Semiparametric, parametric and possibly sparse models for multivariate long-range dependence

Changryong Baek[a], Stefanos Kechagias[b], and Vladas Pipiras[c]

[a]Sungkyunkwan University, Seoul, Korea
[b]SAS Institute, Cary, USA
[c]University of North Carolina, Chapel Hill, USA

ABSTRACT

Several available formulations, parametric models and sparsity settings for multivariate long-range dependence (MLRD) are discussed. Furthermore, a new parametric identifiable model for a general formulation of MLRD is introduced in any dimension, and another sparsity setting is identified of potential interest in MLRD modeling. Estimation approaches for MLRD are also reviewed, including some recent progress and open questions about estimation in higher dimensions and sparse settings.

Keywords: Long-range dependence, multivariate time series, phase parameters, identifiable parametric models, VARFIMA models, sparsity, local Whittle estimation, Gaussian likelihood, regularization.

1. INTRODUCTION

The focus of this work is on multivariate stationary time series exhibiting long-range dependence (LRD, for short). Univariate LRD modeling is well understood and applied widely.[1–5] It relates naturally to scaling, self-similarity, wavelets, fractional Brownian motion and other well-studied phenomena and objects. Though extensions of LRD to multivariate (vector) time series have been considered previously, multivariate long-range dependence (MLRD, for short) still presents a number of only partially resolved or completely unresolved issues. The goal of this work is to draw a broad picture of current MLRD modeling, provide a few new results and raise several open problems.

Definitions and models of MLRD are discussed in Section 2, including several situations where postulating a sparse model is of interest. Several approaches to estimation of model parameters are presented in Section 3, and some open questions are raised. Conclusions can be found in Section 4.

2. MODELS FOR MULTIVARIATE LONG-RANGE DEPENDENCE

We thus focus on (second-order) stationary multivariate series $\{X_n\}_{n\in\mathbb{Z}} = \{(X_{1,n},\dots,X_{\ell,n})'\}_{n\in\mathbb{Z}}$ consisting of ℓ univariate component series $\{X_{j,n}\}_{n\in\mathbb{Z}}$, $j = 1,\dots,\ell$. Stationarity means that the mean vector $\mu_X = (\mu_{X,j})_{j=1,\dots,\ell} = \mathbb{E}X_n$ of the series does not depend on n, and that its autocovariance function (ACVF)

$$\gamma_X(h) := \Big(\gamma_{X,jk}(h)\Big)_{j,k=1,\dots,\ell} := \mathbb{E}X_{n+h}X_n' - \mu_X\mu_X' = \Big(\mathbb{E}X_{j,n+h}X_{k,n} - \mu_{X,j}\mu_{X,k}\Big)_{j,k=1,\dots,\ell} \tag{1}$$

depends only on a time lag $h \in \mathbb{Z}$ (and not on n). Note that the ACVF is matrix-valued. The spectral density, if it exists, is a complex-valued matrix function $f_X(\lambda) = (f_{X,jk}(\lambda))_{j,k=1,\dots,\ell}$, $\lambda \in (-\pi,\pi)$, satisfying

$$\gamma_X(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f_X(\lambda)d\lambda. \tag{2}$$

Further author information: (Send correspondence to V.P.)
C.B.: E-mail: crbaek@skku.edu
S.K.: E-mail: Stefanos.Kechagias@sas.com
V.P.: E-mail: pipiras@email.unc.edu

For each λ, the matrix $f_X(\lambda)$ is Hermitian symmetric and non-negative definite, and the function $f_X(\lambda)$, $\lambda \in (-\pi, \pi)$, is Hermitian (that is, $f_X(-\lambda) = \overline{f_X(\lambda)}$). In contrast to the univariate case, we also note that it is not true in general that $\gamma_X(h) = \gamma_X(-h)$, which is equivalent to $\gamma_X(h) = \gamma_X(h)'$. In fact, it is the relation $\gamma_X(-h) = \gamma_X(h)'$ that holds in general. Finally, we also caution the reader to be aware of what convention is used in the definition of the ACVF and the spectral density (for example, a common alternative definition is to set $\gamma_X(h) = \mathbb{E}X_n X_{n+h}' - \mu_X \mu_X'$), which has been the source of some confusion in the literature and certainly considerable nuisance in our work.

In Section 2.1, we define multivariate long-range dependence (MLRD), focusing on a semiparametric formulation in the spectral domain. Some families of parametric models for MLRD are discussed in Section 2.2, including a new identifiable family for a general dimension ℓ. Section 2.3 concerns some situations where it is natural to consider sparse MLRD models.

2.1 Semiparametric formulation

The most general available formulation of MLRD is given in the following definition.[6,7] Some comments and insight can be found following the definition.

Definition 2.1. *A stationary multivariate series $\{X_n\}_{n\in\mathbb{Z}}$ is long-range dependent (LRD) if its spectral density satisfies:*

$$f_X(\lambda) \sim \lambda^{-D} G \lambda^{-D}, \quad \text{as } \lambda \to 0^+, \tag{3}$$

where $D = diag(d_1, \dots, d_\ell)$ with $d_j \in (0, 1/2)$, $\lambda^{-D} = diag(\lambda^{-d_1}, \dots, \lambda^{-d_\ell})$ and $G \in \mathbb{C}^{\ell\times\ell}$ is a non-negative definite, Hermitian symmetric matrix. Entrywise, the relation (3) reads: for $j, k = 1, \dots, \ell$,

$$f_{X,jk}(\lambda) \sim g_{jk} \lambda^{-(d_j+d_k)} =: \alpha_{j,k} e^{i\phi_{jk}} \lambda^{-(d_j+d_k)}, \quad \text{as } \lambda \to 0^+, \tag{4}$$

where $g_{jk} \in \mathbb{C}$, $\alpha_{jk} \in \mathbb{R}$ and $\phi_{jk} \in (-\pi/2, \pi/2)$.

The parameter $d_j \in (0, 1.2)$ is the usual *long-range dependence (LRD) parameter*. The case $d_j \le 0$ is associated with short-range dependence (SRD), and $d_j < 0$ with the so-called anti-persistence. The representation $g_{jk} = \alpha_{j,k} e^{i\phi_{jk}}$ is just a polar coordinate representation, but with an important convention that ϕ_{jk} is restricted to $(-\pi/2, \pi/2)$ and hence $\alpha_{jk} \in \mathbb{R}$. The other possibility is to assume that $\phi_{jk} \in (-\pi, \pi)$ and $\alpha_{jk} \ge 0$ but this imposes further restrictions on the real-valued matrix (α_{jk}). The parameter ϕ_{jk} is referred to as the *phase parameter* of the component series $\{X_{j,n}\}$ and $\{X_{k,n}\}$, and α_{jk} is their *amplitude parameter*. Note that $\phi_{jj} = 0$ and $\phi_{jk} = -\phi_{kj}$, $j \ne k$, since G is non-negative definite, Hermitian symmetric.

Under mild assumptions, the condition (4) can be shown[4] to be equivalent to: for $j, k = 1, \dots, \ell$,

$$\gamma_{X,jk}(h) \sim r_{jk} h^{(d_j+d_k)-1}, \quad \text{as } h \to \infty. \tag{5}$$

The case $r_{jk} = r_{kj}$ corresponds to $\phi_{jk} = 0$; otherwise, ϕ_{jk} controls the asymmetry of the tails of the ACVF components at large lags.

