JOURNAL OF TIME SERIES ANALYSIS
J. Time. Ser. Anal. **37**: 476–512 (2016)
Published online 10 November 2015 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12170

ORIGINAL ARTICLE

MULTIVARIATE WAVELET WHITTLE ESTIMATION IN LONG-RANGE DEPENDENCE

SOPHIE ACHARD[a,b] AND IRÈNE GANNAZ[c*]

[a] *Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble, France*
[b] *CNRS GIPSA-Lab, F-38000 Grenoble, France*
[c] *Université de Lyon, CNRS UMR 5208, INSA de Lyon, Institut Camille Jordan, Lyon, France*

Multivariate processes with long-range dependent properties are found in a large number of applications including finance, geophysics and neuroscience. For real-data applications, the correlation between time series is crucial. Usual estimations of correlation can be highly biased owing to phase shifts caused by the differences in the properties of autocorrelation in the processes. To address this issue, we introduce a semiparametric estimation of multivariate long-range dependent processes. The parameters of interest in the model are the vector of the long-range dependence parameters and the long-run covariance matrix, also called functional connectivity in neuroscience. This matrix characterizes coupling between time series. The proposed multivariate wavelet-based Whittle estimation is shown to be consistent for the estimation of both the long-range dependence and the covariance matrix and to encompass both stationary and nonstationary processes. A simulation study and a real-data example are presented to illustrate the finite-sample behaviour.

Received 3 November 2014; Revised 5 October 2015; Accepted 13 October 2015

Keywords: Multivariate processes; long memory; fractional integration; semiparametric estimation; covariance matrix; wavelets; neuroscience application; functional connectivity.
JEL 60G22; 62M10; 62M15; 62H20; 92C55.

1. INTRODUCTION

The long-range dependence (LRD) has attracted much interest in statistics and in many applications since the seminal article of Mandelbrot in 1950. First, the fractional Brownian motion model was introduced as the unique Gaussian process having stationary increments and self-similarity index H in $(0, 1)$ (Mandelbrot and Van Ness, 1968). This model is characterized by one parameter called the Hurst exponent. Since then, several extensions were introduced in order to obtain more complex modellings that better match real data, such as autoregressive fractionally integrated moving average (ARFIMA). We refer to Percival and Walden (2006) and references therein for an overview of LRD models. These models were used in a large scope of applications, for example finance (Gençay *et al.*, 2001) (see also the references in Nielsen and Frederiksen, 2005), Internet traffic analysis (Abry and Veitch, 1998), physical sciences (Percival and Walden, 2006; Papanicolaou and Sølna, 2003), geosciences (Whitcher and Jensen, 2000) and neuroimagery (Maxim *et al.*, 2005).

Nowadays, it is common to record data having multiple sensors, such as neuroimagery (functional magnetic resonance imaging or electroencephalography). Each sensor records the activity of a specific part of the brain. However, the brain is a complex system with complex interactions between its different parts, so researchers were interested in modelling the sensors as multivariate time series. A similar representation is suited for data acquired in geosciences where, for example, time series correspond to temperatures in several parts of the earth, like in Whitcher and Jensen (2000). For these two applications, it has been shown that the univariate

* Correspondence to: Irène Gannaz, Université de Lyon, CNRS UMR 5208, INSA de Lyon, Institut Camille Jordan, Lyon, France E-mail: irene.gannaz@insa-lyon.fr

time series present LRD behaviour. Several models accounting for long-memory features have been proposed. In Didier and Pipiras (2011), the multivariate Brownian motion was defined. The values of interactions as defined by the covariance matrix must be carefully chosen so that the model is identifiable (Coeurjolly *et al.*, 2013). Also, the multivariate extension of fractionally difference models was proposed by Chambers (1995), which includes the multivariate extension of ARFIMA models with an explicit expression of the short memory terms. In a recent article, Kechagias and Pipiras (2015) highlighted the difficulties to extend the notion of LRD to multivariate time series and proposed specific linear representations of LRD. Concerning multivariate ARFIMA models, Lobato (1997) and Sela and Hurvich (2008) studied two different classes of extension depending on the order of fractional integration and autoregressive moving average models.

Using these long-memory models, a typical statistical issue is to estimate the long-memory parameter. This characterizes the long-term dependence of the series, which controls many relevant statistical properties. A very large literature exists in the context of univariate time series. First, parametric approaches were considered (Fox and Taqqu, 1986; Dahlhaus, 1989; Giraitis *et al.*, 1997), which provide fast rates of convergence. However, these approaches suffer from inconsistency when the short-term component of the model is misspecified. Semiparametric models were then developed to be robust to model misspecification (Robinson 1994a, 1994b, 1995a, 1995b), where the spectral density is modelled only near zero frequency. In the frequency domain, two popular estimators among the semiparametric ones are the Geweke–Porter-Hudak introduced by Geweke and Porter-Hudak (1983) and the local Whittle estimator of Robinson (1995a). Wavelet-based estimators were also studied and proved to be adequate for studying fractal time series. In Abry and Veitch (1998), the authors developed an estimator using log-regression of the wavelet coefficient variance on the scale index. Moulines *et al.* (2008) derived the asymptotic properties of a wavelet Whittle estimator.

Considering multivariate fractionally integrated processes (Chambers, 1995), the estimation of memory parameters and covariance matrix has been first studied by Robinson (1995b). Then Lobato (1999) proposed a semiparametric two-step estimator. Shimotsu (2007) extended this latter approach including phase-shift consideration. Nielsen (2011) proposed an extension based on the Abadir *et al.* (2007) extended Fourier transform to estimate long-memory parameters for nonstationary time series. In a different approach, Sela and Hurvich (2012) proposed an estimator based on the average periodogram for a power law in coherency. All these approaches were developed using Fourier log-periodogram. In comparison, there are few wavelet-based estimators of long-range memory parameters in multivariate settings. Frías *et al.* (2008) and Wang and Wang (2013) proposed estimation schemes based on multidimensional wavelets. In many real-data applications such as geosciences, Internet traffic or neurosciences, the number of time series is huge, as the real-data example of Section 6 illustrates. The latter works thus do not seem well adapted. Achard *et al.* (2008) studied a two-dimensional estimation, based on univariate wavelets, which defines estimators using a regression of the cross-covariance between the wavelet coefficients. This approach also appears difficult to generalize to high multidimensional settings. The present work defines a wavelet Whittle estimator for multivariate models. The extension to multivariate processes presents two issues. First, a vector of long-memory parameters has to be estimated along with the covariance matrix that is modelling the interactions between the time series. Second, as noted by Robinson (1994b) and Shimotsu (2007), the multivariate extension of the fractional integrated model introduces a phase shift that has to be taken into account in the estimation procedures. The new wavelet-based proposed methodology is shown to be adequate for nonstationary LRD models.

The article is organized as follows. Section 2 introduces the specific framework of multivariate long-memory processes based on the definition of the spectral density matrix. The multivariate wavelet Whittle (MWW) estimators of both the long-memory parameters and the covariance matrix are defined in Section 3. The properties of this new estimation scheme are derived in Section 4 where consistency of both estimations is established. Finally, Section 5 presents a simulation study that illustrates that the wavelet Whittle estimators have comparable performances with the Fourier-based ones. In addition, our method provides a very flexible approach to handle both stationary and nonstationary processes. Section 6 deals with a real-data application in neuroscience.

DOI: 10.1111/jtsa.12170

2. THE SEMIPARAMETRIC MULTIVARIATE LONG-MEMORY FRAMEWORK

Let $\mathbf{X} = \{X_\ell(k), k \in \mathbb{Z}, \ell = 1, \ldots, p\}$ be a multivariate stochastic process. Each process X_ℓ is not necessarily stationary. Denote by ΔX_ℓ the first-order difference, $(\Delta X_\ell)(k) = X_\ell(k) - X_\ell(k-1)$, and by $\Delta^D X_\ell$ the Dth-order difference. For every component X_ℓ, there exists $D_\ell \in \mathbb{N}$ such that the D_ℓth-order difference $\Delta^{D_\ell} X_\ell$ is covariance stationary. Following Achard *et al.* (2008), Chambers (1995) and Moulines *et al.* (2007), we consider a long-memory process $\mathbf{X}$ with memory parameters $\mathbf{d} = (d_1, d_2, \ldots, d_p)$. For any $\mathbf{D} > \mathbf{d} - 1/2$, we suppose that the multivariate process $\mathbf{Z} = \mathrm{diag}(\Delta^{D_\ell}, \ell = 1, \ldots, p)\mathbf{X}$ is covariance stationary with a spectral density matrix given by

$$\text{for all } (\ell, m)\ , \ f_{\ell,m}^{(D_\ell, D_m)}(\lambda) = \frac{1}{2\pi}\Omega_{\ell,m}\left(1 - \mathrm{e}^{-i\lambda}\right)^{-d_\ell^s}\left(1 - \mathrm{e}^{i\lambda}\right)^{-d_m^s} f_{\ell,m}^S(\lambda), \qquad \lambda \in [-\pi, \pi],$$

where the long-memory parameters are given by $d_m^S = d_m - D_m$ for all m. The functions $f_{\ell,m}^S(\cdot)$ correspond to the short-memory behaviour of the process. The generalized cross-spectral density of processes X_ℓ and X_m can be written as

$$f_{\ell,m}(\lambda) = \frac{1}{2\pi}\Omega_{\ell,m}\left(1 - \mathrm{e}^{-i\lambda}\right)^{-d_\ell}\left(1 - \mathrm{e}^{i\lambda}\right)^{-d_m} f_{\ell,m}^S(\lambda), \qquad \lambda \in [-\pi, \pi].$$

As will be explained in Section 2.1, this model corresponds to a subclass of multivariate long-range dependent time series. In a general case, an additional multiplicative term of the form $\mathrm{e}^{i\varphi}$ is required (e.g. Kechagias and Pipiras, 2015; Sela and Hurvich, 2012).

In the case of a multivariate setting, the spectral density of the multivariate process $\mathbf{X}$ is thus

$$\boldsymbol{f}(\lambda) = \boldsymbol{\Omega} \circ \left(\boldsymbol{\Lambda}^0(\mathbf{d})\boldsymbol{f}^S(\lambda)\boldsymbol{\Lambda}^0(\mathbf{d})^*\right), \quad \lambda \in [-\pi, \pi], \quad \text{with } \boldsymbol{\Lambda}^0(\mathbf{d}) = \mathrm{diag}\left(\left(1 - \mathrm{e}^{-i\lambda}\right)^{-\mathbf{d}}\right) \tag{1}$$

where $\mathbf{d} = \mathbf{D} + \mathbf{d}^s$. The exponent $*$ is the conjugate operator and $\circ$ denotes the Hadamard product. The matrix $\boldsymbol{\Omega}$ is supposed to be real symmetric positive definite.

In this semiparametric framework, the spectral density $\boldsymbol{f}^S(\cdot)$ corresponds to the short-memory behaviour, and the matrix $\boldsymbol{\Omega}$ is called *fractal connectivity* by Achard *et al.* (2008) or *long-run covariance* matrix by Robinson (2005). Similar to Moulines *et al.* (2007), we assume that $\boldsymbol{f}^S(\cdot) \in \mathcal{H}(\beta, L)$ with $0 < \beta \leq 2$ and $0 < L$. The space $\mathcal{H}(\beta, L)$ is defined as the class of non-negative symmetric functions $\mathbf{g}(\cdot)$ on $[\pi, \pi]$ such that $g_{\ell,m}(0) = 1$ for all $(\ell, m) \in \{1, \ldots, p\}^2$ and that for all $\lambda \in (-\pi, \pi)$, $\max\{|\mathbf{g}(\lambda) - 1|, (\ell, m) \in \{1, \ldots, p\}^2\}_\infty \leq L|\lambda|^\beta$. The assumption $\boldsymbol{f}_{\ell,m}^S(0) = 1$ for all (ℓ, m) is necessary for $\boldsymbol{\Omega}$ to be identifiable in model (1).

The spectral density specifies that the two processes X_ℓ and X_m have long-memory parameters respectively d_ℓ and d_m. Parameters with an absolute value greater that $1/2$ are allowed, covering nonstationary time series (in this case $D_\ell, D_m \geq 1$). If orders are different, the estimation of the memory parameters is still available, but some bias issues occur for the estimation of the underlying covariance $\boldsymbol{\Omega}$, which is detailed in Section 3.

In order to derive semiparametric estimations of the memory parameters and the matrix $\boldsymbol{\Omega}$, the term inside the matrix $\boldsymbol{\Lambda}^0(\mathbf{d})$ can be simplified using the equality $1 - \mathrm{e}^{-i\lambda} = 2\sin(\lambda/2)\mathrm{e}^{i(\pi-\lambda)/2}$. Consequently, when λ tends to 0, the spectral density matrix is approximated at first order by

$$\boldsymbol{f}(\lambda) \sim \widetilde{\boldsymbol{\Lambda}}(\mathbf{d})\boldsymbol{\Omega}\widetilde{\boldsymbol{\Lambda}}(\mathbf{d})^*, \quad \text{when } \lambda \to 0, \quad \text{with } \widetilde{\boldsymbol{\Lambda}}(\mathbf{d}) = \mathrm{diag}\left(|\lambda|^{-\mathbf{d}}\mathrm{e}^{-i\,\mathrm{sign}(\lambda)\pi\mathbf{d}/2}\right). \tag{2}$$

Here and subsequently, the symbol $\sim$ means that the ratio of the left-hand and right-hand sides converges to 1.

A similar approximation has been carried out by Lobato (1997) or Phillips and Shimotsu (2004), while Shimotsu (2007) derived a second-order approximation. Lobato (1999) used $\widetilde{\boldsymbol{\Lambda}}(\mathbf{d}) = \mathrm{diag}(|\lambda|^{-\mathbf{d}})$ as an approximation of $\boldsymbol{f}(\cdot)$. Whereas Shimotsu (2007) chose to approximate $\boldsymbol{f}(\cdot)$ using $\widetilde{\boldsymbol{\Lambda}}(\mathbf{d}) = \mathrm{diag}(\lambda^{-\mathbf{d}}\mathrm{e}^{-i(\pi-\lambda)\mathbf{d}/2})$, which corresponds to a second-order approximation due to the remaining term λ in the exponential. As mentioned by

DOI: 10.1111/jtsa.12170

Shimotsu (2007), intriguingly, the two defined estimators of long-memory parameters are consistent, but only for the estimation of $\mathbf{d}$. The estimation of the covariance matrix is affected by the choice of $\widetilde{\mathbf{\Lambda}}(\mathbf{d})$. In Section 3, we introduce our estimators using approximation (2), corresponding to a trade-off between Lobato (1999) and Shimotsu (2007). The resulting estimator for $\mathbf{d}$ is equivalent to the one defined by Lobato (1999). However, a specific correction for the estimation of the covariance matrix overcomes the bias caused by the presence of a phase shift through the complex exponential term. This point has also been raised in the context of detecting cointegration, when the cross-spectral density presents an additional phase parameter compared with the case studied in this article.

2.1. Examples of Processes

This section provides some examples of processes that satisfy our semiparametric modelling.

The matrix $\mathbf{\Omega}$ has been defined *via* the spectral representation of the process; the link between $\mathbf{\Omega}$ and the covariance of the multivariate process in the temporal space is detailed hereafter. Let $\bar{\mathbf{X}} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{X}(t)$ be the empirical mean of the process. If the cross-spectral density is defined and continuous at the frequency $\lambda = 0$, Fejer's theorem states that $n^{1/2}\bar{\mathbf{X}}$ converges in distribution to a zero-mean Gaussian distribution with a covariance matrix equal to $2\pi\boldsymbol{f}(0)$. When the cross-spectral density satisfies an approximation (1), Robinson (2005) indicates that

$$\mathbf{D}_n\mathbb{E}\left(\overline{\mathbf{X}}\overline{\mathbf{X}}^{\mathrm{T}}\right)\mathbf{D}_n \xrightarrow[n\to\infty]{} 2\pi\mathbf{\Omega}\circ\mathbf{Q}(\mathbf{d})$$

where $\mathbf{D}_n = \mathrm{diag}(n^{1/2-\mathbf{d}})$ and $Q_{\ell,m}(\mathbf{d}) = \frac{\sin(\pi d_\ell)+\sin(\pi d_m)}{\Gamma(d_\ell+d_m+2)\sin(\pi(d_\ell+d_m))}$. The exponent 'T' denotes the transpose operator.

2.1.1. *Causal Linear Representations of Kechagias and Pipiras (2014)*

Kechagias and Pipiras (2015) define LRD with a more general setting for the phase. Let $\mathbf{X}$ be a p-multivariate time series. We suppose that $\mathbf{X}$ is second-order stationary and that it admits a spectral density. The time series $\mathbf{X}$ is long-range dependent in the sense of Kechagias and Pipiras (2015) if its spectral density $f\cdot$ satisfies

$$\mathbf{f}(\lambda) = \frac{1}{2\pi}\mathrm{diag}\left(|\lambda|^{-\mathbf{d}}\right)\mathbf{G}(\lambda)\,\mathrm{diag}\left(|\lambda|^{-\mathbf{d}}\right)$$

with $\mathbf{d} = (d_1 \dots d_p) \in (0,1/2)^p$ and $\mathbf{G}(\cdot)$ a $\mathbb{C}^{p\times p}$-valued Hermitian non-negative definite matrix function satisfying

$$\mathbf{G}(\lambda) \sim_{\lambda\to 0^+} \mathbf{G} = \left(\Omega_{\ell,m}e^{i\varphi_{\ell,m}}\right)_{\ell,m=1,\dots,p}$$

with $\Omega_{\ell,m}\in\mathbb{R}$, $\ell,m = 1,\dots,p$, and $\varphi_{\ell,m}\in(-\pi,\pi]$. The phases $\varphi_{\ell,m}$ measure the dissymmetry of the process $\mathbf{X}$ at large lags. Indeed, the specificity of multivariate time series is that the autocovariance function may no longer be symmetric compared with the univariate framework [i.e. $\gamma(-h) = \gamma(h)^{\mathrm{T}}$ may not be equal to $\gamma(h)$]. When the process is time reversible, the autocovariance function is symmetric. Time-reversible series will satisfy $\varphi_{\ell,m} = 0$ for all $\ell,m = 1,\dots,p$. We refer to proposition 2.1 of Kechagias and Pipiras (2015) that gives some highlights on the phase parameters. In Proposition 3.1, Kechagias and Pipiras (2015) give examples of LRD linear time series where any combinations of $(\mathbf{d},\mathbf{G})$ can be chosen.

Many results of the present work can be generalized to LRD processes of Kechagias and Pipiras (2015). Yet as our goal is to recover the matrix $\mathbf{\Omega}$, an assumption on the form of the phase φ is necessary (since we consider a real filter that deletes the imaginary part). The most widely used definition of phase is $\varphi_{\ell,m} = -\frac{\pi}{2}(d_\ell - d_m)$, which includes a large scope of models (Section 2.1.2).

Such a definition of phase is verified, for example, using a causal representation of processes described by Kechagias and Pipiras (2015). Let $\{\boldsymbol{\epsilon}_k\}_{k\in\mathbb{Z}}$ be a $\mathbb{R}^p$-valued white noise, satisfying $\mathbb{E}[\boldsymbol{\epsilon}_k] = 0$ and

$\mathbb{E}[\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^{\mathrm{T}}] = \mathbf{I}_p$. Also, let $\{\mathbf{B}_k = (B_{\ell,m,k})_{\ell,m=1,\dots,p}\}_{k\in\mathbb{N}}$ be a sequence of real-valued matrices such that $B_{\ell,m,k} = L_{\ell,m}(k)\, k^{d_a-1}$, $k \in \mathbb{N}$, where $d_\ell \in (0, 1/2)$ and $\mathbf{L}(k)$, $k = 1, \dots, p$, is an $\mathbb{R}^{p\times p}$-valued function satisfying $\mathbf{L}(k) \sim_{k\to+\infty} \mathbf{A}$ for some $\mathbb{R}^{p\times p}$-valued matrix $\mathbf{A}$. We define the time series $\mathbf{X}$ given by the causal linear representation

$$\mathbf{X}(k) = \sum_{j=0}^{+\infty} \mathbf{B}_j \boldsymbol{\epsilon}_{k-j}.$$

Corollary 4.1 of Kechagias and Pipiras (2015) states that the process $\mathbf{X}$ is long-range dependent with

$$\Omega_{\ell,m} = \frac{\Gamma(d_\ell)\Gamma(d_m)}{2\pi} \left(\mathbf{A}\mathbf{A}^*\right)_{\ell,m},$$
$$\varphi_{\ell,m} = -\frac{\pi}{2}\left(d_\ell - d_m\right).$$

Such causal linear representations thus satisfy (2). An example of such a representation is the multivariate ARFIMA$(0, \mathbf{d}, 0)$ model presented in the next subsection.

2.1.2. Multivariate Autoregressive Fractionally Integrated Moving Average of Lobato (1997)

The composition of linear filters does not commute in the multivariate case. Consequently, there are multiple extensions of univariate ARFIMA to the multivariate framework. We detail in this section the multivariate ARFIMA models of Lobato (1997).

Let $\mathbf{u}$ be a p-dimensional white noise with $\mathbb{E}[\mathbf{u}(t) \mid \mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[\mathbf{u}(t)\mathbf{u}(t)^{\mathrm{T}} \mid \mathcal{F}_{t-1}] = \boldsymbol{\Sigma}$, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\mathbf{u}(s),\ s < t\}$ and $\boldsymbol{\Sigma}$ is a positive definite matrix. The spectral density of $\mathbf{u}$ satisfies $\boldsymbol{f}_u(\lambda) = \boldsymbol{\Sigma}/(2\pi)$.

Let $(\mathbf{A}_k)_{k\in\mathbb{N}}$ be a sequence of $\mathbb{R}^{p\times p}$-valued matrices with $\mathbf{A}_0$ being the identity matrix and $\sum_{k=0}^{\infty} \|\mathbf{A}_k\|^2 < \infty$. Let $\mathbf{A}(\cdot)$ be the discrete Fourier transform of the sequence, $\mathbf{A}(\lambda) = \sum_{k=0}^{\infty} \mathbf{A}_k \mathrm{e}^{ik\lambda}$. We assume that all the roots of $|\mathbf{A}(\mathbb{L})|$ are outside the closed unit circle.

Lobato (1997) defines two multivariate ARFIMA models that both satisfy approximation (2).

