



# Multivariate wavelet Whittle estimation in long-range dependence

Sophie Achard, Irène Gannaz

## ► To cite this version:

Sophie Achard, Irène Gannaz. Multivariate wavelet Whittle estimation in long-range dependence. 2014. hal-01079645v1

**HAL Id: hal-01079645**

**<https://hal.science/hal-01079645v1>**

Preprint submitted on 3 Nov 2014 (v1), last revised 30 Oct 2015 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multivariate wavelet Whittle estimation in long-range dependence.

Achard\*, S. and Gannaz<sup>†</sup>, I.

November 3, 2014

## Abstract

This paper deals with the 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. 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. A simulation study confirms the satisfying behaviour of the estimation, which improves the univariate estimation and gives similar results than multivariate Fourier-based procedure. For real data applications, the correlation between time series is an important feature. Usual estimations can be highly biased due to phase-shifts caused by the differences in the properties of autocorrelation in the processes. The long-run covariance matrix provides an interesting estimator for characterizing coupling between time series, also called functional connectivity in neuroscience. A real data application in neuroscience highlights the utility of the wavelets-based method, which is more flexible than Fourier-based procedures. Time series measuring the brain activity are analysed, so as to obtain the characterization of their long-memory behaviour and a measure of the functional connectivity of the brain.

**Keywords:** multivariate processes, long memory, fractional integration, semiparametric estimation, covariance matrix, wavelets, neuroscience application, functional connectivity

**MSC classification:** 60G22, 62M10, 62M15, 62H20, 92C55

## 1 Introduction

The long-range dependence has attracted lots of interest in statistics and in many applications since the seminal paper of Mandelbrot in 1950. First the Brownian motion model was introduced as the unique Gaussian process having stationary increments (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 get more complex modellings that can be closer to real data. One can cite ARIMA, FD, FIN... We refer to Percival and Walden (2006) and references therein for an overview of long-range dependence models in literature. These models were used in a large scope of applications, for example to finance (Gençay et al, 2001) (see also the references in Nielsen and

---

\*CNRS, GIPSA-lab, INSERM U836, Univ. Grenoble Alpes, France

<sup>†</sup>Université de Lyon, CNRS UMR 5208, INSA de Lyon, Institut Camille Jordan, France

Frederiksen (2005)), to internet traffic (Abry and Veitch, 1998), to physical sciences (Papanicolaou and Sølna, 2003; Percival and Walden, 2006), to geosciences (Robinson, 1995a) and to neuroimaging (Maxim et al, 2005).

Nowadays, it is common to record data having multiple sensors, such as neuroimaging with fMRI or EEG. 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 part of the earth, like in Whitcher and Jensen (2000). For these two applications, it has been shown that the univariate time series present long-range dependence behaviour. As a consequence, researchers got interest in studying multivariate models of long-range dependence time series. Several models with the possibility to tackle long memory features have been proposed. In Didier et al (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 identified (Coeurjolly et al, 2013). Also, the multivariate extension of fractionally difference models was proposed in Chambers (1995), which includes the multivariate extension of ARFIMA models with an explicit expression of the short memory terms. In a recent paper, Kechagias and Pipiras (2014) highlighted the difficulties to extend the notion of long range dependence to multivariate time series and proposed specific linear representations of long range dependence. Concerning multivariate extension of ARFIMA models, Sela and Hurvich (2008) studied two different classes of extension depending on the order of fractional integration and ARMA models.

Using these long memory models, a typical statistical issue is to estimate the long memory parameter, since it characterizes the long-term dependence of the series. Another reason is the fact that many relevant statistics behaviours depend on this parameter. A very large literature exists in the context of univariate time series. First, parametric approaches were developed (Dahlhaus, 1989; Fox and Taqqu, 1986; 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, 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 developed 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) studied the asymptotic properties of a wavelet Whittle estimator.

Considering multivariate fractionally integrated processes (Chambers, 1995), the estimation of memory parameters and covariance matrix have been first studied in 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 Abadir et al (2007)'s 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); Wang and Wang

(2013) propose estimation schemes based on multidimensional wavelet basis. 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) study a two-dimensional estimation, based on univariate wavelet basis, which define the estimators through the regression of the cross-covariance between the wavelet coefficients. This approach also appears difficult to generalize to any 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 in Robinson (1994b); 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 proposed methodology using wavelets is shown to be adequate for nonstationary long range dependence models.

The paper is organised as follows. Section 2 is introducing the specific framework of multivariate long memory processes based on the definition of the spectral density matrix. The multivariate wavelet Whittle estimators of the long memory parameters jointly to 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 are established. Finally, simulated results in Section 5 show that the wavelet Whittle estimators have comparable performances as the Fourier-based ones in addition to provide a very flexible approach to handle both stationary and nonstationary processes.

## 2 The semiparametric multivariate long-memory framework

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

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

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  $X_\ell$  and  $X_m$  can be written:

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

As it will be explained in Section 2.1, this model does not deal with cointegrated time series, where one would have an additional multiplicative term of the form  $e^{i\varphi}$ , see e.g. Kechagias and Pipiras (2014); Sela and Hurvich (2012).

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

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

where  $\mathbf{d} = \mathbf{D} + \mathbf{d}^s$ . The exponent  $*$  is the conjugate operator and  $\circ$  denotes the Hadamard product.

In this semiparametric framework, the spectral density  $f^S(\cdot)$  corresponds to the short-memory behaviour and the matrix  $\mathbf{\Omega}$  is called *fractal connectivity* by Achard et al (2008) or *long-run covariance* matrix by Robinson (2005). Similarly to Moulines et al (2007) we assume that  $f^S(\cdot) \in \mathcal{H}(\beta, L)$  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 for all  $\lambda \in (-\pi, \pi)$ ,  $\|\mathbf{g}(\lambda) - \mathbf{g}(0)\|_\infty \leq L\|\mathbf{g}(0)\|_\infty |\lambda|^\beta$ .

The spectral density specifies that the two processes  $X_\ell$  and  $X_m$  have long-memory parameters respectively  $d_\ell$  and  $d_m$ . The parameters are not necessarily taken inside  $(-1/2, 1/2)$  (if  $D \geq 1$ ) allowing nonstationary time series. If orders are different, the estimation of the memory parameters is still available but some bias issues occur for the estimation of the underlying correlation  $\mathbf{\Omega}$ . This will be explained later in the manuscript in Section 3.

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

$$f(\lambda) \sim \mathbf{\Omega} \circ (\tilde{\mathbf{\Lambda}}(\mathbf{d}) f^S(0) \tilde{\mathbf{\Lambda}}(\mathbf{d})^*), \quad \lambda \in [-\pi, \pi], \quad \text{with } \tilde{\mathbf{\Lambda}}(\mathbf{d}) = \text{diag}(|\lambda|^{-\mathbf{d}} e^{-i\pi \mathbf{d}/2}). \quad (2)$$

A similar approximation has been carried out in Lobato (1997) or Phillips et al (2004), while Shimotsu (2007) derived a second order approximation. Lobato (1999) used  $\tilde{\mathbf{\Lambda}}(\mathbf{d}) = \text{diag}(\lambda^{-\mathbf{d}})$  as an approximation of  $f(\cdot)$ . Whereas Shimotsu (2007) chose to approximate  $f(\cdot)$  using  $\tilde{\mathbf{\Lambda}}(\mathbf{d}) = \text{diag}(\lambda^{-\mathbf{d}} e^{-i(\pi-\lambda)\mathbf{d}/2})$ , which corresponds to a second order approximation due to the remaining term  $\lambda$  in the exponential. As mentioned by 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  $\tilde{\mathbf{\Lambda}}(\mathbf{d})$ . The approach proposed in the sequel of the paper is using approximation (2), which is a trade-off between Lobato (1999) and Shimotsu (2007). The resulting estimator for  $\mathbf{d}$  is equivalent to the one defined in 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 comparing to the case studied in this paper.

## 2.1 Examples of processes

Since the papers of Robinson (1994a,b, 1995a,b) the semiparametric approach has known a growing interest. The key point is that any misspecification of the short-range dependence properties of the process may produce inconsistent estimates, while such issues are excluded thanks to the semiparametric formulation. This section provides some examples of processes which satisfy our semiparametric modelling.

The matrix  $\Omega$  has been defined *via* the spectral representation of the process, the link between  $\Omega$  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  $n^{1/2} \bar{\mathbf{X}} \rightarrow 2\pi f(0)$ . When the cross-spectral density satisfies an approximation (1), Robinson (2005) indicates that

$$\mathbf{D}_n \mathbb{E}(\bar{\mathbf{X}} \bar{\mathbf{X}}^T) \mathbf{D}_n \rightarrow 2\pi \Omega \circ \mathbf{Q}(\mathbf{d})$$

where  $\mathbf{D}_n = \text{diag}(n^{1/2-\mathbf{d}})$  and  $Q_{\ell,m}(\mathbf{d}) = \frac{\sin(\pi d_\ell) + \sin(\pi d_m)}{\Gamma(d_\ell + d_m + 2) \sin(\pi(d_\ell + d_m))}$ .

### 2.1.1 Fractionally integrated processes

This section presents some multivariate fractionally integrated processes that satisfy the spectral density approximation (2). Let  $\mathbb{L}$  be the lag-operator and define the  $p$ -vector process  $\mathbf{X}$  by

$$\begin{pmatrix} (1 - \mathbb{L})^{d_1} & & 0 \\ & \ddots & \\ 0 & & (1 - \mathbb{L})^{d_p} \end{pmatrix} \begin{pmatrix} X_1(t) \\ \vdots \\ X_p(t) \end{pmatrix} = \begin{pmatrix} u_1(t) \\ \vdots \\ u_p(t) \end{pmatrix} \quad (3)$$

where  $\mathbf{u}$  is a  $p$ -vector covariance stationary process. In the particular case where  $\mathbf{u}$  is a vector ARMA process,  $\mathbf{X}$  is a multivariate ARFIMA process. We refer to Bollerslev and Wright (2000); Brunetti and Gilbert (2000); Henry and Zaffaroni (2003) for an overview of fractionally integrated processes and applications in finance.

Suppose  $\mathbf{u}$  admits a spectral density  $f_u(\cdot)$  bounded above and away from zero at the zero frequency. Then the spectral density of the process  $\mathbf{X}$  can be written

$$f(\lambda) = \text{diag}((1 - e^{-i\lambda})^{-d_\ell}, \ell = 1, \dots, p) f_u(\lambda) \text{diag}((1 - e^{-i\lambda})^{-d_\ell}, \ell = 1, \dots, p)^*,$$

see Hannan (1970), page 61.

Assume now that at the neighbourhood of the zero frequency the approximation  $f_u(\cdot) \sim \Omega \circ f^S(\cdot)$  holds. This assumption is satisfied whenever each component of the process  $\mathbf{u}$  admits a Cramer-Wold representation: for all  $\ell = 1, \dots, p$ , for any  $t \in \mathbb{Z}$ ,  $u_\ell(t) = C_\ell(\mathbb{L})\varepsilon_\ell(t)$  with a polynomial  $C_\ell(\cdot)$  finite and full rank. Such an expression is available for example if for all  $\ell = 1, \dots, p$  the time series  $\mathbf{u}_\ell$  is an ARMA process with possibly correlated innovations, *i.e.* the correlation matrix of  $\varepsilon_\ell(t)$  is independent of  $t \in \mathbb{Z}$  but possibly not equal to the unity matrix. The spectral density of the process  $\mathbf{X}$  satisfies approximation (2). The matrix  $\Omega$  is related in this model to the correlation between the initial processes  $\varepsilon$ .

A more detailed presentation of multivariate ARFIMA processes can be found in Didier et al (2011); Lobato (1997); Sela (2010); Sela and Hurvich (2012). As it can be seen in these references, multivariate ARFIMA processes are not easy to characterize since it is not equivalent to apply fractional differencing on a vectorial ARMA process or to define a vectorial ARMA process with fractionally integrated noise. In this paper, the model specification does not take into account the possibility of cointegration, namely we cannot consider  $\mathbf{u}(t) = \mathbf{C}(\mathbb{L})\varepsilon(t)$  with  $\mathbf{C}(\cdot)$  a polynomial

defined on  $\mathbb{R}^p$ . Indeed in such configurations the approximation (2) can no longer be written with a diagonal matrix  $\tilde{\Lambda}(\mathbf{d})$ . Cointegration in vectorial ARFIMA processes is difficult to handle with and may concern future research. Some insight on estimation in presence of cointegration can be found in Kechagias and Pipiras (2014); Robinson (2008); Sela and Hurvich (2012).

### 2.1.2 Discrete fractional Brownian motion

Another example is a discrete-time version of fractional Brownian motion. For a given  $\mathbf{H} \in (0, 1)^p$ , let  $(\mathbf{B}_H(k))_{k \in \mathbb{Z}}$  be a sequence of observations of a  $p$ -multivariate Brownian motion (mFBM)  $(\mathbf{B}_H(t))_{t \in \mathbb{R}}$ . Recall that a  $p$ -dimensional mFBM is a process satisfying the three following properties:

- $\mathbf{B}_H(t)$  is Gaussian for any  $t \in \mathbb{R}$ ;
- $\mathbf{B}_H$  is self-similar with parameter  $\mathbf{H}$ , i.e. for every  $t \in \mathbb{R}$  and  $a > 0$ ,  $(\mathbf{B}_{H1}(at), \dots, \mathbf{B}_{Hp}(at))$  has the same distribution as  $(a^{H_1} \mathbf{B}_{H1}(t), \dots, a^{H_p} \mathbf{B}_{Hp}(t))$ ;
- the increments are stationary.

Coeurjolly et al (2013) characterize the spectral behaviour of the increments of a mFBM: if  $f_{\ell,m}$  denotes the cross-spectral density of  $(B_{H\ell}, B_{Hm})$ , then

$$f_{\ell,m}(\lambda) = \tau_{\ell,m}(\text{sign}(\lambda)) \frac{1 - \cos(\lambda)}{|\lambda|^{H_\ell + H_m + 1}},$$

with  $\tau_{\ell,m}(\text{sign}(\lambda)) = \text{re}_{\ell,m}(\tau) - i \text{sign}(\lambda) \text{im}_{\ell,m}(\tau)$ , where  $\text{re}_{\ell,m}(\tau)$  and  $\text{im}_{\ell,m}(\tau)$  are real constants depending on  $H_\ell$ ,  $H_m$  and on the correlations between  $B_{H\ell}(1)$ ,  $B_{Hm}(1)$ ,  $B_{H\ell}(-1)$  and  $B_{Hm}(-1)$ . Actually, the cross-spectral density of the process  $(\mathbf{B}_H(k))_{k \in \mathbb{Z}}$  can be expressed on the form (1), with  $\mathbf{d} = \mathbf{H} - 1/2 \in (-1/2, 1/2)^p$ ,  $\Omega_{\ell,m} = |\tau_{\ell,m}(1)|$  and with

$$f_{\ell,m}^S(\lambda) = 2 \frac{\tau_{\ell,m}(\text{sign}(\lambda))}{|\tau_{\ell,m}(\text{sign}(\lambda))|} \frac{1 - \cos(\lambda)}{\lambda^2} \left( \frac{1 - e^{-i\lambda}}{\lambda} \right)^{H_\ell + H_m + 1}.$$

The spectral density  $f^S(\cdot)$  is bounded and, for any  $(\ell, m)$ ,  $f_{\ell,m}^S(\lambda)$  tends to 1 when  $\lambda$  tends to 0.

