T E C H N I C A L
R E P O R T

0316

LOCALLY ADAPTIVE ESTIMATION OF SPARSE EVOLUTIONARY WAVELET SPECTRA

S. VAN BELLEGEM and R. VON SACHS

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Locally adaptive estimation of sparse evolutionary wavelet spectra

Sébastien Van Bellegem[1,3] Rainer von Sachs[2,3]

May 9, 2003

Abstract

We introduce a wavelet-based model of local stationarity. This model enlarges the class of *locally stationary wavelet processes* and contains processes whose spectral density function may change very suddenly in time. A notion of time-varying *wavelet spectrum* is uniquely defined as a wavelet-type transform of the autocovariance function with respect to so-called *autocorrelation wavelets*. This leads to a natural representation of the autocovariance which is localised on scales. One particularly interesting subcase arises when this representation is sparse, meaning that the nonstationary autocovariance process may be decomposed in the autocorrelation wavelet basis using few coefficients. We present a new test of sparsity for the wavelet spectrum. It is based on a non-asymptotic result on the deviations of a functional of a periodogram. The power of the test is discussed. We also present another application of this result given by the pointwise adaptive estimation of the wavelet spectrum. Properties of this estimator in homogeneous and inhomogeneous regions of the wavelet spectrum are studied.

Keywords: Local stationarity, nonstationary time series, wavelet spectrum, autocorrelation wavelet, change-point, pointwise adaptive estimation

Abbreviated title: Estimation of wavelet spectra

AMS 1991 subject classification: Primary 62M10; secondary 60G15, 62G10, 62G05

[1] *Correponding author.* Research Fellow of the National Fund for Scientific Research (F.N.R.S.). Université catholique de Louvain, Institut de statistique, Louvain-la-Neuve, Belgium. `vanbellegem@stat.ucl.ac.be`

[2] Professor. Université catholique de Louvain, Institut de statistique, Louvain-la-Neuve, Belgium. `vonsachs@stat.ucl.ac.be`

[3] Financial support from the contract 'Projet d'Actions de Recherche Concertes' nr 98/03–217 from the Belgian government, and from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

1 Introduction

The spectral analysis of time series is a large field presenting a great interest from both theoretical and practical viewpoints. The fundamental starting point of this analysis is the *Cramér representation*, stating that all second-order zero-mean stationary process X_t, $t \in \mathbb{Z}$ may be written

$$X_t = \int_{[-\pi,\pi)} A(\omega) \exp(i\omega t) dZ(\omega), \qquad t \in \mathbb{Z} \,, \tag{1.1}$$

where $A(\omega)$ is the amplitude of the process X_t and $dZ(\omega)$ is an orthonormal increment process, i.e. $\mathrm{E}(dZ(\omega), \overline{dZ(\mu)}) = \delta_0(\omega - \mu)$, see Priestley (1981). Correspondingly, under mild conditions, the autocovariance function can be expressed as

$$c_X(\tau) = \int_{[-\pi,\pi)} f_X(\omega) \exp(i\omega\tau) d\omega,$$

where f_X is the *spectral density* of X_t.

There is not a unique way to relax the assumption of stationarity, i.e. to define a second-order process with a time-depending spectrum. However, this modelling is a theoretical challenge which may be helpful in practice, since a lot of studies have shown that models with evolutionary spectrum or time-varying parameters are necessary to explain some observed data, even over short periods of time. Examples may be found in numerous fields, such as economics (Swanson and White, 1997; Los, 2000), biostatistics (Ombao *et al.*, 2002) or meteorology (Nason and Sapatinas, 2002) to name but a few.

Among the different possibilities for modelling nonstationary second-order processes, we can emphasize the approaches consisting in a modification of the Cramér representation (1.1). Different modifications of (1.1) are possible. First, we can replace the process $dZ(\omega)$ by a nonorthonormal process, leading for instance to the *harmonizable* processes (Lii and Rosenblatt, 2002). A second possibility is to replace the amplitude function $A(\omega)$ by a time-varying version $A_t(\omega)$ and assume a slow change of $A_t(\omega)$ over time. Such approach is followed to define *oscillatory* processes (Priestley, 1965).

However, a major statistical drawback of the oscillatory processes is the intrinsic impossibility to construct an asymptotic theory for consistency and inference. To overcome this problem, Dahlhaus (1997) introduced the class of *locally stationary processes*, in which the transfer function is rescaled in time. In this approach, a doubly-indexed process is defined as

$$X_{t,T} = \int_{[-\pi,\pi)} A\left(\frac{t}{T}, \omega\right) \exp(i\omega t) dZ(\omega), \quad t = 0, \dots, T-1, \quad T > 0 \,, \tag{1.2}$$

where the transfer function $A(z, \omega)$ is defined on $(0,1) \times [-\pi, \pi)$. Dahlhaus (1997, 2000) investigated statistical inference for such processes, with a discussion on maximum likelihood, Whittle and least squares estimates, and showed that asymptotic results when T tends to infinity can be considered. However, in this setting, letting T tends to infinity has not the usual meaning of "looking into the future", but means that we have in the sample $X_{0,T}, \dots, X_{T-1,T}$ more information about the local structure of $A(z, \omega)$. This formalism is analogous to the nonparametric regression problems, for which "asymptotic" means an ideal knowledge about the local structure of the underlying curve.

In this article, we focus on a class of doubly-indexed locally stationary processes defined by replacing the harmonic system $\{\exp(i\omega t)\}$ in (1.2) by a wavelet basis. By this way, we move from a *time-frequency* representation to a *time-scale* representation of the nonstationary process. Because wavelets systems are well localized in time and frequency, they appear more natural to model the time-varying spectra of nonstationary processes. Indeed, by the uncertainty principle, allowing the spectrum to be time-varying implies that we lose resolution in the frequency domain. As wavelets decompose the frequency domain into discrete scales, they offer a well-adapted system to achieve the trade-off resolution between time and frequency (Vidakovic, 1999).

The class of locally stationary wavelet processes studied in this article was initially introduced by Nason, von Sachs and Kroisandt (2000). Their definition of wavelet processes involves a time-varying amplitude which is smoothly varying and continuous as a function of time. One first goal of this article is to extend this definition to the the case of time-varying amplitudes with possibly *discontinuous* behaviour in time. This adds some technical difficulties in the proof of our results but we believe the gain due to this extension to be crucial. Our new definition now includes more important examples of nonstationary processes. For instance, this extension of the definition is needed if we wish to model a nonstationary process built as a concatenation of different stationary processes. Moreover, wavelet processes can now be used for the analysis of intermittent phenomena, such as transients followed by regions of smooth behaviour.

Our definition of wavelet processes is presented in Section 2, where we also define their *evolutionary spectrum*. This spectrum is a function of time and scales, and measures the power of the process at a particular time and scale.

In Section 3, we recall that the simultaneous localisation of wavelets in time and scale leads to possibly very sparse representations of evolutionary spectra. A precise explanation of this phenomenon is given in Section 3, but we want now to mention that by "sparsity", we mean that only few segments of the time-scale spectrum are nonzero. This knowledge is important for the exploratory analysis of nonstationary signals, because, at a given time, the analyst can focus on active scales only. Moreover, a scale may be active (i.e. nonzero) at a given time and not active at an other time, and this evolution corresponds to physical changes in the process.

We derive a test statistic, based on a functional of the so-called wavelet periodogram. The test statistic is actually a quadratic form of the increments, which are assumed to be Gaussian, and the test rule is provided through a nonasymptotic result on the deviation of the quadratic form of Gaussian processes. However, the variance of the test statistic depends crucially on the unknown spectrum, and we present a pre-estimator of this nuisance parameter. Finally, we establish a nonasymptotic approximation of the distribution of the test statistics. This approximation is constructed with our pre-estimator of the variance. A theoretical study of the power of the test concludes Section 3. In particular, we discuss the consistency and the local alternatives of the proposed test procedure.

In the following Section 4, we show how the result of Section 3 may be useful for other tasks, as the pointwise adaptive estimation of the wavelet spectrum. In this section, we derive an estimation procedure following the local adaptive method of Lepski (1990). The behaviour of this estimator is briefly studied in the case where the evolutionary wavelet spectrum is regular or irregular near the point of estimation.

Finally, some possible applications and extensions are presented in Section 5. The proofs are deferred to appendices. In particular, Appendix A derives some properties of the autocorrelation wavelet system, which are of independent interest.

2 Locally stationary wavelet processes

The wavelet system used to build locally stationary processes is a non-decimated system of compactly supported and discrete wavelets. We first briefly recall some points about this system of wavelets, and then give a definition of the wavelet processes and wavelet spectra.

2.1 Nondecimated wavelet system

The local basis functions used in the representation of LSW processes are a set of *discrete non-decimated wavelets* $\{\psi_{jk}, j = -1, -2, \dots; k \in \mathbb{Z}\}$. We refer to Vidakovic (1999) for a review on wavelet theory and its applications in statistics, and to Nason and Silverman (1995) for a detailed introduction to the non-decimated wavelet transform. Let us simply recall that, in contrast to the discrete wavelet transform, the discrete non-decimated wavelets at all scales $j < 0$ can be shifted to any location defined by the finest resolution scale, determined by the observed data. As a consequence, this construction leads to an overcomplete system of the space of square summable sequences $\ell^2(\mathbb{Z})$. The wavelets considered in this article are assumed to be compactly supported in time and we will denote by $\mathcal{L}_j$ the length of the support of ψ_{j0}, i.e. $\mathcal{L}_j := |\operatorname{supp}\psi_{j0}|$. Straightforward consequences of the non-decimated wavelet system imply that $|\operatorname{supp}\psi_{jk}| = \mathcal{L}_j = (2^{-j}-1)(\mathcal{L}_{-1}-1)+1$ for all $j < 0$. Observe also that, as in Nason *et al.* (2000), we departed from the usual wavelet numbering scheme. The data live on scale zero, and scale -1 is the scale which contains the finest resolution wavelet detail. Then, the support of the wavelet on the finest scale remains constant with respect to T.

For ease of presentation, recall the simplest discrete non-decimated system, called the *Haar system*, given by

$$\psi_{j0}(t) = 2^{j/2}\mathbb{I}_{\{0,1,\dots,2^{-j-1}-1\}}(t) - 2^{j/2}\mathbb{I}_{\{2^{-j-1},\dots,2^{-j}-1\}}(t) \quad \text{for } j = -1, -2, \dots \text{ and } t \in \mathbb{Z},$$

where $\mathbb{I}_{\mathcal{A}}(t)$ is 1 if $t \in \mathcal{A}$ and 0 otherwise. The shifted version of $\psi_{j0}(t)$ is given by $\psi_{jk}(t) = \psi_{j0}(t-k)$ for all $k \in \mathbb{Z}$.

2.2 The process and its evolutionary wavelet spectrum

As we will note below, our definition of locally stationary wavelet processes differs from the original definition of Nason *et al.* (2000) as we only impose a total variation condition on the amplitudes instead of a Lipschitz condition.

Definition 1. *A sequence of doubly-indexed stochastic processes* $X_{t,T}$ $(t = 0, \dots, T-1, T > 0)$ *with mean zero is in the class of* locally stationary wavelet processes (LSW processes) *if there exists a representation in the mean-square sense*

$$X_{t,T} = \sum_{j=-\infty}^{-1} \sum_{k=0}^{T-1} w_{jk;T}\ \psi_{jk}(t)\ \xi_{jk}, \tag{2.1}$$

where $\{\psi_{jk}(t) = \psi_{j0}(t-k)\}_{jk}$ *with* $j < 0$ *is a discrete non-decimated family of wavelets based on a mother wavelet* $\psi(t)$ *of compact support, and such that:*

1. ξ_{jk} *is a random orthonormal increment sequence with* $\mathrm{E}\xi_{jk} = 0$ *and* $\operatorname{Cov}(\xi_{jk}, \xi_{\ell m}) = \delta_{j\ell}\,\delta_{km}$ *for all* j, ℓ, k, m*, where* $\delta_{j\ell} = 1$ *if* $j = \ell$ *and 0 elsewhere;*

2. *For each* $j \leqslant -1$*, there exists a function* $W_j(z)$ *on* $(0,1)$ *possessing the following properties:*

   (a) $\sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty$ *uniformly in* $z \in (0,1)$*,*

   (b) There exists a sequence of constants C_j *such that for each* T

$$\sup_{k=0,\dots,T-1} \left| w_{jk;T}^2 - W_j^2\left(\frac{k}{T}\right) \right| \leqslant \frac{C_j}{T}\ , \tag{2.2}$$

(c) $W_j^2(z)$ *is bounded by* L_j *in the total variation norm, i.e.*

$$\begin{aligned}\mathrm{TV}\left(W_j^2\right) &:= \sup\left\{\sum_{i=1}^{I}\left|W_j^2\left(a_i\right)-W_j^2\left(a_{i-1}\right)\right| : 0<a_0<\ldots<a_I<1, I\in\mathbb{N}\right\}\\ &\leqslant L_j,\end{aligned} \tag{2.3}$$

(d) The constants C_j *and* L_j *are such that*

$$\sum_{j=-\infty}^{-1}\mathcal{L}_j(\mathcal{L}_jL_j+C_j)\leqslant\rho<\infty \tag{2.4}$$

where $\mathcal{L}_j=|\operatorname{supp}\psi_{j0}|=(2^{-j}-1)(\mathcal{L}_{-1}-1)+1$.

LSW processes use wavelets to decompose a stochastic process with respect to an orthogonal increment process in the time-scale plane. Due to the overcompleteness of the non-decimated basis, LSW processes are not uniquely determined by the sequence $\{w_{jk;T}\}$. However, we can build a theory which ensures the existence of a unique wavelet spectrum. This property is a consequence of the local stationarity setting which introduces a *rescaled time* $z=t/T\in(0,1)$ on which $W_j(z)$ is defined. The rescaled time permits increasing amounts of data about the local structure of $W_j(z)$ to be collected as the observed time T tends to infinity. Even though LSW processes are not uniquely determined by the sequence $\{w_{jk;T}\}$, the model allows to identify (asymptotically) the model coefficients determined by uniquely defined $W_j(z)$. Then, the *evolutionary wavelet spectrum* of an LSW process $\{X_{t,T}\}_{t=0,\ldots,T-1}$, with respect to ψ, is given by

$$S_j(z)=|W_j(z)|^2,\ z\in(0,1) \tag{2.5}$$

and is such that, by definition of the process, $S_j(z)=\lim_{T\to\infty}|w_{j,[zT];T}|^2$ for all $z\in(0,1)$, and by Definition 1, $\sum_{j=-\infty}^{-1}S_j(z)<\infty$ uniformly in $z\in(0,1)$.

The evolutionary wavelet spectrum $S_j(z)$ is related to the time-depending autocorrelation function of the LSW process. Observe that the autocovariance function of an LSW process can be written as

$$c_{X,T}(z,\tau)=\operatorname{Cov}\left(X_{[zT],T},X_{[zT]+\tau,T}\right)$$

for $z\in(0,1)$ and τ in $\mathbb{Z}$, and where $[\,\cdot\,]$ denotes the integer part of a real number. The next result shows that this autocovariance converges asymptotically to a *local covariance* defined by

$$c_X(z,\tau)=\sum_{j=-\infty}^{-1}S_j(z)\Psi_j(\tau) \tag{2.6}$$

where $\Psi_j(\tau)=\sum_{k=-\infty}^{\infty}\psi_{jk}(0)\psi_{jk}(\tau)$ is the *autocorrelation wavelet* function.

Proposition 1. *Under the assumptions of Definition 1, if* $T\to\infty$

$$\sum_{\tau=-\infty}^{\infty}\int_0^1 dz\ |c_{X,T}(z,\tau)-c_X(z,\tau)|=O\left(T^{-1}\right)$$

for all LSW process.

Appendix A presents some properties of the autocorrelation wavelet system appearing in (2.6). Like wavelets themselves, this system enjoys good localisation properties. Consequently, we observe that equation (2.6) is a multiscale decomposition of the autocovariance structure of the process over time: The larger the wavelet spectrum $S_j(z)$ is at a particular scale j and point z in the rescaled time, the more dominant is the contribution of scale j in the variance at time z. Thus, the evolutionary wavelet spectrum describes the distribution of the (co)variance at a particular scale and time location.

It is worth mentioning that a stationary process with an absolutely summable autocovariance function is an LSW process (Nason *et al.*, 2000, Proposition 3). Stationarity is characterized by a wavelet spectrum which is constant over time: $S_j(z) = S_j$ for all $z \in (0, 1)$. However, our motivation to study LSW processes lies in the modelling of time-varying spectra. The regularity of the wavelet spectrum in time is determined by the smoothness of $W_j(z)$ with repect to z. In Nason *et al.* (2000), this function is assumed to be Lipschitz continuous in time. In our definition of LSW processes, we only require the total variation norm of W_j^2 to be bounded. This weaker assumption is not only considered in order to work with minimal assumptions, but also to allow a discontinuous evolution of the wavelet spectrum in time. Consequently, our definition of nonstationary processes includes many interessing processes, as piecewise stationary signals for instance. Figure 1 shows a simulated example of such a nonstationary process.

Figure 1 about here

3 Testing sparsity of a wavelet spectrum

3.1 Sparsity

It is well-known that wavelets are suited to decompose certain inhomogeneous signals into a sparse wavelet coefficients vector. One goal of this article is to study how this important property of wavelets may be exploited in our context of nonstationary covariance modelling.

In the multiscale representation (2.6) the coefficients $S_j(z)$ are depending on the continuous recaled time $z \in (0, 1)$. As $\Psi_j(0) = 1$ for all scales j, (2.6) decomposes the instantaneous variance as

$$c_X(z, 0) = \sum_{j=-\infty}^{-1} S_j(z). \tag{3.1}$$

If we assume this variance to be non zero, it then follows that, at each time z, there exists a scale j where $S_j(z)$ is non zero. If only *few* scales are non zero for each z, we say that the wavelet spectrum is *sparse*. There are many approaches in the literature where the notion of sparsity is quantified. In the context of wavelet decomposition of signals, we refer to Abramovich *et al.* (2000) and the references therein for some possible definitions. In the present work, we do not need a precise definition for quantifying the sparsity. We only think about wavelet spectra $S_j(z)$ which are non zero along only some segments of a few scales as in Figure 1.

As already observed in Nason *et al.* (2000), the possibility of having a sparse representation of wavelet spectra is a major advantage of LSW processes, in comparison with other locally stationary models. Sparsity is a very attractive property since only few coefficients in the multiscale representation have to be estimated. Moreover, the analysis of sparse wavelet spectra is easier, for instance if we want to detect significant variations in the multiscale structure of the process (co)variance.

In this section, we want to develop and study a statistical test of sparsity for a wavelet spectrum. This test is *local* in the sense that we will test the null hypothesis

$$\mathrm{H}_0 : S_j(z) = 0 \text{ for a fixed scale } j < 0 \text{ and for all } z \in \mathcal{R}, \tag{3.2}$$

where $\mathcal{R} \subseteq (0,1)$ is an interval with non zero measure. It is then possible to test if, for instance, a whole scale is "active" or not, or if it is non zero before or after a fixed time point.

The next subsection defines a preliminary estimator of the wavelet spectrum. Then, derivation of a test statistic is considered in subsections 3.3 and 3.4. This test is discussed in Subsection 3.5, where we also study its power under some alternatives.

3.2 A preliminary estimate: The corrected wavelet periodogram

An estimator of the wavelet spectrum is constructed by taking the squared empirical coefficients from the non-decimated transform:

$$I_{j;T}\left(\frac{k}{T}\right) = \left(\sum_{t=0}^{T-1} X_{t,T}\psi_{jk}(t)\right)^2 \qquad j = -1, \dots, -\log_2 T; k = 0, \dots, T-1.$$

$I_{j;T}(z)$ is called the *wavelet periodogram*, as it is analogous to the formula for the classical periodogram in traditional Fourier spectral analysis of stationary processes (Brillinger, 1975).

