Asymptotic results for multivariate local Whittle estimation with applications

Marie-Christine Düker
Fakultät für Mathematik
Ruhr-Universität Bochum, Germany
Email: Marie-Christine.Dueker@ruhr-uni-bochum.de

Vladas Pipiras
Deptartment of Statistics and Operations Research
University of North Carolina at Chapel Hill, USA
Email: pipiras@email.unc.edu

***Abstract*—The asymptotic normality result is obtained for local Whittle estimators of all model parameters in a general formulation of multivariate long memory. The result is then used in devising a global statistical test for the so-called fractal non-connectivity, and in deriving the asymptotics of LASSO estimators of parameters in the so-called long-run variance matrix and its inverse. Some numerical illustrations are also provided.**

I. Introduction

We focus here on local Whittle estimation of model parameters in stationary multivariate short and long memory time series. This is perhaps the estimation method of choice when the possibility of multivariate long memory is considered, with numerous works on the topic over the last several decades (see [1]–[5], to name but a few). We next introduce the setting, the local Whittle estimators and other quantities of interest, and describe our contributions.

Consider a p-dimensional second-order stationary time series $X_n = (X_{1,n}, \dots, X_{p,n})'$, $n \in \mathbb{Z}$, with zero mean and autocovariance matrix function $\Gamma_X(h) = \mathrm{E}\, X_{n+h}X_n'$, $h \in \mathbb{Z}$. Suppose that its spectral density $f_X(\lambda)$, $\lambda \in (-\pi, \pi)$, related to the autocovariance through $\Gamma_X(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f_X(\lambda) d\lambda$, satisfies

$$f_X(\lambda) \sim \lambda^{-D} G \lambda^{-D}, \quad \text{as } \lambda \to 0^+, \tag{1}$$

where $\sim$ denotes componentwise asymptotic equivalence, $D = \mathrm{diag}(d_1, \dots, d_p)$ with $d_k \in [0, 1/2)$, $k = 1, \dots, p$, $\lambda^{-D} = \mathrm{diag}(\lambda^{-d_1}, \dots, \lambda^{-d_p})$ and $G = (g_{kl})_{k,l=1,\dots,p}$ is Hermitian symmetric and positive definite. The case $D \equiv 0$ is associated with short memory, while the case $d_k > 0$ with long memory of the kth component series $X_{k,n}$, $n \in \mathbb{Z}$. See [6], [7] for more details on univariate long and short memory and [8] for a discussion on multivariate long memory.

The local Whittle estimators $(\widehat{D}, \widehat{G})$ for the model parameters in (1) introduced in [3] are given by

$$(\widehat{D}, \widehat{G}) = \underset{(D,G)}{\arg\min}\, \ell(D, G) \tag{2}$$

for the negative log-likelihood

$$\ell(D,G) = \frac{1}{m} \sum_{j=1}^{m} (\log |\lambda_j^{-D} G \lambda_j^{-D}| + \mathrm{tr}(I_X(\lambda_j) \lambda_j^{D} G^{-1} \lambda_j^{D})), \tag{3}$$

where $|\cdot|$ and $\mathrm{tr}(\cdot)$ denote the determinant and the trace of a matrix, $I_X(\lambda) = \frac{1}{2\pi N} (\sum_{n=1}^{N} X_n e^{-in\lambda})(\sum_{n=1}^{N} X_n e^{in\lambda})'$ is the periodogram for sample size N and m is the number of Fourier frequencies $\lambda_j = 2\pi j / N$ used in estimation. In the bivariate case $p = 2$, the asymptotic normality of the local Whittle estimators of memory parameters d_1, d_2 and also of the so-called phase (that can be expressed in terms of the elements of G) was established in [3], and that of all model parameters in [5]. The asymptotic normality results in special cases of (1) but general p appear in [4], [9]. The main goal of this work is to provide such results for all model parameters in the general multivariate formulation (1), as well as to understand its implications. A special feature of the obtained result is the nearly explicit and workable form of the limiting covariance matrix.

Among the parameters of the model (1), we are particularly interested here in G, also called the long-run variance matrix, as well as in its precision matrix $P = G^{-1}$. Furthermore, it is of interest to model both of these matrices sparsely. For the matrix G, a zero element $g_{kl} = r_{1,kl} + i r_{2,kl} = 0$ can be thought as representing uncorrelatedness of the component series $X_{k,n}$ and $X_{l,n}$ at low frequencies (or large time lags). This case is also referred to as fractal non-connectivity of $X_{k,n}$ and $X_{l,n}$; see [10]–[12]. For the matrix $P = G^{-1}$, a zero entry p_{kl} can be thought as representing partial uncorrelatedness of the series $X_{k,n}$ and $X_{l,n}$ at low frequencies. Unsurprisingly perhaps, we will therefore also be interested in LASSO-type estimators of G and P.

