Journal of Statistical Software

May 2019, Volume 89, Issue 6. doi: 10.18637/jss.v089.i06

Wavelet-Based and Fourier-Based Multivariate Whittle Estimation: multiwave

Sophie Achard
Univ. Grenoble Alpes, CNRS,
Grenoble-INP, GIPSA-lab

Irène Gannaz
Université Lyon, INSA de Lyon,
Institut Camille Jordan

Abstract

Multivariate time series with long-dependence are observed in many applications such as finance, geophysics or neuroscience. Many packages provide estimation tools for univariate settings but few are addressing the problem of long-dependence estimation for multivariate settings. The package **multiwave** is providing efficient estimation procedures for multivariate time series. Two semi-parametric estimation methods of the long-memory exponents and long-run covariance matrix of time series are implemented. The first one is the Fourier-based estimation proposed by Shimotsu (2007) and the second one is a wavelet-based estimation described in Achard and Gannaz (2016). The objective of this paper is to provide an overview of the R package **multiwave** with its practical application perspectives.

Keywords: wavelets, multivariate time series, Whittle estimation, long-memory properties, long-run covariance, R.

1. Introduction

Time series with defined autocovariance functions are said to present long-memory or long-range dependency when their autocovariance function is decreasing very slowly, slower than an exponential decay. More precisely, let $g(\cdot)$ be the autocovariance function of a time series X. X is said to be long-memory if there exists α, $0 < \alpha < 1$, such that $g(t)$ is asymptotically equivalent to $|t|^{-\alpha}$ when $t \to +\infty$ (see Beran 1994 and references therein). This definition implies that the covariance function is not summable. Equivalently, the spectral density $f(\cdot)$, if it exists, is such that $f(\lambda)$ is equivalent up to a constant to $|\lambda|^{1-\alpha}$ when $\lambda \to 0^+$. In this case, when trying to estimate the expectation using the empirical mean of long-memory time series, the variance of the estimator is not decreasing to 0 as N^{-1} (where N is the sample size). Hence it is crucial to take into account the presence of long-memory for defining good estimators (Beran 1994). In the case of univariate time series, several very efficient approaches

have been developed and validated. The CRAN Task View "Time Series Analysis" (Hyndman 2019) provides a very exhaustive list of methods and softwares dealing with long-memory time series for R (R Core Team 2019). Among others, we can cite the packages **fracdiff** (Haslett and Raftery 1989; Mächler 2012), **arfima** (Veenstra and McLeod 2018) and **FGN** (Veenstra 2012; McLeod and Veenstra 2014), **longmemo** (Beran 1994) and **forecast** (Hyndman and Khandakar 2008). For example, **fracdiff** is dedicated to simulation of fractional ARIMA time series and to estimation using regression of the periodogram. **longmemo** provides real data examples of time series with long-memory properties.

Approaches for multivariate long-memory time series are less developed. When dealing with multivariate time series, an important quantity to estimate is the covariance or correlation between pairs of time series. The effect of the presence of long-memory on this estimation is obvious, as stated by Robinson (2005). One R package, **waveslim** (Whitcher 2019), is dedicated to the wavelet correlation analysis for pairs of random variables (Whitcher, Guttorp, and Percival 2000) but long-range dependence properties are not considered. Sela and Hurvich (2012) provide R code (freely available at `http://pages.stern.nyu.edu/~rsela/VARFI/code.html`) for bivariate long-range dependent time series with parametric estimations. The objective of this paper is to provide an efficient R package, called **multiwave** (Achard and Gannaz 2019), to estimate the long-memory parameters and covariance matrices for multivariate time series. The estimation procedures are based on a semi-parametric approach, which is robust to model misspecification. The package is available from the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=multiwave`.

The procedures are also suited for dealing with more than two-dimensional data. Indeed they are based on Whittle approximation which provides a simple function to optimize. This function can be used for any dimension of the problem. In comparison, regression of the scalogram or periodogram (Achard, Bassett, Meyer-Lindenberg, and Bullmore 2008) is based on a linear fit of pairs of time series, and thus there does not exist an easy way to extend to more than two dimensions.

multiwave package is based on Achard and Gannaz (2016), where we developed a wavelet-based approach using Whittle approximation for an efficient estimation of the long-memory parameters and the long-run covariance matrices. In addition, **multiwave** proposes an implementation of an alternative method using Fourier decomposition as described in Shimotsu (2007, code available for MATLAB at `http://shimotsu.web.fc2.com/Site/Matlab_Codes.html`).

As described in this paper, **multiwave** is a very versatile package and opens the way to estimation of the long-memory parameters and the long-run covariance matrices using multivariate data sets. It is in particular not necessary to assume the stationarity of the time series as is the case when using Fourier decomposition (Faÿ, Moulines, Roueff, and Taqqu 2009). The Whittle approximation is computed using either the coefficients of wavelet decomposition or the coefficients of Fourier decomposition when the time series are stationary.

The package **multiwave** is divided in three parts. A first group of functions is dedicated to the simulation of multivariate long-memory time series; the main function is `fivarma`. A second group of functions is implementing the wavelet decomposition, through `DWTexact` and associated functions. Finally the computation of the estimators are coded using the Fourier decomposition in `mfw` and its derivatives and using the wavelet decomposition in `mww` and its derivatives.

The mathematical background is detailed in a separate Section 2. The rest of the paper is dedicated to the description of the package **multiwave**. Simple examples of parametric models and real data are presented in Section 3 with corresponding functions of **multiwave** ready to apply. Core estimation functions using wavelets and Fourier transform are detailed in Section 4 along with pieces of code using simulated time series. Finally, practical considerations are discussed in the three last sections. Section 5 is discussing the practical choices of parameters. Comparisons between wavelets and Fourier approaches are described in Section 6. And an application to real data in neuroscience is concluding the paper in Section 7.

2. Theoretical background

As in the univariate case, the definition of long-memory for a p-vector process is based on the asymptotic behavior of the cross-spectral density in the neighborhood of zero (Moulines, Roueff, and Taqqu 2007). We consider N observations of a long-memory p-vector process $\mathbf{X} = \{X_\ell(k), k \in \mathbb{Z}, \ell = 1, \ldots, p\}$, namely $\mathbf{X}(1), \ldots \mathbf{X}(N)$. $\mathbf{X}$ is said to be a multivariate $M(\mathbf{d})$ process when for each $\ell = 1, \ldots, p$ there exists $D_\ell \in \mathbb{N}$ such that the D_ℓth order difference $\Delta^{D_\ell} X_\ell$ is covariance stationary. In addition, let us assume that for any $\ell, m = 1, \ldots, p$ the generalized cross-spectral density of X_ℓ and X_m is

$$f_{\ell,m}(\lambda) = \frac{1}{2\pi}\Omega_{\ell,m}(1 - e^{-i\lambda})^{-d_\ell}(1 - e^{i\lambda})^{-d_m} f^S_{\ell,m}(\lambda), \qquad \lambda \in [-\pi, \pi], \tag{1}$$

with $\mathbf{\Omega} = (\Omega_{\ell,m})_{\ell,m=1,\ldots,p}$ an Hermitian matrix. The functions $f^S_{\ell,m}(\cdot)$ correspond to the short-memory dynamics of the process. The parameter d_ℓ satisfies $-1/2 < d_\ell - D_\ell < 1/2$. More generally, the wavelet-based procedure is available for cross-spectral density satisfying an approximation

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

where the exponent $*$ denotes the conjugate transpose operator. Here and subsequently $\sim$ means that the ratio of the left- and right-hand sides converges to one. Note that the process X_ℓ is not necessarily stationary.

The long-range dependence parameter measures the power-like rate of decay of the autocovariance function. The long-run covariance matrix $\mathbf{\Omega}$ can be seen as the covariance at low frequencies between the time series. It gives a quantification of the link between the components of the multivariate time series. The long-run covariance parameter of the model is free from the difference in the autocorrelation behavior of each component. It is linked with long-run correlations $\left(\Omega_{\ell,m}/\sqrt{\Omega_{\ell,\ell}\Omega_{m,m}}\right)_{\ell,m=1,\ldots,p}$, which are also encountered in literature as power-law coherencies between two time series (Sela and Hurvich 2012) or as fractal connectivities (Achard *et al.* 2008).

2.1. A parametric example: FIVARMA

Fractionally integrated vector auto regressive moving average (FIVARMA) processes are parametric models with a spectral density satisfying approximation (2). They correspond to Model A of Lobato (1997). We refer to this paper for a detailed mathematical description.

Let $\mathbf{u}$ be a p-dimensional white noise with $\mathsf{E}[\mathbf{u}(t) \mid \mathcal{F}_{t-1}] = 0$ and $\mathsf{E}[\mathbf{u}(t)\mathbf{u}(t)^\top \mid \mathcal{F}_{t-1}] = \mathbf{\Sigma}$, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\mathbf{u}(s), s < t\}$, and $\mathbf{\Sigma}$ is a positive definite matrix. Let $(\mathbf{A}_k)_{k\in\mathbb{N}}$ be a sequence of $\mathbb{R}^{p\times p}$-valued matrices with $\mathbf{A}_0$ the identity matrix and

$\sum_{k=0}^{\infty} \|\mathbf{A}_k\|^2 < \infty$. The discrete Fourier transform of the sequence is denoted $\mathbf{A}(\cdot)$, that is $\mathbf{A}(\lambda) = \sum_{k=0}^{\infty} \mathbf{A}_k e^{ik\lambda}$. We assume that all the roots of $|\mathbf{A}(\mathbb{L})|$ are outside the closed unit circle, where $\mathbb{L}$ denotes the lag operator. Let also $(\mathbf{B}_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^{p\times p}$ with $\mathbf{B}_0$ the identity matrix and $\sum_{k=0}^{\infty} \|\mathbf{B}_k\|^2 < \infty$. As defined for $\mathbf{A}$, $\mathbf{B}(\cdot)$ denotes the discrete Fourier transform of the sequence, $\mathbf{B}(\lambda) = \sum_{k=0}^{\infty} B_k e^{ik\lambda}$.

Let $\mathbf{X}$ be defined by

$$\mathbf{A}(\mathbb{L})\,\mathrm{diag}(\mathbb{1} - \mathbb{L})^{\mathbf{d}}\,\mathbf{X}(t) = \mathbf{B}(\mathbb{L})\mathbf{u}(t). \tag{3}$$

The spectral density satisfies

$$f_{\ell,m}(\lambda) \sim_{\lambda\to 0^+} \frac{1}{2\pi}\Omega_{\ell,m} e^{-i\pi/2(d_\ell - d_m)} \lambda^{-(d_\ell + d_m)}$$

with

$$\mathbf{\Omega} = \mathbf{A}(1)^{-1}\mathbf{B}(1)\mathbf{\Sigma}\mathbf{B}(1)^{\top}\mathbf{A}(1)^{\top -1}. \tag{4}$$

Then $\mathbf{X}$ is called a FIVARMA(d, q) process and satisfies approximation (2).

Limits of the model

Note that in definition (3) the operators are applied in a given order, where the lag operator is taken first. Changing the order of the lag operator and autoregression corresponds to model B of Lobato (1997) and VARFI models of Sela and Hurvich (2012) where $\mathbf{X}$ is obtained with equation $\mathrm{diag}(\mathbb{1} - \mathbb{L})^{\mathbf{d}}\,\mathbf{A}(\mathbb{L})\,\mathbf{X}(t) = \mathbf{B}(\mathbb{L})\mathbf{u}(t)$, with similar notations than above. That is, $\mathbf{X}$ is obtained by fractional integration after autoregression, which is also called *cointegration*. The spectral density still satisfies the approximation (2) however the matrix $\mathbf{\Omega}$ may no longer be Hermitian. Clearly the **multiwave** package is not built to deal with such cases. We refer to alternative methods in literature, among others Robinson (2008); Sela and Hurvich (2012); Shimotsu (2012). Taking into account cointegration is a difficult problem that exceeds the scope of this paper. Future work is needed to handle this particular case.