The matrix  $\Omega$  is related to the correlation between the processes but the link does not appear as clearly as in the previous model. To be more precise, when  $d_\ell + d_m \neq 0$ , the following expression holds:

$$\begin{aligned} \Omega_{\ell,m}^2 &= \Gamma(d_\ell + d_m)^2 [\text{Corr}(B_{H\ell}(1), B_{Hm}(1))^2 \cos^2(\pi(d_\ell + d_m)/2) \\ &\quad + (2 - 2^{d_\ell + d_m - 1})^2 (\text{Corr}(B_{H\ell}(1), B_{Hm}(-1)) - \text{Corr}(B_{H\ell}(-1), B_{Hm}(1)))^2 \sin^2(\pi(d_\ell + d_m)/2)]. \end{aligned}$$

The estimation of the matrix  $\Omega$  gives information on the cross-behaviour of the processes. In Amblard and Coeurjolly (2011) an estimation of  $\text{Cov}(B_{H\ell}(1), B_{Hm}(1))$  is proposed but it is said that  $\text{Cov}(B_{H\ell}(1), B_{Hm}(-1)) - \text{Cov}(B_{H\ell}(-1), B_{Hm}(1))$  seems difficult to identify. Using the result

of Amblard and Coeurjolly (2011) together with the estimation of  $\Omega$  and  $\mathbf{d}$  developed in this manuscript can solve this difficulty.

Clearly the semiparametric setting omits the dependence of the function  $f^S(\cdot)$  and of the matrix  $\Omega$  on the Hurst parameters vector  $\mathbf{H}$ . If the fBM is effective, a well-driven estimation under the parametric model will lead to better results than under the semiparametric one. In particular, when considering a mFBM, the between-scale behaviour of the wavelets coefficients is different from our particular setting (see Coeurjolly et al (2013)). Yet our model is more robust to misspecification of the short-run dynamics.

### 3 Multivariate Wavelet Whittle estimation

The estimation is done through a wavelet representation of the time series. 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 take into account the phase-shift.

#### 3.1 The wavelet analysis

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

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  $\psi_{j,k}(\cdot) = 2^{-j/2}\psi(2^{-j} \cdot -k)$ . The same convention as in Moulines et al (2007) and Moulines et al (2008) is used in the sequel of the paper.

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

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

For given  $j \geq 0$  and  $k \in \mathbb{Z}$ ,  $\mathbf{W}_{j,k}$  is a  $p$ -dimensional vector  $\mathbf{W}_{j,k} = (W_{j,k}(1) \ W_{j,k}(2) \ \dots \ W_{j,k}(p))$  where  $W_{j,k}(\ell) = \int_{\mathbb{R}} \tilde{\mathbf{X}}_{\ell}(t)\psi_{j,k}(t)dt$ .

The following assumptions are formulating regularity conditions on the wavelet transform, and they are supposed to be verified in the sequel of the paper:

**(W1)** The functions  $\phi(\cdot)$  and  $\psi(\cdot)$  are integrable, have compact supports,  $\int_{\mathbb{R}} \phi(t)dt = 1$  and  $\int \psi^2(t)dt = 1$ ;

**(W2)** There exists  $\alpha > 1$  such that  $\sup_{\lambda \in \mathbb{R}} |\hat{\psi}(\lambda)|(1 + |\lambda|)^{\alpha} < \infty$ , *i.e.* the wavelet is  $\alpha$ -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  $\ell$  for all  $\ell = 1, \dots, M - 1$ .

(W5) For all  $i = 1, \dots, p$ ,  $(1 + \beta)/2 - \alpha < d_i \leq M$ .

These conditions are not restrictive, and many standard wavelet basis satisfy them. Among them, Daubechies wavelets are compactly supported wavelets parametrized by the number of vanishing moments  $M$ . They are  $\alpha$ -regular with  $\alpha$  an increasing function of  $M$  going to infinity (see 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 Moulines et al (2007), Section 3.

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, see *e.g.* Velasco and Robinson (2000) and references therein. Some recent works propose a procedure that differentiate 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 used by Nielsen (2011) in multivariate analysis.

In practice, a finite number of realisation of the process  $\mathbf{X}$ , say  $\mathbf{X}(1), \dots, \mathbf{X}(N)$ , is 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  $[0, T]$  with  $T \geq 1$ . For every  $j \geq 0$ , define

$$n_j := \max(0, 2^{-j}(N - T + 1)). \quad (4)$$

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 recall some results of Moulines et al (2007) for the wavelet transform of an univariate process. Let  $W_{j,k}$  denotes the wavelet coefficient of an unidimensional process  $X$ , with a spectral density of the form  $f(\lambda) = |1 - e^{i\lambda}|^{-2d_0} f^S(\lambda)$  where  $d_0$  can be outside of the interval  $[-1/2, 1/2]$ . Moulines et al (2007) state that under assumptions (W1)-(W5), the wavelet coefficients 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. Exact decorrelation occurs when the wavelet bases is orthonormal and  $d_0 = 0$  but is not acquired in general framework. Many estimators of long-memory, see among others Abry and Veitch (1998); Gonzaga and Hauser (2011); Jensen (1999); Wornell and Oppenheim (1992), suppose the approximate decorrelation of wavelet coefficient. Following Bardet

et al (2000) or Moulines et al (2008), the present work takes into account the within and between scales behaviour.

Let  $j \geq 0$  and  $j' = j - u \leq j$  be two given scales. In 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 spectral density of this process. Theorem 1 in Moulines et al (2007) establishes that under assumptions (W1)-(W5) there exists  $C$  such that for all  $\lambda \in (-\pi, \pi)$ ,

$$\left| \mathbf{D}_{j,u}(\lambda; d_0) - f^S(0) 2^{2jd_0} \mathbf{D}_{\infty,u}(\lambda; d_0) \right| \leq C f^S(0) 2^{j(2d_0 - \beta)},$$

where

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

with  $\mathbf{e}_u(\xi) = \begin{pmatrix} 1 & e^{-i2^{-u}\xi} & \dots & e^{-i2^{-u}(2^u-1)\xi} \end{pmatrix}^T$ . The exponent  $T$  denotes the transpose operator.

The key point of our estimation is the extension of results obtained by Moulines et al (2007) to the multivariate framework. Due to 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  $\theta_{\ell,m}(j)$  denotes the wavelet covariance at scale  $j$  between processes  $X_\ell$  and  $X_m$ ,  $\theta_{\ell,m}(j) = \text{Cov}(W_{j,k}(\ell), W_{j,k}(m))$  for any position  $k$ . Using the spectral density representation,  $\theta_{\ell,m}(j)$  can be written:

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

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

**Proposition 1.** For any  $j \geq 0$ , let define the function  $K_j(\cdot)$  by

$$K_j(\delta) = \int_{-\pi}^{\pi} \lambda^{-(d_\ell + d_m)} \cos(2^{-j}\lambda(d_\ell - d_m)/2) |\hat{\psi}(2^j\lambda)|^2 d\lambda, \quad \delta \in (-\alpha, M).$$

Under assumptions (W1)-(W5), there exists a constant  $C_0$  depending on  $\beta, \min_i d_i, \max_i d_i, \phi$  and  $\psi$  such that, for all  $j \geq 0$ ,

$$|\theta_{\ell,m}(j) - \Omega_{\ell,m} f_{\ell,m}^S(0) 2^{j(d_\ell + d_m)} \cos(\pi(d_\ell - d_m)/2) K_j(d_\ell + d_m)| \leq C_0 f_{\ell,m}^S(0) L 2^{(d_\ell + d_m - \beta)j}. \quad (5)$$

The proposition above corresponds to a second order approximation of the spectral density at a neighbourhood of zero, such as the one derived in Shimotsu (2007). In the following a first order approximation will be used, given by the following proposition:

**Proposition 2.** Let define the function  $K(\cdot)$  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  $\beta, \min_i d_i, \max_i d_i, \phi$  and  $\psi$  such that, for all  $j \geq 0$ ,

$$|\theta_{\ell,m}(j) - \Omega_{\ell,m} f_{\ell,m}^S(0) 2^{j(d_\ell + d_m)} \cos(\pi(d_\ell - d_m)/2) K(d_\ell + d_m)| \leq C f_{\ell,m}^S(0) L 2^{(d_\ell + d_m - \beta)j}. \quad (6)$$

As mentioned earlier, one crucial issue in the extension to multivariate long-memory processes is the introduction of a phase-shift that has to be taken into account for the estimation of the covariance  $\Omega$ . This can be seen directly using expression (5) where  $\theta_{\ell,m}(j)$  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  $d_m = 2T + 1 + d_\ell^S$  with  $T \in \mathbb{N}$ , 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 to estimate the matrix  $\Omega$  accurately. This example corresponds to a process  $X_\ell$  covariance stationary and a process  $X_m$  such that  $\Delta X_m$  is covariance stationary, both with the same long-memory parameter  $d_\ell$ . This is illustrated in Section 5.3, Figure 1, where correlated processes are simulated, and the computed wavelet coefficients covariances are found to be close to zero. However, it is shown in this paper that even in the case where the estimation of the matrix  $\Omega$  is biased by the consequence of Proposition 2, the long-memory parameters can be estimated consistently.

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

**Proposition 3.** *For all  $u \geq 0$  and for all  $\lambda \in (-\pi, \pi)$ , let define,*

$$\tilde{D}_{u,\tau}(\lambda; \delta) = \sum_{t \in \mathbb{Z}} |\lambda + 2t\pi|^{-\delta} \hat{\psi}^*(\lambda + 2t\pi) 2^{-u/2} \hat{\psi}(2^{-u}(\lambda + 2t\pi)) e^{-i2^{-u}\tau\lambda} \quad (7)$$

and

$$K_{u,\tau}(v; \delta) = \int_{-\pi}^{\pi} \tilde{D}_{u,\tau}(\lambda; \delta) e^{i\lambda v} d\lambda. \quad (8)$$

Under assumptions (W1)-(W5), there exists a constant  $C$  depending on  $\beta, \min_i d_i, \max_i d_i, \phi$  and  $\psi$  such that, for all  $j \geq 0$ , for all  $u, v \geq 0, \tau = 0, \dots, 2^u - 1$ ,

$$\left| \mathbb{E}[W_{j,k}(\ell) W_{j-u, 2^{-u}k' + \tau}(m)] - \Omega_{\ell,m} f_{\ell,m}^S(0) 2^{j(d_\ell + d_m)} \cos(\pi(d_\ell - d_m)/2) K_{u,\tau}(k - k'; d_\ell + d_m) \right| \leq C f_{\ell,m}^S(0) L 2^{(d_\ell + d_m - \beta)j}.$$

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

### 3.3 Wavelet Whittle estimation

Let  $0 \leq j_0 \leq j_1$  be respectively the minimal and the maximal 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  $\mathcal{L}(\cdot)$ . The criterion corresponds to the negative log-likelihood of the vectors  $(W_{j,k}(\ell))_{j,k,\ell}$  when they are supposed to follow a multivariate Gaussian distribution. In addition, it has been shown by Fox and Taqqu (1986); Hannan (1973) that the Whittle approximation is giving satisfactory results for nongaussian processes. In our framework, the wavelet Whittle criterion is defined as,

$$\mathcal{L}(\mathbf{G}(\mathbf{d}), \mathbf{d}) = \frac{1}{n} \sum_{j=j_0}^{j_1} \left[ n_j \log \det (\Lambda_j(\mathbf{d}) \mathbf{G}(\mathbf{d}) \Lambda_j(\mathbf{d})) + \sum_k \mathbf{W}_{j,k}^T (\Lambda_j(\mathbf{d}) \mathbf{G}(\mathbf{d}) \Lambda_j(\mathbf{d}))^{-1} \mathbf{W}_{j,k} \right], \quad (9)$$

where  $\Lambda_j(\mathbf{d})$  and the matrix  $\mathbf{G}(\mathbf{d})$  are obtained with Proposition 2:

$$\Lambda_j(\mathbf{d}) = \text{diag} \left( 2^{j\mathbf{d}} \right)$$

and the  $(\ell, m)$ -th element of the matrix  $\mathbf{G}(\mathbf{d})$  is  $G_{\ell,m}(\mathbf{d}) = f^S(0)\Omega_{\ell,m}K(d_\ell + d_m)\cos(\pi(d_\ell - d_m)/2)$ .

For each  $j \geq 0$ , the quantity  $\sum_k \mathbf{W}_{j,k}^T (\Lambda_j(\mathbf{d})\mathbf{G}(\mathbf{d})\Lambda_j(\mathbf{d}))^{-1} \mathbf{W}_{j,k}$  has a dimension equal to 1 so it is equal to its trace. Thus,

$$\mathcal{L}(\mathbf{G}(\mathbf{d}), \mathbf{d}) = \frac{1}{n} \sum_{j=j_0}^{j_1} \left[ n_j \log \det (\Lambda_j(\mathbf{d})\mathbf{G}(\mathbf{d})\Lambda_j(\mathbf{d})) + \text{trace} \left( (\Lambda_j(\mathbf{d})\mathbf{G}(\mathbf{d})\Lambda_j(\mathbf{d}))^{-1} \mathbf{I}(j) \right) \right], \quad (10)$$

where  $\mathbf{I}(j) = \sum_{k=0}^{n_j} \mathbf{W}_{j,k} \mathbf{W}_{j,k}^T$ . This expression is very similar to the multivariate Fourier Whittle estimator of Shimotsu (2007), replacing the periodogram by the wavelet scalogram  $\mathbf{I}(j)$ .

*Remark.* In Fourier analysis, e.g. in Shimotsu (2007), the periodogram is normalized. In wavelet analysis, the normalization factor may depends on the resolution  $j$ , and the scalogram is not normalized. For every  $j$  the scalogram  $\mathbf{I}(j)$  should be normalized by  $n_j$ . It is more convenient here to keep the initial  $\mathbf{I}(j)$ .

By deriving expression (10) with respect to the matrix  $\mathbf{G}$ , it is written,

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

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

$$\hat{\mathbf{G}}(\mathbf{d}) = \frac{1}{n} \sum_{j=j_0}^{j_1} \Lambda_j(\mathbf{d})^{-1} \mathbf{I}(j) \Lambda_j(\mathbf{d})^{-1}. \quad (11)$$

When replacing  $\mathbf{G}(\mathbf{d})$  by the expression  $\hat{\mathbf{G}}(\mathbf{d})$ , the criterion  $R(\mathbf{d})$  is obtained,

$$\begin{aligned} R(\mathbf{d}) &:= \mathcal{L}(\hat{\mathbf{G}}(\mathbf{d}), \mathbf{d}) \\ &= \frac{1}{n} \sum_{j=j_0}^{j_1} n_j \log \det (\Lambda_j(\mathbf{d}) \hat{\mathbf{G}}(\mathbf{d}) \Lambda_j(\mathbf{d})) \\ &= \log \det(\hat{\mathbf{G}}(\mathbf{d})) + \frac{1}{n} \sum_{j=j_0}^{j_1} n_j \log(\det (\Lambda_j(\mathbf{d}) \Lambda_j(\mathbf{d}))) \\ &= \log \det(\hat{\mathbf{G}}(\mathbf{d})) + 2 \log(2) \left( \frac{1}{n} \sum_{j=j_0}^{j_1} j n_j \right) \left( \sum_{\ell=1}^p d_\ell \right) \end{aligned} \quad (12)$$

The vector of the long-memory parameters  $\mathbf{d}$  is estimated by  $\hat{\mathbf{d}} = \underset{\mathbf{d}}{\text{argmin}} R(\mathbf{d})$ . In a second step of estimation we define  $\hat{\mathbf{G}}(\hat{\mathbf{d}})$ , estimator of  $\mathbf{G}(\mathbf{d})$ . Finally it is needed to correct the phase-shift when estimating the covariance matrix  $\Omega$ , as given in (13),

$$\hat{\Omega}_{\ell,m} = \hat{G}_{\ell,m}(\hat{\mathbf{d}}) / (\cos(\pi(\hat{d}_\ell - \hat{d}_m)/2) K(\hat{d}_\ell + \hat{d}_m)). \quad (13)$$

Equation (13) is correctly defined as the probability that  $\hat{d}_\ell - \hat{d}_m$  is exactly congruent to 1 *modulo* 2 is null. Consequently estimator  $\hat{\Omega}$  is defined almost surely. Yet, in Section 5.3, it is shown empirically that when  $d_\ell - d_m$  is near from to 1 *modulo* 2, the estimation of  $\Omega$  can be strongly biased. In a univariate framework or when the matrix  $\Omega$  is diagonal, the estimator  $\hat{\mathbf{d}}$  is equal to the one introduced in Moulines et al (2008).