Model A. Let $\mathbf{X}$ be defined by $\mathbf{A}(\mathbb{L})\, \mathrm{diag}(\mathbb{1} - \mathbb{L})^{\mathbf{d}}\, \mathbf{X}(t) = \mathbf{B}(\mathbb{L})\mathbf{u}(t)$. The spectral density satisfies

$$f_{\ell,m}(\lambda) \sim_{\lambda\to0^+} \frac{1}{2\pi}\Omega_{\ell,m} e^{-i\pi/2(d_\ell-d_m)} \lambda^{-(d_\ell+d_m)}$$

with $\boldsymbol{\Omega} = \mathbf{A}(1)^{-1}\mathbf{B}(1)\boldsymbol{\Sigma}\mathbf{B}(1)^{\mathrm{T}}\mathbf{A}(1)^{\mathrm{T}^{-1}}$. It is straightforward that Model A satisfies approximation (2).

Model B. Let $\mathbf{X}$ be defined by $\mathrm{diag}(\mathbb{1} - \mathbb{L})^{\mathbf{d}}\, \mathbf{A}(\mathbb{L})\, \mathbf{X}(t) = \mathbf{B}(\mathbb{L})\mathbf{u}(t)$. The spectral density satisfies

$$f_{\ell,m}(\lambda) \sim_{\lambda\to0^+} \frac{1}{2\pi}\sum_{a,b} \beta_{a,b}\alpha_{\ell,a}\alpha_{m,b} e^{-i\pi/2(d_a-d_b)} \lambda^{-(d_a+d_b)}$$

with $\alpha_{\ell,m} = (\mathbf{A}(1)^{-1})_{\ell,m}$ and $\beta_{\ell,m} = (\mathbf{B}(1)\boldsymbol{\Sigma}\mathbf{B}(1)^{\mathrm{T}})_{\ell,m}$. In Model B, the spectral density is equivalent around the zero frequency to the term in $(a, b) = \mathrm{argmax}\{|d_a + d_b|,\ \beta_{a,b}\alpha_{\ell,a}\alpha_{m,b} \neq 0\}$. It gives a more general setting $f_{\ell,m}(\lambda) \sim_{\lambda\to0^+} G_{\ell,m}\lambda^{-d_{\ell,m}}$ with $|d_{\ell,m}| \le (d_\ell + d_m)/2$. [This model is studied more extensively by Sela and Hurvich (2012), where the authors propose an estimator for $d_{1,2}$ in a bivariate framework.] By setting $\Omega_{\ell,m} = 0$ if $|d_{\ell,m}| < (d_\ell + d_m)/2$, equation (2) holds. This means that if there is cointegration, the corresponding long-run covariance value is set to zero.

DOI: 10.1111/jtsa.12170

In particular, Models A and B of Lobato (1997) include FIVAR and VARFI models [where FI stands for fractionally integrated and VAR for vectorial autoregressive] of Sela and Hurvich (2008). These multivariate ARFIMA models admit a causal linear representation. They include short-range dependence behaviour, through the terms $A(\mathbb{L})$ and $B(\mathbb{L})$. When these terms are equal to identity, we obtain an ARFIMA$(0, \mathbf{d}, 0)$ satisfying the causal linear representation given in Section 2.1.1. This (causal) multivariate ARFIMA$(0, \mathbf{d}, 0)$ is a subclass of Kechagias and Pipiras's (2015) (possibly noncausal) definition.

Remark. In the univariate setting, when $A(\cdot)$ and $B(\cdot)$ have no common zeros, Kokoszka and Taqqu (1995) establish that the time series X admits a linear representation $X(k) = \sum_{j=0}^{+\infty} C_j \epsilon_{k-j}$ with $C_j = \frac{B(1)}{A(1)\Gamma(d)} j^{d-1} + O(j^{-1})$ when j goes to infinity. The terms of the linear representation thus satisfy approximately the condition of Kechagias and Pipiras (2015). The extension to multivariate setting would yield information about the link between Lobato's (1997) models and Kechagias and Pipiras's (2015) condition on causal representations. However, it has not been explored in this article since we will not use this fact in any essential way.

3. MULTIVARIATE WAVELET WHITTLE ESTIMATION

This section first defines the wavelet transform of the processes and then gives some results on the cross behaviour of the wavelet coefficients. The main point is the presence of a phase shift caused by the differences in the long-memory parameters. Finally, the proposed estimation scheme is derived, defining simultaneous estimators of the long-memory parameters and of the long-run covariance, which takes into account the phase shift [based on the first-order approximation (2)].

3.1. The Wavelet Analysis

Let $(\phi(\cdot), \psi(\cdot))$ be respectively a father wavelet and a mother wavelet. Their Fourier transforms are given by $\hat{\phi}(\lambda) = \int_{-\infty}^{\infty} \phi(t) \mathrm{e}^{-i\lambda t} \mathrm{d}t$ and $\hat{\psi}(\lambda) = \int_{-\infty}^{\infty} \psi(t) \mathrm{e}^{-i\lambda t} \mathrm{d}t$.

At a given resolution $j \geq 0$, for $k \in \mathbb{Z}$, we define the dilated and translated functions $\phi_{j,k}(\cdot) = 2^{-j/2}\phi(2^{-j} \cdot -k)$ and $\psi_{j,k}(\cdot) = 2^{-j/2}\psi(2^{-j} \cdot -k)$. Throughout the article, we adopt the same convention as in Moulines *et al.* (2007, 2008); that is, large values of the scale index j correspond to coarse scales (low frequencies).

Let $\widetilde{\mathbf{X}}(t) = \sum_{k\in\mathbb{Z}} \mathbf{X}(k)\phi(t-k)$. The wavelet coefficients of the process $\mathbf{X}$ are defined by

$$\mathbf{W}_{j,k} = \int_{\mathbb{R}} \widetilde{\mathbf{X}}(t)\psi_{j,k}(t)\mathrm{d}t \quad j \geq 0, k \in \mathbb{Z}.$$

For given $j \geq 0$ and $k \in \mathbb{Z}$, $\mathbf{W}_{j,k}$ is a p-dimensional vector $\mathbf{W}_{jk} = \left(W_{j,k}(1)\ W_{j,k}(2)\ \ldots\ W_{j,k}(p)\right)$, where $W_{j,k}(\ell) = \int_{\mathbb{R}} \widetilde{X_\ell}(t)\psi_{j,k}(t)\mathrm{d}t$.

The regularity conditions on the wavelet transform are expressed in the following assumptions. They will be needed throughout the article.

- (W1) The functions $\phi(\cdot)$ and $\psi(\cdot)$ are integrable and have compact supports, $\int_{\mathbb{R}} \phi(t)\mathrm{d}t = 1$ and $\int \psi^2(t)\mathrm{d}t = 1$.
- (W2) There exists $\alpha > 1$ such that $\sup_{\lambda\in\mathbb{R}} |\hat{\psi}(\lambda)|(1 + |\lambda|)^\alpha < \infty$, that is, the wavelet is α-regular.
- (W3) The mother wavelet $\psi(\cdot)$ has $M > 1$ vanishing moments.
- (W4) The function $\sum_{k\in\mathbb{Z}} k^\ell \phi(\cdot - k)$ is polynomial with degree ℓ for all $\ell = 1, \ldots, M-1$.
- (W5) For all $i = 1, \ldots, p$, $(1+\beta)/2 - \alpha < d_i \leq M$.

These conditions are not restrictive, and many standard wavelet basis satisfy them. Among them, Daubechies wavelets are compactly supported wavelets parameterized by the number of vanishing moments M. They are α-regular with α an increasing function of M going to infinity (Daubechies, 1992). Assumptions (W1)–(W5) will hold for Daubechies wavelet basis with sufficiently large M.

Remark. The couple of functions $(\phi(\cdot), \psi(\cdot))$ can be associated with a multiresolution analysis, but this condition is not necessary. Similarly, the orthogonality of the family $\{\psi_{j,k}(\cdot)\}$ is not required. See Section 3 of Moulines *et al.* (2007).

DOI: 10.1111/jtsa.12170

Under assumption (W3), the wavelet transform performs an implicit differentiation of order M. Thus, it is possible to apply it on nonstationary processes. In Fourier analysis, tapering procedures are necessary to consider directly nonstationary frameworks (e.g. Velasco and Robinson, 2000, and references therein). Some recent works propose a procedure that differentiates the data before tapering (Hurvich and Chen, 2000, and references therein). Another extension of Fourier to nonstationary frameworks has been proposed by Abadir *et al.* (2007) and applied by Nielsen (2011) in a multivariate analysis.

In practice, a finite number of realizations of the process $\mathbf{X}$, say $\mathbf{X}(1), \dots \mathbf{X}(N)$, are observed. Since the wavelets have a compact support, only a finite number n_j of coefficients are non-null at each scale j. Suppose without loss of generality that the support of $\psi(\cdot)$ is included in $[0, T_\psi]$ with $T_\psi \geq 1$. For every $j \geq 0$, define

$$n_j := \max\left(0, 2^{-j}(N - T_\psi + 1)\right). \tag{3}$$

Then for every $k < 0$ and $k > n_j$, the coefficients $\mathbf{W}_{j,k}$ are set to zero because all the observations are not available. In the following, $n = \sum_{j=j_0}^{j_1} n_j$ denotes the total number of non-zero coefficients used for estimation.

3.2. Spectral Approximation of Wavelet Coefficients

Let us first recall some results of Moulines *et al.* (2007) for the wavelet transform of a univariate process. Let $W_{j,k}$ denote the wavelet coefficient of a unidimensional process X, with spectral density $f(\lambda) = |1 - \mathrm{e}^{i\lambda}|^{-2d_0} f^S(\lambda)$, where $d_0 \in \mathbb{R}$ (note that d_0 can be outside of the interval $[-1/2, 1/2]$). Moulines *et al.* (2007) state that under assumptions (W1)–(W5), the wavelet coefficient process $(W_{j,k})_{k\in\mathbb{Z}}$ is covariance stationary for any given $j \geq 0$. However, they also stress that the between-scale coefficients are not decorrelated. It is shown that the wavelet coefficients are decorrelated when the wavelet basis is orthonormal and $d_0 = 0$, but it is not valid in general settings. Many propositions of estimators of long memory can be found in Wornell and Oppenheim (1992), Abry and Veitch (1998), Jensen (1999) and Gonzaga and Hauser (2011), among others. These works assume that the wavelet coefficients are decorrelated. We follow Bardet *et al.* (2000) or Moulines *et al.* (2008) in taking into account the within-scale and between-scale behaviour.

Let $j \geq 0$ and $j' = j - u \leq j$ be two given scales. Following Moulines *et al.* (2007), the between-scale process is defined as the sequence $\{W_{j,k}, W_{j-u,2^u k+\tau}, \tau = 0, \dots 2^u - 1\}_{k\in\mathbb{Z}}$. Let $\mathbf{D}_{j,u}(\cdot; d_0)$ be the cross-spectral density between $\{W_{j,k}\}_{k\in\mathbb{Z}}$ and $\{W_{j-u,2^u k+\tau}, \tau = 0, \dots 2^u - 1\}_{k\in\mathbb{Z}}$. For any $\lambda \in (-\pi, \pi)$, $\mathbf{D}_{j,u}(\lambda; d_0)$ is a 2^u-dimensional vector. Theorem 1 in Moulines *et al.* (2007) establishes that, under assumptions (W1)–(W5), there exists a positive constant C such that for all $\lambda \in (-\pi, \pi)$,

$$\left|\mathbf{D}_{j,u}(\lambda; d_0) - 2^{2jd_0}\mathbf{D}_{\infty,u}(\lambda; d_0)\right| \leq C 2^{j(2d_0-\beta)},$$

where

$$\mathbf{D}_{\infty,u}(\lambda; d_0) := \sum_{t\in\mathbb{Z}} |\lambda + 2t\pi|^{-2d_0} \hat{\psi}^*(\lambda + 2t\pi) 2^{-u/2} \hat{\psi}\left(2^{-u}(\lambda + 2t\pi)\right) \mathbf{e}_u(\lambda + 2t\pi)$$

with $\mathbf{e}_u(\xi) = \begin{pmatrix} 1 & \mathrm{e}^{-i2^{-u}\xi} & \dots & \mathrm{e}^{-i2^{-u}(2^u-1)\xi} \end{pmatrix}^{\mathrm{T}}$.

The key point of our estimation is the extension of results obtained by Moulines *et al.* (2007) to the multivariate framework. Because of the complexity of the multivariate setting, we choose not to characterize the behaviour of the wavelet coefficients in terms of cross-spectral densities.

First, in order to extend the results of Moulines *et al.* (2007) to a multivariate framework, the covariance behaviour of $\mathbf{W}_{j,k}$ for given (j, k) is derived. Let $\theta_{\ell,m}(j)$ denote the wavelet covariance at scale j between processes X_ℓ and X_m, $\theta_{\ell,m}(j) = Cov(W_{j,k}(\ell), W_{j,k}(m))$ for any position k. Using the spectral density

DOI: 10.1111/jtsa.12170

representation, $\theta_{\ell,m}(j)$ satisfies

$$\theta_{\ell,m}(j) = \int_{-\pi}^{\pi} (1 - e^{-i\lambda})^{-d_\ell} (1 - e^{i\lambda})^{-d_m} \Omega_{\ell,m} f^S_{\ell,m}(\lambda) |\mathbb{H}_j(\lambda)|^2 \, d\lambda,$$

where $\mathbb{H}_j$ is the gain function of the wavelet filter.

The following proposition establishes a second-order approximation of the spectral density in a neighbourhood of zero, such as the one derived by Shimotsu (2007).

Proposition 1. Let assumptions (W1)–(W5) hold. Let $j \geq 0$, $\ell, m = 1, \ldots, p$. Let K_j be defined by

$$K_j(d_\ell, d_m) = \int_{-\infty}^{\infty} |\lambda|^{-(d_\ell + d_m)} \cos\left(2^{-j}\lambda\,(d_\ell - d_m)/2\right) |\hat{\psi}(\lambda)|^2 \, d\lambda.$$

Then there exists a constant C_0 depending on $\beta, min_i\, d_i, \max_i\, d_i, \mathbf{\Omega}, \phi(\cdot)$ and $\psi(\cdot)$ such that

$$|\theta_{\ell,m}(j) - \Omega_{\ell,m} 2^{j(d_\ell + d_m)} \cos(\pi\,(d_\ell - d_m)/2)\, K_j\,(d_\ell + d_m)\,| \leq C_0 L 2^{(d_\ell + d_m - \beta)j}. \tag{4}$$

Note that the second-order approximation of the spectral density depends on j also through the function K_j. The following proposition is deriving a first-order approximation so that its logarithm is linear in j.

Proposition 2. Let $K(\cdot)$ be defined by

$$K(\delta) = \int_{-\infty}^{\infty} |\lambda|^{-\delta} |\hat{\psi}(\lambda)|^2 \, d\lambda, \quad \delta \in (-\alpha, M).$$

Under assumptions (W1)–(W5), there exists a constant C depending on $\beta, min_i\, d_i, \max_i\, d_i, \mathbf{\Omega}, \phi(\cdot)$ and $\psi(\cdot)$ such that, for all $j \geq 0$, for all $\ell, m = 1, \ldots, p$,

$$|\theta_{\ell,m}(j) - \Omega_{\ell,m} 2^{j(d_\ell + d_m)} \cos(\pi\,(d_\ell - d_m)/2)\, K\,(d_\ell + d_m)\,| \leq C L 2^{(d_\ell + d_m - \beta)j}. \tag{5}$$

Observe that the approximation (5) shows that the difficulty with the case of multivariate long-memory processes is the appearance of a phase shift that has to be taken into account for the estimation of the covariance $\mathbf{\Omega}$. Indeed, $\theta_{\ell,m}(j)$ is proved to be close to a term proportional to $\cos(\pi(d_\ell - d_m)/2)$. Then, if $d_\ell \in [-1/2, 1/2]$ and $d_m = 2k + 1 + d_\ell^S$ with $k \in \mathbb{N}$, Proposition 2 implies that for all j, $\theta_{\ell,m}(j)$ is negligible, meaning that the covariance of the wavelet coefficients is close to zero. Consequently, using the covariance of the wavelet coefficients does not allow us to estimate the matrix $\mathbf{\Omega}$ accurately. This example corresponds to a covariance-stationary process X_ℓ and a process X_m such that ΔX_m is covariance stationary, both with the same long-memory parameter d_ℓ. We show in what follows that the consistency of the long-memory parameters is not affected by bias in the estimation of $\mathbf{\Omega}$.

The covariance behaviour for the between-scale process is derived in the following proposition.

Proposition 3. For all $j \geq 0$, $u \geq 0$ and $\lambda \in (-\pi, \pi)$, we define

$$D^{(j)}_{u;\tau}(\lambda; f_{\ell,m}(\cdot)) = \sum_{t \in \mathbb{Z}} f_{\ell,m}\left(2^{-j}(\lambda + 2t\pi)\right) 2^{-j} \mathbb{H}_j\left(2^{-j}(\lambda + 2t\pi)\right) \mathbb{H}^*_{j-u}\left(2^{-j}(\lambda + 2t\pi)\right) e^{-i2^{-u}\tau(\lambda + 2t\pi)}$$
$$\widetilde{D}_{u,\tau}(\lambda; \delta) = \sum_{t \in \mathbb{Z}} |\lambda + 2t\pi|^{-\delta} \hat{\psi}^*(\lambda + 2t\pi) 2^{-u/2} \hat{\psi}\,(2^{-u}(\lambda + 2t\pi))\, e^{-i2^{-u}\tau(\lambda + 2t\pi)}$$

DOI: 10.1111/jtsa.12170

and $K_{u,\tau}(v;\delta) = \int_{-\pi}^{\pi} \widetilde{D}_{u,\tau}(\lambda;\delta)\mathrm{e}^{i\lambda v}\,\mathrm{d}\lambda$.

Then for all $j \geq 0$, for all $u, v \geq 0$, $\tau = 0, \dots, 2^u - 1$,

$$Cov\,(W_{j,k}(\ell), W_{j-u,2^{-u}k'+\tau}(m)) = \int_{-\pi}^{\pi} D_{u;\tau}^{(j)}(\lambda;(\ell,m))\mathrm{e}^{i\lambda(k-k')}\mathrm{d}\lambda.$$

Under assumptions (W1)–(W5), there exists a constant C depending on $\beta, min_i\, d_i$, $\max_i\, d_i$, $\mathbf{\Omega}$,$\phi(\cdot) and d\psi(\cdot)$ such that, for all $j \geq 0$, for all $u, v \geq 0$, $\tau = 0, \dots, 2^u - 1$, and for all $\lambda \in (-\pi, \pi)$,

$$\left| D_{u;\tau}^{(j)}(\lambda; f_{\ell,m}(\cdot)) - \Omega_{\ell,m} 2^{j(d_\ell + d_m)} \cos\left(\pi\left(d_\ell - d_m\right)/2\right) \widetilde{D}_{u,\tau}\left(\lambda; d_\ell + d_m\right)\right| \leq CL2^{(d_\ell + d_m - \beta)j}$$

and

$$\left|Cov[W_{j,k}(\ell) W_{j-u,2^{-u}k'+\tau}(m)] - \Omega_{\ell,m} 2^{j(d_\ell+d_m)} \cos(\pi(d_\ell - d_m)/2) K_{u,\tau}\left(k-k'; d_\ell + d_m\right)\right| \leq CL2^{(d_\ell+d_m-\beta)j}.$$

When $u = 0$ and $2^u k' + \tau = k$, the quantity $K_{0,0}(0; d_\ell + d_m)$ is equal to $\int_{-\infty}^{\infty} |\lambda| - (d_\ell + d_m) |\hat{\psi}(\lambda)|^2 \mathrm{d}\lambda$. Let us remark that $K_{0,0}(0;\cdot)$ is equal to the function $K(\cdot)$ defined in Proposition 2.

3.3. Wavelet Whittle Estimation

Let $j_1 \geq j_0 \geq 1$ be respectively the maximal and minimal resolution levels that are used in the estimation procedure. The estimation is based on the vectors of wavelets coefficients $\{\mathbf{W}_{j,k},\ j_0 \leq j \leq j_1,\ k \in \mathbb{Z}\}$.

The wavelet Whittle approximation of the negative log-likelihood is denoted by $\mathcal{L}(\cdot)$. The criterion corresponds to the negative log-likelihood of Gaussian vectors $(W_{j,k}(\ell))_{j,k,\ell}$. Hannan (1973) and Fox and Taqqu (1986) prove that the Whittle approximation is giving satisfactory results for non-Gaussian processes. In our framework, the wavelet Whittle criterion is defined as

$$\mathcal{L}(\mathbf{G}(\mathbf{d}), \mathbf{d}) = \frac{1}{n} \sum_{j=j_0}^{j_1} \left[n_j \log\det\left(\mathbf{\Lambda}_j(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j(\mathbf{d})\right) + \sum_{k=0}^{n_j} \mathbf{W}_{j,k}^{\mathrm{T}} \left(\mathbf{\Lambda}_j(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j(\mathbf{d})\right)^{-1} \mathbf{W}_{j,k} \right], \quad (6)$$

where $\mathbf{\Lambda}_j(\mathbf{d})$ and the matrix $\mathbf{G}(\mathbf{d})$ are obtained with Proposition 2,

$$\mathbf{\Lambda}_j(\mathbf{d}) = \mathrm{diag}\left(2^{j\mathbf{d}}\right)$$

and the (ℓ, m)th element of the matrix $\mathbf{G}(\mathbf{d})$ is $G_{\ell,m}(\mathbf{d}) = \Omega_{\ell,m} K(d_\ell + d_m)\cos(\pi(d_\ell - d_m)/2)$.

For each $j \geq 0$, the quantity $\sum_k \mathbf{W}_{j,k}^{\mathrm{T}} \left(\mathbf{\Lambda}_j(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j(\mathbf{d})\right)^{-1} \mathbf{W}_{j,k}$ has a dimension equal to 1 and is equal to its trace. Thus,

$$\mathcal{L}(\mathbf{G}(\mathbf{d}), \mathbf{d}) = \frac{1}{n} \sum_{j=j_0}^{j_1} \left[n_j \log\det\left(\mathbf{\Lambda}_j(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j(\mathbf{d})\right) + trace\left(\left(\mathbf{\Lambda}_j(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j(\mathbf{d})\right)^{-1} \mathbf{I}(j) \right) \right], \quad (7)$$

where $\mathbf{I}(j) = \sum_{k=0}^{n_j} \mathbf{W}_{j,k}\mathbf{W}_{j,k}^{\mathrm{T}}$. Note that this expression is very similar to the multivariate Fourier Whittle (MFW) estimator of Shimotsu (2007). Here, we replace the periodogram by the wavelet scalogram $\mathbf{I}(j)$.

Remark. In the Fourier analysis (e.g. Shimotsu, 2007), the periodogram is normalized. In the wavelet analysis, the normalization factor may depend on the resolution j, and the scalogram is not normalized. For every j, the scalogram $\mathbf{I}(j)$ should be normalized by n_j. In the remainder of the article, we will keep the initial $\mathbf{I}(j)$ for convenience.