The rest of the paper is organized as follows. In Section II, the asymptotic normality result for the local Whittle estimators is stated, including a sketch of the proof. Section III is concerned with testing for fractal non-connectivity. Section IV deals with asymptotic results for LASSO-type estimators. Some numerical results appear in Section V. Some of the technical details are omitted and can be found in the technical appendix for this article in [13].

II. Asymptotic normality results

We give here the asymptotic normality result for the local Whittle estimators $(\widehat{D}, \widehat{G})$ in (2) and also sketch its proof.

In order to define the parameter vector of interest, we introduce the matrix $\widetilde{G} = (\widetilde{g}_{kl})_{k,l=1,\dots,p}$, $\widetilde{g}_{kl} = g_{kk} \mathbb{1}_{\{k=l\}} + r_{1,kl} \mathbb{1}_{\{k>l\}} + r_{2,kl} \mathbb{1}_{\{k<l\}}$, where $g_{kl} = r_{1,kl} + i r_{2,kl}$, and the

vector $\widetilde{D} = (d_1, \dots, d_p)'$. Then, the parameter vector can be written as

$$\theta = ((\mathrm{vec}(\widetilde{G}))', \widetilde{D}')'. \tag{4}$$

The respective pairs of matrices $\widetilde{G}$, G and $\widetilde{D}$, D can be related as

$$\mathrm{vec}(\widetilde{G}) = L_p \,\mathrm{vec}(G), \;\; \mathrm{vec}(\widetilde{D}) = E_{p,p^2} \,\mathrm{vec}(D).$$

The matrix $L_p \in \mathbb{C}^{p\times p}$ is defined as

$$L_p = \frac{1}{2}(J_{p^2} I_{p^2} + J_{p^2}^* K_p), \tag{5}$$

where $*$ denotes the Hermitian conjugate and $J_{p^2} = \mathrm{diag}(\mathrm{vec}(\widetilde{J}))$ with $\widetilde{J} = (\mathbb{1}_{\{k\leqslant l\}} + i\mathbb{1}_{\{k>l\}})_{k,l=1,\dots,p}$. The matrix K_p denotes the commutation matrix, which transforms $\mathrm{vec}(M)$ into $\mathrm{vec}(M')$ for a square matrix M; see [14] for more details on these kinds of operations. The matrix E_{p,p^2} is defined as

$$E_{p,p^2} = (e_1, \mathbf{0}_{p\times p}, e_2, \mathbf{0}_{p\times p}, \dots, e_p), \tag{6}$$

where e_i denotes the ith standard basis vector of $\mathbb{R}^p$ and $\mathbf{0}_{p\times p}$ a $p \times p$-matrix with all entries equal to zero.

We introduce the multivariate extensions of the assumptions (C1)$'$, (C2), (C3)$'$ and (C4) in [5], and then state our main asymptotic normality result. It extends Theorem 2.3 in [5] to general p.

Assumption 1: The spectral density f_X satisfies (1). There is $q \in (0, 2]$ such that $\lambda^D \Psi(\lambda) - Q = O(\lambda^q)$, as $\lambda \to 0^+$, where $Q \in \mathbb{C}^{p\times p}$ satisfies $QQ^* = G$ and $\Psi(\lambda)$ is a $\mathbb{C}^{p\times p}$-valued function, differentiable in a neighborhood of $\lambda = 0$ such that $f_X(\lambda) = \Psi(\lambda)\,\Psi(\lambda)^*$ and $\lambda^D d\,\Psi(\lambda)/d\lambda = O(\lambda^{-1})$, as $\lambda \to 0^+$.

Assumption 2: The time series X_n has a linear representation $X_n = \sum_{j\in\mathbb{Z}} \Psi_j\, \varepsilon_{n-j}$ with $\sum_{j\in\mathbb{Z}} \| \Psi_j \|_F^2 < \infty$, where $\| \cdot \|_F$ denotes the Frobenius norm and $\Psi_j = (2\pi)^{-1}\int_{-\pi}^{\pi} \Psi(\lambda) e^{-ij\lambda} d\lambda$. The p-vector series $\{\varepsilon_j\}_{j\in\mathbb{Z}}$ satisfies $\mathrm{E}\,\varepsilon_j = 0$, $\mathrm{E}(\varepsilon_j \varepsilon_j') = I_p$, $\mathrm{E}(\varepsilon_j \varepsilon_i') = \mathbf{0}_{p\times p}$, $i \neq j$ and has almost sure constant first, second, third and fourth moments and respective cross-moments conditionally on $\mathcal{F}_{j-1} = \sigma(\varepsilon_i, i \leqslant j-1)$. Also, $\mathrm{P}(\varepsilon_j'\varepsilon_j > \nu) \leqslant C\,\mathrm{P}(X > \nu)$ for all $\nu > 0$ and an $\mathbb{R}_{>0}$-valued random variable X such that $\mathrm{E}\,X < \infty$.