## 4 Main results

Now that the estimation scheme has been defined, this section deals with the asymptotic behaviour of the estimators. The consistence of the estimators is established, under a condition which 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)

$$\text{For all } \ell, m = 1, \dots, p, \quad \sup_n \sup_{j \geq 0} \frac{1}{n_j} \text{Var} \left( \frac{I_{\ell, m}(j)}{\theta_{\ell, m}(j)} \right) < \infty$$

This condition corresponds to a slightly more restrictive case of condition (9) of Moulines et al (2008) in an univariate framework. The difference lies in their assumption on the spectral density of the process which is only defined on a neighbourhood of zero. Similar assumption could be considered in future work.

In the multivariate setting, Condition (C) is shown to be a valid one. The following proposition gives a class of 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*

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

*with  $\boldsymbol{\varepsilon}(t)$  weak white noise process, in  $\mathbb{R}^p$ . Let  $\mathcal{F}_{t-1}$  denotes the  $\sigma$ -field of events generated by  $\{\boldsymbol{\varepsilon}(s), s \leq t-1\}$ . Assume that  $\boldsymbol{\varepsilon}$  satisfies  $\mathbb{E}[\boldsymbol{\varepsilon}(t)|\mathcal{F}_{t-1}] = 0$ ,  $\mathbb{E}[\varepsilon_a(t)\varepsilon_b(t)|\mathcal{F}_{t-1}] = \mathbb{1}_{a=b}$  and  $\mathbb{E}[\varepsilon_a(t)\varepsilon_b(t)\varepsilon_c(t)\varepsilon_d(t)|\mathcal{F}_{t-1}] = \mu_{a,b,c,d}$  with  $|\mu_{a,b,c,d}| \leq \mu_\infty < \infty$ , for all  $a, b, c, d = 1, \dots, p$ . Then, under assumptions (W1)-(W5), Condition (C) is verified.*

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 has been employed among others by Lobato (1999), Shimotsu (2007), Giraitis et al (1997), or Theorem 1 of Moulines et al (2008). As discussed in Lobato (1999), this assumption does not cover all models with density (1) but is not either very restrictive.

## 4.2 Convergence

We consider  $N$  observations of a multivariate  $p$ -vector process  $\mathbf{X}$ , namely  $\mathbf{X}(1), \dots, \mathbf{X}(N)$  with a spectral density satisfying an approximation (2) around the zero frequency. For given functions  $(\phi, \psi)$ , and for given levels  $0 \leq j_0 \leq j_1$ , the estimator of  $\mathbf{d}$  is the argument minimizing (12) and the matrix  $\mathbf{G}$  is estimated by  $\widehat{\mathbf{G}}(\widehat{\mathbf{d}})$  with  $\widehat{\mathbf{G}}(\cdot)$  defined in (11). The true parameter values are denoted by the superscript 0.

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

**Theorem 5.** *Assume (W1)-(W5) and Condition (C) hold. If in addition  $j_0$  and  $j_1$  are chosen such that  $2^{-j_0\beta} + N^{-1/2}2^{j_1/2} \rightarrow 0$  then*

$$\begin{aligned}\widehat{\mathbf{d}} - \mathbf{d}^0 &= o_{\mathbb{P}}(1), \\ \widehat{\mathbf{G}}(\widehat{\mathbf{d}}) - \mathbf{G}(\mathbf{d}^0) &= o_{\mathbb{P}}(1).\end{aligned}$$

It follows immediately that  $\widehat{\Omega}$  converges in probability to  $\Omega$ .

Moulines et al (2008) obtain a similar result in the univariate case with the assumption  $2^{-j_0} + N^{-1/2}2^{j_0/2} \rightarrow 0$ . Our hypothesis is weaker in a general framework, but since they assume (S-1)  $j_1 - j_0$  is constant, or (S-2)  $j_1 - j_0 \rightarrow \infty$ , both assumptions are equivalent.

**Theorem 6.** *Assume (W1)-(W5) and Condition (C) hold. If in addition  $j_0$  and  $j_1$  are chosen such that  $2^{-j_0\beta} + N^{-1/2}2^{j_0/2} \rightarrow 0$  then*

$$\begin{aligned}\widehat{\mathbf{d}} - \mathbf{d}^0 &= \mathcal{O}_{\mathbb{P}}(2^{-j_0\beta} + N^{-1/2}2^{j_0/2}), \\ \forall (\ell, m) \in \{1, \dots, p\}^2, \widehat{G}_{\ell, m}(\widehat{\mathbf{d}}) - G_{\ell, m}(\mathbf{d}^0) &= \mathcal{O}_{\mathbb{P}}(j_1(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})), \\ \widehat{\Omega}_{\ell, m} - \Omega_{\ell, m} &= \mathcal{O}_{\mathbb{P}}(j_1(2^{-j_0\beta} + N^{-1/2}2^{j_0/2})).\end{aligned}$$

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

$$\widehat{\mathbf{d}} - \mathbf{d}_0 = \mathcal{O}_{\mathbb{P}}(N^{-\beta/(1+2\beta)}).$$

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

Result of Corollary 7 stresses that it is necessary to fix the highest level of the wavelet procedure at a given scale depending on the regularity  $\beta$  of the density  $f^S(\cdot)$ . A possible extension in future

work is to develop an estimation which is adaptive relatively to the parameter  $\beta$ . This is done *e.g.* in univariate Fourier analysis by Iouditsky et al (2001).

Further results on asymptotic normality, and in particular the asymptotic variance of the estimators, would give an important information to quantify the quality of the estimators. In particular it would be a theoretical mean of comparison between the Fourier-based and the wavelet-based approaches or between the univariate and the multivariate estimations of  $\mathbf{d}$ . This work is in progress and will be established in a future paper. In the present paper, the comparison is done hereafter with a simulation study.

## 5 Simulations

In this section, simulated data are used to study the behaviour of the proposed procedure. Three examples of bivariate time series are studied:

- an  $ARFIMA(0, \mathbf{d}, 0)$  with a long-run correlation matrix  $\mathbf{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$  and  $\rho = 0.4$ ;
- the same as previously but with  $\rho = -0.8$ ;
- an  $ARFIMA(1, \mathbf{d}, 0)$  with no cointegration. The AR coefficient is taken equal to  $\mathbf{A} = \begin{pmatrix} 0.8 & 0 \\ 0 & 0.6 \end{pmatrix}$  and the correlation between the innovation processes equal to  $\rho = 0.4$ . More precisely let  $\varepsilon$  be a bivariate white noise process with covariance matrix  $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$  and let  $\mathbf{u}$  be the AR(1) process defined by  $\mathbf{u}(t) + \mathbf{A}\mathbf{u}(t-1) = \varepsilon(t)$ . The time series observation  $\mathbf{X}(t)$  at time  $t$  satisfies  $(1 - \mathbb{L})^d \mathbf{X}(t) = \mathbf{u}(t)$ . The matrix  $\mathbf{\Omega}$  in approximation (2) is equal to  $\begin{pmatrix} 1/1.8^2 & \rho/(1.6 \cdot 1.8) \\ \rho/(1.6 \cdot 1.8) & 1/1.6^2 \end{pmatrix}$ , see *e.g.* Sela (2010), page 34.

The proposed multivariate wavelet Whittle (MWW) estimators are computed for  $N = 512$  observations and 1000 Monte-Carlo replications. Codes are available on request.

For each of these examples, a set of different values of  $\mathbf{d}$  is considered. First the choices are restricted to settings where the two components of the processes share the same order of stationarity. This is motivated by the fact that it seems natural to suppose that time series measuring similar phenomena have similar stationary properties. On the second hand the case of different order of stationarity is studied and more precisely the problem of identifiability of the matrix  $\mathbf{\Omega}$  in such a framework.

A comparison between our estimators and multivariate Fourier Whittle (MFW) estimators developed by Shimotsu (2007) is done. The procedure MFW is only available for stationary processes. 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  $\mathbf{d} < 1.5$ . Due to this restriction we prefer to focus on the comparison with standard Fourier estimates. For nonstationary processes an usual extension of Fourier is tapered Fourier. The comparison between wavelet Whittle estimation and such an extended Fourier-based Whittle estimation, with differencing and tapering, has been done extensively in Faÿ et al (2009) for univariate time series. It shows that both approaches lead to similar results, even if the wavelet-based procedure gives globally better quality estimators. As it is expected to get the same results when working in a multivariate framework, our study focusses on the comparison with standard Fourier estimates

It is worth noticing while comparing to Fourier-based approaches 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. In real data applications, this flexibility represents a crucial asset.

## 5.1 Estimation of the long-memory parameters $\mathbf{d}$

### Parameters in estimation.

The wavelet basis used in estimation is the Daubechies wavelet with  $M = 4$  vanishing moments. Its regularization parameter is  $\alpha = 1.91$ . Note that with a lower number of vanishing moments the quality of the wavelet-based estimators would be improved. This can be seen in Faÿ et al (2009) for univariate estimation. Similar results are observed in multivariate estimation. They are not presented here for the sake of concision. The explication is due to the fact that the number of coefficients affected by bound effects, and thus removed from the estimation, increases with the length of the wavelet support, which is proportional to  $M$ . On the other side, taking a small number of vanishing moments decreases the flexibility of application due to the condition  $\sup_{\ell} d_{\ell} \leq M$ .

The two main parameters that must be chosen are the scales  $j_0$  and  $j_1$ . The finest scale  $j_1$  is fixed equal to  $\log_2(N)$  while the value of  $j_0$  giving the optimal mean square error is taken. In general increasing  $j_0$  leads to a smaller bias but a higher variance since less coefficients are used in the estimation process. As noted previously, an adaptive procedure would be interesting. An insight on the behaviour of the wavelet coefficients at different scales can be found in Achard et al (2008) and in Faÿ et al (2009). Principally, 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  $f^S(\cdot)$ .

A comparison with the quality of estimation of  $\mathbf{d}$  using univariate wavelet Whittle estimators for each components of the multivariate process has also been computed. The multivariate estimations are done on  $\mathbb{R}^p$  while each of the univariate estimations is done on the interval  $[1/2 - \alpha, M]$ .

Concerning MFW estimation, the main parameter is the number  $m$  of frequencies used in the procedure. An usual choice in literature is  $m = N^{0.65}$  (see *e.g.* Shimotsu (2007) or Nielsen and Frederiksen (2005)). Additionally MFW estimators are evaluated using values of  $m$  giving the same number of Fourier coefficients than 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 in Faÿ et al (2009).

### Measures of quality.

The quality of the estimators is measured by the bias, the standard deviation (std) and the RMSE which is equal to the square root of  $(bias^2 + std^2)$ . In order to have an easy comparison between the univariate and the multivariate approaches, we compute the ratio between the RMSE obtained with the multivariate wavelet Whittle estimation and the RMSE obtained with univariate wavelet Whittle estimations. It is denoted *ratio M/U* in the following. A similar quantity is defined for the comparison with MFW estimation. Let *ratio W/S* be the ratio between the RMSE respectively using wavelet-based estimators and Fourier-based estimators.

### Multivariate wavelet Whittle estimation.

Results for the estimation of  $\mathbf{d}$  within the three examples are presented in Tables 1, 3, and 5. The ratio M/U points out that the quality of estimation is increased with the multivariate approach, with respect to the univariate procedure. This is an illustration of the fact that when the series are correlated, it is worth using this correlation to estimate the long-memory parameters.

The estimation is still satisfactory in nonstationary settings but is slightly deteriorated with respect to stationary frameworks. The flexibility of the wavelet-based procedure is noticeable since there is no need of modification of the estimators to apply them. The only adjustment is done through the parameter  $j_0$ , but the calibration of this scale parameter is not specific to the transition from stationary to nonstationary settings. As noticed above, an equivalent Fourier-based procedure would be based on tapered Fourier. Faÿ et al (2009) compared the performance of tapered Fourier-based and wavelet-based Whittle estimators for univariate processes and established on simulations the good quality of wavelet-based estimators. A similar result for multivariate processes can be expected.

### Comparison with Fourier Whittle estimation.

Tables 2, 4, and 6 display the results of the MFW estimators described in Shimotsu (2007). In the first two examples, with the usual number of frequencies  $m = N^{0.65}$  in Fourier-based estimation, our wavelet-based procedure leads to lower RMSEs, as it can be seen with 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.

In the third example, the  $ARFIMA(1, \mathbf{d}, 0)$ , two values of  $j_0$ ,  $j_0 = 2$  or  $j_0 = 3$ , lead to similar RMSE. Table 5 presents the results for  $j_0 = 2$  due to the fact that the quality of the estimation of  $\mathbf{\Omega}$  is better. With  $j_0 = 3$ , the bias of our wavelet-based procedure are similar to the values observed with Fourier-based estimators. With our choice of the parameter  $j_0$ , the variance is decreased but the bias is higher. The quality with the MWW estimation is lower than the MFW estimation with the usual number of frequencies or with an optimized choice. The ratio are lower than 2.2 stressing that nevertheless the order of magnitude are the same.

[Tables 1 to 6]

## 5.2 Estimation of the long-run covariance $\Omega$

This section deals with the estimation of the long-run covariance matrix  $\Omega$ . It is also of interest to look at the quality of estimation of the correlation  $\Omega_{12}/\sqrt{\Omega_{11}\Omega_{22}}$ . This 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 parameters of the procedures ( $j_0$  for MWW and  $m$  for MFW) are taken equal to the values of the previous part. They are thus optimized for the estimation of  $\mathbf{d}$  and not for  $\Omega$ .

### Multivariate wavelet Whittle estimation.

The indicators of the performance of the MWW estimation of the covariance and of the correlation are given respectively in Tables 7, 9 and 11 for the three examples. The quality is satisfactory in all settings, especially in the stationary ones. Actually the RMSE is less than 0.1 for the estimation of the correlation in nearly all simulated frameworks.

### Comparison with Fourier Whittle estimation.

The results for MFW estimation are displayed in Tables 8, 10, and 12. When MFW is applied with  $m = N^{0.65}$  frequencies, the ratio W/F is less than 1 in the  $ARFIMA(0, \mathbf{d}, 0)$  examples. Like for the estimation of  $\mathbf{d}$  the good performance of MWW estimators is principally due to a smaller variance. Recall that the number of wavelet coefficients is much higher than the one considered in MFW procedure in this case, and thus this observation is coherent. In general, the correlation is also estimated with a lower RMSE using the MWW procedure. With  $ARFIMA(0, \mathbf{d}, 0)$  processes, when MFW estimators are implemented with a higher number of frequencies, giving optimal results for the estimation of  $\mathbf{d}$ , the difference between MWW and MFW procedures decreases. The quality of the two estimation schemes are similar, with comparable values for bias and variances.

Concerning the estimation of the long-run covariance matrix  $\Omega$  with  $ARFIMA(1, \mathbf{d}, 0)$  processes, the ratio between the RMSEs of the MWW estimators and of the MFW estimators is always greater than 1, for the optimal as for the usual number of frequencies in the Fourier-based procedure. Yet, with the usual number of frequencies  $m = N^{0.65}$  in MFW, the estimation of the correlation has a higher quality with the wavelet-based estimation. Still, even if Fourier estimators give better results, the quality of estimation with our wavelet-based procedure is satisfactory; the values for RMSEs are very similar comparing Fourier and wavelets multivariate Whittle estimation.

