be listed or plotted.

The Explosion P wave exhibits a higher degree of nonstationarity than the Earthquake P data. Indeed, the Earthquake P data only has one nonstationarity indicator (significant Haar coefficient), so it is quite 'close' to stationarity.

2.4.2 Infant Electrocardiogram Data

The Infant Electrocardiogram data was originally analyzed in Nason et al. (2000, Section 4.2) and their first differences analyzed in Nason (2008, Section 5.3.7). Figure 3 shows a plot of the first differences of the `BabyECG` data as computed in Nason (2008). The subtitle of Figure 3 notes that four Haar wavelet coefficients were deemed significant by the new test of stationarity according to the FDR criterion. The location of the four significant coefficients is shown by the horizontal double-headed arrows. A nonstationarity was discovered in $I_{\ell,k}$ for $\ell = 6, 7$ and 8 for k the Haar wavelet centred on $k = 512$ of length 1024. A further nonstationarity was discovered at about $k = 1400$ in $I_{7,k}$. Further interrogation of the output of the test shows that this last significant Haar wavelet occurred on level $i = 5$ which means that the extent of the nonstationarity exists over $2048/2^5 = 64$ time points.

2.4.3 Size and Power Simulations and Comparisons

We report results from a set of simulations, selected from Nason (2012), comparing the size and power of our new tests of stationarity and also the PSR

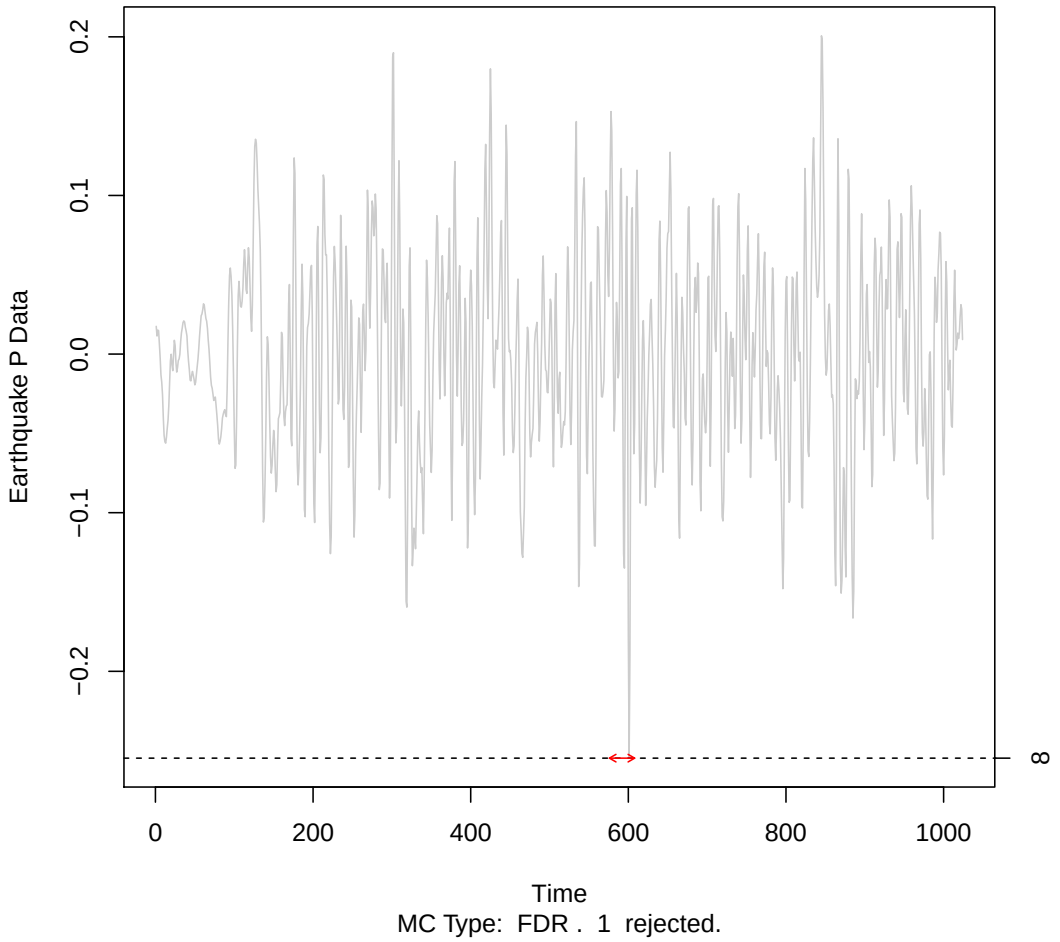


Figure 1: Plot of Earthquake P Data with (single) nonstationarity indicator shown. The horizontal double-headed arrow shows the location of a detected nonstationarity, the right-hand side axis label shows the scale of the wavelet periodogram, $\ell = 8$, where it was detected.

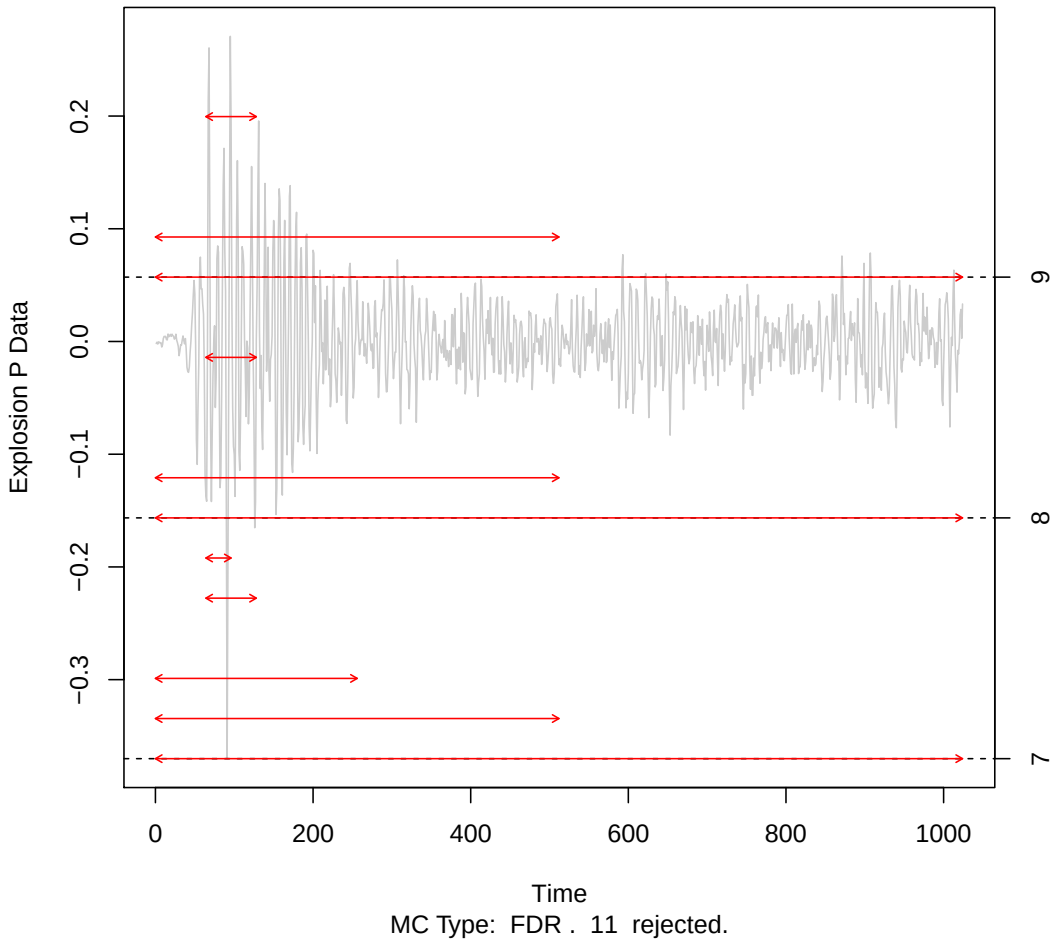


Figure 2: Plot of Explosion P Data with nonstationarity indicators shown. The horizontal double-headed arrow shows the location of a detected nonstationarity, the right-hand side axis label shows the scales of the wavelet periodogram, $\ell = 7, 8, 9$ where they were detected. The scale of the analysing Haar wavelet, i, is indicated by vertical positioning within those scales, but not explicitly shown on the axis.