Some asymptotic properties of this estimator have been studied by Nason *et al.* (2000), who showed that the wavelet periodogram is *not* an asymptotic unbiased estimator of the wavelet spectrum. Indeed, Proposition 4 of Nason *et al.* (2000) states that, for all fixed scales $j < 0$,

$$\mathrm{E}\left[\sum_{\ell=-\log_2 T}^{-1} A_{j\ell}^{-1} I_{\ell;T}(z)\right] - S_j(z) = O(T^{-1}), \tag{3.3}$$

uniformly in $z \in (0,1)$, where the matrix $A = (A_{j\ell})_{j,\ell<0}$ is defined by

$$A_{j\ell} := \langle \Psi_j, \Psi_\ell \rangle = \sum_\tau \Psi_j(\tau)\Psi_\ell(\tau).$$

Note that the matrix $A_{j\ell}$ is not simply diagonal since the autocorrelation wavelet system $\{\Psi_j\}$ is not orthogonal. Nason *et al.* (2000) proved the invertibility of A if $\{\Psi_j\}$ is constructed using Haar wavelets. If other compactly supported wavelets are used, numerical results suggest that the invertibility of A still holds, but a complete proof of this result has not been established yet. As we need the invertibility of A in our following results, from now on we restrict ourselves to Haar wavelets, but we conjecture that all results remain valid for more general Daubechies wavelets.

Equation (3.3) motivates the definition of a *corrected wavelet periodogram*

$$L_{j;T}\left(\frac{k}{T}\right) = \sum_{\ell=-\log_2 T}^{-1} A_{j\ell}^{-1} \left(\sum_{t=0}^{T-1} X_{t,T}\psi_{\ell k}(t)\right)^2 \tag{3.4}$$

which is an asymptotically unbiased estimator for the evolutionary wavelet spectrum.

Remark 1. The asymptotic bias of the wavelet periodogram is a consequence of the overcompleteness of the non-decimated wavelet system $\{\psi_{jk}\}$. One could ask if it would not be easier to define LSW processes using a decimated wavelet system because, for this system, the matrix A reduces to the identity. Unfortunately, the answer is negative: The use of non-decimated wavelets, as described in von Sachs *et al.* (1997), would not allow to write the local autocovariance function as a wavelet-type transform of an evolutionary spectrum, as in (2.6). Moreover, classical stationary processes are not included in the model based on decimated wavelets.

3.3 Derivation of the test statistic and its properties

Suppose we want to test (3.2), i.e. to check if the wavelet spectrum is zero at a fixed scale j and on a given segment of time $\mathcal{R} = (s_1, s_2) \subseteq (0,1)$ for $s_1 < s_2$. Under the null (3.2), the averaged wavelet spectrum

$$Q_{j,\mathcal{R}} = |\mathcal{R}|^{-1} \int_{\mathcal{R}} dz\, S_j(z) \tag{3.5}$$

is zero. If we observe $\underline{X}_T = (X_{0,T}, \ldots, X_{T-1,T})'$, a natural estimate of $Q_{j,\mathcal{R}}$ is

$$Q_{j,\mathcal{R};T} = |\mathcal{R}T|^{-1} \sum_{k \in \mathcal{R}T} L_{j;T}\left(\frac{k}{T}\right) \tag{3.6}$$

where $L_{j;T}(k/T)$ is the corrected wavelet periodogram (3.4) and $k \in \mathcal{R}T$ means $k/T \in \mathcal{R}$. $Q_{j,\mathcal{R};T}$ is the test statistic we use to test H_0. In this section, we will study the statistical properties of $Q_{j,\mathcal{R};T}$ under a set of assumptions.

Assumption 1. The autocovariance function $c_{X,T}$ and the local autocovariance function c_X of the LSW process are such that

$$\|c_{X,T}\|_{1,\infty} := \sum_{\tau=-\infty}^{\infty} \sup_{t=0,\ldots,T-1} \left| c_{X,T}\left(\frac{t}{T}, \tau\right) \right| \text{ is uniformly bounded in } T, \tag{3.7}$$

and

$$\|c_X\|_{1,\infty} := \sum_{\tau=-\infty}^{\infty} \sup_{z \in (0,1)} |c_X(z,\tau)| < \infty. \tag{3.8}$$

This assumption is needed to control the spectral norm of the covariance matrix of the process (Lemma 5 in Appendix B). For a stationary process, it reduces to absolute summability of the autocovariance of the process (short memory property).

Assumption 2. There exists an $\varepsilon > 0$ such that, for all $z \in (0,1)$, $\sum_{j=-\infty}^{-1} S_j(z) \geqslant \varepsilon$.

According to equation (3.1), the sum over scales of $S_j(z)$ is the local variance of the process at time $[zT]$, and this assumption says that the local variance of the process is nowhere zero.

Assumption 3. The increment process $\{\xi_{jk}\}$ in Definition 1 is Gaussian.

This assumption allows substantial simplifications in the proofs. It is also assumed to establish some results in Nason *et al.* (2000) and Fryźlewicz *et al.* (2003). However, Fryźlewicz (2002) mentions that non-Gaussian increment processes would be more appropriate to capture some stylised facts of economic processes, such as the leptokurtic behaviour of the data. To this end, the extension of our results to non-Gaussian processes would be a feasible task using the methodology presented in Neumann and von Sachs (1997) or Spokoiny (2001) for instance.

The following proposition describes the asymptotic properties of $Q_{j,\mathcal{R};T}$.

Proposition 2. *Suppose Assumption 1 to 3 hold true. For all LSW process (Definition 1), and for all* $\mathcal{R} \subseteq (0,1)$,

$$\begin{aligned} \mathrm{E}Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}} &= \frac{K_0\, 2^{j/2}}{\sqrt{T}} \sum_{m=-\log_2 T}^{-1} \mathcal{L}_m\, \mathrm{TV}(S_m) + O\left(2^{j/2}\, |\mathcal{R}T|^{-1}\right) \\ &= O\left(\frac{2^{j/2}}{\sqrt{T}}\right), \end{aligned} \tag{3.9}$$

for all $j = -1, \dots, -J_T$ with $J_T = o_T(\log_2 T)$, and where K_0 is a constant independent of j, T and $|\mathcal{R}|$. Moreover, if Assumptions 1 to 3 hold, then there exists $T_0 > 1$ such that, for all $T \geqslant T_0$,

$$K_1\, 2^{2j}\, |\mathcal{R}T|^{-1}\,(1 + o_T(1)) \leqslant \operatorname{Var} Q_{j,\mathcal{R};T} \leqslant K_2\, 2^j |\mathcal{R}|^{-2} T^{-1}$$

for all $j = -1, \dots, -J_T$ with $J_T = o_T(\log_2 T)$, and where K_1 and K_2 are two constants independent of j, T and $|\mathcal{R}|$.

The proof of this proposition is in Appendix B.2. Note that the squared bias and the variance of the estimator have the same rate of convergence. This phenomenon is due to the nonstationary behaviour of the process. Indeed, for a stationary process, the total variation norm of S_m is zero at all scales, and then the rate of the bias is T^{-1}. This is not the case for a general nonstationary process: When the wavelet spectrum is not constant over time, an additional term resulting from nonstationarity reduces considerably this rate of convergence. Moreover, even we are dealing with a *local* estimator of the wavelet spectrum at a fixed scale $j < 0$ and a fixed time interval $\mathcal{R}$, the nonstationarity term in the bias involves the variation of the *global* wavelet spectrum. This may be observed in equation (3.9), which involves a sum over all scales $m = -1, \dots, -\log_2 T$ and the total variation norm of all S_m over the whole rescaled time interval $(0, 1)$.

This slow rate of convergence of the bias poses a problem to establish the asymptotic normality of $Q_{j,\mathcal{R};T}$. In the next proposition, we circumvent this problem and derive a non asymptotic exponential bound for the deviation of $Q_{j,\mathcal{R};T}$.

Proposition 3. *Assume that (3.7) and Assumption 1 to 3 hold. If $\sigma^2_{j,\mathcal{R},T} = \operatorname{Var} Q_{j,\mathcal{R};T}$, then, for all $\eta > 0$ and for all scales $j = -1, \dots, -J_T$, where $J_T = o_T(\log_2 T)$,*

$$\Pr\left(|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}| > \sigma_{j,\mathcal{R},T}\eta\right) \leqslant c_0 \exp\left\{-\frac{1}{8} \cdot \frac{\eta^2}{1 + \frac{\eta}{|\mathcal{R}T|\sigma_{j,\mathcal{R},T}} L_j + \frac{2^{j/2}\eta\nu}{|\mathcal{R}|\sqrt{T}\sigma_{j,\mathcal{R};T}}\left(\|c_X\|_{1,\infty} + c_1\rho\right)}\right\}$$

with the positive constants $c_0 = 1 + e$ and $c_1 = (2 + \sqrt{2})/2$, where ρ is given in Definition 1 and where ν is a universal positive constant depending only on the wavelet ψ.

The proof of this proposition is to be found in Appendix B.3. This proposition gives a non asymptotic approximation for the distribution of the test statistics $Q_{j,\mathcal{R};T}$. It can be used in order to construct a test rule, i.e. to choose η such that the exponential function in the proposition is the nominal level of the test (see Section 3.5 below, where the test rule is given explicitely). From an asymptotic viewpoint, i.e. as $T \to \infty$, we note that this exponential bound does not tend to zero, meaning that the standardised statistic $Q_{j,\mathcal{R},T}$ is asymptotically non degenerated. This phenomenon is well-known in the context of pointwise estimation, see Lepski (1990) and Brown and Low (1996). In order to have a consistent result when $T \to \infty$, it is then necessary to impose that $\eta = \eta_T$ grows with T. The appropriate rate for η_T is derived in the next corollary. The proof is given in Appendix B.3 and is essentially based on the bounds derived in Proposition 2.

Corollary 1. *Under the assumptions of Propositions 2 and 3, if k_T tends to infinity and is such that*

$$k_T^{-1} 2^{J_T/2} \log_2 J_T = o_T(1), \tag{3.10}$$

then there exists a $T_0 > 1$ such that, for all $T \geqslant T_0$,

$$\Pr\left(\sup_{-J_T \leqslant j < 0} |Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}| \geqslant k_T \sqrt{K_2 |\mathcal{R}T|^{-1}}\right) = o_T(1)$$

where K_2 is as in the assertion of Proposition 2.

Remark 2. An example of admissible rates is $J_T \sim \log_2 \log_2^2 T$ and $k_T \sim \log_2^2 T$. The sequence k_T will play a crucial role in Section 4.

3.4 Estimation of the variance

If we want to use Proposition 3 to test H_0, an estimator of the variance $\sigma^2_{j,\mathcal{R},T} = \operatorname{Var} Q_{j,\mathcal{R};T}$ is needed. This variance depends on the unknown autocovariance function of the LSW process in the following way (see Lemma 3 with equation (B.9)):

$$\sigma^2_{j,\mathcal{R},T} = 2\, \|U'_{j,\mathcal{R};T} \Sigma_T\|_2^2,$$

where Σ_T is the $T \times T$ (non-Toeplitz) covariance matrix of the LSW process $(X_{0,T}, \ldots, X_{T-1,T})'$, and $U_{j,\mathcal{R};T}$ is the $T \times T$ matrix with entry (s,t) equal to

$$U^{(j)}_{st} = |\mathcal{R}T|^{-1} \sum_{\ell=-\log_2 T}^{-1} A^{-1}_{j\ell} \sum_{k \in \mathcal{R}T} \psi_{\ell k}(s) \psi_{\ell k}(t).$$

We also denote by $\sigma_{s,s+u}$ the entry $(s, s+u)$ of the matrix Σ_T.

We will estimate $\sigma^2_{j,\mathcal{R},T}$ by:

$$\tilde{\sigma}^2_{j,\mathcal{R},T} = 2\, \|U'_{j,\mathcal{R};T} \tilde{\Sigma}_T\|_2^2$$

where $\tilde{\Sigma}_T$ is an estimate of the covariance matrix Σ_T. A first idea is to define the elements $\tilde{\sigma}_{s,s+u}$ of $\tilde{\Sigma}_T$ by plugging $Q_{j,\mathcal{R};T}$ into the local autocovariance function (2.6), i.e.

$$\tilde{\sigma}_{s,s+u} = \sum_{j=-\log_2 T}^{-1} Q_{j,\mathcal{R}(s);T} \Psi_j(u),$$

where $\mathcal{R}(s)$ denotes an interval which contains the time point s/T. However, the convergence in probability of $\tilde{\sigma}_{s,s+u}$ to $\sigma_{s,s+u}$ is not faster than the rate of $\sigma_{s,s+u}$ itself, and we need to modify the estimator in two ways.

(i) Assumption 1 indicates that the covariance $|\sigma_{s,s+u}|$ is small for large $|u|$. Then, following the method of Giurcanu and Spokoiny (2002), we set $\tilde{\sigma}_{s,s+u}$ to zero when $|u| \geqslant M_T$, for an appropriate sequence M_T tending to infinity with T;

(ii) It is necessary to control the distance in rescaled time between the spectrum $S_j(z)$, for $z \in \mathcal{R}(s)$, and $S_j(s/T)$. To do so, we allow the window $\mathcal{R}(s)$ to depend on T, which is denoted by $\mathcal{R}_T(s)$, in such a way that its length $|\mathcal{R}_T|$ shrinks to zero when T tends to infinity. This is analogous to the estimation of a regression function by kernel smoothing, where the window usually depends on the length of the data set.

With these two ingredients, we propose to estimate $\sigma_{s,s+u}$ by

$$\tilde{\sigma}_{s,s+u} = \sum_{j=-\log_2 T}^{-1} Q_{j,\mathcal{R}_T(s);T} \Psi_j(u) \mathbb{I}_{|u| \leqslant M_T}, \tag{3.11}$$

and the following assumption makes precise the appropriate rates for the sequences $|\mathcal{R}_T|$ and M_T.

Assumption 4. $|\mathcal{R}_T|$ tends to zero such that $|\mathcal{R}_T| = o_T(2^{-J_T})$, where $J_T = o_T(\log_2 T)$ and M_T tends to infinity such that $2^{J_T}|\mathcal{R}_T T|^{-1/2} M_T k_T \log_2^3 T = o_T(1)$.

With the rates given in Remark 2, admissible rates are $|\mathcal{R}_T| \sim \log_2^{-3} T$ and $M_T \sim \log_2^{\alpha} T$ with $\alpha > 0$. It is worth mentioning that, with this assumption, $|\mathcal{R}_T|$ shrinks to zero in the *rescaled* time, whereas, in the *observed* time, the interval length $|T\mathcal{R}_T|$ tends to infinity. This means that our estimate of $S_j(s/T)$ is built using an increasing amount of data in the observed time, but, at the same time, with an average around $S_j(s/T)$ in the rescaled time on a shrinking segment around s/T.

The next proposition shows that on the random set where the estimator $Q_{j,\mathcal{R}_T(s);T}$ is near $Q_{j,\mathcal{R}_T(s)}$, the estimator (3.11) has a good quality. Our proof of this proposition may be found in Appendix B.4 and needs the following technical assumption, which is a slightly stronger condition than the point 2(a) of Definition 1, in the sense that we need to control the decay of $S_j(z)$ with respect to j and uniformly in z.

Assumption 5. The evolutionary wavelet spectrum $S_j(z)$ is such that

$$\sum_{\ell=-\infty}^{-\log_2(T)-1} \sup_{z\in(0,1)} S_\ell(z) = O\left(T^{-1}\right).$$

Recall that the estimator (3.11) involves the scales -1 up to $-\log_2 T$. However, an explicit computation of the covariance $\mathrm{Cov}(X_{s,T}, X_{s+u,T})$ may be written along the lines of the proof of Proposition 2 and shows that this covariance involves all scales $j = -1, -2, \ldots$. Then, when we estimate this covariance function by $\tilde{\sigma}_{s,s+u}$, Assumption 5 helps to control the reminder of approximation at all scales lower than $-\log_2 T$.

Assumption 6. The local autocovariance function $c(z,\tau)$ is such that

$$\sum_{u=-\infty}^{\infty} \sup_z |c_X(z,u)| \mathbb{I}_{|u|>M_T} = o_T\left(2^{-J_T}\right).$$

This last assumption on the decay of the local autocovariance function uniformly in z, is very sensible in a context of short-memory processes, i.e. when $c(z,u)$ does not depend on z. With the rates specified above, a typical condition is to assume $|c_X(z,u)| \leqslant c \cdot 2^{-|u|}$ uniformly in $z \in (0,1)$.

Proposition 4. *Suppose Assumptions 1 to 6 hold and* k_T *is such that*

$$k_T^{-1} 2^{3J_T/2} \log_2 J_T = o_T(1). \qquad (3.12)$$

Then, there exists a positive number T_0 *and a random set* $\mathcal{A}$ *independent of* j *and such that* $\Pr(\mathcal{A}) \geqslant 1 - o_T(1)$ *and*

$$|Q_{j,\mathcal{R}_T(s);T} - Q_{j,\mathcal{R}_T(s)}| \leqslant k_T \sqrt{\frac{K_2}{|\mathcal{R}_T T|}}$$

for all $T > T_0$. *Moreover, on* $\mathcal{A}$,

$$2^{J_T - j}\, T\, |\tilde{\sigma}^2_{j,\mathcal{R},T} - \sigma^2_{j,\mathcal{R},T}| = o_P(1) \qquad (3.13)$$

holds for all $j = -1, \ldots, -J_T$, *where* $o_P(1)$ *does not depend on* $\mathcal{R}$.

Finally, Proposition 4 together with Proposition 3 leads to the following result, which will be used to construct the test in practice.

Theorem 1. *Suppose Assumptions 1 to 6 hold. Then, there exists a* $\varphi_T = o_T(2^j T^{-1})$ *and a positive number* T_0 *such that, for all* $T > T_0$,

$$\Pr\left(|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}| > \tilde{\sigma}_{j,\mathcal{R},T}\eta'\right) \leqslant c_0 \exp\left\{-\frac{1}{8}\cdot\frac{\eta^2}{1+\frac{\eta}{|\mathcal{R}T|\sigma_{j,\mathcal{R},T}}L_j + \frac{2^{j/2}\eta\nu}{|\mathcal{R}|\sqrt{T}\sigma_{j,\mathcal{R};T}}\left(\|c_X\|_{1,\infty}+c_1\rho\right)}\right\} + o_T(1)$$

for all $j = -1, \ldots, -J_T$, *where* $\eta' = \eta(1-\varphi_T/\sigma^2_{j,\mathcal{R};T})^{1/2}$, *and the positive constants* c_0, c_1 *are defined in the assertion of Proposition 3.*

Remark 3. Theorem 1 gives an approximation of the distribution of the normalized loss $|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}|/\tilde{\sigma}_{j,\mathcal{R},T}$. This depends on the unknown quantities $\|c_X\|_{1,\infty}$ and ρ, cf. (2.4). These two quantities may be understood as nuisance parameters of the problem, depending on the global sprectrum. The estimation of these quantities is based on a preliminary smoothing of $L_{j;T}(z)$ with respect to z, which we denote by $L^*_{j;T}(z)$. Here, we think about using a kernel smoothing procedure, or a wavelet transform shrinkage as studied in Nason *et al.* (2000). Then, a preliminary estimate of $\|c_X\|_{1,\infty}$ is obtained by plugging $L^*_{j;T}(z)$ into $\|c_X\|_{1,\infty}$, cf. (2.6) and (3.8). Next, the preliminary estimation of ρ necessitates the estimation of $\mathrm{TV}(S_j)$, cf. (2.3). We estimate $\mathrm{TV}(S_j)$ by $\sum_i |L^*_{j;T}(z_i^{\max}) - L^*_{j;T}(z_i^{\min})|$, where the sum is over the local minima and maxima of $L^*_{j;T}(z)$, with $z_i^{\max} < z_{i+1}^{\min} < z_{i+1}^{\max}$ for all i.

3.5 Discussion of the test procedure

We now propose our test procedure. Under H_0, see (3.2), the approximate distribution of the test statistic is given by

$$\Pr\left(|Q_{j,\mathcal{R};T}| > \eta'\tilde{\sigma}_{j,\mathcal{R},T}\,\middle|\,\mathrm{H}_0\right) \leqslant h(\eta') \tag{3.14}$$

for T sufficiently large, and where h is the exponential function following from Theorem 1. Let α be the nominal level of the test. We reject H_0 if

$$|Q_{j,\mathcal{R};T}| > \eta_\star\tilde{\sigma}_{j,\mathcal{R},T}, \tag{3.15}$$

where $\eta_\star$ is such that $h(\eta_\star) = \alpha$.