DOI: 10.1111/jtsa.12170

Deriving expression (7) with respect to the matrix $\mathbf{G}$ yields

$$\frac{\partial \mathcal{L}}{\partial \mathbf{G}}(\mathbf{G}, \mathbf{d}) = \frac{1}{n} \sum_{j=j_0}^{j_1} \left[n_j \mathbf{G}^{-1} - \mathbf{G}^{-1} \boldsymbol{\Lambda}_j(\mathbf{d})^{-1} \mathbf{I}(j) \boldsymbol{\Lambda}_j(\mathbf{d})^{-1} \mathbf{G}^{-1} \right].$$

Hence, the minimum for fixed $\mathbf{d}$ is attained at

$$\hat{\mathbf{G}}(\mathbf{d}) = \frac{1}{n} \sum_{j=j_0}^{j_1} \boldsymbol{\Lambda}_j(\mathbf{d})^{-1} \mathbf{I}(j) \boldsymbol{\Lambda}_j(\mathbf{d})^{-1}. \tag{8}$$

Replacing $\mathbf{G}(\mathbf{d})$ by $\hat{\mathbf{G}}(\mathbf{d})$, the objective criterion is defined as

$$\begin{aligned} R(\mathbf{d}) &:= \mathcal{L}\left(\hat{\mathbf{G}}(\mathbf{d}), \mathbf{d}\right) - 1, \\ &= \log\det\left(\hat{\mathbf{G}}(\mathbf{d})\right) + \frac{1}{n} \sum_{j=j_0}^{j_1} n_j \log\left(\det\left(\boldsymbol{\Lambda}_j(\mathbf{d}) \boldsymbol{\Lambda}_j(\mathbf{d})\right)\right), \\ &= \log\det\left(\hat{\mathbf{G}}(\mathbf{d})\right) + 2\log(2) \left(\frac{1}{n} \sum_{j=j_0}^{j_1} j n_j\right) \left(\sum_{\ell=1}^{p} d_\ell\right). \end{aligned} \tag{9}$$

The vector of the long-memory parameter $\mathbf{d}$ is estimated by $\hat{\mathbf{d}} = argmin_{\mathbf{d}}\, R(\mathbf{d})$. The estimator $\hat{\mathbf{d}}$ is exactly equal to the one introduced by Moulines *et al.* (2008) when the matrix $\boldsymbol{\Omega}$ is diagonal corresponding to a univariate setting.

In a second step of estimation, we define $\hat{\mathbf{G}}(\hat{\mathbf{d}})$, estimator of $\mathbf{G}(\mathbf{d})$. Finally, applying the correction of phase shift yields the estimation of the covariance matrix $\boldsymbol{\Omega}$

$$\hat{\Omega}_{\ell,m} = \hat{G}_{\ell,m}\left(\hat{\mathbf{d}}\right) / \left(\cos\left(\pi\left(\hat{d}_\ell - \hat{d}_m\right)/2\right) K\left(\hat{d}_\ell + \hat{d}_m\right)\right). \tag{10}$$

Equation (10) is correctly defined as the probability that $\hat{d}_\ell - \hat{d}_m$ is exactly congruent to 1 *modulo* 2 is null. Consequently, estimator $\hat{\boldsymbol{\Omega}}$ is defined almost surely. Yet, empirically, when $d_\ell - d_m$ is close to 1 *modulo* 2, the estimation of $\boldsymbol{\Omega}$ may be strongly biased.

4. MAIN RESULTS

In the earlier discussion, we have defined the MWW estimator; the following section deals with the asymptotic behaviour of the estimators. The consistency of the estimators is established, under a condition that controls the variance of the empirical wavelet cross-covariances. The first part of this section introduces this condition and characterizes a class of processes for which it is satisfied. The second part details the asymptotic results of convergence.

4.1. Additional Condition

The following condition is an additional assumption, which gives an asymptotic control of the wavelet scalogram.

Condition (C): For all $\ell, m = 1, \ldots, p$, $\sup_n \sup_{j \geq 0} \frac{1}{n_j 2^{2j(d_\ell + d_m)}} \mathrm{Var}\left(I_{\ell,m}(j)\right) < \infty.$

This condition is slightly more restrictive than condition (9) of Moulines *et al.* (2008) in a univariate framework, where their spectral density of the process is only defined on a neighbourhood of zero.

The following proposition gives a class of multivariate processes such that Condition (C) holds.

Proposition 4. Suppose that there exists a sequence $\{\mathbf{A}(u)\}_{u\in\mathbb{Z}}$ in $\mathbb{R}^{p\times p}$ such that $\sum_u \|\mathbf{A}(u)\|_\infty^2 < \infty$ and

$$\forall t, \quad \Delta^D \mathbf{X}(t) = \boldsymbol{\mu} + \sum_{u\in\mathbb{Z}} \mathbf{A}(t+u)\boldsymbol{\epsilon}(t)$$

with $\boldsymbol{\epsilon}(t)$ being the weak white noise process, in $\mathbb{R}^p$. Let $\mathcal{F}_{t-1}$ denote the σ-field of events generated by $\{\boldsymbol{\epsilon}(s),\ s \le t-1\}$. Assume that $\boldsymbol{\epsilon}$ satisfies $\mathbb{E}[\boldsymbol{\epsilon}(t)|\mathcal{F}_{t-1}] = 0$, $\mathbb{E}[\epsilon_a(t)\epsilon_b(t)|\mathcal{F}_{t-1}] = \mathbb{1}_{a=b}$ and $\mathbb{E}[\epsilon_a(t)\epsilon_b(t)\epsilon_c(t)\epsilon_d(t)|\mathcal{F}_{t-1}] = \mu_{a,b,c,d}$ with $|\mu_{a,b,c,d}| \le \mu_\infty < \infty$, for all $a,b,c,d = 1,\ldots,p$.
Then, under assumptions (W1)–(W5), Condition (C) holds.

The proof is given in Appendix B.

This assumption of a Cramer–Wold-type decomposition of the process $\mathbf{X}$ with a linear fourth-order stationary process was made among others by Lobato (1999), Shimotsu (2007), Giraitis *et al.* (1997), or Theorem 1 of Moulines *et al.* (2008). As discussed by Lobato (1999), there exist models with density (1), where Condition (C) is not satisfied; however, it is not particularly restrictive.

4.2. Convergence

We suppose that we have N observations of a multivariate p-vector process $\mathbf{X}$, namely $\mathbf{X}(1),\ldots\mathbf{X}(N)$ with a spectral density satisfying approximation (2) around the zero frequency. For given functions $(\phi(\cdot),\psi(\cdot))$ and for given levels $0 \le j_0 \le j_1$, the estimator of $\mathbf{d}$ is the argument minimizing $R(\cdot)$ defined by (9), and the matrix $\mathbf{G}$ is estimated by $\hat{\mathbf{G}}(\hat{\mathbf{d}})$ defined by (8). From now on, we will add the superscript 0 to denote the true parameter values, $\mathbf{d}^0$ and $\mathbf{G}^0$.

The following assumptions on the resolution levels j_0 and j_1 will be needed throughout the article. We assume that either the difference $j_1 - j_0$ is constant or it tends to infinity as N tends to infinity.

The following results show the consistency of the estimators and the rate of convergence. The proofs are given in the Appendix.

Theorem 5. Assume that (W1)–(W5) and Condition (C) hold. If j_0 and j_1 are chosen such that $2^{-j_0\beta} + N^{-1/2}2^{j_0/2} \to 0$ and $j_0 < j_1 \le j_N$ with $j_N = \max\{j, n_j \ge 1\}$, then

$$\hat{\mathbf{d}} - \mathbf{d}^0 = o_{\mathbb{P}}(1).$$

This result generalizes that of Moulines *et al.* (2008). Our result deals with multivariate settings, with the same assumption on the wavelet filter and on the choice of the scale j_0. The condition in Theorem 5 is equal to the one obtained in Proposition 9 of Moulines *et al.* (2008) in the univariate case, that is, $1/j_0 + N^{-1/2}2^{j_0/2} \to 0$. We choose here to express the condition with the parameter β since the optimal choice of j_0 depends on β as shown in Corollary 7.

The convergence of $\hat{\mathbf{G}}(\hat{d})$ to $\mathbf{G}^0$ is not established under assumptions of Theorem 5. However, we prove it in the following theorem, under more restrictive conditions.

Theorem 6. Assume that (W1)–(W5) and Condition (C) hold. If j_0 and j_1 are chosen such that $\log(N)^2(2^{-j_0\beta} + N^{-1/2}2^{j_0/2}) \to 0$ and $j_0 < j_1 \le j_N$, then

$$\hat{\mathbf{d}} - \mathbf{d}^0 = O_{\mathbb{P}}\left(2^{-j_0\beta} + N^{-1/2}2^{j_0/2}\right),$$

DOI: 10.1111/jtsa.12170

$$\forall(\ell, m) \in \{1, \ldots, p\}^2,\ \hat{G}_{\ell,m}\left(\hat{\mathbf{d}}\right) - G_{\ell,m}\left(\mathbf{d}^0\right) = O_{\mathbb{P}}\left(\log(N)\left(2^{-j_0\beta} + N^{-1/2}2^{j_0/2}\right)\right),$$
$$\hat{\Omega}_{\ell,m} - \Omega_{\ell,m} = O_{\mathbb{P}}\left(\log(N)\left(2^{-j_0\beta} + N^{-1/2}2^{j_0/2}\right)\right).$$

The condition in Theorem 6 is slightly different from Theorem 3 of Moulines *et al.* (2008). Our result presents an additional $\log(N)$ term due to technical simplifications in the proof. However, it may be suppressed by adding technical details. The same arguments apply for the $\log(N)$ term appearing in the rate of the convergence of the matrix.

The optimal rate is then expressed by balancing the two terms appearing in the earlier bound.

Corollary 7. Assume that (W1)–(W5) and Condition (C) hold. Taking $2^{j_0} = N^{1/(1+2\beta)}$,

$$\hat{\mathbf{d}} - \mathbf{d}^0 = O_{\mathbb{P}}\left(N^{-\beta/(1+2\beta)}\right).$$

This corresponds to the optimal rate (Giraitis *et al.*, 1997). Fourier Whittle estimators of Lobato (1999) and Shimotsu (2007) obtained the rate $m^{1/2}$ where m is the number of discrete frequencies used in the Fourier transform. When $m \sim cN^{\zeta}$ with a positive constant c, the convergence is obtained for $0 < \zeta < 2\beta/(1 + 2\beta)$. Wavelet estimators thus give a slightly better rate of convergence.

The result of Corollary 7 stresses that it is necessary to fix the finest frequency j_0 in the wavelet procedure at a given scale depending on the regularity β of the density $\boldsymbol{f}^S(\cdot)$. A possible extension is to develop an estimation that is adaptive relative to the parameter β. This is performed, for example, in univariate Fourier analysis by Iouditsky *et al.* (2001). However, this topic exceeds the scope of this article.

Further results on asymptotic normality, and in particular the asymptotic variance of the estimators, would give important information to quantify the quality of the estimators. In particular, it would give a theoretical mean of comparison between the Fourier-based and wavelet-based approaches or between the univariate and multivariate estimations of **d**. This work is in progress and will be established in a future article. Here, the comparison is carried out with a simulation study.

5. SIMULATIONS

Simulated data are used to study the behaviour of the proposed procedure using one illustrative example. An extensive simulation study would exceed the scope of this article and will be provided in a future article. Here, we consider an ARFIMA(0, **d**, 0) with a long-run correlation matrix $\mathbf{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\rho = 0.4$. The proposed MWW estimators are computed for $N = 512$ observations and 1000 Monte Carlo replications. An R package named *multiwave* is available and MATLAB codes are available on request.

A set of different values of **d** is considered. The choices are restricted to settings where the two components of the processes share the same order of stationarity. Indeed, it seems natural that time series measuring similar phenomena have similar stationary properties. We simulated both stationary and nonstationary ARFIMA processes, $d_1 = 0.2$ and $d_1 = 1.2$. Our MWW estimator is shown to be consistent, and the quality increases when $|\rho|$ increases. We also conducted a comparison between our estimators and MFW estimators developed by Shimotsu (2007) for stationary processes only. In nonstationary cases, Nielsen (2011) proposed a similar approach based on the extended Fourier transform of Abadir *et al.* (2007). However, in our simulations, this approach gives satisfactory results only for $\mathbf{d} < 1.5$.

A wavelet-based procedure with M vanishing moments should be compared with an estimation based on tapered Fourier of order M. Such a comparison has been driven by Faÿ *et al.* (2009) in one-dimensional settings. The authors established that wavelet-based estimation outperforms tapered Fourier estimation. Similar observations are expected in a multivariate framework. Yet as multivariate Whittle estimation based on tapered Fourier transform has not been studied in literature, we choose not to display such a comparison.

DOI: 10.1111/jtsa.12170

It is worth pointing out that the main advantage of wavelets is their flexibility. Wavelet-based estimators can be applied for a large set of data, whatever the degree of stationarity is (if still smaller than the number of vanishing moments) and even if the processes contain polynomial trends, which is very attractive for real-data applications.

Parameters used for Estimation

The quality of estimation by wavelets relies on the choice of the wavelet bases. A trade-off is necessary between the number of vanishing moments and the support size of the wavelets. In a time series analysis, the number of vanishing moments enables us to consider polynomial trends or nonstationary time series, owing to the constraint $\sup_\ell d_\ell \leq M$. Yet the support size of the wavelet is proportional to the number of vanishing moments and increases the variance of estimation.

The wavelet basis used in this section is the Daubechies wavelet with $M = 4$ vanishing moments. Its regularization parameter is $\alpha = 1.91$. In our framework, when considering stationary time series, we could also apply our procedure using Haar bases. Estimation based on Haar wavelet indeed gives better results, possibly better than Fourier (e.g. Gencay and Signori, 2015, in the case of tests of serial correlation). As explained earlier, a lower number of vanishing moments improves the quality of the wavelet-based estimators (see Faÿ *et al.*, 2009, in the univariate case). Similar results are observed in multivariate estimations. They are not presented here for the sake of concision. As our goal is to propose a flexible method for real-data application, we prefer to consider a higher vanishing moments basis to stress its flexibility.

The method is controlled by the scales j_0 and j_1. The scale j_1 is fixed equal to $\log_2(N)$ while j_0 is chosen so that the optimal mean square error is minimal. Increasing j_0 leads to a smaller bias but a higher variance since less coefficients are used in the estimation process, which may be controlled by an adaptive procedure. As stated by Theorem 5, the finest scales have to be removed from estimation to get rid of the presence of the short-range dependence $\boldsymbol{f}^S(\cdot)$. Similar considerations can be found in Achard *et al.* (2008) and Faÿ *et al.* (2009).

Concerning MFW estimation, the main parameter is the number m of frequencies used in the procedure. An usual choice in literature is $m = N^{0.65}$ (e.g. Shimotsu, 2007; Nielsen and Frederiksen, 2005). Additionally, MFW estimators are evaluated using values of m giving the same number of Fourier coefficients as that of wavelet coefficients. The final m kept is the one giving the optimal mean square error. The parallel between the number of wavelet scales and the number of Fourier frequencies has been discussed by Faÿ *et al.* (2009).

Measures of Quality

The quality of the estimators is measured by the bias, the standard deviation (SD) and the root mean square error (RMSE), which is equal to the square root of ($\text{bias}^2 + \text{SD}^2$). In order to display an easy comparison between the univariate and multivariate approaches, we compute the ratio between the RMSE obtained with the MWW estimation and the RMSE obtained with univariate wavelet Whittle estimations. It is denoted by *ratio M/U*. A similar quantity is defined for the comparison with the MFW estimation. We define *ratio W/S* as the ratio between the RMSE respectively using wavelet-based estimators and Fourier-based estimators.

5.1. Estimation of the Long-memory Parameters d

Results for the estimation of **d** are presented in Table I. The ratio M/U points out that the quality of estimation is increased with the multivariate approach with respect to the univariate procedure. When the series are correlated, it is better to use MWW estimators to infer the long-memory parameters. The estimation is still satisfactory in nonstationary settings.

Table II displays the results of the MFW estimators described by Shimotsu (2007). With the usual number of frequencies $m = N^{0.65}$ in Fourier-based estimation, our wavelet-based procedure leads to lower RMSEs, as

DOI: 10.1111/jtsa.12170

Table I. Multivariate Whittle wavelet estimation of **d** for a bivariate ARFIMA$(0, \mathbf{d}, 0)$ with $\rho = 0.4$ and $N = 512$ with 1000 repetitions

d_1	**d**	Bias	SD	RMSE	Ratio M/U
			$j_0 = 1$		
0.2	0.2	−0.0267	0.0413	0.0492	0.9080
	−0.2	0.0379	0.0430	0.0574	1.0595
	0.2	−0.0298	0.0428	0.0522	0.9631
	0.0	−0.0002	0.0438	0.0438	0.9504
	0.2	−0.0330	0.0456	0.0563	0.9713
	0.2	−0.0333	0.0443	0.0554	0.9831
	0.2	−0.0304	0.0429	0.0526	0.9583
	0.4	−0.0571	0.0461	0.0734	0.9701
			$j_0 = 2$		
1.2	1.2	−0.0380	0.0830	0.0913	0.9728
	0.8	−0.0298	0.0775	0.0831	0.9643
	1.2	−0.0360	0.0818	0.0894	0.9702
	1.0	−0.0346	0.0808	0.0879	0.9626
	1.2	−0.0463	0.0853	0.0970	0.9677
	1.2	−0.0393	0.0850	0.0936	0.9688
	1.2	−0.0369	0.0799	0.0880	0.9589
	1.4	−0.0482	0.0863	0.0989	0.9648

ARFIMA, autoregressive fractionally integrated moving average; RMSE, root mean square error; SD, standard deviation.

Table II. Multivariate Whittle Fourier estimation of **d** for a bivariate ARFIMA$(0, \mathbf{d}, 0)$ with $\rho = 0.4$ and $N = 512$ with 1000 repetitions

d	Bias	SD	RMSE	Ratio W/F
		$m = \lfloor N^{0.65} \rfloor = 57$		
0.2	−0.0087	0.0707	0.0712	0.6908
−0.2	−0.0001	0.0824	0.0824	0.6958
0.2	−0.0037	0.0679	0.0680	0.7674
0.0	−0.0010	0.0778	0.0778	0.5630
0.2	−0.0078	0.0691	0.0695	0.8101
0.2	−0.0043	0.0733	0.0735	0.7546
0.2	−0.0038	0.0705	0.0706	0.7445
0.4	0.0012	0.0788	0.0788	0.9320
		$m = \lfloor N^{0.876} \rfloor = 236$		
0.2	−0.0174	0.0318	0.0362	1.3581
−0.2	0.0158	0.0323	0.0359	1.5964
0.2	−0.0170	0.0315	0.0358	1.4558
0.0	−0.0025	0.0318	0.0319	1.3728
0.2	−0.0200	0.0321	0.0378	1.4875
0.2	−0.0189	0.0320	0.0372	1.4905
0.2	−0.0201	0.0325	0.0382	1.3759
0.4	−0.0317	0.0366	0.0484	1.5169

Two numbers of frequencies m are presented: the usual choice $m = \lfloor N^{0.65} \rfloor$ and the value giving the lower RMSE. $\lfloor x \rfloor$ denotes the closest integer smaller than x.
ARFIMA, autoregressive fractionally integrated moving average; RMSE, root mean square error; SD, standard deviation.

DOI: 10.1111/jtsa.12170

Table III. Wavelet Whittle estimation of $\mathbf{\Omega}$ for a bivariate ARFIMA$(0, \mathbf{d}, 0)$ with $\rho = 0.4$ and $N = 512$ with 1000 repetitions

d		Bias	SD	RMSE
		$j_0 = 1$		
$(0.2, -0.2)$	$\Omega_{1,1}$	0.0342	0.0710	0.0788
	$\Omega_{1,2}$	0.0387	0.0605	0.0718
	$\Omega_{2,2}$	−0.0402	0.0709	0.0815
	Correlation	0.0400	0.0496	0.0637
$(0.2, 0.0)$	$\Omega_{1,1}$	0.0309	0.0697	0.0762
	$\Omega_{1,2}$	0.0176	0.0540	0.0568
	$\Omega_{2,2}$	−0.0012	0.0732	0.0733
	Correlation	0.0113	0.0417	0.0432
$(0.2, 0.2)$	$\Omega_{1,1}$	0.0297	0.0733	0.0790
	$\Omega_{1,2}$	0.0116	0.0518	0.0530
	$\Omega_{2,2}$	0.0282	0.0725	0.0778
	Correlation	−0.0003	0.0386	0.0386
$(0.2, 0.4)$	$\Omega_{1,1}$	0.0356	0.0703	0.0788
	$\Omega_{1,2}$	0.0328	0.0568	0.0655
	$\Omega_{2,2}$	0.0707	0.0728	0.1015
	Correlation	0.0106	0.0422	0.0435
		$j_0 = 2$		
$(1.2, 0.8)$	$\Omega_{1,1}$	0.0037	0.1473	0.1474
	$\Omega_{1,2}$	0.0478	0.1199	0.1290
	$\Omega_{2,2}$	0.0052	0.1303	0.1304
	Correlation	0.0462	0.1041	0.1139
$(1.2, 1.0)$	$\Omega_{1,1}$	−0.0031	0.1411	0.1411
	$\Omega_{1,2}$	0.0182	0.1003	0.1019
	$\Omega_{2,2}$	0.0027	0.1357	0.1357
	Correlation	0.0176	0.0781	0.0800
$(1.2, 1.2)$	$\Omega_{1,1}$	0.0055	0.1442	0.1443
	$\Omega_{1,2}$	0.0060	0.0921	0.0923
	$\Omega_{2,2}$	−0.0033	0.1456	0.1456
	Correlation	0.0052	0.0685	0.0687
$(1.2, 1.4)$	$\Omega_{1,1}$	0.0001	0.1496	0.1496
	$\Omega_{1,2}$	0.0155	0.1039	0.1051
	$\Omega_{2,2}$	0.0135	0.1610	0.1615
	Correlation	0.0125	0.0802	0.0812

ARFIMA, autoregressive fractionally integrated moving average; RMSE, root mean square error; SD, standard deviation.

quantified by the ratio W/F. More precisely, the good performance of our scheme of estimation is due to a lower variance, even if the bias is higher. With a higher number of frequencies in Fourier-based estimation, taking a value that minimizes the RMSE, the MWW estimators are no more preferable to MFW. Yet the ratio W/F stays close to 1, and the analysis of the bias and variances reveals similar orders of magnitude.

5.2. Estimation of the Long-run Covariance Ω

This section deals with the estimation of the long-run covariance matrix $\mathbf{\Omega}$ and the estimation of the correlation $\Omega_{12}/\sqrt{\Omega_{11}\Omega_{22}}$. This latter quantity corresponds in literature to the power law coherency between the two time series (Sela and Hurvich, 2012) or to the fractal connectivity (Achard *et al.*, 2008).