2.2. Fourier-based estimation (MFW)

The discrete Fourier transform and the periodogram of $\mathbf{X}$ evaluated at frequency λ are defined as in Shimotsu (2007)'s procedure

$$\mathbf{W}^F(\lambda) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{N} \mathbf{X}(t) e^{it\lambda},$$
$$\mathbf{I}^F(\lambda) = \mathbf{W}^F(\lambda)\mathbf{W}^F(\lambda)^*.$$

Let $\lambda_j = 2\pi j/N$, $j = 1, \ldots, m$, be the Fourier frequencies used in estimation, $m \in \mathbb{N}$. Define $\mathbf{\Lambda}_j^F(\mathbf{d}) = \mathrm{diag}\left(\lambda_j^{\mathbf{d}} e^{i(\pi - \lambda_j)\mathbf{d}/2}\right)$. The estimators $(\widehat{\mathbf{d}}^{\mathrm{MFW}}, \widehat{\mathbf{\Omega}}^{\mathrm{MFW}})$ are minimizers of the criterion

$$\mathcal{L}^{\mathrm{MFW}}(\mathbf{d}, \mathbf{\Omega}) = \frac{1}{m}\sum_{j=1}^{m} \left[\log\det\left(\mathbf{\Lambda}_j^F(\mathbf{d})\mathbf{\Omega}(\mathbf{d})\mathbf{\Lambda}_j^F(\mathbf{d})^*\right) + \mathbf{W}^F(\lambda_j)^* \left(\mathbf{\Lambda}_j^F(\mathbf{d})\mathbf{\Omega}(\mathbf{d})\mathbf{\Lambda}_j^F(\mathbf{d})^*\right)^{-1} \mathbf{W}^F(\lambda_j)\right].$$

The solution satisfies

$$\widehat{\mathbf{d}}^{\mathrm{MFW}} = \mathrm{argmin}_{\mathbf{d}} \log\det(\widehat{\mathbf{\Omega}}^{\mathrm{MFW}}(\mathbf{d})) - 2\log(2)\left(\frac{1}{m}\sum_{j=1}^{m}\lambda_j\right), \tag{5}$$

$$\widehat{\mathbf{\Omega}}^{\mathrm{MFW}} = \widehat{\mathbf{\Omega}}^{\mathrm{MFW}}(\widehat{\mathbf{d}}^{\mathrm{MFW}}), \tag{6}$$

$$\text{with } \widehat{\mathbf{\Omega}}^{\mathrm{MFW}}(\mathbf{d}) = \frac{1}{m}\sum_{j=1}^{m} Re\left(\mathbf{\Lambda}_j^F(\mathbf{d})^{-1}\mathbf{I}^F(j)\mathbf{\Lambda}_j^F(\mathbf{d})^{-1}\right).$$

The dynamics of the frequencies at the neighborhood of the origin is given by the dynamics of the spectral density around the zero frequency. The form of the criterion is justified by a second-order approximation of the spectral density matrix (1), rather than the approximation (2). For an estimation based on the first-order approximation (2), one should replace $\mathbf{\Lambda}_j^F(\mathbf{d})$ by $\mathbf{\Lambda}_j^{F\,(1)}(\mathbf{d}) = \mathrm{diag}\left(\lambda_j^{\mathbf{d}} e^{i\pi\mathbf{d}/2}\right)$.

Shimotsu (2007) established the theoretical performance of this estimation procedure, for both the long-range dependence parameters and the long-run covariance matrix. It is shown that the variance for the estimation of the vector $\mathbf{d}$ is decreased for the multivariate procedure with respect to a univariate one. It is worth mentioning that Lobato (1999) developed a similar estimation procedure, based on a rougher approximation of the cross-spectral density, $\mathbf{\Lambda}_j^F(\mathbf{d}) = \mathrm{diag}\left(\lambda_j^{\mathbf{d}}\right)$. Interestingly, the quality of estimation for the vector $\mathbf{d}$ is similar. Nevertheless, the estimation of the long-run covariance matrix $\mathbf{\Omega}$ is biased since it does not take into account the phase-shift appearing in $\mathbf{\Lambda}_j^F(\mathbf{d})$. We refer to Lobato (1999) and to Shimotsu (2007) for a more detailed study of these estimators and their consistency.

2.3. Wavelet-based estimation (MWW)

Wavelets are providing a very efficient tool because of their high flexibility to deal with nonstationary time series which is particularly useful for real data applications. Their good performances in comparison to Fourier have already been shown for example in univariate settings (Faÿ *et al.* 2009).

Let $(\phi(\cdot), \psi(\cdot))$ be respectively a father and a mother wavelets, satisfying regularity conditions, as stated in Achard and Gannaz (2016).

At a given resolution $j \geqslant 0$, 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 wavelet coefficients of the process $\mathbf{X}$ are defined by

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

where $\widetilde{\mathbf{X}}(t) = \sum_{k\in\mathbb{Z}} \mathbf{X}(k)\phi(t-k)$. For given $j \geqslant 0$ and $k \in \mathbb{Z}$, $\mathbf{W}_{j,k}$ is a p-dimensional vector $\mathbf{W}_{jk} = \begin{pmatrix} W_{j,k}(1) & W_{j,k}(2) & \dots & W_{j,k}(p) \end{pmatrix}$ where $W_{j,k}(\ell) = \int_{\mathbb{R}} \widetilde{\mathbf{X}_\ell}(t)\psi_{j,k}(t)dt$.

For any $j \geqslant 0$, the process $(\mathbf{W}_{j,k})_{k\in\mathbb{Z}}$ is covariance stationary (Achard and Gannaz 2016). Let $\theta_{\ell,m}(j)$ denote the wavelet covariance at scale j between processes X_ℓ and X_m, *i.e.*, $\theta_{\ell,m}(j) = \mathsf{COV}(W_{j,k}(\ell), W_{j,k}(m))$ for any position k. Let us introduce the wavelet scalogram

$$\mathbf{I}^W(j) = \sum_{k\in\mathbb{Z}} \mathbf{W}_{j,k}\mathbf{W}_{j,k}^\top. \tag{7}$$

The wavelet scalogram is the equivalent of the Fourier periodogram. Yet the scalogram is not normalized, contrary to the periodogram. We also introduce the function $K(\cdot)$, defined as

$$K(\delta) = \int_{-\infty}^{\infty} |\lambda|^{-\delta} |\widehat{\psi}(\lambda)|^2 \, d\lambda, \quad \delta \in (-\alpha, M). \tag{8}$$

The wavelet Whittle procedure is described in Achard and Gannaz (2016). Let $\mathbf{\Lambda}_j(\mathbf{d}) = \text{diag}\left(2^{j\mathbf{d}}, \, j_0 \leqslant j \leqslant j_1\right)$. Let $\mathbf{G}(\mathbf{d})$ denote a $p \times p$-matrix with (ℓ, m)th element equal to

$$G_{\ell,m}(\mathbf{d}) = f^S(0)\Omega_{\ell,m}K(d_\ell + d_m)cos(\pi(d_\ell - d_m)/2). \tag{9}$$

The estimators $(\widehat{\mathbf{d}}^{\text{MWW}}, \widehat{\mathbf{G}}^{\text{MWW}})$ are defined by minimization of $\mathcal{L}^{MWM}(\mathbf{d}, \mathbf{G})$, with

$$\mathcal{L}^{MWM}(\mathbf{d}, \mathbf{G}) = \\ \frac{1}{n} \sum_{j=j_0}^{j_1} \left[n_j \log\det\left(\mathbf{\Lambda}_j^W(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j^W(\mathbf{d})\right) + \sum_k \mathbf{W}_{j,k}^\top \left(\mathbf{\Lambda}_j^W(\mathbf{d})\mathbf{G}(\mathbf{d})\mathbf{\Lambda}_j^W(\mathbf{d})\right)^{-1} \mathbf{W}_{j,k} \right].$$

The estimation is here based on a first-order approximation of the spectral density matrix around 0.

The solutions of this problem satisfy

$$\widehat{\mathbf{d}}^{\text{MWW}} = \text{argmin}_{\mathbf{d}} \log\det(\widehat{\mathbf{G}}^{\text{MWW}}(\mathbf{d})) + 2\log(2)\left(\frac{1}{n}\sum_{j=j_0}^{j_1} j n_j\right)\left(\sum_{\ell=1}^{p} d_\ell\right), \tag{10}$$

$$\widehat{\mathbf{G}}^{\text{MWW}}(\mathbf{d}) = \frac{1}{n}\sum_{j=j_0}^{j_1} \mathbf{\Lambda}_j^W(\mathbf{d})^{-1}\mathbf{I}^W(j)\mathbf{\Lambda}_j^W(\mathbf{d})^{-1}, \tag{11}$$

where $\mathbf{I}^W(j)$ is the wavelet scalogram at scale j defined in (7). The long-run covariance matrix can then be estimated by

$$\widehat{\Omega}_{\ell,m}^{\text{MWW}} = \widehat{G}_{\ell,m}^{\text{MWW}}(\widehat{\mathbf{d}}^{\text{MWW}})/(\cos(\pi(\widehat{d}_\ell^{\text{MWW}} - \widehat{d}_m^{\text{MWW}})/2)K(\widehat{d}_\ell^{\text{MWW}} + \widehat{d}_m^{\text{MWW}})). \tag{12}$$

This second step in the estimation of the long-run covariance matrix $\mathbf{\Omega}$ is needed because the wavelets used in this paper are real and cannot correct the phase-shift (given by (12)). This is not the case for the Fourier Whittle estimation described in Shimotsu (2007). Fortunately, the phase-shift can be expressed as a multiplicative cosine term in the covariance of the wavelet coefficients and a correction is still possible.

Achard and Gannaz (2016) established that the MWW estimators (10) and (12) are consistent under non-restrictive conditions. The rate of convergence for the estimation of the long-range parameters $\mathbf{d}$ is similar to the MFW estimator and is minimax. We refer to Achard and Gannaz (2016) for the detailed study of the asymptotic behavior of MWW estimation.

3. Examples of multivariate long-memory time series

This section is describing specific functions of **multiwave** for the user to be able to simulate multivariate long-memory processes. Parametric models are defined and implemented. In addition a data set containing real data from neuroimaging is provided.

3.1. Simulations of FIVARMA

The **multiwave** package proposes simulation functions for time series with a spectral density satisfying approximation (2). The main function is `fivarma` which computes a parametric FIVARMA process defined in Section 2.1.

The input parameters of a FIVARMA(q, d, r) process are the covariance matrix $\mathbf{\Sigma}$ of the innovation process $\mathbf{u}$, the vector AR (autoregressive) $(\mathbf{A}_k)_{k=0,\ldots,q}$, $\mathbf{A}_k \in \mathbb{R}^{p\times p}$, the vector MA (moving average) $(\mathbf{B}_k)_{k=0,\ldots,r}$, $\mathbf{B}_k \in \mathbb{R}^{p\times p}$, and the vector of long-range parameters $\mathbf{d} \in \mathbb{R}^p$. The FIVAR model of Sela and Hurvich (2008) is a subcase, corresponding to MA coefficients equal to zero. The parameters of `fivarma` are thus, in order, `(N, d, cov_matrix, VAR, VMA)` where `cov_matrix` $= \mathbf{\Sigma}$ and `VAR` and `VMA` denote respectively the sequences of matrices $(\mathbf{A}_k)_{k=0,\ldots,q}$ and $(\mathbf{B}_k)_{k=0,\ldots,r}$.

The output of the function `fivarma` is a list with first the values $\mathbf{X}(1), \mathbf{X}(2), \ldots, \mathbf{X}(N)$ obtained by Equation 3 with $\mathbf{u}(t)$ white noise with centered Gaussian distribution and covariance $\mathbf{\Sigma}$. The second element of the list is the value of the matrix $\mathbf{\Omega}$ defined in (4).

`fivarma` is based on two other functions:

- `fracdiff` applies a vectorial fractional differencing procedure and corresponds to a FIVARMA$(0, d, 0)$.
- `varma` computes a realization of a multivariate ARMA process and corresponds to the case $\mathbf{d} = \mathbf{0}$.

Similar functions can be found in other packages (e.g., **fracdiff** and **MST**; Tsay 2013) but were re-implemented in **multiwave** package.