We now discuss the power of this test and, for this, we need to be more specfic about the alternative hypothesis H_1. We will work with the sensible alternative hypothesis that there exists a strictly positive real number θ and a measurable set with a non zero measure $\mathcal{U} \subseteq \mathcal{R}$ such that $S_j(z) \geqslant \theta$ for all z in $\mathcal{U}$. Formaly, if $|\mathcal{U}|$ denotes the Lebesgue measure of $\mathcal{U}$:

$$\mathrm{H}_1 : \exists\,\theta > 0 \text{ and } \mathcal{U} \subseteq \mathcal{R} \text{ with } |\mathcal{U}| > 0 \text{ and } S_j(z) \geqslant \theta \quad \forall z \in \mathcal{U}. \tag{3.16}$$

The next proposition evaluates the type II error of the test. The proof is to be found in Appendix B, Section B.6.

Proposition 5. *Suppose Assumption 1 to 6 hold true. Let the null hypothesis (3.2) against the alternative hypothesis (3.16) be given and consider the test rule (3.15) with*

$$\eta_\star < 2^{(1-j)/2}\|c_X\|^{-1}_{1,\infty}|\mathcal{R}T|^{1/2}Q_{j,\mathcal{R}}.$$

Then, there exists $T_0 > 1$ such that, for all $T \geqslant T_0$, the type II error of the test is bounded as follows:

$$\Pr\left(H_0 \text{ is not rejected} \,\Big|\, H_1\right) \leqslant C' \cdot \exp\left[-c' \frac{T}{\log_2^2 T} \frac{\theta^2 |\mathcal{U}|^2}{|\mathcal{R}|^2}\right] + C'' \cdot \exp\left[-c'' \frac{\sqrt{T}}{\log_2^2 T} \frac{\theta |\mathcal{U}|}{|\mathcal{R}|}\right] + o_T(1),$$

where the positive constants c', C', c'', C'' and the $o_T(1)$ term do not depend on $\mathcal{R}$, $\mathcal{U}$ and θ.

The last result shows the consistency of the test procedure. Moreover, it allows to discuss the local alternative of the test. We first note that the alternative hypothesis (3.16) depends on the two parameters θ and $\mathcal{U}$. Consequently, to study the local alternative of the test, we need to investigate both cases $\theta = \theta_T \to 0$ and $\mathcal{U} = \mathcal{U}_T$ such that $|\mathcal{U}_T| \to 0$. However, the upper bound of the type II error in Proposition 5 depends on the product $\theta_T|\mathcal{U}_T|$, and then the local alternative of the test is studied when this product tends to 0 when $T \to \infty$. By straightforward considerations, we see that if

$$\frac{\log_2^2 T}{\theta_T |\mathcal{U}_T| \sqrt{T}}$$

tends to zero as $T \to \infty$, then the type II error of the test aymptotically vanishes.

4 Pointwise adaptive estimation

Theorem 1 may be also useful for other statistical applications. In this section, we derive one important application given by the pointwise estimation of the wavelet spectrum.

Indeed, the estimator $Q_{j,\mathcal{R};T}$ may be seen as a smoothing over time of the inconsistent corrected wavelet periodogram. It can then be used for the pointwise estimation of the wavelet spectrum. In this problem, we want to estimate $S_j(z_0)$ at a fixed point z_0. This estimation can be done by computing the histogram $Q_{j,\mathcal{R};T}$ constructed on a segment $\mathcal{R}$ containing the fixed time point z_0. Consequently, the question how to choose the best segment $\mathcal{R}$ arises, and the goal of this section is to provide a data-driven procedure to select $\mathcal{R}$ automatically.

The proposed method goes back to the pointwise adaptive estimation theory of Lepski (1990), see also Lepski and Spokoiny (1997) and Spokoiny (1998). Suppose that the wavelet spectrum at $S_j(z_0)$ is well approximated by the averaged spectrum $Q_{j,\mathcal{U}}$ for a given interval $\mathcal{U}$ containing the reference point z_0. The idea of the procedure is to consider a second interval $\mathcal{R}$ containing $\mathcal{U}$ and to test if $Q_{j,\mathcal{R}}$ differs significantly from $Q_{j,\mathcal{U}}$. As we describe below, this test procedure is based on Proposition 3 or Theorem 1. If there exists a subset $\mathcal{U}$ of $\mathcal{R}$ such that $|Q_{j,\mathcal{R}} - Q_{j,\mathcal{U}}|$ is significantly different from zero, then we reject the hypothesis of homogeneity of the wavelet spectrum $S_j(z)$ on $z \in \mathcal{R}$. Finally, the adaptive estimator corresponds to the largest interval $\mathcal{R}$ such that the hypothesis of homogeneity is not rejected.

This section contains a formal description of this algorithm and derives some properties of the estimator.

4.1 Testing homogeneity

Let $\mathcal{R}$ be an interval containing z_0, $\mathcal{U}$ a subset of $\mathcal{R}$ and define

$$\Delta_j(\mathcal{R}, \mathcal{U}) = |Q_{j,\mathcal{R}} - Q_{j,\mathcal{U}}|. \tag{4.1}$$

Under assumptions 1 to 3, Proposition 3 implies

$$\Pr\left[|Q_{j,\mathcal{R},T} - Q_{j,\mathcal{U},T}| > \Delta_j(\mathcal{R}, \mathcal{U}) + \eta\left(\sigma_{j,\mathcal{R},T} + \sigma_{j,\mathcal{U},T}\right) k_T\right] \leqslant h(\mathcal{U}, \eta) + h(\mathcal{R}, \eta)$$

with

$$h(\mathcal{R},\eta) = c_0 \exp\left\{ -\frac{1}{8} \cdot \frac{\eta^2 k_T^2}{1 + \frac{\eta k_T}{|\mathcal{R}T|\sigma_{j,\mathcal{R},T}} L_j + \frac{2^{j/2}\eta k_T \nu}{\sqrt{|\mathcal{R}T|}\sigma_{j,\mathcal{R};T}} \left(\|c_X\|_{1,\infty} + c_1\rho|\mathcal{R}|^{-1/2}\right)} \right\}$$

and where the sequence k_T is such that (3.10) holds. Under the assumption that the wavelet spectrum S_j is homogeneous on the segment $\mathcal{R}$, the difference $\Delta_j(\mathcal{R},\mathcal{U})$ is negligeable. Then, as a test rule, we reject the homogeneity hypothesis on $\mathcal{R}$ if there exists a subset $\mathcal{U} \subset \mathcal{R}$ such that $|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{U};T}| > \eta(\sigma_{j,\mathcal{R},T} + \sigma_{j,\mathcal{U},T})k_T$ for a given η.

In the case where the variances $\sigma_{j,\mathcal{R},T}$ and $\sigma_{j,\mathcal{U},T}$ are unknown, they may be estimated as in Section 3.4 above. In that case, the homogeneity test is based on Theorem 1 and the modification of the following results is straightforward.

In practice, we choose a set Λ of interval-candidates $\mathcal{R}$. Then, for each candidate $\mathcal{R}$, we apply the homogeneity test with respect to a given set $\wp(\mathcal{R})$ of subintervals $\mathcal{U}$ of $\mathcal{R}$.

Assumption 7. In the estimation procedure described below, we assume the following properties on the test sets Λ and $\wp(\mathcal{R})$:

1. For all $\mathcal{R}$, the shortest interval of $\wp(\mathcal{R})$ is of length at least $\delta > 0$,
2. The cardinality of $\wp(\mathcal{R})$ is such that $\sharp(\wp(\mathcal{R})) \leqslant |\mathcal{R}T|^{\kappa\alpha\sqrt{\delta}}$ for some $0 < \alpha < 1$ and $\kappa \leqslant \sqrt{K_1}/[\nu(\|c\|_{1,\infty} + c_1\rho)]$,
3. When we test the homogeneity of the wavelet spectrum on $\mathcal{R}$, we assume that there exists a subinterval $\mathcal{U} \in \wp(\mathcal{R})$ such that $\mathcal{U} \subset \mathcal{R}$ and $\mathcal{U}$ contains z_0.

Remark 4 (Test sets). In this remark, we give one example of sets Λ and $\wp(\mathcal{R})$. For each scale $j < 0$, the corrected wavelet spectrum (3.4) is evaluated on a grid k/T, $r = 0, \dots, T-1$ in time. Then, we can choose the set Λ as

$$\Lambda = \{[r_0/T, r_1/T] : r_0 < [z_0 T] < r_1\}$$

for $r_0, r_1 \in \{0, T-1\}$. Nevertheless, in order to reduce the computational effort, we shrink the cardinality of Λ following the method of Spokoiny (1998). More precisely, we first select two sets $\mathcal{K}_m = \{r_m : r_m \leqslant [z_0 T]\}$ and $\mathcal{K}_n = \{r_n : r_n \geqslant [z_0 T]\}$ which both contain less than T points, and we set

$$\Lambda = \{[r_m/T, r_n/T] : r_m \in \mathcal{K}_m, r_n \in \mathcal{K}_n\}.$$

Then, one possibility to define $\wp(\mathcal{R})$ is to consider

$$\wp(\mathcal{R}) = \{[r_-/T, r_+/T] : r_-, r_+ \in \mathcal{K}_m \cup \mathcal{K}_n\}.$$

We refer to Spokoiny (1998) for details about this construction.

4.2 The estimation procedure

The estimation procedure simply starts with the smallest interval in Λ, assuming that the wavelet spectrum is homogeneous on this short interval. Then, it selects iteratively longer intervals in Λ until the homonegeneity assumption is rejected. Finally, the adaptive segment $\tilde{\mathcal{R}}$ is the longest segment $\mathcal{R}$ of Λ for which the homogeneity test is not rejected:

$$\tilde{\mathcal{R}} = \arg\max_{\mathcal{R}\in\Lambda} \{|\mathcal{R}| \text{ such that } |Q_{j,\mathcal{R};T} - Q_{j,\mathcal{U};T}| \leqslant \eta(\sigma_{j,\mathcal{R},T} + \sigma_{j,\mathcal{U},T})k_T \text{ for all } \mathcal{U} \subset \wp(\mathcal{R})\}. \quad (4.2)$$

The adaptive estimator of $S_j(z_0)$ is then defined by

$$\tilde{S}_j(z_0) = Q_{j,\tilde{\mathcal{R}},T}. \tag{4.3}$$

4.3 Properties of the estimator in homogeneous regions

The next result quantifies the ℓ_p risk ($p \geqslant 2$) when the wavelet spectrum $S_j(z)$ is homogeneous on $z \in \mathcal{R}$. To define this concept of homogeneity, we introduce the bias

$$b(\mathcal{R}) := \sum_{z \in \mathcal{R}} |S_j(z) - Q_{j,\mathcal{R}}|,$$

which measures how well the wavelet spectrum S_j is approximated by $Q_{j,\mathcal{R}}$ on $z \in \mathcal{R}$. We say that the spectrum is *homogeneous* (or *regular*) on $\mathcal{R}$, if the inequality

$$b(\mathcal{R}) \leqslant C_j \ \sigma_{j,\mathcal{R},T} \ k_T \tag{4.4}$$

holds with

$$C_j = 2^{-j/2}\sqrt{\alpha + p}. \tag{4.5}$$

In the inequality (4.4), $\sigma_{j,\mathcal{R},T}$ is the squared root of the variance of the estimator $Q_{j,\mathcal{R};T}$ of $S_j(z)$, $z \in \mathcal{R}$. As in Spokoiny (1998), (4.4) can be viewed as a balance relation between the bias and the variance of this estimate. The k_T term then appears as the correction term necessary in the pointwise estimation in order to bound the normalized loss (see Lepski (1990), Lepski and Spokoiny (1997)). In the following results, we set k_T proportional to $\log_2^2 T$.

To write the ℓ_p risk in the regular case, we need also the following assumption.

Assumption 8. $Q_{j,\mathcal{U},T}$ is uniformly bounded by a constant S.

Observe that, due to the Total Variation constraint on the wavelet spectrum, $S_j(z)$ is also uniformly bounded, and we assume that it is uniformly bounded by S. On the other hand, Proposition 2 shows that $\sigma_{j,\mathcal{U},T}$ is uniformly bounded as well.

Proposition 6. *Let $\mathcal{R}$ be an interval of $(0,1)$ and consider the test rule (4.2). If the wavelet spectrum S_j is regular on $\mathcal{R}$ in the sense of conditions (4.4)—(4.5), then, with $\lambda = \eta = 2^{-j/2}5(2\alpha + p)$ and $k_T \sim \log_2^2 T$,*

$$\Pr\left(\mathcal{R} \text{ is rejected}\right) = O\left(T^{-cp\sqrt{\delta}}\right)$$

for some positive constant $c = c(\nu, \|c\|_{1,\infty}, \rho)$ depending on ν, $\|c\|_{1,\infty}$ and ρ only.

Using this proposition, we can evaluate an upper bound for the ℓ_p risk associated to our estimator.

Theorem 2. *Assume that the wavelet spectrum at scale j, $S_j(z)$, is homogeneous on the segment $\mathcal{R}$ in the sense of (4.4)–(4.5) with*

$$k_T \sim \log_2^2 T.$$

If $\tilde{S}_j(z)$ is the pointwise estimator of the wavelet spectrum obtained by the estimation procedure (4.2)–(4.3) with

$$\eta = 2^{-j/2}5(2\alpha + p),$$

then there exists T_0 such that the pointwise ℓ_p-loss is bounded as follows

$$\mathrm{E}|\tilde{S}_j(z) - S_j(z)|^p \leqslant c|\mathcal{U}T|^{-p/2}\left[2^{1+j/2}\delta^{-1} + 11(2\alpha + p)\log_2^2 T\right]^p$$

with a positive constant c.

The proof is to be found in Appendix B.8.

4.4 Properties of the estimator in inhomogeneous regions

We now describe the behaviour our estimator near a breakpoint located at a time point $z_\star$. We first need to be more specific about the definition of a breakpoint in the evolutionary spectrum.

For a fixed scale $j \in \{-1, \dots, -J_T\}$, assume the evolutionary wavelet spectrum to be homogeneous on $\mathcal{R}_0 = [z_0, z_\star)$ and on $\mathcal{R}_1 = (z_\star, z_1]$. Let us denote $\mathcal{R} = \mathcal{R}_0 \cup \mathcal{R}_1 = [z_0, z_1]$ and

$$\theta_T := \mathrm{E}\left[Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}_0;T}\right]$$

To prove the next proposition, we assume that the estimation procedure is such that $\mathcal{R}_0$ and $\mathcal{R}_1$ are in $\wp(\mathcal{R})$.

Proposition 7. *If the evolutionary wavelet spectrum at scale j contains a breakpoint at $z_\star$ as described above and if $k_T \sim \log_2^2 T$, then*

$$\Pr\left(\mathcal{R} \text{ is not rejected}\right) = O\left(\exp\left[-\frac{T\theta_T^2(|\mathcal{R}_0| \vee |\mathcal{R}_1|)}{\log_2^2 T}\right] + \exp\left[-\frac{\sqrt{T}\theta_T}{\log_2^2 T}\right]\right).$$

where c is a positive constant and $x \vee y = \max(x, y)$.

The proof of this proposition is given in Appendix B.9. In this result, θ_T may be seen as the level of a jump in the wavelet spectrum. Then, Proposition 7 informs about the minimal amplitude of the jump which may be detected by the estimation procedure. If θ_T is such that

$$\frac{\log_2^2 T}{\theta_T\sqrt{T}} \to 0,$$

then the estimation procedure is consistent in the sense that $\Pr(\mathcal{R}$ is not rejected) is asymptotically zero.

5 On the application of the results

Above, we have considered two main problems, namely a local test of sparsity (3.2) for a locally stationary wavelet process, and the pointwise estimation of its evolutionary spectrum. From a theoretical viewpoint, there exists a link between the solution of these two problems, since they are based on a non-asymptotic result on the deviation of a linear functional of the wavelet periodogram (Theorem 1).

However, each solution is devoted to a specific class of statistical problems. On one hand, the *test of sparsity* is of use when we would like to measure a change of regime in an observed process. Many examples arise where the effect of an input is measured on a time series, for instance the effect of a drug on the heart rate measured by an electrocardiogram recording. The resulting time series is expected to be globally nonstationary, and we think our test procedure may be used to detect if a scale that is not active before the input becomes active after the input. Moreover, as already said above, the test of sparsity may be applied on a whole scale, in order to test the significance of one given scale in an observed process. Another application is the use of the results of Section 3 in order to construct *tests of second-order stationarity.* This will lead to a procedure with multiple tests, and a comparison with existing work in this direction could be done.

On the other hand, the *pointwise adaptive estimator* of the wavelet spectrum may be applied time point by time point, leading to an estimator of the whole wavelet spectrum of the process. Another extension is to replace the histogram-based pointwise estimation by a *smooth kernel estimate.*

Behaviour of the estimator in terms of the smoothness of the kernel would be of interest. As our procedure is based on nonasymptotic approximations and is fully adaptive, a practical evaluation of the kernel-based estimator may also be provided through simulations.

Finally, we mention the possibility to combine the two procedures. For instance, the significance of some whole scales over the whole time may be statistically tested, before performing the estimation procedure on the scales which are significantly different from zero.

APPENDICES

A Properties of the autocorrelation wavelet system

This section summaries useful results on the system $\{\Psi_j\}$ and the operator A. Recall that we have denoted by $\mathcal{L}_j$ the length of $|\operatorname{supp}\psi_j|$ for all $j = -1, -2, \dots$ and then it holds $\mathcal{L}_j = (2^{-j} - 1)(\mathcal{L}_{-1} - 1) + 1 \leqslant 2^{-j}\mathcal{L}_{-1}$. We have also recalled the definition of the autocorrelation wavelet system $\{\Psi_j; j = -1, -2, \dots\}$ which is the convolution of the non-decimated wavelet system:

$$\Psi_j(\tau) = \sum_{k=-\infty}^{\infty} \psi_{jk}(0)\psi_{jk}(\tau).$$

It is straightforward to check that Ψ_j is compactly supported for all $j < 0$ and the length of its support is bounded by $2\mathcal{L}_j - 1$.

Our autocorrelation wavelet system is related to the continuous autocorrelation functions of wavelets studied by Saito and Beylkin (1993) and defined as

$$\Psi(x) = \int_{-\infty}^{\infty} du\ \psi(u)\psi(u - x)$$

for a *continuous* compactly supported wavelet ψ. Indeed, following Berkner and Wells (2002, Lemma 4.2), it can be shown that the equation

$$\Psi_j(\tau) = \Psi\left(2^j|\tau|\right)$$

holds for all $j = -1, -2, \dots$ and $\tau \in \mathbb{Z}$.

The following Lemma recalls other useful results on the autocorrelation wavelet system.

Lemma 1. *(a) For all scales j and for all τ, $\Psi_j(\tau) = \Psi_j(-\tau)$.*

(b) The autocorrelation wavelet system $\{\Psi_j; j = -1, -2, \dots\}$ is linearly independent.

(c) The identity

$$\sum_{j=-\infty}^{-1} 2^j\Psi_j(\tau) = \delta_0(\tau) \tag{A.1}$$

holds for all $\tau \in \mathbb{Z}$.

Property (a) is obvious and implies the symmetry of the local autocorrelation function, i.e. $c(z, \tau) = c(z, -\tau)$, as expected. Property (b) is proved in Nason *et al.* (2000, Theorem 1) and shows that the local autocovariance function is univoquely defined. Finally, property (c) is proved in Fryźlewicz *et al.* (2003, Lemma 6) and implies, for instance, that the wavelet spectrum of a White Noise process is proportional to 2^j for all scales $j < -1$.

As the autocorrelation wavelet system is not orthogonal, we introduce the Gram matrix A defined by $A_{j\ell} = \sum_\tau \Psi_j(\tau)\Psi_\ell(\tau)$. The following properties of A are used thereafter.

Lemma 2. *For Haar and Shannon wavelets, there exists a finite positive constant ν such that the matrix A fulfills the following properties for all $j = -1, \ldots, -\log_2 T$:*

$$\sum_{\ell=-\log_2 T}^{-1} A^{-1}_{j\ell} = 2^j + O\left(2^{j/2} T^{-1/2}\right) \tag{A.2}$$

$$\sum_{\ell=-\log_2 T}^{-1} |A^{-1}_{j\ell}| \leqslant \nu(1+\sqrt{2})2^{j/2} \tag{A.3}$$

$$\sum_{\ell=-\log_2 T}^{-1} 2^{-\ell/2}|A^{-1}_{j\ell}| \leqslant \nu \cdot 2^{j/2}\log_2 T, \qquad \sum_{\ell=-\log_2 T}^{-1} 2^{-\ell}|A^{-1}_{j\ell}| \leqslant \nu(2+\sqrt{2})2^{j/2}T^{1/2}. \tag{A.4}$$

For all compactly supported wavelets, the matrix A fulfills the following property:

$$A_{j\ell} \leqslant (2\mathcal{L}_j - 1) \wedge (2\mathcal{L}_\ell - 1) \wedge \sqrt{\mathcal{L}_\ell \mathcal{L}_m} \tag{A.5}$$

where $x \wedge y = \min(x, y)$.

