

## Testing for homogeneity of variance in the wavelet domain.

O. Kouamo<sup>1</sup>, E. Moulines<sup>2</sup>, and F. Roueff<sup>2</sup>

<sup>1</sup> ENSP, LIMSS, BP : 8390 Yaoundé, [olaf.kouamo@telecom-paristech.fr](mailto:olaf.kouamo@telecom-paristech.fr)

<sup>2</sup> Institut Télécom, Télécom ParisTech, CNRS UMR 5181, Paris,  
[eric.moulines,francois.roueff@telecom-paristech.fr](mailto:eric.moulines,francois.roueff@telecom-paristech.fr)

**Summary.** The danger of confusing long-range dependence with non-stationarity has been pointed out by many authors. Finding an answer to this difficult question is of importance to model time-series showing trend-like behavior, such as river runoff in hydrology, historical temperatures in the study of climates changes, or packet counts in network traffic engineering.

The main goal of this paper is to develop a test procedure to detect the presence of non-stationarity for a class of processes whose  $K$ -th order difference is stationary. Contrary to most of the proposed methods, the test procedure has the same distribution for short-range and long-range dependence covariance stationary processes, which means that this test is able to detect the presence of non-stationarity for processes showing long-range dependence or which are unit root.

The proposed test is formulated in the wavelet domain, where a change in the generalized spectral density results in a change in the variance of wavelet coefficients at one or several scales. Such tests have been already proposed in Whitcher et al. (2001), but these authors do not have taken into account the dependence of the wavelet coefficients within scales and between scales. Therefore, the asymptotic distribution of the test they have proposed was erroneous; as a consequence, the level of the test under the null hypothesis of stationarity was wrong.

In this contribution, we introduce two test procedures, both using an estimator of the variance of the scalogram at one or several scales. The asymptotic distribution of the test under the null is rigorously justified. The pointwise consistency of the test in the presence of a single jump in the general spectral density is also be presented.

A limited Monte-Carlo experiment is performed to illustrate our findings.

## 1.1 Introduction

For time series of short duration, stationarity and short-range dependence have usually been regarded to be approximately valid. However, such an assumption becomes questionable in the large data sets currently investigated in geophysics, hydrology or financial econometrics. There has been a long lasting controversy to decide whether the deviations to “short memory stationarity” should be attributed to long-range dependence or are related to the presence of breakpoints in the mean, the variance, the covariance function or other types of more sophisticated structural changes. The links between non-stationarity and long-range dependence (LRD) have been pointed out by many authors in the hydrology literature long ago: Klemes (1974) and Boes and Salas (1978) show that non-stationarity in the mean provides a possible explanations of the so-called Hurst phenomenon. Potter (1976) and later Rao and Yu (1986) suggested that more sophisticated changes may occur, and have proposed a method to detect such changes. The possible confusions between long-memory and some forms of nonstationarity have been discussed in the applied probability literature: Bhattacharya et al. (1983) show that long-range dependence may be confused with the presence of a small monotonic trend. This phenomenon has also been discussed in the econometrics literature. Hidalgo and Robinson (1996) proposed a test of presence of structural change in a long memory environment. Granger and Hyung (1999) showed that linear processes with breaks can mimic the autocovariance structure of a linear fractionally integrated long-memory process (a stationary process that encounters occasional regime switches will have some properties that are similar to those of a long-memory process). Similar behaviors are considered in Diebold and Inoue (2001) who provided simple and intuitive econometric models showing that long-memory and structural changes are easily confused. Mikosch and Stărică (2004) asserted that what had been seen by many authors as long memory in the volatility of the absolute values or the square of the log-returns might, in fact, be explained by abrupt changes in the parameters of an underlying GARCH-type models. Berkes et al. (2006) proposed a testing procedure for distinguishing between a weakly dependent time series with change-points in the mean and a long-range dependent time series. Hurvich et al. (2005) have

proposed a test procedure for detecting long memory in presence of deterministic trends.

The procedure described in this paper deals with the problem of detecting changes which may occur in the spectral content of a process. We will consider a process  $X$  which, before and after the change, is not necessary stationary but whose difference of at least a given order is stationary, so that polynomial trends up to that order can be discarded. Denote by  $\Delta X$  the first order difference of  $X$ ,

$$[\Delta X]_n \stackrel{\text{def}}{=} X_n - X_{n-1}, \quad n \in \mathbb{Z},$$

and define, for an integer  $K \geq 1$ , the  $K$ -th order difference recursively as follows:  $\Delta^K = \Delta \circ \Delta^{K-1}$ . A process  $X$  is said to be  $K$ -th order difference stationary if  $\Delta^K X$  is covariance stationary. Let  $f$  be a non-negative  $2\pi$ -periodic symmetric function such that there exists an integer  $K$  satisfying,  $\int_{-\pi}^{\pi} |1 - e^{-i\lambda}|^{2K} f(\lambda) d\lambda < \infty$ . We say that the process  $X$  admits *generalized spectral density*  $f$  if  $\Delta^K X$  is weakly stationary and with spectral density function

$$f_K(\lambda) = |1 - e^{-i\lambda}|^{2K} f(\lambda). \quad (1.1)$$

This class of process include both short-range dependent and long-range dependent processes, but also unit-root and fractional unit-root processes. The main goal of this paper is to develop a testing procedure for distinguishing between a  $K$ -th order stationary process and a non-stationary process.

In this paper, we consider the so-called *a posteriori* or *retrospective* method (see (Brodsky and Darkhovsky, 2000, Chapter 3)). The proposed test is formulated in the wavelet domain, where a change in the generalized spectral density results in a change in the variance of the wavelet coefficients. Our test is based on a CUSUM statistic, which is perhaps the most extensively used statistic for detecting and estimating change-points in mean. In our procedure, the CUSUM is applied to the partial sums of the squared wavelet coefficients at a given scale or on a specific range of scales. This procedure extends the test introduced in Inclan and Tiao (1994) to detect changes in the variance of an independent sequence of random variables. To describe the idea, suppose that, under the null hypothesis, the time series is  $K$ -th order difference stationary and that, under the alternative, there is one breakpoint where the generalized spectral density of the process changes. We consider the

scalogram in the range of scale  $J_1, J_1 + 1, \dots, J_2$ . Under the null hypothesis, there is no change in the variance of the wavelet coefficients at any given scale  $j \in \{J_1, \dots, J_2\}$ . Under the alternative, these variances takes different values before and after the change point. The amplitude of the change depends on the scale, and the change of the generalized spectral density. We consider the  $(J_2 - J_1 + 1)$ -dimensional W2-CUSUM statistic  $\{T_{J_1, J_2}(t), t \in [0, 1]\}$  defined by (1.41), which is a CUSUM-like statistics applied to the square of the wavelet coefficients. Using  $T_{J_1, J_2}(t)$  we can construct an estimator  $\hat{\tau}_{J_1, J_2}$  of the change point (no matter if a change-point exists or not), by minimizing an appropriate norm of the W2-CUSUM statistics,  $\hat{\tau}_{J_1, J_2} = \text{Argmin}_{t \in [0, 1]} \|T_{J_1, J_2}(t)\|_*$ . The statistic  $T_{J_1, J_2}(\hat{\tau}_{J_1, J_2})$  converges to a well-know distribution under the null hypothesis (see Theorems 1 and 2) but diverges to infinity under the alternative (Theorems 3 and 4). A similar idea has been proposed by Whitcher et al. (2001) but these authors did not take into account the dependence of wavelet coefficient, resulting in an erroneous normalization and asymptotic distributions.

The paper is organized as follows. In Section 1.2, we introduce the wavelet setting and the relationship between the generalized spectral density and the variance of wavelet coefficients at a given scale. In Section 1.3, our main assumptions are formulated and the asymptotic distribution of the W2-CUSUM statistics is presented first in the single scale (sub-section 1.3.1) and then in the multiple scales (sub-section 1.3.2) cases. In Section 1.4, several possible test procedures are described to detect the presence of changes at a single scale or simultaneously at several scales. In Section 1.6, finite sample performance of the test procedure is studied based on Monte-Carlo experiments.

## 1.2 The wavelet transform of $K$ -th order difference stationary processes

In this section, we introduce the wavelet setting, define the scalogram and explain how spectral change-points can be observed in the wavelet domain. The main advantage of using the wavelet domain is to alleviate problems arising when the time series exhibit is long range dependent. We will recall

some basic results obtained in Moulines et al. (2007) to support our claims. We refer the reader to that paper for the proofs of the stated results.

**The wavelet setting.** The wavelet setting involves two functions  $\phi$  and  $\psi$  and their Fourier transforms

$$\widehat{\phi}(\xi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \phi(t)e^{-i\xi t} dt \quad \text{and} \quad \widehat{\psi}(\xi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \psi(t)e^{-i\xi t} dt,$$

and assume the following:

- (W-1)  $\phi$  and  $\psi$  are compactly-supported, integrable, and  $\widehat{\phi}(0) = \int_{-\infty}^{\infty} \phi(t) dt = 1$  and  $\int_{-\infty}^{\infty} \psi^2(t) dt = 1$ .
- (W-2) There exists  $\alpha > 1$  such that  $\sup_{\xi \in \mathbb{R}} |\widehat{\psi}(\xi)| (1 + |\xi|)^\alpha < \infty$ .
- (W-3) The function  $\psi$  has  $M$  vanishing moments, *i.e.*  $\int_{-\infty}^{\infty} t^m \psi(t) dt = 0$  for all  $m = 0, \dots, M - 1$
- (W-4) The function  $\sum_{k \in \mathbb{Z}} k^m \phi(\cdot - k)$  is a polynomial of degree  $m$  for all  $m = 0, \dots, M - 1$ .

The fact that both  $\phi$  and  $\psi$  have finite support (Condition (W-1)) ensures that the corresponding filters (see (1.7)) have finite impulse responses (see (1.9)). While the support of the Fourier transform of  $\psi$  is the whole real line, Condition (W-2) ensures that this Fourier transform decreases quickly to zero. Condition (W-3) is an important characteristic of wavelets: it ensures that they oscillate and that their scalar product with continuous-time polynomials up to degree  $M - 1$  vanishes. Daubechies wavelets and Coiflets having at least two vanishing moments satisfy these conditions.

Viewing the wavelet  $\psi(t)$  as a basic template, define the family  $\{\psi_{j,k}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$  of translated and dilated functions

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k), \quad j \in \mathbb{Z}, k \in \mathbb{Z}. \quad (1.2)$$

Positive values of  $k$  translate  $\psi$  to the right, negative values to the left. The *scale index*  $j$  dilates  $\psi$  so that large values of  $j$  correspond to coarse scales and hence to low frequencies.

Assumptions (W-1)-(W-4) are standard in the context of a multiresolution analysis (MRA) in which case,  $\phi$  is the scaling function and  $\psi$  is the associated wavelet, see for instance Mallat (1998); Cohen (2003). Daubechies wavelets and Coiflets are examples of orthogonal wavelets constructed using an MRA. In this paper, we do not assume the wavelets to be orthonormal nor that

they are associated to a multiresolution analysis. We may therefore work with other convenient choices for  $\phi$  and  $\psi$  as long as (W-1)-(W-4) are satisfied.

