A WAVELET-BASED TEST FOR STATIONARITY

By Rainer von Sachs and Michael H. Neumann

Université catholique de Louvain and Humboldt University

First version received May 1998

Abstract. We develop a test for stationarity of a time series against the alternative of a time-varying covariance structure. Using localized versions of the periodogram, we obtain empirical versions of a reasonable notion of a time-varying spectral density. Coefficients with respect to a Haar wavelet series expansion of such a time-varying periodogram are an indicator of whether there is some deviation from covariance stationarity. We propose a test based on the limit distribution of these empirical coefficients.

Keywords. Locally stationary processes; stationarity; test; time series; wavelets.

1. INTRODUCTION

Often, the simplifying assumption of stationarity of time series, i.e. a second-order structure which is constant over time, is not justified in practice. Examples of non-stationary processes are numerous, and can be found, for instance, in biomedical time series analysis with measurements of blood pressure, enzyme levels, biomechanical movements or EEGs and ECGs etc. Other examples of non-stationary phenomena derive from electrical and acoustical engineering (Doppler signals, speech analysis) and geophysics. Here, sudden changes often arise in the time–frequency structure of these signals where the characterizing quantity of interest is now a time-varying spectrum rather than a changing covariance.

Hence it is important to have a tool for detecting changes in the second-order structure of a stochastic process. As is common practice, in this paper we will use estimates of the spectrum rather than looking at explicit estimates of the covariances. We will develop a new statistical test which is localized, as to our knowledge there exist only a few studies of this kind, and we will now explain what we mean by localization.

For the particular purpose of testing a single change-point in the covariance structure of an otherwise stationary Gaussian time series, Picard (1985) developed a test based on the statistic

$$Z_T = \sup_{\lambda \in [0,\pi]} \sup_{k \in \{1,\ldots,T-1\}} \psi\left(\frac{k}{T}\right) \left| \int_0^{\lambda} \{ I_{\langle 1,k \rangle}(\omega) - I_{\langle k+1,T \rangle}(\omega) \} d\omega \right| \tag{1.1}$$

where ψ is a suitable weight function and $I_{\langle 1,k\rangle}$ and $I_{\langle k+1,T\rangle}$ are periodograms on the segments $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_T$, respectively. This method has been generalized by Giraitis and Leipus (1992) to the case of linear processes, and has been modified by Rozenholc (1995) by using tapered periodograms. The test of Picard is based on estimates of the (possibly time-varying) spectral function $F(\lambda) = \int_0^\lambda f(\omega) d\omega$. In contrast, we intend to use estimates of the (possibly time-varying) spectral density directly. The relative merits of smoothing-based tests based on local characteristics like densities versus non-smoothing tests based on cumulative characteristics are discussed in a different context by Rosenblatt (1975) and Ghosh and Huang (1991). The essential message is that non-smoothing tests look primarily at global deviations, and are therefore well suited for detecting classical Pitman alternatives of the form $f = f_0 + n^{-1/2} g$, where n denotes the sample size. On the other hand, smoothing-based tests focus on more localized deviations, and are consequently more powerful for detecting alternatives of the form $f = f_0 + n^{-\delta} g(./n^{-\gamma})$ for suitable $\delta, \gamma > 0$.

In the present paper we develop a test of stationarity which is based on the following general idea: using special segmentations we test on a change of the autocovariance structure from segment to segment. To be more specific, we use a wavelet decomposition of an appropriate notion of a time-varying spectral density. A model which allows for a rigorous asymptotic theory in this context is developed by Dahlhaus (1997), who introduced the concept of locally stationary processes. However, we think that our proposed test procedure is applicable in more general terms.

Some technical prerequisites for our test can be taken from Neumann and von Sachs (1997), where an estimator of a time-varying spectral density $f(u, \omega)$ was developed. It is a multiple test with the null hypothesis of stationarity H_0: $f(u, \omega) = f(\omega)$, where each subtest checks the significance of a particular coefficient $\alpha_{j,k;j',k'} = \int\int f(u, \omega) \psi_{j,k}(u) \phi_{j',k'}(\omega) du\, d\omega$ in our decomposition with respect to some set of bivariate wavelet functions $\{\psi_{j,k}(u)\phi_{j',k'}(\omega)\}_{j,k;j',k'}$. Note that under H_0 all of these coefficients are equal to zero. To get estimates of the coefficients $\alpha_{j,k;j',k'}$ we consider two natural candidates for empirical versions of $f(u, \omega)$. First, we can use segmented periodograms $I_{\langle K,L\rangle}(\omega)$ as used in Dahlhaus (1997) and von Sachs and Schneider (1996), which are calculated on segments corresponding to the particular wavelet $\psi_{j,k}(u)$ in the time direction. Second, we can employ the so-called pre-periodogram introduced in Neumann and von Sachs (1997). This second method has advantages of adaptivity in the process of estimating the evolutionary spectral density.

Using an asymptotic result for the marginal distributions of our estimates of $\alpha_{j,k;j',k'}$, we obtain an appropriate critical value via Bonferroni's inequality. We prove that the error of the first kind is asymptotically not greater than the nominal error and we give a brief discussion on the power of our test. The practicability of this method for moderate sample sizes is investigated by simulations which are reported in Section 4. In addition we apply our method to a set of tremor data from neurobiology.

2. SOME BASIC CONCEPTS FOR NON-STATIONARY PROCESSES

2.1. *A framework for non-stationary processes: a model of local stationarity*

The null hypothesis is simply that the time series $\{X_t\}$ is covariance stationary. Non-stationarity is basically defined as any arbitrary deviation from covariance stationarity. To define non-stationarity on the level of spectral densities, we have to find an appropriate extension of the definition of the spectral density generalizing from the stationary case. A particular framework which also allows for rigorous asymptotic theory has recently been developed by Dahlhaus (1997). The basic idea of his model may be explained as follows. In order to estimate some object of interest (parameter, function, ...) consistently, one needs an increasing amount of information about each feature of this object. Independence or weak dependence of the observed data is one part of a possible set of sufficient conditions for that. If the object of interest is of infinite dimension, e.g. a curve, we also have to bound its complexity appropriately. In non-parametric regression, an asymptotic with a fixed function on a bounded interval is often taken as the target, and independent observations are made with an increasingly fine grid of design points which then guarantee a growing amount of information about the true function on every subinterval. In order to actually have such an increasing amount of information about the function at any point x_0, we have to be able to gain some information about $f(x_0)$ from the observations corresponding to design points close to x_0. This is guaranteed by appropriate smoothness assumptions on the regression function f. Dahlhaus uses basically this approach to define an appropriate framework for the asymptotic theory of non-stationary processes. He keeps the central parameters of a time series, the mean and the covariance structure, fixed and links them to a set of observations $X_1, \ldots, X_T$ on a growing time horizon by an appropriate rescaling of time. This leads to the following definition.