[Tables 7 to 12]

The simulation study highlights that the multivariate Wavelet estimation have satisfactory results. First the multivariate approach increases the quality of estimation of the long-memory parameters  $\mathbf{d}$  in comparison with an univariate estimation. Second, in stationary frameworks, the performance

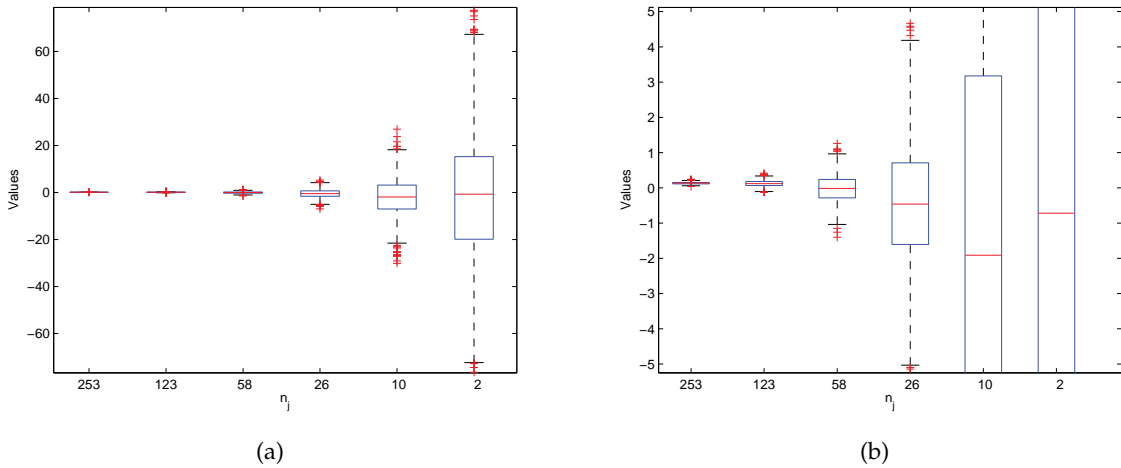
is very similar to multivariate Fourier Whittle estimation, when estimating the vector  $\mathbf{d}$  or the long-run covariance matrix or the fractal connectivity. The main advantage of our wavelet-based procedure is then its flexibility. By contrast with Fourier-based estimation, our estimators can be applied in a larger scope of situations, with nonstationary processes or in the presence of polynomial trends in the time series.

### 5.3 Discussion on bias correction

The simulation study above considers time series with the same order of stationarity. In practical applications, this assumption seems natural. However, as remarked previously, when two time series have long-memory parameters  $d_\ell$  and  $d_m$  satisfying  $d_\ell - d_m = 1$  the long-run covariance matrix  $\Omega$  is no longer identifiable. Indeed, Proposition 2 states that the covariance  $\theta_{\ell,m}(j)$  tends to 0 when the scale  $j$  tends to infinity. Figure 1 illustrates this approximation for a bivariate  $ARFIMA(0, (0.2, 1.2), 0)$ .

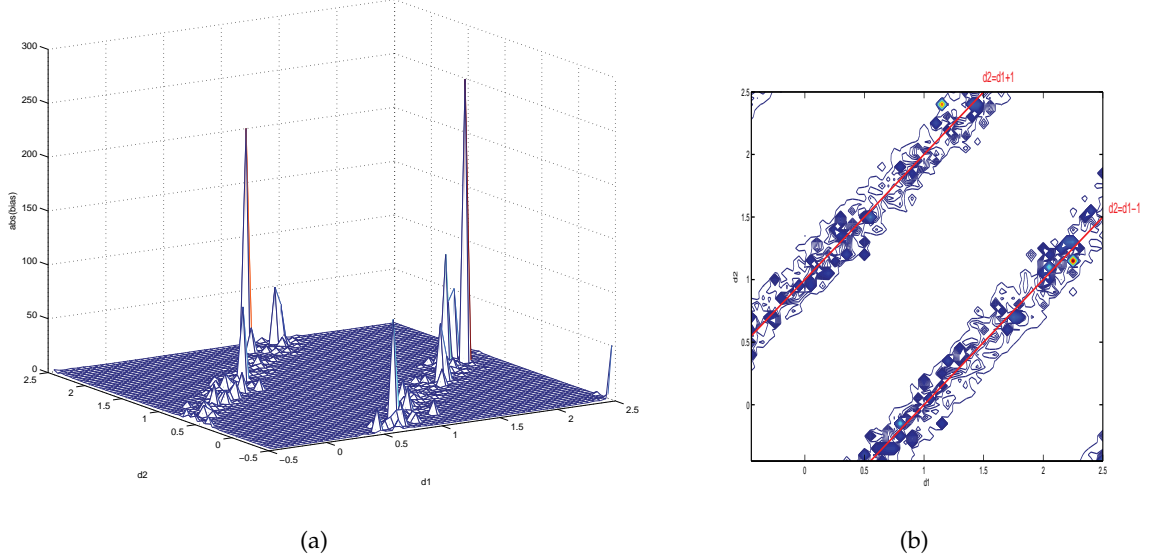
When  $\hat{d}_\ell - \hat{d}_m = 1$ , the estimator (13) is no-longer defined. In practice, the quantity  $\hat{d}_\ell - \hat{d}_m$  cannot be exactly equal to 1. Nevertheless, as dividing by a cosinus of this difference, a small error in the estimation of  $(d_\ell, d_m)$  will lead to an important bias in the estimation of  $\Omega_{\ell,m}$ . As it can be seen in Figure 2, the resulting bias increases in the neighbourhood of the non-identifiable lines  $d_\ell - d_m = \pm 1$ .

Figure 1: Boxplots of the covariance of the wavelet coefficients at different scales for a bivariate  $ARFIMA(0, (d_1, d_2), 0)$ . The index of the horizontal axis displays the number of coefficients available. Subfigures (a) and (b) use different limits on the vertical axis. Calculation was done on  $N = 512$  observations and among 1000 replications.



When this situation occurs, say when the difference between  $d_\ell - d_m$  is between 0.75 and 1.25, the estimation of  $\mathbf{d}$  is not affected. But the user must take care about the estimation of  $\Omega$ . One solution is to differentiate or integrate one of the two processes. For example, Table 13 illustrates the non-identifiability of  $\Omega$  in a bivariate  $ARFIMA(0, (0.2 \ 1.2), 0)$ . When differentiating the second component (with  $d_2 = 1.2$ ) the performance of the estimator is no more affected.

Figure 2: Absolute bias in the estimation of the cross-covariance term  $\Omega_{12}$  with respect to  $(d_1, d_2)$ . Estimation was done using multivariate Wavelet Whittle estimator in a bivariate  $ARFIMA(0, (d_1, d_2), 0)$ . Subfigure (b) represents a contour plot of subfigure (a). Calculation was done on  $N = 512$  observations and among 1000 replications.



[Table 13]

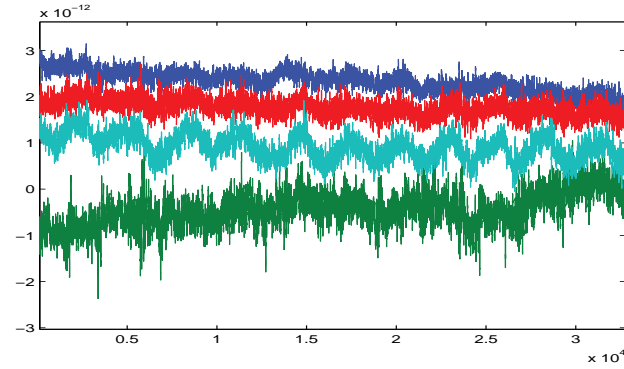
## 6 Application on neuroscience data

So as to illustrate the utility of the method for real data, we apply our approach to neuroscience data. The brain connectivity has attracted lots of interest in the neuroscience community. It is evaluated generally using correlations at different frequencies between time series measuring the brain activity. As shown previously, the presence of long-memory dependence affects the estimation of correlation, through the presence of a phase-shift. It is thus important to correct the estimation.

The study concerns 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 in Achard et al (2008). We consider  $N = 2^{15}$  time points for each of the 274 time series.

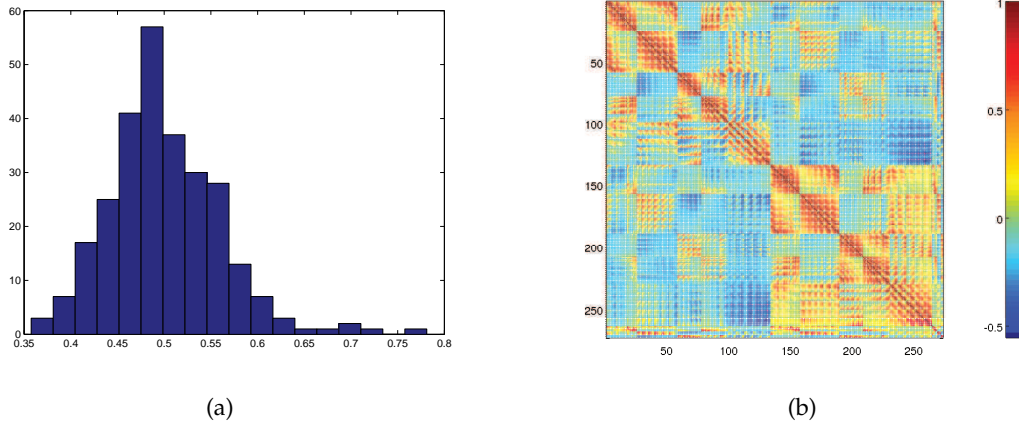
Figure 3 displays the representation of the time series for arbitrary four channels. It is clear that they present nonlinear trends. Consequently, Fourier methods are not adequate to analyse such data. It stresses the advantage of wavelets in real applications.

Figure 3: MEG recordings for 4 arbitrary channels.



Our procedure was applied using scales 4 to 8. It corresponds to frequencies between 1 to 20 Hz. This choice was motivated by discussions with neuroscientists: it takes into account first the presence of high-frequency noise which is modelled by  $f^S(\cdot)$  in our model and second the preprocessing of the data at low-frequencies. Figure 4 presents the results of the estimation of the long-memory parameters  $\mathbf{d}$  and of the long-run covariance matrix  $\mathbf{\Omega}$ .

Figure 4: Results obtained by multivariate wavelet Whittle estimators on the MEG dataset: histogram of the estimated long-memory parameters  $\mathbf{d}$  (a) and estimated fractal connectivity matrix (b).



On the first hand, 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. The problem of identifiability does not occur with the data. Similarly the closeness of long-memory parameters was observed in fMRI data by Maxim et al (2005). 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 others. It would be interesting to relate this result to a neuroscience interpretation. This will be investigated in future work.

## 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 consider a semiparametric multivariate model, including a large class of multivariate processes such as the multivariate Brownian motion and 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 the satisfying behaviour of the procedure on finite samples. Finally we propose an application to the estimation of a human brain functional network based on MEG data sets. 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.** This work was partly supported by the IXXI research institute. This work has been done 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](http://shimotsu.web.fc2.com/Site/Matlab_Codes.html)) and to Fay, Moulines, Roueff and Taqqu for kindly providing the codes of their respective papers.

## A Proof of Propositions 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:  $\theta_{\ell,m}(j) = \int_{\mathbb{R}} \operatorname{Re}(f_{\ell,m}(\lambda)) |\mathbb{H}_j(\lambda)|^2 d\lambda$ , more precisely,

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

Indeed as the cross-spectral density is hermitian, its imaginary part is an odd function. Using again the fact that the sinus function is odd,

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

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

$$\begin{aligned} A_{\ell,m}(j) &= \Omega_{\ell,m} 2^j \cos(\pi(d_\ell - d_m)/2) \int_{-\pi}^{\pi} (2 \sin(\lambda/2))^{-(d_\ell+d_m)} \cos(\lambda(d_\ell - d_m)/2) \\ &\quad f_{\ell,m}^S(\lambda) |\hat{\phi}(\lambda) \hat{\psi}(2^j \lambda)|^2 d\lambda \\ R_{\ell,m}(j) &= \theta_{\ell,m}(j) - A_{\ell,m}(j) \end{aligned}$$

Following the proof of Moulines et al (2007), rewrite the term  $A_{\ell,m}(j)$  as:

$$\begin{aligned} A_{\ell,m}(j) &= \Omega_{\ell,m} 2^j \cos(\pi(d_\ell - d_m)/2) \int_{-\pi}^{\pi} g_{\ell,m}(\lambda) \lambda^{-(d_\ell+d_m)} f_{\ell,m}^S(\lambda) |\hat{\phi}(\lambda)|^2 |\hat{\psi}(2^j \lambda)|^2 d\lambda \\ \text{with } g_{\ell,m}(\lambda) &= \left( \frac{2 \sin(\lambda/2)}{\lambda} \right)^{-(d_\ell+d_m)} \text{ for all } \lambda \in (-\pi, \pi). \end{aligned} \quad (14)$$

- The assumption that  $f^S(\cdot) \in \mathcal{H}(\beta, L)$  states that  $\left| \frac{f_{\ell,m}^S(\lambda)}{f_{\ell,m}^S(0)} - 1 \right| \leq L|\lambda|^\beta$  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  $L_g = \sup_{\ell,m=1,\dots,p} \sup_{\lambda \in (-\pi, \pi)} |g_{\ell,m}''(\lambda)|$ .

Using remarks above imply that there exists a constant  $C$  depending on  $\phi$  such that:

$$\begin{aligned} \left| A_{\ell,m}(j) - \Omega_{\ell,m} 2^j \cos(\pi(d_\ell - d_m)/2) f_{\ell,m}^S(0) \int_{-\pi}^{\pi} \lambda^{-(d_\ell+d_m)} \cos(\lambda(d_\ell - d_m)/2) |\hat{\psi}(2^j \lambda)|^2 d\lambda \right| \\ \leq CL 2^j \int_{-\pi}^{\pi} |\lambda|^{(\beta-d_\ell-d_m)} |\hat{\psi}(2^j \lambda)|^2 d\lambda. \end{aligned}$$

With a change of variable,

$$\begin{aligned} \left| A_{\ell,m}(j) - \Omega_{\ell,m} 2^{j(d_\ell+d_m)} \cos(\pi(d_\ell - d_m)/2) f_{\ell,m}^S(0) K_j(d_\ell + d_m) \right| \\ \leq CL 2^{j(d_\ell+d_m-\beta)} \int_{-\infty}^{\infty} |\lambda|^{(\beta-d_\ell-d_m)} |\widehat{\psi}(\lambda)|^2 d\lambda, \end{aligned}$$

$$\text{with } K_j(\delta) = \int_{-\pi}^{\pi} \lambda^{-(d_\ell+d_m)} \cos(2^{-j}\lambda(d_\ell - d_m)/2) |\widehat{\psi}(2^j\lambda)|^2 d\lambda.$$

Finally, the rest  $R_{\ell,m}(j)$  can be bounded by  $R_{\ell,m}(j) \leq CL 2^{(d_\ell+d_m-\beta)j}$ . This results from the approximation of the squared gain function of the wavelet filter given in Proposition 9 of Moulines et al (2007) and from similar arguments than those given for the term  $A_{\ell,m}(j)$ . We do not detail the proof here for the sake of concision and refer to the proof of Theorem 3 in Moulines et al (2007).

