

# A WAVELET-BASED TEST FOR STATIONARITY

Rainer von Sachs<sup>1</sup> and Michael H. Neumann<sup>2</sup>

September 17, 1997

## Abstract

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

*1991 Mathematics Subject Classification.* Primary 62M07; secondary 62M15.

*Key words and phrases.* Locally stationary processes, stationarity, test, time series, wavelets.

*Short title.* Test for stationarity.

---

<sup>1</sup>Department of Mathematics, University of Kaiserslautern

<sup>2</sup>Humboldt University, SFB 373, Berlin

## 1. INTRODUCTION

The simplifying assumption of *stationarity*, i.e. a second-order structure which is *constant* over time, is very often made in time series analysis. If this is actually adequate, it essentially simplifies the statistical analysis and allows one to use classical methods of data analysis. However, in the case of a strong deviation from stationarity, e.g. sudden or periodic changes in the covariance, this erroneous assumption can cause unexpected effects, and can finally lead to wrong conclusions about the underlying process. Examples of nonstationary processes are numerous, and can be found, for instance, in biomedical time series analysis whether the measurements are of blood pressure, enzyme levels, biomechanical movements or heartbeats, etc. In particular, in the analysis of series with pulsatile components, often a classical model of stationarity (and Gaussianity) is not sufficient to explain the data. We like to refer to Normolle and Brown (1994) who, among others, used a series of luteinizing hormone concentrations to illustrate such phenomena, occurring in the detection of seasonalities in the series. Other examples of nonstationary phenomena derive from electrical and acoustical engineering (Doppler signals, speech analysis, EEG's and ECG's), geophysics and economics.

Therefore, it is important to have some guidelines to assess the adequacy of the assumption of stationarity. When we focus on the mean and the covariance structure as the central characteristics of a time series, we can readily get some impression about possible deviations from stationarity by looking at nonparametric *estimates* of the mean function  $m$  and the time-varying spectral density. The latter has been treated by Neumann and von Sachs (1997), whereas von Sachs and MacGibbon (1997) is an example for estimation of local variation in the trend of the data (e.g., again the luteinizing hormone data) in the presence of a time-changing stochastic fluctuation. If one is interested in a decision rule on a more formal level, then one may employ *tests* for the hypothesis of stationarity.

There are different kinds of deviations from the hypothesis of stationarity one might be interested in. For example, one could suspect that there exists a single change in the mean of the observations or, more generally, a change in the marginal distributions. There already exists a long list of proposals for tests on changes in the mean, whereas the latter case was treated in Dümbgen (1994) in the case of independent observations.

In the present paper we focus on the detection of changes in the covariance structure of a stochastic process. In financial time series, the mean function is often of negligible order of magnitude, whereas the main interest is in the variability, that is the so-called volatility function. Black (1976) observes an effect called *volatility clustering*, i.e. high volatility followed by high, and low volatility tending to follow low one. The common econometric model for this observation is a positively-autocorrelated time series. Because an increasing volatility can be interpreted as an indication for a growing uncertainty on the market, it is of considerable importance to understand the dynamics with respect to possibly very sudden changes in the variance of stochastic stock return models, which consequently cannot be treated as stationary over time.

(Compare, e.g., also Weiner (1996), who performs a detection of jumps in a volatility function as a nonparametric regression function, though.)

Other examples include sudden changes in the time–frequency structure of speech or sound signals (cf. above) where the characterizing quantity of interest is again of second order, usually a time–varying spectrum rather than a changing covariance.