DEFINITION 2.1 (DAHLHAUS, 1997). A sequence of stochastic processes $X_{t,T}(t = 1, \ldots, T)$ is called locally stationary with transfer function A^0 and trend μ if there exists a representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} A^0_{t,T}(\omega)\exp(i\omega t)d\xi(\omega) \tag{2.1}$$

where

(i) $\xi(\omega)$ is a stochastic process on $[-\pi, \pi]$ with $\overline{\xi(\omega)} = \xi(-\omega)$, $E\xi(\omega) = 0$ and orthonormal increments, i.e. $\mathrm{cov}\{d\xi(\omega), d\xi(\omega')\} = \delta(\omega - \omega')d\omega$, $\mathrm{cum}\{d\xi(\omega_1), \ldots, d\xi(\omega_k)\} = \eta(\sum_{j=1}^k \omega_j)h_k(\omega_1, \ldots, \omega_{k-1})d\omega_1 \ldots d\omega_k$, where $\mathrm{cum}\{\ldots\}$ denotes the cumulant of order k, $|h_k(\omega_1, \ldots, \omega_{k-1})| \leqslant \mathrm{const}_k$ for all k (with $h_1 = 0$, $h_2(\omega) = 1$) and $\eta(\omega) = \sum_{j=-\infty}^{\infty}\delta(\omega + 2\pi j)$ is the period 2π extension of the Dirac delta function;

(ii) there exists a positive constant K and a smooth function $A(u, \omega)$ on

$[0, 1] \times [-\pi, \pi]$ which is 2π-periodic in ω, with $A(u, -\omega) = \overline{A}(u, \omega)$, such that for all T

$$\sup_{t,\omega} \left| A^0_{t,T}(\omega) - A\left(\frac{t}{T}, \omega\right) \right| \leqslant KT^{-1}. \tag{2.2}$$

$A(u, \omega)$ and $\mu(u)$ are assumed to be continuous in u.

REMARK 2.1. In (2.1), t denotes a time point in the set $\{1, 2, \ldots, T\}$ while u denotes a time point in the rescaled interval [0, 1], i.e. $u = t/T$. Note that (2.1) does *not* define a finer and finer discretized continuous-time process as T tends to infinity. Rather, it means that more and more data of the same local structure, given by $A(t/T, \omega)$, are observed with increasing T. Simple examples are given by a stationary process modulated by a time-changing variance (as in Dahlhaus, 1996, Example 1.1(i)) and by autoregressive moving-average (ARMA) processes with time-varying coefficients (cf. our simulated examples in Section 4). A more elaborate example for a model of a time-varying spectrum arising from mobile radio communication can be found in von Sachs and Schneider (1996).

This concept now allows for the definition of a time-varying spectral density.

DEFINITION 2.2. As the evolutionary spectrum of $\{X_{t,T}\}$ given in (2.1) we define for $u \in (0, 1)$

$$f(u, \omega) = \lim_{T\to\infty} \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \operatorname{cov}(X_{[uT-s/2],T}; X_{[uT+s/2],T}) \exp(-i\omega s) \tag{2.3}$$

where $X_{t,T}$ is defined according to (2.1) with $A^0_{t,T}(\omega) = A(0, \omega)$ for $t < 1$ and $A^0_{t,T}(\omega) = A(1, \omega)$ for $t > T$.

Under the smoothness assumptions on $A(u, \omega)$ as given in Section 3.1 below, this evolutionary spectrum equals $f(u, \omega) = |A(u, \omega)|^2$, and is uniquely defined. For stationary processes this spectral density becomes constant in time, i.e. $f(u, \omega) = f(\omega)$; hence the class of locally stationary processes is a true generalization including stationary processes.

Note that (2.3) could have been derived also on a purely heuristic level without the theory of locally stationary processes. If we assume that the covariances decay at a sufficiently fast rate as the lag order tends to infinity, and that the covariance structure changes slowly over time (which is in accordance with the idea of rescaling), then Definition 2.2 is obviously a reasonable generalization of the spectral density for stationary processes. Actually, since the covariances decay as the lag length increases, $f(u_0, \omega)$ is mainly determined by covariances of the X_t with $|u_0 T - t|$ small. Hence, the definition of $f(u_0, \omega)$ is already automatically localized in some sense. Furthermore, since the covariance structure is nearly the same over small segments, the definition of $f(u, \omega)$ is also stable in u, which means in turn that it is reasonable to include some X_ts with small $|u_0 T - t|$ in the definition of the spectrum near u_0.

2.2. *Two time-varying periodograms*

We discuss now the two possibilities for defining local periodograms for non-stationary processes which we will use as the main parts of our test on stationarity. Assume for simplicity that $\mu \equiv 0$. In the case of non-stationary time series it is natural to consider fitting time series models on small segments. Accordingly, we can also consider the usual periodogram on small segments as a starting point for further inference. This has been proposed by Dahlhaus (1997) for the purpose of fitting certain time series models locally to a non-stationary process, and by von Sachs and Schneider (1996) as a starting point for a wavelet estimator of the evolutionary spectrum. In the non-tapered case, such a local periodogram has the form

$$I_N(u, \omega) = \frac{1}{2\pi N} \left| \sum_{s=1}^{N} X_{[uT-N/2+s],T} \exp(-i\omega s) \right|^2 .$$

REMARK 2.2. Note that the role of the parameter N, which is usually assumed to obey $N \to \infty$ and $N/T \to 0$ as $T \to \infty$, is twofold. First, it delivers a cut-off point above which higher lags are not incorporated in the definition of the periodogram. Actually, $I_N(u, \omega)$ contains only estimates of the covariances up to lag $N - 1$; hence, too small a value of N will introduce some bias. Second, the definition of $I_N(u, \omega)$ already contains some smoothing in the time direction. In other words, the parameter N acts in two opposite ways as a smoothing parameter: whereas small values of N restrict the resolution in the frequency direction, large N restricts the resolution in the time direction. Of course, according to the uncertainty principle (see, for example, Priestley, 1981, p. 835), there is no loss due to the fact that the number of lags being incorporated in the segmented periodogram is not greater than the time window. Nevertheless, problems can occur with such a global parameter N. At any point u_0 there exists a choice of $N = N(u_0)$ which is connected to the 'bandwidth of stationarity' h_u; for the latter see Dahlhaus (1996). So an in general time-varying h_u over [0, 1] would call for possibly very different segment lengths $N(u)$ over [0, 1]. Moreover, there is the additional important problem of how to perform a data-driven choice of N. Usually, this parameter is chosen before one starts the 'smoothing machinery', i.e. before one gets information about the order of magnitude of the bandwidth of stationarity.