Proof. The following argument shows that the main term in (A.2) is 2^j: Using that $\Psi_\ell(0) = 1$ for all $\ell < 0$ and the identity (A.1), we may write

$$\sum_{\ell=-\infty}^{-1} A^{-1}_{j\ell} = \sum_{\ell=-\infty}^{-1} A^{-1}_{j\ell} \sum_{m,u=-\infty}^{\infty} 2^m \Psi_m(u)\Psi_\ell(u) = \sum_{m=-\infty}^{-1} 2^m \delta_0(j-m) = 2^j$$

from the definition of A. Observe that this argument holds for all compactly supported wavelets. To compute the remainder of (A.2), we introduce the auxiliary matrix $\Gamma = D' \cdot A \cdot D$ with diagonal matrix $D = \mathrm{diag}(2^{\ell/2})_{\ell<0}$, i.e. $\Gamma_{j\ell} = 2^{j/2} A_{j\ell} 2^{\ell/2}$. Nason *et al.* (2000, Theorem 2) have proven that the spectral norm of Γ^{-1} is bounded for Haar and Shannon wavelets. Then, we get

$$\sum_{\ell=-\infty}^{-\log_2(T)-1} A^{-1}_{j\ell} = 2^{j/2} \sum_{\ell=-\infty}^{-\log_2(T)-1} 2^{\ell/2}\Gamma^{-1}_{j\ell} = O\left(2^{j/2}T^{-1/2}\right)$$

To prove (A.3),

$$\sum_{\ell=-\log_2 T}^{-1} |A^{-1}_{j\ell}| = \sum_{\ell=-\log_2 T}^{-1} 2^{j/2}2^{\ell/2}|\Gamma^{-1}_{j\ell}| \leqslant 2^{j/2}(1+\sqrt{2})\nu$$

using $\sup_{j\ell}|\Gamma^{-1}_{j\ell}| \leqslant \nu$. (A.4) is obtained similarly, using the approximation $\sum_{j=-\log_2 T}^{-1} 2^{-j/2} \leqslant (2+\sqrt{2})\sqrt{T}$. (A.5) follows from the definition of $A_{j\ell}$ and the support of the autocorrelation wavelets, using $|\Psi_j(\tau)| \leqslant 1$ uniformly in j and τ. □

B Proofs

Suppose M is an $n \times n$ matrix and $M^\star$ is the conjugate transpose of M. We denote

$$\|M\|_2 := \sqrt{\mathrm{tr}\,(M^\star M)}$$

the Euclidean norm of M and

$$\|M\|_{\mathrm{spec}} := \max\{\sqrt{\lambda} : \lambda \text{ is eigenvalue of } M^\star M\}$$

the spectral norm of M. If M is symmetric and nonnegative definite, by standard theory we have $\|M\|_{\text{spec}} = \sup\{\|Mx\|_2^2 : x \in \mathbb{C}^n, \|x\|_2 = 1\}$. We will also use the following standard relations which hold for all symmetric matrices B, C:

$$\|B\|_{\text{spec}} \leqslant \|B\|_2 \tag{B.1}$$
$$\|B\|_{\text{spec}} = \max\{\lambda : \lambda \text{ is eigenvalue of } B\} \tag{B.2}$$
$$\|BC\|_{\text{spec}} \leqslant \|B\|_{\text{spec}}\|C\|_{\text{spec}} \tag{B.3}$$
$$\|BC\|_2 \leqslant \|B\|_{\text{spec}}\|C\|_2 \leqslant \|B\|_2\|C\|_2 \tag{B.4}$$

Moreover, if we suppose that the elements of the matrix B are continuously differentiable functions of t, then we shall also use

$$\frac{\partial}{\partial t} \log \det B = \operatorname{tr}\left(B^{-1}\frac{\partial}{\partial t}B\right). \tag{B.5}$$

In the sequel, we use the convention $w_{jk;T} = 0$ for $k < 0$ and $k \geqslant T$, which leads to helpful simplifications in the following proofs.

B.1 Proof of Proposition 1

On one hand, due to Definition 1, and equation (2.2), we have

$$\begin{aligned} c_{X,T}(z,\tau) = \operatorname{Cov}\left(X_{[zT],T}, X_{[zT]+\tau,T}\right) &= \sum_{j=-\infty}^{-1}\sum_{k=-\infty}^{\infty} |w_{j,k+[zT];T}|^2 \psi_{jk}(0)\psi_{jk}(\tau) \\ &= \sum_{j=-\infty}^{-1}\sum_{k=-\infty}^{\infty} S_j\left(\frac{k+[zT]}{T}\right)\psi_{jk}(0)\psi_{jk}(\tau) + \operatorname{Rest}_T(z,\tau) \end{aligned}$$

where the remainder is such that $|\operatorname{Rest}_T(z,\tau)| \leqslant T^{-1}\sum_{j=-\infty}^{-1}\sum_{k=-\infty}^{\infty} C_j|\psi_{jk}(0)\psi_{jk}(\tau)|$ by Assumption (2.2). On the other hand, we have $c_X(z,\tau) = \sum_{j=-\infty}^{-1}\sum_{k=-\infty}^{\infty} S_j(z)\,\psi_{jk}(0)\psi_{jk}(\tau)$. Then,

$$\begin{aligned} &\sum_{\tau=-\infty}^{\infty}\int_0^1 dz |c_{X,T}(z,\tau) - c_X(z,\tau)| \\ &\leqslant \sum_{\tau=-\infty}^{\infty}\int_0^1 dz \sum_{k=-\infty}^{\infty}\sum_{j=-\infty}^{-1}\left|S_j\left(\frac{k+[zT]}{T}\right) - S_j(z)\right| |\psi_{jk}(0)\psi_{jk}(\tau)| + \sum_{\tau=-\infty}^{\infty}\int_0^1 dz\,\operatorname{Rest}_T(z,\tau) \end{aligned}$$

With appropriate changes of variables, this bound may be written

$$\sum_{\tau=-\infty}^{\infty}\sum_{j=-\infty}^{-1}\sum_{k=-\infty}^{\infty}\sum_{t=0}^{T-1}\int_0^{1/T} dz\left|S_j\left(\frac{k+[zT]+t}{T}\right) - S_j\left(z+\frac{t}{T}\right)\right||\psi_{jk}(0)\psi_{jk}(\tau)| + \sum_{\tau=-\infty}^{\infty}\int_0^1 dz\,\operatorname{Rest}_T(z,\tau)$$

which is bounded by

$$T^{-1}\sum_{\tau=-\infty}^{\infty}\sum_{j=-\infty}^{-1}\sum_{k=-\infty}^{\infty}|k|\operatorname{TV}(S_j)\,|\psi_{jk}(0)\psi_{jk}(\tau)| + \sum_{\tau=-\infty}^{\infty}\int_0^1 dz\,\operatorname{Rest}_T(z,\tau)$$

where we have used the following property of the Total Variation norm:

$$\sum_{t=0}^{T-1} \left| S_p\left(\frac{t}{T}+\frac{\alpha}{T}\right) - S_p\left(\frac{t}{T}+\frac{\beta}{T}\right) \right| \leqslant |\alpha-\beta| \operatorname{TV}(S_p) \qquad \text{for all } \alpha,\beta \in \mathbb{N}. \tag{B.6}$$

As the support of $\psi_{jk}(0)$ is of length $\mathcal{L}_j$, we get $|k| \leqslant \mathcal{L}_j$. Together with condition (2.3) of Definition 1, this leads to

$$\sum_{\tau=-\infty}^{\infty} \int_0^1 dz |c_{X,T}(z,\tau) - c_X(z,\tau)| \leqslant T^{-1} \sum_{j=-\infty}^{-1} (C_j + \mathcal{L}_j L_j) \sum_{\tau=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |\psi_{jk}(0)\psi_{jk}(\tau)|.$$

The compact support of ψ_{jk} limits the sums over k and τ as follows:

$$\sum_{\tau,k=-\infty}^{\infty} |\psi_{jk}(0)\psi_{jk}(\tau)| = \sum_{\tau=-\mathcal{L}_j+1}^{\mathcal{L}_j-1} \sum_{-\infty}^{\infty} |\psi_{jk}(0)\psi_{jk}(\tau)| \leqslant 2\mathcal{L}_j - 1 \tag{B.7}$$

by the Cauchy-Schwarz inequality for the sum over k. We get the result by Assumption (2.4). □

B.2 Proof of Proposition 2

Our proof of Proposition 2 needs the following Lemma quoted from Neumann and von Sachs (1997).

Lemma 3. *Let* $\underline{Z}_n = (Z_1, \ldots, Z_n)'$ *be a vector of iid Gaussian random variables with zero mean and* $\operatorname{Var} Z_1 = 1$. *If* M_n *is an* $n \times n$ *matrix, then*

$$\begin{aligned} \operatorname{E}\left(\underline{Z}_n' M_n \underline{Z}_n\right) &= \operatorname{tr} M_n, \\ \operatorname{Var}\left(\underline{Z}_n' M_n \underline{Z}_n\right) &= 2 \operatorname{tr} M_n^\star M_n = 2\|M_n\|_2^2, \end{aligned}$$

and, for all $r \geqslant 2$, *if* Cum_r *denotes the* r*th cumulant, we have*

$$|\operatorname{Cum}_r\left(\underline{Z}_n' M_n \underline{Z}_n\right)| \leqslant 2^{r-1}(r-1)!\, \|M_n\|_2^2\, \{\lambda_{\max}(M_n)\}^{r-2}.$$

Define $\underline{X}_T = (X_{0,T}, \ldots, X_{T-1,T})'$. By definition, $Q_{j,\mathcal{R};T}$ is the quadratic form

$$Q_{j,\mathcal{R};T} = \underline{X}_T' U_{j,\mathcal{R};T} \underline{X}_T \tag{B.8}$$

where $U_{j,\mathcal{R};T}$ is the $T \times T$ matrix with entry (s,t) equal to

$$U_{st} = |\mathcal{R}T|^{-1} \sum_{\ell=-\log_2 T}^{-1} A_{j\ell}^{-1} \sum_{k \in \mathcal{R}T} \psi_{\ell k}(s)\psi_{\ell k}(t).$$

For notational convenience, we omit the dependence of U_{st} in j and $\mathcal{R}$. Assuming that the orthonormal increment processes $\{\xi_{jk}\}$ in Definition 1 are Gaussian, $\underline{X}_T$ is a multivariate Gaussian random variable with covariance matrix $\Sigma_T = \operatorname{Cov}(\underline{X}_T \underline{X}_T')$. In that case, $Q_{j,\mathcal{R};T}$ is a quadratic form of Gaussian variables and we can apply Lemma 3 with

$$M_{j,\mathcal{R};T} = \Sigma_T'^{1/2} U_{j,\mathcal{R};T} \Sigma_T^{1/2} \tag{B.9}$$

in order to prove Proposition 2. The following lemmas derive some bounds for the Euclidean and the spectral norm of $U_{j,\mathcal{R};T}$ and Σ_T.

Lemma 4. *With fixed $\mathcal{R} \subseteq (0,1)$, there exists a T_0 such that, uniformly in $T \geqslant T_0$,*

$$K_1\, 2^{2j} |\mathcal{R}T|^{-1}(1 + o_T(1)) \leqslant \|U_{j,\mathcal{R};T}\|_2^2 \leqslant K_2\, 2^j |\mathcal{R}|^{-2} T^{-1}$$

for all $j = -1, \ldots, J_T = o_T(\log_2 T)$, where K_1 and K_2 are two constants independent of j, T and $|\mathcal{R}|$.

Proof. The proof is straightforward when $\mathcal{R} = (0,1)$. However, one technical difficulty is to deal with a general interval $\mathcal{R} = (r_1, r_2) \subset (0,1)$. A simple remark which simplifies the proof is to observe that the quadratic form (B.8) involves $X_{[r_1T],T}, \ldots, X_{T-1,T}$ only. Consequently, the matrix $U_{j,\mathcal{R};T}$ is a $(T - [r_1T] + 1) \times (T - [r_1T] + 1)$ matrix, and we can write, from direct computations,

$$\|U_{j,\mathcal{R};T}\|_2^2 = |\mathcal{R}T|^{-2} \sum_{s,t=[r_1T]}^{T-1} \left(\sum_{\ell=-\log_2 T}^{-1} A_{j\ell}^{-1} \sum_{k \in \mathcal{R}T} \psi_{\ell k}(s)\psi_{\ell k}(t) \right)^2$$

The compact support of $\psi_{\ell k}(s)$ implies that $0 \leqslant k - s$, and, as $k \leqslant [r_2T]$, we can limit the sum over s, t as follows:

$$\|U_{j,\mathcal{R};T}\|_2^2 = |\mathcal{R}T|^{-2} \sum_{s,t \in \mathcal{R}T} \left(\sum_{\ell=-\log_2 T}^{-1} A_{j\ell}^{-1} \sum_{k \in \mathcal{R}T} \psi_{\ell k}(s)\psi_{\ell k}(t) \right)^2. \tag{B.10}$$

If we split the sum over ℓ at point ℓ_T such that $|\mathcal{L}_\ell| \leqslant |\mathcal{R}T|$ for all $\ell = -1, -2, \ldots, \ell_T$, then $|\ell_T| = O(\log_2 |\mathcal{R}T|)$ and

$$\|U_{j,\mathcal{R};T}\|_2^2 = |\mathcal{R}T|^{-2} \sum_{s,t \in \mathcal{R}T} \left(\sum_{\ell=-\ell_T}^{-1} A_{j\ell}^{-1} \Psi_\ell(s-t) + \sum_{m=-\log_2 T}^{-\ell_T - 1} A_{jm}^{-1} \sum_{k \in \mathcal{R}T} \psi_{mk}(s)\psi_{mk}(t) \right)^2.$$

Expanding the square, we get three terms, herewith denoted by I_T, II_T and III_T. By definition of Ψ_ℓ and $A_{j\ell}$, the first squared term is

$$\mathrm{I}_T = |\mathcal{R}T|^{-2} \sum_{s,t \in \mathcal{R}T} \left(\sum_{\ell=-\ell_T}^{-1} A_{j\ell}^{-1} \Psi_\ell(s-t) \right)^2 = |\mathcal{R}T|^{-1} \sum_{\ell,m=-\ell_T}^{-1} A_{j\ell}^{-1} A_{jm}^{-1} A_{m\ell}.$$

If we write the sum over $-\ell_T \leqslant \ell \leqslant -1$ as the sum over $-\log_2 T \leqslant \ell \leqslant -1$ minus the sum over $-\log_2 T \leqslant \ell \leqslant -\ell_T - 1$, then we get

$$\begin{aligned}
\mathrm{I}_T &= |\mathcal{R}T|^{-1} \left[\sum_{m=-\ell_T}^{-1} A_{j\ell}^{-1} \delta_{jm} - \sum_{m=-\ell_T}^{-1} \sum_{\ell=-\log_2 T}^{-\ell_T - 1} A_{j\ell}^{-1} A_{jm}^{-1} A_{m\ell} \right] \\
&= |\mathcal{R}T|^{-1} \left[A_{jj}^{-1} \mathbb{I}_{\{j \geqslant \ell_T\}} - \sum_{\ell=-\log_2 T}^{-\ell_T - 1} A_{j\ell}^{-1} \delta_{j\ell} + \sum_{m,\ell=-\log_2 T}^{-\ell_T - 1} A_{j\ell}^{-1} A_{jm}^{-1} A_{\ell m} \right],
\end{aligned}$$

and, using that the last sum over m, ℓ contains $\log_2(T) - \ell_T = O(\log_2 |\mathcal{R}|)$ elements,

$$\mathrm{I}_T \leqslant |\mathcal{R}T|^{-1} \left[A_{jj}^{-1} \mathbb{I}_{\{j \geqslant \ell_T\}} + A_{jj}^{-1} \mathbb{I}_{\{j < \ell_T\}} + 2^{j+1} \log(|\mathcal{R}|)\nu^2 \right] \leqslant 2^{j+2} \nu |\mathcal{R}T|^{-1}.$$

In order to compute a bound for the double product II_T, we exploit the compact support of $\psi_{\ell k}(s)$ implying that $k \leqslant s + \mathcal{L}_m \leqslant [r_2 T] + \mathcal{L}_m$, and then

$$\text{II}_T = 2|\mathcal{R}T|^{-2} \sum_{s,t \in \mathcal{R}T} \sum_{\ell=-\ell_T}^{-1} \sum_{m=-\log_2 T}^{-\ell_T - 1} A_{j\ell}^{-1} A_{jm}^{-1} \Psi_\ell(s-t) \sum_{k=[r_2 T]+1}^{[r_2 T]+\mathcal{L}_m} \psi_{mk}(s)\psi_{mk}(t).$$

With $u := s - t$,

$$\begin{aligned} \text{II}_T = {} & 2|\mathcal{R}T|^{-2} \sum_{\ell=-\ell_T}^{-1} \sum_{m=-\log_2 T}^{-\ell_T-1} A_{j\ell}^{-1} A_{jm}^{-1} \sum_{u=0}^{|\mathcal{R}T|} \Psi_\ell(u) \sum_{s=u+[r_1 T]}^{[r_2 T]} \sum_{k=[r_2 T]+1}^{[r_2 T]+\mathcal{L}_m} \psi_{mk}(s)\psi_{mk}(s-u) \\ & + 2|\mathcal{R}T|^{-2} \sum_{\ell=-\ell_T}^{-1} \sum_{m=-\log_2 T}^{-\ell_T-1} A_{j\ell}^{-1} A_{jm}^{-1} \sum_{u=-|\mathcal{R}T|}^{-1} \Psi_\ell(u) \sum_{s=[r_1 T]}^{u+[r_2 T]} \sum_{k=[r_2 T]+1}^{[r_2 T]+\mathcal{L}_m} \psi_{mk}(s)\psi_{mk}(s-u) \end{aligned}$$

and, applying the Cauchy-Schwarz inequality for the sum over s, we finally get

$$\text{II}_T \leqslant 4|\mathcal{R}T|^{-2} \left(\sum_{m=-\ell_T}^{-1} \mathcal{L}_m |A_{jm}^{-1}| \right) \left(\sum_{\ell=-\log_2 T}^{-\ell_T - 1} \mathcal{L}_\ell |A_{j\ell}^{-1}| \right) \leqslant 2(6 + 2\sqrt{2})|\mathcal{R}|^{-2} \mathcal{L}_{-1}^2 \nu^2 2^j T^{-1}$$

using (A.4). Similar calculations lead to $\text{III}_T \leqslant (2+\sqrt{2})\nu \mathcal{L}_{-1}^2 |\mathcal{R}|^{-2} 2^j T^{-1}$. Putting these bounds together gives the upper bound of $\|U_{j,\mathcal{R};T}\|_2^2$.