In total, for detection of either type of second-order nonstationarity, it is quite generally a common practice to use estimates of the spectrum rather than looking at explicit estimates of the covariances. This is because the spectral density effectively summarizes the characteristics of the covariance structure and also allows for a more convenient treatment of the problems. For the particular purpose of testing a single change-point in the covariance structure of an otherwise stationary Gaussian time series, Picard (1995) developed a test based on the statistic

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

where  $\psi$  is a suitable weight function and  $I_{\langle 1, k \rangle}$  and  $I_{\langle k+1, T \rangle}$  are periodograms on the segments  $X_1, \dots, X_k$  and  $X_{k+1}, \dots, X_T$ , respectively. This method has been generalized by Giraitis and Leipus (1992) to the case of linear processes, and has been modified by Rosenholc (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 propose to use directly estimates of the (possibly time-varying) spectral density. In different contexts, Rosenblatt (1975) and Ghosh and Huang (1991) discussed the relative merits of non-smoothing tests based on cumulative characteristics versus smoothing-based tests based on local characteristics like densities, for example. 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  $f = f_0 + n^{-1/2}g$ . 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  $\delta, \gamma > 0$ . To be more specific, an  $\alpha$ -test based on the  $L_2$ -distance between a nonparametric estimate with bandwidth  $n^{-\gamma}$  of  $f$  and an analogously smoothed version of  $f_0$  can attain a nontrivial power (greater than  $1 - \alpha$ ) for  $f = f_0 + n^{3\gamma/4 - 1/2}g(. / n^{-\gamma})$ ; see Theorem 2.2 in Ghosh and Huang (1991), whereas the power of the Kolmogorov-Smirnov test will converge to  $1 - \alpha$ . Moreover, a test based on the supremum deviation between a nonparametric estimate with bandwidth  $n^{-\gamma}$  and the correspondingly smoothed version of  $f_0$  may detect even smaller localized deviations.

There are different possibilities to implement a localization for such a smoothing-type test. In the context of testing a presumed additive structure of a regression function, that is  $H_0 : f(x_1, x_2) = f_1(x_1) + f_2(x_2)$ , Eubank, Hart, Simpson and Stefanski (1995) developed a test for the presence of interaction terms based on a Fourier series expansion of  $f$ .

In the present paper we develop a test based on 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. Our test is partially based upon ideas used in Neumann and von Sachs (1997) to develop a wavelet estimator

of the evolutionary spectral density  $f(u, \omega)$ . 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 0. We consider two natural candidates for empirical versions of  $f(u, \omega)$ . First, we can use segmented periodograms  $I_{(K,L)}(\omega)$  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 time direction. Second, we may employ the so-called pre-periodogram introduced in Neumann and von Sachs (1997). This second method has obvious advantages of adaptivity and also of sparsity in the process of *estimating* the evolutionary spectral density. As, in our context of testing for stationarity, our numerical simulations performed on typical test examples show similar results for both methods, we observe that, in the situation of data analysis and modelling, it is simultaneously possible to perform estimation and testing with the *same* nonparametric method. In other words, we can use the empirical wavelet coefficients of the very estimation method we have chosen to perform also the test for stationarity, and will possibly benefit at least on the estimation side of the problem if we choose the admittedly computationally more expensive pre-periodogram based method.

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 one and deliver 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.

## 2. SOME BASIC CONCEPTS FOR NONSTATIONARY PROCESSES

**2.1. A framework for nonstationary processes: A model of local stationarity.** The null hypothesis is simply that the time series  $\{X_t\}$  is covariance stationary. To define this, we do not need any such concept of nonstationarity. For the alternative, we have basically any arbitrary deviation from covariance stationarity in mind. To define nonstationarity 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,...) asymptotically 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, for example a curve, we also have to bound its complexity appropriately. In nonparametric regression, people often consider an asymptotics with a fixed function on a bounded interval as the target, and independent observations at an increasingly fine grid of design points which then guarantees 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 asymptotic theory of nonstationary 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  $X_1, \dots, X_T$  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  $X_{t,T}$  ( $t = 1, \dots, T$ ) is called *locally stationary* with transfer function  $A^o$  and trend  $\mu$  if there exists a representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} A_{t,T}^o(\omega) \exp(i\omega t) d\xi(\omega), \quad (2.1)$$