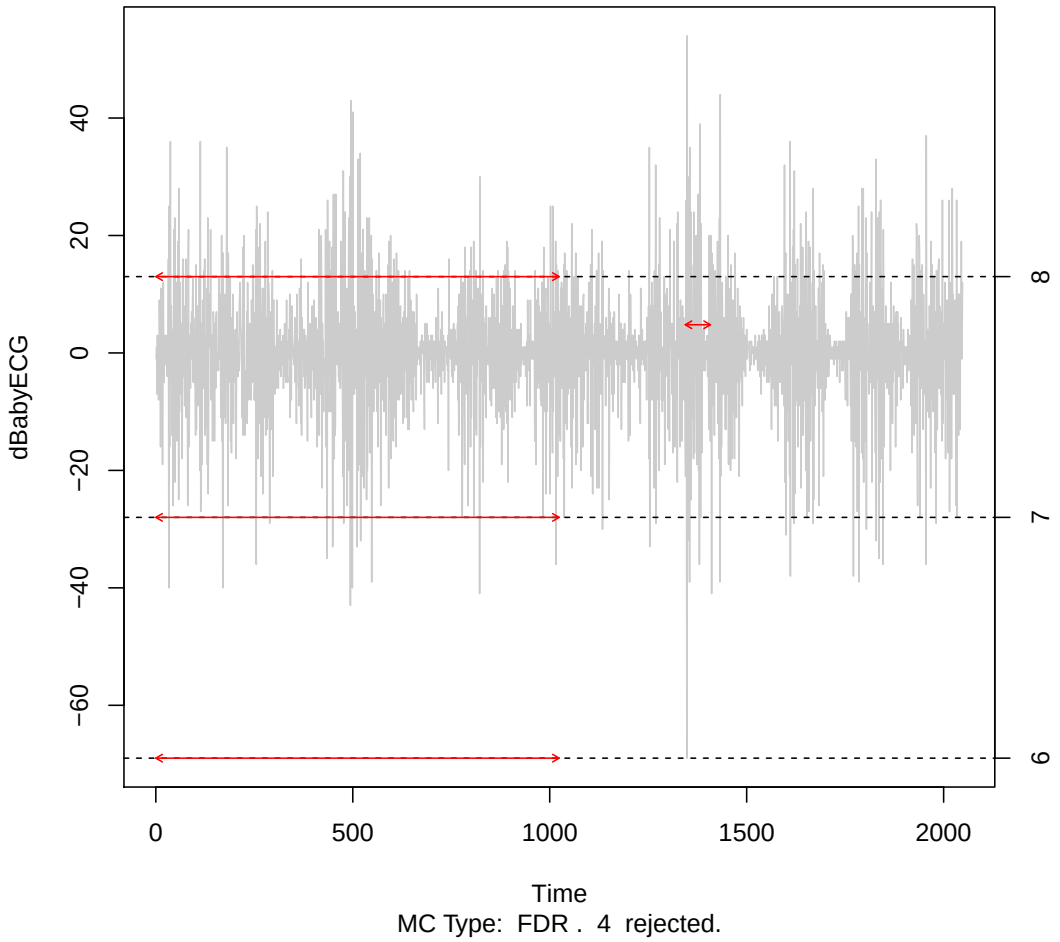


Figure 3: Plot of differenced BabyECG data with nonstationarity indicators shown. The horizontal double-headed arrow shows the location of a detected nonstationarity, the right-hand side axis label indicates both which periodogram, ℓ, and Haar wavelet scale i was involved.

Table 2: Simulated size estimates (%) for stationary Gaussian models $T = 512$.

Model	PSR	HWTOS (Bon)	HWTOS (FDR)
S1	5.6	4.3	4.3
S2	12.4	4.0	4.7
S3	6.2	20.3	20.5
S4	6.0	3.4	3.8
S5	6.5	0.7	0.7
S6	7.5	0.1	0.1
S7	23.9	7.3	7.4

test as implemented in the `fractal` package by Constantine and Percival (2007). In all the simulations below the nominal size of the tests is 5%, empirical values are obtained using $N = 1000$ simulations and $T = 512$. Here, the wavelet-based tests are denoted by HWTOS with "Bon" or "FDR" indicating the method of multiple comparison control.

Size Comparisons. First, we simulate data from a number of stationary models using Gaussian innovations and assess how often our tests of stationarity reject the null hypothesis. The models are:

S1 iid standard normal;

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 2. People should not switch to more complex nonstationary methods unless absolutely necessary so most practitioners would prefer a conservative test. Table 2 shows that, apart from S3 and S7, the HWTOS test's empirical size is less than the nominal size and all the PSR test's sizes are greater than 5% (but only just). The PSR erroneously rejects many realisations in S2 and HWTOS does so with S3.

Heuristics: It is difficult to pin a precise explanation on why PSR does not do well for S2 and S7. The decay of the autocorrelation for S2 is slow leading to a process that does not mix quickly. Hence, the process gets 'stuck' for relatively long periods. PSR divides the series into a fairly small number of blocks and the behaviour of the series appears to be different on those blocks

Table 3: Simulated size estimates (%) for stationary exponential-tailed models with nominal size of 5%

Model	PSR	HWTOS (Bon)	HWTOS (FDR)
SHD1	43.8	7.3	7.9
SHD2	48.9	5.8	7.0
SHD3	40.6	20.5	20.8
SHD4	44.5	7.1	7.8
SHD5	46.8	15	19
SHD6	45.1	11	12
SHD7	57.6	10.6	11.4

due to the slow mixing and hence the excessive empirical size values reported in Table 2. A different mechanism appears to be at work in S7 in this case the variance appears to be quite different in different PSR blocks (realizations look like variance-modulated sinusoids) and this behaviour is also affecting the wavelet tests with empirical sizes of 7.3, 7.4%. The behaviour of S7 is similar to, but not the same as, that in S3 and this is investigated in great detail in Nason (2012).

Table 3 shows empirical size values obtained using the models S1–S7, but with double exponential heavier-tailed innovations, labelled as SHD1–SHD7, which HWTOS should cope with due to Assumption 3. Table 3 shows that PSR does not perform well for these heavy tails. The HWTOS tests perform better and only slightly exceed the nominal size for SHD1, SHD2, SHD4 and is barely tolerable on SHD5, SHD6 and SHD7. Like for the Gaussian innovations above (S3) the HWTOS's empirical size for SHD3 is about 20%, but that for PSR is double, and about eight times what they were for Gaussian innovations.

Nason (2012) documents a further simulation with even heavier tails (Student's t) where the PSR empirical size is uniformly greater than 60% whereas that for the HWTOS ranges from 6.6% to 18.0% for the non-S3 based tests. Although the HWTOS is not theoretically designed for such extreme distributions it performs tolerably well, unlike PSR.

Power Comparisons. To explore statistical power we create processes that are nonstationary and then count the number of times each test reckons a realisation is not stationary over multiple simulations. 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.

Table 4: Simulated power estimates (%) for models P1–P4 with nominal size of 5%

Model	PSR	HWTOS (Bon)	HWTOS (FDR)
P1	37.2	99.7	99.9
P2	100	17.3	19.2
P3	44.3	1.3	1.3
P4	100	94.8	97.8

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.

The spectra and single realisations for these processes are displayed in Nason (2012). The simulation results in Table 4 for power paint an interesting picture. Sometimes the HWTOS tests are good and the PSR is not (P1), sometimes PSR is good and HWTOS is not (P2), sometimes both are not that good (P3) and sometimes both are very good (P4).

Sample size is an important determining factor in power: increasing sample size should increase power of detection of alternatives. For example, with P2 which has a sample size of $T = 512$ the HWTOS tests have fairly low powers of 17.3%/19.2% respectively. For $T = 1024$ the tests are more powerful having powers of 70.7%/75.2% and for $T = 2048$ the powers are 100%.

3 LACV Confidence Bands for Gaussian LSW processes

3.1 Localized Autocovariance

For locally stationary series the autocovariance is a time-varying quantity, $c(t, \tau)$ and Nason et al. (2000, Definition 4) showed that, for LSW processes, it can be obtained via the evolutionary wavelet spectrum, $\{S_j(z)\}$ via the formula

$$c(z, \tau) = \sum_{j=1}^{\infty} S_j(z)\Psi_j(\tau), \tag{10}$$

where $z = t/T$, and $\{\Psi_j(\tau)\}$ are the autocorrelation wavelets. More precisely, Nason et al. (2000, Proposition 1) showed that the 'usual' time-varying

autocovariance for LSW processes: $c_T(z,\tau) = \mathrm{cov}(X_{[zT];T}, X_{[zT]+\tau;T})$ satisfies $|c_T(z,\tau) - c(z,\tau)| = O(T^{-1})$ uniformly in $\tau \in \mathbb{Z}$ as $T \to \infty$. Indeed, proposition 3 of Nason et al. (2000) for all stationary series, with absolutely summable autocovariance, shows $c(z,\tau) = \gamma(\tau)$, the classical autocovariance. A key goal of this section will be to develop rigorously derived CIs for $c(z,\tau)$ based on wavelet models. LSW theory parallels that of locally stationary Fourier models and, in principle, what we introduce here, below, could also be developed for those. Next, an example which shows the perils of computing autocovariances and autocorrelations on informally obtained segments.

3.2 Earthquake Example

Figure 1 shows the Earthquake P wave data from Shumway and Stoffer (2006). The mean of this series is very close to zero and it suitable for direct analysis using locally stationary time series. Figure 4 plots three different estimators of the series autocorrelation around time point $t = 150$: one uses a localized autocovariance estimator computed via wavelets and two others using the regular ACF on two different segment lengths centred on $t = 150$.

On the one hand, it is reassuring that the autocorrelation values, especially for the lower lags, look similar. On the other, the two regular-computed ACFs over segments of 100 and 300 data points can give quite different answers. Although the plotted CIs are computed under white noise assumptions, people do tend to use them as a guide to significance more generally, and hence might be attempted to see the coefficient at lag 9 to be significant when computed using a segment length of 100, but not when computed with a segment length of 300. Further, there is nothing special about using segment lengths of 100 or 300, and other choices will almost certainly lead to a variety of inferences.