**Discrete Wavelet Transform (DWT) in discrete time.** We now describe how the wavelet coefficients are defined in discrete time, that is for a real-valued sequence  $\{x_k, k \in \mathbb{Z}\}$  and for a finite sample  $\{x_k, k = 1, \dots, n\}$ . Using the scaling function  $\phi$ , we first interpolate these discrete values to construct the following continuous-time functions

$$\mathbf{x}_n(t) \stackrel{\text{def}}{=} \sum_{k=1}^n x_k \phi(t-k) \quad \text{and} \quad \mathbf{x}(t) \stackrel{\text{def}}{=} \sum_{k \in \mathbb{Z}} x_k \phi(t-k), \quad t \in \mathbb{R}. \quad (1.3)$$

Without loss of generality we may suppose that the support of the scaling function  $\phi$  is included in  $[-T, 0]$  for some integer  $T \geq 1$ . Then

$$\mathbf{x}_n(t) = \mathbf{x}(t) \quad \text{for all } t \in [0, n - T + 1].$$

We may also suppose that the support of the wavelet function  $\psi$  is included in  $[0, T]$ . With these conventions, the support of  $\psi_{j,k}$  is included in the interval  $[2^j k, 2^j(k+T)]$ . Let  $\tau_0$  be an arbitrary shift order. The wavelet coefficient  $W_{j,k}^{\mathbf{x}}$  at scale  $j \geq 0$  and location  $k \in \mathbb{Z}$  is formally defined as the scalar product in  $L^2(\mathbb{R})$  of the function  $t \mapsto \mathbf{x}(t)$  and the wavelet  $t \mapsto \psi_{j,k}(t)$ :

$$W_{j,k}^{\mathbf{x}} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \mathbf{x}(t) \psi_{j,k}(t) dt = \int_{-\infty}^{\infty} \mathbf{x}_n(t) \psi_{j,k}(t) dt, \quad j \geq 0, k \in \mathbb{Z}, \quad (1.4)$$

when  $[2^j k, 2^j(k+T)] \subseteq [0, n - T + 1]$ , that is, for all  $(j, k) \in \mathcal{I}_n$ , where  $\mathcal{I}_n = \{(j, k) : j \geq 0, 0 \leq k < n_j\}$  with  $n_j = 2^{-j}(n - T + 1) - T + 1$ . It is important to observe that the definition of the wavelet coefficient  $W_{j,k}$  at a given index  $(j, k)$  does not depend on the sample size  $n$  (this is in sharp contrast with Fourier coefficients). For ease of presentation, we will use the convention that at each scale  $j$ , the first available wavelet coefficient  $W_{j,k}$  is indexed by  $k = 0$ , that is,

$$\mathcal{I}_n \stackrel{\text{def}}{=} \{(j, k) : j \geq 0, 1 \leq k \leq n_j\} \quad \text{with} \quad n_j = 2^{-j}(n - T + 1) - T + 1. \quad (1.5)$$

**Practical implementation.** In practice the DWT of  $\{x_k, k = 1, \dots, n\}$  is not computed using (1.4) but by linear filtering and decimation. Indeed the wavelet coefficient  $W_{j,k}^{\mathbf{x}}$  can be expressed as

$$W_{j,k}^{\mathbf{x}} = \sum_{l \in \mathbb{Z}} x_l h_{j, 2^j k - l}, \quad (j, k) \in \mathcal{I}_n, \quad (1.6)$$

where

$$h_{j,l} \stackrel{\text{def}}{=} 2^{-j/2} \int_{-\infty}^{\infty} \phi(t+l)\psi(2^{-j}t) dt. \quad (1.7)$$

For all  $j \geq 0$ , the discrete Fourier transform of the transfer function  $\{h_{j,l}\}_{l \in \mathbb{Z}}$  is

$$H_j(\lambda) \stackrel{\text{def}}{=} \sum_{l \in \mathbb{Z}} h_{j,l} e^{-i\lambda l} = 2^{-j/2} \int_{-\infty}^{\infty} \sum_{l \in \mathbb{Z}} \phi(t+l) e^{-i\lambda l} \psi(2^{-j}t) dt. \quad (1.8)$$

Since  $\phi$  and  $\psi$  have compact support, the sum in (1.8) has only a finite number of non-vanishing terms and,  $H_j(\lambda)$  is the transfer function of a finite impulse response filter,

$$H_j(\lambda) = \sum_{l=-T(2^j+1)+1}^{-1} h_{j,l} e^{-i\lambda l}. \quad (1.9)$$

When  $\phi$  and  $\psi$  are the scaling and the wavelet functions associated to a MRA, the wavelet coefficients may be obtained recursively by applying a finite order filter and downsampling by an order 2. This recursive procedure is referred to as *the pyramidal algorithm*, see for instance Mallat (1998).

**The wavelet spectrum and the scalogram.** Let  $X = \{X_t, t \in \mathbb{Z}\}$  be a real-valued process with wavelet coefficients  $\{W_{j,k}, k \in \mathbb{Z}\}$  and define

$$\sigma_{j,k}^2 = \text{Var}(W_{j,k}).$$

If  $\Delta^M X$  is stationary, by Eq (16) in Moulines et al. (2007), we have that, for all  $j$ , the process of its wavelet coefficients at scale  $j$ ,  $\{W_{j,k}, k \in \mathbb{Z}\}$ , is also stationary. Then, the wavelet variance  $\sigma_{j,k}^2$  does not depend on  $k$ ,  $\sigma_{j,k}^2 = \sigma_j^2$ . The sequence  $(\sigma_j^2)_{j \geq 0}$  is called the *wavelet spectrum* of the process  $X$ .

If moreover  $\Delta^M X$  is centered, the wavelet spectrum can be estimated by using the scalogram, defined as the empirical mean of the squared wavelet coefficients computed from the sample  $X_1, \dots, X_n$ :

$$\hat{\sigma}_j^2 = \frac{1}{n_j} \sum_{k=1}^{n_j} W_{j,k}^2.$$

By (Moulines et al., 2007, Proposition 1), if  $K \leq M$ , then the scalogram of  $X$  can be expressed using the generalized spectral density  $f$  appearing in (1.1) and the filters  $H_j$  defining the DWT in (1.8) as follows:

$$\sigma_j^2 = \int_{-\pi}^{\pi} |H_j(\lambda)|^2 f(\lambda) d\lambda, \quad j \geq 0. \quad (1.10)$$

### 1.3 Asymptotic distribution of the W2-CUSUM statistics

#### 1.3.1 The single-scale case

To start with simple presentation and statement of results, we first focus in this section on a test procedure aimed at detecting a change in the variance of the wavelet coefficients at a single scale  $j$ . Let  $X_1, \dots, X_n$  be the  $n$  observations of a time series, and denote by  $W_{j,k}$  for  $(j, k) \in \mathcal{I}_n$  with  $\mathcal{I}_n$  defined in (1.5) the associated wavelet coefficients. In view of (1.10), if  $X_1, \dots, X_n$  are a  $n$  successive observations of a  $K$ -th order difference stationary process, then the wavelet variance at each given scale  $j$  should be constant. If the process  $X$  is not  $K$ -th order stationary, then it can be expected that the wavelet variance will change either gradually or abruptly (if there is a shock in the original time-series). This thus suggests to investigate the consistency of the variance of the wavelet coefficients.

There are many works aimed at detecting the change point in the variance of a sequence of independent random variables; such problem has also been considered, but much less frequently, for sequences of dependent variables. Here, under the null assumption of  $K$ -th order difference stationarity, the wavelet coefficients  $\{W_{j,k}, k \in \mathbb{Z}\}$  is a covariance stationary sequence whose spectral density is given by (see (Moulines et al., 2007, Corollary 1))

$$\mathbf{D}_{j,0}(\lambda; f) \stackrel{\text{def}}{=} \sum_{l=0}^{2^j-1} f(2^{-j}(\lambda + 2l\pi)) 2^{-j} |H_j(2^{-j}(\lambda + 2l\pi))|^2. \quad (1.11)$$

We will adapt the approach developed in Inlan and Tiao (1994), which uses cumulative sum (CUSUM) of squares to detect change points in the variance.

In order to define the test statistic, we first introduce a change point estimator for the mean of the square of the wavelet coefficients at each scale  $j$ .

$$\hat{k}_j = \operatorname{argmax}_{1 \leq k \leq n_j} \left| \sum_{1 \leq i \leq k} W_{j,i}^2 - \frac{k}{n_j} \sum_{1 \leq i \leq n_j} W_{j,i}^2 \right|. \quad (1.12)$$

Using this change point estimator, the W2-CUSUM statistics is defined as

$$T_{n_j} = \frac{1}{n_j^{1/2} s_{j,n_j}} \left| \sum_{1 \leq i \leq \hat{k}_j} W_{j,i}^2 - \frac{\hat{k}_j}{n_j} \sum_{1 \leq i \leq n_j} W_{j,i}^2 \right|, \quad (1.13)$$



where  $s_{j,n_j}^2$  is a suitable estimator of the variance of the sample mean of the  $W_{j,i}^2$ . Because wavelet coefficients at a given scale are correlated, we use the Bartlett estimator of the variance, which is defined by

$$s_{j,n_j}^2 = \widehat{\gamma}_j(0) + 2 \sum_{1 \leq l \leq q(n_j)} w_l(q(n_j)) \widehat{\gamma}_j(l), \quad (1.14)$$

where

$$\widehat{\gamma}_j(l) \stackrel{\text{def}}{=} \frac{1}{n_j} \sum_{1 \leq i \leq n_j-l} (W_{j,i}^2 - \widehat{\sigma}_j^2)(W_{j,i+l}^2 - \widehat{\sigma}_j^2), \quad (1.15)$$

are the sample autocovariance of  $\{W_{j,i}^2, i = 1, \dots, n_j\}$ ,  $\widehat{\sigma}_j^2$  is the scalogram and, for a given integer  $q$ ,

$$w_l(q) = 1 - \frac{l}{1+q}, l \in \{0, \dots, q\} \quad (1.16)$$

are the so-called Bartlett weights.

The test differs from statistics proposed in Inclan and Tiao (1994) only in its denominator, which is the square root of a consistent estimator of the partial sum's variance. If  $\{X_n\}$  is short-range dependent, the variance of the partial sum of the scalograms is not simply the sum of the variances of the individual square wavelet coefficient, but also includes the autocovariances of these terms. Therefore, the estimator of the averaged scalogram variance involves not only sums of squared deviations of the scalogram coefficients, but also its weighted autocovariances up to lag  $q(n_j)$ . The weights  $\{w_l(q(n_j))\}$  are those suggested by Newey and West (1987) and always yield a positive sequence of autocovariance, and a positive estimator of the (unnormalized) wavelet spectrum at scale  $j$ , at frequency zero using a Bartlett window. We will first established the consistency of the estimator  $s_{j,n_j}^2$  of the variance of the scalogram at scale  $j$  and the convergence of the empirical process of the square wavelet coefficients to the Brownian motion. Denote by  $D([0, 1])$  is the Skorokhod space of functions which are right continuous at each point of  $[0, 1)$  with left limit of  $(0, 1]$  (or *cadlag* functions). This space is, in the sequel, equipped with the classical Skorokhod metric.

**Theorem 1.** *Suppose that  $X$  is a Gaussian process with generalized spectral density  $f$ . Let  $(\phi, \psi)$  be a scaling and a wavelet function satisfying (W-1)-(W-4). Let  $\{q(n_j)\}$  be a non decreasing sequence of integers satisfying*

$$q(n_j) \rightarrow \infty \quad \text{and} \quad q(n_j)/n_j \rightarrow 0 \quad \text{as} \quad n_j \rightarrow \infty. \quad (1.17)$$

Assume that  $\Delta^M X$  is non-deterministic and centered, and that  $\lambda^{2M} f(\lambda)$  is two times differentiable in  $\lambda$  with bounded second order derivative. Then for any fixed scale  $j$ , as  $n \rightarrow \infty$ ,

$$s_{j,n_j}^2 \xrightarrow{P} \frac{1}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda, \quad (1.18)$$

where  $\mathbf{D}_{j,0}(\lambda; f)$  is the wavelet coefficients spectral density at scale  $j$  see (1.11). Moreover, defining  $\sigma_j^2$  by (1.10),

$$\frac{1}{n_j^{1/2} s_{j,n_j}} \sum_{i=1}^{\lfloor n_j t \rfloor} (W_{j,i}^2 - \sigma_j^2) \xrightarrow{\mathcal{L}} B(t) \quad \text{in} \quad D([0,1]), \quad \text{as} \quad n \rightarrow \infty \quad (1.19)$$

where  $(B(t), t \in [0,1])$  is the standard Brownian motion.

*Remark 1.* The fact that  $X$  is Gaussian can be replaced by the more general assumption that the process  $X$  is linear in the strong sense, under appropriate moment conditions on the innovation. The proofs are then more involved, especially to establish the invariance principle which is pivotal in our derivation.

*Remark 2.* By allowing  $q(n_j)$  to increase but at a slower rate than the number of observations, the estimator of the averaged scalogram variance adjusts appropriately for general forms of short-range dependence among the scalogram coefficients. Of course, although the condition (1.17) ensure the consistency of  $s_{j,n_j}^2$ , they provide little guidance in selecting a truncation lag  $q(n_j)$ . When  $q(n_j)$  becomes large relative to the sample size  $n_j$ , the finite-sample distribution of the test statistic might be far from its asymptotic limit. However  $q(n_j)$  cannot be chosen too small since the autocovariances beyond lag  $q(n_j)$  may be significant and should be included in the weighted sum. Therefore, the truncation lag must be chosen ideally using some data-driven procedures. Andrews (1991) and Newey and West (1994) provide a data-dependent rule for choosing  $q(n_j)$ . These contributions suggest that selection of bandwidth according to an asymptotically optimal procedure tends to lead to more accurately sized test statistics than do traditional procedure. The methods suggested by Andrews (1991) for selecting the bandwidth optimally is a plug-in approach. This procedure require the researcher to fit an ARMA model of

given order to provide a rough estimator of the spectral density and of its derivatives at zero frequencies (although misspecification of the order affects only optimality but not consistency). The minimax optimality of this method is based on an asymptotic mean-squared error criterion and its behavior in the finite sample case is not precisely known. The procedure outlined in Newey and West (1994) suggests to bypass the modeling step, by using instead a pilot truncated kernel estimates of the spectral density and its derivative. We use these data driven procedures in the Monte Carlo experiments (these procedures have been implemented in the *R-package sandwich*).