Assumption 3: The parameter vector θ in (4) is such that $\theta \in \Theta_{\widetilde{G}} \times \Theta_{\widetilde{D}}$ with $\Theta_{\widetilde{G}} = (\mathbb{R}_{>0} \times \mathbb{R}^p)^{\times(p-1)} \times \mathbb{R}_{>0}$ and $\Theta_{\widetilde{D}} = [0, 1/2)^p$.

Assumption 4: The number of frequencies m satisfies $(\log m)^2 m^{1+2q}/N^{2q} \to 0$ and $(\log N)^C/m \to 0$ as $N \to \infty$ for any $C < \infty$, where $q \in (0, 2]$ appears in Assumption 1.

Theorem 2.1: Suppose that the above Assumptions 1-4 are satisfied. Then,

$$\sqrt{m}\,\mathrm{diag}((\log(N/m))^{-1} I_{p^2}, I_p)(\widehat{\theta} - \theta) \xrightarrow{\mathrm{d}} \mathcal{N}(0, C), \tag{7}$$

as $N \to \infty$, with

$$C = \begin{pmatrix} L_p R Z_G^{-1} R^* L_p^* & L_p R Z_G^{-1} \\ Z_G^{-1} R^* L_p^* & Z_G^{-1} \end{pmatrix}, \tag{8}$$

where $Z_G = 2(G \odot G^{-1} + I_p)$, $\odot$ denotes the Hadamard product and the matrix R is defined as

$$R = (\mathrm{vec}(G) \odot \mathrm{vec}(Y_1), \dots, \mathrm{vec}(G) \odot \mathrm{vec}(Y_p))$$

with $Y_i = (\mathbb{1}_{\{i=k\}} + \mathbb{1}_{\{i=l\}})_{k,l=1,\dots,p}$.

Sketch of Proof: The result (7) can be derived as in [3] and [5] which considered the case $p = 2$. We focus on calculating the information matrix for all model parameters for arbitrary dimension p. The negative log-likelihood $\ell(\theta) = \ell(D, G)$ of the model is given in (3). The information matrix is a 2×2 block matrix

$$I(\theta) = \begin{pmatrix} M_G & M_{G,D} \\ M_{G,D}^* & M_D \end{pmatrix} \tag{9}$$

with

$$M_{G,D} = \mathrm{E}(\mathbf{D}_{\widetilde{G}}(\mathbf{D}_{\widetilde{D}}\,\ell)), \;\; M_G = M_{G,G}, \;\; M_D = M_{D,D},$$

where $\mathbf{D}_V$ denotes the derivative matrix with respect to a vector V. The resulting blocks of (9) can be shown to be expressed as

$$\begin{aligned} M_G &= (L_p^{-1})^*((G^{-1})' \otimes G^{-1}) L_p^{-1}, \\ M_{G,D} &= -T_1 (L_p^{-1})^* (G^{-1} \oplus (G^{-1})') E_{p,p^2}', \\ M_D &= T_2 E_{p,p^2}(G \oplus G')(G^{-1} \oplus (G^{-1})') E_{p,p^2}', \end{aligned}$$

where $\oplus$ denotes the Kronecker sum defined as $A \oplus B = (I_p \otimes A) + (B \otimes I_p)$, L_p, E_{p,p^2} are as in (5) and (6), and

$$T_1 = \frac{1}{m}\sum_{j=1}^{m} \log\lambda_j, \;\; T_2 = \frac{1}{m}\sum_{j=1}^{m} (\log\lambda_j)^2. \tag{10}$$

The block structure of the information matrix (9) leads to the inverse $I^{-1}(\theta)$ as

$$\begin{pmatrix} M_G^{-1} + M_G^{-1} M_{G,D} S_{G,D}^{-1} M_{G,D}^* M_G^{-1} & -M_G^{-1} M_{G,D} S_{G,D}^{-1} \\ -S_{G,D}^{-1} M_{G,D}^* M_G^{-1} & S_{G,D}^{-1} \end{pmatrix},$$

where the so-called Schur complement is defined as $S_{G,D} = M_D - M_{G,D}^* M_G^{-1} M_{G,D}$ with

$$M_G^{-1} = L_p (G' \otimes G) L_p^*.$$

The Schur complement can be simplified to

$$S_{G,D} = (T_2 - T_1^2)(1/T_2) M_D.$$

Note also the non-obvious relation $M_D = T_2 Z_G$. The asymptotic orders of T_1 and T_2 in (10) are $T_1 \sim \log(m/N)$, $T_2 \sim (\log(m/N))^2$ and $T_2 - T_1^2 = 1 + o(1)$. These relations and further simplifications lead to the limiting covariance matrix C in (8). See [13] for more details. □

Remark 2.1: Theorem 2.1 is stated for the local Whittle estimator of the long-run variance matrix G. However, it can also be written in terms of the precision matrix P. A perhaps surprising fact is that the limiting covariance matrix coincides with (8) when replacing G with P and changing the sign of the off-diagonal blocks.