The two different segment lengths are proxies for different amounts of 'smoothing' in the regular autocovariance estimator. The localized estimator is also subject to smoothing, but with parameters chosen in a manner guided by rigorous theory such as Nason et al. (2000), Van Bellegem and von Sachs (2004), Fryzlewicz and Nason (2006) or Van Bellegem and von Sachs (2008). Further, the localized estimator is designed for locally stationary series with many kinds of smoothness: piecewise, degrees and modes of continuity.

For a stationary series, the usual 'CI' for the regular autocorrelation are $\pm 2/\sqrt{T}$, see Chatfield (2003, Section 2.7.2). These CIs are widely used and are routinely plotted by the `acf` function in `R`. Confidence intervals for the regular autocorrelation, assuming stationarity, are shown in Figure 4. However, for locally stationary series, a *bona fide* time-varying estimate is to be preferred and the usual stationary CIs are not appropriate. We introduce a new set of CIs for the localized quantity next.

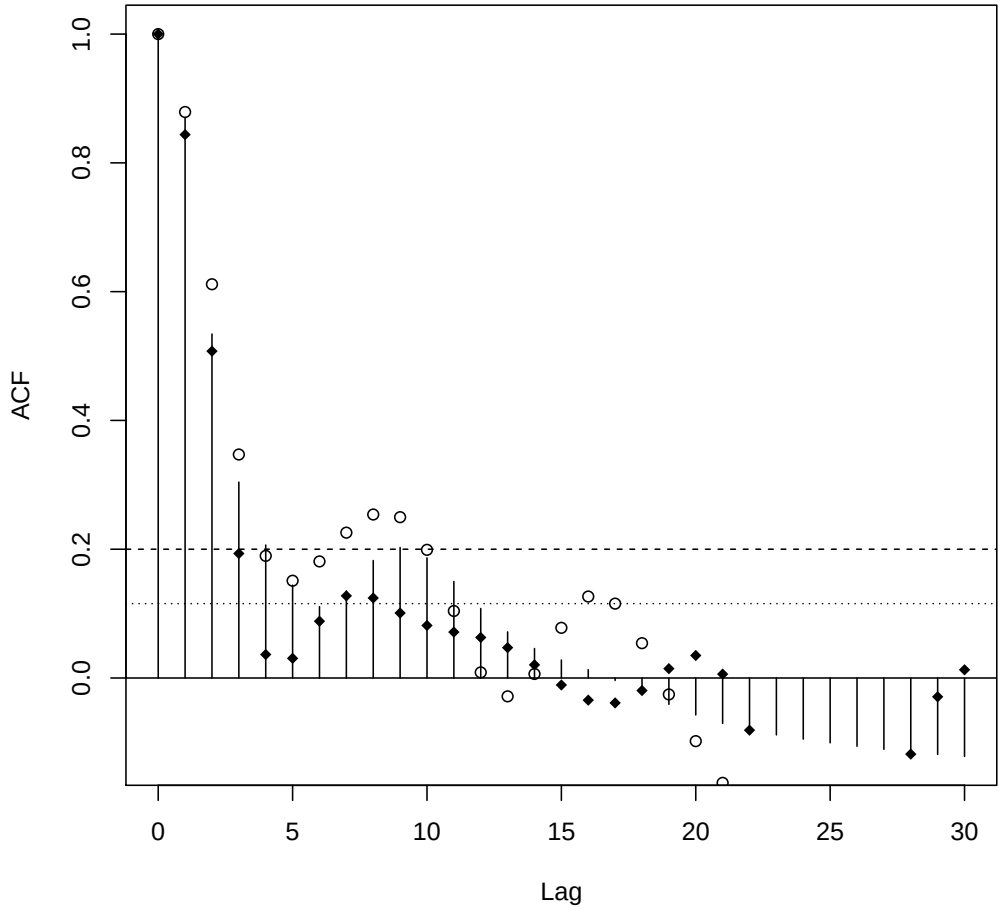


Figure 4: Different Autocorrelation Estimates of Earthquake P Data around $t = 150$. *Vertical line segments:* (default) localized autocorrelation estimate $\hat{c}(150, \tau)$; *Circles:* regular autocorrelation estimator computed on data from $t = 100$ to $t = 200$, dashed line indicates the usual (upper) CI for 100 data points; *Black diamonds:* regular autocorrelation estimator computed on data from $t = 1$ to $t = 300$, dotted line indicates band for 300 data points.

3.3 Confidence Intervals for Localized Autocovariance

We first rewrite the localized autocovariance (10) by substituting the 'biased' spectral estimate $\{\beta_\ell(z)\}$ for $\{S_j(z)\}$:

$$c(z,\tau) = \sum_{j=1}^{\infty}\sum_{\ell=1}^{\infty} A^{-1}_{j,\ell}\beta_\ell(z)\Psi_j(\tau), \tag{11}$$

$$= \sum_{\ell=1}^{\infty}\kappa_\ell(\tau)\beta_\ell(z), \tag{12}$$

where $\kappa_\ell(\tau) = \sum_{j=1}^{\infty} A^{-1}_{j,\ell}\Psi_j(\tau)$. We estimate c by plugging in a smoothed periodogram $\tilde{I}_{\ell,[zT]}$ as an estimate of $\beta_\ell(z)$ in (12) to obtain $\hat{c}$, where

$$\tilde{I}_{\ell,n} = (2s+1)^{-1}\sum_{u=n-s}^{n+s} I_{\ell,u} \tag{13}$$

is a simple running mean smooth of the wavelet periodogram, $I_{\ell,u}$, defined in Section 1.3, with bandwidth s.

Variance-covariance formulae. A useful quantity for constructing the required CI is the variance of the estimator $\hat{c}(z,\tau)$ which can be obtained by:

$$\operatorname{var}\hat{c}(z,\tau) = \sum_{\ell=1}^{\infty}\sum_{j=1}^{\infty}\kappa_\ell(\tau)\kappa_j(\tau)\operatorname{cov}(\tilde{I}_{\ell,[zT]},\tilde{I}_{j,[zT]}), \tag{14}$$

with the covariance of the smoothed periodogram being:

$$\operatorname{cov}(\tilde{I}_{\ell,n},\tilde{I}_{j,n}) = (2s+1)^{-2}\sum_{u=n-s}^{n+s}\sum_{v=n-s}^{n+s}\operatorname{cov}(I_{\ell,u},I_{j,v}). \tag{15}$$

The following result derives the covariance of the wavelet periodogram for Gaussian LSW processes. The assumption of Gaussianity is not limiting and frequently made in both general time series analysis and specifically in locally stationary theory such as Nason et al. (2000), Fryzlewicz et al. (2003), Van Bellegem and von Sachs (2008), Sanderson et al. (2010) and Paparoditis (2010). In particular, Fryzlewicz (2005) provides strong evidence of how Gaussian LSW processes can model data which had traditionally been seen by stationary methods of analysis as heavy-tailed and/or classical models such as GARCH. Fryzlewicz (2005) reviews a growing body of literature providing compelling evidence that models admitting structural change (such as Gaussian LSW) may be more appropriate than the classical stationary models.

Theorem 1. *Let X_t be a Gaussian LSW process with spectrum $\{S_j(z)\}_{j=1}^{\infty}$. Suppose that $S_j(z)$ is Lipschitz continuous for $z \in (0,1)$ for integers $j > 1$*

with Lipschitz constants L_j uniformly bounded in j and $\sum_{j=1}^{\infty} 2^{-j} L_j < \infty$. For integers ℓ, j, m, n with $\ell, j > 0$, asymptotically, the autocovariance of the wavelet periodogram is given by:

$$\operatorname{cov}\left(I_{\ell,m}, I_{j,n}\right) \simeq \begin{cases} 2\beta_\ell^2(m/T) & \text{for } \ell = j, m = n, \\ 2\left\{\sum_k S_k\left(\frac{m+n}{2T}\right) \sum_\tau \Psi_k(\tau)\Psi_{\ell j}(n-m-\tau)\right\}^2 & \text{otherwise,} \end{cases} \tag{16}$$

where $\Psi_{\ell j}(\tau) = \sum_k \psi_{\ell,k}\psi_{j,k+\tau}$ and $x \simeq y$ denotes $x = y + \mathcal{O}(T^{-1})$.

Erratum: statement of this theorem is a correction from previous and published versions and the changes are highlighted in red. The accompanying software, `locits` was updated with the correction from version 1.7

Proof: in appendix. Note, $\Psi_{\ell j}$ was defined in Fryzlewicz and Nason (2006, Section 2.2). Formula (16) coincides with the formula for $\operatorname{var}(I_{j,m})$ in Proposition 4 of Nason et al. (2000), i.e. when $\ell = j$ and $m = n$. Also Theorem 1 operates under the same smoothness regime as in that paper.