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.  $\text{cov}(d\xi(\omega), d\xi(\omega')) = \delta(\omega - \omega')d\omega$ ,  $\text{cum}\{d\xi(\omega_1), \dots, d\xi(\omega_k)\} = \eta(\sum_{j=1}^k \omega_j) h_k(\omega_1, \dots, \omega_{k-1}) d\omega_1 \dots d\omega_k$ , where  $\text{cum}\{\dots\}$  denotes the cumulant of order  $k$ ,  $|h_k(\omega_1, \dots, \omega_{k-1})| \leq \text{const}_k$  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  $2\pi$  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  $2\pi$ -periodic in  $\omega$ , with  $A(u, -\omega) = \overline{A(u, \omega)}$ , such that for all  $T$ ,

$$\sup_{t, \omega} |A_{t,T}^o(\omega) - A(t/T, \omega)| \leq K T^{-1}. \quad (2.2)$$

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

*Remark 2.1.* In (2.1),  $t$  denotes a time point in the set  $\{1, 2, \dots, T\}$  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. It rather means that more and more data of the same local structure, given by  $A(t/T, \omega)$ , are observed with increasing  $T$ . As illustration we cite Example 1.1(i) given in Dahlhaus (1996), which is

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \sigma\left(\frac{t}{T}\right) \cdot Y_t,$$

where  $\{Y_t\}$  is a stationary process and  $\mu, \sigma : [0, 1] \rightarrow \mathbb{R}$  are continuous. Here,  $\{X_{t,T}\}$  is locally stationary with  $A_{t,T}^o(\omega) = A(t/T, \omega)$ . If  $Y_t$  is an autoregressive process of order 2 with (complex) roots close to the unit circle, then  $Y_t$  shows a periodic behaviour and  $\sigma$  may be regarded as a time-varying amplitude function of the process  $\{X_{t,T}\}$ . With increasing  $T$  more and more cycles of the process with  $u = t/T \in [u_o - \varepsilon, u_o + \varepsilon]$ , i.e. with amplitude close to  $\sigma(u_o)$ , are observed.

Further examples are ARMA-processes with time-varying coefficients (cf. our simulated examples in Section 4).

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

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

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

where  $X_{t,T}$  is defined by  $A_{t,T}^o(\omega) = A(0, \omega)$  for  $t < 1$  and  $A_{t,T}^o(\omega) = A(1, \omega)$  for  $t > T$ . Under the smoothness assumptions on  $A(u, \omega)$  as, e.g., given further below in Section 3.1, this evolutionary spectrum equals  $f(u, \omega) = |A(u, \omega)|^2$ , and is uniquely defined, hence. 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 the 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 certain 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  $X_t$ 's with  $|u_0T - t|$  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  $X_t$ 's with small  $|u_0T - t|$  in the definition of the spectrum near  $u_0$ .

**2.2. Two time-varying periodograms.** We discuss now possibilities to define local periodograms in the case of nonstationary processes. Assume for simplicity  $\mu \equiv 0$ . In the case of nonstationary time series it is a natural idea to fit time series models on small segments. Accordingly, we can also consider a usual periodogram on small segments as a starting point for further inference. This has been proposed in Dahlhaus (1997) for the purpose of fitting certain time series models locally to a nonstationary process, and in von Sachs and Schneider (1996) as a starting point for a wavelet estimator of the evolutionary spectrum. In the nontapered case, such a local periodogram has the form

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

Note that the role of the parameter  $N$ , which is usually assumed to obey  $N \rightarrow \infty$  and  $N/T \rightarrow 0$  as  $T \rightarrow \infty$ , is a two-fold one: First, it delivers a cut-off point, from which on 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  $I_N(u, \omega)$  contains already some smoothing in 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 frequency direction, large  $N$  restrict the resolution in 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. 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 a 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”, that means 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 presmoothing 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 in Neumann and von Sachs (1997) to the definition