Example

```
R> N <- 2^8
R> d0 <- c(0.2, 0.4)
R> rho <- 0.8
R> cov <- matrix(c(1, rho, rho, 1), 2, 2)
R> VMA <- diag(c(0.4, 0.7))
R> VAR <- array(c(0.8, 0.2, 0, 0.6), dim = c(2, 2))
R> resp <- fivarma(N, d0, cov_matrix = cov, VAR = VAR, VMA = VMA)
R> x <- resp$x
R> long_run_cov <- resp$long_run_cov
R> long_run_cov

          [, 1]      [, 2]
[1, ] 0.6049383 0.5854938
[2, ] 0.5854938 0.9730806
```

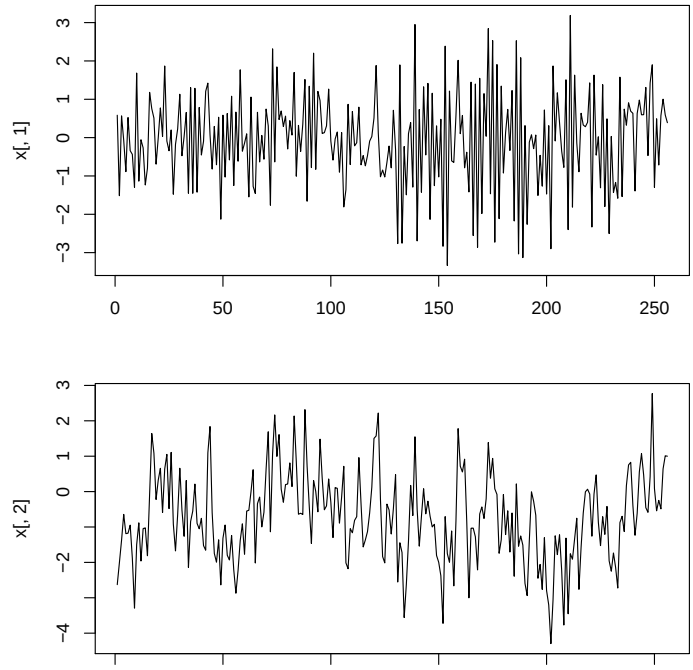

Figure 1: Example for simulating from a FIVARMA model.

```
R> par(mfrow = c(2, 1), mai = c(0.5, 1, 0.5, 0.5))
R> plot(x[, 1], type = "l", lty = 1)
R> plot(x[, 2], type = "l", lty = 1)
```

The resulting plot is shown in Figure 1.

3.2. A real data set

In order to describe how parameters can be chosen from a practical point of view, we provide a real data example (see Section 7).

Noninvasive data recorded from the brain are an example where the proposed methodology is efficient. The data consist of time series recording signals from the brain: electroencephalography (EEG) for the electrical signals, magnetoencephalography (MEG) for the magnetic signals or functional magnetic resonance imaging (fMRI) for the blood oxygen level dependent (BOLD) signals. These data are intrinsically correlated because of the known interactions of the brain areas (also called regions of interest). Furthermore, it has already been shown that these time series present long-memory features (Maxim, Şendur, Fadili, Suckling, Gould, Howard, and Bullmore 2005). Other data sets presenting similar features are coming from finance (e.g., Songsiri and Vandenberghe 2010), where time series are correlated because of links between companies for example, and they also present long-memory characteristics. In this section, we observed time series extracted using fMRI facilities. The whole description of this data sets is detailed in Termenon, Jaillard, Delon-Martin, and Achard (2016). The data set called `brainHCP` contains the time series of 1200 points in time and 89 regions of the brain. Figure 2 displays 6 arbitrary signals from one subject in this data set.

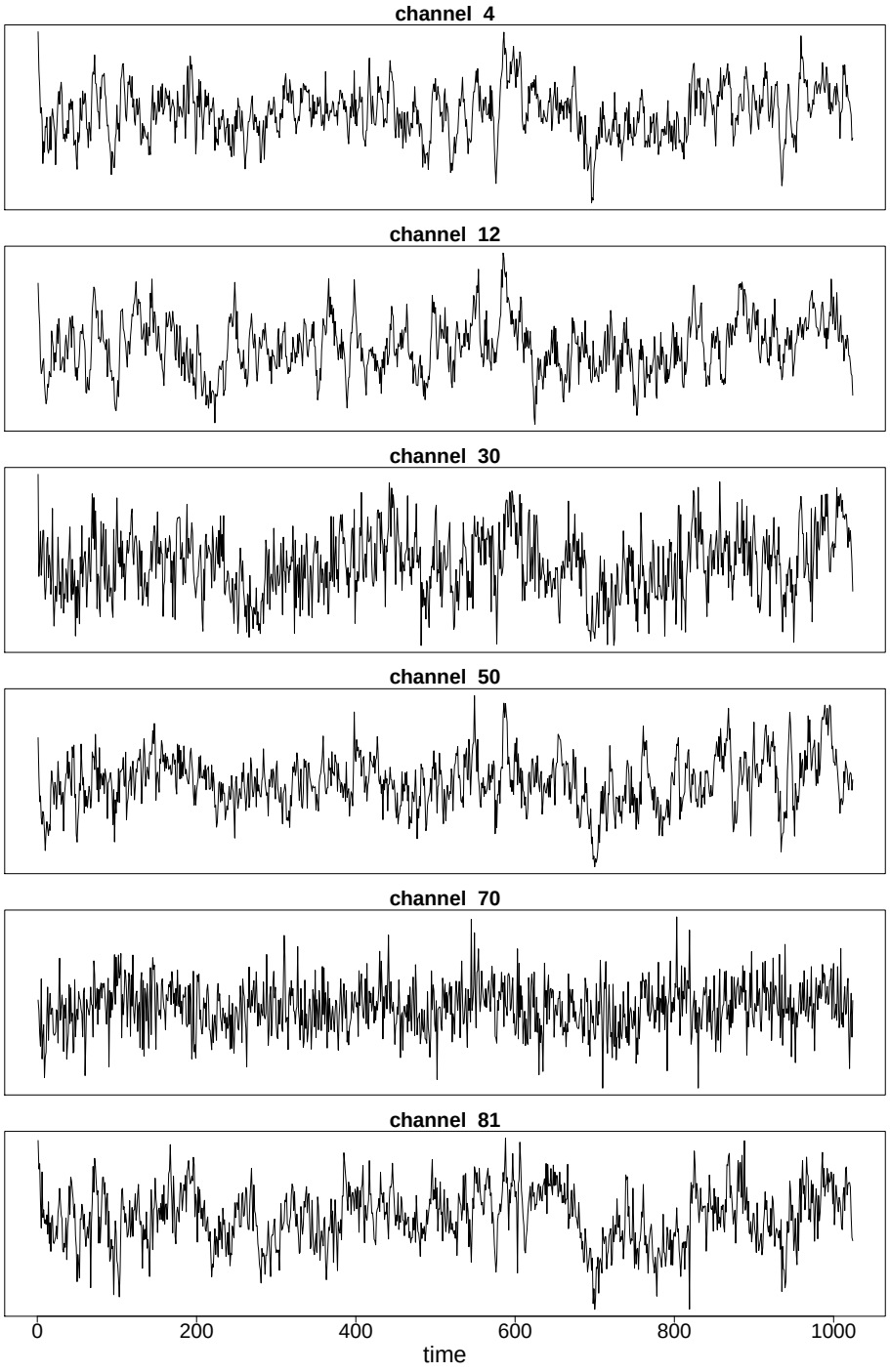

Figure 2: Plot of 6 arbitrary signals from a subject of the fMRI data set `brainHCP`.

```
R> data("brainHCP", package = "multiwave")
R> dim(brainHCP)

[1] 1200 89
```

4. Estimating long-memory parameters and the covariance

The objective of this package is to provide the implementation of two sets of methods based either on Fourier or wavelet decomposition. That is the multivariate Fourier Whittle (MFW) and multivariate wavelet Whittle (MWW) estimation procedures. The corresponding functions are respectively called `mfw` and `mww` in the package.

The output of the implemented methods consists in the estimation of two quantities, $\mathbf{d}$ and $\mathbf{\Omega}$, where $\mathbf{d}$ corresponds to the long-memory parameters of the time series and $\mathbf{\Omega}$ is reflecting the coupling between the pairs of time series.

The two functions `mfw` and `mww` are implementing the semi-parametric Whittle estimation using respectively Fourier decomposition and wavelet decomposition in order to estimate $\mathbf{d}$ and $\mathbf{\Omega}$.

A fast execution of Fourier-based estimation is given below using the default value used in Shimotsu (2007) for m, the number of frequencies.

```
R> N <- nrow(x)
R> m <- floor(N^(0.65))
R> res_mfw <- mfw(x, m)
```

In addition, the wavelet-based estimation is computed using the following code. `res_filter` corresponds to the choice of the wavelet filter. `LU` fixes respectively the lowest and highest scales used in the estimation.

```
R> res_filter <- scaling_filter("Daubechies", 8)
R> filter <- res_filter$h
```

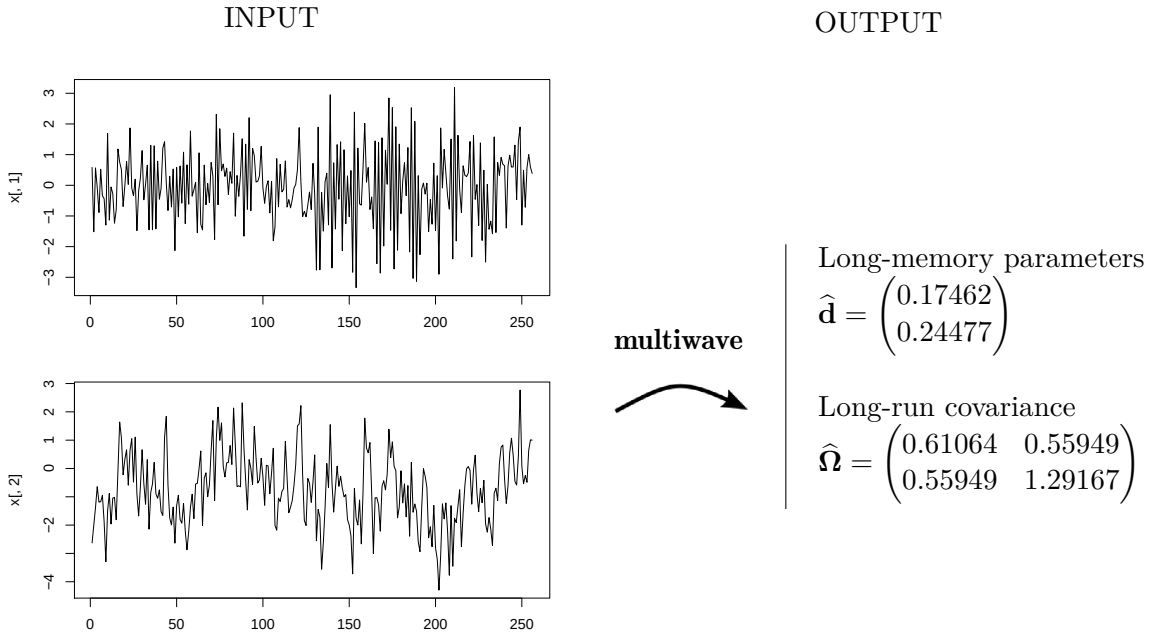

Figure 3: Input and output of the **multiwave** package in a two-dimensional case.

```
R> LU <- c(2, 11)
R> res_mww <- mww(x, filter, LU)
```

4.1. Multivariate Fourier Whittle estimation

As a first implementation, it is natural to use Fourier decomposition to approximate the spectral density of time series.

Package **multiwave** proposes functions to compute MFW estimators:

- `mfw` computes the multivariate Fourier Whittle estimators of both the long-range dependence parameters and the long-run covariance matrix.
- `mfw_cov_eval` computes the multivariate Fourier-based Whittle estimator for the long-run covariance matrix for a given value of the long-range dependence $\mathbf{d}$.
- `mfw_eval` returns the value of the multivariate Fourier Whittle criterion with respect to $\mathbf{d}$ at a given value of $\mathbf{d}$.

The functions `mfw_cov_eval` and `mfw_eval` are internal functions of `mfw`. In `mfw`, we apply first a minimum search of `mfw_eval` with respect to $\mathbf{d}$, and `mfw_cov_eval` is returning the estimation of $\mathbf{\Omega}$ for the estimated value $\mathbf{d}$.

We only detail function `mfw` hereafter and refer to the package description for the other functions.

Let $\mathbf{X}$ be the $p \times N$-matrix of observations, with general term $x_{\ell,i} = X_\ell(i)$, $\ell = 1, \ldots, p$ and $i = 1, \ldots, N$. Let m be the number of frequencies used in the MFW procedure. Given x and m, the function `mfw` computes the MFW estimators defined by (5) and (6), with the frequencies $\lambda_j = 2\pi j/N$, $j = 1, \ldots, m$. The optimization in equation (5) is done using `optimize` function of R in one-dimensional settings and a Newton-type algorithm through `nlm` function of R otherwise. The initialization of the algorithm is set equal to the vector of univariate Fourier-based Whittle estimations. Even if it increases the computational time, such an initialization is important in high-dimensional settings. For example, in the MEG data set studied in Achard and Gannaz (2016), the optimization is done in $\mathbb{R}^{274}$. An initialization at the origin may not be able to reach the minimum even with a high number of iterations, due to the high dimension.