*Proof.* Since  $X$  is Gaussian and  $\mathbf{\Delta}^M X$  is centered, Eq. (17) in Moulines et al. (2007) implies that  $\{W_{j,k}, k \in \mathbb{Z}\}$  is a centered Gaussian process, whose distribution is determined by

$$\gamma_j(h) = \text{Cov}(W_{j,0}, W_{j,h}) = \int_{-\pi}^{\pi} \mathbf{D}_{j,0}(\lambda; f) e^{-i\lambda h} d\lambda .$$

From Corollary 1 and equation (16) in Moulines et al. (2007), we have

$$\begin{aligned} & \mathbf{D}_{j,0}(\lambda; f) \\ &= \sum_{l=0}^{2^j-1} f(2^{-j}(\lambda + 2l\pi)) 2^{-j} \left| \tilde{H}_j(2^{-j}(\lambda + 2l\pi)) \right|^2 \left| 1 - e^{-i2^{-j}(\lambda + 2l\pi)} \right|^{2M} , \end{aligned}$$

where  $\tilde{H}_j$  is a trigonometric polynomial. Using that

$$|1 - e^{-i\xi}|^{2M} = |\xi|^{2M} \left| \frac{1 - e^{-i\xi}}{i\xi} \right|^{2M}$$

and that  $|\xi|^{2M} f(\xi)$  has a bounded second order derivative, we get that  $\mathbf{D}_{j,0}(\lambda; f)$  has also a bounded second order derivative. In particular,

$$\int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda < \infty \quad \text{and} \quad \sum_{s \in \mathbb{Z}} |\gamma_j(s)| < \infty . \quad (1.20)$$

The proof may be decomposed into 3 steps. We first prove the consistency of the Bartlett estimator of the variance of the squares of wavelet coefficients  $s_{j,n_j}^2$ , that is (1.18). Then we determine the asymptotic normality of the finite-dimensional distributions of the empirical scalogram, suitably centered and normalized. Finally a tightness criterion is proved, to establish the convergence

in the Skorokhod space. Combining these three steps completes the proof of (1.19).

**Step 1.** Observe that, by the Gaussian property,  $\text{Cov}(W_{j,0}^2, W_{j,h}^2) = 2\gamma_j^2(h)$ . Using Theorem 3-i in Giraitis et al. (2003), the limit (1.18) follows from

$$2 \sum_{h=-\infty}^{+\infty} \gamma_j^2(h) = \frac{1}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda < \infty, \quad (1.21)$$

and

$$\sup_{h \in \mathbb{Z}} \sum_{r,s=-\infty}^{+\infty} |\mathcal{K}(h, r, s)| < \infty. \quad (1.22)$$

where

$$\mathcal{K}(h, r, s) = \text{Cum}(W_{j,k}^2, W_{j,k+h}^2, W_{j,k+r}^2, W_{j,k+s}^2). \quad (1.23)$$

Equation (1.21) follows from Parseval's equality and (1.20). Let us now prove (1.22). Using that the wavelet coefficients are Gaussian, we obtain

$$\begin{aligned} \mathcal{K}(h, r, s) &= 12 \{ \gamma_j(h)\gamma_j(r-s)\gamma_j(h-r)\gamma_j(s) \\ &\quad + \gamma_j(h)\gamma_j(r-s)\gamma_j(h-s)\gamma_j(r) + \gamma_j(s-h)\gamma_j(r-h)\gamma_j(r)\gamma_j(s) \}. \end{aligned}$$

The bound of the last term is given by

$$\sup_{h \in \mathbb{Z}} \sum_{r,s=-\infty}^{+\infty} |\gamma_j(s-h)\gamma_j(r-h)\gamma_j(r)\gamma_j(s)| \leq \sup_h \left( \sum_{r=-\infty}^{+\infty} |\gamma_j(r)\gamma_j(r-h)| \right)^2$$

which is finite by the Cauchy-Schwarz inequality, since  $\sum_{r \in \mathbb{Z}} \gamma_j^2(r) < \infty$ .

Using  $|\gamma_j(h)| < \gamma_j(0)$  and the Cauchy-Schwarz inequality, we have

$$\sup_{h \in \mathbb{Z}} \sum_{r,s=-\infty}^{+\infty} |\gamma_j(h)\gamma_j(r-s)\gamma_j(h-r)\gamma_j(s)| \leq \gamma_j(0) \sum_{u \in \mathbb{Z}} \gamma_j^2(u) \sum_{s \in \mathbb{Z}} |\gamma_j(s)|,$$

and the same bound applies to

$$\sup_{h \in \mathbb{Z}} \sum_{r,s=-\infty}^{+\infty} |\gamma_j(h)\gamma_j(r-s)\gamma_j(h-s)\gamma_j(r)|.$$

Hence, we have (1.22) by (1.20), which achieves the proof of Step 1.

**Step 2.** Let us define

$$S_{n_j}(t) = \frac{1}{\sqrt{n_j}} \sum_{i=1}^{\lfloor n_j t \rfloor} (W_{j,i}^2 - \sigma_j^2), \quad (1.24)$$

where  $\sigma_j^2 = \mathbb{E}(W_{j,i}^2)$ , and  $\lfloor x \rfloor$  is the entire part of  $x$ . Step 2 consists in proving that for  $0 \leq t_1 \leq \dots \leq t_k \leq 1$ , and  $\mu_1, \dots, \mu_k \in \mathbb{R}$ ,

$$\sum_{i=1}^k \mu_i S_{n_j}(t_i) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \frac{1}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda \times \text{Var} \left( \sum_{i=1}^k \mu_i B(t_i) \right) \right). \quad (1.25)$$

Observe that

$$\begin{aligned} \sum_{i=1}^k \mu_i S_{n_j}(t_i) &= \frac{1}{\sqrt{n_j}} \sum_{i=1}^k \mu_i \sum_{l=1}^{n_j} (W_{j,l}^2 - \sigma_j^2) \mathbb{1}_{\{l \leq \lfloor n_j t_i \rfloor\}} \\ &= \sum_{l=1}^{n_j} W_{j,l}^2 a_{l,n} - E \left( \sum_{l=1}^{n_j} W_{j,l}^2 a_{l,n} \right) \\ &= \xi_{n_j}^T A_{n_j} \xi_{n_j}, \end{aligned}$$

where we set  $a_{l,n} = \frac{1}{\sqrt{n_j}} \sum_{i=1}^k \mu_i \mathbb{1}_{\{l \leq \lfloor n_j t_i \rfloor\}}$ ,  $\xi_{n_j} = (W_{j,1}, \dots, W_{j,n_j})^T$  and  $A_{n_j}$  is the diagonal matrix with diagonal entries  $(a_{1,n_j}, \dots, a_{n_j,n_j})$ . Applying (Moulines et al., 2008, Lemma 12), (1.25) is obtained by proving that, as  $n_j \rightarrow \infty$ ,

$$\rho(A_{n_j}) \rho(\Gamma_{n_j}) \rightarrow 0 \quad (1.26)$$

$$\text{Var} \left( \sum_{i=1}^k \mu_i S_{n_j}(t_i) \right) \rightarrow \frac{1}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda \times \text{Var} \left( \sum_{i=1}^k \mu_i (B(t_i)) \right), \quad (1.27)$$

where  $\rho(A)$  denote the spectral radius of the matrix  $A$ , that is, the maximum modulus of its eigenvalues and  $\Gamma_{n_j}$  is the covariance matrix of  $\xi_{n_j}$ . The process  $(W_{j,i})_{\{i=1, \dots, n_j\}}$  is stationary with spectral density  $\mathbf{D}_{j,0}(\cdot; f)$ . Thus, by Lemma 2 in Moulines et al. (2007) its covariance matrix  $\Gamma_{n_j}$  satisfies  $\rho(\Gamma_{n_j}) \leq 2\pi \sup_{\lambda} \mathbf{D}_{j,0}(\lambda; f)$ . Furthermore, as  $n_j \rightarrow \infty$ ,

$$\rho(A_{n_j}) = \max_{1 \leq l \leq n_j} \frac{1}{\sqrt{n_j}} \left| \sum_{i=1}^k \mu_i \mathbb{1}_{\{l \leq \lfloor n_j t_i \rfloor\}} \right| \leq n_j^{-1/2} \sum_{i=1}^k |\mu_i| \rightarrow 0,$$

and (1.26) holds. We now prove (1.27). Using that  $B(t)$  has variance  $t$  and independent and stationary increments, and that these properties characterize its covariance function, it is sufficient to show that, for all  $t \in [0, 1]$ , as  $n_j \rightarrow \infty$ ,

$$\text{Var} (S_{n_j}(t)) \rightarrow t \int_{-\pi}^{\pi} |\mathbf{D}_{j,0}(\lambda; f)|^2 d\lambda, \quad (1.28)$$

and for all  $0 \leq r \leq s \leq t \leq 1$ , as  $n_j \rightarrow \infty$ ,

$$\text{Cov}(S_{n_j}(t) - S_{n_j}(s), S_{n_j}(r)) \rightarrow 0. \quad (1.29)$$

For any sets  $A, B \subseteq [0, 1]$ , we set

$$V_{n_j}(\tau, A, B) = \frac{1}{n_j} \sum_{k \geq 1} \mathbb{1}_A((k + \tau)/n_j) \mathbb{1}_B(k/n_j).$$

For all  $0 \leq s, t \leq 1$ , we have

$$\begin{aligned} \text{Cov}(S_{n_j}(t), S_{n_j}(s)) &= \frac{1}{n_j} \sum_{i=1}^{\lfloor n_j t \rfloor} \sum_{k=1}^{\lfloor n_j s \rfloor} \text{Cov}(W_{j,i}^2, W_{j,k}^2) \\ &= 2 \sum_{\tau \in \mathbb{Z}} \gamma_j^2(\tau) V_{n_j}(\tau, ]0, t], ]0, s]). \end{aligned}$$

The previous display applies to the left-hand side of (1.28) when  $s = t$  and for  $0 \leq r \leq s \leq t \leq 1$ , it yields

$$\text{Cov}(S_{n_j}(t) - S_{n_j}(s), S_{n_j}(r)) = 2 \sum_{\tau \in \mathbb{Z}} \gamma_j^2(\tau) V_{n_j}(\tau, ]s, t], ]0, r]).$$

Observe that for all  $A, B \subseteq [0, 1]$ ,  $\sup_{\tau} |V_n(j, \tau, A, B)| \leq \frac{k}{n_j} \leq 1$ . Hence, by dominated convergence, the limits in (1.28) and (1.29) are obtained by computing the limits of  $V_n(j, \tau, ]0, t], ]0, t])$  and  $V_n(j, \tau, ]s, t], ]0, r])$  respectively. We have for any  $\tau \in \mathbb{Z}$ ,  $t > 0$ , and  $n_j$  large enough,

$$\sum_{k \geq 1} \mathbb{1}_{\{\frac{k+\tau}{n_j} \in ]0, t]\}} \mathbb{1}_{\{\frac{k}{n_j} \in ]0, t]\}} = \{(n_j t \wedge n_j t - \tau)\}_+ = n_j t - \tau_+.$$

Hence, as  $n_j \rightarrow \infty$ ,  $V_{n_j}(\tau, ]0, t], ]0, t]) \rightarrow t$  and, by (1.21), (1.28) follows. We have for any  $\tau \in \mathbb{Z}$  and  $0 < r \leq s \leq t$ ,

$$\begin{aligned} \sum_{k \geq 1} \mathbb{1}_{\{\frac{k+\tau}{n_j} \in ]s, t]\}} \mathbb{1}_{\{\frac{k}{n_j} \in ]0, r]\}} &= \{(n_j r \wedge \{n_j t - \tau\}) - (0 \vee \{n_j s - \tau\})\}_+ \\ &= (n_j r - n_j s + \tau)_+ \rightarrow \mathbb{1}_{\{r=s\}} \tau_+, \end{aligned}$$

where the last equality holds for  $n_j$  large enough and the limit as  $n_j \rightarrow \infty$ .

Hence  $V_{n_j}(\tau, ]s, t], ]0, r]) \rightarrow 0$  and (1.29) follows, which achieves Step 2.