Remark 2.2. *The case $D = dI_\ell$ with $d \in (0, 1/2)$ (and the identity matrix I_ℓ) and singular matrix G is special and is referred to as fractional cointegration. This case will be excluded from the discussion of this paper. For more information, for example, consult a review in Ref. 8.*

Finally, for later reference, we also note the following time-domain interpretation of the matrix $G = (g_{jk})$ in (3). Note that

$$\mathrm{Var}\Big(\sum_{k=1}^{N} X_k\Big) = \sum_{h=-(N-1)}^{N-1} (N - |h|)\gamma_X(h) = N \sum_{h=-(N-1)}^{N-1} \gamma_X(h) - \sum_{h=-(N-1)}^{N-1} |h|\gamma_X(h). \tag{6}$$

By using (2) and (3), we have

$$N^{-D}\Big(\sum_{h=-(N-1)}^{N-1} \gamma_X(h)\Big)N^{-D} = \int_{-\pi}^{\pi} \Big(\sum_{h=-(N-1)}^{N-1} e^{ih\lambda}\Big) N^{-D} f_X(\lambda) N^{-D} d\lambda$$

$$= 2\int_0^{\pi} \Big(\sum_{h=-(N-1)}^{N-1} e^{ih\lambda} \Big) N^{-D}(f_X(\lambda) + f_X(\lambda)^*)N^{-D} d\lambda$$

$$= 2\int_0^{\pi N} \Big(\sum_{h=-(N-1)}^{N-1} e^{i\frac{h}{N}w}\frac{1}{N} \Big) N^{-D}(f_X(\frac{w}{N}) + f_X(\frac{w}{N})^*)N^{-D} dw$$

$$\sim 2\int_0^{\infty} \Big(\int_{-1}^{1} e^{iuw}du \Big) w^{-D}(G + G^*)w^{-D}dw = 4\int_0^{\infty} \frac{\sin w}{w} w^{-D}(G+G^*)w^{-D}dw$$

$$= \Big(\Re(g_{jk})8\int_0^{\infty}\frac{\sin w}{w}w^{-(d_j+d_k)}dw\Big) =: \Big(\Re(g_{jk})A(d_j+d_k)\Big),$$

as $N \to \infty$, where we leave justification of the asymptotic equivalence above to the reader. Similarly, one expects that $N^{-1}N^{-D}(\sum_{h=-(N-1)}^{N-1}|h|\gamma_X(h))N^{-D} \sim (\Re(g_{jk})B(d_j+d_k))$ and hence, in view of (6), that

$$N^{-(D+1/2)}\mathrm{Var}\Big(\sum_{k=1}^{N} X_k\Big)N^{-(D+1/2)} \sim \Big(\Re(g_{jk})C(d_j+d_k)\Big), \tag{7}$$

as $N \to \infty$, where $B(\cdot)$ and $C(\cdot)$ are some functions. That is, the real part of G can be thought as the covariance matrix of the series aggregated at large time scales. To make an imaginary part emerge in the time domain, one would have to consider a quantity that is not invariant to the reversion of the series in time, for example, $\mathrm{Cov}(\sum_{k=1}^{N} X_k, \sum_{k=N+1}^{2N} X_k)$.

2.2 Parametric models

A common parametric model for MLRD is to take

$$X_n = (I-B)^{-D}\eta_n = \begin{pmatrix} (I-B)^{-d_1}\eta_{1,n} \\ \vdots \\ (I-B)^{-d_\ell}\eta_{\ell,n} \end{pmatrix}, \tag{8}$$

where $\{\eta_n\}$ is a ℓ–variate white noise series satisfying $\mathbb{E}\eta_n = 0$ and $\mathbb{E}\eta_n\eta_n' = \Sigma = (\sigma_{jk})_{j,k=1,\dots,\ell}$, and $(I-B)^{-d}$ is the usual fractional integration operator defined as

$$(I-B)^{-d} = \sum_{m=0}^{\infty} b_m B^m, \tag{9}$$

where B is the backwardshift operator (i.e. $B^k X_n = X_{n-k}$, $k \in \mathbb{Z}$, and $B^0 = I$) and b_m's are the coefficients in the Taylor expansion $(1-z)^{-d} = \sum_{m=0}^{\infty} b_m z^m$, known to satisfy $b_m = \Gamma(m+d)/(\Gamma(d)\Gamma(m+1))$. In other words, according to (8), each component series $\{X_{j,n}\}$ of $\{X_n\}$ is obtained by fractionally integrating a univariate white noise series $\{\eta_{j,n}\}$; the dependence between the component series arises through the dependence of the component white noise series, when the covariance matrix Σ is not diagonal.

The model (8) is known as VARFIMA$(0, D, 0)$ series (also as a FIVARMA$(0, D, 0)$ series). It extends to a VARFIMA(p, D, q) series $\{X_n\}$ satisfying

$$(I-B)^D\Phi(B)X_n = \Theta(B)\eta_n, \tag{10}$$

where $\Phi(z) = I_\ell - \Phi_1 z - \ldots - \Phi_p z^p$ and $\Theta(z) = I_\ell - \Theta_1 z - \ldots - \Theta_q z^q$ are matrix polynomials of integer orders $p, q \geq 0$, modeling short-range dependence effects. (A FIVARMA(p, D, q) series is defined similarly but interchanging the order of $(I-B)^D$ and $\Phi(B)$, with the two definitions not being equivalent since matrices do not commute in general.) Further restrictions are imposed on matrix polynomials $\Phi(z)$ and $\Theta(z)$ to ensure existence of a stationary and causal solution to (10).

A VARFIMA$(0, D, 0)$ series is LRD in the sense of Definition 2.1. Indeed, by using Theorem 11.8.3 in Ref. 9, a VARFIMA $(0, D, 0)$ series has the spectral density given by

$$f_X(\lambda) = \frac{1}{2\pi}(1 - e^{-i\lambda})^{-D}\Sigma(1 - e^{i\lambda})^{-D}. \tag{11}$$

Since $1 - e^{-i\lambda} \sim i\lambda = \lambda e^{i\pi/2}$, as $\lambda \to 0^+$, it follows that the spectral density (11) satisfies (3) with

$$G = \frac{1}{2\pi}e^{-i\frac{\pi}{2}D}\Sigma e^{i\frac{\pi}{2}D}. \tag{12}$$

In particular, the amplitude and phase parameters for a VARFIMA$(0, D, 0)$ series are given by

$$\alpha_{jk} = \sigma_{jk}, \quad \phi_{jk} = \frac{\pi}{2}(d_k - d_j). \tag{13}$$

The same conclusion can be reached for VARFIMA(p, D, q) series.