Approximate confidence interval. An approximate $100(1-\alpha)\%$ pointwise CI for $c(z,\tau)$ can be obtained in the usual way by using the estimate $\hat{c}(z,\tau)$, the variance-covariance formulae in (14), (15) and (16) and noting that the running mean smoother in (13) is asymptotically Gaussian under mild regularity conditions, see Schuster (1972), for example. The CI is

$$[\hat{c}(z,\tau) - z_{\alpha/2} \operatorname{var}\{\hat{c}(z,\tau)\}^{1/2}, \hat{c}(z,\tau) + z_{\alpha/2} \operatorname{var}\{\hat{c}(z,\tau)\}^{1/2}], \tag{17}$$

where $z_q = -\Phi(q)$ is the usual percentage point of the standard normal distribution. E.g. for a 95% CI, $\alpha = 0.05$ and $z_{\alpha/2} \approx 1.96$. Empirical demonstration of approximate normality of $c(z,\tau)$ is demonstrated in Section 3.2 of the associated technical report, Nason (2012).

The intention here is to introduce rigorously defined asymptotic CIs based on well-established smoothing techniques, a precedent for such methods in locally stationary time series estimation, should one be required, is the use of kernel smoothing in Fryzlewicz et al. (2003). An attractive option for future investigation, but beyond the scope of the current work, would be to develop CIs associated with more advanced smoothing techniques, such as the wavelet shrinkage methods used in Nason et al. (2000, Section 3.3). In this case asymptotic Gaussianity might well be established using methods described by Brillinger (1996), Vidakovic (1999, Section 6.5.2), and/or related work building on the asymptotic normality of wavelet coefficients, see e.g., Proposition 1 in the next section.

3.4 Application to Real and Simulated Data

3.4.1 Explosion Q wave

Figure 5 shows two localized autocovariance plots of the Explosion Q wave series (which looks similar to the P wave in Figure 2) at times $t = 50$ and $t = 900$. The vertical scale of plot *b.* is twenty times less than that of plot *a.*: the variance of the series near the end is much smaller than at the beginning. Apart from the scale change, most autocovariance values look reasonably similar. However, $\hat{c}(50/1024, 2)$ looks nonzero and its 95% CI does not cover zero, whereas $\hat{c}(900/1024, 2)$ seems very close to zero and its 95% CI covers zero. Hence, the lag two autocovariance values seem very different in these two parts of the original series. Hence, there is strong evidence to show that the Explosion Q wave's autocovariance function is changing over time both in terms of their overall variances but also the lag two acf.

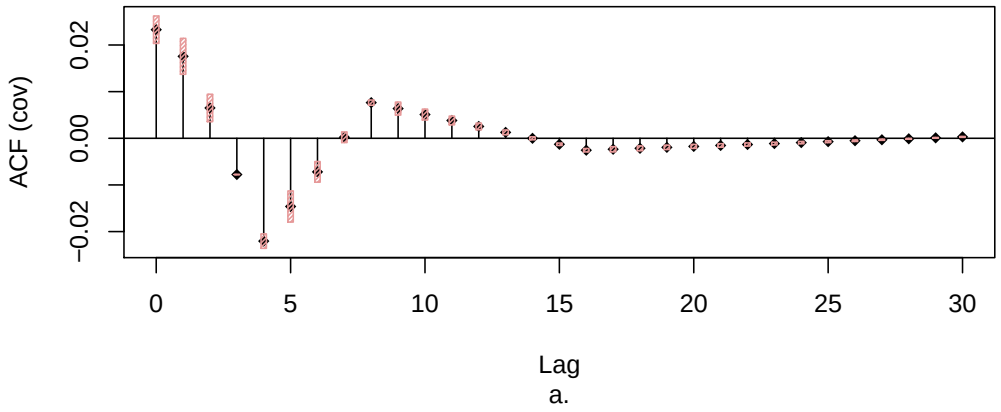
3.4.2 Simulated Example

Figure 6 shows the estimate localized autocovariance function $\hat{c}(z, \tau)$ for $t = 100, 200, 300$ and 400 for the time-varying AR(1) model P1 given in Section 2.4.3. In each of the subplots in Figure 6 the horizontal dashed line shows the true value of the AR(1) parameter at that time point. The changing nature of the AR(1) parameter can be clearly seen in the plots. Further examples can be found in Nason (2012).

4 Conclusion and Future Work

This article introduced a new test of stationarity by assessing the statistical significance of Haar wavelet coefficients of the wavelet periodogram of a time series. The new test is (i) complementary to existing tests searching for departures from stationarity in new directions; (ii) efficient and rigorously established following principles set by von Sachs and Neumann (2000); (iii) capable of identifying the location and time-scale over which nonstationarity exists. The `locits` package has been created to compute the new test, provides functionality to describes the nonstationarities, and superimposes them on a time series plot. Simulations show that the new test works well, complements existing tests and appears to work better on heavy-tailed data compared to the well-established test due to Priestley and Subba Rao (1969).

Theory was developed to compute approximate CIs for the localized autocovariance for Gaussian locally stationary wavelet processes. Again, these ideas have been implemented within `locits` enabling users to view localized autocovariances with approximate CIs. Again, this methodology seems to work well on both real and simulated data.



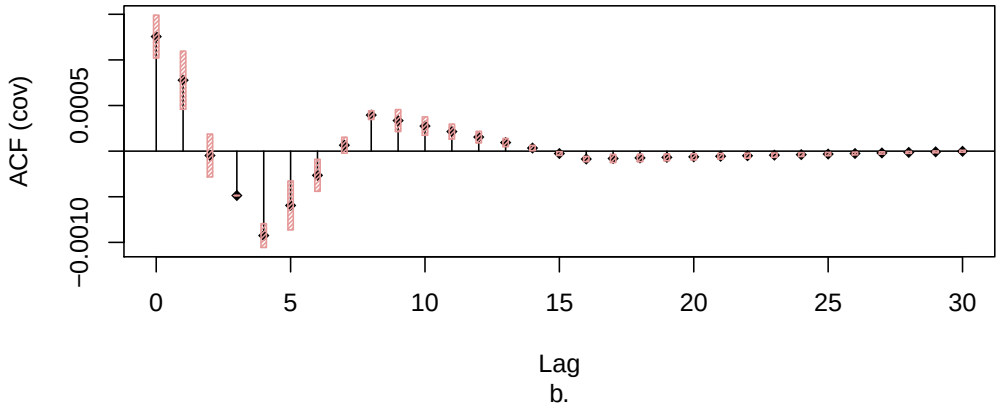


Figure 5: Localized autocovariance estimate $\hat{c}(z, \tau)$ for Explosion Q wave with 95% CI for *a.* $z = 50/1024$ and *b.* $z = 900/1024$ (corresponding to times $t = 50, 900$ respectively). Note: vertical scale of plot *b.* is twenty times less than that in plot *a.* Diamonds correspond to the autocovariance values and the hatched rectangles to the CIs.

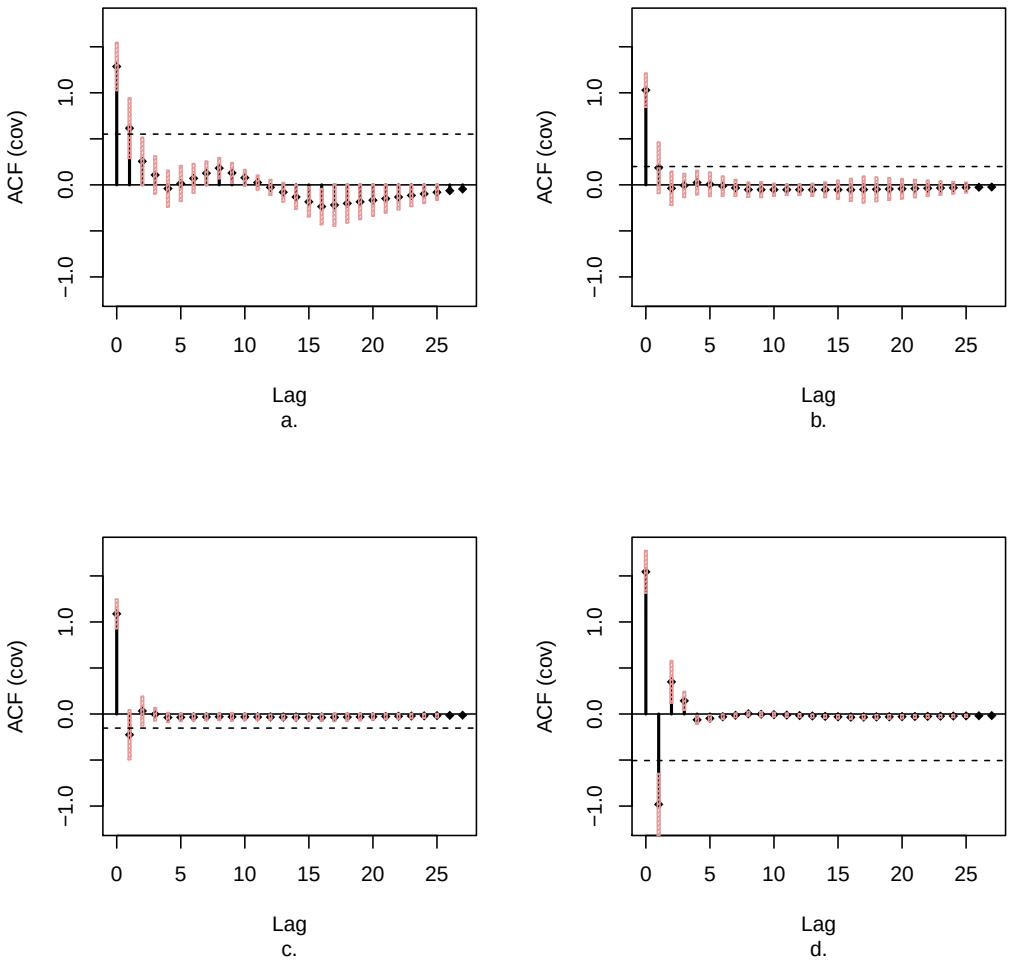


Figure 6: Localized autocovariance estimates, $\hat{c}(z, \tau)$, from a single realisation from nonstationary model P1, a time-varying AR(1) process. Plots *a.* to *d.* correspond to localizing at (rescaled) times of $z = 100/512, 200/512, 300/512, 400/512$ respectively (which corresponds to times $t = 100, 200, 300$ and 400). The horizontal dashed line in each figure corresponds to the true value of the AR parameter at that time point