**Step 3.** We now prove the tightness of  $\{S_{n_j}(t), t \in [0, 1]\}$  in the Skorokhod metric space. By Theorem 13.5 in Billingsley (1999), it is sufficient to prove that for all  $0 \leq r \leq s \leq t$ ,

$$\mathbb{E}[|S_{n_j}(s) - S_{n_j}(r)|^2 |S_{n_j}(t) - S_{n_j}(s)|^2] \leq C|t - r|^2,$$

where  $C > 0$  is some constant independent of  $r, s, t$  and  $n_j$ . We shall prove that, for all  $0 \leq r \leq t$ ,

$$\mathbb{E}[|S_{n_j}(t) - S_{n_j}(r)|^4] \leq C_1 \{n_j^{-1}(\lfloor n_j t \rfloor - \lfloor n_j r \rfloor)\}^2. \quad (1.30)$$

By the Cauchy-Schwarz inequality, and using that, for  $0 \leq r \leq s \leq t$ ,

$$n_j^{-1}(\lfloor n_j t \rfloor - \lfloor n_j s \rfloor) \times n_j^{-1}(\lfloor n_j s \rfloor - \lfloor n_j r \rfloor) \leq 4(t - r)^2,$$

the criterion (1.30) implies the previous criterion. Hence the tightness follows from (1.30), that we now prove. We have, for any  $\mathbf{i} = (i_1, \dots, i_4)$ ,

$$\begin{aligned} \mathbb{E} \left[ \prod_{k=1}^4 (W_{j,i_k}^2 - \sigma_j^2) \right] &= \text{Cum}(W_{j,i_1}^2, \dots, W_{j,i_4}^2) + \mathbb{E}(W_{j,i_1}^2, W_{j,i_2}^2) \mathbb{E}[W_{j,i_3}^2, W_{j,i_4}^2] \\ &\quad + \mathbb{E}[W_{j,i_1}^2, W_{j,i_3}^2] \mathbb{E}[W_{j,i_2}^2, W_{j,i_4}^2] + \mathbb{E}[W_{j,i_1}^2, W_{j,i_4}^2] \mathbb{E}[W_{j,i_2}^2, W_{j,i_3}^2]. \end{aligned}$$

It follows that, denoting for  $0 \leq r \leq t \leq 1$ ,

$$\begin{aligned} \mathbb{E} \left[ |S_{n_j}(t) - S_{n_j}(r)|^4 \right] &= \frac{1}{n_j^2} \sum_{\mathbf{i} \in A_{r,t}^4} \text{Cum}(W_{j,i_1}^2, \dots, W_{j,i_4}^2) \\ &\quad + \frac{3}{n_j^2} \left( \sum_{\mathbf{i} \in A_{r,t}^2} \mathbb{E}[W_{j,i_1}^2, W_{j,i_2}^2] \right)^2 \end{aligned}$$

where  $A_{r,t} = \{\lfloor n_j r \rfloor + 1, \dots, \lfloor n_j t \rfloor\}$ . Observe that

$$0 \leq \frac{1}{n_j} \sum_{\mathbf{i} \in A_{r,t}^2} \mathbb{E}[W_{j,i_1}^2, W_{j,i_2}^2] \leq 2 \sum_{\tau \in \mathbb{Z}} \gamma_j^2(\tau) \times n_j^{-1}(\lfloor n_j t \rfloor - \lfloor n_j r \rfloor).$$

Using that, by (1.23),  $\text{Cum}(W_{j,i_1}^2, \dots, W_{j,i_4}^2) = \mathcal{K}(i_2 - i_1, i_3 - i_1, i_4 - i_1)$ , we have

$$\begin{aligned} \sum_{\mathbf{i} \in A_{r,t}^4} |\text{Cum}(W_{j,i_1}^2, \dots, W_{j,i_4}^2)| &\leq (\lfloor n_j t \rfloor - \lfloor n_j r \rfloor) \sum_{h,s,l=\lfloor n_j r \rfloor - \lfloor n_j t \rfloor + 1}^{\lfloor n_j t \rfloor - \lfloor n_j r \rfloor - 1} |\mathcal{K}(h, s, l)| \\ &\leq 2(\lfloor n_j t \rfloor - \lfloor n_j r \rfloor)^2 \sup_{h \in \mathbb{Z}} \sum_{r,s=-\infty}^{+\infty} |\mathcal{K}(h, r, s)|. \end{aligned}$$

The last three displays and (1.22) imply (1.30), which proves the tightness.

Finally, observing that the variance (1.21) is positive, unless  $f$  vanishes almost everywhere, the convergence (1.19) follows from Slutsky's lemma and the three previous steps.

### 1.3.2 The multiple-scale case

The results above can be extended to test simultaneously changes in wavelet variances occurring simultaneously at multiple time-scales. To construct a multiple scale test, consider the *between-scale* process

$$\{[W_{j,k}^X, \mathbf{W}_{j,k}^X(j-j')^T]^T\}_{k \in \mathbb{Z}}, \quad (1.31)$$

where the superscript  $T$  denotes the transpose and  $\mathbf{W}_{j,k}^X(u)$ ,  $u = 0, 1, \dots, j$ , is defined as follows:

$$\mathbf{W}_{j,k}^X(u) \stackrel{\text{def}}{=} [W_{j-u, 2^u k}^X, W_{j-u, 2^u k+1}^X, \dots, W_{j-u, 2^u k+2^u-1}^X]^T. \quad (1.32)$$

It is a  $2^u$ -dimensional vector of wavelet coefficients at scale  $j' = j - u$  and involves all possible translations of the position index  $2^u k$  by  $v = 0, 1, \dots, 2^u - 1$ . The index  $u$  in (1.32) denotes the scale difference  $j - j' \geq 0$  between the finest scale  $j'$  and the coarsest scale  $j$ . Observe that  $\mathbf{W}_{j,k}^X(0)$  ( $u = 0$ ) is the scalar  $W_{j,k}^X$ . It is shown in (Moulines et al., 2007, Corollary 1) that, when  $\Delta^M X$  is covariance stationary, the between scale process  $\{[W_{j,k}^X, \mathbf{W}_{j,k}^X(j-j')^T]^T\}_{k \in \mathbb{Z}}$  is also covariance stationary. Moreover, for all  $0 \leq u \leq j$ , the *between scale covariance matrix* is defined as

$$\text{Cov}(W_{j,0}^X, \mathbf{W}_{j,k}^X(u)) = \int_{-\pi}^{\pi} e^{i\lambda k} \mathbf{D}_{j,u}(\lambda; f) d\lambda, \quad (1.33)$$

where  $\mathbf{D}_{j,u}(\lambda; f)$  is the cross-spectral density function of the between-scale process given by (see (Moulines et al., 2007, Corollary 1))

$$\begin{aligned} \mathbf{D}_{j,u}(\lambda; f) &\stackrel{\text{def}}{=} \sum_{l=0}^{2^j-1} \mathbf{e}_u(\lambda + 2l\pi) f(2^{-j}(\lambda + 2l\pi)) 2^{-j/2} H_j(2^{-j}(\lambda + 2l\pi)) \\ &\quad \times 2^{-(j-u)/2} \overline{H_{j-u}(2^{-j}(\lambda + 2l\pi))}, \end{aligned} \quad (1.34)$$

where for all  $\xi \in \mathbb{R}$ ,

$$\mathbf{e}_u(\xi) \stackrel{\text{def}}{=} 2^{-u/2} [1, e^{-i2^{-u}\xi}, \dots, e^{-i(2^u-1)2^{-u}\xi}]^T.$$

The case  $u = 0$  corresponds to the spectral density of the *within-scale* process  $\{W_{j,k}\}_{k \in \mathbb{Z}}$  given in (1.11). Under the null hypothesis that  $X$  is  $K$ -th order stationary, a *multiple scale* procedure aims at testing that the scalogram in a range satisfies



$$\mathcal{H}_0 : \sigma_{j,1}^2 = \dots = \sigma_{j,n_j}^2, \text{ for all } j \in \{J_1, J_1 + 1, \dots, J_2\} \quad (1.35)$$

where  $J_1$  and  $J_2$  are the *finest* and the *coarsest* scales included in the procedure, respectively. The wavelet coefficients at different scales are not uncorrelated so that both the *within-scale* and the *between scale* covariances need to be taken into account.

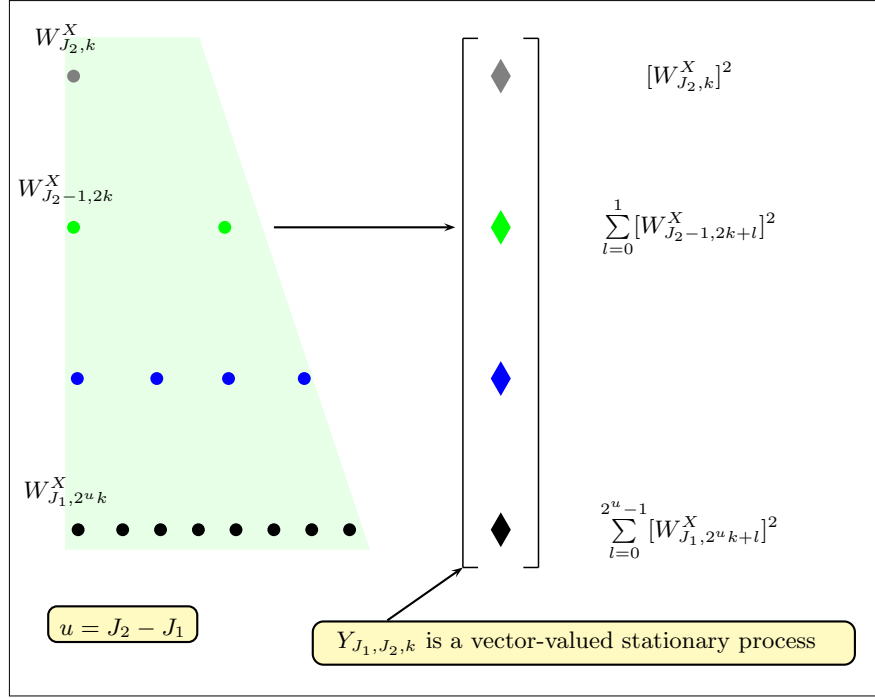


Fig. 1.1. Between scale stationary process.

As before, we use a CUSUM statistic in the wavelet domain. However, we now use multiple scale vector statistics. Consider the following process

$$Y_{J_1, J_2, i} = \left( W_{J_2, i}^2, \sum_{u=1}^2 W_{J_2-1, 2(i-1)+u}^2, \dots, \sum_{u=1}^{2^{(J_2-J_1)}} W_{J_1, 2^{(J_2-J_1)(i-1)+u}}^2 \right)^T .$$

The Bartlett estimator of the covariance matrix of the square wavelet's coefficients for scales  $\{J_1, \dots, J_2\}$  is the  $(J_2 - J_1 + 1) \times (J_2 - J_1 + 1)$  symmetric definite positive matrix  $\hat{\Gamma}_{J_1, J_2}$  given by :

$$\widehat{\Gamma}_{J_1, J_2} = \sum_{\tau=-q(n_{J_2})}^{q(n_{J_2})} w_\tau [q(n_{J_2})] \widehat{\gamma}_{J_1, J_2}(\tau), \quad \text{where} \quad (1.36)$$

$$\widehat{\gamma}_{J_1, J_2}(\tau) = \frac{1}{n_{J_2}} \sum_{i, i+\tau=1}^{n_{J_2}} (Y_{J_1, J_2, i} - \bar{Y}_{J_1, J_2}) (Y_{J_1, J_2, i+\tau} - \bar{Y}_{J_1, J_2})^T. \quad (1.37)$$

where  $\bar{Y}_{J_1, J_2} = \frac{1}{n_{J_2}} \sum_{i=1}^{n_{J_2}} Y_{J_1, J_2, i}$

Finally, let us define the vector of partial sum from scale  $J_1$  to scale  $J_2$  as

$$S_{J_1, J_2}(t) = \frac{1}{\sqrt{n_{J_2}}} \left[ \sum_{i=1}^{\lfloor n_j t \rfloor} W_{j,i}^2 \right]_{j=J_1, \dots, J_2}. \quad (1.38)$$

**Theorem 2.** *Under the assumptions of Theorem 1, we have, as  $n \rightarrow \infty$ ,*

$$\widehat{\Gamma}_{J_1, J_2} = \Gamma_{J_1, J_2} + O_P\left(\frac{q(n_{J_2})}{n_{J_2}}\right) + O_P(q^{-1}(n_{J_2})), \quad (1.39)$$

where  $\Gamma_{J_1, J_2}(j, j') = \sum_{h \in \mathbb{Z}} \text{Cov}(Y_{j,0}, Y_{j',h})$ , with  $1 \leq j, j' \leq J_2 - J_1 + 1$  and,

$$\widehat{\Gamma}_{J_1, J_2}^{-1/2} (S_{J_1, J_2}(t) - \mathbb{E}[S_{J_1, J_2}(t)]) \xrightarrow{\mathcal{L}} B(t) = (B_{J_1}(t), \dots, B_{J_2}(t)), \quad (1.40)$$

in  $D^{J_2 - J_1 + 1}[0, 1]$ , where  $\{B_j(t)\}_{j=J_1, \dots, J_2}$  are independent Brownian motions.

The proof of this result follows the same line as the proof of Theorem 1 and is therefore omitted.