If  $g_{\ell,m}(\cdot)$  in (14) is defined by

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

a similar approximation can be obtained:

$$\begin{aligned} \left| A_{\ell,m}(j) - \Omega_{\ell,m} 2^{j(d_\ell+d_m)} \cos(\pi(d_\ell - d_m)/2) f_{\ell,m}^S(0) K(d_\ell + d_m) \right| \\ \leq CL 2^{j(d_\ell+d_m-\beta)} \int_{-\infty}^{\infty} |\lambda|^{(\beta-d_\ell-d_m)} |\widehat{\psi}(\lambda)|^2 d\lambda, \end{aligned}$$

$$\text{with } K(\delta) = \int_{-\pi}^{\pi} \lambda^{-(d_\ell+d_m)} |\widehat{\psi}(2^j\lambda)|^2 d\lambda.$$

## B Proof of Proposition 4

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

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

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

where  $\mathbf{B}_\ell(j, 2^j k - t) = \widetilde{h}_{j,2^j k - t} (\Delta^{M-D} \mathbf{A}_\ell)_t$ . For all  $\ell = 1, \dots, p$ , 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  $W_{j,k}(\ell)$ .

$$\theta_{j,k}(\ell, \ell) = \mathbb{E}[W_{j,k}(\ell)^2] = \sum_{t, t' \in \mathbb{Z}} \sum_{a, b=1, \dots, p} B_{\ell,a}(j, 2^j k - t) B_{\ell,b}(j, 2^j k - t') \mathbb{E}[\varepsilon_a(t) \varepsilon_b(t')]$$



Using the second-order properties of the process  $\varepsilon$ , the variance is equal to:

$$\theta_{j,k}(\ell, \ell) = \sum_{t \in \mathbb{Z}} \sum_{a=1, \dots, p} B_{\ell,a}(j, 2^j k - t)^2 \quad (15)$$

Let consider now  $E[I_{\ell,m}(j)^2]$ .

$$\begin{aligned} E[I_{\ell,m}(j)^2] &= \mathbb{E} \left[ \left( \sum_k W_{j,k}(\ell) W_{j,k}(m) \right)^2 \right] \\ &= \sum_{k,k'} \sum_{t,t',t'',t''' \in \mathbb{Z}} \sum_{a,b,c,d=1, \dots, p} B_{\ell,a}(j, 2^j k - t) B_{\ell,b}(j, 2^j k - t') B_{\ell,c}(j, 2^j k' - t'') B_{\ell,d}(j, 2^j k' - t''') \\ &\quad \mathbb{E}[\varepsilon_a(t) \varepsilon_b(t') \varepsilon_c(t'') \varepsilon_d(t''')] \end{aligned}$$

The fourth order behaviour of  $\varepsilon$  implies that

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

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

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

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

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

## Bounds $V_2$ and $V_3$

Proposition 3 gives

$$\sum_{k,k'} \mathbb{E}[W_{j,k}(\ell)W_{j,k'}(\ell)]\mathbb{E}[W_{j,k}(m)W_{j,k'}(m)] = \Omega_{\ell,\ell}\Omega_{m,m}f_{\ell,\ell}^S(0)f_{m,m}^S(0)2^{2j(d_\ell+d_m)}\sum_{k,k'}K_{0,0}(k'-k;2d_\ell)K_{0,0}(k'-k;2d_m).$$

The quantity  $K_{0,0}(v;2d_\ell)$  is equal to  $\int_{-\pi}^{\pi}\tilde{D}_{0,0}(\lambda;2d_\ell)e^{-iv\lambda}d\lambda$  which is the  $v$ -th Fourier coefficient of the function  $\tilde{D}_{0,0}(\cdot;2d_\ell)$ . Consequently, Parseval theorem implies that  $\sum_{v\in\mathbb{Z}}K_{0,0}(v;2d_\ell)K_{0,0}(v;2d_m)$  converges to  $I_0(2d_\ell,2d_m) = \int_{-\pi}^{\pi}\tilde{D}_{0,0}(\lambda;2d_\ell)\tilde{D}_{0,0}(\lambda;2d_m)d\lambda$ . On the second hand,  $\Omega_{\ell,\ell}f_{\ell,\ell}^S(0)2^{2jd_\ell}K(2d_\ell) \sim \theta_{\ell,\ell}(j)$  via Proposition 2. Then,

$$\frac{1}{n_j}\frac{V_2}{\theta_{\ell,m}(j)^2} \leq \frac{\theta_{\ell,\ell}(j)\theta_{m,m}(j)}{\theta_{\ell,m}(j)^2} \frac{I_0(2d_\ell,2d_m)}{K(2d_\ell)K(2d_m)}.$$

Proposition 2 states that  $\frac{\theta_{\ell,\ell}(j)\theta_{m,m}(j)}{\theta_{\ell,m}(j)^2}$  tends to a constant independent from  $j$ . As  $\frac{I_0(2d_\ell,2d_m)}{K(2d_\ell)K(2d_m)}$  is only depending on  $d_\ell$  and  $d_m$ , it follows that  $\frac{1}{n_j}\frac{V_2}{\theta_{\ell,m}(j)^2}$  is bounded. With a similar argumentation,

$$\frac{1}{n_j}\frac{V_3}{\theta_{\ell,m}(j)^2} \leq \frac{I_0(d_\ell+d_m,d_\ell+d_m)}{K(d_\ell+d_m)^2}.$$

And so  $\frac{1}{n_j}\frac{V_3}{\theta_{\ell,m}(j)^2}$  is bounded by a constant depending only on  $\mathbf{d}$ ,  $\phi$  and  $\psi$ .

## Bound $V_1$

The quantity  $V_1 = (1 + \mu_\infty)$  can be written:

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

Applying Minkowski inequality on  $V_1$ ,

$$V_1 \leq (1 + \mu_\infty) \sum_k \sum_{t \in \mathbb{Z}} \sum_{a,b=1,\dots,p} |B_{\ell,a}(j,2^jk-t)B_{m,b}(j,2^jk-t)| \left( \sum_{t' \in \{t+2^jk', k' \in \mathbb{Z}\}} \sum_{c,d=1,\dots,p} B_{\ell,c}(j,2^jk-t')^2 \right)^{1/2} \left( \sum_{t' \in \{t-2^jk', k' \in \mathbb{Z}\}} \sum_{c,d=1,\dots,p} B_{m,d}(j,2^jk-t')^2 \right)^{1/2}.$$

Hence,

$$V_1 \leq (1 + \mu_\infty) \sum_{k=0} \left( \sum_{t \in \mathbb{Z}} \sum_{a,b=1,\dots,p} B_{\ell,a}(j,2^jk-t)^2 \right)^{1/2} \left( \sum_{t \in \mathbb{Z}} \sum_{a,b=1,\dots,p} B_{m,b}(j,2^jk-t)^2 \right)^{1/2} \left( \sum_{t' \in \mathbb{Z}} \sum_{c,d=1,\dots,p} B_{\ell,c}(j,2^jk-t')^2 \right)^{1/2} \left( \sum_{t' \in \mathbb{Z}} \sum_{c,d=1,\dots,p} B_{m,d}(j,2^jk-t')^2 \right)^{1/2}.$$

Together with (15) the following inequality comes:  $V_1 \leq (1 + \mu_\infty) n_j p^2 \theta_{\ell, \ell}(j) \theta_{m, m}(j)$ . To conclude

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

Condition (C) follows, since  $\frac{\theta_{\ell, \ell}(j) \theta_{m, m}(j)}{\theta_{\ell, m}(j)^2}$  tends to a constant independent from  $j$  thanks to Proposition 2.

## C Preliminary results

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

$$S_{\ell, m}(\mu) = \sum_{j, k} \mu_j \left( \frac{W_{j, k}(\ell) W_{j, k}(m)}{G_{\ell, m}^0 2^{j(d_\ell^0 + d_m^0)}} - 1 \right) = \sum_{j=j_0}^{j_1} \mu_j \left( \frac{I_{\ell, m}(j)}{G_{\ell, m}^0 2^{j(d_\ell^0 + d_m^0)}} - n_j \right). \quad (16)$$

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

$$\sup_{\mu \in \{\mu_{j \geq 0}, |\mu_j| \leq \frac{1}{n_j}\}} \sup_{\mathbf{d} \in \mathbb{R}^p} \{S_{\ell, m}(\mu(\mathbf{d}))\} = \mathcal{O}_{\mathbb{P}}(2^{-j_0 \beta} + N^{-1/2} 2^{j_1/2}).$$

*Proof.* The term  $S_{\ell, m}(\mu)$  is decomposed in two terms  $S_{\ell, m}^{(0)}(\mu)$  and  $S_{\ell, m}^{(1)}(\mu)$ :

$$\begin{aligned} S_{\ell, m}^{(0)}(\mu) &= \sum_{j=j_0}^{j_1} \mu_j \frac{\theta_{\ell, m}(j)}{G_{\ell, m}^0 2^{j(d_\ell^0 + d_m^0)}} \sum_k \left( \frac{W_{j, k}(\ell) W_{j, k}(m)}{\theta_{\ell, m}(j)} - 1 \right), \\ S_{\ell, m}^{(1)}(\mu) &= \sum_{j=j_0}^{j_1} n_j \mu_j \left[ \frac{\theta_{\ell, m}(j)}{G_{\ell, m}^0 2^{j(d_\ell^0 + d_m^0)}} - 1 \right]. \end{aligned}$$

Following Proposition 2 which derives an approximation of the covariance of the wavelet coefficients,

$$|S_{\ell, m}^{(0)}(\mu)| \leq (1 + C) \sum_{j=j_0}^{j_1} |\mu_j| \left| \sum_k \frac{W_{j, k}(\ell) W_{j, k}(m)}{\theta_{\ell, m}(j)} - n_j \right|, \quad (17)$$

$$|S_{\ell, m}^{(1)}(\mu)| \leq C \sum_{j=j_0}^{j_1} 2^{-\beta j} n_j |\mu_j|. \quad (18)$$

Under the assumption that  $|\mu_j| \leq \frac{1}{n_j}$ , it comes the inequality:  $|S_{\ell, m}^{(1)}(\mu)| \leq C \sum_{j=j_0}^{j_1} 2^{-\beta j}$ . The right-hand bound is equivalent up to a constant to  $2^{-j_0 \beta}$ . Concerning  $S_{\ell, m}^{(0)}(\mu)$ , using Condition (C) gives the inequality:

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

As  $n_j \sim N2^{-j}$ , the right-hand side of the inequality is equivalent to  $(1 + C)N^{-1/2}2^{j_1/2}$ .

□

**Proposition 9.** Assume that the sequences  $\mu$  belong to the set

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

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

$$\sup_{\mu \in \mathcal{S}(q, \gamma, c)} \sup_{\mathbf{d} \in \mathbb{R}^p} \{S_{\ell, m}(\mu)\} = \mathcal{O}_{\mathbb{P}}(2^{-j_0\beta} + N^{-1/2}2^{j_0/2}).$$

*Proof.* Under the assumptions of the proposition, one can deduce from inequality (18) that,

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

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 (17) is rewritten as,

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

Proposition follows using the assumption  $0 < \gamma < 1/2$ .

□

## D Proof of Theorem 5

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  $\Omega$  will be detailed in a second step. The proof is based on the following inequality:

$$R(\mathbf{d}) - R(\mathbf{d}_0) \geq L(\mathbf{d} - \mathbf{d}^0) + \Delta(\mathbf{d}, \mathbf{d}^0), \quad (19)$$

where  $L$  is a deterministic and convex function of  $\mathbf{d}$  and the remaining term  $\Delta$  tends uniformly to zero in probability. The proof of Theorem 5 will follow directly from this inequality.

We first establish inequality (19). The difference between the criterion at a point  $\mathbf{d}$  and the value at the true long-memory parameters is equal to

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

where  $\langle \mathcal{J} \rangle = \frac{1}{n} \sum_{j=j_0}^{j_1} j n_j$  and  $n = \sum_{j=j_0}^{j_1} n_j$ .

The equality can be rewritten:

$$\begin{aligned} & R(\mathbf{d}) - R(\mathbf{d}^0) \\ &= \log \det \left( \frac{1}{n} \sum_{j=j_0}^{j_1} \Lambda_{\langle \mathcal{J} \rangle} (\mathbf{d} - \mathbf{d}^0) \Lambda_j (\mathbf{d} - \mathbf{d}^0)^{-1} I(j) \Lambda_j (\mathbf{d} - \mathbf{d}^0)^{-1} \Lambda_{\langle \mathcal{J} \rangle} (\mathbf{d} - \mathbf{d}^0) \right) \\ &\quad - \log \det(\widehat{\mathbf{G}}(\mathbf{d}^0)) \\ &= \log \det \left( \frac{1}{n} \sum_{j=j_0}^{j_1} \Lambda_{j-\langle \mathcal{J} \rangle} (\mathbf{d} - \mathbf{d}^0)^{-1} \Lambda_j (\mathbf{d}^0)^{-1} I(j) \Lambda_j (\mathbf{d}^0)^{-1} \Lambda_{j-\langle \mathcal{J} \rangle} (\mathbf{d} - \mathbf{d}^0)^{-1} \right) \\ &\quad - \log \det(\widehat{\mathbf{G}}(\mathbf{d}^0)). \end{aligned}$$

Let also define  $\lambda_j(\delta) = 2^{-(j-\langle \mathcal{J} \rangle)\delta}$  for any  $j \geq 0$  and  $\delta \in \mathbb{R}$ .

Recall Oppenheim's inequality:

**Proposition 10** (Oppenheim's inequality). *Let  $\mathbf{E}$  and  $\mathbf{B}$  be two semi-definite positive matrices. Then  $\det(\mathbf{E} \circ \mathbf{B}) \geq \det(\mathbf{E}) \prod_{\ell} B_{\ell,\ell}$ .*

Let  $\mathbf{A} = \frac{1}{n} \sum_{j=j_0}^{j_1} \Lambda_{j-\langle \mathcal{J} \rangle} (\mathbf{d} - \mathbf{d}^0)^{-1} \Lambda_j (\mathbf{d}^0)^{-1} I(j) \Lambda_j (\mathbf{d}^0)^{-1} \Lambda_{j-\langle \mathcal{J} \rangle} (\mathbf{d} - \mathbf{d}^0)^{-1}$  be the following matrix: Oppenheim's inequality will be applied to matrices  $\mathbf{B}$  and  $\mathbf{E}(\mathbf{d} - \mathbf{d}^0)$  where the  $(\ell, m)$ -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  $(\ell, m)$ -th element of  $\mathbf{E}(\mathbf{d} - \mathbf{d}^0)$  is equal to

$$\begin{aligned} E_{\ell,m}(\mathbf{d}) &= \sum_{j=j_0}^{j_1} \mu_{j,\ell,m}(\mathbf{d} - \mathbf{d}^0) I_{\ell,m}(j) 2^{-j(d_\ell^0 + d_m^0)} \\ \text{avec } \mu_{j,\ell,m}(\delta) &= \frac{2^{-j(\delta_\ell + \delta_m)} 2^{\langle \mathcal{J} \rangle(\delta_\ell + \delta_m)}}{\sum_{a=j_0}^{j_1} n_a 2^{-a(\delta_\ell + \delta_m)} 2^{\langle \mathcal{J} \rangle(\delta_\ell + \delta_m)}} = \frac{2^{-j(\delta_\ell + \delta_m)}}{\sum_{a=j_0}^{j_1} n_a 2^{-a(\delta_\ell + \delta_m)}}. \end{aligned}$$

- The matrix  $\mathbf{E}$  can be written on the form  $\mathbf{E} = \sum_{j,k} \widetilde{W}_{j,k} \widetilde{W}_{j,k}$ . Consequently  $\mathbf{E}$  is positive semi-definite being the sum of positive semi-definite matrices.
- The matrix  $\mathbf{B}$  satisfies  $\mathbf{B} = \sum_{j=j_0}^{j_1} \mathbf{M}_j \mathbf{M}_j$  with  $\mathbf{M}_j = \left(\frac{n_j}{n}\right)^{1/2} \Lambda_{j-\langle \mathcal{J} \rangle} (\mathbf{d} - \mathbf{d}^0)^{-1}$ . Thus  $\mathbf{B}$  is also positive semi-definite.

Oppenheim's inequality implies  $\log \det(\mathbf{A}) \geq \log \det(\mathbf{E}(\mathbf{d} - \mathbf{d}^0)) + \sum_{\ell} \log B_{\ell,\ell}$ .

Let define  $L(\mathbf{d} - \mathbf{d}^0)$  by  $L(\mathbf{d} - \mathbf{d}^0) := \sum_{\ell=1}^p \log B_{\ell,\ell}$ . As it can be seen that

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

the function  $L$  satisfies the following equality:

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

One can easily recognize that each term of the sum corresponds to the criterion defined in Proposition 6 of Moulines et al (2008).

Inequality (19) follows with  $\Delta(\mathbf{d}, \mathbf{d}^0) = \log \det(\mathbf{E}(\mathbf{d} - \mathbf{d}^0)) - \log \det(\widehat{\mathbf{G}}(\mathbf{d}^0))$ . We will now control the two terms in the right hand side inequality (19).

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

- The  $(\ell, m)$ -element of the matrix  $\mathbf{E}(\mathbf{d} - \mathbf{d}^0)$  is written as:

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

As  $\sum_{j=j_0}^{j_1} n_j \mu_{j,\ell,m}(\delta) = 1$ , the quantity  $\mathbf{E}_{\ell,m}(\mathbf{d} - \mathbf{d}^0)$  is equal to:

$$\mathbf{E}_{\ell,m}(\mathbf{d} - \mathbf{d}^0) = G_{\ell,m}^0 \left[ 1 + \sum_{j,k} \mu_{j,\ell,m}(\mathbf{d} - \mathbf{d}^0) \left( \frac{W_{j,k}(\ell) W_{j,k}(m)}{G_{\ell,m}^0 2^{j(d_\ell^0 + d_m^0)}} - 1 \right) \right]$$

where  $G_{\ell,m}^0 = f_{\ell,m}^S(0) \Omega_{\ell,m} K(d_\ell + d_m) \cos(\pi(d_\ell^0 - d_m^0)/2)$ .