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

which was called “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 also applied in Dahlhaus (1995) to establish local likelihood methods as a tool for fitting semiparametric time series models to locally stationary processes. The obvious advantage of this new definition is, that the choice of the appropriate bandwidths in time and frequency direction is completely left to the major smoothing step. In contrast to  $I_{(K,L)}(\omega)$ , the pre-periodogram has a diverging variance as  $T \rightarrow \infty$ . However, it turns out that smoothing in time and frequency direction lead both 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 as 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

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

and

$$\phi(\omega) = 1/\sqrt{\pi} \quad \text{for } 0 \leq \omega \leq \pi.$$

Further, we set

$$\psi_{j,k}(u) = 2^{j/2} \psi(2^j u - k) \quad \text{and} \quad \phi_{j,k}(\omega) = 2^{j/2} \phi(2^j \omega - k\pi), \quad k = 0, \dots, 2^j - 1.$$

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 time direction. Under the null hypothesis,  $H_0 : f(u, \omega) = f(\omega)$ , all of these coefficients are equal to 0. That means that we have to test the hypothesis of  $\alpha_{j,k;j',k'} = 0$  for all  $(j, k; j', k')$ . A natural estimate based on the segmented periodogram is

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

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  $X_K, \dots, X_L$ .

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

**(A0)**  $2^j = o(T^{1/2})$ .

Analogously, we obtain for the pre-periodogram

$$\begin{aligned}
\tilde{\alpha}_{j,k;j',k'}^{(2)} &= \int_0^1 \int_0^\pi I(u, \omega) \psi_{j,k}(u) \phi_{j',k'}(\omega) dud\omega \\
&= 2^{(j+j')/2} \frac{1}{\sqrt{\pi}} \int_{k'2^{-j'}\pi}^{(k'+1)2^{-j'}\pi} \left[ \int_{k2^{-j}}^{(k+1/2)2^{-j}} I(u, \omega) du - \int_{(k+1/2)2^{-j}}^{(k+1)2^{-j}} I(u, \omega) du \right] d\omega.
\end{aligned} \tag{3.2}$$

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

$$\mathcal{I}_T = \left\{ I \mid 0 \leq j + j' \leq \log_2(J_T), \quad 0 \leq k \leq 2^j - 1, \quad 0 \leq k' \leq 2^{j'} - 1 \right\} \tag{3.3}$$

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

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

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 nonparametric 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:

- (A1) a)  $\sup_{u,\omega} |A(u, \omega)| < \infty$ .  
 b)  $\inf_{u,\omega} |A(u, \omega)| \geq \kappa$  for some  $\kappa > 0$ .  
 c)  $A(u, \omega)$  has a uniformly bounded total variation both wrt time  $u$  and frequency  $\omega$ , i.e.  $\sup_u TV_{[0,\pi]}(A(u, \cdot)) < \infty$  and  $\sup_\omega TV_{[0,1]}(A(\cdot, \omega)) < \infty$ .
- (A2) Let  $\hat{A}(u, s) := 1/(2\pi) \int A(u, \omega) \exp(i\omega s) d\omega$ ,  $s \in \mathbb{Z}$ ,  $u \in [0, 1]$ . Then assume,  
 a)  $\sum_s \sup_u |\hat{A}(u, s)| < \infty$ .  
 b)  $\sum_s TV_{[0,1]}(\hat{A}(\cdot, s)) < \infty$ ,  
 where  $TV_{[0,1]}(\hat{A}(\cdot, \ell))$  denotes the total variation of the Fourier transform  $\hat{A}(\cdot, \ell)$  of  $A(\cdot, \omega)$  as a function in the first argument  $u \in [0, 1]$ .
- (A3)  $\psi(u)$  and  $\phi(\omega)$  have bounded total variation on  $[0, 1]$  and  $[0, \pi]$ , respectively. Then it is in particular fulfilled that, for the Fourier transform  $\hat{\phi}(s)$  of  $\phi(\omega)$ , we have  $|\hat{\phi}(s)| = O(|s|^{-1})$ .
- (A4)  $\sup_{1 \leq t_1 \leq T} \left\{ \sum_{t_2, \dots, t_k=1}^T |\text{cum}(X_{t_1, T}, \dots, X_{t_k, T})| \right\} \leq C^k (k!)^{1+\gamma}$  for all  $k = 2, 3, \dots$ , where  $\gamma \geq 0$ .