Function `mfw` returns a list with first the p-dimensional vector $\widehat{\mathbf{d}}^{\mathrm{MFW}}$ and second the $p \times p$-matrix $\widehat{\mathbf{\Omega}}^{\mathrm{MFW}}$. The quality of the estimation is depending on the parameter m. Theoretical results show that m must be small enough so that the short-range properties of the time series do not bias the estimation. On the contrary a too small value will introduce variance in the estimation since it decreases the number of frequencies used in the procedure. Lobato (1999), Shimotsu (2007) or Nielsen (2011) propose as default value $m = N^{0.65}$. This choice is discussed in the simulation study in Section 6.

Example

```
R> N <- 2^8
R> d0 <- c(0.2, 0.4)
```

```
R> rho <- 0.8
R> cov <- matrix(c(1, rho, rho, 1), 2, 2)
R> VMA <- diag(c(0.4, 0.7))
R> VAR <- array(c(0.8, 0.2, 0, 0.6), dim = c(2, 2))
R> resp <- fivarma(N, d0, cov_matrix = cov, VAR = VAR, VMA = VMA)
R> x <- resp$x
R> m <- floor(N^(0.65))
R> res_mfw <- mfw(x, m)
R> res_mfw

$d
[1] 0.1669139 0.4601785

$cov
          [,1]      [,2]
[1,] 0.5678479 0.5645579
[2,] 0.5645579 1.0308766
```

4.2. Multivariate wavelet Whittle estimation

The functions applying MWW estimation in package **multiwave** are the following:

- `mww` computes the multivariate wavelet Whittle estimators of the long-range dependence parameters and the long-run covariance matrix.
- `mww_cov_eval` computes the multivariate wavelet-based Whittle estimator for the long-run covariance matrix for a given value of the long-range dependence $\mathbf{d}$.
- `mww_eval` returns the value of the multivariate wavelet Whittle criterion with respect to $\mathbf{d}$ at a given value of $\mathbf{d}$.

`mww_cov_eval` and `mww_eval` are internal functions of `mww`. In `mww`, we apply first a minimum search of `mww_eval` with respect to $\mathbf{d}$, and `mww_cov_eval` is returning the estimation of $\mathbf{\Omega}$ for the estimated value of $\mathbf{d}$. MWW estimation is based on the wavelet transform of time series and `mww` needs the definition of a wavelet filter. The computation of a filter and of a wavelet transform are described below.

Wavelet transform and scalogram

The wavelet decomposition in package **multiwave** is implemented using an exact discrete wavelet transform.

- `scalingfilter` defines the wavelet filter (only Daubechies' wavelets are available).
- `computenj` computes the number of wavelet coefficients for each individual scale.
- `DWTexact` provides the wavelet transform of the data.
- `psi_hat_exact` gives the Fourier transform of the wavelet function.
- `K_eval` evaluates the value of the integral (8).

Example

To obtain the wavelet filter of a Daubechies' wavelet of order 4, that is with 2 vanishing moments, one should write:

```
R> res_filter <- scaling_filter("Daubechies", 4);
R> filter <- res_filter$h
R> filter

[1]  0.4829629  0.8365163  0.2241439 -0.1294095
```

Next, given an N-dimensional vector `x`, the wavelet coefficients of `x` are given by function `DWTexact`:

```
R> N <- 2^8
R> d0 <- 0.2
R> resp <- fivarma(N, d0)
R> x <- resp$x
R> resw <- DWTexact(x, filter)
R> xwav <- resw$dwt
R> index <- resw$indmaxband
R> Jmax <- resw$Jmax
R> Jmax

[1] 6

R> index

[1] 127 189 219 233 239 241

R> length(x)

[1] 256
```

`index` gives the index of the last coefficient at each scale and `Jmax` gives the maximal scale. The vector of coefficients `xwav` is m-dimensional, with m the maximum of `index`, equal to `index[Jmax]`. The coefficients of the third scale are for example given by:

```
R> xwav[seq(index[2] + 1, index[3]), 1]
```

Finally it is useful to compute the quantity $K(\delta)$ defined in (8). Thus one needs to recover the Fourier transform of the wavelet, $\widehat{\psi}(\cdot)$. This is done using the function `psi_hat_exact`. Its inputs are the filter defined previously and an index of precision. It returns $(\widehat{\psi}(u_i))_{i=1,\ldots,q\cdot 2^J}$ where q is the length of the filter and u_i are equally spaced points on the interval $[-\pi 2^{J-3}(q-1)/2;\ \pi 2^{J-3}(q-1)/2]$.

```
R> res_psi <- psi_hat_exact(filter, J = 10)
R> psih <- res_psi$psih
R> gridh <- res_psi$grid
```

where `res_psi$grid` returns the values of the grid (u_i) and `res_psi$psih` returns the corresponding values of $\widehat{\psi}(u_i)$. It is recommended to take $J \leqslant 15$ in practice and the default value is $J = 10$. Indeed, a large value of `J` is increasing the computational time.
Given the function $\widehat{\psi}(\cdot)$, we are now able to evaluate $K(\mathbf{d})$ for a given value of $\mathbf{d}$:

```
R> K <- K_eval(psih, gridh, d0)
```

Estimation

`mww` is now described in detail. Let $\mathbf{X}$ be the $p \times N$-matrix of observations, with general entries $x_{\ell,i} = X_\ell(i)$, $\ell = 1, \ldots, p$ and $i = 1, \ldots, N$. Let LU be the bivariate vector giving the lowest scale j_0 and the upper scale j_1 of the wavelet coefficients used in estimation. Given $\mathbf{X}$ and LU, the function `mww` computes the MWW estimators defined by (10) and (12). As previously, the optimization in (10) is done using the `optimize` function of R in the one-dimensional settings and using a Newton-type algorithm through the `nlm` function of R otherwise. The initialization of the algorithm is set equal to the vector of univariate wavelet-based Whittle estimations. The reasons are identical to the ones given for the function `mfw`.

Function `mww` returns a list with first the p-dimensional vector $\widehat{\mathbf{d}}^{\mathrm{MWW}}$ and second the $p \times p$-matrix $\widehat{\mathbf{\Omega}}^{\mathrm{MWW}}$. The quality of the estimation is depending on the parameters j_0 and j_1 appearing in (10) and (11). The default value for j_0 is set to 2 and for j_1 to the highest integer lower than $\log_2(N)$. The critical value to choose for estimation is j_0, as can be seen in the theoretical conditions for consistency (Achard and Gannaz 2016) and in simulations studies. Similarly to the choice of the parameter m for the MFW procedure, a compromise is required between choosing a small value of j_0, which would introduce a bias due to the short-range properties of the time series, and a high value that will reduce the number of frequencies and thus increase variance.

Example

```
R> N <- 2^8
R> d0 <- c(0.2, 0.4)
R> rho <- 0.8
R> cov <- matrix(c(1, rho, rho, 1), 2, 2)
R> VMA <- diag(c(0.4, 0.7))
R> VAR <- array(c(0.8, 0.2, 0, 0.6), dim = c(2, 2))
R> resp <- fivarma(N, d0, cov_matrix = cov, VAR = VAR, VMA = VMA)
R> x <- resp$x
R> res_filter <- scaling_filter("Daubechies", 8);
R> filter <- res_filter$h
R> LU <- c(2, 8)
R> res_mww <- mww(x, filter, LU)
R> res_mww

$d
[1] 0.01222985 0.32778571
```

```
$cov
          [,1]      [,2]
[1,] 0.6614721 0.7003364
[2,] 0.7003364 1.1873687
```

If one wants to apply several times the estimation on the same data set, or modifying the parameters of estimation, it is useful to separate the wavelet transform and the estimation scheme. Wavelet-based estimation can be evaluated directly on the wavelet transform of the data using the following functions:

- `mww_wav` computes the multivariate wavelet Whittle estimators of the long-range dependence parameters and the long-run covariance matrix, given the wavelet transform of the data.
- `mww_wav_cov_eval` computes the MWW estimator for the long-run covariance matrix for a given value of the long-range dependence $\mathbf{d}$, given the wavelet transform of the data.
- `mww_wav_eval` returns the value of the multivariate wavelet-based Whittle criterion with respect to $\mathbf{d}$ at a given value of $\mathbf{d}$, for a specific wavelet transform of the data.

We refer to the description of the functions in the package for more details.

Example

The following gives code for data simulation.

```
R> N <- 2^8
R> d0 <- c(0.2, 0.4)
R> rho <- 0.8
R> cov <- matrix(c(1, rho, rho, 1), 2, 2)
R> VMA <- diag(c(0.4, 0.7))
R> VAR <- array(c(0.8, 0.2, 0, 0.6), dim = c(2, 2))
R> resp <- fivarma(N, d0, cov_matrix = cov, VAR = VAR, VMA = VMA)
R> x <- resp$x
R> N <- dim(x)[1]
R> k <- dim(x)[2]
```

The following code performs wavelet decomposition.

```
R> res_filter <- scaling_filter("Daubechies", 8)
R> filter <- res_filter$h
R> LU <- c(2, 8)
R> xwav <- matrix(0, N, k)
R> for (j in 1:k) {
+    xx <- x[, j]
+    resw <- DWTexact(xx, filter)
+    xwav_temp <- resw$dwt
+    index <- resw$indmaxband
```

```
+    Jmax <- resw$Jmax
+    xwav[1:index[Jmax], j] <- xwav_temp
+  }
R> new_xwav <- matrix(0, min(index[Jmax], N), k)
R> if (index[Jmax] < N) {
+     new_xwav[(1:(index[Jmax])), ] <- xwav[(1:(index[Jmax])), ]
+  }
R> xwav <- new_xwav
R> index <- c(0, index)
R> res_psi <- psi_hat_exact(filter, Jmax)
R> psih <- res_psi$psih
R> grid <- res_psi$grid
```

Next we provide the code for estimation.

```
R> res_mww_wav <- mww_wav(xwav, index, psih, grid, LU)
R> res_mww_wav

$d
[1] -0.0245506  0.3372074

$cov
          [,1]      [,2]
[1,] 0.6856003 0.5001938
[2,] 0.5001938 1.0366141
```

5. Practical choices of parameters for MWW estimation

The MWW procedure implemented in `mww` depends on mainly two parameters: the choice of the wavelet bases `filter` and the choice of the wavelet scales `LU`. These parameters are not fixed in the package as the performances of the estimation may be improved by a careful choice. The possibility to choose the wavelet scales is particularly of interest when dealing with short-range dependence. We show that a simple graphical representation of the scalogram is able to guide the user in the choice of the wavelet scales.

5.1. Choice of the wavelet bases

Actually **multiwave** only proposes Daubechies' wavelets, which satisfy theoretical properties of Achard and Gannaz (2016). Other bases are possible but not implemented. The wavelet bases is imputed *via* the parameter `filter` in function `mww`, where `filter` is obtained by `filter <- scaling_filter("Daubechies", 2 * M)$h`. The main parameter characterizing the Daubechies' bases is thus the number of vanishing moments, M. MWW estimation presents the advantage to be available even if the time series are nonstationary or with polynomial trends, as soon as $\max \mathbf{d} \leqslant M$. For example, for stationary time series, a parameter $M = 1$ is sufficient (which is equivalent to considering Haar bases). For real data applications, when nonstationarity or trends are suspected, a higher value of M is necessary.

As discussed in Faÿ *et al.* (2009), when M increases, the quality of estimation (slightly) decreases. Depending on the data, a compromise is then needed between choosing a large enough number of vanishing moments M to handle nonstationarity in the data and the quality of estimation.

On the contrary, MFW estimators are only suited to stationary time series. Some extensions of Fourier-based estimation were proposed in univariate settings such as tapered Fourier (see, e.g., Faÿ *et al.* 2009 and references therein). For multivariate estimation Nielsen (2011) proposes an extension of Shimotsu (2007) based on the transform defined in Abadir, Distaso, and Giraitis (2007). However, this approach gives satisfactory results only for $d < 1.5$ and we decided not to implement it in the package for simplicity.

5.2. Choice of wavelet scales

The second parameter we need to tune is `LU`, which corresponds to the range of scales used in estimation. `LU` is a two-dimensional vector, that is `LU <- c(j0, j1)`, with `j0` the lowest scale and `j1` the upper scale. Parameters `j0` and `j1` are respectively j_0 and j_1 defined in (10) and (11) in the estimation procedure.