Above expression can be rewritten as  $\mathbf{E}_{\ell,m}(\mathbf{d} - \mathbf{d}^0) = G_{\ell,m}^0 (1 + S_{\ell,m}(\mu_{\ell,m}(\mathbf{d} - \mathbf{d}^0)))$  with  $S_{\ell,m}(\mu)$  as defined previously in equation (16). Since  $\sup_{\mathbf{d}} |\mu_j(\mathbf{d} - \mathbf{d}^0)| \leq \frac{1}{n_j}$ , Proposition 8 states that  $\mathbf{E}_{\ell,m}(\mathbf{d} - \mathbf{d}^0) \rightarrow G_{\ell,m}^0$  uniformly in  $\mathbf{d}$  when  $2^{-j_1\beta} + N^{-1/2} 2^{j_0/2} \rightarrow 0$ .

- Finally we shall establish that  $\log \det \widehat{\mathbf{G}}(\mathbf{d}^0)$  tends uniformly to  $\log \det(\mathbf{G}^0)$  for  $\mathbf{d} \in \mathbb{R}^p$ . Recall

$$\widehat{G}_{\ell,m}(\mathbf{d}^0) = G_{\ell,m}^0 (1 + S_{\ell,m}(v)) \text{ where } v_j = \frac{1}{n}.$$

The sequence  $v$  satisfies  $|v_j| \leq 1/n_j$ . Applying Proposition 8, the convergence is proved when  $2^{-j_0\beta} + N^{-1/2} 2^{j_1/2} \rightarrow 0$ .

## Consistence of the estimation of $\mathbf{G}$ and $\Omega$

Let  $\ell, m$  be given indexes in  $\{1, \dots, p\}$ . Any  $(\ell, m)$ -th element of the matrix  $\widehat{\mathbf{G}}(\mathbf{d})$  satisfies:

$$\widehat{G}_{\ell,m}(\mathbf{d}) = G_{\ell,m}^0 2^{\langle \mathcal{J} \rangle_{(d_\ell - d_\ell^0 + d_m - d_m^0)}} \left[ \sum_{j=j_0}^{j_1} n_j v_{j,\ell,m} + S_{\ell,m}(v_{\ell,m}(\mathbf{d} - \mathbf{d}^0)) \right]$$

where  $v_{j,\ell,m}(\mathbf{d} - \mathbf{d}^0) = \frac{1}{n} 2^{-(j - \langle \mathcal{J} \rangle)(d_\ell - d_\ell^0 + d_m - d_m^0)}$ .

Recall that  $\langle \mathcal{J} \rangle \sim j_0 + \eta_{j_1 - j_0}$  with  $0 \leq \eta_{j_1 - j_0} \leq 1$ . As  $\sup_\ell |\hat{d}_\ell - d_\ell^0| = o_{\mathbb{P}}(1)$ , for any  $\gamma \in (0, 1/2)$ , there exists  $N_\gamma$  such that for any  $N \geq N_\gamma$ ,  $2^{-(j - \langle \mathcal{J} \rangle)(d_\ell - d_\ell^0 + d_m - d_m^0)} \leq 2^\gamma 2^{(j - j_0)\gamma}$ . For  $N \geq N_\gamma$ , the sequence  $v_{\ell,m}(\hat{\mathbf{d}} - \mathbf{d}^0)$  belongs to  $\mathcal{S}(0, \gamma, 2^\gamma)$ . Proposition 9 shows that  $S_{\ell,m}(v_{\ell,m}(\hat{\mathbf{d}} - \mathbf{d}^0))$  tends to zero when  $2^{-j_0\beta} + N^{-1/2} 2^{j_1/2} \rightarrow 0$ .

Finally, as  $\sum_j n_j v_{j,\ell,m} \rightarrow 1$  and as the quantity  $2^{\langle \mathcal{J} \rangle(d_\ell - d_\ell^0 + d_m - d_m^0)}$  tends to 0,  $\hat{G}_{\ell,m}(\hat{\mathbf{d}}) \xrightarrow{\mathbb{P}} G_{\ell,m}^0$  when  $\hat{\mathbf{d}} - \mathbf{d}^0 = o_{\mathbb{P}}(1)$ . The rate of convergence of  $\hat{G}_{\ell,m}(\hat{\mathbf{d}})$  is derived from the fact that  $2^{\langle \mathcal{J} \rangle(\hat{d}_\ell - d_\ell^0 + \hat{d}_m - d_m^0)} - 1 \sim j_0(\hat{d}_\ell - d_\ell^0 + \hat{d}_m - d_m^0) \log(2)$ .

The convergence of  $\mathbf{\Omega}$  is straightforward, thanks to the fact that  $K(\cdot)$  is a continuous function of  $\mathbf{d}$ .

## E Proof of Theorem 6

The criterion  $R$  is equal to  $R(\mathbf{d}) = \log \det \left( \mathbf{\Lambda}_{\langle \mathcal{J} \rangle}(\mathbf{d}) \hat{\mathbf{G}}(\mathbf{d}) \mathbf{\Lambda}_{\langle \mathcal{J} \rangle}(\mathbf{d}) \right)$ . It is immediate that  $\hat{\mathbf{d}} = \underset{\mathbf{d}}{\operatorname{argmin}} R(\mathbf{d})$  satisfies

$$\hat{\mathbf{d}} = \underset{\mathbf{d}}{\operatorname{argmin}} \bar{R}(\mathbf{d}) \quad \text{with} \quad \bar{R}(\mathbf{d}) = \log \det \bar{\mathbf{G}}(\mathbf{d}) \quad (20)$$

$$\text{and} \quad \bar{\mathbf{G}}(\mathbf{d}) = \mathbf{\Lambda}_{\langle \mathcal{J} \rangle}(\mathbf{d} - \mathbf{d}^0) \hat{\mathbf{G}}(\mathbf{d}) \mathbf{\Lambda}_{\langle \mathcal{J} \rangle}(\mathbf{d} - \mathbf{d}^0)$$

When  $\mathbf{d} - \mathbf{d}^0 = o_{\mathbb{P}}(1)$ , it has been proven that the matrix  $\hat{\mathbf{G}}(\mathbf{d})$  converges to  $\mathbf{G}(\mathbf{d}^0)$ . So does  $\bar{\mathbf{G}}(\mathbf{d})$ .

The Taylor expansion at first order of the derivative of  $\bar{R}$  at  $\mathbf{d}_0$  gives:

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

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

### E.1 Derivatives of the criterion

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

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

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

$$\frac{\partial^2 \bar{\mathbf{G}}(\mathbf{d})}{\partial d_a \partial d_b} = \log(2)^2 \Lambda_{<\mathcal{J}>}(\mathbf{d} - \mathbf{d}_0) \frac{1}{n} \sum_{j=j_0}^{j_1} (j - <\mathcal{J}>)^2 \Lambda_j(\mathbf{d})^{-1} (\mathbf{i}_b \mathbf{i}_a \mathbf{I}(j) + \mathbf{I}(j) \mathbf{i}_a \mathbf{i}_b + \mathbf{i}_b \mathbf{I}(j) \mathbf{i}_a + \mathbf{i}_a \mathbf{I}(j) \mathbf{i}_b) \Lambda_j(\mathbf{d})^{-1} \Lambda_{<\mathcal{J}>}(\mathbf{d} - \mathbf{d}_0) \quad (21)$$

Consequently, the derivatives of the criterion  $\bar{R}(\mathbf{d})$  are equal to:

$$\frac{\partial \bar{R}(\mathbf{d})}{\partial d_a} = \text{trace} \left( \bar{\mathbf{G}}(\mathbf{d})^{-1} \frac{\partial \bar{\mathbf{G}}(\mathbf{d})}{\partial d_a} \right) \quad (22)$$

$$\frac{\partial^2 \bar{R}(\mathbf{d})}{\partial d_a \partial d_b} = -\text{trace} \left( \bar{\mathbf{G}}(\mathbf{d})^{-1} \frac{\partial \bar{\mathbf{G}}(\mathbf{d})}{\partial d_b} \bar{\mathbf{G}}(\mathbf{d})^{-1} \frac{\partial \bar{\mathbf{G}}(\mathbf{d})}{\partial d_a} \right) + \text{trace} \left( \bar{\mathbf{G}}(\mathbf{d})^{-1} \frac{\partial^2 \bar{\mathbf{G}}(\mathbf{d})}{\partial d_a \partial d_b} \right) \quad (23)$$

**Convergence of  $\frac{\partial^2 \bar{\mathbf{G}}_{a,b}(\mathbf{d})}{\partial d_a \partial d_b}$**

The proof is derived for  $\left( \frac{\partial^2 \bar{\mathbf{G}}_{a,b}(\mathbf{d})}{\partial d_a \partial d_b} \right)_{a,b}$ , for  $a \neq b$ . The argumentation is similar for the diagonal terms. The expression (21) is rewritten by introducing a sequence  $S_{a,b}$ :

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

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

First,  $\sum_{j=j_0}^{j_1} n_j \omega_{j,a,b}^{(2)}(\mathbf{d} - \mathbf{d}^0) \sim \frac{1}{n} \sum_{j=j_0}^{j_1} (j - <\mathcal{J}>)^2 n_j = \kappa_{j_1 - j_0}$  when  $\mathbf{d} - \mathbf{d}^0 \rightarrow 0$ .

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  $\omega_{a,b}^{(2)}$  belongs to the set  $\mathcal{S}(2, \gamma, 2^{\gamma})$ . Using Proposition 9,  $S_{a,b}(\omega_{a,b}^{(2)}(\mathbf{d} - \mathbf{d}^0)) \leq C(2^{j_1 \beta} + N^{-1/2} 2^{-j_1/2})$  uniformly on the neighbourhood of  $\mathbf{d}$  for  $N \geq N_0$ . As a consequence,

$$\left( \frac{\partial^2 \bar{\mathbf{G}}_{a,b}(\mathbf{d})}{\partial d_a \partial d_b} \right)_{a,b} \sim G_{a,b}^0 \log(2)^2 (\kappa_{j_1 - j_0} + C_{a,b}^{(2)}), \quad (24)$$

with  $C_{a,b}^{(2)} = \mathcal{O}_{\mathbb{P}}(2^{j_1 \beta} + N^{-1/2} 2^{-j_1/2})$ .

With matrix notations,

$$\frac{\partial^2 \bar{\mathbf{G}}(\mathbf{d})}{\partial d_a \partial d_b} = (\kappa_{j_1 - j_0} + C^{(2)}) \log(2)^2 \left( \mathbf{i}_b \mathbf{i}_a \mathbf{G}^0 + \mathbf{i}_b \mathbf{G}^0 \mathbf{i}_a + \mathbf{i}_a \mathbf{G}^0 \mathbf{i}_b + \mathbf{G}^0 \mathbf{i}_a \mathbf{i}_b \right) + o_{\mathbb{P}}(1)$$

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



## Convergence of $\frac{\partial \bar{G}^{(d)}}{\partial d_a}$

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

$$\begin{aligned} \left(\frac{\partial \bar{G}(\mathbf{d})}{\partial d_a}\right)_{a,b} &= \log(2) 2^{<\mathcal{J}>(d_a+d_b-d_a^0-d_b^0)} \frac{1}{n} \sum_{j=j_0}^{j_1} (j - <\mathcal{J}>) 2^{-j(d_a+d_b)} I_{a,b}(j) \\ &= G_{a,b}^0 \log(2) \left[ S_{a,b}(\omega_{a,b}^{(1)}(\mathbf{d} - \mathbf{d}^0)) + \sum_{j=j_0}^{j_1} n_j \omega_{j,a,b}^{(1)}(\mathbf{d} - \mathbf{d}^0) \right], \end{aligned} \quad (25)$$

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

First,  $\sum_{j=j_0}^{j_1} n_j \omega_{j,a,b}^{(1)}(\mathbf{d} - \mathbf{d}^0) = o(1)$  when  $d - d^0 \rightarrow 0$ .

As it was done previously, let  $\mathbf{d}$  be in a neighbourhood of  $\mathbf{d}^0$  such that  $\sup_{\ell} |d_{\ell} - d_{\ell}^0| < \gamma$  with  $0 \leq \gamma < 1/2$ . Then for  $N \geq N_0$ , the sequence  $\omega_{a,b}^{(1)}$  belongs to  $\mathcal{S}(1, \gamma, 2^{\gamma})$ . Thanks to Proposition 9 it comes that  $S_{a,b}(\omega_{a,b}^{(1)}(d - d^0)) = \mathcal{O}_{\mathbb{P}}(2^{\beta j_1} + N^{-1/2} 2^{-j_1/2})$  uniformly on the neighbourhood. Consequently,

$$\left(\frac{\partial \bar{G}(\mathbf{d})}{\partial d_a}\right)_{a,b} \sim G_{a,b}^0 \log(2) C_{a,b}^{(1)} \quad (26)$$

where  $C^{(1)} = \mathcal{O}_{\mathbb{P}}(2^{\beta j_1} + N^{-1/2} 2^{-j_1/2})$ .

Finally,

$$\frac{\partial \bar{G}(\mathbf{d})}{\partial d_a} = C^{(1)} \log(2) \left( \mathbf{i}_a \mathbf{G}^0 + \mathbf{G}^0 \mathbf{i}_a \right) + o_{\mathbb{P}}(1)$$

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

## E.2 Second derivative

Let us detail the expression of the second derivative  $\frac{\partial^2 \bar{R}(\mathbf{d})}{\partial d_a \partial d_b}$ :

$$\begin{aligned} \frac{\partial^2 \bar{R}(\mathbf{d})}{\partial d_a \partial d_b} &= -\text{trace} \left( \bar{G}(\mathbf{d})^{-1} \frac{\partial \bar{G}(\mathbf{d})}{\partial d_b} \bar{G}(\mathbf{d})^{-1} \frac{\partial \bar{G}(\mathbf{d})}{\partial d_a} \right) + \text{trace} \left( \bar{G}(\mathbf{d})^{-1} \frac{\partial^2 \bar{G}(\mathbf{d})}{\partial d_a \partial d_b} \right) \\ &\sim -\log(2)^2 C^{(1)2} \text{trace} \left( \mathbf{G}^{0-1} (\mathbf{i}_b \mathbf{G}^0 + \mathbf{G}^0 \mathbf{i}_b) \mathbf{G}^{0-1} (\mathbf{i}_a \mathbf{G}^0 + \mathbf{G}^0 \mathbf{i}_a) \right) \\ &\quad + \log(2)^2 (\kappa_{j_1-j_0} + C^{(2)}) \text{trace} \left( \mathbf{G}^{0-1} (\mathbf{i}_b \mathbf{i}_a \mathbf{G}^0 + \mathbf{i}_b \mathbf{G}^0 \mathbf{i}_a + \mathbf{i}_a \mathbf{G}^0 \mathbf{i}_b + \mathbf{G}^0 \mathbf{i}_a \mathbf{i}_b) \right) \\ &\sim \log(2)^2 (\kappa_{j_1-j_0} + C^{(2)} - C^{(1)2}) \text{trace} \left( \mathbf{G}^{0-1} (\mathbf{i}_b \mathbf{i}_a \mathbf{G}^0 + \mathbf{i}_b \mathbf{G}^0 \mathbf{i}_a + \mathbf{i}_a \mathbf{G}^0 \mathbf{i}_b + \mathbf{G}^0 \mathbf{i}_a \mathbf{i}_b) \right) \end{aligned}$$

When  $a \neq b$ ,

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

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

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

Finally,

$$\frac{\partial^2 \bar{R}(\mathbf{d})}{\partial \mathbf{d} \partial \mathbf{d}^T} \sim \tilde{\kappa} 2 \log(2)^2 \left( \mathbf{G}^{0-1} \circ \mathbf{G}^0 + \mathbf{I}_p \right) \quad (27)$$

with  $\tilde{\kappa} = \kappa_{j_1-j_0} + C^{(2)} - C^{(1)2}$ .