To avoid these shortcomings, Neumann and von Sachs (1997) introduced a different empirical version of $f(u, \omega)$. The basic idea is to avoid any kind of pre-smoothing at this stage, which amounts to choosing the time window as small as possible and the lag window as large as possible. These considerations led Neumann and von Sachs (1997) to the definition

$$I(u, \omega) = \frac{1}{2\pi} \sum_{k:1 \leqslant [uT-k/2],[uT+k/2] \leqslant T} X_{[uT-k/2],T} X_{[uT+k/2],T} \exp(i\omega k) \tag{2.4}$$

which was called the 'pre-periodogram'. It was used in Neumann and von Sachs (1997) as a starting point for a wavelet estimator of the evolutionary spectral density and was also applied in Dahlhaus (1999) to establish local likelihood methods as a tool for fitting semiparametric time series models to locally stationary processes.

REMARK 2.3. (i) In contrast to the usual periodogram, both non-stationary versions provide an appropriate localization in time. While this is obvious for the segmented periodogram $I_N(u, \omega)$, the localization is achieved in a more implicit way by $I(u, \omega)$. Actually, since $EX_{[uT-k/2],T}X_{[uT+k/2],T}$ tends to zero as $|k| \to \infty$, $I(u, \omega)$ also reflects the local covariance structure for t around uT.

(ii) An obvious advantage of the pre-periodogram over the segmented periodogram is that the choice of the appropriate bandwidths in the time and frequency directions is completely left to the major smoothing step. In contrast to $I_N(u, \omega)$, the pre-periodogram has a diverging variance as $T \to \infty$. However, it turns out that smoothing in time and smoothing in frequency both lead to a variance reduction; see also the calculations in Neumann and von Sachs (1997). This fact also explains why the wavelet estimator of the evolutionary spectral density considered in Neumann and von Sachs (1997) attains similar rates of convergence to the estimator based on the segmented periodogram considered in von Sachs and Schneider (1996).

3. THE TEST

3.1. *Derivation of the test statistic*

The test we intend to devise will be based on a decomposition of an empirical version of $f(u, \omega)$ with respect to a certain system of Haar wavelet functions. Anticipating their later use on the intervals [0, 1] and $[0, \pi]$, respectively, we define

$$\psi(u) = \begin{cases} 1 & \text{if } 0 \leqslant u \leqslant 1/2 \\ -1 & \text{if } 1/2 < u \leqslant 1 \end{cases}$$

and

$$\phi(\omega) = 1/\pi^{1/2} \qquad \text{for } 0 \leqslant \omega \leqslant \pi.$$

Furthermore, we set

$$\psi_{j,k}(u) = 2^{j/2}\psi(2^j u - k) \qquad \text{and} \qquad \phi_{j,k}(\omega) = 2^{j/2}\phi(2^j\omega - k\pi)$$

$$k = 0, \ldots, 2^j - 1.$$

In the following we estimate the coefficients

$$\alpha_{j,k;j',k'} = \int_0^1 \int_0^\pi f(u, \omega)\psi_{j,k}(u)\phi_{j',k'}(\omega) du\, d\omega$$

which may be interpreted as measures for the local contrast in the time direction. Under the null hypothesis H_0: $f(u, \omega) = f(\omega)$, all of these coefficients are equal to zero. This means that we have to test the hypothesis $\alpha_{j,k;j',k'} = 0$ for all $(j, k; j', k')$.

In sharp contrast to the problem of *estimating* $f(u, \omega)$, where the use of wavelets of higher regularity may lead to better rates of convergence than that of Haar wavelets, there is no such gain in the context of testing. The functions $\psi_{j,k}$ do a perfect job of detecting differences, whereas the scaling functions $\phi_{j,k}$ are used to stabilize the spectral estimate.

A natural estimate based on the segmented periodogram is

$$\begin{aligned}
\tilde{\alpha}^{(1)}_{j,k;j',k'} &= \int_0^1 \int_0^\pi [I_{\langle [k2^{-j}T],[(k+1/2)2^{-j}T]\rangle}(\omega)\chi\{u \in [k2^{-j}, (k+\tfrac{1}{2})2^{-j}]\} \\
&\quad + I_{\langle [\{(k+1/2)2^{-j}+1\}T],[(k+1)2^{-j}T]\rangle}(\omega)\chi\{u \in [(k+\tfrac{1}{2})2^{-j}, (k+1)2^{-j}]\}] \\
&\quad \times \psi_{j,k}(u)\phi_{j',k'}(\omega) du\, d\omega \\
&= 2^{(j+j')/2} \frac{1}{\pi^{1/2}} \int_{k'2^{-j'}\pi}^{(k'+1)2^{-j'}\pi} \{I_{\langle [k2^{-j}T],[(k+1/2)2^{-j}T]\rangle}(\omega) \\
&\quad - I_{\langle [\{(k+1/2)2^{-j}+1\}T],[(k+1)2^{-j}T]\rangle}(\omega)\} d\omega
\end{aligned} \tag{3.1}$$

where

$$I_{\langle K,L\rangle}(\omega) = \frac{1}{2\pi(L-K+1)} \left| \sum_{t=K}^{L} X_t \exp(-i\omega t) \right|^2$$

is the ordinary periodogram for the segment $X_K, \ldots, X_L$. In words, we pick a segment in the time–frequency plane which is determined by the indices $(j, k; j', k')$ of wavelet and scaling functions, respectively. On the two halves of this segment in time we calculate integrated periodograms, and if these show a significant difference we reject the null hypothesis of stationarity.