The results obtained in simulations for MWW estimation of the covariance and correlation are given in Table III. The quality is satisfactory in all settings, especially in the stationary ones.

The results for MFW estimation are displayed in Table IV. When MFW is applied with $m = N^{0.65}$ frequencies, the ratio W/F is less than 1. Like for the estimation of **d**, the good performance of MWW estimators is principally due to a smaller variance. When MFW estimators are implemented with a higher number of frequencies, giving optimal results for the estimation of **d**, the difference between MWW and MFW procedures decreases. The quality of the two estimation schemes is similar, with comparable values for bias and variances.

  J. Time. Ser. Anal. **37**: 476–512 (2016)
DOI: 10.1111/jtsa.12170

Table IV. Fourier Whittle estimation of $\mathbf{\Omega}$ for a bivariate ARFIMA$(0, \mathbf{d}, 0)$ with $\rho = 0.4$ and $N = 512$ with 1000 repetitions

		$m = \lfloor N^{0.65} \rfloor = 57$				$m = \lfloor N^{0.876} \rfloor = 236$			
d		Bias	SD	RMSE	Ratio W/F	Bias	SD	RMSE	ratio W/F
$(0.2, -0.2)$	$\Omega_{1,1}$	0.0394	0.2253	0.2287	0.3444	0.0492	0.0679	0.0839	0.9395
	$\Omega_{1,2}$	0.0091	0.1156	0.1160	0.6189	0.0009	0.0498	0.0498	1.4414
	$\Omega_{2,2}$	0.0145	0.2308	0.2313	0.3525	−0.0470	0.0640	0.0794	1.0273
	Correlation	−0.0002	0.0774	0.0774	0.8229	0.0006	0.0387	0.0387	1.6464
$(0.2, 0.0)$	$\Omega_{1,1}$	0.0245	0.2245	0.2259	0.3373	0.0449	0.0666	0.0803	0.9486
	$\Omega_{1,2}$	0.0124	0.1154	0.1161	0.4892	0.0105	0.0506	0.0517	1.0985
	$\Omega_{2,2}$	0.0163	0.2341	0.2347	0.3121	−0.0008	0.0677	0.0677	1.0819
	Correlation	0.0061	0.0793	0.0795	0.5428	0.0014	0.0383	0.0383	1.1259
$(0.2, 0.2)$	$\Omega_{1,1}$	0.0319	0.2319	0.2341	0.3376	0.0450	0.0708	0.0839	0.9417
	$\Omega_{1,2}$	0.0141	0.1191	0.1199	0.4423	0.0176	0.0520	0.0549	0.9666
	$\Omega_{2,2}$	0.0236	0.2331	0.2343	0.3321	0.0438	0.0690	0.0818	0.9517
	Correlation	0.0041	0.0781	0.0782	0.4935	−0.0006	0.0382	0.0382	1.0099
$(0.2, 0.4)$	$\Omega_{1,1}$	0.0264	0.2255	0.2271	0.3470	0.0489	0.0682	0.0839	0.9392
	$\Omega_{1,2}$	0.0107	0.1232	0.1237	0.5298	0.0313	0.0531	0.0616	1.0632
	$\Omega_{2,2}$	0.0276	0.2462	0.2478	0.4096	0.1052	0.0705	0.1267	0.8012
	Correlation	0.0001	0.0783	0.0783	0.5548	0.0002	0.0384	0.0384	1.1307

Two number of frequencies m are presented: the usual choice $m = \lfloor N^{0.65} \rfloor$ and the value giving the lower RMSE.
ARFIMA, autoregressive fractionally integrated moving average; RMSE, root mean square error; SD, standard deviation.

To conclude, the multivariate approach increases the quality of estimation of the long-memory parameters $\mathbf{d}$ in comparison with a univariate estimation. In stationary frameworks, the performance is very similar to MFW estimation, when estimating the vector $\mathbf{d}$ or the long-run covariance matrix. The main advantage of our wavelet-based procedure is then its flexibility. By contrast with Fourier-based estimation, our estimators can be applied in a larger scope of situations, with nonstationary processes or in the presence of polynomial trends in the time series.

6. APPLICATION ON NEUROSCIENCE DATA

We apply our approach to neuroscience data where the recorded data are a typical example of multivariate long-range dependent time series. Researchers are interested in characterizing the brain connectivity. Usually, the connectivity is evaluated using correlations at different frequencies between time series measuring the brain activity. We will show in this section that our method is perfectly adequate to deal with these real data.

The study concerns magnetoencephalography (MEG) data acquired from a healthy 43-year-old woman studied during rest with eyes open at the National Institute of Mental Health Bethesda, MD, using a 274-channel CTF MEG system VSM MedTech, Coquitlam, BC, Canada, operating at 600 Hz. The data were previously used by Achard *et al.* (2008). We consider $N = 2^{15}$ time points for each of the 274 time series.

Figure 1 displays the time series for four arbitrary channels. It is clear that they present nonlinear trends. Consequently, Fourier methods are not adequate to analyse such data, and methods based on wavelets are better to use.

Our procedure was applied using scales 4–8. It corresponds to frequencies between 1 and 20 Hz. This choice was motivated by discussions with neuroscientists. It takes into account the presence of high-frequency noise, which is modelled by $f^S(\cdot)$. The data were preprocessed and the low frequencies were removed. Figure 2 presents the results of the estimation of the long-memory parameters $\mathbf{d}$ and of the long-run covariance matrix $\mathbf{\Omega}$.

First, the histogram of the estimate $\hat{\mathbf{d}}$ shows that the maximal difference between the values of the long-memory parameters is less than 0.5, so the problem of identifiability of $\mathbf{\Omega}$ does not occur with these data. This allows us to give an estimate of the fractal connectivity. It is worth noticing that clusters appear in the correlation matrix. Most of them are situated along the diagonal, corresponding to spatially closed channels. Some channels are still correlated, even far from each other. It would be interesting to relate this result to a neuroscience interpretation. This will be investigated in future work.

DOI: 10.1111/jtsa.12170

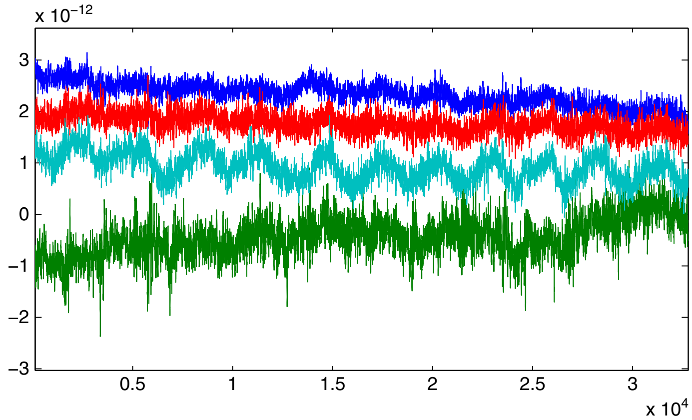
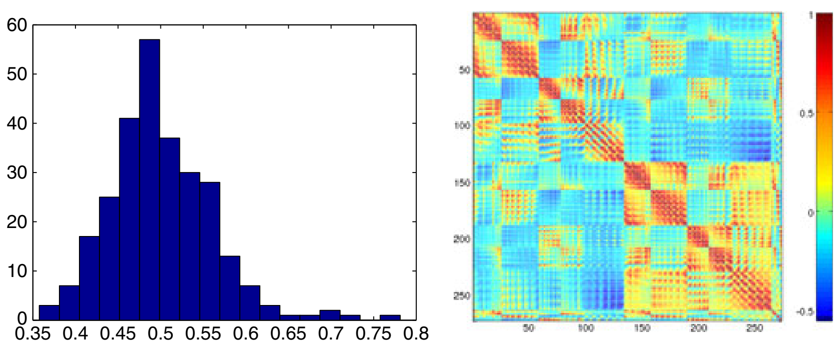


Figure 1. Magnetoencephalography recordings for four arbitrary channels



(a) (b)

Figure 2. Results obtained by multivariate wavelet Whittle estimators on the magnetoencephalogram dataset: histogram of the estimated long-memory parameters **d** (a) and estimated fractal connectivity matrix (b)

7. CONCLUSION

Many application fields are concerned with high-dimensional time series. A challenge is to characterize their long-memory properties and their correlation structure. The present work considers a semiparametric multivariate model, including a large class of multivariate processes such as some fractionally integrated processes. We propose an estimation of the long-dependence parameters and of the fractal connectivity, based on the Whittle approximation and on a wavelet representation of the time series. The theoretical properties of the estimation show the asymptotic optimality. A simulation study confirms that the estimation is well behaved on finite samples. Finally, we propose an application to the estimation of a human brain functional network based on MEG datasets. Our study highlights the benefit of the multivariate analysis, namely improved efficiency of estimation of dependence parameters and estimation of long-term correlations. Future work may concern the asymptotic normality of the estimators, since the development of tests may present a significant benefit for real-data applications.

ACKNOWLEDGEMENTS

The authors thank the co-editor and two anonymous referees for their comments that led to substantial improvements in the article. This work was partly supported by the IXXI Research Institute and the ANR project

GRAPHSIP ANR-14-CE27-0001. This work has been performed during the provisional assignment of I. Gannaz within CNRS at GIPSA-Lab, Grenoble. The authors are grateful to Shimotsu (http://shimotsu.web.fc2.com/Site/Matlab_Codes.html) and to Faÿ, Moulines, Roueff and Taqqu for kindly providing the codes of their respective articles.

REFERENCES

Abadir KM, Distaso W, Giraitis L. 2007. Nonstationarity-extended local Whittle estimation. *Journal of Econometrics* **141**(2): 1353–1384.

Abry P, Veitch D. 1998. Wavelet analysis of long-range-dependent traffic. *Information Theory, IEEE Transactions on* **44**(1): 2–15.

Achard S, Bassett DS, Meyer-Lindenberg A, Bullmore Ed. 2008. Fractal connectivity of long-memory networks. *Physical Review E* **77**(3): 036104.

Bardet JM, Lang G, Moulines E, Soulier Ph. 2000. Wavelet estimator of long-range dependent processes. *Statistical Inference for Stochastic Processes* **3**(1-2): 85–99.

Chambers MJ. 1995. The simulation of random vector time series with given spectrum. *Mathematical and Computer Modelling* **22**(2): 1–6.

Coeurjolly JF, Amblard PO, Achard S. 2013. Wavelet analysis of the multivariate fractional Brownian motion. *ESAIM: Probability and Statistics* **17**: 592–604.

Dahlhaus R. 1989. Efficient parameter estimation for self-similar processes. *The Annals of Statistics* **17**(4): 1749–1766.

Daubechies I. 1992. *Ten lectures on wavelets*, Vol. 61. Philadelphia: SIAM.

Didier G, Pipiras V. 2011. Integral representations and properties of operator fractional Brownian motions. *Bernoulli* **17**(1): 1–33.

Faÿ G, Moulines E, Roueff F, Taqqu MS. 2009. Estimators of long-memory: Fourier versus wavelets. *Journal of Econometrics* **151**(2): 159–177.

Fox R, Taqqu MS. 1986. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics* **14**(2): 517–532.

Frías MP, Alonso FJ, Ruiz-Medina MD, Angulo JM. 2008. Semiparametric estimation of spatial long-range dependence. *Journal of Statistical Planning and Inference* **138**(5): 1479–1495.

Gençay R, Selçuk F, Whitcher BJ. 2001. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. New York: Academic Press.

Gencay R, Signori D. 2015. Multi-scale tests for serial correlation. *Journal of Econometrics* **184**(1): 62–80.

Geweke J, Porter-Hudak S. 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**(4): 221–238.

Giraitis L, Robinson PM, Samarov A. 1997. Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long-range dependence. *Journal of Time Series Analysis* **18**(1): 49–60.

Gonzaga A, Hauser M. 2011. A wavelet Whittle estimator of generalized long-memory stochastic volatility. *Statistical Methods & Applications* **20**(1): 23–48.

Hannan EJ. 1973. The asymptotic theory of linear time-series models. *Journal of Applied Probability* **10**(1): 130–145.

Horn RA, Johnson CR. 1990. *Matrix Analysis*. New York: Cambridge University Press.

Hurvich CM, Chen WW. 2000. An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis* **21**(2): 155–180.

Iouditsky A, Moulines E, Soulier P, *et al.* 2001. Adaptive estimation of the fractional differencing coefficient. *Bernoulli* **7**(5): 699–731.

Jensen MJ. 1999. An approximate wavelet MLE of short and long memory parameters. *Studies in Nonlinear Dynamics and Econometrics* **3**(4): 239–253.

Kechagias S, Pipiras V. 2015. Definitions and representations of multivariate long-range dependent time series. *Journal of Time Series Analysis* **36**(1): 1–25.

Kokoszka PS, Taqqu MS. 1995. Fractional ARIMA with stable innovations. *Stochastic Processes and Their Applications* **60**(1): 19–47.

Lobato IN. 1997. Consistency of the averaged cross-periodogram in long memory series. *Journal of Time Series Analysis* **18**(2): 137–155.

Lobato IN. 1999. A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics* **90**(1): 129–153.

Mandelbrot BB, Van Ness JW. 1968. Fractional Brownian motions, fractional noises and applications. *SIAM Review* **10**(4): 422–437.

Maxim V, Şendur L, Fadili MJ, Suckling J, Gould R, Howard R, Bullmore ET. 2005. Fractional Gaussian noise, functional MRI and Alzheimer's disease. *Neuro Image* **25**: 141–158.

DOI: 10.1111/jtsa.12170

Moulines E, Roueff F, Taqqu MS. 2007. On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter. *Journal of Time Series Analysis* **28** (2): 155–187.

Moulines E, Roueff F, Taqqu MS. 2008. A wavelet Whittle estimator of the memory parameter of a nonstationary gaussian time series. *The Annals of Statistics* **36**(4): 1925–1956.

Nielsen FS. 2011. Local Whittle estimation of multi-variate fractionally integrated processes. *Journal of Time Series Analysis* **32**(3): 317–335.

Nielsen MØ, Frederiksen PH. 2005. Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews* **24**(4): 405–443.

Papanicolaou GC, Sølna K. 2003. *Wavelet based estimation of local Kolmogorov turbulence*. In *Theory and Applications of Long-range Dependence*, Doukhan P, Oppenheim G, Taqqu MS (eds.) Boston: Birkhäuser, pp. 473–505.

Percival DB, Walden AT. 2006. *Wavelet Methods for Time Series Analysis*, Vol. 4. New York: Cambridge University Press.

Phillips PCB, Shimotsu K. 2004. Local Whittle estimation in nonstationary and unit root cases. *The Annals of Statistics* **32**(2): 656–692.

Robinson PM. 1994a. Rates of convergence and optimal spectral bandwidth for long range dependence. *Probability Theory and Related Fields* **99**(3): 443–473.

Robinson PM. 1994b. Semiparametric analysis of long-memory time series. *The Annals of Statistics* **22**: 515–539.

Robinson PM. 1995a. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics* **23** (5): 1630–1661.

Robinson PM. 1995b. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics* **23**(3): 1048–1072.

Robinson PM. 2005. Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction. *Econometric Theory* **21**(01): 171–180.

Sela RJ, Hurvich CM. 2008. Computationally efficient methods for two multivariate fractionally integrated models. *Journal of Time Series Analysis* **30**: 6. DOI: 10.1111/jtsa.12086.

Sela RJ, Hurvich CM. 2012. The averaged periodogram estimator for a power law in coherency. *Journal of Time Series Analysis* **33**(2): 340–363.

Shimotsu K. 2007. Gaussian semiparametric estimation of multivariate fractionally integrated processes. *Journal of Econometrics* **137**(2): 277–310.

Velasco C, Robinson PM. 2000. Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association* **95**(452): 1229–1243.

Wang L, Wang J. 2013. Wavelet estimation of the memory parameter for long range dependent random fields. *Statistical Papers* **55**(4): 1–14.

Whitcher B, Jensen MJ. 2000. Wavelet estimation of a local long memory parameter. *Exploration Geophysics* **31** (1/2): 94–103.

Wornell GW, Oppenheim AV. 1992. Estimation of fractal signals from noisy measurements using wavelets. *Signal Processing, IEEE Transactions on* **40**(3): 611–623.

APPENDIX A: PROOF OF PROPOSITIONS 1, 2 AND 3

This section deals with the proof of Propositions 1 and 2. The proof of Proposition 3 is based on similar arguments and is omitted.

The covariance between $W_{j,k}(\ell)$ and $W_{j,k}(m)$ can be written with the cospectrum, $\theta_{\ell,m}(j) = \int_{\mathbb{R}} Re(f_{\ell,m}(\lambda))|\mathbb{H}_j(\lambda)|^2 \,\mathrm{d}\lambda$. Indeed, as the cross-spectral density is Hermitian, its imaginary part is an odd function,

$$\theta_{\ell,m}(j) = \Omega_{\ell,m} \int_{\mathbb{R}} |2\sin(\lambda/2)|^{-(d_\ell+d_m)} \cos\left((\pi \operatorname{sign}(\lambda) - \lambda)\left(d_\ell - d_m\right)/2\right) f^S_{\ell,m}(\lambda)|\mathbb{H}_j(\lambda)|^2 \,\mathrm{d}\lambda.$$

The sinus function being odd,

$$\theta_{\ell,m}(j) = \Omega_{\ell,m} \cos\left(\pi\left(d_\ell - d_m\right)/2\right) \int_{\mathbb{R}} |2\sin(\lambda/2)|^{-(d_\ell+d_m)} \cos\left(\lambda\left(d_\ell - d_m\right)/2\right) f^S_{\ell,m}(\lambda)|\mathbb{H}_j(\lambda)|^2 \,\mathrm{d}\lambda.$$

  J. Time. Ser. Anal. **37**: 476–512 (2016)
DOI: 10.1111/jtsa.12170

The proof is very similar to Theorem 1 of Moulines *et al.* (2007). Define the quantities $A_{\ell,m}(j)$ and $R_{\ell,m}(j)$,

$$A_{\ell,m}(j) = \Omega_{\ell,m} 2^j \cos(\pi(d_\ell - d_m)/2) \int_{-\pi}^{\pi} |2\sin(\lambda/2)|^{-(d_\ell+d_m)} \cos(\lambda(d_\ell - d_m)/2) \, f^S_{\ell,m}(\lambda)|\hat{\phi}(\lambda)\hat{\psi}(2^j\lambda)|^2 \,\mathrm{d}\lambda$$

$$R_{\ell,m}(j) = \theta_{\ell,m}(j) - A_{\ell,m}(j)$$

Following the proof of Moulines *et al.* (2007), we can rewrite $A_{\ell,m}(j)$,

$$A_{\ell,m}(j) = \Omega_{\ell,m} 2^j \cos(\pi(d_\ell - d_m)/2) \int_{-\pi}^{\pi} g_{\ell,m}(\lambda)|\lambda|^{-(d_\ell+d_m)} \cos(\lambda(d_\ell - d_m)/2) \, f^S_{\ell,m}(\lambda)|\hat{\phi}(\lambda)|^2 |\hat{\psi}\left(2^j\lambda\right)|^2 \,\mathrm{d}\lambda$$

$$\text{with } g_{\ell,m}(\lambda) = \left|\frac{2\sin(\lambda/2)}{\lambda}\right|^{-(d_\ell+d_m)} \text{ for all } \lambda \in (-\pi, \pi).$$

- The assumption $\boldsymbol{f}^S(\cdot) \in \mathcal{H}(\beta, L)$ states that $\left|f^S_{\ell,m}(\lambda) - 1\right| \le L|\lambda|^\beta$ for all $\lambda \in (-\pi, \pi)$.
- Under assumption (W1), the function $|\hat{\phi}(\cdot)|^2$ is infinitely differentiable and bounded on $(-\pi, \pi)$.
- Using a Taylor expansion, the function $\mathbf{g}(\cdot)$ belongs to $\mathcal{H}(2, L_g)$ with $L_g = \sup_{\ell,m=1,\ldots,p} \sup_{\lambda\in(-\pi,\pi)} |g''_{\ell,m}(\lambda)|$ where $\mathbf{g}''(\cdot)$ denotes the second derivative of $\mathbf{g}(\cdot)$.

This implies that there exists a constant $C_{\phi,d}$ depending on $\phi(\cdot)$ and $\mathbf{d}$ such that

$$\left|A_{\ell,m}(j) - \Omega_{\ell,m} 2^j \cos(\pi(d_\ell - d_m)/2) \int_{-\pi}^{\pi} |\lambda|^{-(d_\ell+d_m)} \cos(\lambda(d_\ell - d_m)/2) \left|\hat{\psi}\left(2^j\lambda\right)\right|^2 \mathrm{d}\lambda\right| \le C_{\phi,d} L 2^j \int_{-\pi}^{\pi} |\lambda|^{(\beta-d_\ell-d_m)} |\hat{\psi}\left(2^j\lambda\right)|^2 \mathrm{d}\lambda.$$

With a change of variable,

$$\left|A_{\ell,m}(j) - \Omega_{\ell,m} 2^{j(d_\ell+d_m)} \cos(\pi(d_\ell - d_m)/2) \int_{-2^j\pi}^{2^j\pi} |\lambda|^{-(d_\ell+d_m)} \cos\left(2^{-j}\lambda(d_\ell - d_m)/2\right) \left|\hat{\psi}(\lambda)\right|^2 \mathrm{d}\lambda\right| \le C_{\phi,d} L 2^{j(d_\ell+d_m-\beta)} \int_{-2^j\pi}^{-2^j\pi} |\lambda|^{(\beta-d_\ell-d_m)} \left|\hat{\psi}(\lambda)\right|^2 \mathrm{d}\lambda.$$

Under assumption (W2), there exists a positive constant C_ψ such that $\int_{-\pi}^{\pi} |\lambda|^{(\beta-d_\ell-d_m)}|\hat{\psi}(\lambda)|^2 \mathrm{d}\lambda \le C_\psi \int_{-\infty}^{\infty} |\lambda|^{(\beta+2\alpha)-d_\ell-d_m} \mathrm{d}\lambda$. (W5) states that $(\beta + 2\alpha) - d_\ell - d_m < 1$; the right-hand side of the inequality is bounded by a constant depending on $\psi(\cdot)$, β and $min_{i=1,\ldots,p} d_i$. Using assumptions (W2) and (W5), we also have $\left|\int_{|\lambda|>2^j\pi} |\lambda|^{-(d_\ell+d_m)} \cos(2^{-j}\lambda(d_\ell - d_m)/2)|\hat{\psi}(\lambda)|^2 \mathrm{d}\lambda\right| \le C_\psi \int_{|\lambda|>2^j\pi} |\lambda|^{-(1+\beta)} \mathrm{d}\lambda$. The right-hand side is bounded by a constant depending on ψ and β. We obtain that there exists a constant C_0 depending on α, β, $\phi(\cdot)$, $\psi(\cdot)$, $\min_{i=1,\ldots,p} d_i$ and $\Omega_{\ell,m}$ such that

$$\left|A_{\ell,m}(j) - \Omega_{\ell,m} 2^{j(d_\ell+d_m)} \cos(\pi(d_\ell - d_m)/2) K_j(d_\ell + d_m)\right| \le C_0 L 2^{j(d_\ell+d_m-\beta)},$$

with $K_j(d_\ell, d_m) = \int_{-\infty}^{\infty} |\lambda|^{-(d_\ell+d_m)} \cos(2^{-j}\lambda(d_\ell - d_m)/2)|\hat{\psi}(\lambda)|^2 \mathrm{d}\lambda$.