Note from (13) that VARFIMA$(0, D, 0)$ series have very special phase parameters, which depend directly on the values of LRD parameters. One parametric class of models allowing for general phases consists of two-sided VARFIMA$(0, D, 0)$ series defined by

$$X_n = \Big((I - B)^{-D}Q_+ + (I - B^{-1})^{-D}Q_-\Big)\epsilon_n, \tag{14}$$

where $\{\epsilon_n\}$ is a ℓ–variate white noise series satisfying $\mathbb{E}\epsilon_n = 0$ and $\mathbb{E}\epsilon_n\epsilon_n' = I_\ell$, and $Q_+, Q_- \in \mathbb{R}^{\ell\times\ell}$ are matrices. It is called two-sided because of the presence of forward shift operator B^{-1} in (14). When $Q_- = 0$, a two-sided VARFIMA$(0, D, 0)$ series becomes a one-sided VARFIMA$(0, D, 0)$ series with $\Sigma = Q_+Q_+'$. An appealing feature of the model (14), for example, in simulation or inference, is that its ACVF can be computed explicitly.[6]

The fact that two-sided VARFIMA$(0, D, 0)$ series allow for general phases will become apparent from the result given below. In fact, a real issue with two-sided VARFIMA$(0, D, 0)$ series is not in proving that they yield general phases but that a fixed semiparametric specification (3) can be achieved by multiple choices of the parameters D, Q_+Q_+' and Q_-Q_-' (e.g. in the bivariate case $\ell = 2$, note that (3) has 6 parameters while D, Q_+Q_+' and Q_-Q_-' consist of 8 parameters). The next proposition specifies an identifiable parametrization of two-sided VARFIMA$(0, D, 0)$ series.

Proposition 2.3. *There is a one-to-one correspondance between the parameters D and G in the semiparametric specification (3) and the parameters D and lower-triangular matrices*

$$Q_+ = \begin{pmatrix} q_{11} & 0 & \dots & 0 \\ q_{+.21} & q_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ q_{+,\ell 1} & q_{+,\ell 2} & \dots & q_{\ell\ell} \end{pmatrix}, \quad Q_- = \begin{pmatrix} 0 & 0 & \dots & 0 \\ q_{-.21} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ q_{-,\ell 1} & q_{-,\ell 2} & \dots & 0 \end{pmatrix} \tag{15}$$

in the specification (14) of two-sided VARFIMA$(0, D, 0)$ series, with $q_{jj} > 0$.

The one-to-one correspondence stated in the proposition is detailed in the proof below.

Proof. Since G in (3) is Hermitian positive definite, it has a unique Cholesky decomposition

$$G = ZZ^* = (Ze^{-i\frac{\pi}{2}D})(Ze^{-i\frac{\pi}{2}D})^*, \tag{16}$$

where a lower-triangular matrix

$$Z = \begin{pmatrix} z_{11} & 0 & \dots & 0 \\ z_{21} & z_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ z_{\ell 1} & z_{\ell 2} & \dots & z_{\ell\ell} \end{pmatrix} \in \mathbb{C}^{\ell\times\ell} \tag{17}$$

has real-valued and positive diagonal entries $z_{jj} > 0$. On the other hand, by arguing as for (11)–(12), the spectral density of a two-sided VARFIMA$(0, D, 0)$ series satisfies

$$f_X(\lambda) \sim \lambda^{-D} W W^* \lambda^{-D}, \quad \text{as } \lambda \to 0^+, \tag{18}$$

where

$$W = \frac{1}{\sqrt{2\pi}}(e^{-i\frac{\pi}{2}D} Q_+ + e^{i\frac{\pi}{2}D} Q_-). \tag{19}$$

To relate (3) with G in (16)–(17) and (18) with Q_+, Q_- in (15), one can then set

$$\frac{1}{\sqrt{2\pi}}(e^{-i\frac{\pi}{2}D} Q_+ + e^{i\frac{\pi}{2}D} Q_-) = Z e^{-i\frac{\pi}{2}D}. \tag{20}$$

Entrywise, this is equivalent to

$$\frac{1}{\sqrt{2\pi}} q_{jj} = z_{jj} \tag{21}$$

(when considering the diagonal elements of (20)), and

$$\frac{1}{\sqrt{2\pi}}(e^{-i\frac{\pi}{2}d_j} q_{+,jk} + e^{i\frac{\pi}{2}d_j} q_{-,jk}) = z_{jk} e^{-i\frac{\pi}{2}d_k} \tag{22}$$

(when considering the off-diagonal elements of (20)). Note that (21)–(22) has a unique solution for q_{jj}, $q_{+,jk}$ and $q_{-,jk}$, since $e^{-i\pi d_j/2}$ and $e^{i\pi d_j/2}$ are linearly independent. That is, as stated in the proposition, there is a one-to-one correspondence between D, G and D, Q_+, Q_-. □

In the bivariate case $\ell = 2$, another identifiable specification (14) was proposed in Ref. 10, where no restrictions were placed on Q_+ but Q_- was taken as

$$Q_- = \begin{pmatrix} c & 0 \\ 0 & -c \end{pmatrix} Q_+ \tag{23}$$

with $c \in (-1, 1)$. In this case, there is a one-to-one correspondence between the parameter $c \in (-1, 1)$ and the phase parameter ϕ_{12} (the only uniquely defined phase parameter in the bivariate case). The specification (23) has a number of appealing features over (15), especially in terms of interpretation and symmetry, but such a construction does not extend to general dimension $\ell \geq 2$.

In order to capture SRD effects, similarly to (10), one can introduce two-sided VARFIMA(p, D, q) series $\{X_n\}$ satisfying

$$\Phi(B) X_n = \Delta_{D,Q}(B)^{-1} \Theta(B) \epsilon_n, \tag{24}$$

where $\Phi(B)$ and $\Theta(B)$ are matrix polynomials of respective orders p and q as in (10), and

$$\Delta_{D,Q}(B)^{-1} = \Big((I - B)^{-D} Q_+ + (I - B^{-1})^{-D} Q_- \Big). \tag{25}$$

For identifiability purposes, one could assume that $\Phi(B)$ is diagonal – see a discussion in Ref. 10, Section 3.2. Moroever, the case of non-diagonal $\Phi(B)$ is associated with fractional cointegration, which is not the focus here (see Remark 2.2).

Finally, we also note that constructing practical one-sided models with general phase has proved challenging, and remains an open problem. Several attempts and discussion on this issue can be found in Ref. 6.

2.3 Sparsity

There are a number of situations where one might be interested in a sparse[*] semiparametric or parametric model for MLRD, especially as the dimension ℓ increases. One natural direction is to consider sparse matrix polynomials $\Phi(z)$ and $\Theta(z)$ in the parametric specifications (10) and (24), which has certainly been a central assumption in the case of SRD time series modeling in higher dimension. Our goal here is to consider instead two contexts of sparsity specific to MLRD, namely, related to fractal connectivity and LRD precision matrices. Both contexts concern the matrix $G = (g_{jk})$ in the specification (3) of MLRD.

Fractal connectivity matrix: When

$$g_{jk} = 0, \tag{26}$$

the pair of the component series $\{X_{j,n}\}$ and $\{X_{k,n}\}$, $j \neq k$, is called *fractally non-connected*; and *fractally connected* if $g_{jk} \neq 0$. Moreover, in the fractally non-connected case, a further modeling assumption is often made as

$$f_{X,jk}(\lambda) \sim c_{jk}\lambda^{-d_{jk}}, \quad \text{as } \lambda \to 0^+, \tag{27}$$

with $c_{jk} \neq 0$ and

$$d_{jk} < d_j + d_k. \tag{28}$$

Likewise, the case $d_{jk} = d_j + d_k$ is associated with fractal connectivity.[11–13] The matrix G is referred to as the *fractal connectivity matrix*.