Several interesting avenues lead from this current article. On tests of stationarity the case of the AR(1) process with parameter near to -1 is of interest (this is example S3 in Section 2.1 of Nason (2012)). Here, visually, such processes look extremely similar to, e.g., modulated white noise processes and confuse both our and the Priestley and Subba Rao (1969) test. It is likely that combining tests working from different principles, e.g. Dwivedi and Subba Rao (2011), might result in less false positives in these extreme situations. A referee suggested the interesting possibility of examining the statistic $\sum_{i,p} w_{i,p} \hat{v}_{i,p}^2$, with appropriate weights $w_{i,p}$, rejecting H_0 for large values, as the basis for a further test of stationarity, which might mitigate issues arising from the autocorrelation structure of the $\hat{v}$s

For the CIs, further work could extend them to the non-Gaussian situation. In principle, this might not be too complex providing the underlying spectrum satisfies reasonable smoothness constraints and the localized autocovariance smoothing can be used to induce asymptotic Gaussianity. Further improvement of the existing methodology could be achieved by better modelling the asymptotic distribution of the localized autocovariance, especially for small sample sizes.

The Associate Editor kindly pointed out that the tests of stationarity here (and elsewhere) give local information about the location of nonstationarities and that this information could feed into the estimation of the localized autocovariances. This good idea might work by applying the regular classical autocorrelation estimator to those portions of the series where the wavelet test of stationarity did not identify nonstationarities, but using the localized estimate on areas where nonstationarities were discovered. This is an intriguing idea for further work.

5 Acknowledgements

An `R` package called `locits` exists to implement the ideas in this paper (in due course will be available on the CRAN archive). The author was partially supported by the Research Councils UK Energy Programme. The Energy Programme is an RCUK cross-council initiative led by EPSRC and contributed to by ESRC, NERC, BBSRC and STFC.

A Appendix

A.1 Lemma for Proof of Theorem 1

Lemma 2. *Let $\{\psi_{\ell,v}\}$ and $\{\Psi_\ell(\tau)\}$ be the discrete wavelet and discrete autocorrelation wavelet for some integer $\ell > 0$. Suppose $0 \leq u \leq N_\ell - 1$ for integer u and N_ℓ be the (integer) support length of the discrete wavelet.*

For Haar wavelets and the Daubechies' compactly supported wavelet with two vanishing moments, for i sufficiently large, $E_{i,\ell,u} = \sum_{v=0}^{N_\ell-1} \Psi_i(u-v)\psi_{\ell,v} = -3 \times 2^{-i-\ell/2} C(u, N_\ell)$ where

$$C(u, N_\ell) = \begin{cases} u^2 + u - (N_\ell/2)^2 & \text{for } 0 \le u \le N_\ell/2 - 1 \\ -u^2 - u + 2uN_\ell + N_\ell - 3\,(N_\ell/2)^2 & \text{for } N_\ell/2 \le u \le N_\ell - 1. \end{cases} \tag{18}$$

For Daubechies' wavelets with three or more vanishing moments $\exists i$ such that $E_{k,\ell,u} = 0$ for $k \ge i$. For Shannon wavelets, $E_{i,\ell,u} = 0$ for all $i \ne \ell$.

Proof: Nason et al. (2000, Remark 5) show that $\Psi_i(\tau) = \Psi(2^{-i}\tau)$ where $\Psi(\tau)$ is the continuous autocorrelation wavelet.

Haar wavelets: It is well known that, for Haar wavelets, $\Psi(u) = \Psi_H(u) = 1 - 3|u|$, for $0 \le u \le 1/2$, also shown in Nason et al. (2000, Equation (13)). Hence, for i sufficiently large we have $-2^{i-1} \le u - v \le 2^{i-1}$ for any values of $0 \le u, v \le N_\ell - 1$ and hence

$$\Psi_i(u-v) = 1 - 3 \times 2^{-i}|u-v|, \tag{19}$$

for these values of u, v, i. (We retain N_ℓ in its general form here, to aid with the development for general wavelets later, but for Haar wavelets $N_\ell = 2$). Now substituting (19) into the formula for $E_{i,\ell,u}$ yields:

$$\begin{aligned} E_{i,\ell,u} &= \sum_{v=0}^{N_\ell-1} \psi_{\ell,v} - 3 \times 2^{-i} \sum_{v=0}^{N_\ell-1} \psi_{\ell,v}|u-v| \\ &= 0 - 3 \times 2^{-i-\ell/2} \left(\sum_{v=0}^{N_\ell/2-1} |u-v| - \sum_{v=N_\ell/2}^{N_\ell-1} |u-v| \right), \end{aligned} \tag{20}$$

the first part summing to zero as $\psi_{\ell,v}$ is a wavelet, and the sum splits in the second part as $\psi_{\ell,v} = 2^{-\ell/2}$ for $v = 0, \ldots, N_\ell/2 - 1$ and $\psi_{\ell,v} = -2^{\ell/2}$ for $v = N_\ell/2, \ldots, N_\ell - 1$.

For the first sum in (20), for $0 \le u \le N_\ell/2 - 1$ we have

$$\begin{aligned} \sum_{v=0}^{N_\ell/2-1} |u-v| &= \sum_{v=0}^{u} |v-u| + \sum_{u+1}^{N_\ell/2-1} |v-u| \\ &= \sum_{r=1}^{u} r + \sum_{r=1}^{N_\ell/2-1-u} r \\ &= u(u+1)/2 + (N_\ell/2 - 1 - u)\,(N_\ell/2 - u)\,/2 \\ &= u^2 + u - uN_\ell/2 + \tfrac{1}{2}\,(N_\ell/2 - 1)\,N_\ell/2. \end{aligned} \tag{21}$$

Similarly, for the second sum in (20) it can be shown that

$$\sum_{v=N_\ell/2}^{N_\ell-1} |u-v| = \tfrac{1}{2}\left(N_\ell/2-1\right) N_\ell/2 + \left(N_\ell/2\right)^2 - uN_\ell/2. \qquad (22)$$

Putting (21) and (22) together gives

$$\left(\sum_{v=0}^{N_\ell/2-1} |u-v| - \sum_{v=N_\ell/2}^{N_\ell-1} |u-v| \right) = u^2 + u - \left(N_\ell/2\right)^2, \qquad (23)$$

for $0 \leq u \leq N_\ell/2 - 1$.

Similarly, it can be shown that

$$\left(\sum_{v=0}^{N_\ell/2-1} |u-v| - \sum_{v=N_\ell/2}^{N_\ell-1} |u-v| \right) = -u^2 - u + 2uN_\ell + N_\ell - 3\left(N_\ell/2\right)^2, \qquad (24)$$

for $N_\ell/2 \leq u \leq N_\ell - 1$.

General Daubechies compactly supported wavelets. The case for general Daubechies follows the development above, but with the following modifications. Saito and Beylkin (1993), formula (3.45) show that the continuous autocorrelation wavelet satisfies $\Psi(x) = 2\Phi(x) - \Phi(x)$, where $\Phi(x)$ is the continuous autocorrelation function of the father wavelet and further point out that $\Phi(x)$ is the fundamental function of the symmetric iterative interpolation process from Deslauriers and Dubuc (1989). Deslauriers and Dubuc (1989) undertake a detailed examination of the smoothness properties of fundamental functions corresponding to the Daubechies' wavelets and indeed demonstrate that $\Phi(x)$ for the Daubechies compactly supported wavelet with two vanishing moments has one continuous derivative and for such wavelets with $M \geq 3$ vanishing moments $\Phi(x)$ has at least two continuous derivatives.