## 1.4 Test statistics

Under the assumption of Theorem 1, the statistics

$$T_{J_1, J_2}(t) \stackrel{\text{def}}{=} (S_{J_1, J_2}(t) - tS_{J_1, J_2}(1))^T \widehat{\Gamma}_{J_1, J_2}^{-1} (S_{J_1, J_2}(t) - tS_{J_1, J_2}(1)) \quad (1.41)$$

converges in weakly in the Skorokhod space  $D([0, 1])$

$$T_{J_1, J_2}(t) \xrightarrow{\mathcal{L}} \sum_{\ell=1}^{J_2 - J_1 - 1} [B_\ell^0(t)]^2 \quad (1.42)$$

where  $t \mapsto (B_1^0(t), \dots, B_{J_2 - J_1 + 1}^0(t))$  is a vector of  $J_2 - J_1 + 1$  independent Brownian bridges

Nominal S.	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$
0.95	0.4605	0.7488	1.0014	1.2397	1.4691	1.6848
0.99	0.7401	1.0721	1.3521	1.6267	1.8667	2.1259

**Table 1.1.** Quantiles of the distribution  $C(d)$  (see (1.44)) for different values of  $d$

For any continuous function  $F : D[0, 1] \rightarrow \mathbb{R}$ , the continuous mapping Theorem implies that

$$F[T_{J_1, J_2}(\cdot)] \xrightarrow{\mathcal{L}} F \left[ \sum_{\ell=1}^{J_2 - J_1 - 1} [B_\ell^0(\cdot)]^2 \right].$$

We may for example apply either integral or max functionals, or weighted versions of these. A classical example of integral function is the so-called Cramér-Von Mises functional given by

$$\text{CVM}(J_1, J_2) \stackrel{\text{def}}{=} \int_0^1 T_{J_1, J_2}(t) dt, \tag{1.43}$$

which converges to  $C(J_2 - J_1 + 1)$  where for any integer  $d$ ,

$$C(d) \stackrel{\text{def}}{=} \int_0^1 \sum_{\ell=1}^d [B_\ell^0(t)]^2 dt. \tag{1.44}$$

The test rejects the null hypothesis when  $\text{CVM}_{J_1, J_2} \geq c(J_2 - J_1 + 1, \alpha)$ , where  $c(d, \alpha)$  is the  $1 - \alpha$ th quantile of the distribution of  $C(d)$ . The distribution of the random variable  $C(d)$  has been derived by Kiefer (1959) (see also Carmona et al. (1999) for more recent references). It holds that, for  $x > 0$ ,

$$\mathbb{P}(C(d) \leq x) = \frac{2^{(d+1)/2}}{\pi^{1/2} x^{d/4}} \sum_{j=0}^{\infty} \frac{\Gamma(j + d/2)}{j! \Gamma(d/2)} e^{-(j+d/4)^2/x} \text{Cyl}_{(d-2)/2} \left( \frac{2j + d/2}{x^{1/2}} \right)$$

where  $\Gamma$  denotes the gamma function and Cyl are the parabolic cylinder functions. The quantile of this distribution are given in table 1.1 for different values of  $d = J_2 - J_1 + 1$ . It is also possible to use the max. functional leading to an analogue of the Kolmogorov-Smirnov statistics,

$$\text{KSM}(J_1, J_2) \stackrel{\text{def}}{=} \sup_{0 \leq t \leq 1} T_{J_1, J_2}(t) \tag{1.45}$$

which converges to  $D(J_2 - J_1 + 1)$  where for any integer  $d$ ,

$d$	1	2	3	4	5	6
0.95	1.358	1.58379	1.7472	1.88226	2.00	2.10597
0.99	1.627624	1.842726	2.001	2.132572	2.24798	2.35209

**Table 1.2.** Quantiles of the distribution  $D(d)$  (see (1.46)) for different values of  $d$ .

$$D(d) \stackrel{\text{def}}{=} \sup_{0 \leq t \leq 1} \sum_{\ell=1}^d [B_\ell^0(t)]^2. \quad (1.46)$$

The test reject the null hypothesis when  $\text{KSM}_{J_1, J_2} \geq \delta(J_2 - J_1 + 1, \alpha)$ , where  $\delta(d, \alpha)$  is the  $(1 - \alpha)$ -quantile of  $D(d)$ . The distribution of  $D(d)$  has again be derived by Kiefer (1959) (see also Pitman and Yor (1999) for more recent references). It holds that, for  $x > 0$ ,

$$\mathbb{P}(D(d) \leq x) = \frac{2^{1+(2-d)/2}}{\Gamma(d/2)a^d} \sum_{n=1}^{\infty} \frac{j_{\nu,n}^{2\nu}}{J_{\nu+1}^2(j_{\nu,n})} \exp\left(-\frac{j_{\nu,n}^2}{2x^2}\right),$$

where  $0 < j_{\nu,1} < j_{\nu,2} < \dots$  is the sequence of positive zeros of  $J_\nu$ , the Bessel function of index  $\nu = (d - 2)/2$ . The quantiles of this distribution are given in Table 1.2.

## 1.5 Power of the W2-CUSUM statistics

### 1.5.1 Power of the test in single scale case

In this section we investigate the power of the test. A minimal requirement is to establish that the test procedure is pointwise consistent in a presence of a breakpoint, *i.e.* that under a fixed alternative, the probability of detection converges to one as the sample size goes to infinity. We must therefore first define such alternative. For simplicity, we will consider an alternative where the process exhibit a single breakpoint, though it is likely that the test does have power against more general class of alternatives.

The alternative that we consider in this section is defined as follows. Let  $f_1$  and  $f_2$  be two given generalized spectral densities and suppose that, at a given scale  $j$ ,  $\int_{-\pi}^{\pi} |H_j(\lambda)|^2 f_i(\lambda) d\lambda < \infty$ ,  $i = 1, 2$ , and

$$\int_{-\pi}^{\pi} |H_j(\lambda)|^2 (f_1(\lambda) - f_2(\lambda)) d\lambda \neq 0. \quad (1.47)$$

Define by  $(X_{l,i})_{l \in \mathbb{Z}}$ ,  $i = 1, 2$ , be two Gaussian processes, defined on the same probability space, with generalized spectral density  $f_1$ . We do not specify the dependence structure between these two processes, which can be arbitrary. Let  $\kappa \in ]0, 1[$  be a breakpoint. We consider a sequence of Gaussian processes  $(X_k^n)_{k \in \mathbb{Z}}$ , such that

$$X_k^{(n)} = X_{k,i} \text{ for } k \leq \lfloor n\kappa \rfloor \text{ and } X_k^{(n)} = X_{k,2} \text{ for } k \geq \lfloor n\kappa \rfloor + 1. \quad (1.48)$$

**Theorem 3.** Consider  $\{X_k^n\}_{k \in \mathbb{Z}}$  be a sequence of processes specified by (1.47) and (1.48). Assume that  $q(n_j)$  is non decreasing and :

$$q(n_j) \rightarrow \infty \text{ and } \frac{q(n_j)}{n_j} \rightarrow 0 \text{ as } n_j \rightarrow \infty. \quad (1.49)$$

Then the statistic  $T_{n_j}$  defined by (1.13) satisfies

$$\frac{\sqrt{n_j}}{\sqrt{2q(n_j)}} \sqrt{\kappa(1-\kappa)}(1 + o_p(1)) \leq T_{n_j} \xrightarrow{P} \infty. \quad (1.50)$$

*Proof.* Let  $k_j = \lfloor n_j \kappa \rfloor$  the change point in the wavelet spectrum at scale  $j$ . We write  $q$  for  $q(n_j)$  and suppress the dependence in  $n$  in this proof to alleviate the notation. By definition  $T_{n_j} = \frac{1}{s_{j,n_j}} \sup_{0 \leq t \leq 1} (S_{n_j}(t) - tS_{n_j}(1))$ , where the process  $t \mapsto S_{n_j}(t)$  is defined in (1.24). Therefore,  $T_{n_j} \geq \frac{1}{s_{j,n_j}} (S_{n_j}(\kappa) - \kappa S_{n_j}(1))$ . The proof consists in establishing that  $\frac{1}{s_{j,n_j}} (S_{n_j}(\kappa) - \kappa S_{n_j}(1)) = \frac{\sqrt{n_j}}{\sqrt{2q(n_j)}} \sqrt{\kappa(1-\kappa)}(1 + o_p(1))$ . We first decompose this difference as follows

$$\begin{aligned} S_{n_j}(\kappa) - \kappa S_{n_j}(1) &= \frac{1}{\sqrt{n_j}} \left| \sum_{i=1}^{\lfloor n_j \kappa \rfloor} W_{j,i}^2 - \kappa \sum_{i=1}^{n_j} W_{j,i}^2 \right| \\ &= B_{n_j} + f_{n_j} \end{aligned}$$

where  $B_{n_j}$  is a fluctuation term

$$B_{n_j} = \frac{1}{\sqrt{n_j}} \left| \sum_{i=1}^{k_j} (W_{j,i}^2 - \sigma_{j,i}^2) - \kappa \sum_{i=1}^{n_j} (W_{j,i}^2 - \sigma_{j,i}^2) \right| \quad (1.51)$$

and  $f_{n_j}$  is a bias term

$$f_{n_j} = \frac{1}{\sqrt{n_j}} \left| \sum_{i=1}^{k_j} \sigma_{j,i}^2 - \kappa \sum_{i=1}^{n_j} \sigma_{j,i}^2 \right|. \quad (1.52)$$

Since support of  $h_{j,l}$  is included in  $[-T(2^j+1), 0]$  where  $h_{j,l}$  is defined in (1.7), there exists a constant  $a > 0$  such that

$$W_{j,i} = W_{j,i;1} = \sum_{l \leq k} h_{j,2^j i - l} X_{l,1}, \quad \text{for } i < k_j, \quad (1.53)$$

$$W_{j,i} = W_{j,i;2} = \sum_{l > k} h_{j,2^j i - l} X_{l,2} \quad \text{for } i > k_j + a, \quad (1.54)$$

$$W_{j,i} = \sum_l h_{j,2^j i - l} X_l, \quad \text{for } k_j \leq i < k_j + a. \quad (1.55)$$

Since the process  $\{X_{l,1}\}_{l \in \mathbb{Z}}$  and  $\{X_{l,2}\}_{l \in \mathbb{Z}}$  are both  $K$ -th order covariance stationary, the two processes  $\{W_{j,i;1}\}_{i \in \mathbb{Z}}$  and  $\{W_{j,i;2}\}_{i \in \mathbb{Z}}$  are also covariance stationary. The wavelet coefficients  $W_{j,i}$  for  $i \in \{k_j, \dots, k_j + a\}$  are computed using observations from the two processes  $X_1$  and  $X_2$ . Let us show that there exists a constant  $C > 0$  such that, for all integers  $l$  and  $\tau$ ,

$$\text{Var} \left( \sum_{i=l}^{l+\tau} W_{j,i}^2 \right) \leq C\tau. \quad (1.56)$$

Using (1.21), we have, for  $\epsilon = 1, 2$ ,

$$\text{Var} \left( \sum_{i=l}^{l+\tau} W_{j,i;\epsilon}^2 \right) \leq \frac{\tau}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0;\epsilon}(\lambda)|^2 d\lambda$$

where,  $\mathbf{D}_{j,0;1}(\lambda)$  and  $\mathbf{D}_{j,0;2}(\lambda)$  denote the spectral density of the stationary processes  $\{W_{j,i;1}\}_{i \in \mathbb{Z}}$  and  $\{W_{j,i;2}\}_{i \in \mathbb{Z}}$  respectively. Using Minkovski inequality, we have for  $l \leq k_j \leq k_j + a < l + \tau$  that  $\left( \text{Var} \sum_{i=l}^{l+\tau} W_{j,i}^2 \right)^{1/2}$  is at most

$$\begin{aligned} & \left( \text{Var} \sum_{i=l}^{k_j} W_{j,i}^2 \right)^{1/2} + \sum_{i=k_j+1}^{k_j+a} (\text{Var} W_{j,i}^2)^{1/2} + \left( \text{Var} \sum_{i=k_j+a+1}^{l+\tau} W_{j,i}^2 \right)^{1/2} \\ & \leq \left( \text{Var} \sum_{i=l}^{k_j} W_{j,i;1}^2 \right)^{1/2} + a \sup_i (\text{Var} W_{j,i}^2)^{1/2} + \left( \text{Var} \sum_{i=k_j+a+1}^{l+\tau} W_{j,i;2}^2 \right)^{1/2}. \end{aligned}$$

Observe that  $\text{Var}(W_{j,i}^2) \leq 2(\sum_l |h_{j,l}|)^2 (\sigma_{j,1}^2 \vee \sigma_{j,2}^2)^2 < \infty$  for  $k_j \leq i < k_j + a$ , where

$$\sigma_{j;1}^2 = \mathbb{E} [W_{j,i;1}^2], \quad \text{and} \quad \sigma_{j;2}^2 = \mathbb{E} [W_{j,i;2}^2] \quad (1.57)$$