One can think of fractal non-connectivity as asymptotic decorrelation between two component series at large time scales (see (7) and the surrounding discussion). For example, a fractally non-connected model can be constructed as

$$X_n = Y_n + Z_n, \tag{29}$$

where $\{Y_n\}$ and $\{Z_n\}$ are uncorrelated, $\{Y_n\}$ is a one-sided VARFIMA$(0, D, 0)$ series (8) with the innovation covariance matrix Σ being sparse, and $\{Z_n\}$ is a one-sided VARFIMA$(0, D_1, 0)$ series with $d_{1,j} < d_j$, $j = 1, \dots, k$, and the innovation covariance matrix Σ_1 having non-zero entries. Those components $X_{j,n}$ and $X_{k,n}$ of X_n for which $\Sigma_{jk} = 0$ are fractally non-connected and satisfy (27) with $d_{jk} = d_{1,j} + d_{1,k} < d_j + d_k$. It should also be mentioned though that the model (29) may not be most representative of fractally connected pairs since their cross spectral densities are mixtures of two scaling laws.

Despite some effort, it is fair to say that a number of questions on fractal connectivity remain (at least partially) unresolved, for example, related to testing in the spectral domain, constructing practical families of parametric models, or relevance in applications.

LRD precision matrices: As pointed out above, zero elements of the matrix G are associated with component pairs of the series $\{X_n\}$ that can be thought as uncorrelated at large time lags. As with all correlation matrices,[14] their inverses, known as precision matrices, are associated with partial correlations between component pairs, while zero entries of precision matrices with partially uncorrelated pairs.

Several matrices P could be called LRD precision matrices in the case of MLRD series, in the sense that they would carry information about partial correlations (or the lack thereof) of component series at large time scales. The most natural candidate is perhaps the inverse of the matrix G itself, that is,

$$P = G^{-1}. \tag{30}$$

The motivation for this choice stems from the way partial (also known as conditional) correlations are commonly considered in the multivariate time series context. That is, first, residual series are defined for components j and k as

$$\begin{aligned} \epsilon_{j,n} &= X_{j,n} - \textstyle\sum_{m=-\infty}^{\infty} D_{j,m} X_{-jk,n-m}, \\ \epsilon_{k,n} &= X_{k,n} - \textstyle\sum_{m=-\infty}^{\infty} D_{k,m} X_{-jk,n-m}, \end{aligned} \tag{31}$$

where $\{X_{-jk,n}\}$ is the $(\ell-2)$–vector series obtained from $\{X_n\}$ by removing the jth and kth components, and $\{D_{j,m}\}$ ($\{D_{k,m}\}$, resp.) are the weights in the regression (prediction) of $X_{j,n}$ ($X_{j,k}$, resp.) on $\{X_{-jk,n-m}\}$. The

[*]As usual, model sparsity refers to a number of model coefficients being zero.

partial correlation between the component series $\{X_{j,n}\}$ and $\{X_{k,n}\}$ is characterized by the correlations between the residual series $\{\epsilon_{j,n}\}$ and $\{\epsilon_{k,n}\}$. These correlations are also reflected in the cross spectral density $f_{\epsilon,jk}(\lambda)$, $\lambda \in (-\pi, \pi)$, and the so-called partial spectral coherence

$$\mathrm{PSC}_{jk}(\lambda) = \frac{f_{\epsilon,jk}(\lambda)}{\sqrt{f_{\epsilon,jj}(\lambda) f_{\epsilon,kk}(\lambda)}}, \tag{32}$$

which can be shown to be equal to[15, 16]

$$\mathrm{PSC}_{jk}(\lambda) = -\frac{g_{jk}(\lambda)}{\sqrt{g_{jj}(\lambda) g_{kk}(\lambda)}}, \tag{33}$$

where $g(\lambda) = f_X(\lambda)^{-1}$. The component pair $\{X_{j,n}\}$ and $\{X_{k,n}\}$ is said to be partially uncorrelated if $\mathrm{PSC}_{jk}(\lambda) = 0$ for all $\lambda \in (-\pi, \pi)$. It is then natural to call this pair partially uncorrelated at large time scales (low frequencies) if

$$\mathrm{PSC}_{jk}(0^+) = 0. \tag{34}$$

In view of (33) and (3), this is equivalent to

$$(G^{-1})_{jk} = 0. \tag{35}$$

That is, the matrix G^{-1} can indeed be viewed as a LRD precision matrix as stated around (30). We also note that similarly to (27) in the case of fractal non-connectivity, a further modeling assumption could be made under (35), namely, $f^{-1}_{X,jk}(\lambda) \sim b_{jk}\lambda^{\tilde{d}_{jk}}$, where $\tilde{d}_{jk} < d_j + d_k$.

An example of MLRD model with some partially uncorrelated component pairs can be constructed through the one-sided VARFIMA$(0, D, 0)$ model in (8) by taking a sparse Σ^{-1}, since in this case G is given by (12) and thus G^{-1} is (up to constant entrywise) Σ^{-1}.

Other LRD precision matrix could be introduced as well. For example, in view of the discussion around (7), it might also be of interest to define a LRD precision matrix as $P = (\Re(g_{jk})C(d_j + d_k))^{-1}$.

3. ESTIMATION APPROACHES

We discuss here the estimation approaches commonly used in the settings of Sections 2.1 and 2.2, with those for sparsity (Section 2.3) appearing here for first time. Some simulations and data illustrations are also included.

3.1 Semiparametric formulation

A common way to estimate the parameters D and G in the semiparametric formulation (3) is through the so-called local Whittle approach, namely,

$$\begin{aligned} (\widehat{D}, \widehat{G}) &= \underset{(D,G)}{\operatorname{argmin}}\, Q(D, G) \quad \text{with} \\ Q(D, G) &= \frac{1}{m}\sum_{s=1}^{m} \Big(\log\det(\lambda_s^{-D} G \lambda_s^{-D}) + \mathrm{tr}(\lambda_s^{D} I_X(\lambda_s) \lambda_s^{D} G^{-1}) \Big), \end{aligned} \tag{36}$$

where $\lambda_s = (2\pi s)/N$ are the Fourier frequencies for sample size N, $I_X(\lambda) = N^{-1}(\sum_{n=1}^{N} X_n e^{in\lambda})(\sum_{n=1}^{N} X_n e^{-in\lambda})'$ is the periodogram matrix and m is the number of Fourier frequencies used in estimation. The optimization problem (36) can be reduced to that over D only as

$$\begin{aligned} \widehat{D} &= \underset{D}{\operatorname{argmin}}\, R(D) \quad \text{with} \\ R(D) &= \log\det(\widehat{G}(D)) - 2\mathrm{tr}(D)\frac{1}{m}\sum_{s=1}^{m} \log \lambda_s, \end{aligned} \tag{37}$$

where

$$\widehat{G}(D) = \frac{1}{m}\sum_{s=1}^{m} \lambda_s^D I_X(\lambda_s)\lambda_s^D. \tag{38}$$

A theoretical study of the local Whittle estimators in the bivariate case $\ell = 2$ was undertaken by Robinson,[7] who established the asymptotic normality result for the estimates of d_1, d_2 and the phase parameter ϕ_{12}. Baek et al.[17] raised and clarified a number of other issues for the local Whittle estimation, still in the bivariate case $\ell = 2$. We are not aware of theoretical developments for arbitrary dimension ℓ and general semiparametric formulation (3), except work in some special cases of G.[18] Future work in arbitrary dimension should take into account the issues raised in Ref. 17.