For all Daubechies wavelets with $M \geq 2$ vanishing moments we can use a Taylor expansion for $\Psi(x)$ around $x = 0$ by $\Psi(x) = 1 + \Psi^{(1)}(\xi)|x|$, $0 < \xi < x$, where necessarily $\Psi^{(1)}(\xi) < 0$ as $\Psi(x)$ achieves its global maximum at $x = 0$, where $\Psi^{(1)}$ is the first derivative of Ψ. Hence, we are essentially in the same situation as in (20) and the same result holds. For Daubechies wavelets with $M \geq 3$, vanishing moments we can exploit a second order Taylor expansion around $x = 0$ which is $\Psi(x) = 1 + \Psi^{(2)}(\xi)x^2/2$, the first

order term is zero as $\Psi(x)$ has a maximum at $x = 0$. Then, for i suitably large,

$$E_{i,\ell,u} = \sum_{v=0}^{N_\ell-1} \psi_{\ell,v} + \Psi^{(2)}(\xi) 2^{-1-2i} \sum_{v=0}^{N_\ell-1} (u-v)^2 \psi_{\ell,v} \quad (25)$$

$$= 2^{-1-2i}\Psi^{(2)}(\xi) \left(u^2 \sum_{v=0}^{N_\ell-1} \psi_{\ell,v} - 2u \sum_{v=0}^{N_\ell-1} v\psi_{\ell,v} + \sum_{v=0}^{N_\ell-1} v^2 \psi_{\ell,v} \right)$$

$$= 0, \quad (26)$$

since the wavelet has ≥ 3 vanishing moments which means $\sum \psi_{\ell,v} = \sum v\psi_{\ell,v} = \sum v^2\psi_{\ell,v} = 0$. Note, these results have been verified numerically.

Shannon wavelets. For Shannon wavelets it is the case that $E_{i,\ell,u} = 0$ for all $i \neq \ell$. To see this, let $\hat{E}_{i,\ell}(\omega)$ be the Fourier transform of $E_{i,\ell,u}$. Nason et al. (2000, Section A.5) show that the Fourier transforms of $\Psi_i(\tau)$ and $\psi_{\ell,v}$ are $\hat{\Psi}_i(\omega) = 2^i \chi_{C_i}(\omega)$ and $\hat{\psi}_i(\omega) = 2^{i/2} \exp(-2^{i-1}\omega)\chi_{C_i}(\omega)$ respectively, where $\chi_A(\omega)$ is the indicator function of the set A and $C_i = \left[-\frac{\pi}{2^{i-1}}, -\frac{\pi}{2^i}\right] \cup \left[\frac{\pi}{2^i}, \frac{\pi}{2^{i-1}}\right]$. Since $E_{i,\ell,u}$ is a convolution, $E_{i,\ell}(\omega)$ is the product of $\hat{\Psi}_i$ and $\hat{\psi}_\ell$ which is equal to zero if $i \neq \ell$ as their supports do not overlap.

A.2 Proof of Theorem 1

Recall that X_t is Gaussian with evolutionary wavelet spectrum $\{S_j(z)\}_{j=1}^\infty$ for $z \in (0,1)$. It is known from Nason et al. (2000, Proposition 4) that $\mathbb{E}I_{\ell,m} = \sum_i A_{i\ell} S_i(m/T)$. Thus, to obtain the covariance of $I_{\ell,m}$ with $I_{j,n}$, irrespective of whether $m = n$ or not, we examine:

$$\mathbb{E}(I_{\ell,m}I_{j,n}) = \mathbb{E}(d^2_{\ell,m}d^2_{j,n})$$

$$= \mathbb{E}\left(\sum_t X_t\psi_{\ell,m-t} \sum_s X_s\psi_{\ell,m-s} \sum_p X_p\psi_{j,n-p} \sum_q X_q\psi_{j,n-q} \right)$$

$$= \sum_{t,s,p,q} \psi_{\ell,m-t}\psi_{\ell,m-s}\psi_{j,n-p}\psi_{j,n-q}\mathbb{E}(X_tX_sX_pX_q). \quad (27)$$

For the fourth-order expectation, with X_t Gaussian, we use Isserlis (1918) which states:

$$\mathbb{E}(X_tX_sX_pX_q) = \mathbb{E}(X_tX_s)\mathbb{E}(X_pX_q) + \mathbb{E}(X_tX_p)\mathbb{E}(X_sX_q) + \mathbb{E}(X_tX_q)\mathbb{E}(X_sX_p). \quad (28)$$

Nason et al. (2000), equation (14), introduce the following formula for the (regular) time-varying autocovariance at time z with lag τ in terms of the evolutionary wavelet spectrum of an LSW process by: $c_T(z,\tau) = \sum_{j=1}^\infty S_j(z)\Psi_j(\tau) + \mathcal{O}(T^{-1})$.

For our problem, we need to evaluate each of the terms in (28), and so by

using the substitution $z = (t+s)/2$ and $\tau = s - t$ we can show

$$\mathbb{E}(X_t X_s) \simeq c\left(\frac{t+s}{2T}, s-t\right) = \sum_{j=1}^{\infty} S_j\left(\frac{t+s}{2T}\right)\Psi_j(s-t), \qquad (29)$$

where $x \simeq y$ means $x = y + \mathcal{O}(T^{-1})$.

Hence,

$$\begin{aligned}
\mathbb{E}(X_t X_s X_p X_q) &\simeq \sum_i S_i\left(\frac{t+s}{2T}\right)\Psi_i(s-t)\sum_h S_h\left(\frac{q+p}{2T}\right)\Psi_h(q-p) \\
&+ \sum_g S_g\left(\frac{t+p}{2T}\right)\Psi_g(p-t)\sum_f S_f\left(\frac{s+q}{2T}\right)\Psi_f(q-s) \\
&+ \sum_e S_e\left(\frac{t+q}{2T}\right)\Psi_e(q-t)\sum_d S_d\left(\frac{s+p}{2T}\right)\Psi_d(p-s) \\
&= A + B + C. \qquad (30)
\end{aligned}$$

Now substitute $\mathbb{E}(X_t X_s X_p X_q)$ back into (27) and considering the first term, A, only gives:

$$\begin{aligned}
&\sum_{i,s,t} S_i\left(\frac{t+s}{2T}\right)\psi_{\ell,m-t}\psi_{\ell,m-s}\Psi_i(s-t) \quad \times \\
&\sum_{h,p,q} S_h\left(\frac{q+p}{2T}\right)\psi_{j,n-p}\psi_{j,n-q}\Psi_h(q-p) \\
&\qquad = A1 \times A2. \qquad (31)
\end{aligned}$$

In $A1$ the substitution $u = m - t, v = m - s$ gives:

$$\begin{aligned}
A1 &= \sum_{i,u,v} S_i\left(\frac{m}{T} - \frac{u+v}{2T}\right)\Psi_i(u-v)\psi_{\ell,u}\psi_{\ell,v} \\
&= \sum_{i,u,v}\left\{S_i(m/T) + \mathcal{O}\left(\frac{|u+v|}{2T}\right)\right\}\Psi_i(u-v)\psi_{\ell,u}\psi_{\ell,v}, \qquad (32)
\end{aligned}$$

due to the Lipschitz continuity of S.

Remainder term in (32)*:* From Nason et al. (2000), formula (5), that the discrete wavelets $\psi_{\ell,u}$ are compactly supported and are non-zero for $u = 0, \ldots, N_\ell - 1$ where $N_\ell = (2^\ell - 1)(N_h - 1) + 1$, where N_h is the number of non-zero elements of $\{h_k\}$ the low pass quadrature mirror filter used in the construction of the Daubechies (1992) compactly supported continuous time wavelets. As a consequence, the autocorrelation wavelet $\Psi_i(\tau)$ is also compactly supported on $\tau = 1 - N_i, \ldots, N_i - 1, N_i$ as noted in Eckley and Nason (2005, Section 2.1). Putting in the appropriate bounds on the sums in

the remainder term (32) implies:

$$\begin{aligned}(32) &\leq (2T)^{-1}\sum_{i=1}^{\infty}\sum_{u=0}^{N_\ell-1}\sum_{v=\max(0,u-N_i+1)}^{\min(N_\ell-1,u+N_i-1)}|u+v|\Psi_i(u-v)\psi_{\ell,u}\psi_{l,v}\\ &= (2T)^{-1}\sum_{i=1}^{\infty}R_\ell(i), \qquad (33)\end{aligned}$$

where $R_\ell(i)$ is the inner sum over u, v in (33) for fixed ℓ. We now show that the summation in (32) is finite and hence the whole remainder is $\mathcal{O}(T^{-1})$. To do this consider a single term in the initial sum, i.e. fix on some i. Then for i suitably large

$$\begin{aligned}R_\ell(i) &= \sum_{u=0}^{N_\ell-1}\sum_{v=0}^{N_\ell-1}|u+v|\Psi_i(u-v)\psi_{\ell,u}\psi_{\ell,v}\\ &= \sum_{u=0}^{N_\ell-1}\psi_{\ell,u}\sum_{v=0}^{N_\ell-1}(u+v)\Psi_i(u-v)\psi_{\ell,v}\\ &= \sum_{u=0}^{N_\ell-1}u\psi_{\ell,u}\sum_{v=0}^{N_\ell-1}\Psi_i(u-v)\psi_{\ell,v}+\sum_{u=0}^{N_\ell-1}\psi_{\ell,u}\sum_{v=0}^{N_\ell-1}v\Psi_i(u-v)\psi_{\ell,v}\\ &= \sum_{u=0}^{N_\ell-1}u\psi_{\ell,u}\sum_{v=0}^{N_\ell-1}\Psi_i(u-v)\psi_{\ell,v}+\sum_{v=0}^{N_\ell-1}v\psi_{\ell,v}\sum_{u=0}^{N_\ell-1}\Psi_i(v-u)\psi_{\ell,u}\\ &= 2\sum_{u=0}^{N_\ell-1}u\psi_{\ell,u}\sum_{v=0}^{N_\ell-1}\Psi_i(u-v)\psi_{\ell,v} \qquad (34)\\ &= 2\sum_{u=0}^{N_\ell-1}u\psi_{\ell,u}E_{i,\ell,u}, \qquad (35)\end{aligned}$$

where $E_{i,\ell,u}$ is derived in Lemma 2. For Shannon wavelets and Daubechies' compactly supported wavelets with three or more vanishing moments Lemma 2 shows that $\exists i$ such that $E_{k,\ell,u} = 0$ for all $k \geq i$. Hence, $R_\ell(k) = 0$ for all $k > i$ also.