**Proposition 3.1.** *Suppose that Assumptions (A0) to (A4) are fulfilled. Let  $\Delta_T = C\sqrt{\log T}$  for any fixed  $C < \infty$ . Then*

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

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

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

The proof of this proposition is analogous to that of Proposition 3.1 in Neumann and von Sachs (1997), and therefore omitted. Note however, that here, by (A2)(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 least prerequisite for our test we have to define a consistent estimate of the variance  $\sigma_I^2$  of  $\tilde{\alpha}_I$ . Under  $H_0$ , with  $\int \psi_{j,k}^2(u) du = 1$ , we have

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

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

$$EI_{(1,T)}^2(\omega) = 2f^2(\omega) + o(1) \quad \text{as } T \rightarrow \infty,$$

we propose to estimate  $\sigma_I^2$  simply by

$$\hat{\sigma}_I^2 = \frac{\pi}{T} \int_0^\pi I_{\langle 1, T \rangle}^2(\omega) \phi_{j', k'}^2(\omega) d\omega. \quad (3.5)$$

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

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

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

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

$$|\tilde{\alpha}_I| > \hat{\sigma}_I \Phi^{-1}(1 - \alpha_T/2) \quad \text{for any } I \in \mathcal{I}_T. \quad (3.6)$$

Although our simple approach of estimating  $\sigma_I^2$  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 one. This is obviously the case if the time series is Gaussian. Further, since the term from the fourth-order cumulants becomes negligible when  $j' \rightarrow \infty$ , our test is also correct if  $j' \rightarrow \infty$ . The assumption of  $j' \rightarrow \infty$  appears to be very natural in our fully nonparametric context.

**Theorem 3.1.** *Suppose that (A1) to (A4) are fulfilled. Further, suppose either that*

(i)  $\{X_t\}$  *is Gaussian,*

*or that*

(ii)  $j' \rightarrow \infty$ .

*Then*

$$P_{H_0}\left(|\tilde{\alpha}_I| > \hat{\sigma}_I \Phi^{-1}(1 - \alpha_T/2) \quad \text{for any } I \in \mathcal{I}_T\right) \leq \alpha + o(1).$$

**3.2. A brief discussion on the power of the test.** Although the focus is often primarily at the error of the first kind, the power is an important quantity to compare different competing tests. Since there is no result on the *joint* distribution of the  $\tilde{\alpha}_I$ 's 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

$$P\left(|\tilde{\alpha}_I| \leq \hat{\sigma}_I \Phi^{-1}(1 - \alpha_T/2) \quad \text{for all } I \in \mathcal{I}_T\right) \leq \beta \quad (3.7)$$

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

$$|\alpha_I| \geq C_\beta T^{-1/2} \sqrt{\log(T)} \quad (3.8)$$

with  $C_\beta < \infty$ . 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  $\varepsilon \sim N(0, 1)$  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

$$H_0 : \quad \sigma(u) \equiv \sigma_0$$

against

$$H_1 : \quad \sigma(u) = \begin{cases} \sigma_0, & u \in [0, 1/2) \cup (1/2 + 2^{-j_0}, 1], \\ \sigma_0 + c_T, & u \in [1/2, 1/2 + 2^{-j_0}] \end{cases}$$