For the purposes of asymptotic considerations, in order to control the bias of the coefficients $\tilde{\alpha}^{(1)}_{j,k;j',k'}$, we assume that the dyadic segment lengths $N_j = N_j(T) = 2^{-(j+1)}T$ fulfil $N_j \gg T^{1/2}$. Note that, in our situation, there is no additional segmentation bias of order $O(N/T)$ as we estimate integrals of the spectrum over pre-defined dyadic segments. We define our assumption as follows.

(A1) $2^j = o(T^{1/2})$.

Analogously, we obtain for the pre-periodogram

$$\tilde{\alpha}^{(2)}_{j,k;j',k'} = \int_0^1 \int_0^{\pi} I(u,\omega)\psi_{j,k}(u)\phi_{j',k'}(\omega)du\,d\omega$$

$$= 2^{(j+j')/2}\frac{1}{\pi^{1/2}}\int_{k'2^{-j'}\pi}^{(k'+1)2^{-j'}\pi}$$

$$\times \left\{ \int_{k2^{-j}}^{(k+1/2)2^{-j}} I(u,\omega)du - \int_{(k+1/2)2^{-j}}^{(k+1)2^{-j}} I(u,\omega)du \right\} d\omega. \quad (3.2)$$

For simplicity of notation, we use the multi-index $I = (j, k; j', k')$. Let

$$\mathscr{T}_T = \{I | 0 \leqslant j + j' \leqslant \log_2(J_T), \quad 0 \leqslant k \leqslant 2^j - 1, \quad 0 \leqslant k' \leqslant 2^{j'} - 1\} \quad (3.3)$$

be the set of indices that correspond to those $\tilde{\alpha}_I^{(i)}$s to be used in our test, where

$$J_T = O(T^{1-\rho}) \quad (3.4)$$

for some $\rho > 0$.

To complete the construction of the test, we have to know at least the asymptotic distribution of the $\tilde{\alpha}_I$s. It will turn out that, under natural conditions, a large number of these $\tilde{\alpha}_I$ are asymptotically normally distributed. Since it is natural in this non-parametric context to base the test on an increasing number of $\tilde{\alpha}_I$s, we need an appropriate formulation of this fact in terms of probabilities of large deviations.

We use the following assumptions.

(A2) (a) $\sup_{u,\omega}|A(u,\omega)| < \infty$.
(b) $\inf_{u,\omega}|A(u,\omega)| \geqslant \kappa$ for some $\kappa > 0$.
(c) $A(u,\omega)$ has a uniformly bounded total variation with respect to both time u and frequency ω, i.e. $\sup_u TV_{[0,\pi]}\{A(u,.)\} < \infty$ and $\sup_\omega TV_{[0,1]}\{A(.,\omega)\} < \infty$.

(A3) Let $\hat{A}(u,s) := (1/2\pi)\int A(u,\omega)\exp(i\omega s)d\omega, \quad s \in \mathbb{Z}, \quad u \in [0,1]$. Then assume

(a) $\Sigma_s \sup_u |\hat{A}(u,s)| < \infty$
(b) $\Sigma_s TV_{[0,1]}\{\hat{A}(\cdot,s)\} < \infty$
where $TV_{[0,1]}\{\hat{A}(.,l)\}$ denotes the total variation of the Fourier transform $\hat{A}(.,l)$ of $A(.,\omega)$ as a function of the first argument $u \in [0,1]$.

(A4) $\sup_{1\leqslant t_1 \leqslant T}\{\sum_{t_2,\ldots,t_k=1}^T |\mathrm{cum}(X_{t_1,T},\ldots,X_{t_k,T})|\} \leqslant C^k(k!)^{1+\gamma}$ for all $k = 2, 3, \ldots,$ where $\gamma \geqslant 0$.

REMARK 3.1. Conditions (A2) and (A3) are fulfilled in particular if $A(u, \omega)$ is differentiable in both arguments with uniformly bounded partial derivatives. We decided to use these weaker assumptions to allow for jumps of $f(u, \omega) = |A(u, \omega)|^2$ in time u. An important class of processes which fulfil these assumptions are time-varying ARMA processes with coefficient functions which can have jumps in time. In the case of MA, for example, condition (A3) simplifies considerably, being only a condition on the summability and the time variation of the MA coefficients. Furthermore, it was shown in Neumann (1994) that (A4) is fulfilled if $\{X_t\}$ is α-mixing with coefficients $\alpha(s) \leqslant K \exp(-b|s|)$ and

$$E|X_t|^k \leqslant C^k (k!)^\rho \qquad \text{for all } k. \tag{3.5}$$

Relation (3.5) is known to be satisfied for many distributions that can be found in the literature for an appropriate choice of ρ. In Johnson and Kotz (1970) we can find closed forms of higher order cumulants of the exponential, gamma and inverse Gaussian distributions, which show that this condition is satisfied for $\rho = 0$. The need for a positive ρ occurs in the case of a heavier-tailed distribution, which could arise as the distribution of a sum of weakly dependent random variables.

PROPOSITION 3.1. Suppose that Assumptions (A1)–(A4) are fulfilled. Let $\Delta_T = C(\log T)^{1/2}$ for any fixed $C < \infty$. Then

$$P\{\pm(\tilde{\alpha}_I - \alpha_I)/\sigma_I \geqslant x\} = \{1 - \Phi(x)\}\{1 + \mathrm{o}(1)\}$$

holds uniformly in $-\infty \leqslant x \leqslant \Delta_T$ and $I \in \mathscr{T}_T$, where $\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt$ denotes the standard normal cumulative distribution function and

$$\sigma_I^2 = 2\pi T^{-1} \int_0^1 \int_0^\pi f^2(u, \omega)\psi_{j,k}^2(u)\phi_{j',k'}^2(\omega)du\, d\omega + \mathrm{o}(T^{-1}) + \mathrm{O}(2^{-j'}T^{-1}).$$

The proof of this proposition is analogous to that of Proposition 3.1 in Neumann and von Sachs (1997), and is therefore omitted. Note, however, that here, by (A3), part (b), we use slightly stronger assumptions on the smoothness of $A(u, \omega)$, basically to allow for the use of less regular Haar scaling functions $\phi(\omega)$.