One advantage of wavelets is to be able to qualitatively evaluate the choice of wavelet scales to estimate the long-memory parameters and correlation by inspection of the wavelet scalogram. As mentioned in Abry and Veitch (1998); Faÿ *et al.* (2009) for univariate settings, the first and last scales may have to be discarded from the analysis. The first scale may be affected by the presence of short-memory phenomena. In the example of the FIVARMA model, this is driven by the AR and MA coefficients. For the last scales, the impact is different and it comes from the finite length of the time series. Indeed, as derived in Whitcher *et al.* (2000), the variance of the estimator is increasing with the wavelet scales.

The usual log-scalogram diagram used in univariate settings is showing the linear behavior of the log variance with respect to the wavelet scales (Abry and Veitch 1998). This is also true for the covariances as shown in Proposition 2 of Achard and Gannaz (2016). For all $k \in \mathbb{Z}$, $\mathsf{COV}(W_{j,k}(\ell), W_{j,k}(m))$ is equivalent to $2^{j(d_\ell + d_m)} G_{\ell,m}(\mathbf{d})$ when j goes to infinity, with $G_{\ell,m}(\mathbf{d})$ defined in (9). This property is illustrated in Figure 4 with a bivariate FIVARMA processes, as described in Section 2.1. This figure represents the boxplots of the variance of the wavelet coefficients of each component of the time series at each scales (Figures 4(a) and 4(b)). These plots correspond to the usual log-scalogram diagram (Abry and Veitch 1998). Figure 4(c) displays the analog representation of the covariance between the components. For both variance and covariance, the points satisfying the above approximation are aligned. Scales corresponding to non-aligned points should be removed from estimation as can be seen in Figures 4(a) and 4(c) where the highest frequencies are modified by the presence of short-range dependence. Thus, one may discard the first scale from estimation to improve its quality.

Using the results on the wavelet variance and covariance in terms of scales, we showed that the wavelet correlation is asymptotically constant with respect to the wavelet scales. Indeed, when j goes to infinity, for all $\ell, m = 1, \ldots, p$, for all $k \in \mathbb{Z}$, $\mathsf{COR}(W_{j,k}(\ell), W_{j,k}(m))$ is equivalent to $G_{\ell,m}(\mathbf{d}) / \sqrt{G_{\ell,\ell}(\mathbf{d}) G_{m,m}(\mathbf{d})}$ (Achard and Gannaz 2016). As for the log-scalogram diagram, the correlation between wavelet coefficients with respect to the scales can be plotted and scales where the observed correlation is not equal to the value obtained for the majority of scales should be removed from estimation. The wavelet correlation spectrum is a complementary

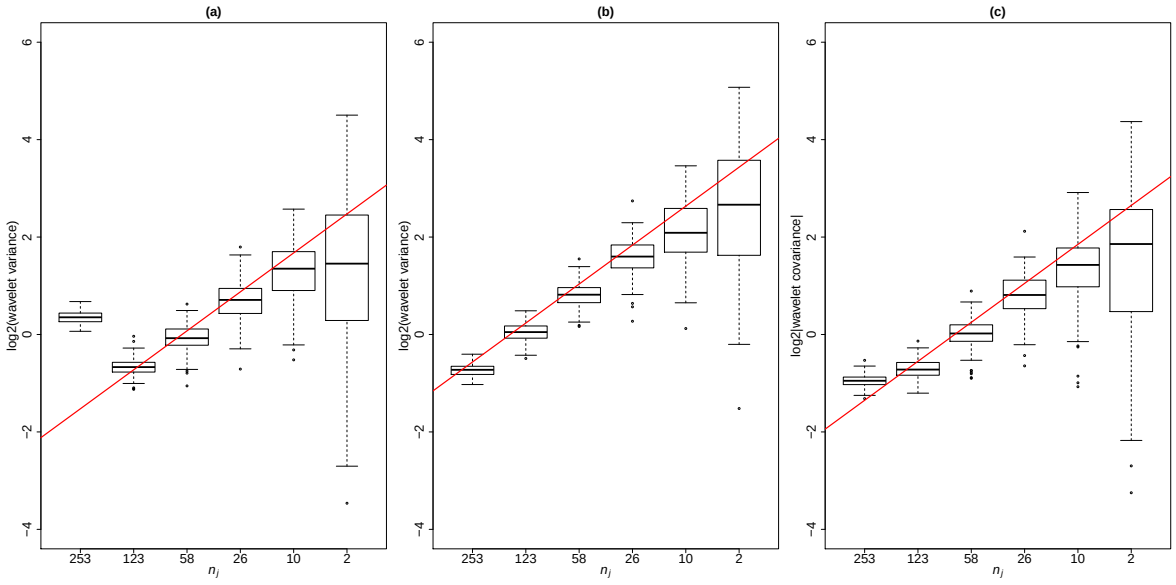

Figure 4: Boxplots of the log2-variance of the wavelet coefficients at different scales for simulated bivariate FIVARMA$(1,(0.4,0.4),1)$; (a) for the first component and (b) for the second component. (c) Boxplots of the log2-absolute covariance for the same data. The red lines represent the theoretical linear prediction given by $j(d_\ell+d_m)+\log_2(G_{\ell,m}(\mathbf{d}))$, with $\ell=m=1$ for Figure (a), $\ell=m=2$ for Figure (b) and $\ell=1$, $m=2$ for Figure (c). The horizontal axis corresponds to increasing scales, that is, decreasing frequencies. The indexes of the horizontal axis display the number of coefficients available. Parameters in FIVARMA model were the following: the white noise is Gaussian with a covariance equal to $\mathbf{\Sigma}=\begin{pmatrix}1 & 0.8\\ 0.8 & 1\end{pmatrix}$, the AR coefficient is set equal to $A=\begin{pmatrix}0.8 & 0\\ 0.2 & 0.6\end{pmatrix}$ and the MA coefficient is set equal to $B=\begin{pmatrix}0.4 & 0\\ 0.2 & 0.7\end{pmatrix}$. Calculation was done on $N=512$ observations for 100 replications.

way to qualitatively evaluate the range of scales where the analysis should be carried out. Figure 5 illustrates this on four different data sets. Four different simulations of bivariate processes are applied using finite difference processes, FIVAR and FIVARMA processes, as described in Section 2.1. Figure 5 represents the boxplots of the correlation between wavelet coefficients of the two components of the time series at each scales. Again the presence of short-range dependence alters the highest frequencies (Figures 5(b) and 5(c)). This is also observed with nonstationarity (Figure 5(d)). The first scales should then be removed from estimation. This free parameter of the package is particularly useful with the presence of short-range dependence or nonstationarity. Visual comparison to constant values may be easier for selection of the correct range of wavelet scale to use in the estimation.

A similar discussion is detailed in Section 7 for a real neuroscience data set. With real data sets a bootstrap procedure is necessary to obtain boxplots, as will be explained in Section 7. For the Fourier procedure, the equivalent parameter is the number of frequencies m. However, wavelets are providing a graphical way to choose the upper and lower scales. To our knowledge, no equivalent qualitative evaluation for the Fourier procedure is available.

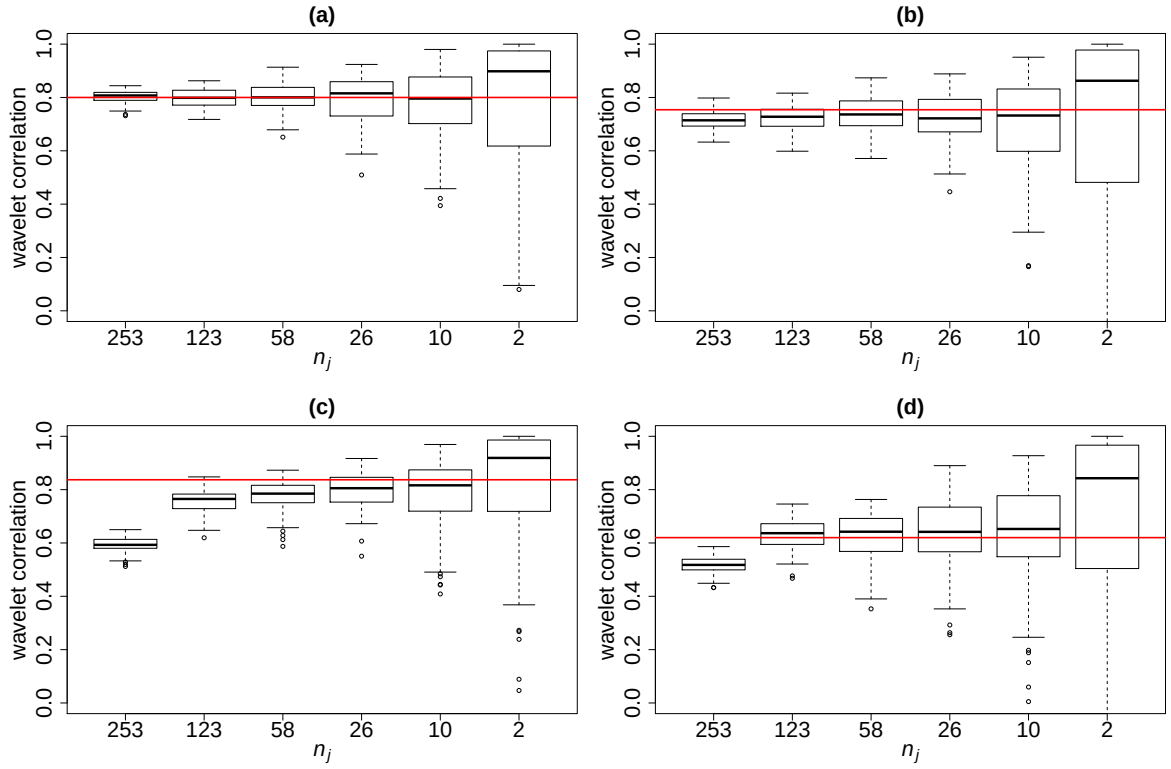

Figure 5: Boxplots of the correlation of the wavelet coefficients at different scales for four different simulated data: (a) bivariate FIVARMA$(0,(0.4,0.4),0)$ (b) bivariate FIVARMA$(1,(0.4,0.4),0)$; (c) bivariate FIVARMA$(1,(0.4,0.4),1)$; and finally, (d) an example with nonstationary time series, FIVARMA$(0,(0.8,1.2),0)$. The horizontal red lines represent the true long-run correlation for each simulation. The horizontal axis corresponds to increasing scales, that is, decreasing frequencies. The indexes of the horizontal axis display the number of coefficients available. Parameters in FIVARMA models were, if needed, the following: the white noise covariance is set equal to $\mathbf{\Sigma} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$, the AR coefficient is set equal to $A = \begin{pmatrix} 0.8 & 0 \\ 0.2 & 0.6 \end{pmatrix}$ and the MA coefficient is set equal to $B = \begin{pmatrix} 0.4 & 0 \\ 0.2 & 0.7 \end{pmatrix}$. Calculation was done on $N = 512$ observations for 100 replications.

5.3. Numerical examples

In order to quantify the quality of the choice of the parameters, numerical examples are provided. The first tables illustrate the quality of the estimators for a long-memory process with no short-range dependence. Then, the simulations are complexified by adding short-range behavior or nonstationarity.

In each example we simulated 1000 Monte Carlo replications of $N = 512$ observations from FIVARMA(q,d,r), with dimension $p = 2$, for a set of different parameter values. Simulations are done using the function `fivarma`. MWW estimators are computed using function `mww`. The MWW procedure is applied using a Daubechies' wavelet bases with $M = 4$ vanishing

moments. This choice is motivated by the discussion in Section 5.1, because it can handle different settings, including nonstationary ones.

The quality of estimation is measured *via* the bias, the standard deviation (std) and the root mean square error (RMSE) which is equal to $\sqrt{\text{bias}^2 + \text{std}^2}$. For clarity, all tables are displayed in Appendix A.

A reference example

We first consider a simple example, with neither short-range dependence, nor nonstationarity. Time series were simulated using a bivariate FIVARMA$(0, \mathbf{d}, 0)$ with a long-run correlation matrix $\mathbf{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\rho = 0.8$. The bivariate vector $\mathbf{d}$ is chosen in $[0, 0.5)^2$, such that the time series are stationary.

As shown in Figure 5(a), and discussed in Section 5.2, all scales can be kept for estimation. Table 2 displays results for the MWW estimation of $\mathbf{d}$. This illustrates that multivariate estimation improves the quality of estimation for $\mathbf{d}$. Indeed, the last column gives the ratio between the RMSEs of the multivariate wavelet-based estimation and of the univariate wavelet-based estimation (ratio M/U). This ratio is always smaller than 1, that is, multivariate RMSE is always lower than univariate RMSE.