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

$$\max_{k, k'} \{|\alpha_{j, k; j', k'}|\} \asymp c_T 2^{-j'/2} [2^{-j_0} 2^{j/2} \wedge 2^{-j/2}], \quad (3.9)$$

where the maximum is attained for  $(k + 1/2)2^{-j} = 1/2$ . The right-hand side of (3.9) is maximized by the choice  $j' = 0$  and  $j = j_0$ . Hence, (3.8) implies that  $c_T = 2^{j_0/2} T^{-1/2} \sqrt{\log(T)}$  is a sufficient height of a transient jump in  $\sigma(u)$  of length  $2^{-j_0}$ . Deviations from  $\sigma_0$  of short duration are modelled by  $j_0 = j_0(T) \rightarrow \infty$  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.

#### 4. A NUMERICAL STUDY

Finally we want to apply our new test procedures to some simulated examples which give an idea about 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  $\tilde{\alpha}_{j, k; j', k'}^{(1)}$  as given in (3.1), though we also give some comparison to the use of the pre-periodogram as in (3.2).

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

$$I_7 = \left\{ (j, k; 0, 0) \mid 0 \leq j \leq 2, 0 \leq k \leq 2^j - 1 \right\},$$

i.e. in frequency direction we start with using only the Haar scaling function  $\phi_{00}(\omega)$  on the coarsest scale. This also enters in equation (3.5) for determining an estimate  $\hat{\sigma}_I^2$  for the unknown variance  $\sigma_I^2$ .

Fixing the nominal level of our test to  $\alpha = 0.1$  we measure the error of first kind  $e_0$  in counting the exceedences in (3.6) based on the normal quantile  $q_7$  for  $\alpha_T = \alpha/|I_7| = 1/70$  which is  $q_7 = 2.45$ .

Our examples 1 – 3 are the standard normal white noise  $\varepsilon_t$ , 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 pretty close to the nominal level of  $\alpha = 0.1$ .

In the now 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 second kind. For this we simulated two situations of time-varying autoregressive processes, which can be considered as quite typical examples of realizations of a nonstationary process motivated from model (2.1),

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

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 cartoon for a transient, i.e. a short nonstationarity 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  $Y_t = C \cdot X_t$  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 set-up as above for the simulations of the stationary examples 1-3, the error rates of second kind  $e_1 = e_1(C)$  depend, of course, on  $C$ :

For  $C$  varying between  $C = 1.50$  (“small jump”),  $C = 1.65$  (“medium jump”) and  $C = 1.75$  (“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  $t \leq 0.6 T$  and  $a_1 = -0.99$  for  $t > 0.6 T$ . So we have a 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.

Running the same set-up as for examples 1-4, with testing only seven wavelet coefficients will lead to a high error of second kind of  $e_1 = e_1(7) = 0.543$ . It seems that for this, and the next example to follow, too, it is no more sufficient to only use the scaling function in frequency  $\phi_{00}(\omega)$  on the coarsest level  $j' = 0$ : We need to include also some splitting in the frequency domain, to detect significant differences by wavelet coefficients where we integrate over smaller segments in frequency direction. This seems to be necessary whenever the time-changing spectrum shows a higher spatial variability, like 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

$$I_{14} = \left\{ (j, k; j', k') \mid 0 \leq j \leq 2, 0 \leq k \leq 2^j - 1; j' = 1, 0 \leq k' \leq 2^{j'} - 1 \right\} ,$$

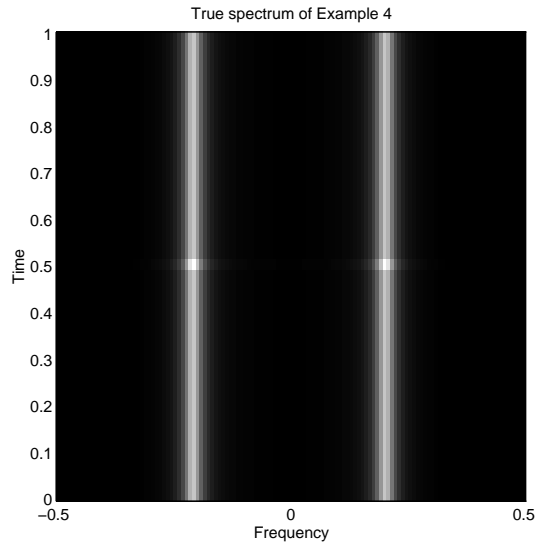


FIGURE 1. Example 4: true spectrum of AR(2)–process “transient with medium–size jump” ( $C = 1.65$ )