As a minimum prerequisite for our test we have to define a consistent estimate of the variance σ_I^2 of $\tilde{\alpha}_I$. Under H_0, with $\int \psi_{j,k}^2(u)du = 1$, we have

$$\sigma_I^2 = 2\pi T^{-1} \int_0^\pi f^2(\omega)\phi_{j',k'}^2(\omega)d\omega + \mathrm{o}(T^{-1}) + \mathrm{O}(2^{-j'}T^{-1}).$$

Whereas it is rather laborious to estimate the $\mathrm{O}(2^{-j'}T^{-1})$ term caused by fourth-order cumulants, the estimation of the first term is quite easy. Since, for the ordinary periodogram $I_{\langle 1,T\rangle}(\omega)$,

$$EI_{\langle 1,T\rangle}^2(\omega) = 2f^2(\omega) + \mathrm{o}(1) \qquad \text{as } T \to \infty$$

we propose to estimate σ_I^2 simply by

$$\hat{\sigma}_I^2 = \frac{\pi}{T}\int_0^{\pi} I^2_{\langle 1,T\rangle}(\omega)\phi^2_{j',k'}(\omega)d\omega. \tag{3.6}$$

From well-known properties of the periodogram, it follows that

$$P(T|\hat{\sigma}_I^2 - \sigma_I^2| > T^{-\delta}) = \mathrm{O}(T^{-\lambda})$$

for suitable $\delta > 0$ and arbitrarily large $\lambda < \infty$.

Let α be the nominal level of our test. Since a result on the joint distribution of the $\tilde{\alpha}_I$s does not exist, we use a slightly conservative approach via Bonferroni's inequality and define $\alpha_T = \alpha/|\mathscr{T}_T|$. Then our test rejects H_0 if

$$|\tilde{\alpha}_I| > \hat{\sigma}_I \Phi^{-1}(1 - \alpha_T/2) \qquad \text{for any } I \in \mathscr{T}_T. \tag{3.7}$$

Although our simple approach for estimating σ_I^2 neglects terms caused by fourth-order cumulants, there are basically two settings under which the asymptotic error of the first kind does not exceed the desired error. This is obviously the case if the time series is Gaussian. Furthermore, since the term from the fourth-order cumulants becomes negligible when $j' \to \infty$, our test is also correct if $j' \to \infty$. The assumption of $j' \to \infty$ appears to be very natural in our fully non-parametric context.

THEOREM 3.1. *Suppose that (A1)–(A4) are fulfilled. Furthermore, suppose either (i) that $\{X_t\}$ is Gaussian, or (ii) that $j' \to \infty$. Then*

$$P_{\mathrm{H}_0}\{|\tilde{\alpha}_I| > \hat{\sigma}_I \Phi^{-1}(1 - \alpha_T/2) \quad \text{for any } I \in \mathscr{T}_T\} \leqslant \alpha + \mathrm{o}(1).$$

3.2. *A brief discussion on the power of the test*

Although the focus is often primarily on errors of the first kind, the power is an important quantity to compare different competing tests. Since no result on the joint distribution of the $\tilde{\alpha}_I$s is available, asymptotically exact power calculations seem to be out of reach. Nevertheless, some insight into the power properties of our test is provided by looking at certain special cases in the space of alternatives. First, from Proposition 3.1 we obtain, for $0 < \beta < 1 - \alpha$, that

$$P\{|\tilde{\alpha}_I| \leqslant \hat{\sigma}_I \Phi^{-1}(1 - \alpha_T/2) \quad \text{for all } I \in \mathscr{T}_T\} \leqslant \beta \tag{3.8}$$

if there exists an $I \in \mathscr{T}_T$ such that

$$|\alpha_I| \geqslant C_\beta T^{-1/2}\{\log(T)\}^{1/2} \tag{3.9}$$

for an appropriate constant C_β. We consider the simple case of a jump in $f(u, \omega)$. For simplicity of presentation, suppose that $X_t = \sigma(t/T)\varepsilon_t$, where $\varepsilon \sim \mathrm{N}(0, 1)$ are independent and identically distributed. To simplify our considerations further, we assume that the jumps are located at dyadic points. We look at the case of testing

$$\mathrm{H}_0: \qquad \sigma(u) \equiv \sigma_0$$

against

$$\mathrm{H}_1: \qquad \sigma(u) = \begin{cases} \sigma_0 & u \in [0,\, 1/2) \cup (1/2 + 2^{-j_0},\, 1] \\ \sigma_0 + c_T & u \in [1/2,\, 1/2 + 2^{-j_0}] \end{cases}$$

where $\sigma_0 > 0$ is fixed, and we look for the minimal value of c_T that guarantees a 'non-trivial power', i.e. $\beta < 1 - \alpha$. The parameter $j_0 > 0$ is used to model a transient deviation from stationarity. Now it is easy to see that

$$\max_{k,k'} \{|\alpha_{j,k;j',k'}|\} \asymp c_T 2^{-j'/2}[2^{-j_0} 2^{j/2} \wedge 2^{-j/2}] \tag{3.10}$$

where the maximum is attained for $(k + 1/2)2^{-j} = 1/2$. The right-hand side of (3.10) is maximized by the choice $j' = 0$ and $j = j_0$. Hence, (3.9) implies that $c_T = \tilde{C}_\beta 2^{j_0/2} T^{-1/2} \{\log(T)\}^{1/2}$ is a sufficient height of transient jump in $\sigma(u)$ of length 2^{-j_0}. Deviations from σ_0 of short duration are modelled by $j_0 = j_0(T) \to \infty$ in our simplified context. Hence, it is necessary to incorporate wavelets on these fine scales j_0 in order to be able to detect such a jump of minimal height.

So far, bivariate wavelet bases with mixed scale indices have rarely been used in the statistical literature. However, in our opinion their use has important advantages for detecting localized deviations from a time-constant spectral density. We refer to the examples of the next section.