On the other hand, from (B.10), we can write

$$\|U_{j,\mathcal{R};T}\|_2^2 \geqslant |\mathcal{R}T|^{-2} \sum_{s \in \mathcal{R}T} \left(\sum_{\ell=-\log_2 T}^{-1} A_{j\ell}^{-1} \sum_{k \in \mathcal{R}T} \psi_{\ell k}^2(s) \right)^2.$$

If we split the sum over ℓ at point ℓ_T such that $|\mathcal{L}_\ell| \leqslant |\mathcal{R}T|$ for all $\ell = -1, -2, \ldots, \ell_T$, then $|\ell_T| = O(\log_2 |\mathcal{R}T|)$ and, as $|k - s| \leqslant |\mathcal{R}T|$ and by definition of ℓ_T, $\sum_{k \in \mathcal{R}T} \psi_{\ell k}^2(s) = 1$ in the first term of the parenthesis, and we obtain

$$\|U_{j,\mathcal{R};T}\|_2^2 \geqslant |\mathcal{R}T|^{-2} \sum_{s \in \mathcal{R}T} \left(\sum_{\ell=\ell_T}^{-1} A_{j\ell}^{-1} + \text{Rest}_T \right)^2.$$

with $|\,\text{Rest}_T\,| \leqslant \sum_{\ell=-\log_2 T}^{\ell_T - 1} |A_{j\ell}^{-1}| = O(\nu \cdot 2^{j/2} |\mathcal{R}T|^{-1/2}|)$, where the rate follows using the same techniques to prove (A.3), except that here the sum over ℓ goes from $-\log_2 T$ to $\ell_T - 1$ with $\ell_T = O(\log_2 |\mathcal{R}T|)$. On the other hand, (A.2) implies that $\sum_{\ell=\ell_T}^{-1} A_{j\ell}^{-1} = 2^j + O(|\mathcal{R}T|^{-1})$, and we get the result. □

Lemma 5. *Under Assumption (3.7)*

$$\|\Sigma_T\|_{\text{spec}} = \|\Sigma_T^{1/2}\|_{\text{spec}}^2 \leqslant \|c_X\|_{1,\infty} < \infty.$$

On the other hand, under Assumption 2, $\|\Sigma_T^{-1}\|_{\text{spec}}$ *is uniformly bounded in* T*.*

Proof. For $x \in \mathbb{C}^T$ with $\|x\|_2 = 1$ and if $\sigma_{s,t}$ denotes the element (s,t) of the matrix Σ_T,

$$\|\Sigma^{1/2} x\|_2^2 = \sum_{s,t=0}^{T-1} x_s \overline{x}_t \sigma_{s,t} \leqslant \sum_{s=0}^{T-1} \sum_{u=-(T-1)}^{T-1} |x_s \overline{x}_{s+u} \sigma_{s,s+u}| \leqslant \sum_{u=-(T-1)}^{T-1} \sup_{s=0,\ldots,T-1} |\sigma_{s,s+u}|$$

which gives the first result using (3.7). The proof on the inverse matrix is similar to the proof of Fryźlewicz *et al.* (2003, Lemmas A.1 and A.2), itself based on a technique developed in Dahlhaus (1996, Section 4) for approximating covariance matrices of locally stationary Fourier processes. In the present case, an additional difficulty occurs since we are assuming a total variation constraint on the spectrum, instead of the Lipschitz continuity in the cited articles. To deal with the total variation norm, we can use the techniques of approximation developed in our proof of Proposition 4 below. For the sake of presentation, we do not explicitly give the details of this proof here. □

We can now prove Proposition 2.

Expectation

$$\begin{aligned}
\mathrm{E}Q_{j,\mathcal{R};T} &= |\mathcal{R}T|^{-1} \sum_{k\in\mathcal{R}T} \sum_{\ell=-\log_2 T}^{-1} A_{j\ell}^{-1} \sum_{s,t=0}^{T-1} \psi_{\ell k}(s)\psi_{\ell k}(t) \sum_{m=-\infty}^{-1} \sum_{n=-\infty}^{\infty} w_{mn;T}^2 \psi_{mn}(s)\psi_{mn}(t) \\
&= |\mathcal{R}T|^{-1} \sum_{k\in\mathcal{R}T} \sum_{\ell=-\log_2 T}^{-1} A_{j\ell}^{-1} \sum_{m=-\infty}^{-1} \sum_{n=-\infty}^{\infty} w_{mn;T}^2 \left(\sum_{s=0}^{T-1} \psi_{\ell k}(s)\psi_{mn}(s) \right)^2
\end{aligned}$$

defining $u := n - k$,

$$= |\mathcal{R}T|^{-1} \sum_{k\in\mathcal{R}T} \sum_{m=-\infty}^{-1} \sum_{u=-\infty}^{\infty} w_{m,u+k,T}^2 \sum_{\ell=-\log_2 T}^{-1} A_{j\ell}^{-1} \left(\sum_{s=-\infty}^{\infty} \psi_{\ell k}(s)\psi_{m,u+k}(s) \right)^2 .$$

By Definition 1, we can write $w_{m,u+k,T}^2 = S_m(k/T) + R_T(m,u,k)$ with

$$|R_T(m,u,k)| \leqslant \left| S_m\left(\frac{u+k}{T}\right) - S_m\left(\frac{k}{T}\right) \right| + \frac{C_m}{T}$$

which leads to

$$\mathrm{E}Q_{j,\mathcal{R};T} = |\mathcal{R}T|^{-1} \sum_{k\in\mathcal{R}T} \sum_{m=-\infty}^{-1} S_m\left(\frac{k}{T}\right) \sum_{\ell=-\log_2 T}^{-1} A_{j\ell}^{-1} \sum_{u=-\infty}^{\infty} \left(\sum_{s=-\infty}^{\infty} \psi_{\ell k}(s)\psi_{m,u+k}(s) \right)^2 + \mathrm{Rest}_T$$

By construction of the matrix A, we observe that

$$A_{\ell m} = \sum_{u=-\infty}^{\infty} \left(\sum_{s=-\infty}^{\infty} \psi_{\ell k}(s)\psi_{m,u+k}(s) \right)^2 \tag{B.11}$$

which implies

$$\mathrm{E}Q_{j,\mathcal{R};T} = |\mathcal{R}T|^{-1} \sum_{k\in\mathcal{R}T} S_j\left(\frac{k}{T}\right) + \mathrm{Rest}_T = |\mathcal{R}|^{-1} \int_{\mathcal{R}} dz\ S_j(z) + O\left(|\mathcal{R}T|^{-1} L_j\right) + \mathrm{Rest}_T \tag{B.12}$$

where the last equality is a standard result on the Total Variation norm (see Brillinger (1975, Lemma P5.1) for instance).

We now bound $|\operatorname{Rest}_T|$. As s goes from $-\infty$ to ∞, we have

$$|\operatorname{Rest}_T| \leqslant \sum_{m=-\infty}^{-1} \sum_{\ell=-\log_2 T}^{-1} |A_{j\ell}^{-1}| \sum_{u=-\infty}^{\infty} |\mathcal{R}T|^{-1} \sum_{k\in\mathcal{R}T} \left[\left|S_m\left(\frac{u+k}{T}\right) - S_m\left(\frac{k}{T}\right)\right| + \frac{C_m}{T}\right] \left(\sum_{s=-\infty}^{\infty} \psi_{\ell 0}(s)\psi_{mu}(s)\right)^2 .$$

Using (B.6) for the sum over k, $|\operatorname{Rest}_T|$ is bounded by

$$\sum_{m=-\infty}^{-1} \sum_{u=-\infty}^{\infty} \left[|u|\frac{\operatorname{TV}(S_m)}{|\mathcal{R}T|} + \frac{C_m}{T}\right] \sum_{\ell=-\log_2 T}^{-1} |A_{j\ell}^{-1}| \left(\sum_{s=-\infty}^{\infty} \psi_{\ell 0}(s)\psi_{mu}(s)\right)^2$$

In this last expression, the compact support of $\psi_{\ell 0}$ and ψ_{mu} implies that $|u| \leqslant \mathcal{L}_\ell \vee \mathcal{L}_m$. Together with (B.11), we get

$$|\operatorname{Rest}_T| \leqslant |\mathcal{R}T|^{-1} \sum_{m=-\infty}^{-1} \sum_{\ell=-\log_2 T}^{-1} (\operatorname{TV}(S_m)(\mathcal{L}_\ell \vee \mathcal{L}_m) + C_m)|A_{j\ell}^{-1}|A_{\ell m}$$

with (A.5) and where $x \vee y = \max(x, y)$,

$$\begin{aligned} &\leqslant |\mathcal{R}T|^{-1} \sum_{m=-\infty}^{-1} \sum_{\ell=-\log_2 T}^{-1} (\operatorname{TV}(S_m)\mathcal{L}_\ell(2\mathcal{L}_m - 1) + \operatorname{TV}(S_m)\mathcal{L}_m(2\mathcal{L}_\ell - 1) + C_m(2\mathcal{L}_m - 1))|A_{j\ell}^{-1}| \\ &= (2+\sqrt{2})\nu 2^{j/2}|\mathcal{R}T|^{-1}\sqrt{T} \sum_{m=-\infty}^{-1} (2\mathcal{L}_m - 1)\operatorname{TV}(S_m) + O\left(2^{j/2}|\mathcal{R}T|^{-1}\right) \qquad \text{(B.13)} \end{aligned}$$

using (A.4) and (2.4).

Variance

Using Lemma 3 with (B.4), Lemma 4 and Lemma 5, we get the upper bound as follows

$$\begin{aligned} \operatorname{Var} Q_{j,\mathcal{R};T} &= 2\|M_{j,\mathcal{R};T}\|_2^2 \leqslant 2\|\Sigma_T^{1/2}\|_{\text{spec}}^4 \|U_{j,\mathcal{R};T}\|_2^2 \\ &\leqslant \|c_X\|_{1,\infty}^2\, 2^j\, |\mathcal{R}T|^{-1} \qquad \text{(B.14)} \end{aligned}$$

To obtain the lower bound, we use twice (B.4) on $U_{j,\mathcal{R};T} = \Sigma_T^{-1/2} M_{j,\mathcal{R};T} \Sigma_T^{-1/2}$:

$$\operatorname{Var} Q_{j,\mathcal{R};T} = 2\|M_{j,\mathcal{R};T}\|_2^2 \geqslant 2\|\Sigma_T^{-1/2}\|_{\text{spec}}^{-4} \|U_{j,\mathcal{R};T}\|_2^2. \qquad \text{(B.15)}$$

Since Σ_T is a symmetric matrix, $\|\Sigma_T^{-1/2}\|_{\text{spec}}^2 = \|\Sigma_T^{-1}\|_{\text{spec}}$ and Lemma 5 shows that $\|\Sigma_T^{-1}\|$ is uniformly bounded under Assumption 2. A lower bound for $\|U_{j,\mathcal{R};T}\|_2^2$ is stated in Lemma 4 and the result follows. □

B.3 Proof of Proposition 3 and its consequences

Our proof of Proposition 3 needs the use of an exponential bound for quadratic forms of Gaussian random variables. For sake of presentation, we recall now this result and refer to Dahlhaus and Polonik (2002, Proposition 6.1).

Proposition 8. *Let* $\underline{Z}_n = (Z_1, \dots, Z_n)'$ *be a vector of iid Gaussian random variables with zero mean and* $\operatorname{Var} Z_1 = 1$. *If* M_n *is an* $n \times n$ *matrix such that* $\|M_n\|_{\text{spec}} \leqslant \tau_\infty$ *and* $\sigma_n^2 = 2\|M_n\|_2^2$, *then for all* $\lambda > 0$

$$\Pr\left((\underline{Z}_n' M_n \underline{Z}_n - \operatorname{tr} M_n) > \sigma_n \lambda\right) \leqslant \exp\left(-\frac{1}{2} \cdot \frac{\lambda^2}{1 + 2\lambda\frac{\tau_\infty}{\sigma_n}}\right).$$

As in the proof of Proposition 2, equation (B.9), we write $Q_{j,\mathcal{R};T}$ as a quadratic form of Gaussian variables in order to apply Proposition 8 with

$$M_{j,\mathcal{R};T} = \Sigma_T'^{1/2} U_{j,\mathcal{R};T} \Sigma_T^{1/2} \tag{B.16}$$

to prove the assertion.

Proof of Proposition 3. Lemma 4 and 5 imply with (B.1) and (B.3):

$$\|M_{j,\mathcal{R};T}\|_{\text{spec}} \leqslant 2^{j/2}\nu\|c_X\|_{1,\infty}|\mathcal{R}|^{-1}T^{-1/2} \tag{B.17}$$

which, using Proposition 8, implies

$$\begin{aligned}
&\Pr\left((Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}) > \eta\sigma_{j,\mathcal{R},T}\right) \\
&\quad \leqslant \Pr\left((Q_{j,\mathcal{R};T} - \mathrm{E}Q_{j,\mathcal{R};T}) > \eta\sigma_{j,\mathcal{R},T}/2\right) + \Pr\left((\mathrm{E}Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}) > \eta\sigma_{j,\mathcal{R},T}/2\right) \\
&\quad \leqslant \exp\left(-\frac{1}{8} \cdot \frac{\eta^2}{1 + \eta\frac{2^{j/2}\nu\|c_X\|_{1,\infty}}{|\mathcal{R}|T^{1/2}\sigma_{j,\mathcal{R},T}}}\right) + \Pr\left((\mathrm{E}Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}) \geqslant \eta\sigma_{j,\mathcal{R},T}/2\right).
\end{aligned}$$

To bound the second probability, we observe that

$$|\mathrm{E}Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}| \leqslant |\mathcal{R}T|^{-1}\left\{L_j + 2(2+\sqrt{2})\rho\nu 2^{j/2}\sqrt{T}\right\}$$

is obtained using (B.12) and (B.13). This implies

$$\begin{aligned}
&\Pr\left((Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}) \geqslant \eta\sigma_{j,\mathcal{R},T}\right) \\
&\quad \leqslant \exp\left(-\frac{1}{8} \cdot \frac{\eta^2\sigma_{j,\mathcal{R},T}}{\sigma_{j,\mathcal{R},T} + \eta\frac{2^{j/2}\nu\|c_X\|_{1,\infty}}{|\mathcal{R}|T^{1/2}}}\right) + \exp\left(1 - \frac{1}{2\eta}\frac{\eta^2\sigma_{j,\mathcal{R},T}}{\frac{L_j + 2(2+\sqrt{2})\rho\nu 2^{j/2}\sqrt{T})}{|\mathcal{R}T|}}\right)
\end{aligned}$$

and the result follows. □

Proof of Corollary 1. In the following proof, K denotes a generic constant and k_T is an increasing function of T. By Proposition 2, $\sigma_{j,\mathcal{R},T}^2 := \operatorname{Var} Q_{j,\mathcal{R};T} \leqslant K_2|\mathcal{R}T|^{-1}$ uniformly in j, which implies

$$\begin{aligned}
&\Pr\left(\sup_{-J_T \leqslant j < 0} |Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}| \geqslant k_T\sqrt{K_2|\mathcal{R}T|^{-1}}\right) \\
&\quad \leqslant \sum_{j=-J_T}^{-1} \Pr\left(|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}| \geqslant k_T\sqrt{K_2|\mathcal{R}T|^{-1}}\right) \leqslant \sum_{j=-J_T}^{-1} \Pr\left(|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}| \geqslant k_T\sigma_{j,\mathcal{R},T}\right).
\end{aligned}$$

Using Proposition 3, this probability is bounded by

$$c_0\ J_T \max_{-J_T\leqslant j<0} \exp\left(-\frac{1}{8}\cdot\frac{k_T^2}{1+\frac{k_T L_j}{|\mathcal{R}T|\sigma_{j,\mathcal{R};T}}+\frac{k_T 2^{j/2}\nu}{|\mathcal{R}|\sqrt{T}\sigma_{j,\mathcal{R};T}}\left(\|c_X\|_{1,\infty}+c_1\rho\right)}\right)$$

Proposition 2 shows that, for T sufficiently large, $\sigma_{j,\mathcal{R},T}\geqslant 2^j\sqrt{K_1|\mathcal{R}T|^{-1}}$. Moreover, by equation (2.4), there exists a positive constant ρ' such that $L_j\leqslant 2^{j/2}\rho'$. This leads to the bound

$$c_0\ J_T \max_{-J_T\leqslant j<0} \exp\left(-\frac{1}{8}\cdot\frac{k_T^2}{1+\frac{k_T L_j}{2^j\sqrt{|\mathcal{R}T|K_1}}+\frac{k_T\nu}{2^{j/2}\sqrt{K_1}|\mathcal{R}|}\left(\|c_X\|_{1,\infty}+c_1\rho\right)}\right)$$

Since the maximum is attained for $j=-J_T$ and using $2^{-j}L_j<\rho\mathcal{L}_0$, we get the upper bound

$$c_0\ J_T \max_{-J_T\leqslant j<0} \exp\left(-\frac{1}{8}\cdot\frac{k_T^2}{1+\rho\mathcal{L}_0 k_T\left(K_1|\mathcal{R}T|\right)^{-1/2}+2^{J_T/2}\nu k_T K_1^{-1/2}|\mathcal{R}|^{-1/2}\left(\|c_X\|_{1,\infty}+c_1\rho\right)}\right)$$

which is $o_T(1)$ by the assumption on k_T. □

B.4 Proof of Proposition 4

In the following proof K is a generic constant.

Lemma 6. *If* $U_{ts}^{(j)}=|\mathcal{R}T|^{-1}\sum_{\ell=-\log_2 T}^{-1}A_{j\ell}^{-1}\sum_{k\in\mathcal{R}T}\psi_{\ell k}(s)\psi_{\ell k}(t)$, *then*

$$\sum_{t=-\infty}^{\infty}\sum_{s,u=-\infty}^{\infty}U_{ts}^{(j)}U_{tu}^{(j)}\mathbb{I}_{|s-u|\leqslant N_T}\leqslant|\mathcal{R}|^{-2}N_T T^{-1}2\mathcal{L}_0\nu^2 2^j\log_2^2 T=O\left(2^j\frac{N_T\log_2^2 T}{T}\right)$$

Proof. Direct calculations yields

$$\sum_{t=-\infty}^{\infty}\sum_{s,u=-\infty}^{\infty}U_{ts}^{(j)}U_{tu}^{(j)}\mathbb{I}_{|s-u|\leqslant N_T}\leqslant|\mathcal{R}T|^{-2}\sum_{\ell,m=-\log_2 T}^{-1}|A_{j\ell}^{-1}||A_{jm}^{-1}|\sum_{s,u=-\infty}^{\infty}\mathbb{I}_{|s-u|\leqslant N_T}$$
$$\sum_{t=-\infty}^{\infty}\left(\sum_{k\in\mathcal{R}T}|\psi_{\ell k}(s)\psi_{\ell k}(t)|\right)\left(\sum_{n\in\mathcal{R}T}|\psi_{mn}(u)\psi_{mn}(t)|\right).$$

Using the Cauchy-Schwarz inequality for the sum over t, we get a product between two terms similar to

$$\left\{\sum_t\left(\sum_{k\in\mathcal{R}T}\psi_{\ell k}(s)\psi_{\ell k}(t)\right)^2\right\}^{1/2}=\left\{\sum_{k,r\in\mathcal{R}T}\Psi_\ell(k-r)\psi_{\ell k}(s)\psi_{\ell r}(s)\right\}^{1/2}$$

if $k=r+u$, then the range of u is included in $\{-|\mathcal{R}T|,\dots,0,\dots,|\mathcal{R}T|\}$:

$$\leqslant\left\{\sum_{u=-|\mathcal{R}T|}^{|\mathcal{R}T|}\sum_{r\in\mathcal{R}T}|\Psi_\ell(u)|\cdot|\psi_{\ell,r+u}(s)\psi_{\ell r}(s)|\right\}^{1/2}\leqslant\left\{\sum_{u=-|\mathcal{R}T|}^{|\mathcal{R}T|}|\Psi_\ell(u)|\right\}^{1/2}$$

using Cauchy-Schwarz inequality for the sum over r and $\sum_r \psi_{\ell r}^2(s) = 1$. Finally, using the basic properties of $\Psi_j(\tau)$ described in Appendix A, we get

$$\left\{ \sum_t \left(\sum_{k \in \mathcal{R}T} \psi_{\ell k}(s) \psi_{\ell k}(t) \right)^2 \right\}^{1/2} \leqslant \sqrt{2\mathcal{L}_\ell - 1}. \tag{B.18}$$

Then

$$\sum_{t=-\infty}^{\infty} \sum_{s,u=-\infty}^{\infty} U_{ts}^{(j)} U_{tu}^{(j)} \mathbb{I}_{|s-u| \leqslant N_T} \leqslant T N_T |\mathcal{R}T|^{-2} \sum_{\ell,m=-\log_2 T}^{-1} |A_{j\ell}^{-1}||A_{jm}^{-1}| \sqrt{2\mathcal{L}_\ell - 1} \sqrt{2\mathcal{L}_m - 1}$$

and we obtain the result by (A.4). □

In the proof of Proposition 4, we need a modification of Corollary 1, in which $\mathcal{R}$ is replaced by $\mathcal{R}_T$. The proof of the following result is along the lines of the proof of Corollary 1.