The three last displays imply (1.56) and thus that  $B_{n_j}$  is bounded in probability. Moreover, since  $f_{n_j}$  reads

$$\begin{aligned} \frac{1}{\sqrt{n_j}} \left| \sum_{i=1}^{\lfloor n_j \kappa \rfloor} \sigma_{j;1}^2 - \kappa \sum_{i=1}^{\lfloor n_j \kappa \rfloor} \sigma_{j;1}^2 - \kappa \sum_{i=\lfloor n_j \kappa \rfloor + 1}^{\lfloor n_j \kappa \rfloor + a} \sigma_{j,i}^2 - \kappa \sum_{i=\lfloor n_j \kappa \rfloor + a + 1}^{n_j} \sigma_{j;2}^2 \right| \\ = \sqrt{n_j} \kappa (1 - \kappa) |\sigma_{j;1}^2 - \sigma_{j;2}^2| + O(n_j^{-1/2}), \end{aligned}$$

we get

$$S_{n_j}(\kappa) - \kappa S_{n_j}(1) = \sqrt{n_j} \kappa (1 - \kappa) (\sigma_{j;1}^2 - \sigma_{j;2}^2) + O_P(1). \quad (1.58)$$

We now study the denominator  $s_{j,n_j}^2$  in (1.13). Denote by

$$\bar{\sigma}_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} \sigma_{j,i}^2$$

the expectation of the scalogram (which now differs from the wavelet spectrum). Let us consider for  $\tau \in \{0, \dots, q(n_j)\}$   $\hat{\gamma}_j(\tau)$  the empirical covariance of the wavelet coefficients defined in (1.15).

$$\begin{aligned} \hat{\gamma}_j(\tau) = \frac{1}{n_j} \sum_{i=1}^{n_j - \tau} (W_{j,i}^2 - \bar{\sigma}_j^2)(W_{j,i+\tau}^2 - \bar{\sigma}_j^2) - (1 + \frac{\tau}{n_j}) (\bar{\sigma}_j^2 - \hat{\sigma}_j^2)^2 \\ + \frac{1}{n_j} (\hat{\sigma}_j^2 - \bar{\sigma}_j^2) \left\{ \sum_{i=n_j - \tau + 1}^{n_j} (W_{j,i}^2 - \bar{\sigma}_j^2) + \sum_{i=1}^{\tau} (W_{j,i}^2 - \bar{\sigma}_j^2) \right\}. \end{aligned}$$

Using Minkowski inequality and (1.56), there exists a constant  $C$  such that for all  $1 \leq l \leq l + \tau \leq n_j$ ,

$$\begin{aligned} \left\| \sum_{i=l}^{l+\tau} (W_{j,i}^2 - \bar{\sigma}_j^2) \right\|_2 \leq \left\| \sum_{i=l}^{l+\tau} (W_{j,i}^2 - \sigma_{j,i}^2) \right\|_2 + \left\| \sum_{i=l}^{l+\tau} (\sigma_{j,i}^2 - \bar{\sigma}_j^2) \right\|_2 \\ \leq C(\tau^{1/2} + \tau), \end{aligned}$$

and similarly

$$\|\hat{\sigma}_j^2 - \bar{\sigma}_j^2\|_2 \leq \frac{C}{\sqrt{n_j}}.$$

By combining these two latter bounds, the Cauchy-Schwarz inequality implies that

$$\left\| \frac{1}{n_j} (\hat{\sigma}_j^2 - \bar{\sigma}_j^2) \sum_{i=l}^{l+\tau} (W_{j,i}^2 - \bar{\sigma}_j^2) \right\|_1 \leq \frac{C(\tau^{1/2} + \tau)}{n_j^{3/2}}.$$

Recall that  $s_{j,n_j}^2 = \sum_{\tau=-q}^q w_\tau(q) \hat{\gamma}_j(\tau)$  where  $w_\tau(q)$  are the so-called Bartlett weights defined in (1.16). We now use the bounds above to identify the limit

of  $s_{j,n_j}^2$  as the sample size goes to infinity. The two previous identities imply that

$$\sum_{\tau=0}^q w_\tau(q) \left(1 + \frac{\tau}{n_j}\right) \|\bar{\sigma}_j^2 - \hat{\sigma}_j^2\|_2 \leq C \frac{q^2}{n_j^{3/2}}$$

and

$$\sum_{\tau=0}^q w_\tau(q) \left\| \frac{1}{n_j} (\hat{\sigma}_j^2 - \bar{\sigma}_j^2) \sum_{i=l}^{l+\tau} (W_{j,i}^2 - \bar{\sigma}_j^2) \right\|_1 \leq C \frac{q^2}{n_j^{3/2}},$$

Therefore, we obtain

$$s_{j,n_j}^2 = \sum_{\tau=-q}^q w_\tau(q) \tilde{\gamma}_j(\tau) + O_P\left(\frac{q^2}{n_j}\right), \quad (1.59)$$

where  $\tilde{\gamma}_j(\tau)$  is defined by

$$\tilde{\gamma}_j(\tau) = \frac{1}{n_j} \sum_{i=1}^{n_j-\tau} (W_{j,i}^2 - \bar{\sigma}_j^2)(W_{j,i+\tau}^2 - \bar{\sigma}_j^2). \quad (1.60)$$

Observe that since  $q = o(n_j)$ ,  $k_j = \lfloor n_j \kappa \rfloor$  and  $0 \leq \tau \leq q$ , then for any given integer  $a$  and  $n$  large enough  $0 \leq \tau \leq k_j \leq k_j + a \leq n_j - \tau$  thus in (1.60) we may write  $\sum_{i=1}^{n_j-\tau} = \sum_{i=1}^{k_j-\tau} + \sum_{i=k_j-\tau+1}^{k+a} + \sum_{i=k_j+a+1}^{n_j-\tau}$ . Using  $\sigma_{j;1}^2$  and  $\sigma_{j;2}^2$  in (1.60) and straightforward bounds that essentially follow from (1.56), we get  $s_{j,n_j}^2 = \bar{s}_{j,n_j}^2 + O_P\left(\frac{q^2}{n_j}\right)$ , where

$$\begin{aligned} \bar{s}_{j,n_j}^2 &= \sum_{\tau=-q}^q w_\tau(q) \left( \frac{k}{n_j} \tilde{\gamma}_{j;1}(\tau) + \frac{n_j - k_j - a}{n_j} \tilde{\gamma}_{j;2}(\tau) \right. \\ &\quad \left. + \frac{k_j - |\tau|}{n_j} (\sigma_{j;1}^2 - \bar{\sigma}_j^2)^2 + \frac{n_j - k_j - a - |\tau|}{n_j} (\sigma_{j;2}^2 - \bar{\sigma}_j^2)^2 \right) \end{aligned}$$

with

$$\begin{aligned} \tilde{\gamma}_{j;1}(\tau) &= \frac{1}{k_j} \sum_{i=1}^{k_j-\tau} (W_{j,i}^2 - \sigma_{j;1}^2) (W_{j,i+\tau}^2 - \sigma_{j;1}^2), \\ \tilde{\gamma}_{j;2}(\tau) &= \frac{1}{n_j - k_j - a} \sum_{i=k_j+a+1}^{n_j-\tau} (W_{j,i}^2 - \sigma_{j;2}^2) (W_{j,i+\tau}^2 - \sigma_{j;2}^2). \end{aligned}$$

Using that  $\bar{\sigma}_j^2 \rightarrow \kappa \sigma_{j;1}^2 + (1 - \kappa) \sigma_{j;2}^2$  as  $n_j \rightarrow \infty$ , and that, for  $\epsilon = 1, 2$ ,



$$s_{j,n_j;\epsilon}^2 \stackrel{\text{def}}{=} \sum_{\tau=-q}^q w_\tau(q) \tilde{\gamma}_{j;\epsilon}(\tau) \xrightarrow{P} \frac{1}{\pi} \int_{-\pi}^{\pi} |\mathbf{D}_{j,0;\epsilon}(\lambda)|^2 d\lambda,$$

we obtain

$$\begin{aligned} s_{j,n_j}^2 &= \frac{1}{\pi} \int_{-\pi}^{\pi} \left\{ \kappa |\mathbf{D}_{j,0;1}(\lambda)|^2 + (1-\kappa) |\mathbf{D}_{j,0;2}(\lambda)|^2 \right\} d\lambda \\ &= 2q\kappa(1-\kappa) (\sigma_{j;1}^2 - \sigma_{j;2}^2)^2 + o_p(1) + O_P\left(\frac{q^2}{n_j}\right). \end{aligned} \quad (1.61)$$

Using (1.58), the last display and that  $o_p(1) + O_P\left(\frac{q^2}{n_j}\right) = o_p(q)$ , we finally obtain

$$\begin{aligned} S_{n_j}(\kappa) - \kappa S_{n_j}(1) &= \frac{\sqrt{n_j} \kappa (1-\kappa) |\sigma_{j;1}^2 - \sigma_{j;2}^2| + O_P(1)}{\sqrt{2q(\kappa(1-\kappa))} |\sigma_{j;1}^2 - \sigma_{j;2}^2| + o_p(\sqrt{q})} \\ &= \frac{\sqrt{n_j}}{\sqrt{2q}} \sqrt{\kappa(1-\kappa)} (1 + o_p(1)), \end{aligned}$$

which concludes the proof of Theorem 3.

### 1.5.2 Power of the test in multiple scales case

The results obtained in the previous Section in the single scale case easily extend to the test procedure designed to handle the multiple scales case. The alternative is specified exactly in the same way than in the single scale case but instead of considering the square of the wavelet coefficients at a given scale, we now study the behavior of the between-scale process. Consider the following process for  $\epsilon = 1, 2$ ,

$$Y_{J_1, J_2, i; \epsilon} = \left( W_{J_2, i; \epsilon}^2, \sum_{u=1}^2 W_{J_2-1, 2(i-1)+u; \epsilon}^2, \dots, \sum_{u=1}^{2^{(J_2-J_1)}} W_{J_1, 2^{(J_2-J_1)}(i-1)+u; \epsilon}^2 \right)^T,$$

where  $J_1$  and  $J_2$  are respectively the finest and the coarsest scale considered in the test,  $W_{j,i;\epsilon}$  are defined in (1.53) and (1.54) and  $\Gamma_{J_1, J_2; \epsilon}$  the  $(J_2 - J_1 + 1) \times (J_2 - J_1 + 1)$  symmetric non negative matrix such that

$$\Gamma_{J_1, J_2; \epsilon}(j, j') = \sum_{h \in \mathbb{Z}} \text{Cov}(Y_{j,0;\epsilon}, Y_{j',h;\epsilon}) = \int_{-\pi}^{\pi} \|\mathbf{D}_{j,u;\epsilon}(\lambda; f)\|^2 d\lambda, \quad (1.62)$$

with  $1 \leq j, j' \leq J_2 - J_1 + 1$  for  $\epsilon = 1, 2$ .