Wavelet estimation in the semiparametric setting was considered by Achard and Gannaz.[19]

3.2 Parametric models

A parametric model such as the two-sided VARFIMA(p, D, q) model in (24)–(25) can be fit to data through a Gaussian maximum likelihood (ML), as long as its ACVF can be computed efficiently for different parameter values. ACVF is known explicitly for VARFIMA(p, D, q) series when $p = 0$, but not when $p \geq 1$.[6,20] For this reason, it is convenient to use a conditional ML estimation, namely,

$$\underset{\Phi(B),\Theta(B),D,\Sigma}{\operatorname{argmax}} \mathcal{L}(\Theta(B), D, \Sigma; \{\Phi(B)X_n\}), \tag{39}$$

where $\mathcal{L} := \mathcal{L}(\Theta(B), D, \Sigma; \{Y_n\})$ is the Gaussian log-likelihood for the VARFIMA$(0, D, q)$ series $Y_n = \Phi(B)X_n$, that is,

$$\mathcal{L} = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\Omega_N| - \frac{1}{2}Y'\Omega_N^{-1}Y, \tag{40}$$

where $Y = (Y_{1,1}, \ldots, Y_{1,N}, \ldots, Y_{\ell,1}, \ldots, Y_{\ell,N})'$ is the vector of each component series stacked one under the other, $\Omega_N = \mathbb{E}YY'$ is the covariance matrix of the vector Y and $|\Omega_N|$ is the determinant of Ω_N. Note that Ω_N is comprised of ℓ^2 Toeplitz blocks of size $N \times N$.

Computing the log-likelihood function $\mathcal{L}$ directly is stable and efficient for small and moderate dimensions ($\ell N \leq 40,000$), despite the high order of complexity $O(\ell^3 N^3)$ associated with evaluating the log-determinant $\log|\Omega_N|$ and the product $\Omega_N^{-1}Y$. For longer series, on the other hand, a more efficient and robust strategy is to express the log-likelihood function as

$$\mathcal{L} = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{n=0}^{N-1}\log|V_n| - \frac{1}{2}\sum_{n=0}^{N-1}(Y_{n+1} - \widehat{Y}_{n+1})'V_n^{-1}(Y_{n+1} - \widehat{Y}_{n+1}), \tag{41}$$

where $\widehat{Y}_{n+1} = \mathbb{E}(Y_{n+1}|Y_1, \ldots, Y_n)$ and V_n, $n = 0, \ldots, N-1$, are the one-step-ahead finite sample predictors and their corresponding error covariance matrices. These quantities can be obtained via the multivariate Durbin Levinson (DL) algorithm, a recursive algorithm with order of complexity $O(\ell^3 N^2)$. In the bivariate case $\ell = 2$, the DL method was used in Ref. 10 for inference under the two-sided VARFIMA(p, D, q) series.

Further improvements in numerical efficiency of evaluating $\mathcal{L}$ can be attained by using the preconditioned conjugate gradient (PCG) method, a popular iterative algorithm that can be used to solve the system $\Omega_N x = Y$ up to a desired accuracy (equivalent to computing $\Omega_N^{-1}Y$).[20] The number of iterations of the PCG method, however, increases with the condition number $\kappa(\Omega_N)$ of Ω_N which is known to be large, for example, for d's close to 0.5. To reduce the number of iterations, each Toeplitz block of Ω_N is embedded into a $\widetilde{N} \times \widetilde{N}$ circulant matrix, leading to a system $C\widetilde{x} = \widetilde{Y}$, where C is a matrix with ℓ^2 circulant blocks and $\widetilde{Y}$ is the vector resulting from padding the individual components of Y with $\widetilde{N} - N$ zeros. Circulant block matrices have two important properties that allow $C\widetilde{x} = \widetilde{Y}$ to be solved at $O(\ell^2 N \log N)$ cost. First, they have known preconditioning matrices,[21] that is, nonsingular matrices M such that $\kappa(M^{-1}C) << \kappa(C)$ (hence $C\widetilde{x} = \widetilde{Y}$ can be replaced by

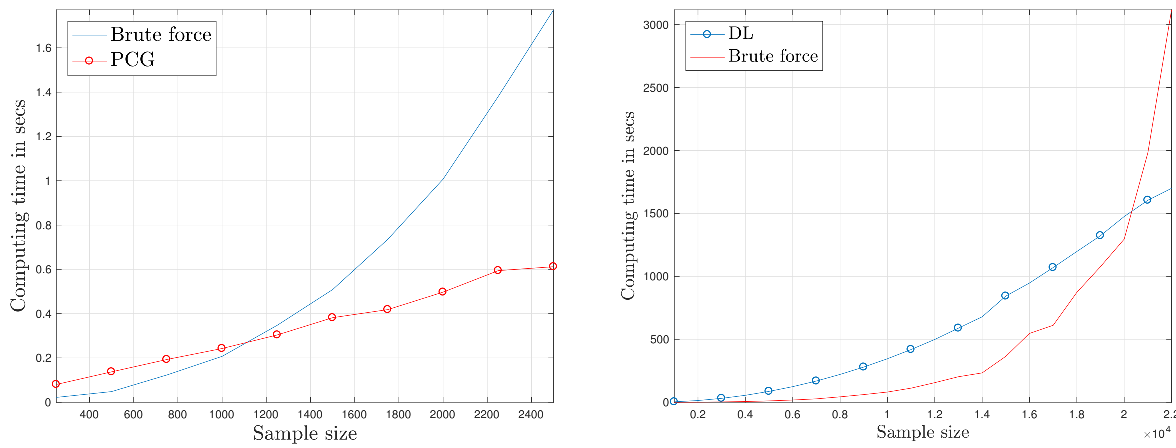

Figure 1. Log-likelihood evaluation times of bivariate VARFIMA$(0, D, 0)$ series with $D = \text{diag}(0.2, 0.4)$, $\Sigma_{11} = \Sigma_{22} = 1, \Sigma_{12} = 0.5$.

$M^{-1}C\widetilde{x} = M^{-1}\widetilde{Y}$). Second, the major computational tasks of PCG, which are matrix-vector products of the form Cz and $M^{-1}z$, can be carried out very efficiently using the fast Fourier transform.

Ref. 20 also uses the PCG method to approximate $\log|\Omega_N|$ (at a satisfying accuracy) using the following interpolation scheme. First, compute a small number of $|V_r|, r = 1, \dots, s$, say $s = 50$ via the multivariate DL algorithm. Then, use the PCG method ℓ times to compute V_{N-1} from Sowell's equations.[22] The intermediate determinants can then be approximated by fitting the relationship $r\sqrt{|V_r|} = \alpha + \beta r$ for $r = 1, \dots, s, N-1$ and taking the fitted values for $r = s+1, \dots, N-2$.