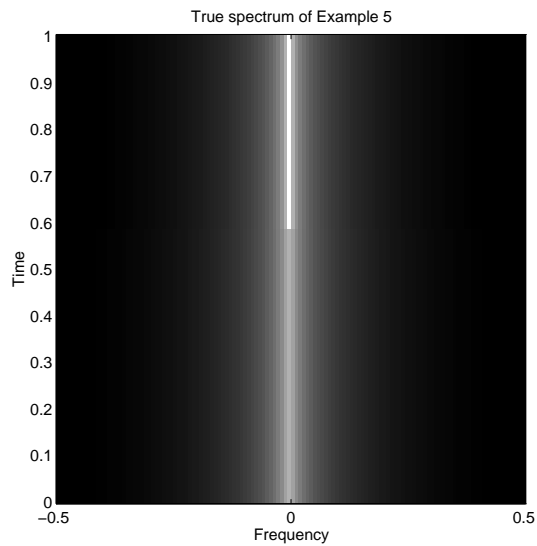


FIGURE 2. Example 5: true spectrum of piecewise constant AR(1)–process.

with 14 coefficients, and

$$I_{28} = \left\{ (j, k; j', k') \mid 0 \leq j \leq 2, 0 \leq k \leq 2^j - 1; j' = 2, 0 \leq k' \leq 2^{j'} - 1 \right\} ,$$

counting 28 coefficients, which is a sufficiently large subset of  $\mathcal{I}_T$  for this purpose here.

We get the following errors of second kind, again based on 1000 simulation runs, now with quantiles  $q_{14} = 2.69$  and  $q_{28} = 2.91$ , respectively:

For  $I_{14}$ ,  $e_1(14) = 0.203$ , and for  $I_{28}$ ,  $e_1(28) = 0.200$ , which is quite a bit of an improvement compared to  $e_1(7) = 0.543$  with the use of  $I_7$ .

In this case we also checked the errors of first kind  $e_0(14)$  and  $e_0(28)$  for the use of  $I_{14}$  and  $I_{28}$  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 got a bit high, compared to 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 is 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 the smaller the larger the range of integration in frequency is.

In another, final example, Example 6, yet, again with  $T = 1024$ , we simulated a piecewise stationary autoregressive process of order 2, now, with parameters  $a_1 = -0.60$  for  $t \leq 0.6 T$  and  $a_1 = -0.208$  for  $t > 0.6 T$ , and a constant  $a_2 = 0.36$ . This amounts to a sudden shift from 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  $I_{14}$  and  $I_{28}$ . 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  $\tilde{\alpha}_{j,k;j',k'}^{(1)}$  in  $I_7$  of similar size, and only integration over finer segments in frequency allows for significant differences.

As mentioned in the beginning of this section, it is also of considerable interest to compare these results with the performance of the test with the use of the pre-periodogram instead, i.e. with coefficients  $\tilde{\alpha}_{j,k;j',k'}^{(2)}$ . We repeated the experiments 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  $I_7$  (and to give a comparison with the performance on  $H_1$ , below) also  $e_0(14) = 0.124$  for  $I_{14}$ . 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 typical example for the performance on the alternative  $H_1$  we chose Example 4 with  $C = 1.65$ , where we got  $e_1(7) = 0.106$  and, also to use 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 we could observe already for