Lemma 7. *Under the assumptions of Propositions 2 and 3, and under Assumption 4, there exists $T_0 \geqslant 1$ such that, for all $T \geqslant T_0$,*

$$\begin{aligned} &\Pr\left(\sup_{-J_T \leqslant j < 0} |Q_{j,\mathcal{R}_T(s);T} - Q_{j,\mathcal{R}_T(s)}| \geqslant k_T \sqrt{\frac{K_2}{|\mathcal{R}_T T|}} \right) \\ &\quad \leqslant c_0 J_T \exp\left[-\frac{1}{8} \cdot \frac{k_T^2}{1 + \rho \mathcal{L}_0 k_T \left(K_1 |\mathcal{R}_T T|\right)^{-1/2} + 2^{J_T/2} \nu k_T K_1^{-1/2} |\mathcal{R}_T|^{-1/2} \left(\|c_X\|_{1,\infty} + c_1 \rho\right)} \right] \\ &\quad = o_T(1) \end{aligned}$$

where k_T is such that (3.12) holds true.

Proof of Proposition 4. Define $\bar{\sigma}_{s,s+u} := \sum_{\ell=-\log_2 T}^{-1} Q_{\ell,\mathcal{R}_T(s)} \Psi_\ell(u) \mathbb{I}_{|u| \leqslant M_T}$ the entries of a matrix $\bar{\Sigma}$, and define $\bar{\sigma}_{j,\mathcal{R},T}^2 := 2\|U_{j,\mathcal{R};T}' \bar{\Sigma}_T\|_2^2$. Our proof is based on the decomposition

$$\tilde{\sigma}_{j,\mathcal{R},T}^2 - \sigma_{j,\mathcal{R},T}^2 = \left(\tilde{\sigma}_{j,\mathcal{R},T}^2 - \bar{\sigma}_{j,\mathcal{R},T}^2\right) + \left(\bar{\sigma}_{j,\mathcal{R},T}^2 - \sigma_{j,\mathcal{R},T}^2\right)$$

where the first term is stochastic while the second term is deterministic.

We will first show that the deterministic term $|\bar{\sigma}_{j,\mathcal{R},T}^2 - \sigma_{j,\mathcal{R},T}^2|$ is $o(2^{j-J_T} T^{-1})$. Using (B.4), we can write

$$\begin{aligned} \frac{1}{2}\left(\bar{\sigma}_{j,\mathcal{R},T}^2 - \sigma_{j,\mathcal{R},T}^2\right) &= \|U_{j,\mathcal{R};T}' \bar{\Sigma}_T\|_2^2 - \|U_{j,\mathcal{R};T}' \Sigma_T\|_2^2 \\ &\leqslant \|U_{j,\mathcal{R};T}'(\bar{\Sigma}_T - \Sigma_T)\|_2^2 + 2 \cdot \|U_{j,\mathcal{R};T}' \Sigma_T\|_2 \cdot \|U_{j,\mathcal{R};T}'(\bar{\Sigma}_T - \Sigma_T)\|_2 \\ &\leqslant \|U_{j,\mathcal{R};T}\|_2^2 \cdot \|\bar{\Sigma}_T - \Sigma_T\|_{\text{spec}}^2 + 2 \cdot \|U_{j,\mathcal{R};T}\|_2^2 \cdot \|\Sigma_T\|_{\text{spec}} \cdot \|\bar{\Sigma}_T - \Sigma_T\|_{\text{spec}} \end{aligned}$$

where we know by lemmas 4 and 5 that $\|U_{j,\mathcal{R};T}\|_2^2 = O(2^j T^{-1})$ and $\|\Sigma_T\|_{\text{spec}} \leqslant \|c_X\|_{1,\infty}$. Moreover, we can write:

$$\begin{aligned} \|\bar{\Sigma}_T - \Sigma_T\|_{\text{spec}} &\leqslant \sum_{u=-\infty}^{\infty} \sup_s (\sigma_{s,s+u} - \bar{\sigma}_{s,s+u}) \\ &= \sum_{u=-\infty}^{\infty} \sup_s \sum_{\ell=-\infty}^{-1} \sum_{n=-\infty}^{\infty} \left(w_{\ell n;T}^2 - Q_{\ell,\mathcal{R}_T(s)}\right) \cdot \psi_{\ell n}(s) \psi_{\ell n}(s+u) + \mathrm{R}_1 + \mathrm{R}_2 \end{aligned} \tag{B.19}$$

with

$$\mathrm{R}_1 = \sum_{u=-\infty}^{\infty} \sup_s \sum_{\ell=-\infty}^{-1} Q_{\ell,\mathcal{R}_T(s)} \Psi_\ell(u) \, \mathbb{I}_{|u|>M_T}$$

and

$$\mathrm{R}_2 = \sum_{u=-\infty}^{\infty} \sup_s \sum_{\ell=-\infty}^{-\log_2(T)-1} Q_{\ell,\mathcal{R}_T(s)} \Psi_\ell(u) \, \mathbb{I}_{|u|<M_T}.$$

As

$$\sum_{u=-\infty}^{\infty} \sup_s \sum_{\ell=-\infty}^{-1} Q_{\ell,\mathcal{R}_T(s)} \Psi_\ell(u) = \sum_{u=-\infty}^{\infty} \sup_s |\mathcal{R}_T|^{-1} \int_{\mathcal{R}_T(s)} dz \; c_X\,(z,u) \;\; ,$$

the rate of R_1 is $O(2^{-J_T})$ by Assumption 6. Next, using $|\Psi_\ell(u)| \leqslant 1$ uniformly in $\ell < 0$, we get

$$|\,\mathrm{R}_2\,| \leqslant \sum_{u=-\infty}^{\infty} \sup_s |\mathcal{R}_T|^{-1} \int_{\mathcal{R}_T(s)} dz \sum_{\ell=-\infty}^{-\log_2(T)-1} S_\ell(z) \, \mathbb{I}_{|u|<M_T} \leqslant 2M_T \sum_{\ell=-\infty}^{-\log_2(T)-1} \sup_z S_\ell(z) = o_T(2^{-J_T})$$

using Assumption 5. The main term of (B.19) is bounded by

$$\sum_{u=-\infty}^{\infty} \sup_s \sum_{\ell=-\log_2 T}^{-1} \sum_{n=-\infty}^{\infty} |\mathcal{R}_T|^{-1} \int_{\mathcal{R}_T(s)} dz \; |w^2_{\ell n;T} - S_\ell(z)| \cdot |\psi_{\ell n}(s)\psi_{\ell n}(s+u)|. \tag{B.20}$$

By Definition 1, we can write

$$|w^2_{\ell n;T} - S_\ell(z)| \leqslant \frac{C_\ell}{T} + \left| S_\ell\left(\frac{n}{T}\right) - S_\ell\left(\frac{n-s}{T} + z\right) \right| + \left| S_\ell(z) - S_\ell\left(\frac{n-s}{T} + z\right) \right|$$

which, when replaced in (B.20), leads to three terms. By (B.7) and (2.4), the first term is $O(T^{-1})$. For the second term, with a change of variable z to $z + s/T$, we get:

$$\sum_{u=-\infty}^{\infty} \sup_s \sum_{\ell=-\infty}^{-1} \sum_{n=-\infty}^{\infty} |\mathcal{R}_T|^{-1} \int_{\mathcal{R}_T(0)} dz \left| S_\ell\left(\frac{n}{T}\right) - S_\ell\left(\frac{n}{T} + z\right) \right| \cdot |\psi_{\ell n}(s)\psi_{\ell n}(s+u)|,$$

where $\mathcal{R}_T(0)$ denotes the interval $\mathcal{R}_T(s)$ shifted by $-s$. If we use that $|\psi_{\ell n}(s)|$ is uniformly bounded and that $\sum_{u=-\infty}^{\infty} |\psi_{\ell n}(s+u)| = O(\mathcal{L}_\ell)$, the second term is bounded (up to a multiplicative constant) by

$$\begin{aligned}
&|\mathcal{R}_T|^{-1} \sum_{\ell=-\infty}^{-1} \mathcal{L}_\ell \int_{\mathcal{R}_T(0)} dz \sum_{n=-\infty}^{\infty} \left| S_\ell\left(\frac{n}{T}\right) - S_\ell\left(\frac{n}{T} + z\right) \right| \\
&\qquad \leqslant |\mathcal{R}_T|^{-1} \sum_{\ell=-\infty}^{-1} \mathcal{L}_\ell \int_{\mathcal{R}_T(0)} dz \; |z| \, \mathrm{TV}(S_\ell) \leqslant |\mathcal{R}_T| \sum_{\ell=-\infty}^{-1} \mathcal{L}_\ell L_\ell \\
&\qquad = O(|\mathcal{R}_T|)
\end{aligned}$$

by assumptions (2.3) and (2.4). The third term is

$$\sum_{u=-\infty}^{\infty} \sup_s \sum_{\ell=-\infty}^{-1} \sum_{n=-\infty}^{\infty} |\mathcal{R}_T|^{-1} \int_{\mathcal{R}_T(s)} dz \left| S_\ell(z) - S_\ell\left(\frac{n-s}{T} + z\right) \right| \cdot |\psi_{\ell n}(s)\psi_{\ell n}(s+u)|.$$

If s_0 denotes the infimum of $\mathcal{R}_T(s)$, we decompose the integral as follows:

$$\sum_{u=-\infty}^{\infty} \sup_s \sum_{\ell=-\infty}^{-1} \sum_{n=-\infty}^{\infty} |\mathcal{R}_T|^{-1} \sum_{k=0}^{|\mathcal{R}_T T|-1} \int_{s_0+\frac{k}{T}}^{s_0+\frac{k+1}{T}} dz \left| S_\ell(z) - S_\ell\left(\frac{n-s}{T} + z\right) \right| \cdot |\psi_{\ell n}(s)\psi_{\ell n}(s+u)|$$

which can be rewritten with the change of variables $y := z - s_0 - k/T$,

$$\sum_{u=-\infty}^{\infty} \sup_s \sum_{\ell=-\infty}^{-1} \sum_{n=-\infty}^{\infty} |\mathcal{R}_T|^{-1} \sum_{k=0}^{|\mathcal{R}_T T|-1} \int_0^{1/T} dy \left| S_\ell\left(y + s_0 + \frac{k}{T}\right) - S_\ell\left(y + s_0 + \frac{n-s+k}{T}\right) \right| \cdot |\psi_{\ell n}(s)\psi_{\ell n}(s+u)|.$$

Assumption (2.3) for the sum over k with (B.6) leads to the bound

$$\sum_{u=-\infty}^{\infty} \sup_s \sum_{\ell=-\infty}^{-1} L_\ell \sum_{n=-\infty}^{\infty} |\mathcal{R}_T T|^{-1} |n-s| |\psi_{\ell n}(s)\psi_{\ell n}(s+u)|.$$

The compact support of $\psi_{\ell n}(s)$ implies $|n-s| < \mathcal{L}_\ell$, which leads to $O(|\mathcal{R}_T T|^{-1})$ by (B.7), (2.3) and (2.4). Finally, we get

$$2^{-j} T \cdot \left(\bar{\sigma}^2_{j,\mathcal{R},T} - \sigma^2_{j,\mathcal{R},T}\right) = O\left(T^{-1} + |\mathcal{R}_T|\right) + o_T(2^{-J_T}) = o_T(2^{-J_T})$$

by Assumption 4.

Let us now turn to the stochastic term $|\tilde{\sigma}^2_{j,\mathcal{R},T} - \bar{\sigma}^2_{j,\mathcal{R},T}|$. Lemma 7 implies the existence of a random set $\mathcal{A}$ which does not depend on j and such that $\Pr(\mathcal{A}) \geqslant 1 - o_T(1)$ and

$$|Q_{j,\mathcal{R}_T(s);T} - Q_{j,\mathcal{R}_T(s)}| \leqslant k_T \sqrt{K_2 |\mathcal{R}_T T|^{-1}} \tag{B.21}$$

almost surely on $\mathcal{A}$, for all $T > T_0$ and $j = -1, \ldots, -J_T$. We can write

$$|\bar{\sigma}^2_{j,\mathcal{R},T} - \tilde{\sigma}^2_{j,\mathcal{R},T}| \leqslant 2 \sum_{h,t=0}^{T-1} \left| \sum_{s,u=0}^{T-1} U^{(j)}_{ts} U^{(j)}_{tu} \sum_{\ell,m=-\log_2 T}^{-1} \left(Q_{\ell,\mathcal{R}_T(s);T} Q_{m,\mathcal{R}_T(s);T} - Q_{\ell,\mathcal{R}_T(s)} Q_{m,\mathcal{R}_T(s)}\right) \Psi_\ell(s-h)\Psi_m(u-h) \right| \cdot \mathbb{I}_{|s-h| \leqslant M_T} \mathbb{I}_{|u-h| \leqslant M_T} \tag{B.22}$$

almost surely on $\mathcal{A}$. Using the decomposition

$$Q_{\ell,\mathcal{R}_T(s);T} Q_{m,\mathcal{R}_T(s);T} - Q_{\ell,\mathcal{R}_T(s)} Q_{m,\mathcal{R}_T(s)} = \left(Q_{m,\mathcal{R}_T(s);T} - Q_{m,\mathcal{R}_T(s)}\right) Q_{\ell,\mathcal{R}_T(s)} + \left(Q_{\ell,\mathcal{R}_T(s);T} - Q_{\ell,\mathcal{R}_T(s)}\right) Q_{m,\mathcal{R}_T(s);T},$$

the first term of the right hand side of (B.22) is split into two terms. On $\mathcal{A}$, the first of these two terms is bounded as follows (the other term is bounded similarly):

$$
\begin{aligned}
&2\sum_{h,t=0}^{T-1}\sum_{s,u=0}^{T-1}\Big|U_{ts}^{(j)}U_{tu}^{(j)}\sum_{m=-\log_2 T}^{-1}\left(Q_{m,\mathcal{R}_T(s);T}-Q_{m,\mathcal{R}_T(s)}\right)\Psi_m(u-h)\sum_{\ell=-M_T}^{-1}Q_{\ell,\mathcal{R}}\Psi_\ell(s-h)\Big|\mathbb{I}_{|s-u|\leqslant 2M_T}\\
&\quad\leqslant 2k_T\log_2(T)\sqrt{K_2|\mathcal{R}_TT|^{-1}}\sum_{h,t=0}^{T-1}\sum_{s,u=0}^{T-1}|U_{ts}^{(j)}U_{tu}^{(j)}|\cdot\Big|\sum_{\ell=-M_T}^{-1}Q_{\ell,\mathcal{R}_T(s)}\Psi_\ell(s-h)\Big|\mathbb{I}_{|s-u|\leqslant 2M_T}\\
&\quad\leqslant 2k_T\log_2(T)\sqrt{K_2|\mathcal{R}_TT|^{-1}}\sum_{t=0}^{T-1}\sum_{s,u=0}^{T-1}|U_{ts}^{(j)}U_{tu}^{(j)}|\mathbb{I}_{|s-u|\leqslant 2M_T}\cdot\sum_{h=-\infty}^{\infty}\sup_z\Big|\sum_{\ell=-\log_2 T}^{-1}S_\ell(z)\Psi_\ell(h)\Big|\\
&\quad= O\left(2^jM_Tk_T|\mathcal{R}_TT|^{-1/2}\,T^{-1}\log_2^3 T\right)\qquad\text{a.s. on }\mathcal{A}
\end{aligned}
$$

using Assumption (3.8) and Lemma 6. The result follows from Assumption 4. □

B.5 Proof of Theorem 1

By Lemma 7 and Proposition 4 and for T large enough, there exists of a random set $\mathcal{A}$ such that $1-\Pr(\mathcal{A})=o_T(1)$ and (3.13) holds on $\mathcal{A}$. Then, if $\mathcal{A}^c$ denotes the complementary random set of $\mathcal{A}$, we can write:

$$
\begin{aligned}
\Pr\left(|Q_{j,\mathcal{R};T}-Q_{j,\mathcal{R}}|>\tilde\sigma_{j,\mathcal{R},T}\eta\right)&=\Pr\left(|Q_{j,\mathcal{R};T}-Q_{j,\mathcal{R}}|>\tilde\sigma_{j,\mathcal{R},T}\eta|\mathcal{A}\right)\Pr(\mathcal{A})\\
&\quad+\Pr\left(|Q_{j,\mathcal{R};T}-Q_{j,\mathcal{R}}|>\tilde\sigma_{j,\mathcal{R},T}\eta|\mathcal{A}^c\right)(1-\Pr(\mathcal{A}))\,.
\end{aligned}
$$

The second term of this sum is bounded using Lemma 7. To bound the first term, we observe that Proposition 4 implies $\tilde\sigma^2_{j,\mathcal{R},T}\geqslant\sigma^2_{j,\mathcal{R},T}-\varphi_T$ on $\mathcal{A}$ with $\varphi_T=o_T(2^{j-J_T}T^{-1})$. Together with Proposition 2, this implies

$$
\frac{\tilde\sigma^2_{j,\mathcal{R},T}}{\sigma^2_{j,\mathcal{R},T}}\geqslant 1-\frac{\varphi_T}{\sigma^2_{j,\mathcal{R},T}}=1-o_T(1)\to 1 \tag{B.23}
$$

for all $j=-1,\dots,-J_T$, as T tends to infinity. Then, we can write:

$$
\Pr\left(|Q_{j,\mathcal{R};T}-Q_{j,\mathcal{R}}|>\tilde\sigma_{j,\mathcal{R},T}\eta\right)\ \leqslant\ \Pr\left(|Q_{j,\mathcal{R};T}-Q_{j,\mathcal{R}}|>\sigma_{j,\mathcal{R},T}\eta\sqrt{1-\frac{\varphi_T}{\sigma^2_{j,\mathcal{R},T}}}\,\Big|\mathcal{A}\right)\ +\ o_T(1).
$$

and Proposition 3 leads to the result. □

B.6 Proof of Proposition 5

We first prove the following lemma, stating an exponential inequality for quadratic forms of Gaussian random variables. This result is a generalisation of a similar result obtained by Laurent and Massart (2000) for chi-squared distributions, and is proved in the spirit of Spokoiny (2001, Appendix).

Lemma 8. *Let $\underline{Z}_T=(Z_1,\dots,Z_T)'$ be a vector of iid Gaussian random variables with zero mean and $\operatorname{Var} Z_1=1$. If M_T is a $T\times T$ symmetric and positive definite matrix, then*

$$
\Pr\left(\underline{Z}_T'M_T\underline{Z}_T\leqslant\eta\right)\leqslant\exp\left(-\frac{(\eta-\operatorname{tr}M_T)^2}{4\|M_T\|_2^2}\right).
$$

provided that $\eta\leqslant\operatorname{tr}M_T$.

Proof. By assumption on the matrix M_T, the decomposition $M_T = O_T'\Lambda_T O_T$ holds with a diagonal $T\times T$ matrix Λ_T and an orthonormal matrix O_T. If we denote $\underline{Y}_T = U_T'\underline{Z}_T$, then $\underline{Y}_T$ is a vector of iid Gaussian random variables with zero mean and $\operatorname{Var} Y_1 = 1$. We can write $\underline{Z}_T' M_T \underline{Z}_T = \underline{Y}_T'\Lambda_T\underline{Y}_T = \sum_{i=1}^T \lambda_i Y_i^2$ with $\lambda_i > 0$. Moreover, $\operatorname{tr} M_T = \operatorname{tr}\Lambda_T$, $\operatorname{tr}\Lambda_T^2 = \operatorname{tr} M_T^2 = \|M_T\|_2^2$ and $\|M_T\|_{\text{spec}} = \max\{\lambda_1,\dots,\lambda_T\}$. A Chernoff bound on $\underline{Y}_T$ leads to

$$\begin{aligned}\Pr\left(\underline{Z}_T' M_T \underline{Z}_T \leqslant \eta\right) = \Pr\left(\underline{Y}_T'\Lambda_T\underline{Y}_T \leqslant \eta\right) &\leqslant \exp\left[\inf_{t<0}\left(-t\eta + \log \operatorname{E}\exp(t\underline{Y}_T'\Lambda_T\underline{Y}_T)\right)\right]\\ &= \exp\left[\inf_{t<0}\left(-t\eta + \sum_{i=1}^T \log\operatorname{E}\exp(\lambda_i t Y_i^2)\right)\right]\end{aligned}$$

and, using that

$$\log\operatorname{E}\exp(\alpha_i Y_i^2) = -\frac{1}{2}\log(1-2\alpha_i) \leqslant \alpha_i + \alpha_i^2$$

holds for $\alpha_i \leqslant 0$, we get

$$\Pr\left(\underline{Z}_T' M_T \underline{Z}_T \leqslant \eta\right) \leqslant \exp\left[\inf_{t<0}\left(-t\eta + t\operatorname{tr}\Lambda_T + t^2\operatorname{tr}\Lambda_T^2\right)\right].$$

The result follows by taking $t = (\eta - \operatorname{tr}\Lambda_T)/(2\operatorname{tr}\Lambda_T^2)$. □

Lemma 8 is not directly applicable on the quadratic form $Q_{j,\mathcal{R};T} = \underline{Z}_T' M_{j,\mathcal{R};T}\underline{Z}_T$ because the matrix $M_{j,\mathcal{R};T}$ is not definite positive in general. In the next lemma, we show how this matrix can be approximated by the matrix $M_{j,\mathcal{R};T}^\star$, defined as

$$M_{j,\mathcal{R};T}^\star = \Sigma_T^{1/2\,\prime} U_{j,\mathcal{R};T}^\star \Sigma_T^{1/2},$$

where the entry (s,t) of the matrix $U_{j,\mathcal{R};T}^\star$ is

$$u_{st}^\star = \gamma_0 |\mathcal{R}T|^{-1} \sum_{\ell=-\log_2 T}^{-1} 2^{\ell/2}\Psi_\ell(s-t),$$

with $\gamma_0 \geqslant \sup_{j<0}\sup_{\ell<0} 2^{-\ell/2}|A_{j\ell}^{-1}| > 0$. The matrix $M_{j,\mathcal{R};T}^\star$ is clearly symmetric. It is also positive definite because $U_{j,\mathcal{R};T}^\star$ is positive definite: For all sequences $\underline{x} = (x_1,\dots,x_T)'$ of ℓ^2, the quadratic form

$$x' U_{j,\mathcal{R};T}^\star x = \gamma_0|\mathcal{R}T|^{-1} \sum_{\ell=-\log_2 T}^{-1} 2^{\ell/2} \sum_{s=0}^{T-1}\left(\sum_{k\in\mathcal{R}T} x_s \psi_{\ell k}(s)\right)^2$$

is strictly positive.