Short-range dependence

Consider a FIVARMA$(1, \mathbf{d}, 0)$ obtained with the model described in Section 2.1 and given by the function `fivarma`. This case corresponds to a FIVAR model of Sela and Hurvich (2012). The AR coefficient is taken equal to $\mathbf{A} = \begin{pmatrix} 0.8 & 0 \\ 0.2 & 0.6 \end{pmatrix}$ and the correlation between the innovation processes equal to $\rho = 0.8$. More precisely let $\boldsymbol{\varepsilon}$ be a bivariate white noise process with covariance matrix $\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and let $\mathbf{u}$ be the AR(1) process defined by $\mathbf{u}(t) + \mathbf{A}\mathbf{u}(t-1) = \boldsymbol{\varepsilon}(t)$. 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 (4) is equal to

$$\mathbf{\Omega} = (I + \mathbf{A})^{-1} \mathbf{\Sigma} (I + \mathbf{A})^{-1\,T} \simeq \begin{pmatrix} 0.3086 & 0.2392 \\ 0.2392 & 0.3260 \end{pmatrix}.$$

The corresponding long-run correlation is thus equal to 0.754.

As explained above, the finest scales are influenced by the short-range dependence and they have to be discarded from the estimation. As can be seen in Figure 5(c), the first two scales should be removed. We obtained accordingly that the lowest RMSE in estimation is obtained taking $j_0 = 3$.

Numerical results for estimation are given in Tables 4 and 5. We can observe that estimation of $\widehat{\mathbf{d}}$ and $\widehat{\mathbf{\Omega}}$ is still satisfactory. The RMSE is very similar to the previous case with no short-range behavior.

Nonstationarity

Nonstationary examples are simulated using values of $\mathbf{d}$ higher than 0.5. We consider order 1 or 2 of nonstationarity, that is $\mathbf{d} \in [0.5, 2.5)^2$.

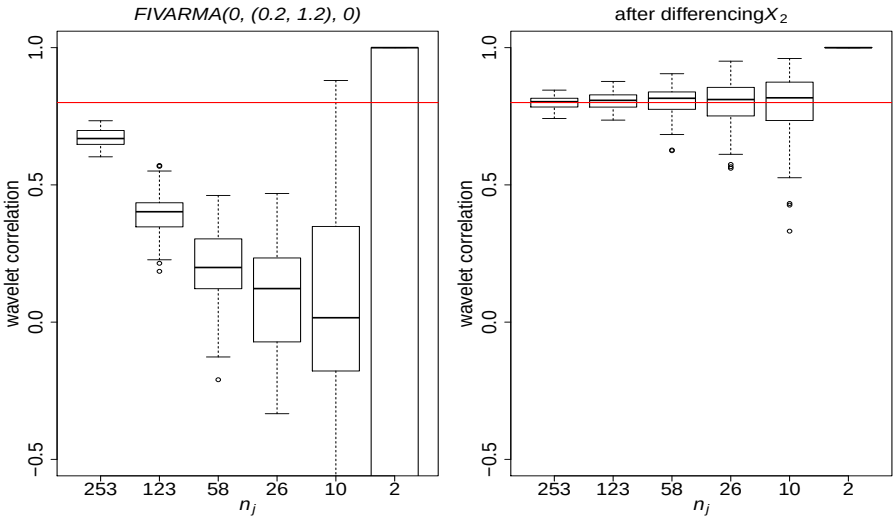

Figure 6: Boxplots of the covariance of the wavelet coefficients at different scales for a bivariate FIVARMA$(0, (0.2, 1.2), 0)$ with $\mathbf{\Omega} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$. The index of the horizontal axis displays the number of coefficients available. Calculation was done on $N = 512$ observations for 1000 replications.

The behavior of the wavelet correlations at each scale is illustrated in Figure 5(d). Contrary to stationary simulations where the optimal choice of j_0 was equal to $j_0 = 1$, the optimal choice of the parameter j_0 is $j_0 = 2$. Results are given in Tables 6 and 7.

Comparing Tables 2 and 6, the quality of estimation of $\mathbf{d}$ is still accurate in nonstationary settings, with similar values for the RMSE. As for the estimation of $\mathbf{\Omega}$, Tables 3 and 7 indicate that MWW still provides a good quality estimation of the long-run covariance matrix. The quality is slightly lower but still satisfactory.

5.4. Discussion on identifiability

In practical applications, it seems natural to assume that time series have the same order of stationarity. However, when two time series have long-memory parameters d_ℓ and d_m satisfying $d_\ell - d_m = 1$ the long-run covariance matrix $\mathbf{\Omega}$ is no longer identifiable with the wavelet-based procedure. Indeed, Proposition 2 in Achard and Gannaz (2016) states that in this particular case the covariance $\mathsf{COV}(W_{j,k}(\ell), W_{j,k}(m))$ tends to 0 when the scale j tends to infinity. Figure 6 illustrates this approximation for a bivariate FIVARMA$(0, (0.2, 1.2), 0)$, a correlation matrix $\mathbf{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\rho = 0.8$.

When $\widehat{d}_\ell - \widehat{d}_m = 1$, the estimator (12) is no longer defined. In practice, the quantity $\widehat{d}_\ell - \widehat{d}_m$ cannot be exactly equal to 1. Nevertheless, as dividing by a cosine function of this difference, a small error in the estimation of (d_ℓ, d_m) will lead to an important bias in the estimation of $\Omega_{\ell,m}$. As can be seen in Figure 7, the resulting bias increases in the neighborhood of the non-identifiable lines $d_\ell - d_m = \pm 1$.

When this situation occurs, say when the difference between $d_\ell - d_m$ is between 0.75 and

(a) (b)

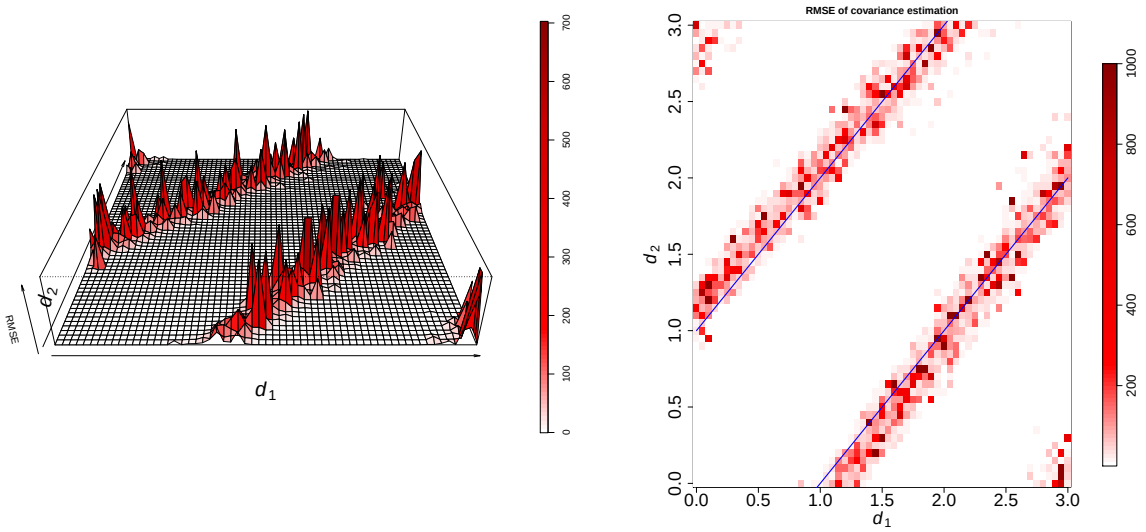

Figure 7: RMSE in the estimation of the cross-covariance term Ω_{12} with respect to (d_1, d_2). Estimation was done using the multivariate wavelet Whittle estimator for a bivariate FIVARMA$(0, (d_1, d_2), 0)$ with $\mathbf{\Omega} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$. Figure (b) represents an image plot of Figure (a) (with a different color scale to improve visual quality). Blue lines in Figure (b) correspond to $d_2 - d_1 = \pm 1$. Calculation was done on $N = 512$ observations for 1000 replications.

	Without differentiation			With differentiation		
$\mathbf{\Omega}$	*bias*	*std*	*RMSE*	*bias*	*std*	*RMSE*
$\Omega_{1,1}$	0.0935	0.0762	0.1206	0.0349	0.0678	0.0762
$\Omega_{1,2}$	4.1863	7.2103	8.3375	0.0261	0.06	0.0654
$\Omega_{2,2}$	0.2215	0.0819	0.2362	0.0292	0.07	0.0758
correlation	3.5255	6.2935	7.2137	0.0003	0.0155	0.0155

Table 1: Multivariate wavelet Whittle estimation of $\mathbf{\Omega}$ for a bivariate FIVARMA$(0, (0.2, 1.2), 0)$ with $\rho = 0.8$, $N = 512$ with 1000 repetitions. As the memory parameter of the second component is greater than one, one possibility is to differentiate the second component. j_0 is chosen to be equal to 1.

1.25, the estimation of $\mathbf{d}$ is not affected. But the user must be careful for the estimation of $\mathbf{\Omega}$. One solution is to differentiate or integrate one of the two processes. For example, Table 1 illustrates the non-identifiability of $\mathbf{\Omega}$ in a bivariate FIVARMA$(0, \begin{pmatrix} 0.2 & 1.2 \end{pmatrix}, 0)$. When differentiating the second component (with $d_2 = 1.2$) the estimator has again good performances.

6. MFW estimation and comparison with MWW

The comparison between Fourier-based and wavelet-based approach is presented now. Time series were simulated using a bivariate FIVARMA$(0, \mathbf{d}, 0)$ with a long-run correlation matrix $\mathbf{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\rho = 0.8$. The bivariate vector $\mathbf{d}$ is chosen in $[0, 0.5)^2$, such that the time series are stationary. In such a setting, Fourier-based estimators are available. For Fourier-based approach, the parameter to choose is `m` corresponding to the number of frequencies taken into account in the estimation. The default value in Shimotsu (2007) is $m = N^{0.65}$. We also make comparisons with an optimal value computed by minimizing the RMSE.

Estimation of the long-memory parameters

Table 8 gives the results obtained for the estimation of $\mathbf{d}$ using the MFW procedure (to be compared to Table 2 for wavelets). Comparison between Fourier and wavelet-based procedures is summarized by the ratio between the RMSE given by MWW estimation and the RMSE given by MFW estimation, denoted by the ratio W/F. Taking the same number of frequencies as Shimotsu (2007), that is, $m = N^{0.65}$, Table 8 shows that the quality of the MWW and MFW procedures are comparable, even if wavelet-based estimation slightly improves Fourier-based estimation with such a choice of m.

Next we also consider the number of frequencies leading to the minimal RMSE for MFW estimation. As can be seen in Table 8, qualities of both procedures are very similar but MFW estimation then (slightly) surpasses MWW estimation.

Very precise comparisons of Fourier-based and wavelet-based approaches are described in Faÿ *et al.* (2009) for a univariate setting. In particular, it is shown that since the time series are stationary, the use of Haar bases should improve the MWW quality. The authors indeed obtained better results with the Haar-based procedure than with the Fourier-based procedure. To highlight the versatility of wavelet-based procedures, we choose here wavelet bases with four vanishing moments. A fair comparison with a Fourier-based method should consider tapered Fourier of order 4 as detailed in Faÿ *et al.* (2009). They quantify the influence of the regularity of the wavelet bases and discuss the comparison with (tapered) Fourier bases. Similar results are expected to be obtained in the case of multivariate time series, however, this topic exceeds the scope of this paper.

Estimation of the long-run covariance

Finally, Table 9 displays results for the estimation of $\mathbf{\Omega}$ with the MFW method (to be compared to Table 3 for wavelets). When the MFW estimation is applied with the usual number $m = N^{0.65}$ of frequencies, one can see that the wavelet-based procedure still estimates better the long-run covariance and the long-run correlation, with a ratio W/F always lower than 0.6 for the estimation of the $\mathbf{\Omega}$ terms. When the number of frequencies in the Fourier-based estimation is chosen optimally, the MFW and MWW procedures behave similarly and none appears significantly better than the other.

To conclude, the MFW and MWW estimation procedures give very similar results. The slight improvement of Fourier-based procedure for the estimation of $\mathbf{d}$ can be explained by the choice of the wavelet bases, however wavelets are efficient for a large set of applications, including time series with trends and nonstationarity features.