Figure 1 shows the time required for one likelihood evaluation of white noise series of several sample sizes using the three approaches discussed above under a VARFIMA$(0, D, 0)$ covariance structure. The software used is Matlab R2017a on a 2.8 GHz processor with 16GB of RAM. The brute force evaluation outperformed the PCG method (with $s = 50$) for $N \leq 1100$ and the multivariate DL approach for $N \leq 20,000$. Moreover, we did not encounter any instability issues even for parameters schemes with larger d's. Finally, to verify the computational costs of the DL and brute force approaches, we fitted power-law equations of the form $t(N) = aN^b$ to the corresponding evaluation times and obtained $t_{DL}(N) = 3.14^{-6}N^{2.01}$ and $t_{BF}(N) = 2.6^{-11}N^{3.15}$, which agree with the computational orders listed above.

Performance of the various approaches as ℓ increases remains to be explored.

3.3 Sparsity

Several situations were discussed in Section 2.3 where sparse modeling in MLRD series might be of interest. We focus here on estimation of a sparse fractal connectivity matrix G and a sparse LRD precision matrix $P = G^{-1}$ through regularization, and present some preliminary discussion, simulation results and an application. Some work on sparse estimation in parametric MLRD models can be found in Ref. 23, though more systematic studies of related issues are certainly called for.

Matrices G^{-1} and G (denoted G^k with $k = -1$ and 1 below) can be estimated sparsely through a penalized version of the "local" negative log-likelihood in (36), that is,

$$(\widehat{D}, \widehat{G}) = \underset{(D,G)}{\text{argmin}}\, Q_{p,k}(D, G) \quad \text{with } k = -1, 1, \tag{42}$$

$$Q_{p,k}(D, G) = -2\text{tr}(D)\frac{1}{m}\sum_{s=1}^{m}\log\lambda_s - \log\det(G^{-1}) + \text{tr}\Big(\Big(\frac{1}{m}\sum_{s=1}^{m}\lambda_s^D I_X(\lambda_s)\lambda_s^D\Big)G^{-1}\Big) + \rho\|G^k\|_1,$$

where we slightly rewrote the first term of $Q(D, G)$ in (36), $\rho > 0$ is a penalization parameter and $\|G^k\|_1$ is the l_1–norm of G^k (the sum of the absolute values of the elements of G^k). For fixed D and real-valued G,

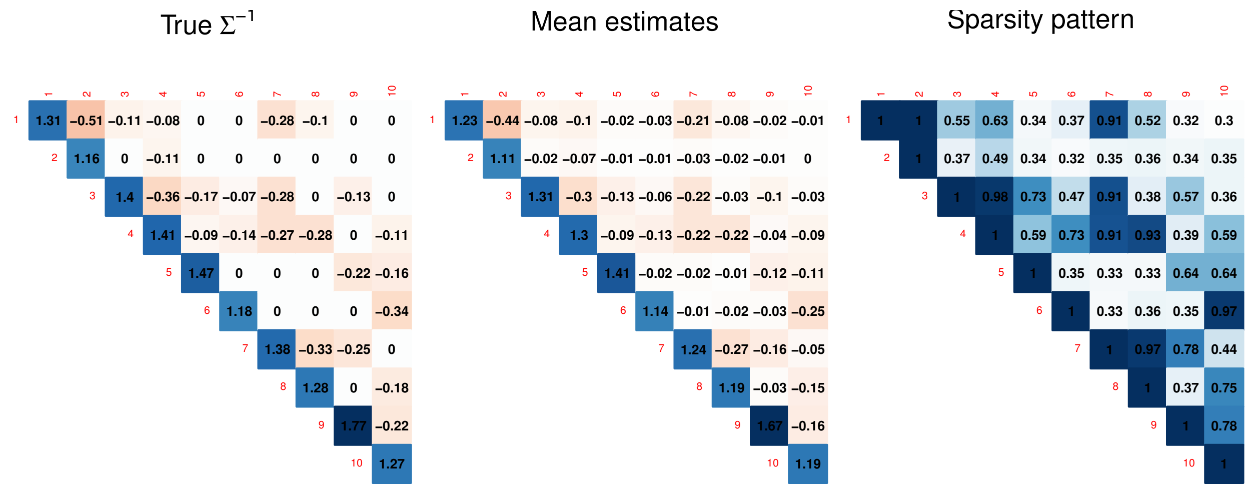

Figure 2. Regularized estimation based on the graphical lasso for Σ^{-1}. True values (left), mean estimated values (middle) and the proportions of non-zero coefficients out of 500 replications (right) are shown.

the objective function $Q_{p,k}(D,G)$ is that used in the so-called graphical lasso[24] when $k=-1$, and in sparse estimation of covariance matrices[25] when $k=1$.

We shall examine (42) in the case of the one-sided VARFIMA$(0,D,0)$ model in (8), for which G is given by (12) with real-valued Σ. In this case, the objective function $Q_p(D,G)$ in (42) can be replaced by

$$Q_{p,k}(D,\Sigma) = -2\mathrm{tr}(D)\frac{1}{m}\sum_{s=1}^{m}\log\lambda_s - \log\det(\Sigma^{-1}) + \mathrm{tr}\Big(\Re\Big(\frac{1}{m}\sum_{s=1}^{m}\lambda_s^D e^{i\pi D/2} I_X(\lambda_s) e^{-i\pi D/2}\lambda_s^D\Big)\Sigma^{-1}\Big) + \rho\|\Sigma^k\|_1. \tag{43}$$

Again, for fixed D and up to an additive constant, this is exactly the objective function used in the graphical lasso when $k=-1$, and in sparse estimation of covariance matrix when $k=1$. To assess the performance of estimation of Σ through (43), we shall minimize $Q_{p,k}(D,\Sigma)$ over Σ^k through the methods in Refs. 24, 25 by using D estimated through local Whittle. In simulations (not reported here), we found this approach to be comparable to the ideal situation where true D was considered.

Simulation results are reported in Figure 2 concerning LRD precision matrix and Figure 3 concerning fractal connectivity matrix. The true sparse matrix Σ^{-1} is depicted in the left plot of Figure 2. The sample size of the one-sided VARFIMA$(0,D,0)$ series is taken as $N=250$, and the true diagonal matrix D has diagonal entries .36, .39, .39, .36, .42, .32, .34, .38, .45, .35 (the estimated values for LRD parameters in the real data considered below). The number of frequencies m was taken as $N^{2/3}$, and the penalization parameter is chosen by cross validation. The middle plot of Figure 2 depicts the average estimates of Σ^{-1} over 500 realizations. The right plot shows the proportions of realizations that the (j,k)th element of Σ^{-1} is estimated as non-zero. Observe that the average value is slightly biased but this is due to biasedness of the regularization method. Note also that larger coefficients tend not to be estimated as zero most of the time as expected, and zero coefficients are estimated as zero around 60 to 70 percent. These proportions improve as the sample size increases (not reported here for brevity).