For Haar wavelets, or Daubechies compactly supported wavelets with two vanishing moments one can show, after some algebra, that, for i suitably large,

$$R_\ell(i) = -2^{-i-\ell-2}\left(\tfrac{5}{4}N_\ell^4 + N_\ell^2\right). \qquad (36)$$

Hence $R_\ell(i)$ is subject to exponential decay in i for fixed ℓ.

Therefore, in all cases. the series in (33) converges by the comparison test for series.

Main term in (32)*:* is

$$\begin{aligned} A1 &\simeq \sum_i S_i(m/T) \sum_u \sum_v \Psi_i(u-v)\psi_{\ell,u}\psi_{\ell,v} \\ &= \sum_i S_i(m/T) \sum_r \Psi_i(r) \sum_u \psi_{\ell,u}\psi_{\ell,u-r} \\ &= \sum_i S_i(m/T) A_{i,\ell}. \end{aligned} \tag{37}$$

Similarly, $A2 \simeq \sum_h S_h(n/T)A_{h,j}$.
For the second term in (30), substituting B into (27) we have

$$\begin{aligned} &\sum_{g,t,p} S_g\left(\frac{t+p}{2T}\right)\psi_{\ell,m-t}\psi_{j,n-p}\Psi_g(p-t) \quad \times \\ &\sum_{f,s,q} S_f\left(\frac{s+q}{2T}\right)\psi_{\ell,m-s}\psi_{j,n-q}\Psi_g(q-s) \quad = \quad B1 \times B2. \end{aligned} \tag{38}$$

For $B1$ substitute $u = m - t, v = n - p$ to obtain

$$B1 = \sum_{g,u,v} S_g\left\{\frac{m+n-(u+v)}{2T}\right\}\psi_{\ell,u}\psi_{j,v}\Psi_g(n-m+u-v). \tag{39}$$

Then use the Lipschitz continuity of S_g to obtain:

$$B1 = \sum_{g,u,v}\left[S_g\{(m+n)/2T\} + \mathcal{O}\left\{\frac{|u+v|}{2T}\right\}\right]\psi_{\ell,u}\psi_{j,v}\Psi_g(n-m+u-v). \tag{40}$$

Remainder term in (40)*:* is precisely the same as that in (32) except that $\psi_{\ell,v}$ is replaced by $\psi_{j,v}$ and the argument of Ψ_g is $n-m+u-v$ instead of $u-v$. The remainder term can be dealt with in exactly the same way as for $A1$ by appealing to Lemma 2 (where N_ℓ is replaced by N_j and similarly $\psi_{\ell,v}$ by $\psi_{j,v}$) and seeing that, for Haar wavelets, $E_{g,j,n-m+u} = -3\times 2^{-g}K \times P_2(n-m+u)$ where K is some constant, $P_2(x)$ some polynomial of order 2 and n, m, j are fixed. Similar arguments apply for more general wavelets.

Main term in (40)*:* By substituting $\tau = n - m + u - v$ for the main term in (40), $B1$ can be written as

$$B1 \simeq \sum_g S_g\left(\frac{n+m}{2T}\right)\sum_\tau \Psi_g(\tau)\sum_u \psi_{\ell,u}\psi_{j,u+n-m-\tau} \tag{41}$$

$$= \sum_g S_g\left(\frac{n+m}{2T}\right)\sum_\tau \Psi_g(\tau)\Psi_{\ell,j}(n-m-\tau). \tag{42}$$

Clearly, $B2 = B1$ and the component involving C can be established using

precisely the same argument.

Finally, for $m \neq n$ we have that $A1 \times A2$ cancels exactly with $\mathbb{E}(I_{\ell,m})\mathbb{E}(I_{j,n})$ the second term in the covariance, so the final covariance term is $B1 \times B2$ and the C component is identical resulting in the factor 2 in (16).

For $n = m$ formula (38) becomes identical to (31) and the third term is also the same. Hence, the covariance in this case is $A1 \times A2 + B1 \times B2 + C1 \times C2 - E(I_{\ell,m})\mathbb{E}(I_{j,m}) = 2\sum_i S_i(m/T)A_{i,\ell} \sum_k S_k(m/T)A_{k,j} = 2\beta_\ell(m/T)\beta_j(m/T)$.

A.3 Proof of Lemma 1

For clarity of exposition we drop the (ℓ) superscript.

(i) We first deal with the asymptotic expectation of $\hat{v}_{i,p}$. We first rewrite equation (4) in a form suitable for more direct comparison with the formula for $\hat{v}_{i,p}$ given in (6). Using the wavelet dilation equation for Haar wavelets $\psi^H(z) = -\phi(2z) + \phi(2z-1)$, where $\phi(z)$ is the Haar father wavelet, see e.g. Nason (2008, (2.38)), we obtain:

$$\begin{aligned} v_{i,p} &= \int_0^1 \beta_\ell(z)\psi^H_{i,p}(z)\,dz \qquad (43) \\ &= 2^{\frac{J-i}{2}} \int_0^1 \beta_\ell(z)\left\{-\phi(2^{J-i+1}z - 2p) + \phi(2^{J-i+1}z - 2p - 1)\right\}, \qquad (44) \end{aligned}$$

where $J = \log_2(T)$ and some $\ell \in \mathbb{N}$. Noting that $\phi(z) = 1$ for $z \in (0,1)$ and $\phi(z) = 0$ otherwise, $2^{-J} = T$, formula (44) becomes, after rearrangement:

$$v_{i,p} = 2^{-i/2}\left\{\int_{2^{i-1}(2p-1)}^{2^{i-1}2p} \beta_\ell(z/T)\,dz - \int_{2^{i-1}(2p-2)}^{2^{i-1}(2p-1)} \beta_\ell(z/T)\,dz\right\}, \quad (45)$$

for $p = 1, 2, \ldots, 2^i$.

The proof of Proposition 4 in Nason et al. (2000) shows that $I_{j,k}$ is asymptotically unbiased and does not rely on Gaussianity (the remainder of that proof for variance does) and the result carries over easily to the case of bounded variation. Hence

$$\mathbb{E}\hat{v}_{i,p} = 2^{-i/2}\left\{\sum_{r=0}^{2^{i-1}-1} \beta_\ell\left(\frac{2^i p - r}{T}\right) - \sum_{q=2^{i-1}}^{2^i-1} \beta_\ell\left(\frac{2^i p - q}{T}\right)\right\} + o(T^{-1/2}), \quad (46)$$

for $p = 1, 2, \ldots, 2^i$. The $1/2$ in the error rate is due to the combination of the T^{-1} rate from the bias of $I_{j,k}$ combined with the $2^{-i/2}$ factor, summing over 2^i terms and Assumption 1. This result agrees with Nason et al. (2000, Theorem 3) for Gaussian processes and Lipschitz continuity.

Now to examine $\mathbb{E}\hat{v}_{i,p} - v_{i,p}$ we look at the difference of the first terms of

(45) and (46), f, by partitioning the integral into pieces that cover subintervals of size commensurate with the summation:

$$f = \sum_{m=0}^{2^{i-1}-1} \int_{2^i p-2^{i-1}+m}^{2^i p-2^{i-1}+m+1} \beta_\ell(z/T)\,dz - \sum_{m=0}^{2^{i-1}-1} \beta_\ell\left(\frac{2^i p - 2^{i-1}+1+m}{T}\right) \quad (47)$$

$$= \sum_{m=0}^{2^{i-1}-1} \left[\int_{2^i p-2^{i-1}+m}^{2^i p-2^{i-1}+m+1} \left\{\beta_\ell(z/T) - \beta_l\left(\frac{2^i p - 2^{i-1}+m+1}{T}\right)\right\} dz\right] \quad (48)$$

Hence: $|f| \leq 2^{i-1}O(T^{-1})\,\mathrm{TV}_{[0,1]}(\beta_\ell)$, by the same argument behind the bounding of $|R_T^{(1,1)}|$ on page 66 of Neumann and von Sachs (1997). Putting together the first and second term bounds gives $|\mathbb{E}\hat{v}_{i,p} - v_{i,p}| \leq O(2^{i/2}T^{-1}) = O(T^{-1/2})$ by Assumption 1. [Note: if we made the more restrictive assumption of β_ℓ being Lipschitz continuous, as in Nason et al. (2000), then writing $z = 2^i - 2^{i-1} + m + u$ for $u \in (0,1)$ gives a remainder in the inner integral as $O(uT^{-1})$ which when integrated gives $O(T^{-1})$, and then when summed gives the same order $O(T^{-1/2})$.]