DOI: 10.1111/jtsa.12170

On the other hand, we can consider a first-order approximation. Let

$$A_{\ell,m}(j) = \Omega_{\ell,m} 2^j \cos\left(\pi\left(d_\ell - d_m\right)/2\right) \int_{-\pi}^{\pi} g_{\ell,m}(\lambda) |\lambda|^{-(d_\ell+d_m)} f^S_{\ell,m}(\lambda) \left|\hat{\phi}(\lambda)\right|^2 \left|\hat{\psi}\left(2^j\lambda\right)\right|^2 \mathrm{d}\lambda$$

$$\text{with now } g_{\ell,m}(\lambda) = \left|\frac{2\sin(\lambda/2)}{\lambda}\right|^{-(d_\ell+d_m)} \cos\left(\lambda\left(d_\ell - d_m\right)/2\right) \text{ for all } \lambda \in (-\pi, \pi).$$

Then a similar approximation is obtained,

$$\left|A_{\ell,m}(j) - \Omega_{\ell,m} 2^{j(d_\ell+d_m)} \cos\left(\pi\left(d_\ell - d_m\right)/2\right) \int_{-2^j\pi}^{2^j\pi} |\lambda|^{-(d_\ell+d_m)} \left|\hat{\psi}(\lambda)\right|^2 \mathrm{d}\lambda\right| \leq C_{\phi,d} L 2^{j(d_\ell+d_m-\beta)} \int_{-2^j\pi}^{2^j\pi} |\lambda|^{(\beta-d_\ell-d_m)} \left|\hat{\psi}(\lambda)\right|^2 \mathrm{d}\lambda.$$

Using assumptions (W2) and (W5),

$$\left|A_{\ell,m}(j) - \Omega_{\ell,m} 2^{j(d_\ell+d_m)} \cos\left(\pi\left(d_\ell - d_m\right)/2\right) K(\delta)\right| \leq C L 2^{j(d_\ell+d_m-\beta)},$$

with $K(\delta) = \int_{-\infty}^{\infty} |\lambda|^{-(\delta)} |\hat{\psi}(\lambda)|^2 \, \mathrm{d}\lambda$.

Finally, $R_{\ell,m}(j)$ is bounded by $R_{\ell,m}(j) \leq C L 2^{(d_\ell+d_m-\beta)j}$. This inequality follows from the approximation of the squared gain function of the wavelet filter given in Proposition 3 of Moulines *et al.* (2007) and from similar arguments to those given for $A_{\ell,m}(j)$. We do not detail the proof here for the sake of concision, and we refer to the proof of Theorem 1 in Moulines *et al.* (2007).

APPENDIX B: PROOF OF PROPOSITION 4

Since the wavelet ψ admits M vanishing moments, at each scale $j \geq 0$ the associated filter $\mathbb{H}_j$ is factorized as $\mathbb{H}_j(\lambda) = (1 - \mathrm{e}^{i\lambda})^M \widetilde{\mathbb{H}}_j(\lambda)$, with $\widetilde{\mathbb{H}}_j$ trigonometric polynomial, $\widetilde{\mathbb{H}}_j(\lambda) = \sum_{t \in \mathbb{Z}} \widetilde{h}_{j,t} \mathrm{e}^{it\lambda}$.

Since $M \geq D$, the wavelet coefficients may be written as

$$W_{j,k}(\ell) = \sum_{t \in \mathbb{Z}} \widetilde{h}_{j,2^jk-t} \left(\Delta^D X_\ell\right)(t) = \sum_{t \in \mathbb{Z}} \mathbf{B}_\ell\left(j, 2^jk - t\right) \boldsymbol{\epsilon}(t),$$

where $\mathbf{B}_\ell(j, 2^jk - t) = \widetilde{h}_{j,2^jk-t}(\Delta^{M-D}\mathbf{A}_\ell)(t)$. For all $\ell = 1, \ldots, p$, the sequence $\{\mathbf{B}_\ell(j,u)\}_{u \in \mathbb{Z}}$ belongs to $\ell^2(\mathbb{Z})$.

We first give a preliminary result on the second-order moment of $W_{j,k}(\ell)$,

$$\mathbb{E}\left[W_{j,k}(\ell)^2\right] = \sum_{t,t' \in \mathbb{Z}} \sum_{a,b=1,\ldots p} B_{\ell,a}\left(j, 2^jk - t\right) B_{\ell,b}\left(j, 2^jk - t'\right) \mathbb{E}\left[\epsilon_a(t)\epsilon_b\left(t'\right)\right].$$

Using the second-order properties of the process $\boldsymbol{\epsilon}$, the variance is equal to

$$\mathbb{E}\left[W_{j,k}(\ell)^2\right] = \sum_{t \in \mathbb{Z}} \sum_{a=1,\ldots p} B_{\ell,a}\left(j, 2^jk - t\right)^2. \tag{B1}$$

Consider now $\mathbb{E}[I_{\ell,m}(j)^2]$,

$$\begin{aligned}\mathbb{E}\left[I_{\ell,m}(j)^2\right] &= \mathbb{E}\left[\left(\sum_k W_{j,k}(\ell)W_{j,k}(m)\right)^2\right] \\ &= \sum_{k,k'}\sum_{t,t',t'',t'''\in\mathbb{Z}}\sum_{a,b,c,d=1,\ldots p} B_{\ell,a}\left(j,2^jk-t\right)B_{m,b}\left(j,2^jk-t'\right)B_{\ell,c}\left(j,2^jk'-t''\right) \\ &\quad B_{m,d}\left(j,2^jk'-t'''\right) \\ &\quad \mathbb{E}\left[\epsilon_a(t)\epsilon_b\left(t'\right)\epsilon_c\left(t''\right)\epsilon_d(t''')\right].\end{aligned}$$

The fourth-order behaviour of $\boldsymbol{\epsilon}$ implies that

$$\begin{aligned}\mathbb{E}\left[I_{\ell,m}(j)^2\right] &= \sum_{k,k'}\sum_{t\in\mathbb{Z}}\sum_{a,b,c,d=1,\ldots p}\mu_{a,b,c,d}B_{\ell,a}\left(j,2^jk-t\right)B_{m,b}\left(j,2^jk+t\right)B_{\ell,c}(j,2^jk'-t)B_{m,d}\left(j,2^jk'-t\right) \\ &+\sum_{k,k'}\sum_{t\neq t'}\sum_{a,b=1,\ldots p}B_{\ell,a}\left(j,2^jk-t\right)B_{m,a}\left(j,2^jk-t\right)B_{\ell,b}\left(j,2^jk'-t'\right)B_{m,b}\left(j,2^jk'-t'\right) \\ &+\sum_{k,k'}\sum_{t\neq t'}\sum_{a,b=1,\ldots p}B_{\ell,a}\left(j,2^jk-t\right)B_{m,b}\left(j,2^jk-t'\right)B_{\ell,a}\left(j,2^jk'-t\right)B_{m,b}\left(j,2^jk'-t'\right) \\ &+\sum_{k,k'}\sum_{t\neq t'}\sum_{a,b=1,\ldots p}B_{\ell,a}(j,2^jk-t)B_{m,b}\left(j,2^jk-t'\right)B_{\ell,b}\left(j,2^jk'-t'\right)B_{m,a}\left(j,2^jk'-t\right).\end{aligned}$$

As $\mathbb{E}[I_{\ell,m}(j)]^2 = \sum_{k,k'}\sum_{t,t'}\sum_{a,b}B_{\ell,a}(j,2^jk-t)B_{m,a}(j,2^jk-t)B_{\ell,b}(j,2^jk'-t')B_{m,b}(j,2^jk'-t')$, the variance of the scalogram satisfies

$$\begin{aligned}&\mathrm{Var}(I_{\ell,m}(j)) \\ &= \sum_{k,k'}\sum_{t\in\mathbb{Z}}\sum_{a,b,c,d=1,\ldots p}\mu_{a,b,c,d}B_{\ell,a}\left(j,2^jk-t\right)B_{m,b}\left(j,2^jk-t\right)B_{\ell,c}\left(j,2^jk'-t\right)B_{m,d}\left(j,2^jk'-t\right) \\ &+\sum_{k,k'}\mathbb{E}\left[W_{j,k}(\ell)W_{j,k'}(\ell)\right]\mathbb{E}\left[W_{j,k}(m)W_{j,k'}(m)\right]+\sum_{k,k'}\mathbb{E}\left[W_{j,k}(\ell)W_{j,k'}(m)\right]\mathbb{E}\left[W_{j,k}(m)W_{j,k'}(\ell)\right] \\ &-\sum_{k,k'}\sum_{t}\sum_{a,b=1,\ldots p}B_{\ell,a}\left(j,2^jk-t\right)B_{m,a}\left(j,2^jk-t\right)B_{\ell,b}\left(j,2^jk'-t\right)B_{m,b}\left(j,2^jk'-t\right) \\ &-\sum_{k,k'}\sum_{t}\sum_{a,b=1,\ldots p}B_{\ell,a}\left(j,2^jk-t)B_{m,b}(j,2^jk-t\right)B_{\ell,a}\left(j,2^jk'-t\right)B_{m,b}\left(j,2^jk'-t\right) \\ &-\sum_{k,k'}\sum_{t}\sum_{a,b=1,\ldots p}B_{\ell,a}\left(j,2^jk-t\right)B_{m,b}\left(j,2^jk-t\right)B_{\ell,b}\left(j,2^jk'-t\right)B_{m,a}\left(j,2^jk'-t\right).\end{aligned}$$

Finally, $\mathrm{Var}(I_{\ell,m}(j)) \leq V_1+V_2+V_3$ with

$$\begin{aligned}V_1 &= \left|\sum_{k,k'}\mathbb{E}\left[W_{j,k}(\ell)W_{j,k'}(\ell)\right]\mathbb{E}\left[W_{j,k}(m)W_{j,k'}(m)\right]\right|, \\ V_2 &= \left|\sum_{k,k'}\mathbb{E}\left[W_{j,k}(\ell)W_{j,k'}(m)\right]\mathbb{E}\left[W_{j,k}(m)W_{j,k'}(\ell)\right]\right|, \\ V_3 &= (1+\mu_\infty)\sum_{k,k'}\sum_{t\in\mathbb{Z}}\sum_{a,b,c,d=1,\ldots p}\left|B_{\ell,a}\left(j,2^jk-t\right)B_{m,b}\left(j,2^jk-t\right)B_{\ell,c}\left(j,2^jk'-t\right)B_{m,d}\left(j,2^jk'-t\right)\right|.\end{aligned}$$

B1. BOUNDS V_1 AND V_2

Proposition 3 states that $Cov(W_{j,k}(\ell), W_{j,k'}(\ell)) = \int_{-\pi}^{\pi} e^{i\lambda(k-k')} D_{0;0}^{(j)}(\lambda; (\ell,\ell))\mathrm{d}\lambda$. The quantity $\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda v} D_{0;0}^{(j)}(\lambda; (\ell,\ell))\mathrm{d}\lambda$ is the v-th Fourier coefficient of the function $D_{0;0}^{(j)}(\cdot; 2d_\ell)$. Consequently, Parseval theorem implies that $\sum_{v\in\mathbb{Z}} \mathbb{E}[W_{j,k}(\ell)W_{j,k+v}(\ell)]\mathbb{E}[W_{j,k}(m)W_{j,k+v}(m)]$ converges to $I_0^{(j)}(2d_\ell, 2d_m) = \int_{-\pi}^{\pi} D_{0;0}^{(j)}(\lambda; (\ell,\ell)) D_{0;0}^{(j)}(\lambda; (m,m))\mathrm{d}\lambda$. The approximation given in Proposition 3 yields $D_{0;0}^{(j)}(\lambda; (\ell,\ell)) \leq \Omega_{\ell,\ell} 2^{j2d_\ell} \widetilde{D}_{0,0}(\lambda; 2d_\ell) + CL\pi 2^{(2d_\ell-\beta)j}$. Then, using Minkowski inequality

$$\frac{1}{n_j 2^{2j(d_\ell+d_m)}} V_2 \leq 2\left(\Omega_{\ell,\ell}^2 \widetilde{I}_0\left(2d_\ell\right) + C^2L^2\pi^2 2^{-2\beta j}\right)^{1/2} \left(\Omega_{m,m}^2 \widetilde{I}_0\left(2d_m\right) + C^2L^2\pi^2 2^{-2\beta j}\right)^{1/2},$$

where $\widetilde{I}_0(\delta) = \int_{-\pi}^{\pi} \widetilde{D}_{0,0}(\lambda; \delta)^2\mathrm{d}\lambda$. It follows that $\frac{1}{n_j 2^{2j(d_\ell+d_m)}} V_1$ is bounded by a constant independent of j and depending only on $\mathbf{d}$, $\mathbf{\Omega}$, β, $\phi(\cdot)$ and $\psi(\cdot)$.

Similar arguments apply to V_2. Therefore, $\frac{1}{n_j 2^{2j(d_\ell+d_m)}} V_2$ is bounded by $\frac{\int_{-\pi}^{\pi} \widetilde{D}_{0;0}(\lambda;(\ell,m))^2\mathrm{d}\lambda}{2^{2j(d_\ell+d_m)}}$. By Proposition 3, $\frac{1}{n_j 2^{2j(d_\ell+d_m)}} V_2$ is bounded by a constant depending only on $\mathbf{d}$, $\mathbf{\Omega}$, β, $\phi(\cdot)$ and $\psi(\cdot)$.

B2. BOUND V_3

The quantity V_3 is equal to

$$(1+\mu_\infty)\sum_k \sum_{\substack{t\in\mathbb{Z} \\ t'\in\{t+2^j(k-k'),k'\in\mathbb{Z}\}}} \sum_{a,b,c,d=1,\dots p} \left|B_{\ell,a}\left(j, 2^jk-t\right) B_{m,b}\left(j, 2^jk-t\right) B_{\ell,c}\left(j, 2^jk-t'\right) B_{m,d}\left(j, 2^jk-t'\right)\right|.$$

Applying Minkowski inequality on V_3,

$$\begin{aligned} V_3 \leq{}& (1+\mu_\infty)\sum_k \sum_{t\in\mathbb{Z}} \sum_{a,b=1,\dots p} \left|B_{\ell,a}\left(j, 2^jk-t\right) B_{m,b}\left(j, 2^jk-t\right)\right| \\ & \left(\sum_{t'\in\{t+2^jk',k'\in\mathbb{Z}\}} \sum_{c,d=1,\dots p} B_{\ell,c}\left(j, 2^jk-t'\right)^2\right)^{1/2} \left(\sum_{t'\in\{t-2^jk',k'\in\mathbb{Z}\}} \sum_{c,d=1,\dots p} B_{m,d}\left(j, 2^jk-t'\right)^2\right)^{1/2}. \end{aligned}$$

Hence,

$$\begin{aligned} V_3 \leq (1+\mu_\infty)\sum_k & \left(\sum_{t\in\mathbb{Z}} \sum_{a,b=1,\dots p} B_{\ell,a}\left(j, 2^jk-t\right)^2\right)^{1/2} \left(\sum_{t\in\mathbb{Z}} \sum_{a,b=1,\dots p} B_{m,b}\left(j, 2^jk-t\right)^2\right)^{1/2} \\ & \left(\sum_{t'\in\mathbb{Z}} \sum_{c,d=1,\dots p} B_{\ell,c}\left(j, 2^jk-t'\right)^2\right)^{1/2} \left(\sum_{t'\in\mathbb{Z}} \sum_{c,d=1,\dots p} B_{m,d}\left(j, 2^jk-t'\right)^2\right)^{1/2}. \end{aligned}$$

Together with (B1), the following inequality is obtained, $V_3 \leq (1+\mu_\infty) n_j p^2 \theta_{\ell,\ell}(j)\theta_{m,m}(j)$. To conclude,

$$\frac{1}{n_j 2^{2j(d_\ell+d_m)}} V_1 \leq (1+\mu_\infty)\, p^2 \frac{\theta_{\ell,\ell}(j)\theta_{m,m}(j)}{2^{2j(d_\ell+d_m)}}.$$

Condition (C) is proved since $\frac{\theta_{\ell,\ell}(j)\theta_{m,m}(j)}{2^{2j(d_\ell+d_m)}}$ tends to a constant independent of j thanks to Proposition 2.

DOI: 10.1111/jtsa.12170

APPENDIX C: PRELIMINARY RESULTS

Let us take ℓ and m in $1, \dots, p$ and define, for any sequence $\mu = \{\mu_j,\ j \geq 0\}$,

$$S_{\ell,m}(\mu) = \sum_{j,k} \mu_j \left(\frac{W_{j,k}(\ell) W_{j,k}(m)}{2^{j(d_\ell^0 + d_m^0)}} - G^0_{\ell,m} \right) = \sum_{j=j_0}^{j_1} \mu_j \left(\frac{I_{\ell,m}(j)}{2^{j(d_\ell^0 + d_m^0)}} - n_j G^0_{\ell,m} \right). \tag{C1}$$

Proposition 8. Assume that the sequences μ belong to the set $\{\{\mu_j\}_{j\geq 0}, |\mu_j| \leq \frac{1}{n_j}\}$. Under condition (C), $\sup_{\{\mu, |\mu_j| \leq \frac{1}{n_j}\}} S_{\ell,m}(\mu)$ is uniformly bounded by $2^{-j_0\beta} + N^{-1/2} 2^{j_1/2}$ up to a multiplicative constant, that is,

$$\sup_{\mu \in \left\{(\mu_j)_{j\geq 0}, |\mu_j| \leq \frac{1}{n_j}\right\}} \{S_{\ell,m}(\mu)\} = O_{\mathbb{P}}\left(2^{-j_0\beta} + N^{-\frac{1}{2}} 2^{\frac{j_1}{2}}\right).$$

Proof
$S_{\ell,m}(\mu)$ is decomposed into two terms $S^{(0)}_{\ell,m}(\mu)$ and $S^{(1)}_{\ell,m}(\mu)$,

$$S^{(0)}_{\ell,m}(\mu) = \sum_{j=j_0}^{j_1} \mu_j \frac{1}{2^{j(d_\ell^0 + d_m^0)}} \sum_k \left(W_{j,k}(\ell) W_{j,k}(m) - \theta_{\ell,m}(j)\right),$$
$$S^{(1)}_{\ell,m}(\mu) = \sum_{j=j_0}^{j_1} n_j \mu_j \left[\frac{\theta_{\ell,m}(j)}{2^{j(d_\ell^0 + d_m^0)}} - G^0_{\ell,m} \right].$$

From Proposition 2,

$$|S^{(0)}_{\ell,m}(\mu)| \leq \sum_{j=j_0}^{j_1} |\mu_j| \left| \sum_k W_{j,k}(\ell) W_{j,k}(m) - n_j \theta_{\ell,m}(j) \right|, \tag{C2}$$

$$\left|S^{(1)}_{\ell,m}(\mu)\right| \leq C \sum_{j=j_0}^{j_1} 2^{-\beta j} n_j |\mu_j|. \tag{C3}$$

Under the assumption $|\mu_j| \leq \frac{1}{n_j}$, we have the inequality $|S^{(1)}_{\ell,m}(\mu)| \leq C \sum_{j=j_0}^{j_1} 2^{-\beta j}$. The right-hand bound is equivalent to $2^{-j_0\beta}$ up to a constant. Condition (C) gives

$$\mathbb{E}\left[\sup_{\left\{\mu, |\mu_j| \leq \frac{1}{n_j}\right\}} \left| S^{(0)}_{\ell,m}(\mu) \right| \right] \leq C' \sum_{j=j_0}^{j_1} n_j^{-1/2},$$

with a positive constant C'. As $n_j = N 2^{-j}(1 + o(1))$, the right-hand side of the inequality is equivalent to $C' N^{-1/2} 2^{j_1/2}$. □

Proposition 9. Let $0 < j_0 \leq j_1 \leq j_N$. Assume that the sequences μ belong to the set

$$\mathcal{S}(q, \gamma, c) = \left\{ \{\mu_j\}_{j\geq 0}, |\mu_j| \leq \frac{c}{n} |j - j_0 + 1|^q\, 2^{(j-j_0)\gamma} \forall j = j_0, \dots j_1 \right\}$$

DOI: 10.1111/jtsa.12170

with $0 \le \gamma < 1$. Under condition (C), $\sup_{\mu\in\mathcal{S}(q,\gamma,c)} S_{\ell,m}(\mu)$ is uniformly bounded by $2^{-j_0\beta} + H(N^{-1/2}2^{j_0/2})$ up to a constant,

$$\sup_{\mu\in\mathcal{S}(q,\gamma,c)} \{S_{\ell,m}(\mu)\} = O_{\mathbb{P}}\left(2^{-j_0\beta} + H_\gamma\left(N^{-\frac{1}{2}}2^{\frac{j_0}{2}}\right)\right)$$

with $H_\gamma(u) = \begin{cases} u & \textit{if } 0 \le \gamma < 1/2, \\ \log(1+u^{-2})^{q+1}\, u & \textit{if } \gamma = 1/2, \\ \log(1+u^{-2})^{q}\, u^{2(1-\gamma)} & \textit{if } 1/2 < \gamma < 1. \end{cases}$

In particular, for any $0 \le \gamma < 1$, under the assumption $2^{-j_0\beta} + N^{-1/2}2^{j_0/2} \to 0$, we have $\sup_{\mu\in\mathcal{S}(q,\gamma,c)}\{S_{\ell,m}(\mu)\} = o_{\mathbb{P}}(1)$

Proof
Under the assumptions of the proposition, one deduces from inequality (C3) that

$$\sup_{\mu\in\mathcal{S}(q,\gamma,c)} \left|S^{(1)}_{\ell,m}(\mu)\right| \le cC\frac{1}{n}\sum_{j=j_0}^{j_1} n_j 2^{(-\beta j+\gamma(j-j_0))}\,(j-j_0+1)^q \le cC2^{-\beta j_0}\sum_{i=0}^{j_1-j_0} 2^{-(1+\beta-\gamma)i}(i+1)^q.$$

The sum on the right-hand side of the inequality tends to 0 because $1+\beta-\gamma > 0$.