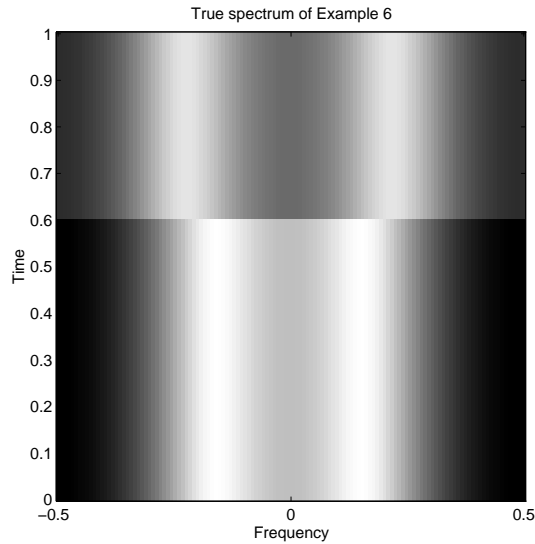


FIGURE 3. Example 6: true spectrum of piecewise constant AR(2)-process.

the segmented periodogram.

Summarizing, both test procedures do 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 advertise the use of the algorithmically much faster method based on the segmented periodogram. However, an estimator based on 2-d 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. Hence, we believe nonetheless 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 of incorporating a long range of lags, even for a locally changing spectrum in time. As a certain advantage, in any case, though, we consider the possibility of performing both estimation and testing simultaneously with the *same* nonparametric method. I.e., we can use the empirical wavelet coefficients of the very estimation method we have chosen to perform also the test for stationarity, and will possibly benefit for at least the estimation if we use the pre-periodogram.

### Acknowledgement

The authors thank the University of Kaiserslautern and the SFB 373 at Humboldt University for making mutual research visits possible.

## REFERENCES

- Black, F. (1976). Studies in Stock Price Volatility Changes. Proceedings on the 1976 Business and Economic Statistic Section, Am. Stat. Assoc., 177–181.
- Dahlhaus, R. (1995). Local likelihood methods for locally stationary time series. Manuscript.
- Dahlhaus, R. (1996). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. Athens Conf. On Applied Prob. and Time Series Anal., Vol. II (P. M. Robinson, M. Rosenblatt, eds.), *Lect. Notes Statist.*, **115**, 145–159. Springer-Verlag: New York.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1–37.
- Dümbgen, L. (1994). The asymptotic behaviour of some nonparametric change-point estimators. *Ann. Statist.* **22**, 1471–1495.
- Eubank, R. L., Hart, J. D., Simpson, D. G. and Stefanski, L. A. (1995). Testing for additivity in nonparametric regression. *Ann. Statist.* **23**, 1896–1920.
- Ghosh, B. K. and Huang, W.-M. (1991). The power and optimal kernel of the Bickel-Rosenblatt test for goodness of fit. *Ann. Statist.* **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.
- Neumann, M. H. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.* **25**, 38–76.
- Normolle, D. P. and Brown, M. B. (1994). Identification of aperiodic seasonality in non-Gaussian time series. *Biometrics* **50**, 798–812.
- Picard, D. (1985). Testing and estimating change-points in time series. *Adv. Appl. Prob.* **17**, 841–867.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*. Vol. 2, Academic Press, London.
- Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3**, 1–14.
- Rozenholc Y. (1995). Non-parametric tests of change-point with tapered data, Manuscript.
- von Sachs, R. and MacGibbon, B. (1997). Nonparametric curve estimation by wavelet thresholding with locally stationary errors. Technical Report “Berichte der AG Technomathematik” 179, Universität Kaiserslautern.
- von Sachs, R. and Schneider, K. (1996). Wavelet smoothing of evolutionary spectra by non-linear thresholding. *Appl. Comp. Harmonic Anal.* **3**, 268–282.
- Weiner, Ch. (1996). Logarithmic Stock Returns: leptokurtosis, heteroskedasticity and change-points. Preprint Institut de statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.