7. Application on real neuroscience data

As already shown in Figure 5 for simulated data, the advantage of representing the wavelet correlation in terms of scale is to qualitatively determine the scales necessary to estimate the long-memory parameters and long-range covariance matrix. When dealing with real data, the bootstrap is providing a way to assess the variability of the estimators. Using the real data described in Section 3.2, sliding overlapping windows of the time series were extracted containing 512 points and we repeated the estimation until reaching the final point of the time series. This is illustrated in Figure 8, where an example of four pairs of fMRI data from one subject is presented. Boxplots are constructed using the sliding window extractions. From these plots, and taking into account neuroscientific hypothesis stating that the signal of interest for resting state is occurring for frequencies below 0.1Hz, we chose to compute the long-memory parameters between scales 3 and 6.

Figures 9 and 10 display an example of long-memory parameter and long-run correlation estimated for one subject.

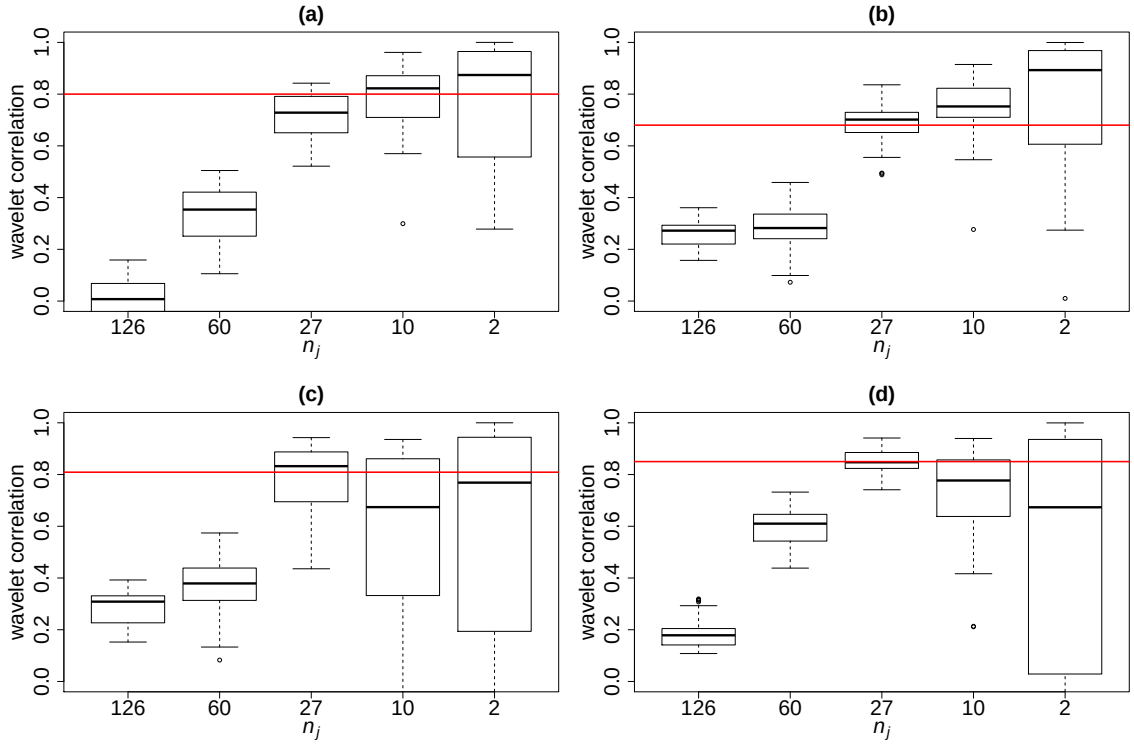

Figure 8: Boxplots of the correlation of the wavelet coefficients at different scales for real time series from fMRI data sets: (a) Time series 1 and 2; (b) Time series 13 and 14; (c) Time series 31 and 32; (d) Time series 47 and 48. Boxplots were obtained using sliding windows with $N = 512$ points, extracted from two fMRI time series with length equal to 1200 points, from a single subject. The estimated long parameters d of the two time series are equal. The fMRI data set is described in Section 7. The index of the horizontal axis displays the number of coefficients available. The horizontal red lines represent the estimated long-run correlation. Calculation was done on $N = 512$ observations for 100 replications using sliding windows (with overlap).

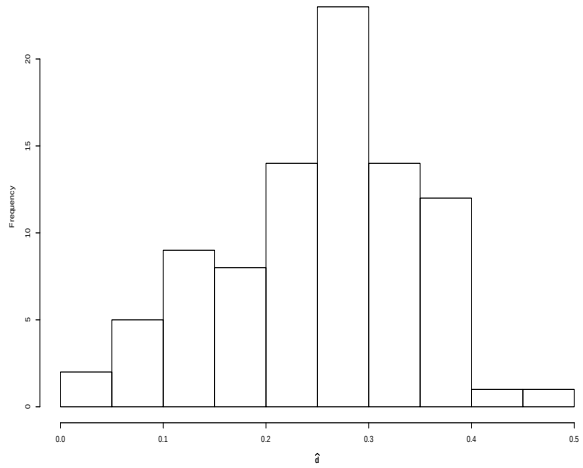

Figure 9: Histogram of $\widehat{\mathbf{d}}$ from a subject of the fMRI data set `brainHCP`.

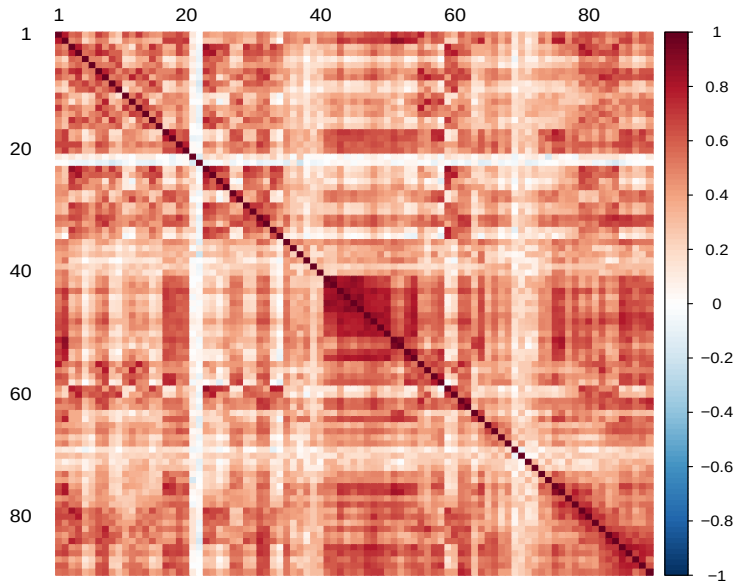

Figure 10: Estimation of $\boldsymbol{\Omega}$ from a subject of the fMRI data set `brainHCP`.

8. Conclusion

The R package **multiwave** provides a versatile wavelet-based approach, as well as a Fourier-based approach, for estimating long-memory parameters and long-run covariance matrices of multivariate time series. The two estimation procedures are based on semi-parametric

approaches proposed by Shimotsu (2007) and Achard and Gannaz (2016). The added value of the package is to provide estimations in long-range dependence multidimensional settings, which is not proposed presently by any R package to our knowledge. This paper describes the functions of the package **multiwave** and discusses some practical points for applications, including an application on a real data set. A simulation study shows first that multivariate estimation improves univariate estimation. The advantage of the wavelet-based procedure with respect to the Fourier-based estimation is its flexibility, allowing to take into account trends or nonstationarity.

Acknowledgments

The authors are grateful to Shimotsu (http://shimotsu.web.fc2.com/Site/Matlab_Codes.html) and to Faÿ, Moulines, Roueff and Taqqu for kindly providing the codes of their respective papers. This work was partly supported by the project *Graphsip* from Agence Nationale de la Recherche (ANR-14-CE27-0001). S.A. was partly funded by a grant from la Région Rhône-Alpes and a grant from AGIR-PEPS, Université Grenoble Alpes-CNRS.

References

Abadir KM, Distaso W, Giraitis L (2007). "Nonstationarity-Extended Local Whittle Estimation." *Journal of Econometrics*, **141**(2), 1353–1384. doi:10.1016/j.jeconom.2007.01.020.

Abry P, Veitch D (1998). "Wavelet Analysis of Long-Range-Dependent Traffic." *IEEE Transactions on Information Theory*, **44**(1), 2–15. doi:10.1109/18.650984.

Achard S, Bassett DS, Meyer-Lindenberg A, Bullmore E (2008). "Fractal Connectivity of Long-Memory Networks." *Physical Review E*, **77**(3), 036104. doi:10.1103/physreve.77.036104.

Achard S, Gannaz I (2016). "Multivariate Wavelet Whittle Estimation in Long-Range Dependence." *Journal of Time Series Analysis*, **37**(4), 476–512. doi:10.1111/jtsa.12170.

Achard S, Gannaz I (2019). **multiwave**: *Estimation of Multivariate Long-Memory Models Parameters*. R package version 1.4, URL https://CRAN.R-project.org/package=multiwave.

Beran J (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York.

Faÿ G, Moulines E, Roueff F, Taqqu MS (2009). "Estimators of Long-Memory: Fourier Versus Wavelets." *Journal of Econometrics*, **151**(2), 159–177. doi:10.1016/j.jeconom.2009.03.005.

Haslett J, Raftery AE (1989). "Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource." *Journal of the Royal Statistical Society C*, **38**(1), 1–50. doi:10.2307/2347679.

Hyndman RJ (2019). *CRAN Task View: Time Series Analysis*. Version 2019-04-14, URL https://CRAN.R-project.org/view=TimeSeries.

Hyndman RJ, Khandakar Y (2008). "Automatic Time Series Forecasting: The **forecast** Package for R." *Journal of Statistical Software*, **27**(3), 1–22. doi:10.18637/jss.v027.i03.

Lobato IN (1997). "Consistency of the Averaged Cross-Periodogram in Long Memory Series." *Journal of Time Series Analysis*, **18**(2), 137–155. doi:10.1111/1467-9892.00043.

Lobato IN (1999). "A Semiparametric Two-Step Estimator in a Multivariate Long Memory Model." *Journal of Econometrics*, **90**(1), 129–153. doi:10.1016/s0304-4076(98)00038-4.

Mächler M (2012). ***fracdiff**: Fractionally Differenced ARIMA aka ARFIMA(p, d, q) Models*. R package version 1.4-2, URL https://CRAN.R-project.org/package=fracdiff.

Maxim V, Şendur L, Fadili J, Suckling J, Gould R, Howard R, Bullmore E (2005). "Fractional Gaussian Noise, Functional MRI and Alzheimer's Disease." *Neuroimage*, **25**(1), 141–158. doi:10.1016/j.neuroimage.2004.10.044.

McLeod AI, Veenstra J (2014). ***FGN**: Fractional Gaussian Noise and Power Law Decay Time Series Model Fitting*. R package version 2.0-12, URL https://CRAN.R-project.org/package=FGN.

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. doi:10.1111/j.1467-9892.2006.00502.x.

Nielsen FS (2011). "Local Whittle Estimation of Multi-Variate Fractionally Integrated Processes." *Journal of Time Series Analysis*, **32**(3), 317–335. doi:10.1111/j.1467-9892.2010.00702.x.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robinson PM (2005). "Robust Covariance Matrix Estimation: HAC Estimates with Long Memory/Antipersistence Correction." *Econometric Theory*, **21**(1), 171–180. doi:10.1017/s0266466605050115.

Robinson PM (2008). "Multiple Local Whittle Estimation in Stationary Systems." *The Annals of Statistics*, **36**(5), 2508–2530. doi:10.1214/07-aos545.

Sela RJ, Hurvich CM (2008). "Computationaly Efficient Methods for Two Multivariate Fractionally Integrated Models." *Journal of Time Series Analysis*, **30**(6), 631–651. doi:10.1111/j.1467-9892.2009.00631.x.

Sela RJ, Hurvich CM (2012). "The Averaged Periodogram Estimator for a Power Law in Coherency." *Journal of Time Series Analysis*, **33**(2), 340–363. doi:10.1111/j.1467-9892.2011.00770.x.

Shimotsu K (2007). "Gaussian Semiparametric Estimation of Multivariate Fractionally Integrated Processes." *Journal of Econometrics*, **137**(2), 277–310. doi:10.1016/j.jeconom.2006.01.003.

Shimotsu K (2012). "Exact Local Whittle Estimation of Fractionally Cointegrated Systems." *Journal of Econometrics*, **169**(2), 266–278. doi:10.1016/j.jeconom.2012.01.028.

Songsiri J, Vandenberghe L (2010). "Topology Selection in Graphical Models of Autoregressive Processes." *The Journal of Machine Learning Research*, **11**, 2671–2705.