A fractally non-connected model (29) is considered for Figure 3. We used the same D as above and D_1 given by $D_1 = \mathrm{diag}(.11, .10, .08, .14, .09, .09, .10, 0.10, .06, .08)$. The true sparse Σ and non-sparse Σ_1 used in the simulation are depicted in the left and middle plots of the figure. The sample size $N=1000$ is considered and tuning parameters are chosen by cross validation. Out of 500 replications, it is observed from the right plot of Figure 3 that sparse covariance estimation finds the true sparsity pattern reasonably well. Some of the difficulties (e.g. the (4,7)th element of Σ is detected as zero) might be related to the issue with the chosen model discussed following (29).

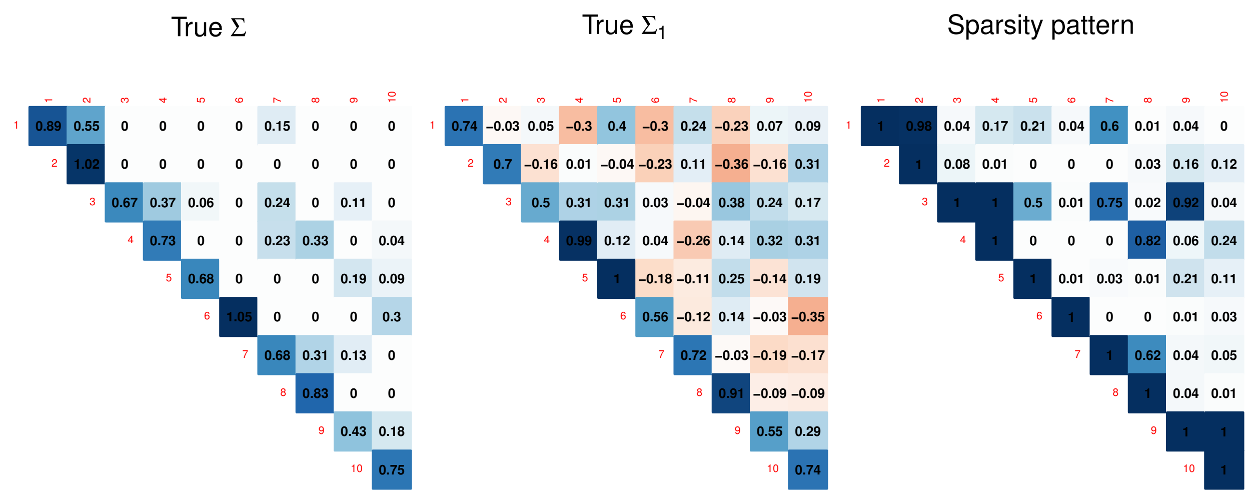

Figure 3. Regularized estimation based on sparse covariance estimation for Σ. True values (left and middle) and the proportions of non-zero coefficients out of 500 replications (right) are shown.

We also applied sparse estimation to the real fMRI data supplied in the R package "multiwave",[†] which implements multivariate wavelet Whittle estimation. For illustration purposes, we consider here only the first 10 ROIs, and first look into estimating sparse Σ^{-1}. In contrast to the simulations above, we examine how parameter estimates change depending on the number of frequencies m used in estimation. (The tuning parameter ρ is selected by cross validation as above.) This is depicted in Figure 4 through heatmap plots for vectorized matrix Σ^{-1} elements (vertical axis) according to the number of frequencies m (horizontal axis). The left plot represents the estimated values, coded blue for positive values, red for negative values and white for values close to zero. The right plot indicates whether the estimated values are zero (white) or non-zero (black). Observe that the estimated coefficients are stable over a wide range of frequencies m considered. The stability can also be expressed by counting the number of non-zero elements in the estimated Σ as in Figure 5, the left plot. The total number of non-zero elements is stable over a wide range of frequencies considered, and the choice of $m = N^{2/3}$ seems quite reasonable. The estimate of Σ^{-1} for this choice of m is given in the right plot of Figure 5. The local Whittle estimates of the LRD parameters are .36, .39, .39, .36, .42, .32, .34, .38, .45, .35.

The analogous plots for regularized estimation of sparse Σ are presented in Figures 6 and 7. For this particular data, less sparsity is attained by regularizing Σ, rather than the inverse Σ^{-1}.

4. CONCLUSIONS

In this work, we described the various modeling approaches for multivariate long-range dependence (MLRD), focusing on semiparametric, parametric and sparse model formulations. A new parametric identifiable model for MLRD was introduced. Several sparsity contexts were identified (one of them appearing here for the first time) and examined through popular regularized estimation approaches. A number of open questions were also raised, setting the stage for future work.

ACKNOWLEDGMENTS

The work of the first author was supported in part by the Basic Science Research Program from the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT & Future Planning (NRF-2017R1A1A1A05000831). The third author was supported in part by the NSF grant DMS-1712966.

[†]See https://cran.r-project.org/web/packages/multiwave/

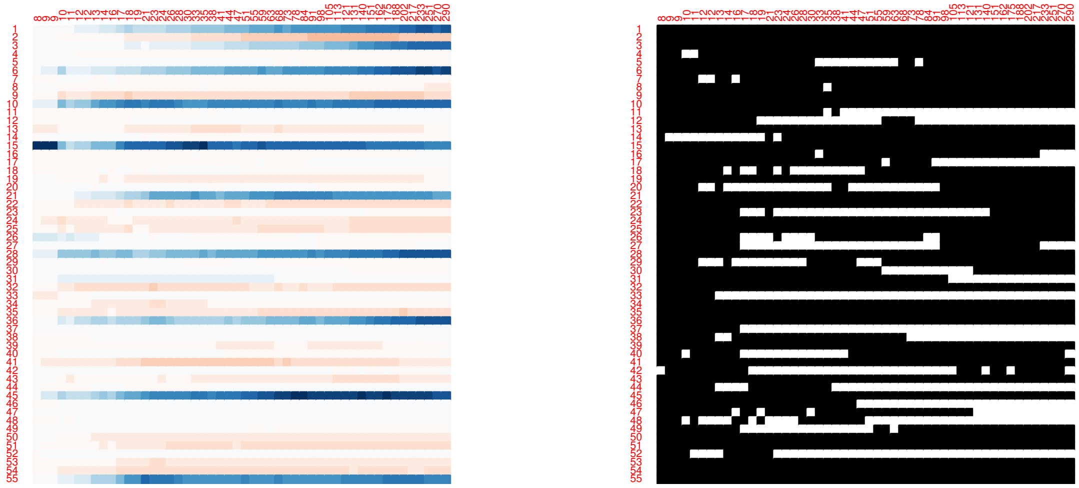

Figure 4. Heatmap plot of parameter estimates (y-axis) of Σ^{-1} as the number of frequencies used in estimation increases (x-axis). Left panel shows estimated coefficients in color raging from red to blue indicating negative to positive values. The right panel shows either non-zero (black) or zero (white).