**Theorem 4.** Consider  $\{X_k^n\}_{k \in \mathbb{Z}}$  be a sequence of processes specified by (1.47) and (1.48). Finally assume that for at least one  $j \in \{J_1, \dots, J_2\}$  and that at least one of the two matrices  $\Gamma_{J_1, J_2; \epsilon}$   $\epsilon = 1, 2$  defined in (1.62) is positive definite. Assume in addition that Finally, assume that the number of lags  $q(n_{J_2})$  in the Barlett estimate of the covariance matrix (1.36) is non decreasing and:

$$q(n_{j_2}) \rightarrow \infty \quad \text{and} \quad \frac{q^2(n_{J_2})}{n_{J_2}} \rightarrow 0, \quad \text{as } n_{J_2} \rightarrow \infty, \quad . \quad (1.63)$$

Then, the W2-CUSUM test statistics  $T_{J_1, J_2}$  defined by (1.41) satisfies

$$\frac{n_{J_2}}{2q(n_{J_2})} \kappa(1 - \kappa)(1 + o_p(1)) \leq T_{J_1, J_2} \xrightarrow{P} \infty \quad \text{as } n_{J_2} \rightarrow \infty$$

*Proof.* As in the single scale case we drop the dependence in  $n_{J_2}$  in the expression of  $q$  in this proof section. Let  $k_j = \lfloor n_j \kappa \rfloor$  the change point in the wavelet spectrum at scale  $j$ . Then using (1.38), we have that  $T_{J_1, J_2} \geq S_{J_1, J_2}(\kappa) - \kappa S_{J_1, J_2}(1)$  where

$$S_{J_1, J_2}(\kappa) - \kappa S_{J_1, J_2}(1) = \frac{1}{\sqrt{n_{J_2}}} [n_j(B_{n_j} + f_{n_j})]_{j=J_1, \dots, J_2},$$

where  $B_{n_j}$  and  $f_{n_j}$  are defined respectively by (1.51) and (1.52). Hence as in (1.58), we have

$$S_{J_1, J_2}(\kappa) - \kappa S_{J_1, J_2}(1) = \sqrt{n_{J_2}} \kappa(1 - \kappa) \Delta + O_P(1),$$

where  $\Delta = [\sigma_{J_1, J_2; 1}^2 - \sigma_{J_1, J_2; 2}^2]^T$  and

$$\sigma_{J_1, J_2; \epsilon}^2 = (\sigma_{J_2; \epsilon}^2, \dots, 2^{J_2 - J_1} \sigma_{J_1; \epsilon}^2)^T.$$

We now study the asymptotic behavior of  $\widehat{\Gamma}_{J_1, J_2}$ . Using similar arguments as those leading to (1.61) in the proof of Theorem 3, we have

$$\begin{aligned} \widehat{\Gamma}_{J_1, J_2} &= 2q\kappa(1 - \kappa)\Delta\Delta^T + \kappa\Gamma_{J_1, J_2; 1} + (1 - \kappa)\Gamma_{J_1, J_2; 2} \\ &\quad + O_P\left(\frac{q}{n_{J_2}}\right) + O_P(q^{-1}) + O_P\left(\frac{q^2}{n_{J_2}}\right). \end{aligned}$$

For  $\Gamma$  a positive definite matrix, consider the matrix  $\mathbf{M}(\Gamma) = \Gamma + 2q\kappa(1 - \kappa)\Delta\Delta^T$ . Using the matrix inversion lemma, the inverse of  $\mathbf{M}(\Gamma)$  may be expressed as

$$\mathbf{M}^{-1}(\Gamma) = \left( \Gamma^{-1} - \frac{2q\kappa(1 - \kappa)\Gamma^{-1}\Delta\Delta^T\Gamma^{-1}}{1 + 2q\kappa(1 - \kappa)\Delta^T\Gamma^{-1}\Delta} \right),$$

which implies that

$$\Delta^T \mathbf{M}^{-1}(\Gamma) \Delta = \frac{\Delta^T \Gamma^{-1} \Delta}{1 + 2q\kappa(1 - \kappa)\Delta^T \Gamma^{-1} \Delta}.$$

Applying these two last relations to  $\Gamma_0 = \kappa\Gamma_{J_1, J_2}^{(1)} + (1 - \kappa)\Gamma_{J_1, J_2}^{(2)}$  which is symmetric and definite positive (since, under the stated assumptions at least one of the two matrix  $\Gamma_{J_1, J_2; \epsilon}$ ,  $\epsilon = 1, 2$  is positive) we have

$$\begin{aligned} T_{J_1, J_2} &\geq \kappa^2(1 - \kappa)^2 n_{J_2} \Delta^T \mathbf{M}^{-1} \left( \Gamma_0 + O_P \left( \frac{q^2}{n_{J_2}} \right) + O_P(q^{-1}) \right) \Delta + O_P(1) \\ &= n_{J_2} \kappa^2(1 - \kappa)^2 \frac{\Delta^T \Gamma_0^{-1} \Delta + O_P \left( \frac{q^2}{n_{J_2}} \right) + O_P(q^{-1})}{2q\kappa(1 - \kappa)\Delta^T \Gamma_0^{-1} \Delta(1 + o_p(1))} + O_P(1) \\ &= \frac{n_{J_2}}{2q} \kappa(1 - \kappa) (1 + o_p(1)) . \end{aligned}$$

Thus  $T_{J_1, J_2} \xrightarrow{P} \infty$  as  $n_{J_2} \rightarrow \infty$ , which completes the proof of Theorem 4.

*Remark 3.* The term corresponding to the "bias" term  $\kappa\Gamma_{J_1, J_2; 1} + (1 - \kappa)\Gamma_{J_1, J_2; 2}$  in the single case is  $\frac{1}{\pi} \int_{-\pi}^{\pi} \{\kappa|\mathbf{D}_{j, 0; 1}(\lambda)|^2 + (1 - \kappa)|\mathbf{D}_{j, 0; 2}(\lambda)|^2\} d\lambda = O(1)$ , which can be neglected since the main term in  $s_{j, n_j}^2$  is of order  $q \rightarrow \infty$ . In multiple scale case, the main term in  $\widehat{\Gamma}_{J_1, J_2}$  is still of order  $q$  but is no longer invertible (the rank of the leading term is equal to 1). A closer look is thus necessary and the term  $\kappa\Gamma_{J_1, J_2; 1} + (1 - \kappa)\Gamma_{J_1, J_2; 2}$  has to be taken into account. This is also explains why we need the more stringent condition (1.63) on the bandwidth size in the multiple scales case.

## 1.6 Some examples

In this section, we report the results of a limited Monte-Carlo experiment to assess the finite sample property of the test procedure. Recall that the test rejects the null if either  $\text{CVM}(J_1, J_2)$  or  $\text{KSM}(J_1, J_2)$ , defined in (1.43) and (1.45) exceeds the  $(1 - \alpha)$ -th quantile of the distributions  $C(J_2 - J_1 + 1)$  and  $D(J_2 - J_1 + 1)$ , specified in (1.44) and (1.46). The quantiles are reported in Tables (1.1) and (1.2), and have been obtained by truncating the series expansion of the cumulative distribution function. To study the influence on the test procedure of the strength of the dependency, we consider different classes of Gaussian processes, including white noise, autoregressive

moving average (ARMA) processes as well as fractionally integrated ARMA (ARFIMA( $p, d, q$ )) processes which are known to be long range dependent. In all the simulations we set the lowest scale to  $J_1 = 1$  and vary the coarsest scale  $J_2 = J$ . We used a wide range of values of sample size  $n$ , of the number of scales  $J$  and of the parameters of the ARMA and FARIMA processes but, to conserve space, we present the results only for  $n = 512, 1024, 2048, 4096, 8192$ ,  $J = 3, 4, 5$  and four different models: an AR(1) process with parameter 0.9, a MA(1) process with parameter 0.9, and two ARFIMA(1,d,1) processes with memory parameter  $d = 0.3$  and  $d = 0.4$ , and the same AR and MA coefficients, set to 0.9 and 0.1. In our simulations, we have used the Newey-West estimate of the bandwidth  $q(n_j)$  for the covariance estimator (as implemented in the R-package *sandwich*).

### Asymptotic level of *KSM* and *CVM*.

We investigate the finite-sample behavior of the test statistics  $CVM(J_1, J_2)$  and  $KSM(J_1, J_2)$  by computing the number of times that the null hypothesis is rejected in 1000 independent replications of each of these processes under  $\mathcal{H}_0$ , when the asymptotic level is set to 0.05.

White noise					
$n$	512	1024	2048	4096	8192
$J = 3$ <i>KSM</i>	0.02	0.01	0.03	0.02	0.02
$J = 3$ <i>CVM</i>	0.05	0.045	0.033	0.02	0.02
$J = 4$ <i>KSM</i>	0.047	0.04	0.04	0.02	0.02
$J = 4$ <i>CVM</i>	0.041	0.02	0.016	0.016	0.01
$J = 5$ <i>KSM</i>	0.09	0.031	0.02	0.025	0.02
$J = 5$ <i>CVM</i>	0.086	0.024	0.012	0.012	0.02

**Table 1.3.** Empirical level of *KSM* – *CVM* for a white noise.

MA(1)[ $\theta = 0.9$ ]					
$n$	512	1024	2048	4096	8192
$J = 3$ <i>KSM</i>	0.028	0.012	0.012	0.012	0.02
$J = 3$ <i>CVM</i>	0.029	0.02	0.016	0.016	0.01
$J = 4$ <i>KSM</i>	0.055	0.032	0.05	0.025	0.02
$J = 4$ <i>CVM</i>	0.05	0.05	0.03	0.02	0.02
$J = 5$ <i>KSM</i>	0.17	0.068	0.02	0.02	0.02
$J = 5$ <i>CVM</i>	0.13	0.052	0.026	0.021	0.02

**Table 1.4.** Empirical level of KSM – CVM for a  $MA(q)$  process.

AR(1)[ $\phi = 0.9$ ]					
$n$	512	1024	2048	4096	8192
$J = 3$ <i>KSM</i>	0.083	0.073	0.072	0.051	0.04
$J = 3$ <i>CVM</i>	0.05	0.05	0.043	0.032	0.03
$J = 4$ <i>KSM</i>	0.26	0.134	0.1	0.082	0.073
$J = 4$ <i>CVM</i>	0.14	0.092	0.062	0.04	0.038
$J = 5$ <i>KSM</i>	0.547	0.314	0.254	0.22	0.11
$J = 5$ <i>CVM</i>	0.378	0.221	0.162	0.14	0.093

**Table 1.5.** Empirical level of KSM – CVM for an  $AR(1)$  process.

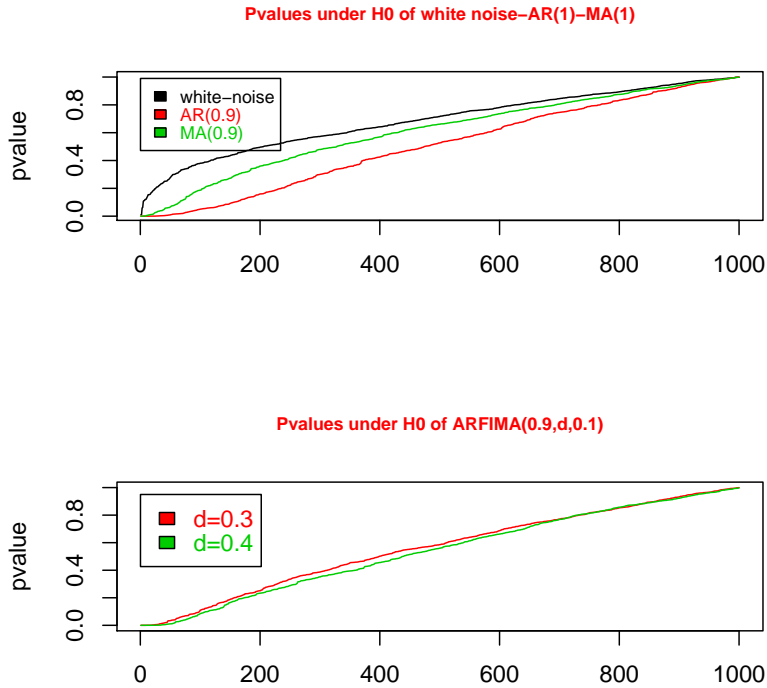
ARFIMA(1,0.3,1)[ $\phi = 0.9, \theta = 0.1$ ]					
$n$	512	1024	2048	4096	8192
$J = 3$ <i>KSM</i>	0.068	0.047	0.024	0.021	0.02
$J = 3$ <i>CVM</i>	0.05	0.038	0.03	0.02	0.02
$J = 4$ <i>KSM</i>	0.45	0.42	0.31	0.172	0.098
$J = 4$ <i>CVM</i>	0.39	0.32	0.20	0.11	0.061
$J = 5$ <i>KSM</i>	0.57	0.42	0.349	0.229	0.2
$J = 5$ <i>CVM</i>	0.41	0.352	0.192	0.16	0.11

**Table 1.6.** Empirical level of KSM – CVM for an  $ARFIMA(1, 0.3, 1)$  process.

ARFIMA(1,0.4,1)[ $\phi = 0.9, \theta = 0.1$ ]					
$n$	512	1024	2048	4096	8192
$J = 3$ KSM	0.11	0.063	0.058	0.044	0.031
$J = 3$ CVM	0.065	0.05	0.043	0.028	0.02
$J = 4$ KSM	0.512	0.322	0.26	0.2	0.18
$J = 4$ CVM	0.49	0.2	0.192	0.16	0.08
$J = 5$ KSM	0.7	0.514	0.4	0.321	0.214
$J = 5$ CVM	0.59	0.29	0.262	0.196	0.121

**Table 1.7.** Empirical level of KSM – CVM for an  $ARFIMA(1, 0.3, 1)$  process.

We notice that in general the empirical levels for the CVM are globally more accurate than the ones for the KSM test, the difference being more significant when the strength of the dependence is increased, or when the number of scales that are tested simultaneously get larger. The tests are slightly too conservative in the white noise and the MA case (tables (1.3) and (1.4)); in the AR(1) case and in the ARFIMA cases, the test rejects the null much too often when the number of scales is large compared to the sample size (the difficult problem being in that case to estimate the covariance matrix of the test). For  $J = 4$ , the number of samples required to meet the target rejection rate can be as large as  $n = 4096$  for the CVM test and  $n = 8192$  for the KSM test. The situation is even worse in the ARFIMA case (tables (1.6) and (1.7)). When the number of scales is equal to 4 or 5, the test rejects the null hypothesis much too often.