Termenon M, Jaillard A, Delon-Martin C, Achard S (2016). "Reliability of Graph Analysis of Resting State fMRI Using Test-Retest Dataset from the Human Connectome Project." *NeuroImage*, **142**, 172–187. doi:10.1016/j.neuroimage.2016.05.062.

Tsay RS (2013). *Multivariate Time Series Analysis: With R and Financial Applications*. John Wiley & Sons.

Veenstra JQ (2012). *Persistence and Anti-Persistence: Theory and Software*. Ph.D. thesis, Western University.

Veenstra JQ, McLeod AI (2018). **arfima**: *Fractional ARIMA (and Other Long Memory) Time Series Modeling*. R package version 1.7-0, URL https://CRAN.R-project.org/package=arfima.

Whitcher B (2019). **waveslim**: *Basic Wavelet Routines for One-, Two- And Three-Dimensional Signal Processing*. R package version 1.7.5.1, URL https://CRAN.R-project.org/package=waveslim.

Whitcher B, Guttorp P, Percival DB (2000). "Wavelet Analysis of Covariance with Application to Atmospheric Time Series." *Journal of Geophysical Research*, **105**(D11), 14941–14962. doi:10.1029/2000jd900110.

A. Additional tables

d	*bias*	*std*	*RMSE*	*ratio M/U*
0.2	−0.0170	0.0390	0.0425	0.7786
0.0	0.0130	0.0391	0.0412	0.9013
0.2	−0.0313	0.0376	0.0490	0.8960
0.2	−0.0316	0.0378	0.0493	0.8805
0.2	−0.0170	0.0383	0.0419	0.7674
0.4	−0.0442	0.0395	0.0592	0.7902

Table 2: Multivariate Whittle wavelet estimation of $\mathbf{d}$ for a bivariate FIVARMA$(0, \mathbf{d}, 0)$ with $\rho = 0.8$, $N = 512$ with 1000 repetitions. For the estimation, $j_0 = 1$.

	$\mathbf{d} = (0.2, 0)$			$\mathbf{d} = (0.2, 0.2)$			$\mathbf{d} = (0.2, 0.4)$		
$\mathbf{\Omega}$	*bias*	*std*	*RMSE*	*bias*	*std*	*RMSE*	*bias*	*std*	*RMSE*
$\Omega_{1,1}$	0.0417	0.0724	0.0836	0.0343	0.0711	0.0789	0.0417	0.0717	0.0830
$\Omega_{1,2}$	0.0382	0.0657	0.0759	0.0279	0.0626	0.0686	0.0673	0.0684	0.0959
$\Omega_{2,2}$	0.0048	0.0709	0.0710	0.0323	0.0714	0.0784	0.0748	0.0748	0.1057
corr.	0.0191	0.0227	0.0296	0.0010	0.0164	0.0164	0.0194	0.0235	0.0304

Table 3: Wavelet Whittle estimation of $\mathbf{\Omega}$ for a bivariate FIVARMA$(0, \mathbf{d}, 0)$ with $\rho = 0.8$, $N = 512$ with 1000 repetitions. For the estimation, $j_0 = 1$.

d	*bias*	*std*	*RMSE*	*ratio M/U*
0.2	−0.0473	0.1213	0.1302	0.8472
0.0	−0.0371	0.1266	0.1320	0.8511
0.2	−0.0623	0.1209	0.1360	0.8848
0.2	−0.0526	0.1258	0.1364	0.8714
0.2	−0.066	0.1244	0.1408	0.9161
0.4	−0.0584	0.1293	0.1418	0.8935

Table 4: Multivariate Whittle wavelet estimation of $\mathbf{d}$ for a bivariate FIVARMA$(1, \mathbf{d}, 0)$ with $\rho = 0.8$, $N = 512$ with 1000 repetitions. For the estimation, $j_0 = 3$.

	$\mathbf{d} = (0.2, 0)$			$\mathbf{d} = (0.2, 0.2)$			$\mathbf{d} = (0.2, 0.4)$		
$\mathbf{\Omega}$	*bias*	*std*	*RMSE*	*bias*	*std*	*RMSE*	*bias*	*std*	*RMSE*
$\Omega_{1,1}$	−0.0067	0.0828	0.0831	0.0024	0.0859	0.0860	0.0050	0.0889	0.0890
$\Omega_{1,2}$	0.0551	0.0828	0.0995	0.0495	0.0794	0.0936	0.0507	0.0875	0.1012
$\Omega_{2,2}$	0.1363	0.1310	0.1891	0.1412	0.1384	0.1977	0.1391	0.1426	0.1992
corr.	−0.0088	0.0805	0.0810	−0.0386	0.0527	0.0653	−0.0358	0.0985	0.1047

Table 5: Wavelet Whittle estimation of $\mathbf{\Omega}$ for a bivariate FIVARMA$(1, \mathbf{d}, 0)$ with $\rho = 0.8$, $N = 512$ with 1000 repetitions. For the estimation, $j_0 = 3$.

d	*bias*	*std*	*RMSE*	*ratio M/U*
1.2	−0.0338	0.0762	0.0834	0.8509
1	−0.0276	0.0725	0.0776	0.8316
1.2	−0.0430	0.0732	0.0849	0.8672
1.2	−0.0411	0.0743	0.0849	0.8591
1.2	−0.0338	0.0741	0.0814	0.8309
1.4	−0.0356	0.0797	0.0873	0.8344
2.2	−0.0421	0.0884	0.0979	0.8718
2	−0.0403	0.0862	0.0951	0.8516
2.2	−0.0503	0.0860	0.0996	0.8875
2.2	−0.0490	0.0823	0.0958	0.8566
2.2	−0.0436	0.0868	0.0971	0.8652
2.4	−0.0429	0.0831	0.0935	0.8400

Table 6: Multivariate Whittle wavelet estimation of $\mathbf{d}$ for a bivariate FIVARMA$(0, \mathbf{d}, 0)$ with $\rho = 0.8$, $N = 512$ with 1000 repetitions. Nonstationary cases. For the estimation, $j_0 = 2$.

	$\mathbf{d} = (1.2, 1)$			$\mathbf{d} = (1.2, 1.2)$			$\mathbf{d} = (1.2, 1.4)$		
$\boldsymbol{\Omega}$	*bias*	*std*	*RMSE*	*bias*	*std*	*RMSE*	*bias*	*std*	*RMSE*
$\Omega_{1,1}$	−0.0048	0.1362	0.1363	0.0059	0.1369	0.1370	−0.0047	0.1361	0.1361
$\Omega_{1,2}$	0.0180	0.1168	0.1182	0.0093	0.1155	0.1158	0.0159	0.1266	0.1276
$\Omega_{2,2}$	−0.0027	0.1277	0.1277	0.0042	0.1386	0.1386	−0.0113	0.1487	0.1491
corr.	0.0214	0.0475	0.0521	0.0051	0.0286	0.0291	0.0227	0.0506	0.0555
	$\mathbf{d} = (2.2, 2)$			$\mathbf{d} = (2.2, 2.2)$			$\mathbf{d} = (2.2, 2.4)$		
$\boldsymbol{\Omega}$	*bias*	*std*	*RMSE*	*bias*	*std*	*RMSE*	*bias*	*std*	*RMSE*
$\Omega_{1,1}$	−0.0383	0.1795	0.1835	−0.0253	0.1789	0.1807	−0.0361	0.1776	0.1812
$\Omega_{1,2}$	−0.0043	0.1565	0.1565	−0.0129	0.1493	0.1498	−0.0097	0.1602	0.1605
$\Omega_{2,2}$	−0.0318	0.1776	0.1804	−0.0276	0.1809	0.1830	−0.0481	0.1813	0.1876
corr.	0.0251	0.0604	0.0654	0.0087	0.0374	0.0384	0.0249	0.0626	0.0674

Table 7: Wavelet Whittle estimation of $\boldsymbol{\Omega}$ for a bivariate FIVARMA$(0, \mathbf{d}, 0)$ with $\rho = 0.8$, $N = 512$ with 1000 repetitions. Nonstationary cases. For the estimation, $j_0 = 2$.

	$m = \lfloor N^{0.65} \rfloor = 57$				$m = \lfloor N^{\eta} \rfloor$				
d	*bias*	*std*	*RMSE*	*W/F*	η	*bias*	*std*	*RMSE*	*W/F*
0.2	−0.002	0.0576	0.047	0.9050	0.90	−0.0197	0.0271	0.0335	1.2688
0	−0.0009	0.0593	0.0238	1.7294		−0.0033	0.0264	0.0267	1.5475
0.2	−0.0033	0.0574	0.0531	0.9214	0.85	−0.0130	0.0305	0.0332	1.4751
0.2	−0.0031	0.0591	0.0522	0.9434		−0.0123	0.0293	0.0318	1.5502
0.2	0.0008	0.0576	0.0579	0.7239	0.85	−0.0136	0.0308	0.0337	1.2456
0.4	0.0009	0.0595	0.0889	0.6666		−0.0192	0.0299	0.0355	1.6689

Table 8: Multivariate Whittle Fourier estimation of $\mathbf{d}$ for a bivariate FIVARMA$(0, \mathbf{d}, 0)$ with $\rho = 0.8$, $N = 512$ with 1000 repetitions depending on the number of frequencies m. $\lfloor x \rfloor$ denotes the closest integer smaller than x.

	$\mathbf{d} = (0.2, 0)$			$\mathbf{d} = (0.2, 0.2)$			$\mathbf{d} = (0.2, 0.4)$		
$\boldsymbol{\Omega}$	*bias*	*RMSE*	*W/F*	*bias*	*RMSE*	*W/F*	*bias*	*RMSE*	*W/F*
$\Omega_{1,1}$	0.0186	0.2027	0.4123	0.0220	0.2036	0.3876	0.0113	0.2010	0.4127
$\Omega_{1,2}$	0.0124	0.1733	0.4381	0.0197	0.1748	0.3923	0.0149	0.1730	0.5545
$\Omega_{2,2}$	0.0146	0.2138	0.3323	0.0263	0.2163	0.3624	0.0291	0.2162	0.4891
corr.	−0.0013	0.0357	0.8298	−0.0004	0.0355	0.4621	−0.0012	0.0359	0.8483

Table 9: Fourier Whittle estimation of $\boldsymbol{\Omega}$ for a bivariate FIVARMA$(0, \mathbf{d}, 0)$ with $\rho = 0.8$, $N = 512$ with 1000 repetitions. The number of frequencies is $m = \lfloor n^\eta \rfloor$ with $\eta = 0.65$ as chosen in Shimotsu (2007).

	$\mathbf{d} = (0.2, 0)$ $\eta = 0.9$			$\mathbf{d} = (0.2, 0.2)$ $\eta = 0.85$			$\mathbf{d} = (0.2, 0.4)$ $\eta = 0.85$		
$\boldsymbol{\Omega}$	*bias*	*RMSE*	*W/F*	*bias*	*RMSE*	*W/F*	*bias*	*RMSE*	*W/F*
$\Omega_{1,1}$	0.0622	0.0940	0.8896	0.0387	0.0832	0.9484	0.0389	0.0833	0.9962
$\Omega_{1,2}$	0.0222	0.0638	1.1907	0.0304	0.0731	0.9380	0.0470	0.0826	1.1615
$\Omega_{2,2}$	−0.0031	0.0637	1.1149	0.0373	0.0839	0.9341	0.0812	0.1132	0.9338
corr.	−0.0013	0.0163	1.8223	−0.0004	0.0177	0.9237	−0.0012	0.0179	1.6969

Table 10: Fourier Whittle estimation of $\boldsymbol{\Omega}$ for a bivariate FIVARMA$(0, \mathbf{d}, 0)$ with $\rho = 0.8$, $N = 512$ with 1000 repetitions. The number of frequencies is $m = \lfloor n^\eta \rfloor$ with η such that RMSE of $\widehat{\mathbf{d}}$ is minimized.

Affiliation:

Sophie Achard
Univ. Grenoble Alpes, CNRS,
Grenoble-INP, GIPSA-lab
38000 Grenoble, France
E-mail: sophie.achard@gipsa-lab.fr
URL: http://www.gipsa-lab.grenoble-inp.fr/~sophie.achard/

Irène Gannaz
Université Lyon, INSA de Lyon,
Institut Camille Jordan
20 avenue Albert Einstein, 69621 Villeurbanne Cedex, France
E-mail: irene.gannaz@insa-lyon.fr
URL: http://math.univ-lyon1.fr/~gannaz/

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
May 2019, Volume 89, Issue 6 *Submitted:* 2016-11-25
doi:10.18637/jss.v089.i06 *Accepted:* 2018-01-28