Lemma 9. *Assume Assumptions 1 to 3 and Assumption 5 hold true. Define γ_1 such that*

$$0 < \gamma_1 < \gamma_0 \inf_{m<0} \sum_{\ell=-\log_2 T}^{-1} 2^{\ell/2} A_{m\ell}.$$

The following properties hold true for T sufficiently large:

$$\gamma_1 |\mathcal{R}|^{-1}\varepsilon \leqslant \operatorname{tr}(M_{j,\mathcal{R};T}^\star - M_{j,\mathcal{R};T}) \leqslant 2\|c_{X,T}\|_{1,\infty}\gamma_0|\mathcal{R}|^{-1} \tag{B.24}$$

where ε is defined in Assumption 2,

$$\|M^{\star}_{j,\mathcal{R};T} - M_{j,\mathcal{R};T}\|^2_{\text{spec}} \leqslant \|M^{\star}_{j,\mathcal{R};T} - M_{j,\mathcal{R};T}\|^2_2 \\ \leqslant 4\mathcal{L}_{-1}\gamma_0^2|\mathcal{R}|^{-2}\|c_X\|^2_{1,\infty}T^{-1}\log_2^2(T) + O(T^{-1}), \tag{B.25}$$

and, if $\underline{Z}_T = (Z_1, \ldots, Z_T)'$ is a vector of iid Gaussian random variables with zero mean and $\operatorname{Var} Z_1 = 1$*, then*

$$\Pr\left(\underline{Z}'_T(M^{\star}_{j,\mathcal{R};T} - M_{j,\mathcal{R};T})\underline{Z}_T > \lambda_T\right) = O\left(\exp\left[-\frac{\sqrt{T}\operatorname{tr} M_{j,\mathcal{R};T}}{\log_2^2 T}\right]\right) \tag{B.26}$$

where $\lambda_T = \operatorname{tr} M^{\star}_{j,\mathcal{R};T} - \operatorname{tr} M_{j,\mathcal{R};T} + \operatorname{tr} M_{j,\mathcal{R};T}\log_2^{-1} T$.

Proof. 1. We prove (B.24). Write $\operatorname{tr}(M^{\star}_{j,\mathcal{R};T} - M_{j,\mathcal{R};T}) = \operatorname{tr}(M^{\star}_{j,\mathcal{R};T}) - \operatorname{tr}(M_{j,\mathcal{R};T})$, where the second term is $\operatorname{E}(\underline{Z}'_T M_{j,\mathcal{R};T}\underline{Z}_T) = Q_{j,\mathcal{R}} + O(2^{j/2}T^{-1/2})$ from Lemma 3 and Proposition 2. Moreover,

$$\begin{aligned}\operatorname{tr}(M^{\star}_{j,\mathcal{R};T}) &= \operatorname{tr}\left(\Sigma'_T U^{\star}_{j,\mathcal{R};T}\right) \\ &= \gamma_0|\mathcal{R}T|^{-1}\sum_{s,u=\infty}^{\infty} c_{X,T}\left(\frac{s}{T},u\right)\sum_{\ell=-\log_2 T}^{-1} 2^{\ell/2}\Psi_\ell(u) \qquad (B.27)\\ &= \gamma_0|\mathcal{R}T|^{-1}\sum_{s,u=\infty}^{\infty} c_X\left(\frac{s}{T},u\right)\sum_{\ell=-\log_2 T}^{-1} 2^{\ell/2}\Psi_\ell(u) + \operatorname{Rest}_T. \qquad (B.28)\end{aligned}$$

We now derive a bound for Rest_T. Denote $\Delta_T(s/T,u) := c_{X,T}(s/T,u) - c_X(s/T,u)$. We first show that $\operatorname{TV}(\Delta_T(\cdot,u))$ is uniformly bounded in u. For all $I \in \{1,\ldots,T\}$ and for every sequence $0 < a_1 < a_2 < \ldots < a_I < 1$, we can write

$$\begin{aligned}&\Delta_T(a_i,u) - \Delta_T(a_{i-1},u)\\ &= \sum_{j=-\infty}^{-1}\sum_{k=-\infty}^{\infty}\left\{S_j\left(\frac{k}{T}\right) - S_j(a_i)\right\}\psi_{jk}([a_iT])\psi_{jk}([a_iT]+u)\\ &\quad - \sum_{j=-\infty}^{-1}\sum_{k=-\infty}^{\infty}\left\{S_j\left(\frac{k}{T}\right) - S_j(a_{i-1})\right\}\psi_{jk}([a_{i-1}T])\psi_{jk}([a_{i-1}T]+u) + O(T^{-1}),\end{aligned}$$

where the $O(T^{-1})$ term comes from the approximation (2.2). Now, substitute k by $k+[a_iT]$ in the first sum, and by $k+[a_{i-1}T]$ in the second one. This leads to

$$\begin{aligned}&\Delta_T(a_i,u) - \Delta_T(a_{i-1},u)\\ &= \sum_{j=-\infty}^{-1}\sum_{k=-\infty}^{\infty}\left\{S_j\left(\frac{k}{T}+a_i\right) - S_j\left(\frac{k}{T}+a_{i-1}\right) + S_j(a_{i-1}) - S_j(a_i)\right\}\psi_{jk}(0)\psi_{jk}(u)\\ &\quad + O(T^{-1}).\end{aligned}$$

Consequently, using the Cauchy-Schwarz inequality and (2.4),

$$\sum_{i=1}^{I}\left[\Delta_T(a_i,u) - \Delta_T(a_{i-1},u)\right] \leqslant 2\sum_{j=-\log_2 T}^{-1} L_j \sum_{k=-\infty}^{\infty}|\psi_{jk}(0)\psi_{jk}(u)| \leqslant 2\rho + K,$$

where K is a constant (because $I \leqslant T$), leading to $\mathrm{TV}(\Delta_T(\cdot, u)) \leqslant 2\rho + K$ uniformly in u. We can now bound Rest_T in (B.28) as follows:

$$\mathrm{Rest}_T = \gamma_0 |\mathcal{R}T|^{-1} \sum_{s,u=-\infty}^{\infty} \Delta_T\left(\frac{s}{T}, u\right) \sum_{\ell=-\log_2 T}^{-1} 2^{\ell/2} \Psi_\ell(u)$$

$$= \gamma_0 |\mathcal{R}|^{-1} \sum_{s,u=-\infty}^{\infty} \int_{s/T}^{(s+1)/T} dz \, \left\{ \Delta_T(z,u) + \Delta_T\left(\frac{s}{T}, u\right) - \Delta_T(z,u) \right\} \sum_{\ell=-\log_2 T}^{-1} 2^{\ell/2} \Psi_\ell(u) \ ,$$

as $|\Psi_\ell(u)|$ is uniformly bounded by 1,

$$\leqslant \gamma_0 |\mathcal{R}|^{-1} \int_0^1 dz \sum_{u=-\infty}^{\infty} |\Delta_T(z,u)|$$

$$+ \gamma_0 |\mathcal{R}|^{-1} \sum_{s,u=-\infty}^{\infty} \int_0^{1/T} dz \, \left| \Delta_T\left(\frac{s}{T}, u\right) - \Delta_T\left(z + \frac{s}{T}, u\right) \right| \ .$$

From Proposition 1, the first term is $O(|\mathcal{R}T|^{-1})$. Using (B.6) and that $\mathrm{TV}(\Delta_T(\cdot, u))$ is uniformly bounded in u, the second term is also $O(|\mathcal{R}T|^{-1})$.

Now, using (2.6) and the definition of the matrix A, the first term of (B.28) is bounded from below as follows:

$$\gamma_0 |\mathcal{R}T|^{-1} \sum_{s,u=-\infty}^{\infty} c_X\left(\frac{s}{T}, u\right) \sum_{\ell=-\log_2 T}^{-1} 2^{\ell/2} \Psi_\ell(u)$$

$$= \gamma_0 |\mathcal{R}T|^{-1} \sum_{s,u=-\infty}^{\infty} \sum_{m=-\infty}^{-1} S_m\left(\frac{s}{T}\right) \Psi_m(u) \sum_{\ell=-\log_2 T}^{-1} 2^{\ell/2} \Psi_\ell(u)$$

$$= \gamma_0 |\mathcal{R}T|^{-1} \sum_{s=0}^{T-1} \sum_{m=-\infty}^{-1} S_m\left(\frac{s}{T}\right) \inf_{m<0} \left[\sum_{\ell=-\log_2 T}^{-1} 2^{\ell/2} A_{m\ell} \right] \ ,$$

and we get the lower bound with Assumption 2. The upper bound is derived from (B.27), using Assumption 1, and $|\Psi_\ell(u)| \leqslant 1$ uniformly in $\ell < 0$ and $u \in \mathbb{Z}$.

2. We prove (B.25). The first inequality is (B.1). From (B.4), we write $\|M^\star_{j,\mathcal{R};T} - M_{j,\mathcal{R};T}\|_2^2 \leqslant \|\Sigma^{1/2}\|^4_{\mathrm{spec}} \|U^\star_{j,\mathcal{R};T} - U_{j,\mathcal{R};T}\|_2^2$. Then, with Lemma 4,

$$\|U^\star_{j,\mathcal{R};T} - U_{j,\mathcal{R};T}\|_2^2 \leqslant 2\|U^\star_{j,\mathcal{R};T}\|_2^2 + 2\|U_{j,\mathcal{R};T}\|_2^2$$

$$\leqslant \gamma_0^2 |\mathcal{R}|^{-2} T^{-1} \sum_{m,\ell=-\log_2 T}^{-1} 2^{(\ell+m)/2} A_{\ell m} + K_2 2^j |\mathcal{R}|^{-2} T^{-1}$$

with (A.5) and $\sqrt{\mathcal{L}_\ell \mathcal{L}_m} \leqslant 2^{-(\ell+m)/2} 4\mathcal{L}_{-1}$,

$$\leqslant 4\mathcal{L}_{-1} \gamma_0^2 |\mathcal{R}|^{-2} T^{-1} \log_2^2(T) + O(T^{-1}).$$

The result follows from Lemma 5.

3. We prove (B.26). For T large enough, λ_T is strictly positive. Using Proposition 8 and, from Lemma 3, using $p_T^2 = \mathrm{Var}(\underline{Z}'_T(M^\star_{j,\mathcal{R};T} - M_{j,\mathcal{R};T})\underline{Z}_T) = 2\|M^\star_{j,\mathcal{R};T} - M_{j,\mathcal{R};T}\|_2^2$ and denoting $q_T = \|M^\star_{j,\mathcal{R};T} - M_{j,\mathcal{R};T}\|_{\mathrm{spec}}$,

$$\Pr\left(\underline{Z}'_T(M^\star_{j,\mathcal{R};T} - M_{j,\mathcal{R};T})\underline{Z}_T > \lambda_T\right) \leqslant \exp\left(-\frac{1}{2}\cdot\frac{(\operatorname{tr} M_{j,\mathcal{R};T})^2}{p_T^2\log_2^2 T + 2q_T\operatorname{tr}(M_{j,\mathcal{R};T})\log_2 T}\right).$$

(B.25) gives the rates for p_T and q_T, leading to the result. □

Proof of Proposition 5. The type II error is

$$\Pr\left(|Q_{j,\mathcal{R};T}| \leqslant \eta_\star\tilde{\sigma}_{j,\mathcal{R},T}\middle|\,\mathrm{H}_1\right) \leqslant \Pr\left(Q_{j,\mathcal{R};T} \leqslant \eta_\star\tilde{\sigma}_{j,\mathcal{R},T}\middle|\,\mathrm{H}_1\right).$$

In the proof of Proposition 4, we define a random set $\mathcal{A}$ such that $\Pr(\mathcal{A}) \geqslant 1 - o_T(1)$ and (3.13) holds. Proposition 4 implies that $\tilde{\sigma}^2_{j,\mathcal{R},T} \leqslant \sigma^2_{j,\mathcal{R},T} + \gamma_T$ on $\mathcal{A}$ with $\gamma_T = o(2^{j-J_T}T^{-1})$. Together with Proposition 2, this implies

$$\frac{\tilde{\sigma}^2_{j,\mathcal{R},T}}{\sigma^2_{j,\mathcal{R},T}} \leqslant 1 + \frac{\gamma_T}{\sigma^2_{j,\mathcal{R},T}} \to 1,$$

and then the type II error is bounded by

$$\Pr\left(Q_{j,\mathcal{R};T} \leqslant \eta_T\sigma_{j,\mathcal{R},T}\middle|\,\mathrm{H}_1\right) + o_T(1)$$

with $\eta_T = \eta_\star\sqrt{1 + \gamma_T/\sigma^2_{j,\mathcal{R},T}}$.

With the notations of Lemma 9, we can write the type II error as

$$\Pr\left(Q_{j,\mathcal{R};T} \leqslant \eta_T\sigma_{j,\mathcal{R},T}\middle|\,\mathrm{H}_1\right) = \Pr\left(\underline{Z}'_T M^\star_{j,\mathcal{R};T}\underline{Z}_T \leqslant \eta_T\sigma_{j,\mathcal{R},T} + Z'_T(M^\star_{j,\mathcal{R};T} - M_{j,\mathcal{R};T})Z_T\middle|\,\mathrm{H}_1\right) \tag{B.29}$$

where $\underline{Z}_T = (Z_1, \ldots, Z_T)'$ is a vector of iid Gaussian random variables with zero mean and $\mathrm{Var}\, Z_1 = 1$. We now define the random set $\mathcal{P}_T = \{Z'_T(M^\star_{j,\mathcal{R};T} - M_{j,\mathcal{R};T})Z_T \leqslant \lambda_T\}$ with $\lambda_T = \operatorname{tr} M^\star_{j,\mathcal{R};T} - \operatorname{tr} M_{j,\mathcal{R};T}(1 - \log_2^{-1} T)$. Lemma 9, equation (B.26), gives an upper bound for $\Pr(\mathcal{P}^c)$. Conditioning on $\mathcal{P}_T$, we can write

$$\Pr\left(Q_{j,\mathcal{R};T} \leqslant \eta_T\sigma_{j,\mathcal{R},T}\middle|\,\mathrm{H}_1\right) \leqslant \Pr\left(\underline{Z}'_T M^\star_{j,\mathcal{R};T}\underline{Z}_T \leqslant \eta_T\sigma_{j,\mathcal{R},T} + \lambda_T\middle|\,\mathrm{H}_1\right) + O\left(\exp\left[-\frac{\sqrt{T}\operatorname{tr} M_{j,\mathcal{R};T}}{\log_2^2 T}\right]\right).$$

Then, we are in position to apply Lemma 8 and we get

$$\Pr\left(\underline{Z}'_T M^\star_{j,\mathcal{R};T}\underline{Z}_T \leqslant \eta_T\sigma_{j,\mathcal{R},T} + \lambda_T\middle|\,\mathrm{H}_1\right) \leqslant \exp\left[-\frac{\left(\eta_T\sigma_{j,\mathcal{R},T} - \operatorname{tr} M_{j,\mathcal{R};T}(1 - \log_2^{-1} T)\right)^2}{4\|M^\star_{j,\mathcal{R};T}\|_2^2}\right]$$

provided that $\eta_T\sigma_{j,\mathcal{R};T} + \lambda_T \leqslant \operatorname{tr} M^\star_T$ (which holds true for T large enough by definition of λ_T). Proposition 2 allows to write $\mathrm{E}Q_{j,\mathcal{R};T} = \operatorname{tr} M_{j,\mathcal{R};T} = Q_{j,\mathcal{R}} + r_T$ with $r_T = O(T^{-1/2})$, and then the first term of the type II error is bounded in order by

$$\exp\left[-\|M^\star_{j,\mathcal{R};T}\|_2^{-2} Q^2_{j,\mathcal{R}}\right] + \exp\left[-\sqrt{T} Q_{j,\mathcal{R}}\log_2^{-2} T\right].$$

We conclude using that, under H_1, we can write

$$Q_{j,\mathcal{R}} = |\mathcal{U}||\mathcal{R}|^{-1}Q_{j,\mathcal{U}} + (|\mathcal{R}| - |\mathcal{U}|)|\mathcal{R}|^{-1}Q_{j,\mathcal{R}\setminus\mathcal{U}} > |\mathcal{U}||\mathcal{R}|^{-1}\theta,$$

from (3.16) and applying Lemma 9, equation (B.25) to $\|M^\star_{j,\mathcal{R};T}\|_2$. □

B.7 Proof of Proposition 6

Let $\mathcal{U}$ be a segment of $\wp(\mathcal{R})$. Consider the a.s. inequality

$$|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{U};T}| \leqslant |Q_{j,\mathcal{U};T} - Q_{j,\mathcal{U}}| + |Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}| + \Delta_j(\mathcal{R},\mathcal{U})$$

where $\Delta_j(\mathcal{R},\mathcal{U})$ is defined in (4.1). In the regular case, $\Delta_j(\mathcal{R},\mathcal{U}) \leqslant b(\mathcal{U}) + b(\mathcal{R}) \leqslant C_j(\sigma_{j,\mathcal{U},T} + \sigma_{j,\mathcal{R},T})k_T$. Consequently, in the regular case,

$$\begin{aligned}\Pr(\mathcal{R} \text{ is rejected}) &\leqslant \sum_{\mathcal{U}\in\wp(\mathcal{R})} \Pr\left[|Q_{j,\mathcal{U};T} - Q_{j,\mathcal{R};T}| > (\eta\sigma_{j,\mathcal{U},T} + \eta\sigma_{j,\mathcal{R};T})\, k_T\right] \\ &\leqslant \sum_{\mathcal{U}\in\wp(\mathcal{R})} \Pr\left(|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}| > -C_j\sigma_{j,\mathcal{U};T}k_T + \eta\sigma_{j,\mathcal{U};T}k_T\right) \\ &\quad + \sum_{\mathcal{U}\in\wp(\mathcal{R})} \Pr\left(|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{R}}| > -C_j\sigma_{j,\mathcal{R},T}k_T + \eta\sigma_{j,\mathcal{R},T}k_T\right)\end{aligned}$$

Proposition 3 implies

$$\begin{aligned}\Pr(\mathcal{R} \text{ is rejected}) \leqslant (\sharp\wp(\mathcal{R}))\, c_0 \exp&\left\{-\frac{1}{8}\cdot\frac{\eta_T^2}{1 + \frac{\eta_T L_j}{|\mathcal{R}T|\sigma_{j,\mathcal{R},T}} + \frac{2^{j/2}\eta_T\nu(\|c_X\|_{1,\infty}+c_1\rho)}{\sigma_{j,\mathcal{R};T}|\mathcal{R}_T|\sqrt{T}}}\right\} \\ + c_0 \sum_{\mathcal{U}\in\wp(\mathcal{R})} \exp&\left\{-\frac{1}{8}\cdot\frac{\eta_T^2}{1 + \frac{\eta_T L_j}{|\mathcal{U}T|\sigma_{j,\mathcal{U},T}} + \frac{2^{j/2}\eta_T\nu(\|c_X\|_{1,\infty}+c_1\rho)}{\sigma_{j,\mathcal{U};T}|\mathcal{U}|\sqrt{T}}}\right\}\end{aligned}$$

with

$$\begin{aligned}\eta_T &:= \eta k_T\sqrt{1-\varphi_T} - C_j k_T \\ &= k_T 2^{-j/2}\left[5(2\alpha+p) - \sqrt{\alpha+p}\right].\end{aligned}$$

Proposition 2 leads to $\sigma_{j,\mathcal{R},T}^{-1} \leqslant 2^{-j}\sqrt{K_1^{-1}|\mathcal{R}T|}$ and similarly for $\sigma_{j,\mathcal{U},T}^{-1}$. As $\delta \leqslant |\mathcal{R}| \leqslant |\mathcal{U}| \leqslant 1$, we consider the dominant terms in the sum, and we can write, for T large enough, and with $2^{-j}L_j \leqslant \rho\mathcal{L}_{-1}$,

$$\Pr(\mathcal{R} \text{ is rejected}) \leqslant 2c_0\,(\sharp\wp(\mathcal{R}))\exp\left\{-\frac{1}{8}\cdot\frac{\eta_T^2}{1 + \frac{\eta_T\rho\mathcal{L}_{-1}}{\sqrt{K_1|\mathcal{R}T|}} + \frac{\eta_T\nu(\|c_X\|_{1,\infty}+c_1\rho)}{\sqrt{2^j K_1\delta}}}\right\}.$$

Replacing η_T and using $2\alpha+p \geqslant \sqrt{\alpha+p}$ lead to the result. □

B.8 Proof of Theorem 2

We first prove the following technical lemma.