### E.3 Proof of the Theorem

Using the results established previously,

$$2 \log(2)^2 \tilde{\kappa} (\hat{\mathbf{d}} - \mathbf{d}^0) = - \left( \frac{1}{2\tilde{\kappa}} \frac{\partial^2 \bar{R}(\mathbf{d})}{\partial \mathbf{d} \partial \mathbf{d}^T} \Big|_{\hat{\mathbf{d}}} \right)^{-1} \frac{\partial \bar{R}(\mathbf{d})}{\partial \mathbf{d}} \Big|_{\mathbf{d}^0} \sim (\mathbf{G}^{0-1} \circ \mathbf{G}^0 + \mathbf{I}_p)^{-1} \frac{\partial \bar{R}(\mathbf{d})}{\partial \mathbf{d}} \Big|_{\mathbf{d}^0}$$

We now study the convergence of  $\frac{\partial \bar{R}(\mathbf{d})}{\partial \mathbf{d}} \Big|_{\mathbf{d}^0}$ . Using equation (25), we have the equivalence:

$$\frac{\partial \bar{R}(\mathbf{d})}{\partial d_a} \Big|_{\mathbf{d}^0} \sim C^{(1)} 2 \log(2) \sum_b \bar{G}_{a,b}^{(-1)}(\mathbf{d}_0) G_{a,b}^0.$$

So the asymptotic behaviour of the first derivative of the criterion is:  $\frac{\partial \bar{R}(\mathbf{d})}{\partial d_a} \Big|_{\mathbf{d}^0} = \mathcal{O}_{\mathbb{P}}(2^{j_1 \beta} + N^{-1/2} 2^{-j_1/2})$ .

When plugging this result into the expression above, it comes:

$$\tilde{\kappa} \left( 2^{j_1 \beta} + N^{-1/2} 2^{-j_1/2} \right)^{-1} (\hat{\mathbf{d}} - \mathbf{d}^0) = \mathcal{O}_{\mathbb{P}}(1).$$

Recall that  $\tilde{\kappa} = \kappa_{j_1-j_0} + C^{(2)} - C^{(1)2}$  tends to 0. Since  $\kappa_\ell > 0$  for  $\ell \geq 1$  and  $\kappa_\ell \rightarrow 2$  when  $\ell \rightarrow \infty$ , the sequence  $\tilde{\kappa}$  is bounded below by a positive constant. Theorem 6 follows.

### E.4 Additional tools

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

Let define the sequences  $\eta_L$  and  $\kappa_L$  for any  $L \geq 0$  by:

$$\eta_L := \sum_{i=0}^L i \frac{2^{-i}}{2 - 2^{-L}} \quad (28)$$

$$\kappa_L := \sum_{i=0}^L (i - \eta_L)^2 \frac{2^{-i}}{2 - 2^{-L}} \quad (29)$$

It is straightforward that:

$$n \sim N 2^{-j_0} (2 - 2^{-(j_1-j_0)}) \quad (30)$$

$$< \mathcal{J} > \sim j_0 + \eta_{j_1-j_0} \quad (31)$$

$$\frac{1}{n} \sum_{j=j_0}^{j_1} (j - < \mathcal{J} >)^2 n_j \sim \kappa_{j_1-j_0} \quad (32)$$

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

And for all  $u \geq 0$ ,

$$\frac{1}{\kappa_L} \sum_{i=0}^{L-u} \frac{2^{-i}}{2 - 2^{-L}} (i - \eta_L)(i + u - \eta_L) \rightarrow 1 \text{ when } L \rightarrow \infty \quad (33)$$

## 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 E (2008) Fractal connectivity of long-memory networks. *Physical Review E* 77(3):036,104
- Amblard PO, Coeurjolly JF (2011) Identification of the multivariate fractional brownian motion. *Signal Processing, IEEE Transactions on* 59(11):5152–5168
- Bardet J, Lang G, Moulines E, Soulier P (2000) Wavelet estimator of long-range dependent processes. *Statistical Inference for Stochastic Processes* 3(1-2):85–99
- Bollerslev T, Wright JH (2000) Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data. *Journal of Econometrics* 98(1):81–106
- Brunetti C, Gilbert CL (2000) Bivariate figarch and fractional cointegration. *Journal of Empirical Finance* 7(5):509–530
- Chambers M (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* pp 1749–1766
- Daubechies I (1992) Ten lectures on wavelets, vol 61. SIAM
- Didier G, Pipiras V, et al (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* pp 517–532
- Frías M, Alonso F, Ruiz-Medina M, Angulo J (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. Academic Press
- 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 (1970) *Multiple time series*. John Wiley & Sons
- Hannan EJ (1973) The asymptotic theory of linear time-series models. *Journal of Applied Probability* pp 130–145
- Henry M, Zaffaroni P (2003) The long range dependence paradigm for macroeconomics and finance. *Theory and Applications of Long-range Dependence* pp 417–438
- 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 (2014) Definitions and representations of multivariate long-range dependent time series. *Journal of Time Series Analysis* DOI 10.1111/jtsa.12086
- 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. *NeuroImage* 25:141–158
- 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* pp 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. *Theory and Applications of Long-range Dependence* pp 473–505

- Percival DB, Walden AT (2006) Wavelet methods for time series analysis, vol 4. Cambridge University Press
- Phillips PC, Shimotsu K, et al (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* pp 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* pp 1048–1072
- Robinson PM (2005) Robust covariance matrix estimation: Hac estimates with long memory/antipersistence correction. *Econometric Theory* 21(01):171–180
- Robinson PM (2008) Multiple local Whittle estimation in stationary systems. *The Annals of Statistics* 36(05):2508–2530
- Sela RJ (2010) Three essays in econometrics: multivariate long memory time series and applying regression trees to longitudinal data. PhD thesis, Citeseer
- 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* pp 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

Table 1: Multivariate Whittle wavelet estimation of  $\mathbf{d}$  for a bivariate  $ARFIMA(0, \mathbf{d}, 0)$  with  $\rho = 0.4$ ,  $N = 512$  with 1000 repetitions.

$d_1$	$\mathbf{d}$	$j_0$	bias	std	RMSE	ratio M/U
0.2	0.2	1	-0.0267	0.0413	0.0492	0.9080
	-0.2		0.0379	0.0430	0.0574	1.0595
	0.2	1	-0.0298	0.0428	0.0522	0.9631
	0.0		-0.0002	0.0438	0.0438	0.9504
	0.2	1	-0.0330	0.0456	0.0563	0.9713
	0.2		-0.0333	0.0443	0.0554	0.9831
	0.2	1	-0.0304	0.0429	0.0526	0.9583
	0.4		-0.0571	0.0461	0.0734	0.9701
1.2	1.2	2	-0.0380	0.0830	0.0913	0.9728
	0.8		-0.0298	0.0775	0.0831	0.9643
	1.2	2	-0.0360	0.0818	0.0894	0.9702
	1.0		-0.0346	0.0808	0.0879	0.9626
	1.2	2	-0.0463	0.0853	0.0970	0.9677
	1.2		-0.0393	0.0850	0.0936	0.9688
	1.2	2	-0.0369	0.0799	0.0880	0.9589
	1.4		-0.0482	0.0863	0.0989	0.9648
2.2	2.2	2	-0.0452	0.0950	0.1052	0.9642
	1.8		-0.0418	0.0935	0.1025	0.9657
	2.2	2	-0.0432	0.0962	0.1054	0.9712
	2.0		-0.0459	0.0930	0.1037	0.9524
	2.2	2	-0.0516	0.0953	0.1084	0.9815
	2.2		-0.0513	0.0955	0.1084	0.9812
	2.2	2	-0.0489	0.0962	0.1079	0.9840
	2.4		-0.0573	0.0986	0.1141	0.9797

Table 2: Multivariate Whittle Fourier estimation of  $\mathbf{d}$  for a bivariate  $ARFIMA(0, \mathbf{d}, 0)$  with  $\rho = 0.4$ ,  $N = 512$  with 1000 repetitions.  $\lfloor x \rfloor$  denotes the closest integer smaller than  $x$ .

$\mathbf{d}$	bias	std	RMSE	ratio W/F
Number of frequencies $m = \lfloor N^{0.65} \rfloor = 57$ .				
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
Number of frequencies $m = \lfloor N^{0.876} \rfloor = 236$ .				
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



Table 3: Multivariate Whittle wavelet estimation of  $\mathbf{d}$  for a bivariate  $ARFIMA(0, \mathbf{d}, 0)$  with  $\rho = -0.8$ ,  $N = 512$  with 1000 repetitions.

$d_1$	$\mathbf{d}$	$j_0$	bias	std	RMSE	ratio M/U
0.2	0.2	1	0.0127	0.0396	0.0416	0.7685
	-0.2		0.0797	0.0405	0.0894	1.6425
	0.2	1	-0.0161	0.0380	0.0413	0.7625
	0.0		0.0129	0.0371	0.0393	0.8980
	0.2	1	-0.0334	0.0384	0.0509	0.8780
	0.2		-0.0331	0.0391	0.0512	0.8966
	0.2	1	-0.0164	0.0392	0.0425	0.7742
	0.4		-0.0439	0.0387	0.0585	0.7836
1.2	1.2	2	-0.0130	0.0773	0.0784	0.8360
	0.8		-0.0048	0.0703	0.0704	0.8248
	1.2	2	-0.0277	0.0744	0.0794	0.8614
	1.0		-0.0255	0.0699	0.0744	0.8363
	1.2	2	-0.0440	0.0730	0.0852	0.8498
	1.2		-0.0411	0.0714	0.0824	0.8475
	1.2	2	-0.0321	0.0700	0.0770	0.8394
	1.4		-0.0371	0.0732	0.0820	0.8538
2.2	2.2	2	-0.0267	0.0898	0.0937	0.8587
	1.8		-0.0282	0.0866	0.0911	0.8266
	2.2	2	-0.0372	0.0856	0.0933	0.8599
	2.0		-0.0366	0.0838	0.0915	0.8337
	2.2	2	-0.0507	0.0812	0.0957	0.8669
	2.2		-0.0529	0.0838	0.0991	0.8762
	2.2	2	-0.0419	0.0832	0.0931	0.8492
	2.4		-0.0430	0.0869	0.0969	0.8395

Table 4: Multivariate Whittle Fourier estimation of  $\mathbf{d}$  for a bivariate  $ARFIMA(0, \mathbf{d}, 0)$  with  $\rho = -0.8$ ,  $N = 512$  with 1000 repetitions.

$\mathbf{d}$	bias	std	RMSE	ratio W/F
Number of frequencies $m = \lfloor N^{0.65} \rfloor = 57$ .				
0.2	0.0008	0.0628	0.0628	0.6626
-0.2	0.0033	0.0641	0.0642	1.3934
0.2	-0.0000	0.0595	0.0595	0.6941
0.0	-0.0017	0.0582	0.0583	0.6746
0.2	-0.0057	0.0580	0.0583	0.8730
0.2	-0.0054	0.0582	0.0585	0.8764
0.2	0.0034	0.0652	0.0653	0.6506
0.4	0.0053	0.0684	0.0686	0.8536
Number of frequencies $m = \lfloor N^{0.876} \rfloor = 236$ .				
0.2	-0.0098	0.0273	0.0290	1.4366
-0.2	0.0133	0.0274	0.0304	2.9392
0.2	-0.0138	0.0271	0.0304	1.3583
0.0	-0.0043	0.0264	0.0268	1.4686
0.2	-0.0191	0.0266	0.0327	1.5565
0.2	-0.0196	0.0269	0.0333	1.5398
0.2	-0.0194	0.0275	0.0337	1.2613
0.4	-0.0296	0.0293	0.0416	1.4064

Table 5: Multivariate Whittle wavelet estimation of  $\mathbf{d}$  for a bivariate  $ARFIMA(1, \mathbf{d}, 0)$  with  $\rho = 0.4$ ,  $N = 512$  with 1000 repetitions.

$d_1$	$\mathbf{d}$	$j_0$	bias	std	RMSE	ratio M/U
0.2	0.2	2	-0.1457	0.0755	0.1641	0.9660
	-0.2		-0.1274	0.0744	0.1476	0.9597
	0.2	2	-0.1485	0.0749	0.1663	0.9862
	0.0		-0.1337	0.0717	0.1517	0.9851
	0.2	2	-0.1499	0.0765	0.1683	0.9950
	0.2		-0.1393	0.0749	0.1581	0.9898
	0.2	2	-0.1512	0.0747	0.1687	0.9898
	0.4		-0.1402	0.0721	0.1576	0.9914
1.2	1.2	2	-0.1352	0.0859	0.1602	0.9675
	0.8		-0.1299	0.0774	0.1513	0.9649
	1.2	2	-0.1397	0.0845	0.1633	0.9891
	1.0		-0.1310	0.0756	0.1512	0.9903
	1.2	2	-0.1391	0.0861	0.1636	0.9920
	1.2		-0.1305	0.0844	0.1554	0.9857
	1.2	2	-0.1399	0.0839	0.1631	0.9899
	1.4		-0.1296	0.0889	0.1572	0.9917
2.2	2.2	2	-0.1081	0.0973	0.1455	0.9737
	1.8		-0.1158	0.0947	0.1496	0.9752
	2.2	2	-0.1103	0.0965	0.1465	0.9844
	2.0		-0.1070	0.0947	0.1429	0.9869
	2.2	2	-0.1110	0.0947	0.1459	0.9858
	2.2		-0.1084	0.0945	0.1439	0.9789
	2.2	2	-0.1102	0.0939	0.1448	0.9879
	2.4		-0.1014	0.0943	0.1385	0.9793

Table 6: Multivariate Whittle Fourier estimation of  $\mathbf{d}$  for a bivariate  $ARFIMA(1, \mathbf{d}, 0)$  with  $\rho = 0.4$ ,  $N = 512$  with 1000 repetitions.

$\mathbf{d}$	bias	std	RMSE	ratio W/F
Number of frequencies $m = \lfloor N^{0.65} \rfloor = 57$ .				
0.2	0.0438	0.1862	0.1913	0.8580
-0.2	0.0738	0.2702	0.2801	0.5268
0.2	0.0350	0.1831	0.1864	0.8924
0.0	0.0166	0.1603	0.1612	0.9410
0.2	0.0321	0.1785	0.1813	0.9279
0.2	-0.0025	0.1035	0.1035	1.5278
0.2	0.0364	0.1769	0.1806	0.9340
0.4	0.0055	0.0907	0.0908	1.7351
Number of frequencies $m = \lfloor N^{0.754} \rfloor = 110$ .				
0.2	-0.0669	0.0612	0.0907	1.8087
-0.2	-0.0492	0.0487	0.0692	2.1312
0.2	-0.0686	0.0661	0.0953	1.7457
0.0	-0.0564	0.0474	0.0736	2.0600
0.2	-0.0716	0.0521	0.0885	1.9009
0.2	-0.0597	0.0483	0.0768	2.0604
0.2	-0.0667	0.0552	0.0866	1.9484
0.4	-0.0548	0.0582	0.0800	1.9711

Table 7: Wavelet Whittle estimation of  $\Omega$  for a bivariate  $ARFIMA(0, \mathbf{d}, 0)$  with  $\rho = 0.4$ ,  $N = 512$  with 1000 repetitions.