**Fig. 1.2.** Pvalue under  $\mathcal{H}_0$  of the distribution  $D(J)$   $n = 1024$  for white noise and MA(1) processes and  $n = 4096$  for AR(1) and ARFIMA(1,d,1) processes; the coarsest scale is  $J = 4$  for white noise, MA and AR processes and  $J = 3$  for the ARFIMA process. The finest scale is  $J_1 = 1$ .

**Power of *KSM* and *CVM*.**

We assess the power of test statistic by computing the test statistics in presence of a change in the spectral density. To do so, we consider an observation obtained by concatenation of  $n_1$  observations from a first process and  $n_2$  observations from a second process, independent from the first one and having a different spectral density. The length of the resulting observations is  $n = n_1 + n_2$ . In all cases, we set  $n_1 = n_2 = n/2$ , and we present the results for  $n_1 = 512, 1024, 2048, 4096$  and scales  $J = 4, 5$ . We consider the following situations: the two processes are white Gaussian noise with two different variances, two AR processes with different values of the autoregressive coefficient, two MA processes with different values of the moving average coefficient and two ARFIMA with same moving average and same autoregressive coefficients

but different values of the memory parameter  $d$ . The scenario considered is a bit artificial but is introduced here to assess the ability of the test to detect abrupt changes in the spectral content. For 1000 simulations, we report the number of times  $\mathcal{H}_1$  was accepted, leading the following results.

white-noise		[ $\sigma_1^2 = 1, \sigma_2^2 = 0.7$ ]			
$n_1 = n_2$		512	1024	2048	4096
$J = 4$	<i>KSM</i>	0.39	0.78	0.89	0.95
$J = 4$	<i>CVM</i>	0.32	0.79	0.85	0.9
$J = 5$	<i>KSM</i>	0.42	0.79	0.91	0.97
$J = 5$	<i>CVM</i>	0.40	0.78	0.9	0.9

**Table 1.8.** Power of KSM – CVM on two white noise processes.

MA(1)+MA(1)		[ $\theta_1 = 0.9, \theta_2 = 0.5$ ]			
$n_1 = n_2$		512	1024	2048	4096
$J = 4$	<i>KSM</i>	0.39	0.69	0.86	0.91
$J = 4$	<i>CVM</i>	0.31	0.6	0.76	0.93
$J = 5$	<i>KSM</i>	0.57	0.74	0.84	0.94
$J = 5$	<i>CVM</i>	0.46	0.69	0.79	0.96

**Table 1.9.** Power of KSM – CVM on a concatenation of two different *MA* processes.

AR(1)+AR(1)		[ $\phi_1 = 0.9, \phi_2 = 0.5$ ]			
$n_1 = n_2$		512	1024	2048	4096
$J = 4$	<i>KSM</i>	0.59	0.72	0.81	0.87
$J = 4$	<i>CVM</i>	0.53	0.68	0.79	0.9
$J = 5$	<i>KSM</i>	0.75	0.81	0.94	0.92
$J = 5$	<i>CVM</i>	0.7	0.75	0.89	0.91

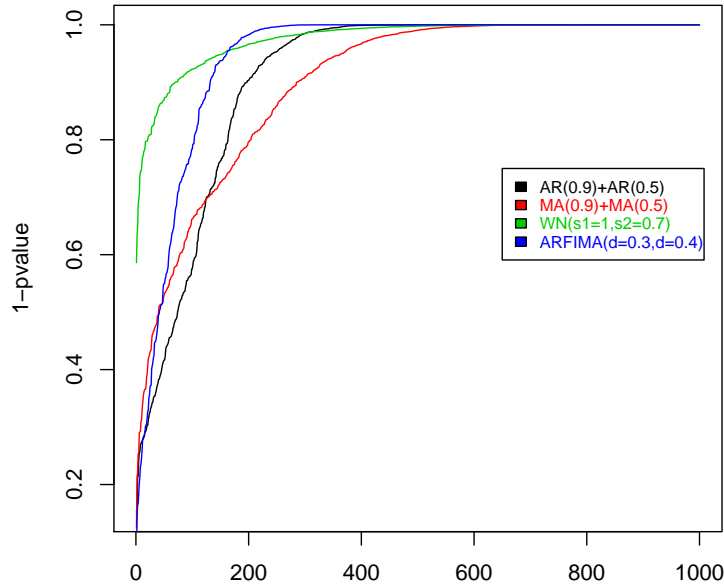
**Table 1.10.** Power of KSM – CVM on a concatenation of two different *AR* processes.

The power of our two statistics gives us satisfying results for the considered processes, especially if the sample size tends to infinity.



ARFIMA(1,0.3,1) + ARFIMA(1,0.4,1) [ $\phi = 0.9, \theta = 0.1$ ]					
$n_1 = n_2$		512	1024	2048	4096
$J = 4$	<i>KSM</i>	0.86	0.84	0.8	0.81
$J = 4$	<i>CVM</i>	0.81	0.76	0.78	0.76
$J = 5$	<i>KSM</i>	0.94	0.94	0.9	0.92
$J = 5$	<i>CVM</i>	0.93	0.92	0.96	0.91

**Table 1.11.** Power of KSM – CVM two ARFIMA(1,d,1) with same AR and MA part but two different values of memory parameter  $d$ .



**Fig. 1.3.** Empirical power of KSM( $d = 4$ ) for white noise, AR, MA and ARFIMA processes.

**Estimation of the change point in the original process.**

We know that for each scale  $j$ , the number  $n_j$  of wavelet coefficients is  $n_j = 2^{-j}(n - T + 1) - T + 1$ . If we denote by  $k_j$  the change point in the wavelet coefficients at scale  $j$  and  $k$  the change point in the original signal, then  $k = 2^j(k_j + T - 1) + T - 1$ . In this paragraph, we estimate the change point in the generalized spectral density of a process when it exists and give its 95% confidence interval. For that, we proceed as before. We consider an observation obtained by concatenation of  $n_1$  observations from a first process and  $n_2$

observations from a second process, independent from the first one and having a different spectral density. The length of the resulting observations is  $n = n_1 + n_2$ . we estimate the change point in the process and we present the result for  $n_1 = 512, 1024, 4096, 8192, n_2 = 512, 2048, 8192, J = 3$ , the statistic  $CVM$ , two  $AR$  processes with different values of the autoregressive coefficient and two  $ARFIMA$  with same moving average and same autoregressive coefficients but different values of the memory parameter  $d$ . For 10000 simulations, the bootstrap confidence intervals obtained are set in the tables below. we give also the empirical mean and the median of the estimated change point.

- $[AR(1), \phi = 0.9]$  and  $[AR(1), \phi = 0.5]$

$n_1$	512	512	512	1024	4096	8192
$n_2$	512	2048	8192	1024	4096	8192
$MEAN_{CVM}$	478	822	1853	965	3945	8009
$MEDIAN_{CVM}$	517	692	1453	1007	4039	8119
$IC_{CVM}$	[283,661]	[380,1369]	[523,3534]	[637,1350]	[3095,4614]	[7962,8825]

**Table 1.12.** Estimation of the change point and confidence interval at 95% in the generalized spectral density of a process which is obtain by concatenation of two  $AR(1)$  processes.

- $[ARFIMA(1, 0.2, 1)]$  and  $[ARFIMA(1, 0.3, 1)]$ , with  $\phi = 0.9$  and  $\theta = 0.2$

$n_1$	512	512	512	1024	4096	8192
$n_2$	512	2048	8192	1024	4096	8192
$MEAN_{CVM}$	531	1162	3172	1037	4129	8037
$MEDIAN_{CVM}$	517	1115	3215	1035	4155	8159
$IC_{CVM}$	[227,835]	[375,1483]	[817,6300]	[527,1569]	[2985,5830]	[6162,9976]

**Table 1.13.** Estimation of the change point and confidence interval at 95% in the generalized spectral density of a process which is obtain by concatenation of two  $ARFIMA(1,d,1)$  processes.

We remark that the change point belongs always to the considered confidence interval excepted for  $n_1 = 512, n_2 = 8192$  where the confidence interval is

[523, 3534] and the change point  $k = 512$  doesn't belong it. One can noticed that when the size of the sample increases and  $n_1 = n_2$ , the interval becomes more accurate. However, as expected, this interval becomes less accurate when the change appears either at the beginning or at the end of the observations.

## References

- D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991. ISSN 0012-9682.
- I. Berkes, L. Horvatz, Kokoszka P., and Qi-Man Shao. On discriminating between long-range dependence and change in mean,. *The annals of statistics*, 34(3):1140–1165, 2006.
- R. N. Bhattacharya, Vijay K. Gupta, and Ed Waymire. The Hurst effect under trends. *Journal of Applied Probability*, 20(3):649–662, 1983. ISSN 00219002. URL <http://www.jstor.org/stable/3213900>.
- P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.
- D. C. Boes and J. D. Salas. Nonstationarity of the Mean and the Hurst Phenomenon. *Water Resources Research*, 14:135–143, 1978. doi: 10.1029/WR014i001p00135.
- B. E. Brodsky and B. S. Darkhovsky. *Non-parametric statistical diagnosis*, volume 509 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2000. ISBN 0-7923-6328-0. Problems and methods.
- P. Carmona, F. Petit, J. Pitman, and M. Yor. On the laws of homogeneous functionals of the Brownian bridge. *Studia Sci. Math. Hungar.*, 35(3-4): 445–455, 1999. ISSN 0081-6906.
- A. Cohen. *Numerical analysis of wavelet methods*, volume 32 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 2003. ISBN 0-444-51124-5.
- F. X. Diebold and A. Inoue. Long memory and regime switching. *J. Econometrics*, 105(1):131–159, 2001. ISSN 0304-4076.

- L. Giraitis, P. Kokoszka, R. Leipus, and G. Teyssière. Rescaled variance and related tests for long memory in volatility and levels. *Journal of econometrics*, 112(4):265–294, 2003. ISSN 0090-5364.
- C. W. Granger and N. Hyung. Occasional structural breaks and long memory. *Journal of Empirical Finance*, 11:99–114, 1999.
- J. Hidalgo and P. M. Robinson. Testing for structural change in a long-memory environment. *J. Econometrics*, 70(1):159–174, 1996. ISSN 0304-4076.
- C. M. Hurvich, G. Lang, and P. Soulier. Estimation of long memory in the presence of a smooth nonparametric trend. *J. Amer. Statist. Assoc.*, 100(471):853–871, 2005. ISSN 0162-1459.
- C. Inclan and G. C. Tiao. Use of cumulative sums of squares for retrospective detection of changes of variance. *American Statistics*, 89(427):913–923, 1994. ISSN 0012-9682.
- J. Kiefer.  $K$ -sample analogues of the Kolmogorov-Smirnov and Cramér-V. Mises tests. *Ann. Math. Statist.*, 30:420–447, 1959. ISSN 0003-4851.
- V. Klemes. The Hurst Phenomenon: A Puzzle? *Water Resources Research*, 10:675–688, 1974. doi: 10.1029/WR010i004p00675.
- S. Mallat. *A wavelet tour of signal processing*. Academic Press Inc., San Diego, CA, 1998. ISBN 0-12-466605-1.
- T. Mikosch and C. Stărică. Changes of structure in financial time series and the Garch model. *REVSTAT*, 2(1):41–73, 2004. ISSN 1645-6726.
- E. Moulines, F. Roueff, and M. S. Taqqu. On the spectral density of the wavelet coefficients of long memory time series with application to the log-regression estimation of the memory parameter. *J. Time Ser. Anal.*, 28(2), 2007.
- E. Moulines, F. Roueff, and M.S. Taqqu. A wavelet Whittle estimator of the memory parameter of a non-stationary Gaussian time series. *Ann. Statist.*, 36(4):1925–1956, 2008.
- W. K. Newey and K. D. West. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987. ISSN 0012-9682.
- W. K. Newey and K. D. West. Automatic lag selection in covariance matrix estimation. *Rev. Econom. Stud.*, 61(4):631–653, 1994. ISSN 0034-6527.

- J. Pitman and M. Yor. The law of the maximum of a Bessel bridge. *Electron. J. Probab.*, 4:no. 15, 35 pp. (electronic), 1999. ISSN 1083-6489.
- K.W. Potter. Evidence of nonstationarity as a physical explanation of the Hurst phenomenon. *Water Resources Research*, 12:1047–1052, 1976.
- A. Ramachandra Rao and G. H. Yu. Detection of nonstationarity in hydrologic time series. *Management Science*, 32(9):1206–1217, 1986. ISSN 00251909. URL <http://www.jstor.org/stable/2631546>.
- B. Whitcher, S. D. Byers, P. Guttorp, and Percival D. Testing for homogeneity of variance in time series : Long memory, wavelets and the Nile river. *Water Resources Research*, 38(5), 2002.