Lemma 10. *Let $\underline{Z}_T = (Z_1,\dots,Z_T)'$ be a vector of iid Gaussian random variables with zero mean and* $\operatorname{Var} Z_1 = 1$*. If $M_{j,\mathcal{R},T}$ is the matrix (B.16), v is a positive constant and $p \geqslant 2$, then, there exists T_0 such that*

$$\mathrm{E}\left(\underline{Z}_T' M_{j,\mathcal{R},T}\underline{Z}_T - \operatorname{tr} M_{j,\mathcal{R},T} + vk_T T^{-1/2}\right)^p \leqslant C\left(\kappa,\nu,\|c_X\|_{1,\infty},p\right) T^{-p/2}\left(2^{1+j/2}|\mathcal{R}|^{-1} + vk_T\right)^p$$

for all $T \geqslant T_0$.

Proof. First, we write

$$\mathrm{E}\left(\underline{Z}_T' M_{j,\mathcal{R},T}\underline{Z}_T - \operatorname{tr} M_{j,\mathcal{R},T} + vk_T T^{-1/2}\right)^p$$
$$= \sum_{r=0}^{p} \begin{pmatrix} p \\ r \end{pmatrix} \mathrm{E}\left(\underline{Z}_T M_{j,\mathcal{R},T}\underline{Z}_T - \operatorname{tr} M_{j,\mathcal{R},T}\right)^r v^{p-r} k_T^{p-r} T^{-(p-r)/2} \ . \quad \text{(B.30)}$$

Due to the relationship between the centered moments of a random variable and its cumulants, we can write

$$\mathrm{E}\left(\underline{Z}_T' M_{j,\mathcal{R},T}\underline{Z}_T - \operatorname{tr} M_{j,\mathcal{R},T}\right)^r = \sum_{m=0}^{r}\sum C(p_1,\ldots,p_m,m,\pi_1,\ldots,\pi_m,r)\kappa_{p_1}^{\pi_1}\ldots\kappa_{p_m}^{\pi_m} \ ,$$

where the second sum is over $p_1,\ldots,p_m,\pi_1,\ldots,\pi_m$ in $\{1,\ldots,r\}$ such that $\sum_{i=1}^m p_i\pi_i = r$, κ_{p_i} is the p_ith cumulant of $\underline{Z}_T M_{j,\mathcal{R},T}\underline{Z}_T$ and C denotes a generic constant in this proof. From Lemma 3, we can write, using (B.17) and Proposition 2:

$$\begin{aligned}\kappa_{p_i} &\leqslant 2^{p_i-2}(p_i-1)!\operatorname{Var}(\underline{Z}_T M_{j,\mathcal{R},T}\underline{Z}_T)\|M_{j,\mathcal{R},T}\|_{\text{spec}}^{p_i-2} \\ &\leqslant 2^{p_i-2}(p_i-1)!K_2\nu^{p_i-2}\|c_X\|_{1,\infty}^{p_i-2}2^{jp_i/2}|\mathcal{R}|^{-p_i}T^{-p_i/2}.\end{aligned}$$

Consequently,

$$\mathrm{E}\left(\underline{Z}_T' M_{j,\mathcal{R},T}\underline{Z}_T - \operatorname{tr} M_{j,\mathcal{R},T}\right)^r \leqslant C(\kappa,\|c_X\|_{1,\infty},r,\nu)2^{r(1+j/2)}|\mathcal{R}|^{-r}T^{-r/2} \ ,$$

and using this inequality in (B.30) leads to the result. □

Proof of Theorem 2. Let $\tilde{\mathcal{R}}$ be the interval selected by the estimation procedure. We consider two cases: $|\tilde{\mathcal{R}}| < |\mathcal{R}|$ or $|\tilde{\mathcal{R}}| \geqslant |\mathcal{R}|$. In the first case, and since the estimator and the wavelet spectrum are uniformly bounded by S,

$$\mathrm{E}|\tilde{S}_j(z_0) - S_j(z_0)|^p \, \mathbb{I}_{|\tilde{\mathcal{R}}|<|\mathcal{R}|} \leqslant (2S)^p \ \Pr\left(|\tilde{\mathcal{R}}| < |\mathcal{R}|\right) .$$

As $\Pr(|\tilde{\mathcal{R}}| < |\mathcal{R}|) \leqslant \Pr(\mathcal{R} \text{ is rejected})$, Proposition 6 allows to write $\mathrm{E}|\tilde{S}_j(z_0) - S_j(z_0)|^p \, \mathbb{I}_{|\tilde{\mathcal{R}}|<|\mathcal{R}|} = O(T^{-cp\sqrt{\delta}})$. We consider now the second case. Select a subinterval $\mathcal{U}$ in $\mathcal{R}\cap\wp(\tilde{\mathcal{R}})$ containing z_0. Then, consider the decomposition

$$\mathrm{E}|\tilde{S}_j(z_0) - S_j(z_0)|^p \, \mathbb{I}_{|\tilde{\mathcal{R}}|\geqslant|\mathcal{R}|} \leqslant \mathrm{E}\left[|Q_{j,\mathcal{U}} - S_j(z_0)| + |Q_{j,\mathcal{U};T} - Q_{j,\mathcal{U}}| + |Q_{j,\tilde{\mathcal{R}};T} - Q_{j,\mathcal{U};T}|\right]^p .$$

As the wavelet spectrum is regular on $\mathcal{U}\subset\mathcal{R}$, the term $|Q_{j,\mathcal{U}} - S_j(z_0)|$ is bounded by $C_j\sigma_{j,\mathcal{U},T}k_T$. On the other hand, using Proposition 2, $|Q_{j,\mathcal{U};T} - Q_{j,\mathcal{U}}| = |Q_{j,\mathcal{U};T} - \operatorname{tr} M_{j,\mathcal{U};T}| + R_T$ with $R_T = O(2^{j/2}T^{-1/2})$. Moreover, as $\tilde{\mathcal{R}}$ is selected by the estimation procedure, it holds $|Q_{j,\tilde{\mathcal{R}};T} - Q_{j,\mathcal{U};T}| \leqslant (\eta\tilde{\sigma}_{j,\tilde{\mathcal{R}},T} + \eta\tilde{\sigma}_{j,\mathcal{U},T})k_T$. Finally,

$$\begin{aligned}&\mathrm{E}|\tilde{S}_j(z_0) - S_j(z_0)|^p \, \mathbb{I}_{|\tilde{\mathcal{R}}|\geqslant|\mathcal{R}|} \\ &\qquad\leqslant \mathrm{E}\left[|Q_{j,\mathcal{U};T} - \operatorname{tr} M_{j,\mathcal{U};T}| + R_T + C_j\sigma_{j,\mathcal{U},T}k_T + \left(\eta\sigma_{j,\tilde{\mathcal{R}},T} + \eta\sigma_{j,\mathcal{U},T}\right)k_T\right]^p \ .\end{aligned}$$

With $2\alpha + p \geqslant \sqrt{\alpha + p}$, we can write

$$C_j\sigma_{j,\mathcal{U},T}k_T + \left(\eta\sigma_{j,\tilde{\mathcal{R}},T} + \eta\sigma_{j,\mathcal{U},T}\right)k_T \leqslant 11(2\alpha+p)K_2^{1/2}|\mathcal{U}T|^{-1/2}k_T \ .$$

Lemma 10 proves the existence of a constant c_5 depending on κ, ν, p, K_2 and on $\|c_X\|_{1,\infty}$, such that, for $T \geqslant T_0$,

$$\mathrm{E}\left[|Q_{j,\mathcal{U};T} - \operatorname{tr} M_{j,\mathcal{U};T}| + R_T + C_j\sigma_{j,\mathcal{U},T}\alpha_T + \left(\eta\sigma_{j,\tilde{\mathcal{R}},T} + \lambda\sigma_{j,\mathcal{U},T}\right)k_T\right]^p \leqslant c_5|\mathcal{U}T|^{-p/2}\left[2^{1+j/2}\delta^{-1} + 11(2\alpha+p)k_T\right]^p$$

since $|\tilde{\mathcal{R}}| \geqslant |\mathcal{U}| \geqslant \delta$, and the result follows. □

B.9 Proof of Proposition 7

Suppose $|\mathcal{R}_0| \vee |\mathcal{R}_1| = |\mathcal{R}_0|$ w.l.o.g. and write

$$\Pr(\mathcal{R} \text{ is not rejected}) \leqslant \Pr\left[Q_{j,\mathcal{R},T} - Q_{j,\mathcal{R}_0,T} \leqslant \eta(\sigma_{j,\mathcal{R},T} + \sigma_{j,\mathcal{R}_0,T})k_T\right] .$$

As in the proof of Proposition 5, we approximate $M_T = \Sigma_T'(U_{j,\mathcal{R};T} - U_{j,\mathcal{R}_0;T})\Sigma_T$ by $M_T^\star = \Sigma_T'(U^\star_{j,\mathcal{R};T} - U^\star_{j,\mathcal{R}_0;T})\Sigma_T$, where $U^\star_{j,\mathcal{R};T}$ is defined in Lemma 9. Define the random set $\mathcal{P}_T = \{Z_T'(M_T^\star - M_T)Z_T \leqslant \lambda_T\}$ with $\lambda_T = \operatorname{tr} M_T^\star - (1 - \log_2^{-1} T)\operatorname{tr} M_T$, where $\underline{Z}_T = (Z_1, \ldots, Z_T)'$ is a vector of iid Gaussian random variables. As in the proof of Lemma 9, equation (B.26), we have

$$\Pr(\mathcal{P}_T^c) = O\left(\exp\left[-\frac{\sqrt{T}\operatorname{tr} M_T}{\log_2^2 T}\right]\right) .$$

Conditioning on $\mathcal{P}_T$, we can write

$$\Pr(\mathcal{R} \text{ is not rejected}) \leqslant \Pr\left[\underline{Z}_T' M_T^\star \underline{Z} \leqslant \eta(\sigma_{j,\mathcal{R},T} + \sigma_{j,\mathcal{R}_0,T})k_T + \lambda_T\right] + O\left(\exp\left[-\frac{\sqrt{T}\operatorname{tr} M_T}{\log_2^2 T}\right]\right) .$$

We are now in position to apply Lemma 8 with the symmetric, positive definite matrix $M^\star_{j,\mathcal{R};T}$. As $\eta(\sigma_{j,\mathcal{R},T} + \sigma_{j,\mathcal{R}_0,T})k_T + \lambda_T \leqslant \operatorname{tr} M_T^\star$ for T large enough, we can write, with Lemma 8,

$$\Pr(\mathcal{R} \text{ is not rejected}) \leqslant \exp\left[-\frac{1}{2}\frac{(\eta(\sigma_{j,\mathcal{R},T} + \sigma_{j,\mathcal{R}_0,T})k_T + \lambda_T - \operatorname{tr} M_T^\star)^2}{\operatorname{Var}\left(\underline{Z}_T' M_T^\star \underline{Z}\right)}\right] + O\left(\exp\left[-\frac{\sqrt{T}\operatorname{tr} M_T}{\log_2^2 T}\right]\right).$$

Lemma 9 leads to $\operatorname{Var}(\underline{Z}_T' M_T^\star \underline{Z}) = O(|\mathcal{R}_0|T^{-1}\log_2^2 T)$, and then, replacing λ_T, the rate of the probability becomes

$$O\left(\exp\left[-\frac{T(\operatorname{tr} M_T)^2|\mathcal{R}_0|}{\log_2^2 T}\right] + \exp\left[-\frac{\sqrt{T}\operatorname{tr} M_T}{\log_2^2 T}\right]\right).$$

The result follows using $\operatorname{tr} M_T = \theta_T$. □

Acknowledgements

The authors would like to thank Rainer Dahlhaus and Suhasini Subba Rao for helpful remarks.

References

Abramovich, F., Benjamini, Y., Donoho, D. and Johnstone, I. (2000). *Adapting to unknown sparsity by controlling the false discovery rate* (Technical Report No. 2000-19). Stanford University: Dept. of Statistics. (http://www.math.tau.ac.il/~felix/ltx/PAPERS/Annals.ps.gz)

Berkner, K. and Wells, R. (2002). Smoothness estimates for soft-threshold denoising via translation-invariant wavelet transforms. *Appl. Comput. Harmon. Anal.*, *12*, 1–24.

Brillinger, D. (1975). *Time Series. Data Analysis and Theory.* Holt, Rinehart and Winston, Inc.

Brown, L. D. and Low, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.*, *24*, 2524–2535.

Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.*, *62*, 139–168.

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

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.*, *28*, 1762–1794.

Dahlhaus, R. and Polonik, W. (2002). Empirical spectral processes and nonparametric maximum likelihood estimation for time series. In H. Dehling, T. Mikosch and M. Sørensen (Eds.), *Empirical Process Techniques for Dependent Data.* New York: Springer-Verlag.

Fryźlewicz, P. (2002). *Modelling and forecasting financial log-returns as locally stationary wavelet processes* (Research Report). Department of Mathematics, University of Bristol. (http://www.stats.bris.ac.uk/pub/ResRept/2002/14.pdf)

Fryźlewicz, P., Van Bellegem, S. and von Sachs, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Ann. Inst. Statist. Math.* (To appear)

Giurcanu, M. and Spokoiny, V. (2002). *Confidence estimation of the covariance function of stationary and locally stationary processes* (Preprint No. 726). Berlin: WIAS. (http://www.math.tu-berlin.de/~giurcanu/document.pdf)

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, *28*, 1302–1338.

Lepski, O. (1990). On a problem of adaptive estimation in Gaussian white noise. *Theory of Probab. Appl.*, *35*, 454–470.

Lepski, O. and Spokoiny, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.*, *25*, 2512–2546.

Lii, K.-S. and Rosenblatt, M. (2002). Spectral analysis for harmonizable processes. *Ann. Statist.*, *30*, 258–297.

Los, C. A. (2000). Nonparametric efficiency testing of Asian stock markets using weekly data. *Adv. in Econometrics*, *14*, 329–363.

Nason, G. and Sapatinas, T. (2002). Wavelet packet transfer function modelling of nonstationary time series. *Statist. Comput.*, *12*, 45–56.

Nason, G. and Silverman, B. (1995). The stationary wavelet transform and some statistical applications. In A. Antoniadis and G. Oppenheim (Eds.), *Wavelets and Statistics* (Vol. 103, pp. 281–299). New York: Springer-Verlag.

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

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

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

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

Priestley, M. (1981). *Spectral Analysis of Time Series.* Academic Press, London.

von Sachs, R., Nason, G. and Kroisandt, G. (1997). *Adaptive estimation of the evolutionary wavelet spectrum* (Technical Report No. 516). Stanford: Department of Statistics. (`http://www.stats.bris.ac.uk/pub/reports/Wavelets/StanTechRep516.ps.gz`)

Saito, N. and Beylkin, G. (1993). Multiresolution representations using the autocorrelation functions of compactly supported wavelets. *IEEE Trans. Signal Process.*, *41*, 3584–3590.

Spokoiny, V. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive choice. *Ann. Statist.*, *26*, 1356–1378.

Spokoiny, V. (2001). Data driven testing the fit of linear models. *Math. Methods Statist.*, *10*, 465–497.

Swanson, N. R. and White, H. (1997). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *Int. J. Forecasting*, *13*, 439–461.

Vidakovic, B. (1999). *Statistical modeling by wavelets.* John Wiley and Sons.

Figure

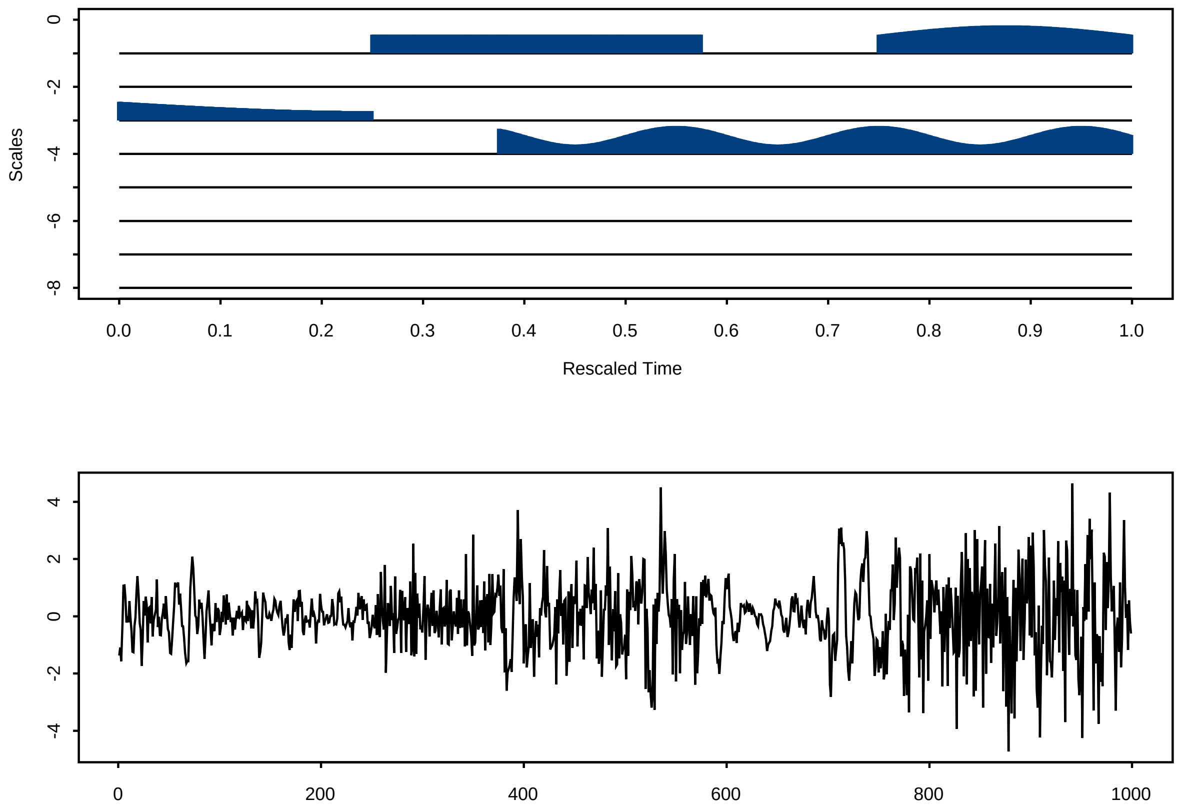

Figure 1: The first figure is an example of theoretical spectrum $S_j(z)$. This spectrum is used in the second figure to simulate a locally stationary wavelet process. This simulation uses Gaussian innovations ξ_{jk} and non-decimated Haar wavelets.