(ii) For the second part we can write $v_{i,p}$ (again for some $\ell \in \mathbb{N}$) as

$$\hat{v}_{i,p} = \sum_r \psi_{i,r}^H I_{\ell,2^i p-r} = \sum_r \psi_{i,r}^H d^2_{\ell,2^i p-r} \quad (49)$$

$$= \sum_r \psi_{i,r}^H \left(\sum_t X_t \psi_{\ell,2^i p-r-t} \sum_s X_s \psi_{\ell,2^i p-r-s}\right) \quad (50)$$

$$= \sum_t \sum_s X_t X_s \sum_r \psi_{i,r}^H \psi_{\ell,2^i p-r-t}\psi_{\ell,2^i p-r-s} \quad (51)$$

$$= \sum_t \sum_s X_t X_s m_{t,s}, \quad (52)$$

where ψ^H are Haar wavelets and

$$m_{t,s} = \sum_r \psi_{i,r}^H \psi_{\ell,2^i p-r-t}\psi_{\ell,2^i p-r-s}, \quad (53)$$

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

Hence, $\hat{v}_{i,p}$, satisfies the conditions for Lemma 3.1 of Neumann and von Sachs (1997) which says that if X_t satisfies Assumptions 1–3 then $\hat{v}_{i,p} = \eta_T = X^T M X$, and if we let $\xi_T = Y^T M Y$ where $Y \sim N(0, \mathrm{cov}(X))$ then:

$$\mathrm{cum}_n(\eta_T) = \mathrm{cum}_n(\xi_T) + R_n, \quad (54)$$

where

$$R_n \leq 2^{n-2}C^{2n}((2n)!)^{1+\gamma} \max_{s,t}\{|M_{s,t}|\}\tilde{M}||M||_\infty^{n-2}, \quad (55)$$

and $\mathrm{cum}_n(\xi_T)|$ is bounded, see Neumann and von Sachs (1997).

As in Neumann and von Sachs (1997), our M matrix 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 length of the discrete wavelet $\psi_{\ell,\cdot}$ from Nason et al. (2000, (5)). The quantities relating the M in the remainder term R_n are as follows.

(a) For $\max_{s,t}\{M_{s,t}\}$.

$$|M_{t,s}| \leq \sum_r |\psi^H_{i,r}||\psi_{\ell,2^i p-r-t}||\psi_{\ell,2^i p-r-s}| = 2^{-i/2}\sum_r \kappa_\ell, \qquad (56)$$

where $\kappa_\ell = \max_{u,v} \psi_{\ell,v}\psi_{\ell,u}$. Due to the finite support of the discrete wavelets it can be seen that the range of r in (56) can be restricted to:

$$\max(2^i p - t - N_\ell + 1, 2^i - s - N_\ell + 1) \leq r \leq \min(2^i p - t, 2^i - s), \quad (57)$$

and if $t > s$ is assumed, $t = s + u$ then the restriction simplifies to $u \leq r \leq N_\ell - 1$. Hence, $\max_{t,s}\{M_{s,t}\} \leq 2^{-i/2}\{N_\ell - 1 - (t-s)\}\kappa_\ell = O(T^{-1/2})$. This is a crude estimate of the rate and could be improved by more careful examination of the decay of $\psi_{\ell,\cdot}$ rather than the rough estimate of κ_ℓ.

(b) For $||M||_\infty = \max_s\{\sum_t |M_{s,t}|\}$. Using (56) we have:

$$\begin{aligned}
\max_s \sum_t |M_{t,s}| &\leq \max_s 2^{-i/2}\sum_r |\psi_{\ell,2^i p-r-s}| \sum_t |\psi_{\ell,2^p p-r-t}| && (58)\\
&\leq \max_s 2^{-i/2}\sum_r |\psi_{\ell,2^i p-r-s}|\psi^*_\ell N_\ell && (59)\\
&\leq 2^{-i/2}\psi^*_\ell N_\ell \max_s \sum_r |\psi_{\ell,2^i p-r-s}| && (60)\\
&\leq 2^{-i/2}\psi^{*2}_\ell N^2_\ell = O(T^{-1/2}), && (61)
\end{aligned}$$

where $\psi^*_\ell = \max_s \psi_{\ell,s}$. Note that, for part (iii) $||M||_2 \leq ||M||_\infty$.

(c) For $\tilde{M} = \sum_s \max_t |M_{s,t}|$. Similar arguments to (b) show that this rate is at least $O(T^{-1/2})$.

Finally, Nason et al. (2000) showed that, for Gaussian processes,

$$\mathrm{var}(\hat{v}_{i,p}) = 2T^{-1}\int_0^1 \beta_\ell(z)^2\{\psi^H(z)\}^2\,dz + \mathcal{O}(2^i T^{-2}). \qquad (62)$$

Using Assumption 1 gives that the order in (62) is $\mathcal{O}(T^{-1})$. Putting together the matrix norm remainder terms to form R_n in (55) gives at least $\mathcal{O}(T^{-1})$, hence result.

(iii) As in Neumann and von Sachs (1997) we use their Lemma 3.1 to-

gether with Assumption 3 to show that

$$\lambda_{\max}(M)\lambda_{\max}(\mathrm{cov}(X)) = \mathcal{O}(T^{-1/2}) \sum_{1\leq t\leq T} \left\{ \sum_{s} \mathrm{cov}(X_s, X_t) \right\} \quad (63)$$

$$= \mathcal{O}(T^{-1/2}), \quad (64)$$

where $\lambda_{\max}(M)$ is the maximum eigenvalue of A and using the estimates form norm-quantities of M given in part (ii). See also Neumann (1994).

References

Adak, S. (1998) Time-dependent spectral analysis of nonstationary time series, *J. Am. Statist. Ass.*, **93**, 1488–1501.

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.

Brillinger, D. (1996) Uses of cumulants in wavelet analysis, *J. Nonparam. Statist.*, **6**, 93–114.

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

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

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.

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

Davies, S. and Bland, D. (2010) Interestingness detection in sports audio broadcasts, in *IEEE Ninth International Conference on Machine Learning and Applications (ICMLA)*, pp. 643–649.

Davis, R. A., Lee, T., and Rodriguez-Yam, G. (2006) Structural break estimation for nonstationary time series models, *J. Am. Statist. Ass.*, **101**, 223–239.

Deslauriers, G. and Dubuc, S. (1989) Symmetric iterative interpolation processes, *Constr. Approx.*, **5**, 49–68.

Draghicescu, D., Guillas, S., and Wu, W.-B. (2009) Quantile curve estimation and visualization for nonstationary time series, *J. Comp. Graph. Stat.*, **18**, 1–20.

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. A. and Nason, G. P. (2005) Efficient computation of the discrete autocorrelation wavelet inner product matrix., *Statistics and Computing*, **15**, 83–92.

Falkenberg, J., McNames, J., Aboy, M., and Burchiel, K. (2003) Segmentation of extracellular microelectrode recordings with equal power, in *Proceedings of the 25th Annual International Conference of the IEEE*, pp. 2475–2478.

Florian, G. and Pfurtscheller, G. (2000) Dynamic spectral analysis of event-related EEG data, *Electroencephalography and Clinical Neurophysiology*, **95**, 393–396.

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

Isserlis, L. (1918) On a formula for the product moment coefficient of any order of a normal frequency distribution in any number of variables, *Biometrika*, **12**, 134–139.

Lio, P. (2003) Wavelets in bioinformatics and computational biology: state of art and perspectives, *Bioinformatics*, **19**, 2–9.

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, arXiv:1603.06415 [stat.ME].

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.

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.

Ombao, H. C., Raz, J., von Sachs, R., and Malow, B. A. (2001) Automatic statistical analysis of bivariate nonstationary time series, *J. Am. Statist. Ass.*, **96**, 543–560.

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

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.

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

R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rosen, O., Stoffer, D., and Wood, S. (2009) Local spectral analysis via a Bayesian mixture of smoothing splines, *J. Am. Statist. Ass.*, **104**, 249–262.

Saito, N. and Beylkin, G. (1993) Multiresolution representations using the autocorrelation functions of compactly supported wavelets., *IEEE Trans. Sig. Proc*, **41**.

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.

Schuster, E. (1972) Joint asymptotic distribution of the estimated regression function at a finite number of distinct points., *Ann. Statist.*, **43**, 84–88.

Shumway, R. and Stoffer, D. (2006) *Time Series Analysis and Its Applications with R Examples*, Springer, New York.

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

Spanias, A., Loizou, P., and Chen, Y. (1993) Analysis/synthesis of speech using the short-time Fourier transform and a time-varying ARMA process, *IEICE Transactions on Fundamentals*, **E76-A**, 645–652.

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.

Vannucci, M. and Lio, P. (2001) Non-decimated wavelet analysis of biological sequences: Applications to protein structure and genomics., *Sankhyā* B, **63**, 218–233.

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

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