4. A NUMERICAL STUDY

We now apply our new test procedures to some simulated examples which give an indication of the performance both on the null hypothesis of stationarity and on the alternative, i.e. with a spectrum $f(u, \omega)$ which is not constant in time u. Most of our simulations concern the test based on the segmented periodogram, i.e. with coefficients $\tilde{\alpha}^{(1)}_{j,k;j',k'}$ as given in (3.1), though we also show the use of the pre-periodogram as in (3.2). Note that in all our simulated examples the technical conditions of the previous section are fulfilled, as we consider Gaussian autoregressive processes. However, as this specific choice is due only to the convenience of the simulation algorithms, we would like to emphasize that the chosen examples have some generic character and could have been realized by simulating other classes of weakly dependent processes. This robustness is also confirmed by the successful application of our proposed method to our final example of a real data set.

We start with simulation of some stationary processes, all of length $T = 1024$, by generation of pseudo-random standard normal $\{\varepsilon_t\}$ and, possibly, transformation to some low-order autoregressive process with time-constant spectral density $f(\omega)$. For testing the null hypothesis H_0 we use the following set of seven Haar wavelet coefficients:

$$I_{7,1} = \{(j, k; 0, 0) | 0 \leqslant j \leqslant 2, 0 \leqslant k \leqslant 2^j - 1\}.$$

That is, in the frequency direction we start with only the Haar scaling function $\phi_{00}(\omega)$ on the coarsest scale. This also enters in Equation (3.6) for determining an estimate $\hat{\sigma}_I^2$ for the unknown variance σ_I^2. Fixing the nominal level of our test to $\alpha = 0.1$ we measure the error of the first kind e_0 in counting the exceedences in (3.7) based on the normal quantile $q_{7,1}$ for $\alpha_T = \alpha/|I_{7,1}| = 1/70$, which is $q_{7,1} = 2.45$.

The first three examples are the standard normal white noise ε_t, an autoregressive process of order 1, $X_t + a_1 X_{t-1} = \varepsilon_t$ with parameter $a_1 = -0.9$, and an AR(1) process with $a_1 = 0.9$. In 1000 simulation runs we observed the following rates e_0 of false rejection: for Example 1 (white noise), $e_0 = 0.105$; for Example 2 ($a_1 = -0.9$), $e_0 = 0.109$; and for Example 3 ($a_1 = 0.9$), $e_0 = 0.134$. We note that the number of simulation runs is large enough to ensure a small enough standard deviation over these pseudo-independent runs, and we observe empirical levels which are quite close to the nominal level $\alpha = 0.1$.

In the following examples of non-stationary processes, we simulate the performance on the alternative H_1 to get an idea about the error e_1 of the second kind. For this we simulated two time-varying autoregressive processes, which can be considered as quite typical examples of realizations of a non-stationary process motivated from Model (2.1):

$$X_{t,T} + \sum_{i=1}^{p} a_i\left(\frac{t}{T}\right) X_{t-i,T} = \varepsilon_t$$

with autoregressive parameters $a_i = a_i(t/T)$ being functions which change over time.

The first, our Example 4, can be considered as a symbolic transient, i.e. short non-stationarity of considerable size but short duration (note the similarity to the simplified example of Section 3.2). We start from a stationary AR(2) process $X_t + a_1 X_{t-1} + a_2 X_{t-2} = \varepsilon_t$ with $a_1 = -0.5$, $a_2 = 0.2$, up to time $t = T/2$. Then, for the short interval $t \in [T/2, T/2 + T/64]$ we switch to $Y_t = CX_t$ with some parameter $C > 1$ that we will vary appropriately. Finally for $t > T/2 + T/64$ we jump back to the original process X_t. In exactly the same way as above for the simulations of the stationary Examples 1–3, the error rates of the second kind $e_1 = e_1(C)$ depend, of course, on C. For C varying between 1.50 ('small jump'), 1.65 ('medium jump') and 1.75 ('large jump') we observe a monotonically falling error $e_1(C = 1.50) = 0.275$, $e_1(C = 1.65) = 0.121$, $e_1(C = 1.75) = 0.055$, by counting the frequency of failure of detection of the jump. This is compatible with the performance on the null hypothesis H_0. We display the time-varying (piecewise in u constant) spectrum of this example, with $C = 1.65$, in Figure 1. Observe the higher intensity in the short-duration segment in time.

The next example, Example 5, is a piecewise constant AR(1) process with parameter $a_1 = -0.95$ for $t \leqslant 0.6T$ and $a_1 = -0.99$ for $t > 0.6T$. So we have a

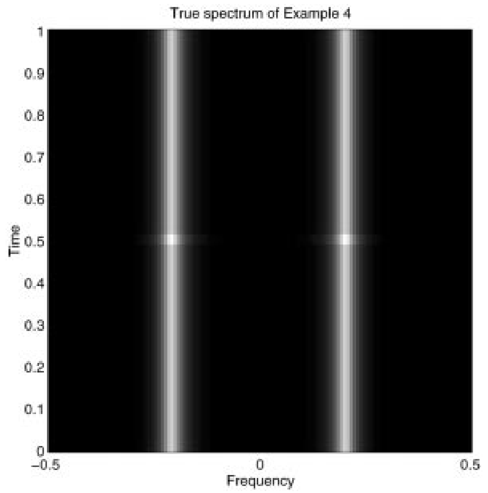

FIGURE 1. Example 4: True spectrum of AR(2) process 'transient with medium-size jump' ($C = 1.65$).

peak in the spectrum at zero frequency which gets sharper for the second segment of piecewise stationarity. Observe the plot in Figure 2 with higher intensity in the second time segment. Using the same set-up as for Examples 1–4 and testing only seven wavelet coefficients will lead to a high error of the second kind of $e_1 = e_1(7) = 0.543$. It seems that for this and the next example, it is no longer sufficient to only use the scaling function in frequency $\phi_{00}(\omega)$

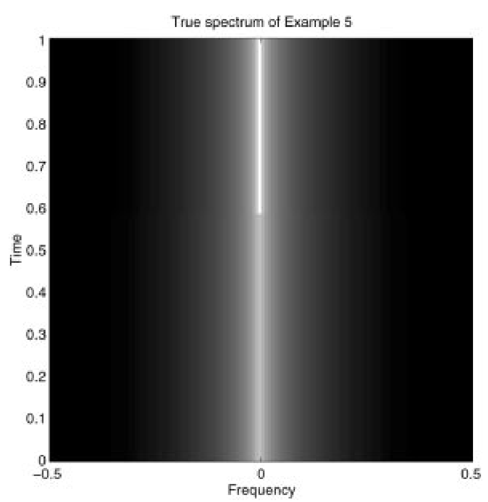

FIGURE 2. Example 5: True spectrum of piecewise constant AR(1) process.