Similarly, under the additional Condition (C), inequality (C2) is rewritten as

$$\begin{aligned}\mathbb{E}\left[\sup_{\mu\in\mathcal{S}(q,\gamma,c)}\left|S^{(0)}_{\ell,m}(\mu)\right|\right] &\le cC'\frac{1}{n}\sum_{j=j_0}^{j_1} n_j^{\frac{1}{2}} 2^{\gamma(j-j_0)}\,(j-j_0+1)^q \\ &\le cC'N^{-\frac{1}{2}}2^{\frac{j_0}{2}}\sum_{i=0}^{j_1-j_0} 2^{-(1/2-\gamma)i}(i+1)^q.\end{aligned}$$

We distinguish three cases depending on the values of γ.

- The result is straightforward when $0 \le \gamma < 1/2$.
- If $\gamma = 1/2$, the right-hand side is bounded by $cC'N^{-1/2}2^{j_0/2}(j_1-j_0+1)^{q+1}$. The parameter j_1 satisfies $2^{j_1} \le N$. Consequently, $j_1 - j_0 \le \log_2(N) + \log_2(2^{-j_0}) = \log_2(N2^{-j_0})$, and the result is proved.
- When $1/2 < \gamma < 1$, the right-hand side admits the upper bound $cC''N^{-1/2}2^{j_0/2}(j_1-j_0+1)^q 2^{(\gamma-1/2)(j_1-j_0)}$ with a positive constant C''. Since $2^{j_1} \le N$, it is inferior to $cC''(j_1-j_0+1)^q(N^{-1}2^{j_0})^{(1-\gamma)}$, which concludes the proof. □

APPENDIX D: WEAK CONSISTENCY

We first establish the convergence under the condition $2^{-j_0\beta} + N^{-1/2}2^{j_1/2} \to 0$. This assumption is more restrictive than the condition $2^{-j_0\beta} + N^{-1/2}2^{j_0/2} \to 0$ given in Theorem 5. Both conditions are equivalent when $j_1 - j_0$ is finite but not in a general case.

We then prove Theorem 5 in two steps: first, we establish a lower bound for $\hat{\mathbf{d}}$ and second, we develop a proof similar to the first one that has been given in Section 7 but with a weaker assumption thanks to the previous bound.

DOI: 10.1111/jtsa.12170

D1. Consistency Under Non-optimal Assumptions

We give a first result of consistency, with an assumption on j_0 and j_1 that can be weakened. This result is not necessary to obtain Theorem 5, but the scheme of the proof is similar, and it points out why an additional step is necessary.

Proposition 10. Assume that (W1)–(W5) and Condition (C) hold. If in addition j_0 and j_1 are chosen such that $2^{-j_0\beta} + N^{-1/2}2^{j_1/2} \to 0$, then

$$\hat{\mathbf{d}} - \mathbf{d}^0 = o_{\mathbb{P}}(1),$$

$$\hat{\mathbf{G}}\left(\hat{\mathbf{d}}\right) - \mathbf{G}\left(\mathbf{d}^0\right) = o_{\mathbb{P}}(1).$$

Proof

In order to evaluate the performances of the estimator of the long-memory parameters, the first step consists in proving that the proposed estimator for $\mathbf{d}$ is consistent. The equivalent properties for $\boldsymbol{\Omega}$ will be detailed in a second step. The proof is based on the following inequality,

$$R(\mathbf{d}) - R\left(\mathbf{d}^0\right) \geq L\left(\mathbf{d} - \mathbf{d}^0\right) + \Delta\left(\mathbf{d}, \mathbf{d}^0\right), \tag{D1}$$

where L is a deterministic and convex function of $\mathbf{d}$ and the remaining term Δ tends uniformly to zero in probability.

We first establish inequality (D1). The difference between the criterion evaluated at $\mathbf{d}$ and at the true long-memory parameters is equal to

$$R(\mathbf{d}) - R\left(\mathbf{d}^0\right) = \log\det\left(\hat{\mathbf{G}}(\mathbf{d})\right) - \log\det\left(\hat{\mathbf{G}}(\mathbf{d}^0)\right) + 2\log(2) < \mathcal{J} > \left(\sum_{\ell} d_\ell - d_\ell^0\right)$$

where $< \mathcal{J} >= \frac{1}{n}\sum_{j=j_0}^{j_1} j\, n_j$ and $n = \sum_{j=j_0}^{j_1} n_j$.

The equality can be rewritten as

$$\begin{aligned}
&R(\mathbf{d}) - R\left(\mathbf{d}^0\right)\\
&= \log\det\left(\frac{1}{n}\sum_{j=j_0}^{j_1} \boldsymbol{\Lambda}_{<\mathcal{J}>}\left(\mathbf{d}-\mathbf{d}^0\right)\boldsymbol{\Lambda}_j(\mathbf{d})^{-1} I(j)\boldsymbol{\Lambda}_j(\mathbf{d})^{-1}\boldsymbol{\Lambda}_{<\mathcal{J}>}\left(\mathbf{d}-\mathbf{d}^0\right)\right) - \log\det\left(\hat{\mathbf{G}}\left(\mathbf{d}^0\right)\right)\\
&= \log\det\left(\frac{1}{n}\sum_{j=j_0}^{j_1} \boldsymbol{\Lambda}_{j-<\mathcal{J}>}\left(\mathbf{d}-\mathbf{d}^0\right)^{-1}\boldsymbol{\Lambda}_j\left(\mathbf{d}^0\right)^{-1} I(j)\boldsymbol{\Lambda}_j\left(\mathbf{d}^0\right)^{-1}\boldsymbol{\Lambda}_{j-<\mathcal{J}>}\left(\mathbf{d}-\mathbf{d}^0\right)^{-1}\right)\\
&\quad - \log\det\left(\hat{\mathbf{G}}\left(\mathbf{d}^0\right)\right).
\end{aligned}$$

Define $\lambda_j(\delta) = 2^{-(j-<\mathcal{J}>)\delta}$ for any $j \geq 0$ and $\delta \in \mathbb{R}$.

Let us first recall Oppenheim's inequality (e.g. Horn and Johnson, 1990, p. 480).

Proposition 11 (Oppenheim's inequality)**.** Let $\mathbf{E}$ and $\mathbf{B}$ be two semidefinite positive matrices. Then $\det(\mathbf{E}\circ\mathbf{B}) \geq \det(\mathbf{E})\prod_\ell B_{\ell,\ell}$.

Let $\mathbf{A}$ be the following matrix,

$$\mathbf{A} = \frac{1}{n}\sum_{j=j_0}^{j_1} \mathbf{\Lambda}_{j-<\mathcal{J}>}\left(\mathbf{d}-\mathbf{d}^0\right)^{-1}\mathbf{\Lambda}_j\left(\mathbf{d}^0\right)^{-1} I(j)\mathbf{\Lambda}_j\left(\mathbf{d}^0\right)^{-1}\mathbf{\Lambda}_{j-<\mathcal{J}>}\left(\mathbf{d}-\mathbf{d}^0\right)^{-1}.$$

Oppenheim's inequality will be applied to matrices $\mathbf{B}$ and $\mathbf{E}(\mathbf{d}-\mathbf{d}^0)$ where the (ℓ,m)th element of $\mathbf{B}$ is defined by $B_{\ell,m} = \frac{1}{n}\sum_{j=j_0}^{j_1} n_j\lambda_j(d_\ell - d_\ell^0)\lambda_j(d_m - d_m^0)$ and $\mathbf{E}(\mathbf{d}-\mathbf{d}^0) = \mathbf{A}\circ\widetilde{\mathbf{B}}$ where $\widetilde{B}_{\ell,m} = B_{\ell,m}^{-1}$. The relation $\mathbf{A} = \mathbf{E}(\mathbf{d}-\mathbf{d}^0)\circ\mathbf{B}$ holds. The (ℓ,m)th element of $\mathbf{E}(\mathbf{d}-\mathbf{d}^0)$ is equal to

$$E_{\ell,m}(\mathbf{d}) = \sum_{j=j_0}^{j_1}\mu_{j,\ell,m}\left(\mathbf{d}-\mathbf{d}^0\right)I_{\ell,m}(j)2^{-j\left(d_\ell^0+d_m^0\right)}$$

$$\text{with } \mu_{j,\ell,m}(\boldsymbol{\delta}) = \frac{2^{-j(\delta_\ell+\delta_m)}2^{<\mathcal{J}>(\delta_\ell+\delta_m)}}{\sum_{a=j_0}^{j_1} n_a 2^{-a(\delta_\ell+\delta_m)}2^{<\mathcal{J}>(\delta_\ell+\delta_m)}} = \frac{2^{-j(\delta_\ell+\delta_m)}}{\sum_{a=j_0}^{j_1} n_a 2^{-a(\delta_\ell+\delta_m)}}.$$

- The matrix $\mathbf{E}$ can be expressed as $\mathbf{E} = \sum_{j,k}\widetilde{W}_{j,k}\widetilde{W}_{j,k}$. Consequently, $\mathbf{E}$ is positive semidefinite, being the sum of positive semidefinite matrices.
- The matrix $\mathbf{B}$ satisfies $\mathbf{B} = \sum_{j=j_0}^{j_1}\mathbf{M}_j\mathbf{M}_j$ with $\mathbf{M}_j = \left(\frac{n_j}{n}\right)^{1/2}\mathbf{\Lambda}_{j-<\mathcal{J}>}(\mathbf{d}-\mathbf{d}^0)^{-1}$. Thus, $\mathbf{B}$ is also positive semidefinite.

Oppenheim's inequality implies $\log\det(\mathbf{A}) \geq \log\det(\mathbf{E}(\mathbf{d}-\mathbf{d}^0)) + \sum_\ell \log B_{\ell,\ell}$.
Define $L(\mathbf{d}-\mathbf{d}^0) := \sum_{\ell=1}^p \log B_{\ell,\ell}$. As we have

$$\sum_{j=j_0}^{j_1} n_j\lambda_j(\delta)\lambda_j(\delta) = \sum_j n_j 2^{-2j\delta}2^{2<\mathcal{J}>\delta} = 2^{2<\mathcal{J}>\delta}\sum_{j=j_0}^{j_1} n_j 2^{-2j\delta},$$

the function L satisfies the following equality:

$$L\left(\mathbf{d}-\mathbf{d}^0\right) = \sum_{\ell=1}^p\left[\log\left(2^{2<\mathcal{J}>\left(d_\ell-d_\ell^0\right)}\right) + \log\left(\frac{1}{n}\sum_{j=j_0}^{j_1} n_j 2^{-2j\left(d_\ell-d_\ell^0\right)}\right)\right].$$

It is easily seen that each term of the sum corresponds to the criterion defined in Proposition 6 of Moulines *et al.* (2008).

Inequality (D1) follows with $\Delta(\mathbf{d},\mathbf{d}^0) = \log\det(\mathbf{E}(\mathbf{d}-\mathbf{d}^0)) - \log\det(\hat{\mathbf{G}}(\mathbf{d}^0))$. We will now control the two terms in the right-hand side inequality (D1).

Control of L. $L(\mathbf{d}-\mathbf{d}^0)$ is a multivariate extension of the criterion studied in Proposition 6 of Moulines *et al.* (2008). It is convex, positive and minimal at $\mathbf{d}=\mathbf{d}^0$.

Control of Δ. We shall prove that both $\log\det\mathbf{E}(\mathbf{d}-\mathbf{d}^0)$ and $\log\det(\hat{\mathbf{G}}(\mathbf{d}^0))$ tend uniformly to $\log\det(\mathbf{G}^0)$ for $\mathbf{d}\in\mathbb{R}^p$.

- The $(\ell,m)th$-element of the matrix $\mathbf{E}(\mathbf{d}-\mathbf{d}^0)$ is equal to

$$E_{\ell,m}\left(\mathbf{d}-\mathbf{d}^0\right) = \sum_{j=j_0}^{j_1}\mu_{j,\ell,m}\left(\mathbf{d}-\mathbf{d}^0\right)I_{\ell,m}(j)2^{-j\left(d_\ell^0+d_m^0\right)} \text{ where } \mu_{j,\ell,m}(\boldsymbol{\delta}) = \frac{2^{-j(\delta_\ell+\delta_m)}}{\sum_a n_a 2^{-a(\delta_\ell+\delta_m)}}.$$

DOI: 10.1111/jtsa.12170

As $\sum_{j=j_0}^{j_1} n_j \mu_{j,\ell,m}(\boldsymbol{\delta}) = 1$, the quantity $E_{\ell,m}(\mathbf{d}-\mathbf{d}^0)$ is written as

$$E_{\ell,m}\left(\mathbf{d}-\mathbf{d}^0\right) = G^0_{\ell,m} + \sum_{j,k} \mu_{j,\ell,m}\left(\mathbf{d}-\mathbf{d}^0\right)\left(\frac{W_{j,k}(\ell)W_{j,k}(m)}{2^{j\left(d^0_\ell+d^0_m\right)}} - G^0_{\ell,m}\right)$$

where $G^0_{\ell,m} = \Omega_{\ell,m}K(d_\ell+d_m)\cos(\pi(d^0_\ell-d^0_m)/2)$.

The preceding expression is equal to $E_{\ell,m}(\mathbf{d}-\mathbf{d}^0) = G^0_{\ell,m} + S_{\ell,m}(\mu_{\ell,m}(\mathbf{d}-\mathbf{d}^0))$ with $S_{\ell,m}(\mu)$ defined previously in equation (C1). Since $\sup_{\mathbf{d}} |\mu_j(\mathbf{d}-\mathbf{d}^0)| \le \frac{1}{n_j}$, Proposition 8 states that $E_{\ell,m}(\mathbf{d}-\mathbf{d}^0) \to G^0_{\ell,m}$ uniformly in $\mathbf{d}$ when $2^{-j_0\beta} + N^{-1/2}2^{j_1/2} \to 0$.

- Finally, we shall establish that $\log\det\hat{\mathbf{G}}(\mathbf{d}^0)$ tends to $\log\det(\mathbf{G}^0)$. Recall

$$\hat{G}_{\ell,m}\left(\mathbf{d}^0\right) = G^0_{\ell,m} + S_{\ell,m}(\nu) \text{ where } \nu_j = \frac{1}{n}.$$

The sequence ν belongs to the set $\mathcal{S}(0,0,1)$. Applying Proposition 9, the convergence is proved when $2^{-j_0\beta} + N^{-1/2}2^{j_0/2} \to 0$. □

The consistency has been established in Proposition 10 under the condition $2^{-j_0\beta} + N^{-1/2}2^{j_1/2} \to 0$. The objective is to weaken this condition in order to prove Theorem 5. The scheme of the proof is a generalization of the proof of Proposition 9 of Moulines *et al.* (2008) to multivariate cases.

The only step in the proof of Proposition 10 that needs the assumption $2^{-j_0\beta} + N^{-1/2}2^{j_1/2} \to 0$ is the convergence study of $E_{\ell,m}(\mathbf{d}-\mathbf{d}^0)$ to $G^0_{\ell,m}$. The proof of Theorem 5 consists in proving that $\hat{\mathbf{d}} > \mathbf{d}^0 - 1/2$ in probability in order to obtain a weaker convergence assumption for $E_{\ell,m}(\mathbf{d}-\mathbf{d}^0)$ applying Proposition 9.

D2. Lower Bound of the Estimate

We proceed to show first that there exists $\mathbf{d}^{min}$ such that for all $\ell = 1,\dots,p$, we have $d^0_\ell - 1/2 < d^{min}_\ell < d^0_\ell$ and $\mathbb{P}(\inf_{j_1\ge j_0+2}\hat{d}_\ell \le d^{min}_\ell)$ tends to 0 when N goes to infinity. The proof is recursive.

Step 1.

We introduce $\widetilde{\alpha}$ defined by

$$\widetilde{\alpha}_{j,\ell,m} = \begin{cases} \frac{1}{n}2^{-(j-<\mathcal{J}>)(d_\ell-d^0_\ell+d_m-d^0_m)} & \text{if } j_0 < j \le <\mathcal{J}>, \\ \frac{1}{n}2^{-(j-<\mathcal{J}>)(d^{min}_\ell-d^0_\ell+d^{min}_m-d^0_m)} & \text{if } <\mathcal{J}> < j \le j_1, \\ 0 & \text{otherwise} \end{cases}$$

where the vector $\mathbf{d}^{min}$ is taken such that

$$\forall \ell = 1,\dots,p,\ d^0_\ell - 1/2 < d^{min}_\ell < d^0_\ell \text{ and } \liminf_{n\to\infty} \inf_{\{\mathbf{d},\ d_\ell \le d^{min}_\ell\}} \inf_{j_1=j_0,\dots,j_N} \sum_{j=j_0}^{j_1} n_j\widetilde{\alpha}_{j,\ell,\ell} > 1. \tag{D2}$$

Let ℓ be a given index in $\{1,\dots,p\}$. Following similar arguments as in Moulines *et al.* (2008) when showing their formula (59), one can find d^{min}_ℓ satisfying condition (D2).

The quantity $R(\mathbf{d}) - R(\mathbf{d}^0)$ is equal to

$$R(\mathbf{d}) - R\left(\mathbf{d}^0\right) = \log\det\mathbf{A}\left(\alpha\left(\mathbf{d}-\mathbf{d}^0\right)\right) - \log\det\hat{\mathbf{G}}\left(\mathbf{d}^0\right)$$

$$\text{with } A_{\ell,m}(\alpha) = \sum_{j=j_0}^{j_1} \alpha_{j,\ell,m}I_{\ell,m}(j) \text{ and } \alpha_{j,\ell,m}(\delta) = \frac{1}{n}2^{-(j-<\mathcal{J}>)(\delta_\ell+\delta_m)}.$$

We first want to establish that for all $\mathbf{d}$ in $\{\mathbf{d},\ \forall \ell\ d_\ell \le d_\ell^{min}\}$, we have $R(\mathbf{d}) - R(\mathbf{d}^0) \ge \log\det \mathbf{A}(\widetilde{\alpha}) - \log\det \hat{\mathbf{G}}(\mathbf{d}^0)$. To this aim, we will use a generalization of Corollary 7.7.4 of Horn and Johnson (1990).

Proposition 12. Let $\mathbf{A}$ and $\mathbf{B}$ be two positive semidefinite matrices of $\mathbb{C}^{p\times p}$. Suppose $\mathbf{A}-\mathbf{B}$ is positive semidefinite. Then $\det\mathbf{A} \ge \det\mathbf{B}$.

Proof

We distinguish two cases:

- Suppose $\det\mathbf{A}=0$. Then there exists a unitary vector $x\in\mathbb{R}^p$ such that $x^{\mathrm{T}}\mathbf{A}x=0$. Using the fact that $\mathbf{A}-\mathbf{B}$ is positive semidefinite, we have $x^{\mathrm{T}}(\mathbf{A}-\mathbf{B})x = -x^{\mathrm{T}}\mathbf{B}x \ge 0$ which implies $x^{\mathrm{T}}\mathbf{B}x=0$ since $\mathbf{B}$ is positive semidefinite. Thus, $\det\mathbf{B}=0=\det\mathbf{A}$.
- Suppose $\det\mathbf{A}>0$. If $\det\mathbf{B}=0$, inequality $\det\mathbf{B}\le\det\mathbf{A}$ holds. If $\det\mathbf{B}>0$, since $\mathbf{A}-\mathbf{B}$ is positive semidefinite, we apply Corollary 7.7.4 of Horn and Johnson (1990), which concludes the proof.

□

Moreover, to prove the positive semidefiniteness of the matrices, we will use the Schur product theorem (e.g. Horn and Johnson, 1990, p. 458).

Proposition 13 (Schur product theorem)**.** Let $\mathbf{B}_1$ and $\mathbf{B}_2$ be two positive semidefinite matrices of $\mathbb{C}^{p\times p}$, then $\mathbf{B}_1\circ\mathbf{B}_2$ is also positive semidefinite.

Let $j\ge 0$. The matrix $\mathbf{B}_1(j) = (2^{j(d_\ell^0+d_m^0)} I_{\ell,m}(j))_{\ell,m}$ is positive semidefinite since it can be written as $\widetilde{W}(j)^{\mathrm{T}}\widetilde{W}(j)$. The matrix $\mathbf{B}_2 = (\alpha_{j,\ell,m}(\mathbf{d}-\mathbf{d}^0) - \widetilde{\alpha}_{j,\ell,m})_{\ell,m}$ has positive terms for all $\mathbf{d}\in\{\mathbf{d},\forall\ell\ d_\ell\le d_\ell^{min}\}$, and it is thus positive semidefinite. Applying Proposition 13, we obtain that $\mathbf{B}(j)=\mathbf{B}_1(j)\circ\mathbf{B}_2(j)$ is positive semidefinite. $\mathbf{A}(\alpha(\mathbf{d}-\mathbf{d}^0)) - \mathbf{A}(\widetilde{\alpha})$ is then positive semidefinite, being the sum of positive semidefinite matrices. Similarly, it is easy to check that both $\mathbf{A}(\widetilde{\alpha})$ and $\mathbf{A}(\alpha(\mathbf{d}-\mathbf{d}^0))$ are positive semidefinite. Consequently, Proposition 12 gives $\log\det\mathbf{A}(\alpha(\mathbf{d}-\mathbf{d}^0)) \ge \log\det\mathbf{A}(\widetilde{\alpha})$. This result holds for all $\mathbf{d}$ satisfying $\forall\ell,\ d_\ell\le d_\ell^{min}$.

We now study the behaviour of $\log\det\mathbf{A}(\widetilde{\alpha}) - \log\det\hat{\mathbf{G}}(\mathbf{d}^0)$. First, we have proved previously (see end of Section 7 D.1) that $\log\det\hat{\mathbf{G}}(\mathbf{d}^0)$ tends uniformly in $\mathbf{d}$ to $\log\det\mathbf{G}^0$. Second, we decompose $\mathbf{A}(\widetilde{\alpha})$ in $\mathbf{A}(\widetilde{\alpha}) = \widetilde{\mathbf{G}} + \mathbf{S}(\widetilde{\alpha})$ with the elements of $\mathbf{S}(\widetilde{\alpha})$ defined in equation (C1) and $\widetilde{\mathbf{G}}_{\ell,m} = \sum_j n_j\widetilde{\alpha}_{j,\ell,m}G^0_{\ell,m}$. We distinguish the study of the two terms.