$d_1$	$\mathbf{d}$		$j_0$	bias	std	RMSE
0.2	(0.2,-0.2)	$\Omega_{1,1}$	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)	$\Omega_{1,1}$	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)	$\Omega_{1,1}$	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)	$\Omega_{1,1}$	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
1.2	(1.2,0.8)	$\Omega_{1,1}$	2	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)	$\Omega_{1,1}$	2	-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)	$\Omega_{1,1}$	2	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)	$\Omega_{1,1}$	2	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
	(2.2,1.8)	$\Omega_{1,1}$	2	-0.0300	0.1906	0.1930
		$\Omega_{1,2}$		0.0408	0.1527	0.1581
		$\Omega_{2,2}$		-0.0167	0.1739	0.1747
		correlation		0.0537	0.1375	0.1476
	(2.2,2.0)	$\Omega_{1,1}$	2	-0.0351	0.1957	0.1988
		$\Omega_{1,2}$		0.0080	0.1159	0.1162
		$\Omega_{2,2}$		-0.0197	0.1923	0.1933
		correlation		0.0204	0.0883	0.0906
	(2.2,2.2)	$\Omega_{1,1}$	2	-0.0342	0.1945	0.1975
		$\Omega_{1,2}$		-0.0085	0.1081	0.1085
		$\Omega_{2,2}$		-0.0284	0.2016	0.2036
		correlation		0.0054	0.0804	0.0806
	(2.2,2.4)	$\Omega_{1,1}$	2	-0.0270	0.1961	0.1980
		$\Omega_{1,2}$		0.0039	0.1253	0.1253
		$\Omega_{2,2}$		-0.0225	0.2001	0.2014
		correlation		0.0145	0.0962	0.0973

Table 8: Fourier Whittle estimation of  $\Omega$  for a bivariate  $ARFIMA(0, \mathbf{d}, 0)$  with  $\rho = 0.4$ ,  $N = 512$  with 1000 repetitions.

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

Table 9: Wavelet Whittle estimation of  $\Omega$  for a bivariate  $ARFIMA(0, \mathbf{d}, 0)$  with  $\rho = -0.8$ ,  $N = 512$  with 1000 repetitions.

$d_1$	$\mathbf{d}$		$j_0$	bias	std	RMSE
0.2	(0.2,-0.2)	$\Omega_{1,1}$	1	0.0342	0.0710	0.0788
		$\Omega_{1,2}$		-0.0749	0.0777	0.1079
		$\Omega_{2,2}$		-0.0381	0.0686	0.0784
		correlation		-0.0768	0.0425	0.0878
	(0.2,0.0)	$\Omega_{1,1}$	1	0.0309	0.0697	0.0762
		$\Omega_{1,2}$		-0.0284	0.0641	0.0701
		$\Omega_{2,2}$		-0.0062	0.0667	0.0670
		correlation		-0.0182	0.0236	0.0298
	(0.2,0.2)	$\Omega_{1,1}$	1	0.0297	0.0733	0.0790
		$\Omega_{1,2}$		-0.0263	0.0656	0.0706
		$\Omega_{2,2}$		0.0318	0.0731	0.0798
		correlation		-0.0013	0.0172	0.0173
	(0.2,0.4)	$\Omega_{1,1}$	1	0.0356	0.0703	0.0788
		$\Omega_{1,2}$		-0.0632	0.0659	0.0913
		$\Omega_{2,2}$		0.0708	0.0706	0.1000
		correlation		-0.0195	0.0232	0.0303
1.2	(1.2,0.8)	$\Omega_{1,1}$	2	0.0037	0.1473	0.1474
		$\Omega_{1,2}$		-0.0847	0.1519	0.1739
		$\Omega_{2,2}$		0.0076	0.1307	0.1309
		correlation		-0.0807	0.0982	0.1271
	(1.2,1.0)	$\Omega_{1,1}$	2	-0.0031	0.1411	0.1411
		$\Omega_{1,2}$		-0.0178	0.1219	0.1232
		$\Omega_{2,2}$		-0.0039	0.1314	0.1315
		correlation		-0.0207	0.0480	0.0523
	(1.2,1.2)	$\Omega_{1,1}$	2	0.0055	0.1442	0.1443
		$\Omega_{1,2}$		-0.0072	0.1200	0.1202
		$\Omega_{2,2}$		0.0022	0.1413	0.1413
		correlation		-0.0040	0.0305	0.0308
	(1.2,1.4)	$\Omega_{1,1}$	2	0.0001	0.1496	0.1496
		$\Omega_{1,2}$		-0.0177	0.1363	0.1374
		$\Omega_{2,2}$		-0.0017	0.1543	0.1543
		correlation		-0.0181	0.0499	0.0531
	(2.2,1.8)	$\Omega_{1,1}$	2	-0.0300	0.1906	0.1930
		$\Omega_{1,2}$		-0.0712	0.2183	0.2296
		$\Omega_{2,2}$		-0.0103	0.1839	0.1842
		correlation		-0.0900	0.1441	0.1699
	(2.2,2.0)	$\Omega_{1,1}$	2	-0.0351	0.1957	0.1988
		$\Omega_{1,2}$		0.0075	0.1653	0.1655
		$\Omega_{2,2}$		-0.0303	0.1915	0.1939
		correlation		-0.0200	0.0600	0.0633
	(2.2,2.2)	$\Omega_{1,1}$	2	-0.0342	0.1945	0.1975
		$\Omega_{1,2}$		0.0212	0.1592	0.1606
		$\Omega_{2,2}$		-0.0277	0.1962	0.1982
		correlation		-0.0040	0.0389	0.0391
	(2.2,2.4)	$\Omega_{1,1}$	2	-0.0270	0.1961	0.1980
		$\Omega_{1,2}$		0.0069	0.1677	0.1679
		$\Omega_{2,2}$		-0.0366	0.2050	0.2083
		correlation		-0.0210	0.0622	0.0656

Table 10: Fourier Whittle estimation of  $\Omega$  for a bivariate  $ARFIMA(0, \mathbf{d}, 0)$  with  $\rho = -0.8$ ,  $N = 512$  with 1000 repetitions.

$\mathbf{d}$		bias	std	RMSE	ratio W/F
Number of frequencies $m = \lfloor N^{0.65} \rfloor = 57$ .					
(0.2,-0.2)	$\Omega_{1,1}$	0.0132	0.2036	0.2040	0.3862
	$\Omega_{1,2}$	-0.0029	0.1662	0.1662	0.6493
	$\Omega_{2,2}$	0.0016	0.1987	0.1988	0.3945
	correlation	0.0034	0.0349	0.0351	2.5023
(0.2,0.0)	$\Omega_{1,1}$	0.0125	0.1980	0.1984	0.3840
	$\Omega_{1,2}$	-0.0058	0.1638	0.1639	0.4279
	$\Omega_{2,2}$	0.0048	0.2001	0.2002	0.3348
	correlation	0.0014	0.0350	0.0350	0.8519
(0.2,0.2)	$\Omega_{1,1}$	0.0233	0.2054	0.2067	0.3823
	$\Omega_{1,2}$	-0.0179	0.1692	0.1701	0.4153
	$\Omega_{2,2}$	0.0199	0.2009	0.2019	0.3949
	correlation	-0.0001	0.0366	0.0366	0.4716
(0.2,0.4)	$\Omega_{1,1}$	0.0083	0.2078	0.2079	0.3789
	$\Omega_{1,2}$	-0.0061	0.1714	0.1715	0.5324
	$\Omega_{2,2}$	0.0129	0.2104	0.2108	0.4741
	correlation	0.0027	0.0357	0.0358	0.8462
Number of frequencies $m = \lfloor N^{0.876} \rfloor = 236$ .					
(0.2,-0.2)	$\Omega_{1,1}$	0.0512	0.0677	0.0849	0.9281
	$\Omega_{1,2}$	0.0001	0.0592	0.0592	1.8216
	$\Omega_{2,2}$	-0.0463	0.0631	0.0783	1.0019
	correlation	0.0014	0.0170	0.0171	5.1496
(0.2,0.0)	$\Omega_{1,1}$	0.0459	0.0665	0.0808	0.9424
	$\Omega_{1,2}$	-0.0149	0.0581	0.0599	1.1703
	$\Omega_{2,2}$	-0.0057	0.0635	0.0638	1.0512
	correlation	0.0012	0.0165	0.0165	1.8025
(0.2,0.2)	$\Omega_{1,1}$	0.0458	0.0706	0.0841	0.9393
	$\Omega_{1,2}$	-0.0378	0.0640	0.0743	0.9504
	$\Omega_{2,2}$	0.0467	0.0694	0.0836	0.9536
	correlation	-0.0004	0.0171	0.0171	1.0093
(0.2,0.4)	$\Omega_{1,1}$	0.0497	0.0680	0.0842	0.9353
	$\Omega_{1,2}$	-0.0614	0.0635	0.0883	1.0344
	$\Omega_{2,2}$	0.1061	0.0710	0.1277	0.7831
	correlation	0.0010	0.0169	0.0170	1.7875



Table 11: Wavelet Whittle estimation of  $\Omega$  for a bivariate  $ARFIMA(1, \mathbf{d}, 0)$  with  $\rho = 0.4$ ,  $N = 512$  with 1000 repetitions.

$d_1$	$\mathbf{d}$		$j_0$	bias	std	RMSE
0.2	(0.2,-0.2)	$\Omega_{1,1}$	2	0.1508	0.0484	0.1584
		$\Omega_{1,2}$		0.1003	0.0562	0.1150
		$\Omega_{2,2}$		0.1852	0.0620	0.1953
		correlation		0.0647	0.0958	0.1157
	(0.2,0.0)	$\Omega_{1,1}$	2	0.1502	0.0484	0.1578
		$\Omega_{1,2}$		0.0741	0.0429	0.0856
		$\Omega_{2,2}$		0.1806	0.0634	0.1914
		correlation		0.0154	0.0674	0.0692
	(0.2,0.2)	$\Omega_{1,1}$	2	0.1532	0.0501	0.1611
		$\Omega_{1,2}$		0.0678	0.0404	0.0789
		$\Omega_{2,2}$		0.1768	0.0633	0.1878
		correlation		0.0029	0.0602	0.0603
	(0.2,0.4)	$\Omega_{1,1}$	2	0.1529	0.0504	0.1610
		$\Omega_{1,2}$		0.0716	0.0432	0.0836
		$\Omega_{2,2}$		0.1715	0.0617	0.1823
		correlation		0.0125	0.0678	0.0690
1.2	(1.2,0.8)	$\Omega_{1,1}$	2	0.1134	0.0606	0.1286
		$\Omega_{1,2}$		0.0846	0.0569	0.1019
		$\Omega_{2,2}$		0.1524	0.0640	0.1653
		correlation		0.0678	0.1072	0.1269
	(1.2,1.0)	$\Omega_{1,1}$	2	0.1149	0.0567	0.1281
		$\Omega_{1,2}$		0.0604	0.0430	0.0741
		$\Omega_{2,2}$		0.1423	0.0654	0.1566
		correlation		0.0194	0.0732	0.0757
	(1.2,1.2)	$\Omega_{1,1}$	2	0.1166	0.0573	0.1299
		$\Omega_{1,2}$		0.0537	0.0415	0.0678
		$\Omega_{2,2}$		0.1349	0.0727	0.1532
		correlation		0.0070	0.0659	0.0662
	(1.2,1.4)	$\Omega_{1,1}$	2	0.1149	0.0551	0.1274
		$\Omega_{1,2}$		0.0549	0.0459	0.0715
		$\Omega_{2,2}$		0.1278	0.0748	0.1480
		correlation		0.0136	0.0795	0.0806
	(2.2,1.8)	$\Omega_{1,1}$	2	0.0665	0.0701	0.0966
		$\Omega_{1,2}$		0.0649	0.0717	0.0967
		$\Omega_{2,2}$		0.1039	0.0865	0.1352
		correlation		0.0736	0.1449	0.1625
	(2.2,2.0)	$\Omega_{1,1}$	2	0.0655	0.0714	0.0969
		$\Omega_{1,2}$		0.0389	0.0506	0.0638
		$\Omega_{2,2}$		0.0869	0.0854	0.1218
		correlation		0.0210	0.0914	0.0938
	(2.2,2.2)	$\Omega_{1,1}$	2	0.0673	0.0688	0.0963
		$\Omega_{1,2}$		0.0314	0.0453	0.0551
		$\Omega_{2,2}$		0.0811	0.0859	0.1181
		correlation		0.0051	0.0806	0.0807
	(2.2,2.4)	$\Omega_{1,1}$	2	0.0662	0.0705	0.0967
		$\Omega_{1,2}$		0.0336	0.0502	0.0604
		$\Omega_{2,2}$		0.0688	0.0905	0.1137
		correlation		0.0161	0.0924	0.0938

Table 12: Fourier Whittle estimation of  $\Omega$  for a bivariate  $ARFIMA(1, \mathbf{d}, 0)$  with  $\rho = 0.4$ ,  $N = 512$  with 1000 repetitions.

$\mathbf{d}$		bias	std	RMSE	ratio W/F
Number of frequencies $m = \lfloor N^{0.65} \rfloor = 57$ .					
(0.2,-0.2)	$\Omega_{1,1}$	0.0204	0.0895	0.0918	1.7243
	$\Omega_{1,2}$	-0.0025	0.0612	0.0613	1.8775
	$\Omega_{2,2}$	0.0132	0.1193	0.1200	1.6270
	correlation	-0.0319	0.1300	0.1339	0.8641
(0.2,0.0)	$\Omega_{1,1}$	0.0257	0.0937	0.0971	1.6242
	$\Omega_{1,2}$	0.0034	0.0572	0.0573	1.4948
	$\Omega_{2,2}$	0.0275	0.1060	0.1095	1.7486
	correlation	-0.0234	0.1143	0.1167	0.5930
(0.2,0.2)	$\Omega_{1,1}$	0.0260	0.0895	0.0932	1.7288
	$\Omega_{1,2}$	0.0066	0.0540	0.0544	1.4496
	$\Omega_{2,2}$	0.0357	0.1011	0.1072	1.7519
	correlation	-0.0179	0.1014	0.1030	0.5856
(0.2,0.4)	$\Omega_{1,1}$	0.0219	0.0871	0.0898	1.7922
	$\Omega_{1,2}$	0.0051	0.0523	0.0526	1.5901
	$\Omega_{2,2}$	0.0282	0.1027	0.1065	1.7124
	correlation	-0.0146	0.1010	0.1021	0.6758
Number of frequencies $m = \lfloor N^{0.754} \rfloor = 110$ .					
(0.2,-0.2)	$\Omega_{1,1}$	0.0957	0.0473	0.1067	1.4834
	$\Omega_{1,2}$	0.0383	0.0332	0.0507	2.2682
	$\Omega_{2,2}$	0.0956	0.0563	0.1109	1.7601
	correlation	-0.0003	0.0585	0.0585	1.9785
(0.2,0.0)	$\Omega_{1,1}$	0.0970	0.0469	0.1077	1.4650
	$\Omega_{1,2}$	0.0395	0.0346	0.0525	1.6292
	$\Omega_{2,2}$	0.1006	0.0562	0.1153	1.6604
	correlation	-0.0010	0.0595	0.0595	1.1628
(0.2,0.2)	$\Omega_{1,1}$	0.0977	0.0452	0.1076	1.4973
	$\Omega_{1,2}$	0.0418	0.0343	0.0541	1.4597
	$\Omega_{2,2}$	0.1105	0.0574	0.1245	1.5087
	correlation	-0.0000	0.0584	0.0584	1.0323
(0.2,0.4)	$\Omega_{1,1}$	0.0943	0.0449	0.1045	1.5410
	$\Omega_{1,2}$	0.0428	0.0338	0.0545	1.5328
	$\Omega_{2,2}$	0.1188	0.0593	0.1328	1.3727
	correlation	0.0007	0.0578	0.0578	1.1943

Table 13: Multivariate Wavelet Whittle estimation of  $\Omega$  for a bivariate  $ARFIMA(0, (0.2, 1.2), 0)$  with  $\rho = 0.4$ ,  $N = 512$  with 1000 repetitions.

	bias	std	RMSE
Without differentiation of the second component and $j_0 = 2$ .			
$\Omega_{1,1}$	0.0039	0.1108	0.1109
$\Omega_{1,2}$	12.464	432.66	432.84
$\Omega_{2,2}$	0.0016	0.1487	0.1488
correlation	9.7231	349.41	349.55
With differentiation of the second component and $j_0 = 1$ .			
$\Omega_{1,1}$	0.0411	0.0709	0.0819
$\Omega_{1,2}$	0.0163	0.0548	0.0571
$\Omega_{2,2}$	0.0103	0.0745	0.0752
correlation	0.0055	0.0410	0.0414