on the coarsest level $j' = 0$: we need to include also some splitting in the frequency domain, to detect significant differences by wavelet coefficients obtained by integrating over smaller segments in the frequency direction. This seems to be necessary whenever the time-changing spectrum shows a higher spatial variability, as for the process of this example 5 with a lot of spectral mass concentrated around zero frequency. So now our new sets of incorporated indices are

$$I_{7,2} = \{(j, k; j', k')|0 \leqslant j \leqslant 2, 0 \leqslant k \leqslant 2^j - 1; j' = 1, 0 \leqslant k' \leqslant 2^{j'} - 1\}$$

with 14 coefficients, and

$$I_{7,4} = \{(j, k; j', k')|0 \leqslant j \leqslant 2, 0 \leqslant k \leqslant 2^j - 1; j' = 2, 0 \leqslant k' \leqslant 2^{j'} - 1\}$$

counting 28 coefficients, which is a sufficiently large subset of $\mathscr{T}_T$ for this purpose. We get the following errors of the second kind, again based on 1000 simulation runs, now with quantiles $q_{7,2} = 2.69$ and $q_{7,4} = 2.91$, respectively. For $I_{7,2}$, $e_1(14) = 0.203$, and for $I_{7,4}$, $e_1(28) = 0.200$, which is a big improvement compared with $e_1(7) = 0.543$ with the use of $I_{7,1}$. In this case we also checked the errors of the first kind $e_0(14)$ and $e_0(28)$ for the use of $I_{7,2}$ and $I_{7,4}$ in Examples 1 and 2, which, with sample size $T = 1024$, were the following: for Example 1 (standard Gaussian white noise), $e_0(14) = 0.111$ and $e_0(28) = 0.117$; and for Example 2 (AR(1) process with $a_1 = -0.9$), $e_0(14) = 0.176$ and $e_0(28) = 0.214$. As these values were quite high compared with those with I_7, we repeated the simulations with increased sample size $T = 2048$, and got $e_0(14) = 0.116$ and $e_0(28) = 0.158$. We conjecture that the higher the scale j' of the wavelet coefficients in frequency, the more data are needed to get close enough to the asymptotic normality of the empirical coefficients. This is not too surprising as the variance of the integrated classical periodogram is smaller the larger the range of integration in frequency is. In another, final simulation example, Example 6, again with $T = 1024$, we simulated a piecewise stationary autoregressive process of order 2, with parameters $a_1 = -0.60$ for $t \leqslant 0.6T$ and $a_1 = -0.208$ for $t > 0.6T$, and a constant $a_2 = 0.36$. This amounts to a sudden shift of the autoregressive peak from frequency $\pi/3$ to frequency $4\pi/9$, which is comparatively close. Compare the plot of the true spectrum in Figure 3. Here we suspect again the need for using the sets $I_{7,2}$ and $I_{7,4}$. Our simulations confirmed this conjecture as $e_1(7) = 0.694$, $e_1(14) = 0.093$ and $e_1(28) = 0.092$. Obviously, integration in frequency over the whole domain leads to wavelet coefficients $\tilde{\alpha}^{(1)}_{j,k;j',k'}$ in $I_{7,1}$ of similar size, and only integration over finer segments in frequency allows for significant differences. This would not be possible with the use of existing methods such as the one by Picard, as their test statistic (cf. Equation (1.1)) is based on uniform integration over frequency. So the real improvement of our new method shows up in Examples 5 and 6.

As mentioned at the beginning of this section, it is also of considerable interest to compare these results with the performance of the test using the pre-periodogram instead, i.e. with coefficients $\tilde{\alpha}^{(2)}_{j,k;j',k'}$. We repeated the experiments

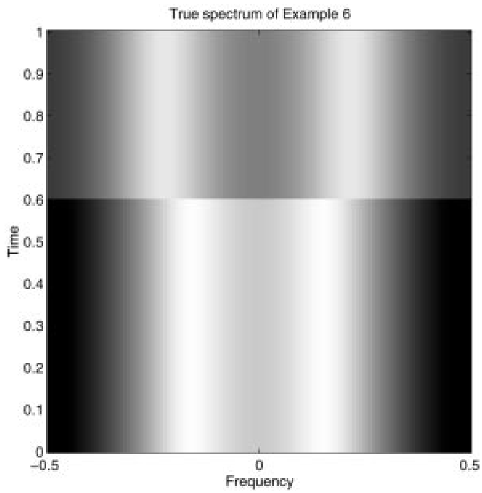

FIGURE 3. Example 6: True spectrum of piecewise constant AR(2) process.

done for Example 2 (i.e. under the null hypothesis H_0 of a stationary AR(1) with coefficient $a_1 = -0.9$) with $e_0(7) = 0.117$ for the use of $I_{7,1}$ and (to give a comparison with the performance on H_1, below) also $e_0(14) = 0.124$ for $I_{7,2}$. That is, the performance of the test on H_0, based on the pre-periodogram, is as close to the nominal level of $\alpha = 0.100$ as it was for the segmented periodogram.

As a typical illustration of the performance on the alternative H_1, we chose Example 4 with $C = 1.65$, where we got $e_1(7) = 0.106$ and, as an example with a higher number of coefficients needed, Example 6: here, $e_1(7) = 0.312$, whereas $e_1(14) = 0.102$, which is roughly the same improvement as that observed already for the segmented periodogram.