- As $<\mathcal{J}>\sim j_0$, for sufficiently large N, there exists a positive constant c such that
$$|\widetilde{\alpha}_{j,\ell,m}| \le \frac{c}{n}2^{(j-j_0)(d_\ell^0-d_\ell^{min}+d_m^0-d_m^{min})}.$$
Consequently, for all (ℓ,m), the sequence $\widetilde{\alpha}_{\ell,m}$ belongs to $\mathcal{S}(0,\gamma,c)$ with $\gamma = 2\sup_a(d_a^0-d_a^{min})$. As the vector $\mathbf{d}^{min}$ satisfies that for any $\ell=1,\dots,p$, we have $d_\ell^0-1/2<d_\ell^{min}\le d_\ell^0$, then $0\le\gamma<1/2$. Applying Proposition 9, we deduce that $\mathbf{S}(\widetilde{\alpha})$ tends to 0 in probability uniformly in $\mathbf{d}$.
- As $\mathbf{G}^0$ is a covariance matrix, it is positive semidefinite. We can apply Oppenheim's inequality (Proposition 11), $\log\det\widetilde{\mathbf{G}} \ge \log\det(\mathbf{G}^0) + \sum_\ell \log(\sum_{j=j_0}^{j_1} n_j\widetilde{\alpha}_{j,\ell,\ell})$. As we defined $\mathbf{d}^{min}$ such that (D2) holds, it follows that $\liminf_{N\to\infty}\inf_{\{\mathbf{d},\forall\ell\ d_\ell\le d_\ell^{min}\}}\inf_{j_1=j_0,\dots,j_N}\log\det\mathbf{A}(\widetilde{\alpha}) - \log\det\mathbf{G}^0>0$.

We thus obtain
$$\lim_{N\to\infty}\mathbb{P}\left(\inf_{\{\mathbf{d},\forall\ell\ d_\ell\le d_\ell^{min}\}}\ \inf_{j_1=j_0,\dots,j_N}\log\det\mathbf{A}(\widetilde{\alpha}) - \log\det\hat{\mathbf{G}}(\mathbf{d}^0)>0\right)=1. \qquad \text{(D3)}$$
Suppose $\hat{d}\in\{\mathbf{d},\forall\ell\ d_\ell\le d_\ell^{min}\}$. By definition of $\hat{\mathbf{d}}$, inequality $\inf_{\{\mathbf{d},\ \forall m\ d_m\le d_m^{min}\}} R(d) - R(d^0)\le 0$ holds, which is in contradiction with result (D3). Finally, with a probability tending to 1, there exists $\ell_1\in\{1,\dots,p\}$ such that $d_{\ell_1}\ge d_{\ell_1}^{min}$.

DOI: 10.1111/jtsa.12170

STEP 2.
Suppose there exist $\ell_1, \ell_2, \dots \ell_k$ with $k < p$ such that, with a probability tending to 1, $\hat{d}_{\ell_i} \geq d^{min}_{\ell_i}$ for all $i = 1, \dots, k$. We introduce $\widetilde{\alpha}^{(k)}$ defined by $\widetilde{\alpha}^{(k)}_{j,\ell,m} = \alpha_{j,\ell,m}(\mathbf{d} - \mathbf{d}^0)$ if $j_0 \leq j \leq \mathcal{J}$,

$$\widetilde{\alpha}^{(k)}_{j,\ell,m} = \begin{cases} \alpha_{j,\ell,m}\left(\mathbf{d} - \mathbf{d}^0\right) & \text{if } \ell, m \in \{\ell_1, \ell_2, \dots \ell_k\} \\ \alpha_{j,\ell,m}\left(\mathbf{d}^{min} - \mathbf{d}^0\right) & \text{if } \ell \notin \{\ell_1, \ell_2, \dots \ell_k\} \text{ and } m \notin \{\ell_1, \ell_2, \dots \ell_k\} \\ \frac{1}{n} 2^{-(j - <\mathcal{J}>)(d_\ell - d^0_\ell + d^{min}_m - d^0_m)} & \text{if } \ell \in \{\ell_1, \ell_2, \dots \ell_k\} \text{ and } m \notin \{\ell_1, \ell_2, \dots \ell_k\} \\ \frac{1}{n} 2^{-(j - <\mathcal{J}>)(d^{min}_\ell - d^0_\ell + d_m - d^0_m)} & \text{if } \ell \notin \{\ell_1, \ell_2, \dots \ell_k\} \text{ and } m \in \{\ell_1, \ell_2, \dots \ell_k\} \end{cases}$$

if $< \mathcal{J} >< j \leq j_1$ and $\widetilde{\alpha}^{(k)}_{j,\ell,m} = 0$ otherwise. It is straightforward that for such $\widetilde{\alpha}^{(k)}$ the three following points hold:

1. For all $\mathbf{d} \in \{\mathbf{d}, \forall \ell \notin \{\ell_1, \dots, \ell_k\}\ d_\ell \leq d^{min}_\ell\}$, for all $j \geq 0$, for all $(\ell, m) \in \{1, \dots, p\}^2$, we have $\alpha_{j,\ell,m}(\mathbf{d} - \mathbf{d}^0) - \widetilde{\alpha}^{(k)}_{j,\ell,m} \geq 0$.
2. For all $(\ell, m) \in \{1, \dots, p\}^2$, the sequence $(\widetilde{\alpha}^{(k)}_{j,\ell,m})_{j \geq 0}$ belongs to $\mathcal{S}(0, \gamma, c)$ with $0 \leq \gamma < 1/2$.
3. $\liminf_{n \to \infty} \inf_{\{\mathbf{d}, \forall \ell \notin \{\ell_1, \dots, \ell_k\}\ d_\ell \leq d^{min}_\ell\}} \inf_{j_1 = j_0, \dots, j_N} \sum_\ell \log(\sum_{j=j_0}^{j_1} n_j \widetilde{\alpha}^{(k)}_{j,\ell_0,\ell_0}) > 0$.

Analysis similar to STEP 1 shows that there exists $\ell_{k+1} \notin \{\ell_1, \ell_2, \dots \ell_k\}$ such that, with a probability tending to 1, $d_{\ell_{k+1}} \geq d^{min}_{\ell_{k+1}}$.

Steps 1 and 2 imply that $\mathbb{P}(\hat{\mathbf{d}} \in \{\mathbf{d}, \forall \ell = 1, \dots, p, d_\ell \geq d^{min}_\ell\}) \to 1$ when $N \to \infty$.

D3. Proof of Theorem 5

Suppose first that for sufficiently large N, we have $j_1 \geq j_0 + 2$.

We have established that there exists $\mathbf{d}^{min}$ such that for all $\ell = 1, \dots, p$, $d^0_\ell - 1/2 < d^{min}_\ell < d^0_\ell$ and $\mathbb{P}(\sup_{j_1 \geq j_0 + 2} \hat{d}_\ell \leq d^{min}_\ell)$ tends to 0 when N goes to infinity.

Let $\mathbf{d}$ be a $\mathbb{R}^p$-vector satisfying $d_\ell \geq d^{min}_\ell$ for all $\ell = 1, \dots, p$. Recall that

$$E_{\ell,m}(\mathbf{d}) = G^0_{\ell,m} + S_{\ell,m}\left(\mu_{\ell,m}\left(\mathbf{d} - \mathbf{d}^0\right)\right) \text{ with } \mu_{j,\ell,m}(\boldsymbol{\delta}) = \frac{2^{-j(\delta_\ell + \delta_m)}}{\sum_{j' \geq j_0} n_{j'} 2^{-j'(\delta_\ell + \delta_m)}}. \tag{D4}$$

We bound the sequence μ as follows:

$$\left|\mu_{j,\ell,m}\left(\mathbf{d} - \mathbf{d}^0\right)\right| \leq \frac{2^{(j-j_0)(d^0_\ell - d^{min}_\ell + d^0_m - d^{min}_m)}}{\sum_{j'=j_0}^{j_1} n_{j'} 2^{(j'-j_0)(d^0_\ell - d_\ell + d^0_m - d_m)}} \leq n^{-1}_{j_0} 2^{(j-j_0)(d^0_\ell - d^{min}_\ell + d^0_m - d^{min}_m)}.$$

Since $n_{j_0} \sim N 2^{-j_0}$ and $n \sim N 2^{-j_0}(2 - 2^{j_1 - j_0})$, there exists a sufficiently large N such that $n^{-1}_{j_0} \leq 2n^{-1}$. We introduce $\gamma = \sup_{\ell=1,\dots,p} d^0_\ell - d^{min}_\ell$. Then for all (ℓ, m), $\mu_{\ell,m}(\mathbf{d} - \mathbf{d}^0)$ belongs to $\mathcal{S}(0, 2\gamma, 2)$. We have $0 \leq \gamma < 1/2$ with a probability tending to 1. Applying Proposition 9, we deduce that $E_{\ell,m}(\mathbf{d} - \mathbf{d}^0) \to G^0_{\ell,m}$ uniformly in $\mathbf{d}$ with a probability tending to 1. Following the proof of Proposition 10, with this result, we obtain Theorem 5.

It remains to consider the case where we do not have $j_1 \geq j_0 + 2$. Since we supposed $j_1 - j_0$ is an increasing sequence of N, Proposition 10 holds, which concludes the proof.

APPENDIX E: RATE OF CONVERGENCE

In order to obtain the optimal rate of convergence, we first prove the convergence with a suboptimal rate. Based on this result, we are able to obtain feasible conditions under which the rates of Theorem 6 hold.

DOI: 10.1111/jtsa.12170

E1. Convergence with a Suboptimal Rate

We shall now establish a first rate of convergence that is not optimal but that will be useful to derive conditions for the optimal rate.

Proposition 14. Assume that (W1)–(W5) and Condition (C) hold. If in addition j_0 is chosen such that $2^{-j_0\beta} + N^{-1/2}2^{j_0/2} \to 0$, then

$$\hat{\mathbf{d}} - \mathbf{d}^0 = O_{\mathbb{P}}\left(N^{-\frac{1}{4}}2^{\frac{j_0}{4}} + 2^{-\beta\frac{j_0}{2}}\right).$$

Proof
The proof is based on inequality (D1). The procedure is to find a lower bound for $R(\mathbf{d}) - R(\mathbf{d}^0)$ on the set $\{\mathbf{d},\ \max_{\ell=1,\dots,p}|d_\ell - d_\ell^0| < 1/4\}$. By Theorem 5, $\hat{\mathbf{d}} - \mathbf{d}^0$ goes to 0 in probability. Therefore, for sufficiently large N, we have $\max_{\ell=1,\dots,p}|\hat{d}_\ell - d_\ell^0| < 1/4$.

First, a second-order Taylor expansion of $L(\mathbf{d}-\mathbf{d}^0)$ at the neighbourhood of 0 for $\mathbf{d}^0 - 1/4 < \mathbf{d} \leq \mathbf{d}^0$ is $L(\mathbf{d}-\mathbf{d}^0) = \left.\frac{d^2L(u)}{du^2}\right|_{\bar{u}}(\mathbf{d}-\mathbf{d}^0)^2 + o(\max_{\ell=1,\dots,p}|d_\ell - d_\ell^0|^2)$ with $|\bar{u}| \leq 1/4$. Proposition 6 of Moulines *et al.* (2008) states that $\liminf_{N\to\infty}\inf_{\bar{u}\in[-1/2,0]}\inf_{j_1=j_0+1,\dots,j_N}\left.\frac{d^2L(u)}{du^2}\right|_{\bar{u}} > 0$. Thus, for $\mathbf{d}^0 - 1/4 < \mathbf{d} \leq \mathbf{d}^0 + 1/4$, there exists $c > 0$ such that $L(\mathbf{d}-\mathbf{d}^0) > c\sum_{\ell=1}^p(d_\ell - d_\ell^0)^2 + o(\max_{\ell=1,\dots,p}|d_\ell - d_\ell^0|^2)$.

We now want to establish an upper bound for $\Delta(\mathbf{d},\mathbf{d}^0)$. This quantity satisfies

$$\Delta\left(\mathbf{d},\mathbf{d}^0\right) = \log\det\left(I + \mathbf{G}^{0-1}\mathbf{S}\left(\mu\left(\mathbf{d}-\mathbf{d}^0\right)\right)\right),$$

where $\mathbf{S}(\mu(\mathbf{d}-\mathbf{d}^0))$ is the matrix with (ℓ,m)th element $S_{\ell,m}(\mu(\mathbf{d}-\mathbf{d}^0))$ defined in (C1) and $\mu_{j,\ell,m}(\cdot)$ defined in (D4). Thus, $\Delta(\mathbf{d},\mathbf{d}^0) = \log(\prod_{i=1}^p\lambda_i)$ where $(\lambda_i)_{i=1,\dots,p}$ denote the eigenvalues of the semidefinite positive matrix $\mathbf{G}^{0-1}\mathbf{S}(\mu(\mathbf{d}-\mathbf{d}^0))$. Hence,

$$0 \leq \Delta\left(\mathbf{d},\mathbf{d}^0\right) \leq trace\left(\mathbf{G}^{0-1}\mathbf{S}\left(\mu\left(\mathbf{d}-\mathbf{d}^0\right)\right)\right).$$

Since $\max_{\ell=1,\dots,p}|d_\ell - d_\ell^0| \to 0$, for sufficiently large N, the quantity $\gamma = \max_{\ell=1,\dots,p}|d_\ell - d_\ell^0|$ satisfies $0 \leq \gamma < 1/4$. As established previously, for all (ℓ,m), $\mu_{\ell,m}(\mathbf{d}-\mathbf{d}^0)$ belongs to $\mathcal{S}(0,2\gamma,2)$. Then for all (ℓ,m), $S_{\ell,m}(\mu_{\ell,m}(\mathbf{d}-\mathbf{d}^0)) = O_{\mathbb{P}}(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})$ uniformly in $\mathbf{d} \in [\mathbf{d}^0 - 1/4, \mathbf{d}^0]$ applying Proposition 9. Hence, $\Delta(\mathbf{d},\mathbf{d}^0) = O_{\mathbb{P}}(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})$.

Inequality (D1) thus gives

$$R(\mathbf{d}) - R\left(\mathbf{d}^0\right) \geq c\sum_{\ell=1}^p\left(d_\ell - d_\ell^0\right)^2 + o\left(\max_\ell\left|d_\ell - d_\ell^0\right|^2\right) + O_{\mathbb{P}}\left(2^{-j_0\beta} + N^{-\frac{1}{2}}2^{\frac{j_0}{2}}\right)$$

with $c > 0$. It follows that for all $\ell = 1,\dots,p$, $(\hat{d}_\ell - d_\ell^0)^2 = O_{\mathbb{P}}(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})$. □

E2. Proof of Theorem 6

The criterion R is equal to $R(\mathbf{d}) = \log\det(\mathbf{\Lambda}_{<\mathcal{J}>}(\mathbf{d})\hat{\mathbf{G}}(\mathbf{d})\mathbf{\Lambda}_{<\mathcal{J}>}(\mathbf{d}))$ –1. It is straightforward that $\hat{\mathbf{d}} = argmin_{\mathbf{d}}\,R(\mathbf{d})$ satisfies

$$\hat{\mathbf{d}} = \underset{\mathbf{d}}{argmin}\,\overline{R}(\mathbf{d}) \text{ with } \overline{R}(\mathbf{d}) = \log\det\overline{\boldsymbol{G}}(\mathbf{d}) \text{ and } \overline{\boldsymbol{G}}(\mathbf{d}) = \mathbf{\Lambda}_{<\mathcal{J}>}\left(\mathbf{d}-\mathbf{d}^0\right)\hat{\mathbf{G}}(\mathbf{d})\mathbf{\Lambda}_{<\mathcal{J}>}\left(\mathbf{d}-\mathbf{d}^0\right) \tag{E1}$$

DOI: 10.1111/jtsa.12170

The Taylor expansion of $\overline{R}$ at $\hat{\mathbf{d}}$ at the neighbourhood of $\mathbf{d}^0$ gives

$$R\left(\hat{\mathbf{d}}\right) - R\left(\mathbf{d}^0\right) = \left.\frac{\partial \overline{R}(\mathbf{d})}{\partial \mathbf{d}}\right|_{\mathbf{d}^0} \left(\mathbf{d} - \mathbf{d}^0\right) + \left(\hat{\mathbf{d}} - \mathbf{d}^0\right)^{\mathrm{T}} \left.\frac{\partial^2 \overline{R}(\mathbf{d})}{\partial \mathbf{d} \partial \mathbf{d}^T}\right|_{\bar{\mathbf{d}}} \left(\hat{\mathbf{d}} - \mathbf{d}^0\right), \tag{E2}$$

where $\bar{\mathbf{d}}$ is such that $\|\bar{\mathbf{d}} - \mathbf{d}^0\| \le \|\hat{\mathbf{d}} - \mathbf{d}^0\|$.

The derivatives of the criterion $\overline{R}(\mathbf{d})$ are equal to

$$\frac{\partial \overline{R}(\mathbf{d})}{\partial d_a} = trace\left(\overline{\boldsymbol{G}}(\mathbf{d})^{-1} \frac{\partial \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a}\right) \tag{E3}$$

$$\frac{\partial^2 \overline{R}(\mathbf{d})}{\partial d_a \partial d_b} = -trace\left(\overline{\boldsymbol{G}}(\mathbf{d})^{-1} \frac{\partial \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_b} \overline{\boldsymbol{G}}(\mathbf{d})^{-1} \frac{\partial \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a}\right) + trace\left(\overline{\boldsymbol{G}}(\mathbf{d})^{-1} \frac{\partial^2 \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a \partial d_b}\right) \tag{E4}$$

when $\overline{\boldsymbol{G}}(\mathbf{d})^{-1}$ exists.

E.2.1. Study of $\overline{\boldsymbol{G}}$ and its Derivatives

To study the asymptotic behaviour of the derivatives of the criterion, it is necessary first to study the asymptotic behaviour of $\overline{\boldsymbol{G}}$ and of its derivatives.

For any $a = 1, \ldots, p$, let $\mathbf{i}_a$ be a $p \times p$ matrix whose ath diagonal element is one and all other elements are zero. Let a and b be two indexes in $1, \ldots, p$. The first derivative of $\overline{\boldsymbol{G}}(\mathbf{d})$ with respect to d_a, $\frac{\partial \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a}$, is equal to

$$-\log(2)\frac{1}{n}\sum_{j=j_0}^{j_1}(j - <\mathcal{J}>)\boldsymbol{\Lambda}_{<\mathcal{J}>}\left(\mathbf{d} - \mathbf{d}^0\right)\boldsymbol{\Lambda}_j(\mathbf{d})^{-1}\left(\mathbf{i}_a\mathbf{I}(j) + \mathbf{I}(j)\mathbf{i}_a\right)\boldsymbol{\Lambda}_j(\mathbf{d})^{-1}\boldsymbol{\Lambda}_{<\mathcal{J}>}\left(\mathbf{d} - \mathbf{d}^0\right).$$

And the second derivative, with respect to d_a and d_b,

$$\begin{aligned}\frac{\partial^2 \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a \partial d_b} = \log(2)^2 \boldsymbol{\Lambda}_{<\mathcal{J}>}\left(\mathbf{d} - \mathbf{d}^0\right)\frac{1}{n}\sum_{j=j_0}^{j_1}(j - <\mathcal{J}>)^2 \boldsymbol{\Lambda}_j(\mathbf{d})^{-1} \\ \left(\mathbf{i}_b\mathbf{i}_a\mathbf{I}(j) + \mathbf{I}(j)\mathbf{i}_a\mathbf{i}_b + \mathbf{i}_b\mathbf{I}(j)\mathbf{i}_a + \mathbf{i}_a\mathbf{I}(j)\mathbf{i}_b\right)\boldsymbol{\Lambda}_j(\mathbf{d})^{-1}\Lambda_{<\mathcal{J}>}\left(\mathbf{d} - \mathbf{d}^0\right)\end{aligned} \tag{E5}$$

E.2.2. Convergence of $\overline{\boldsymbol{G}}(\mathbf{d})$

Let ℓ, m be given indexes in $\{1, \ldots, p\}$. Any (ℓ, m)th element of the matrix $\overline{\boldsymbol{G}}(\mathbf{d})$ satisfies

$$\bar{G}_{\ell,m}(\mathbf{d}) = G^0_{\ell,m}\sum_{j=j_0}^{j_1} n_j \omega^{(0)}_{j,\ell,m} + S_{\ell,m}\left(\omega^{(0)}_{\ell,m}\left(\mathbf{d} - \mathbf{d}^0\right)\right)$$

where $\omega^{(0)}_{j,\ell,m}(\mathbf{d} - \mathbf{d}^0) = \frac{1}{n}2^{-(j-<\mathcal{J}>)(d_\ell - d^0_\ell + d_m - d^0_m)}$.

Recall that $<\mathcal{J}> = (j_0 + \eta_{j_1-j_0})(1 + o(1))$ with $0 \le \eta_{j_1-j_0} \le 1$. Let $\mathbf{d}$ be a $\mathbb{R}^p$-vector such that for all $\ell = 1, \ldots, p$, $\hat{d}_\ell \le d_\ell \le d^0_\ell$. As $\sup_\ell |\hat{d}_\ell - d^0_\ell| = o_{\mathbb{P}}(1)$, for any $\gamma \in (0, 1/2)$, there exists N_γ such that for any $N \ge N_\gamma$, $2^{-(j-<\mathcal{J}>)(d_\ell - d^0_\ell + d_m - d^0_m)} \le 2^\gamma 2^{(j-j_0)\gamma}$. For $N \ge N_\gamma$, the sequence $\omega^{(0)}_{\ell,m}(\hat{\mathbf{d}} - \mathbf{d}^0)$ belongs to $\mathcal{S}(0,\gamma,2^{\gamma})$. Proposition 9 shows that $S_{\ell,m}(\omega^{(0)}_{\ell,m}(\mathbf{d}-\mathbf{d}^0))$ tends to zero when $2^{-j_0\beta}+N^{-1/2}2^{j_0/2}\to 0$ uniformly in $\mathbf{d}$.

Finally, we shall prove that $\sum_{j=j_0}^{j_1} n_j\omega^{(0)}_{j,\ell,m}(\mathbf{d}-\mathbf{d}^0)\to 1$. Since $|2^a-1|\le 2^{|a|}-1\le \log(2)|a|2^{|a|}$ for all $a\in\mathbb{R}$,

$$|\sum_{j=j_0}^{j_1} n_j\omega^{(0)}_{j,\ell,m}\left(\mathbf{d}-\mathbf{d}^0\right)-1|\le \frac{1}{n}\sum_j n_j\log(2)|j-<\mathcal{J}>|\max_{\ell=1,\dots,p}|d_\ell-d^0_\ell|2^{|j-<\mathcal{J}>|\max_{\ell=1,\dots,p}\left|d_\ell-d_\ell^0\right|}$$
$$\le 2\log(N)\max_{\ell=1,\dots,p}\left|d_\ell-d_\ell^0\right|2^{2\log_2(N)\max_{\ell=1,\dots,p}\left|d_\ell-d_\ell^0\right|}.$$

The last inequality is a consequence of $j_0\le j_1\le j_N=\log_2(N)$. Under assumption $\log(N)\max_{\ell=1,\dots,p}|\hat{d}_\ell-d^0_\ell|\to 0$, the right-hand side of the inequality goes to 0 when N goes to infinity.