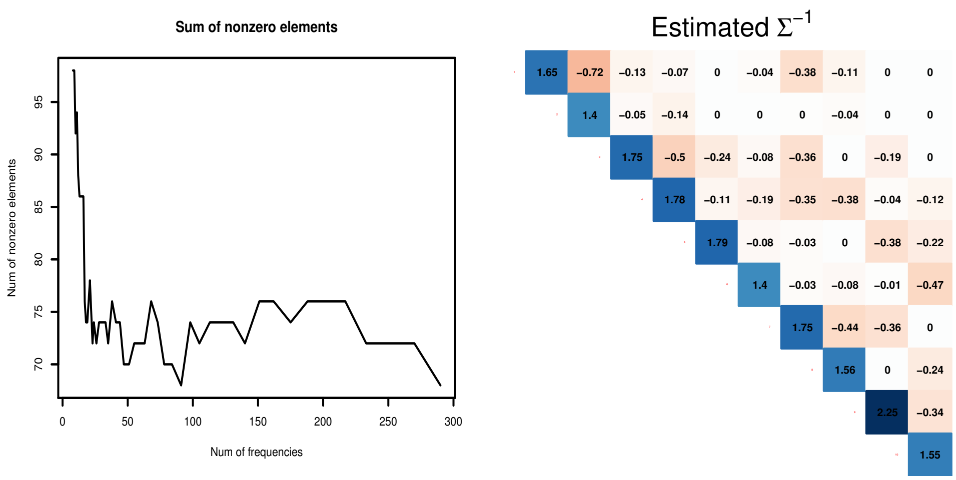

Figure 5. Plot for total number of non-zero elements in Σ^{-1} as m increases, and the estimated Σ^{-1} from real fMRI data.

REFERENCES

[1] Beran, J., Feng, Y., Ghosh, S., and Kulik, R., [*Long-Memory Processes: Probabilistic Properties and Statistical Methods*], Springer (2013).

[2] Doukhan, P., Oppenheim, G., and Taqqu, M. S., eds., [*Theory and Applications of Long-Range Dependence*], Birkhäuser, Boston (2003).

[3] Giraitis, L., Koul, H. L., and Surgailis, D., [*Large Sample Inference for Long Memory Processes*], Imperial College Press, London (2012).

[4] Pipiras, V. and Taqqu, M. S., [*Long-Range Dependence and Self-Similarity*], Cambridge University Press, Cambridge (2017).

[5] Rangarajan, G. and Ding, M., [*Processes with Long-Range Correlations: Theory and Applications*], vol. 621, Springer Science & Business Media (2003).

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

Figure 6. Heatmap plot of parameter estimates (y-axis) for Σ as the number of frequencies used in estimation increases (x-axis). Left panel shows estimated coefficients and right panel indicates whether it is non-zero (black) or zero (white).

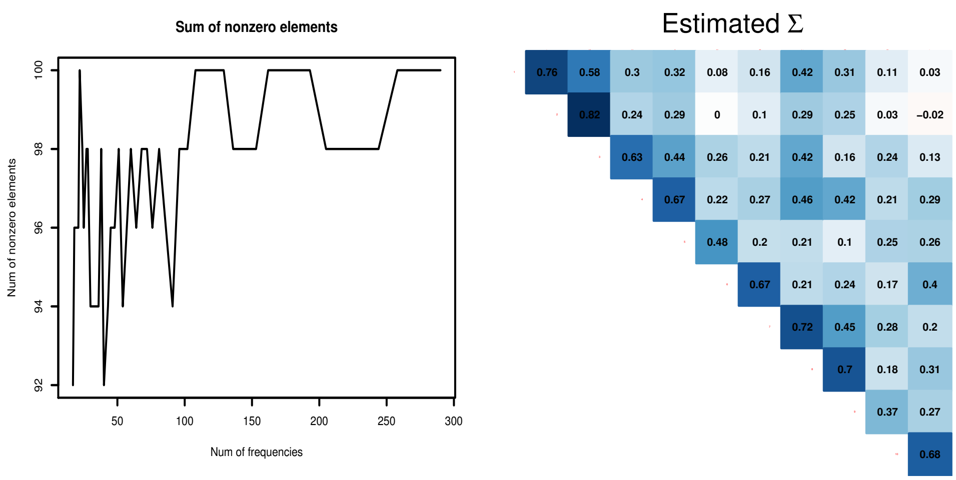

Figure 7. Plot for total number of non-zero elements in Σ as m increases, and the estimated Σ from real fMRI data.

[7] Robinson, P. M., "Multiple local Whittle estimation in stationary systems," *The Annals of Statistics* **36**(5), 2508–2530 (2008).

[8] Chen, W. W. and Hurvich, C. M., "Fractional cointegration," in [*Handbook of Financial Time Series*], Mikosch, T., Kreiß, J.-P., Davis, R. A., and Andersen, T. G., eds., 709–726, Springer Berlin Heidelberg, Berlin, Heidelberg (2009).

[9] Brockwell, P. J. and Davis, R. A., [*Time Series: Theory and Methods*], Springer Series in Statistics, Springer-Verlag, New York, second ed. (1991).

[10] Kechagias, S. and Pipiras, V., "Identification, estimation and applications of a bivariate long-range dependent time series model with general phase," Preprint (2017).

[11] Achard, S., Bassett, D. S., Meyer-Lindenberg, A., and Bullmore, E., "Fractal connectivity of long-memory networks," *Physical Review E* **77**, 036104 (Mar 2008).

[12] Wendt, H., Scherrer, A., Abry, P., and Achard, S., "Testing fractal connectivity in multivariate long memory processes," in [*Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*], 2913–2916, IEEE (2009).

[13] Sela, R. J. and Hurvich, C. M., "The averaged periodogram estimator for a power law in coherency," *Journal of Time Series Analysis* **33**(2), 340–363 (2012).

[14] Pourahmadi, M., [*High-Dimensional Covariance Estimation*], Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ (2013).

[15] Brillinger, D. R., [*Time series*], vol. 36 of *Classics in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2001). Data analysis and theory, Reprint of the 1981 edition.

[16] Dahlhaus, R., "Graphical interaction models for multivariate time series," *Metrika* **51**(2), 157–172 (2000).

[17] Baek, C., Kechagias, S., and Pipiras, V., "On local Whittle estimation in bivariate long-memory systems," Preprint (2017).

[18] Nielsen, F. S., "Local Whittle estimation of multi-variate fractionally integrated processes," *Journal of Time Series Analysis* **32**(3), 317–335 (2011).

[19] Achard, S. and Gannaz, I., "Multivariate wavelet Whittle estimation in long-range dependence," *Journal of Time Series Analysis* **37**(4), 476–512 (2016).

[20] Sela, R., *Three essays in econometrics: multivariate long memory time series and applying regression trees to longitudinal data*, PhD thesis, New York University (2010).

[21] Chan, T. F. and Olkin, J. A., "Circulant preconditioners for Toeplitz-block matrices," *Numerical Algorithms* **6**(1-2), 89–101 (1994).

[22] Sowell, F., "A decomposition of block Toeplitz matrices with applications to vector time series," *Unpublished manuscript* (1989).

[23] Sun, Y., "Modeling long-memory time series with sparse autoregressive processes," *Journal of Uncertain Systems* **6**(4), 289–298 (2012).

[24] Friedman, J., Hastie, T., and Tibshirani, R., "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics* **9**(3), 432–441 (2008).

[25] Bien, J. and Tibshirani, R. J., "Sparse estimation of a covariance matrix," *Biometrika* **98**(4), 807–820 (2011).