Finally we apply our procedure to a data set which is typical for the kind of applications we have in mind. The data shown in Figure 4(a) are the first 3072 observations of a set of tremor data recorded in the Cognitive Neuroscience Laboratory of the University of Quebec at Montreal in 1995 by Professors Anne Beuter and Roderick Edwards. The object of this study is to compare different regions of tremor activity coming from a subject with Parkinson's disease. Note that we added a Gaussian white noise of standard deviation 0.01 to the original data, in order to break the discrete nature of the data (where the level of the additional noise was kept small enough not to mask the variance structure in the original data). As our aim is to detect possible changes in the second-order structure of the data (and not in the trend), we investigated three consecutive segments of length 1024 of the first-order differenced series, shown in Figure 4(b). Using the simplest of our test procedures, i.e. the same configuration as for the simulated Examples 1–3 (based on $I_{7,1}$), separately on one segment after another, the results were as follows. The test of level $\alpha = 0.1$

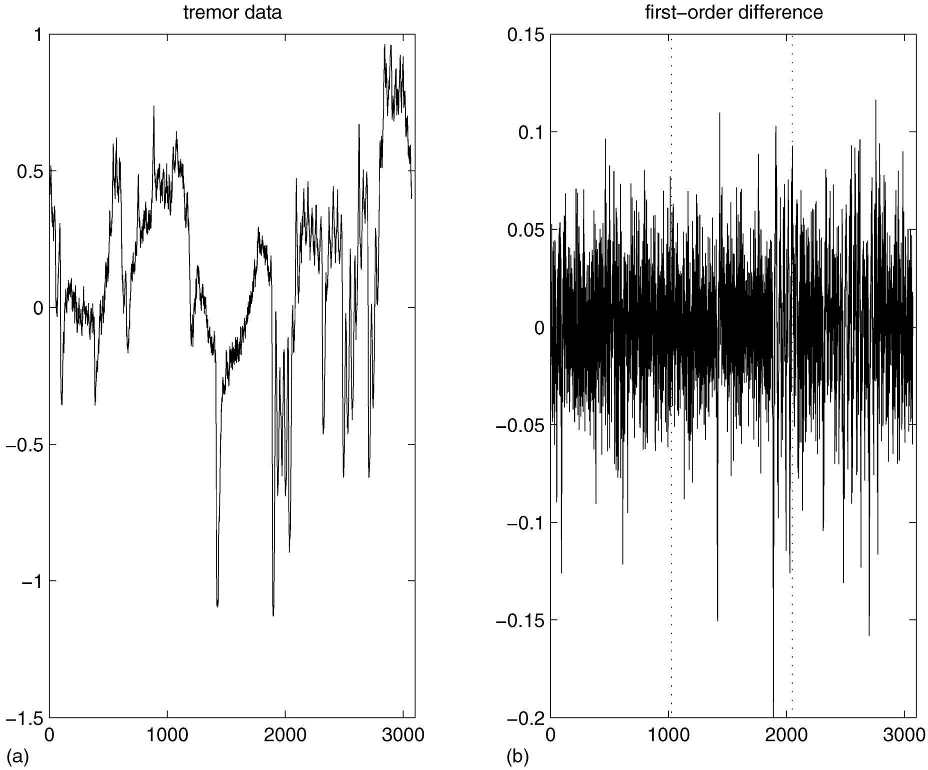

FIGURE 4. Example 7: Tremor data.

did not reject stationarity in the first and the third segment, whereas it rejected the null hypothesis for the second segment. This seems to be highly plausible as by inspection a possible break point can be anticipated in the region shortly before data point 2000, whereas the oscillation behaviour in the segments outside this neighbourhood seems to be rather homogeneous. This is in accordance with the findings of the neurologists who attributed only two different regimes of tremor activity to this particular part of the data.

Summarizing, our (simulated) examples have shown that the two test procedures not only seem to keep the nominal level on H_0, but also show sufficient power on H_1. As the method based on the pre-periodogram did not lead to a significant improvement for most simple simulated examples, we recommend the use of the algorithmically much faster method based on the segmented periodogram. However, an estimator based on two-dimensional tensor wavelet coefficients of the pre-periodogram investigated in Neumann and von Sachs (1997) proved useful for situations of considerably different regularity of the time-dependent spectrum $f(u, \omega)$ in time and frequency. We believe none the less that there are situations where it might be necessary to run a pre-periodogram based test, e.g. if a lot of frequency resolution would be necessary, or where a situation of long-range dependence might call for the need to incorporate a long range of lags, even for a locally changing spectrum in time. A definite advantage, however, is that we can consider the possibility of performing both estimation and testing simultaneously with the same non-parametric method. That is, we can use the empirical wavelet coefficients of the very estimation method we have chosen to perform the test for stationarity, and will possibly benefit, at least in the estimation, if we use the pre-periodogram.

ACKNOWLEDGEMENTS

The authors thank the University of Kaiserslautern and the SFB 373 at Humboldt University for making mutual research visits possible. Special thanks go to Professor Anne Beuter, Cognitive Neuroscience Laboratory of the University of Quebec at Montreal, and to Professor Roderick Edwards, now at the Department of Mathematics and Statistics, University of Victoria, Canada, for collecting and providing the tremor data (Example 7 of Section 4).

REFERENCES

DAHLHAUS, R. (1996) Asymptotic statistical inference for nonstationary processes with evolutionary spectra. *Athens Conference on Applied Probability and Time Series Analysis*, Vol. II (ed. P. M. Robinson, M. Rosenblatt), *Lect. Notes Stat.* 115, 145–59. New York: Springer.

—— (1997) Fitting time series models to nonstationary processes. *Ann. Stat.* 25, 1–37.

—— (1999) A likelihood approximation for locally stationary processes. *Beiträge zur Statistik 56*, Universität Heidelberg.

GHOSH, B. K. and HUANG, W.-M. (1991) The power and optimal kernel of the Bickel–Rosenblatt test for goodness of fit. *Ann. Stat.* 19, 999–1009.

GIRAITIS, L. and LEIPUS, R. (1992) Testing and estimating in the change-point problem of the spectral function. *Lithuanian Math. J.* 32, 15–29.

JOHNSON, N. L. and KOTZ, S. (1970) *Distributions in Statistics. Continuous Univariate Distributions—1, 2*. New York: Wiley.

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

—— and VON SACHS, R. (1997) Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Stat.* 25, 38–76.

PICARD, D. (1985) Testing and estimating change-points in time series. *Adv. Appl. Probab.* 17, 841–67.

PRIESTLEY, M. B. (1981) *Spectral Analysis and Time Series*, Vol. 2. London: Academic Press.

ROSENBLATT, M. (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Stat.* 3, 1–14.

ROZENHOLC, Y. (1995) Non-parametric tests of change-point with tapered data. Technical Report PMA-553. Université Paris Jussieu.

VON SACHS, R. and SCHNEIDER, K. (1996) Wavelet smoothing of evolutionary spectra by non-linear thresholding. *Appl. Comp. Harmonic Anal.* 3, 268–82.