Consequently,

$$\sup_{\{\mathbf{d},\,\|\mathbf{d}-\mathbf{d}^0\|\le\|\hat{\mathbf{d}}-\mathbf{d}^0\|\}}\bar{G}_{a,b}(\mathbf{d})=G^0_{a,b}\left(1+o_{\mathbb{P}}\left(\log(N)\max_\ell\left|d_\ell-d^0_\ell\right|\right)\right)+o_{\mathbb{P}}\left(2^{-j_0\beta}+N^{-\frac{1}{2}}2^{\frac{j_0}{2}}\right).\tag{E6}$$

Because of Proposition 14, it is sufficient that $\log(N)^2(2^{-j_0\beta}+N^{-1/2}2^{j_0/2})\to 0$ to obtain $\sup_{\{\mathbf{d},\,\|\mathbf{d}-\mathbf{d}^0\|\le\|\hat{\mathbf{d}}-\mathbf{d}^0\|\}}\bar{G}_{a,b}(\mathbf{d})=G^0_{a,b}+o_{\mathbb{P}}(1)$. Thus, for sufficiently large N, $\overline{\boldsymbol{G}}(\mathbf{d})$ is invertible, and $\overline{\boldsymbol{G}}(\mathbf{d})^{-1}$ converges in probability to $\mathbf{G}^{0-1}$ on the set $\{\mathbf{d},\ \|\mathbf{d}-\mathbf{d}^0\|\le\|\hat{\mathbf{d}}-\mathbf{d}^0\|\}$.

E.2.3. Convergence of $\left.\frac{\partial\overline{G}(d)}{\partial d_a}\right|_{\mathbf{d}}$

This section concerns the convergence in probability of $\left(\left.\frac{\partial\overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a}\right|_{\mathbf{d}}\right)_{a,b}$, which is equal to

$$\left(\left.\frac{\partial\overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a}\right|_{\mathbf{d}}\right)_{a,b}=\log(2)\frac{1}{n}\sum_{j=j_0}^{j_1}(j-<\mathcal{J}>)2^{-j\left(d_a^0+d_b^0\right)}I_{a,b}(j)$$
$$=\log(2)\left[G^0_{a,b}\sum_{j=j_0}^{j_1}n_j\omega^{(1)}\left(\mathbf{d}-\mathbf{d}^0\right)+S_{a,b}\left(\omega^{(1)}\left(\mathbf{d}-\mathbf{d}^0\right)\right)\right],$$

where $\omega^{(1)}_j(\boldsymbol{\delta})=\frac{1}{n}(j-<\mathcal{J}>)2^{-(j-<\mathcal{J}>)(\delta_\ell+\delta_m)}$.

We first study the behaviour of $\sum_{j=j_0}^{j_1}n_j\omega^{(1)}(\mathbf{d}-\mathbf{d}^0)$.

$$\left|\sum_{j=j_0}^{j_1}n_j\omega^{(1)}(\mathbf{d}-\mathbf{d}^0)\right|\le\frac{1}{n}\sum_j n_j\log(2)\,|j-<\mathcal{J}>|\max_{\ell=1,\dots,p}\left|d_\ell-d^0_\ell\right|2^{|j-<\mathcal{J}>|\max_{\ell=1,\dots,p}\left|d_\ell-d^0_\ell\right|}$$
$$\le\log(N)\max_{\ell=1,\dots,p}\left|d_\ell-d^0_\ell\right|2^{\log_2(N)\max_{\ell=1,\dots,p}\left|d_\ell-d^0_\ell\right|}$$

It is thus sufficient that $\log(N)\max_{\ell=1,\dots,p}|\hat{d}_\ell-d^0_\ell|\to 0$ to have $\sum_{j=j_0}^{j_1}n_j\omega^{(1)}(\mathbf{d}-\mathbf{d}^0)=o_{\mathbb{P}}(1)$.

Let $\mathbf{d}$ be in a neighbourhood of $\mathbf{d}^0$ such that $\sup_\ell|d_\ell-d^0_\ell|<\gamma$ with $0\le\gamma<1/2$. As $<\mathcal{J}>\sim j_0+\eta_{j_1-j_0}$ with $0\le\eta_{j_1-j_0}\le 1$, there exists N_0 such that for any $N\ge N_0$, the sequence $\omega^{(1)}(\mathbf{d}-\mathbf{d}^0)$ belongs to $\mathcal{S}(1,\gamma,2^{\gamma})$. Thanks to Proposition 9, it comes that $S_{a,b}\omega^{(1)}(\mathbf{d}-\mathbf{d}^0)=O_{\mathbb{P}}\left(2^{-j_0\beta}+N^{-1/2}2^{j_0/2}\right)$ uniformly on the neighbourhood. Consequently, $\left(\left.\frac{\partial\overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a}\right|_{\mathbf{d}}\right)_{a,b}=o_{\mathbb{P}}(1)$. Finally, considering similarly the other terms, when $\log(N)\max_{\ell=1,\dots,p}|\hat{d}_\ell-d^0_\ell|\to 0$, the equivalence $\left.\frac{\partial\overline{\mathbf{G}}}{\partial d_a}\right|_{\mathbf{d}}=o_{\mathbb{P}}(1)$ is fulfilled under assumptions of Theorem 6.

DOI: 10.1111/jtsa.12170

In addition, when $\mathbf{d} = \mathbf{d}^0$, we have $\sum_{j=j_0}^{j_1} n_j \omega_j^{(1)}(\mathbf{0}) = 0$. Hence, $\left(\frac{\partial \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a}\Big|_{\mathbf{d}^0} \right)_{a,b} = C_{a,b}^{(1)}$ where $C_{a,b}^{(1)} = O_{\mathbb{P}}(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})$. Or more generally,

$$\frac{\partial \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a}\Bigg|_{\mathbf{d}^0} = \mathbf{i}_a \mathbf{C}^{(1)} + \mathbf{C}^{(1)} \mathbf{i}_a, \tag{E7}$$

where each term of the matrix $\mathbf{C}^{(1)}$ is $O_{\mathbb{P}}(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})$.

E.2.4. Convergence of $\frac{\partial^2 \overline{\boldsymbol{G}}_{a,b}(\mathbf{d})}{\partial d_a \partial d_b}\Big|_{\mathbf{d}}$

Let $\mathbf{d}$ be a $\mathbb{R}^p$-vector such that $|d_\ell - d_\ell^0| \leq |\hat{d}_\ell - d_\ell^0|$ for all $\ell = 1, \ldots, p$. The proof is derived for $\left(\frac{\partial^2 \overline{\boldsymbol{G}}_{a,b}(\mathbf{d})}{\partial d_a \partial d_b}\Big|_{\mathbf{d}} \right)_{a,b}$, for $a \neq b$. The argumentation is similar for the diagonal terms. Introducing a sequence $S_{a,b}(\cdot)$, the expression (E5) is rewritten as

$$\begin{aligned}\left(\frac{\partial^2 \overline{\boldsymbol{G}}_{a,b}(\mathbf{d})}{\partial d_a \partial d_b}\Bigg|_{\mathbf{d}} \right)_{a,b} &= \log(2)^2 \left[2^{<\mathcal{J}>(d_a+d_b-d_a^0-d_b^0)} \frac{1}{n} \sum_{j=j_0}^{j_1} (j - <\mathcal{J}>)^2 \frac{I_{a,b}(j)}{2^{j(d_a+d_b)}} \right] \\ &= \log(2)^2 \left[G_{a,b}^0 \sum_{j=j_0}^{j_1} n_j \omega_{j,a,b}^{(2)} \left(\mathbf{d} - \mathbf{d}^0\right) + S_{a,b}\left(\omega_{a,b}^{(2)} \left(\mathbf{d} - \mathbf{d}^0\right) \right) \right]\end{aligned}$$

where $\omega_{j,a,b}^{(2)}(\boldsymbol{\delta}) = \frac{1}{n}(j - <\mathcal{J}>)^2 2^{-(j-<\mathcal{J}>)(\delta_a+\delta_b)}$.

First, we want to prove that

$$\sum_{j=j_0}^{j_1} n_j \omega_{j,a,b}^{(2)} \left(\mathbf{d} - \mathbf{d}^0\right) = \kappa_{j_1-j_0} \left(1 + o_{\mathbb{P}}(1)\right). \tag{E8}$$

Recall $\kappa_{j_1-j_0} = \frac{1}{n} \sum_{j=j_0}^{j_1} (j - <\mathcal{J}>)^2 n_j$. Since $|2^a - 1| \leq 2^{|a|} - 1 \leq \log(2)|a|2^{|a|}$ for all $a \in \mathbb{R}$, we have the inequality

$$\left| \sum_{j=j_0}^{j_1} n_j \omega_{j,a,b}^{(2)} \left(\bar{\mathbf{d}} - \mathbf{d}^0\right) - \kappa_{j_1-j_0} \right| \leq 2\log(2)\kappa_{j_1-j_0} \log_2(N) \left| d - d^0 \right| 2^{2\log_2(N)\left|d-d^0\right|}.$$

Equation (E8) holds when $\log(N) \max_{\ell=1,\ldots,p} |d_\ell - d_\ell^0| \to 0$. Because of Proposition 14, it is sufficient that $\log(N)^2 (2^{-j_0\beta} + N^{-1/2}2^{j_0/2}) \to 0$.

Let $\mathbf{d}$ be in a neighbourhood of $\mathbf{d}^0$ such that $\sup_\ell |d_\ell - d_\ell^0| < \gamma$ with $0 \leq \gamma < 1/2$. As $<\mathcal{J}> \sim j_0 + \eta_{j_1-j_0}$ with $0 \leq \eta_{j_1-j_0} \leq 1$, there exists N_0 such that for any $N \geq N_0$, the sequence $\omega_{a,b}^{(2)}(\mathbf{d} - \mathbf{d}^0)$ belongs to the set $\mathcal{S}(2, \gamma, 2^\gamma)$. Using Proposition 9, $S_{a,b}(\omega_{a,b}^{(2)}(\mathbf{d} - \mathbf{d}^0)) \leq C(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})$ uniformly on the neighbourhood of $\mathbf{d}$ for $N \geq N_0$. As a consequence,

$$\left(\frac{\partial^2 \overline{\boldsymbol{G}}_{a,b}(\mathbf{d})}{\partial d_a \partial d_b}\Bigg|_{\bar{\mathbf{d}}} \right)_{a,b} = \log(2)^2 \kappa_{j_1-j_0} G_{a,b}^0 \left(1 + o_{\mathbb{P}}(1)\right) + C_{a,b}^{(2)},$$

with $C_{a,b}^{(2)} = O_{\mathbb{P}}(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})$.

DOI: 10.1111/jtsa.12170

Finally, when N goes to infinity,

$$\sup_{\{\mathbf{d},\,\|\mathbf{d}-\mathbf{d}^0\|\leq\|\hat{\mathbf{d}}-\mathbf{d}^0\|\}} \left.\frac{\partial^2 \overline{\boldsymbol{G}}_{a,b}(\mathbf{d})}{\partial d_a \partial d_b}\right|_{\bar{\mathbf{d}}} = \log(2)^2 \kappa_{j_1-j_0} \left(\mathbf{i}_b \mathbf{i}_a \mathbf{G}^0 + \mathbf{i}_b \mathbf{G}^0 \mathbf{i}_a + \mathbf{i}_a \mathbf{G}^0 \mathbf{i}_b + \mathbf{G}^0 \mathbf{i}_a \mathbf{i}_b\right) + o_{\mathbb{P}}(1) \tag{E9}$$

when $\log(N) \max_{\ell=1,\dots,p} |d_\ell - d_\ell^0| = o_{\mathbb{P}}(1)$ and $2^{-j_0\beta} + N^{-1/2} 2^{j_0/2} \to 0$.

E.2.5. Second Derivative of the Criterion

Let us detail the expression of the second derivative of R with respect to $\mathbf{d}$ at $\bar{\mathbf{d}}$,

$$\left.\frac{\partial^2 \overline{R}(\mathbf{d})}{\partial d_a \partial d_b}\right|_{\bar{\mathbf{d}}} = -trace\left(\overline{\boldsymbol{G}}(\bar{\mathbf{d}})^{-1} \left.\frac{\partial \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_b}\right|_{\bar{\mathbf{d}}} \overline{\boldsymbol{G}}\left(\bar{\mathbf{d}}\right)^{-1} \left.\frac{\partial \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a}\right|_{\bar{\mathbf{d}}}\right) + trace\left(\overline{\boldsymbol{G}}\left(\bar{\mathbf{d}}\right)^{-1} \left.\frac{\partial^2 \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a \partial d_b}\right|_{\bar{\mathbf{d}}}\right).$$

Using (E6) and the previous study that established that $\left.\frac{\partial \overline{\boldsymbol{G}}(\mathbf{d})}{\partial d_a}\right|_{\bar{\mathbf{d}}} = o_{\mathbb{P}}(1)$, the first term tends to 0 under assumptions of Theorem 6. Combining (E6) and (E9), we can assert that $\left.\frac{\partial^2 \overline{R}(\mathbf{d})}{\partial d_a \partial d_b}\right|_{\bar{\mathbf{d}}}$ tends in probability to

$$\log(2)^2 \kappa_{j_1-j_0} trace\left(\mathbf{G}^{0-1}\left(\mathbf{i}_b \mathbf{i}_a \mathbf{G}^0 + \mathbf{i}_b \mathbf{G}^0 \mathbf{i}_a + \mathbf{i}_a \mathbf{G}^0 \mathbf{i}_b + \mathbf{G}^0 \mathbf{i}_a \mathbf{i}_b\right)\right).$$

Let $G_{\ell,m}^{0(-1)}$ denotes the (ℓ, m)th element of $\mathbf{G}^{0-1}$. When $a \neq b$,

$$trace\left(\mathbf{G}^{0-1}\left(\mathbf{i}_b \mathbf{i}_a \mathbf{G}^0 + \mathbf{i}_b \mathbf{G}^0 \mathbf{i}_a + \mathbf{i}_a \mathbf{G}^0 \mathbf{i}_b + \mathbf{G}^0 \mathbf{i}_a \mathbf{i}_b\right)\right) = G_{a,b}^{0(-1)} G_{a,b}^0 + G_{b,a}^{0(-1)} G_{b,a}^0 = 2 G_{a,b}^{0(-1)} G_{a,b}^0.$$

When $a = b$,

$$trace\left(\mathbf{G}^{0-1}\left(\mathbf{i}_b \mathbf{i}_a \mathbf{G}^0 + \mathbf{i}_b \mathbf{G}^0 \mathbf{i}_a + \mathbf{i}_a \mathbf{G}^0 \mathbf{i}_b + \mathbf{G}^0 \mathbf{i}_a \mathbf{i}_b\right)\right) = 2(1 + G_{a,a}^{0(-1)} G_{a,a}^0).$$

Finally,

$$\left.\frac{\partial^2 \overline{R}(\mathbf{d})}{\partial \mathbf{d} \partial \mathbf{d}^T}\right|_{\bar{\mathbf{d}}} = \kappa_{j_1-j_0} 2 \log(2)^2 \left(\mathbf{G}^{0-1} \circ \mathbf{G}^0 + \mathbf{I}_p\right) + o_{\mathbb{P}}(1). \tag{E10}$$

The matrix $\mathbf{G}^{0-1} \circ \mathbf{G}^0$ is positive definite using Schur product theorem (Proposition 13), and hence, the matrix $\mathbf{G}^{0-1} \circ \mathbf{G}^0 + \mathbf{I}_p$ is invertible.

E.2.6. End of the Proof

The Taylor expansion (E2) together with (E10) implies

$$\begin{aligned} 2\log(2)^2 \kappa_{j_1-j_0} \left(\hat{\mathbf{d}} - \mathbf{d}^0\right) &= -\left(\frac{1}{2\log(2)^2 \kappa_{j_1-j_0}} \left.\frac{\partial^2 \overline{R}(\mathbf{d})}{\partial \mathbf{d} \partial \mathbf{d}^T}\right|_{\bar{\mathbf{d}}}\right)^{-1} \left.\frac{\partial \overline{R}(\mathbf{d})}{\partial \mathbf{d}}\right|_{\mathbf{d}^0} \\ &= \left(\mathbf{G}^{0-1} \circ \mathbf{G}^0 + \mathbf{I}_p\right)^{-1} \left.\frac{\partial \overline{R}(\mathbf{d})}{\partial \mathbf{d}}\right|_{\mathbf{d}^0} (1 + o_{\mathbb{P}}(1)) \end{aligned}$$

DOI: 10.1111/jtsa.12170

We now study the convergence of $\left.\frac{\partial \overline{R}(\mathbf{d})}{\partial \mathbf{d}}\right|_{\mathbf{d}^0}$. Using equations (E3) and (E7), we have the equation

$$\left.\frac{\partial \overline{R}(\mathbf{d})}{\partial d_a}\right|_{\mathbf{d}^0} = 2\sum_{b=1}^{p} \overline{\boldsymbol{G}}^{(-1)}_{a,b}\left(\mathbf{d}^0\right) C^{(1)}_{a,b}.$$

So the asymptotic behaviour of the first derivative of the criterion is

$$\left.\frac{\partial \overline{R}(\mathbf{d})}{\partial d_a}\right|_{\mathbf{d}^0} = O_{\mathbb{P}}\left(2^{-j_0\beta} + N^{-\frac{1}{2}}2^{\frac{j_0}{2}}\right).$$

Plugging this result into the earlier expression, it becomes

$$\kappa_{j_1-j_0}\left(2^{-j_0\beta} + N^{-1/2}2^{j_0/2}\right)^{-1}\left(\hat{\mathbf{d}} - \mathbf{d}^0\right) = O_{\mathbb{P}}(1).$$

Since $\kappa_\ell > 0$ for $\ell \geq 1$ and $\kappa_\ell \to 2$ when $\ell \to \infty$, the sequence $\kappa_{j_1-j_0}$ is bounded below by a positive constant. The rate of convergence for $\hat{\mathbf{d}} - \mathbf{d}^0$ in Theorem 6 follows.

E.3. Convergence of $\hat{\mathbf{G}}(\hat{\mathbf{d}})$ and of $\hat{\boldsymbol{\Omega}}$

Recall $\hat{\mathbf{G}}_{\ell,m}(\hat{\mathbf{d}}) = 2^{<\mathcal{J}>(\hat{d}_\ell - d^0_\ell + \hat{d}_m - d^0_m)}\bar{G}_{\ell,m}(\hat{\mathbf{d}})$. Equation (E6) with the rate obtained for the convergence of $\hat{\mathbf{d}} - \mathbf{d}^0$ states that $\bar{G}_{\ell,m}(\hat{\mathbf{d}}) = G^0_{\ell,m}(1 + O_{\mathbb{P}}(\log(N)(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})) + O_{\mathbb{P}}(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})$ under assumptions of Theorem 6. The rate of convergence of $\hat{G}_{\ell,m}(\hat{\mathbf{d}})$ in Theorem 6 is then derived from the fact that $2^{<\mathcal{J}>u} - 1 = j_0 u \log(2)(1 + o(1))$ when $u \to 0$.

The convergence of $\hat{\boldsymbol{\Omega}}$ is straightforward, thanks to the fact that $K(\cdot)$ is a continuous function of $\mathbf{d}$. To obtain the rate of convergence, we observe first that $\cos(u) = 1 + o(u^2)$ when u goes to 0. Second,

$$\begin{aligned} K\left(\hat{d}_\ell + \hat{d}_m\right) - K\left(d^0_\ell + d^0_m\right) &= \int_{-\infty}^{\infty}\left(|\lambda|^{-\hat{d}_\ell - \hat{d}_m} - |\lambda|^{-d^0_\ell - d^0_m}\right)|\hat{\psi}(\lambda)|^2 \mathrm{d}\lambda \\ &\leq \left|\hat{d}_\ell + \hat{d}_m - d^0_\ell - d^0_m\right| \int_{-\infty}^{\infty}\left|\log|\lambda||\lambda|^{-d^0_\ell - d^0_m}\left|\hat{\psi}(\lambda)\right|\right|^2 \mathrm{d}\lambda. \end{aligned}$$

Using assumptions (W2) and (W5),

$$|K(\hat{d}_\ell + \hat{d}_m) - K(d^0_\ell + d^0_m)| \leq |\hat{d}_\ell + \hat{d}_m - d^0_\ell - d^0_m|\, C\int_{-\infty}^{\infty} |\log|\lambda||\, |\lambda|^{-(1+\beta)} d\lambda$$

with C positive constant. The integral on the right-hand side is finite and thus $K(\hat{d}_\ell + \hat{d}_m) - K(d^0_\ell + d^0_m) = O_{\mathbb{P}}(\max_{i=1,\dots,p}|\hat{d}_i - d^0_i|)$. When $\max_{i=1,\dots,p}|\hat{d}_i - d^0_i| = O_{\mathbb{P}}\left(2^{-j_0\beta} + N^{-1/2}2^{j_0/2}\right)$, we have $1/(K(\hat{d}_\ell + \hat{d}_m)cos(\frac{\pi}{2}(\hat{d}_\ell - \hat{d}_m))) = (1 + O_{\mathbb{P}}\left(2^{-j_0\beta} + N^{-1/2}2^{j_0/2}\right))$, which concludes the proof.

E.4. Additional tools

These results correspond to Lemma 13 of Moulines *et al.* (2008)

Define the sequences η_L and κ_L for any $L \geq 0$ by

$$\eta_L := \sum_{i=0}^{L} i\,\frac{2^{-i}}{2 - 2^{-L}} \tag{E11}$$

$$\kappa_L := \sum_{i=0}^{L} (i - \eta_L)^2 \frac{2^{-i}}{2 - 2^{-L}} \tag{E12}$$

It is straightforward that

$$n \sim N 2^{-j_0} \left(2 - 2^{-(j_1 - j_0)}\right) \tag{E13}$$

$$< \mathcal{J} > \sim j_0 + \eta_{j_1 - j_0} \tag{E14}$$

$$\frac{1}{n} \sum_{j=j_0}^{j_1} (j - < \mathcal{J} >)^2 n_j \sim \kappa_{j_1 - j_0} \tag{E15}$$

For every $L \geq 1$ the quantities η_L and κ_L are strictly positive. When L goes to infinity, the sequences η_L and κ_L respectively converge to 1 and 2.

And for all $u \geq 0$,

$$\frac{1}{\kappa_L} \sum_{i=0}^{L-u} \frac{2^{-i}}{2 - 2^{-L}} (i - \eta_L)(i + u - \eta_L) \to 1 \text{ when } L \to \infty \tag{E16}$$

DOI: 10.1111/jtsa.12170