Remark 2.2: When $p = 1$, the asymptotic variance of the estimator $\widehat{d}$ of the memory parameter is $1/4$. When $p = 2$, these variances of $\widehat{d}_1$, $\widehat{d}_2$ take values in the range $(\frac{1}{8}, \frac{1}{4}]$; see Remark 2.1 in [5]. As a consequence of the asymptotic covariance matrix Z_G^{-1} in (8) associated with D, the asymptotic variance $(Z_G^{-1})_{kk}$ of $\widehat{d}_k$ is expected as

$$\frac{1}{4p} \leqslant \left[2(G \odot G^{-1} + I_p)\right]_{kk}^{-1} \leqslant \frac{1}{4}, \quad k = 1, \ldots, p, \tag{11}$$

for general p. The upper bound in (11) is a consequence of Corollary 7.7.4 in [15]. The lower bound is conjectured as discussed in [13].

Remark 2.3: In [3] and [5], the matrix G in (1) is also parametrized in terms of polar coordinates. For general p, this parametrization reads

$$G = (\omega_{kl} e^{\operatorname{sign}(k-l) i \phi_{kl}})_{k,l=1,\ldots,p} \tag{12}$$

with the so-called phase parameter $\phi_{kl} \in (-\pi/2, \pi/2)$ and $\omega_{kl} \in \mathbb{R}$. In this parametrization, one cannot test for fractal non-connectivity, since the respective phase parameter ϕ_{kl} is not identifiable for $\omega_{kl} = 0$, $k \neq l$; see [5] for a related discussion when $p = 2$. Furthermore, in contrast to the case $p = 2$ (see [5]), the phase parameter estimates $\widehat{\phi}_{kl}$ are generally not asymptotically uncorrelated with respect to the estimators $\widehat{\Omega} = (\widehat{\omega}_{kl})_{k,l=1,\ldots,p}$ and $\widehat{D}$. This leads to a 3×3 block matrix written in the same manner as (9). It is possible to derive Theorem 2.1 in terms of the parametrization in (12), but due to the 3×3 block structure, it seems impossible to get an explicit expression for the limiting covariance matrix.

III. Testing for fractal non-connectivity

As motivated in Section I, we are interested here in testing for fractal non-connectivity on a global level, that is, the hypothesis testing problem

$$H_0 : r_{1,kl} = r_{2,kl} = 0 \quad \text{for all} \quad k \neq l. \tag{13}$$

The parameters entering the null hypothesis (13) can be obtained from $\widetilde{G}$ by eliminating the diagonal elements of G. Therefore, we introduce the matrix

$$\mathcal{E}_{p(p-1),p^2} = (\mathbf{0}_{p(p-1)}, V_1, \mathbf{0}_{p(p-1)}, \ldots, \mathbf{0}_{p(p-1)}, V_{p-1}, \mathbf{0}_{p(p-1)})$$

with $V_q = (v_{1+(q-1)p}, \ldots, v_{p+(q-1)p})$, where v_i denotes the ith standard basis vector of $\mathbb{R}^{(p(p-1))^2}$ and $\mathbf{0}_{p(p-1)}$ is a $p(p-1)$-vector with all entries equal to zero. Then, the vector of parameters of interest can be written as

$$\vartheta = \mathcal{E}_{p(p-1),p^2} \operatorname{vec}(\widetilde{G}),$$

and similarly $\widehat{\vartheta}$ for the local Whittle estimators. The next result is a direct consequence of Theorem 2.1.

Corollary 3.1: Suppose that the above Assumptions 1-4 are satisfied. Then, under the hypothesis H_0 in (13),

$$\sqrt{m}(\widehat{\vartheta} - \vartheta) \xrightarrow{\mathrm{d}} \mathcal{N}(0, C_0), \tag{14}$$

as $N \to \infty$, where

$$C_0 = \frac{1}{2} \mathcal{E}_{p(p-1),p^2} (\mathcal{G} \otimes \mathcal{G}) \mathcal{E}'_{p(p-1),p^2} \tag{15}$$

with $\mathcal{G} = \operatorname{diag}(g_{11}, \ldots, g_{pp})$.

Note that C_0 is a diagonal matrix and $C_0^{-1} = 2\mathcal{E}_{p(p-1),p^2} (\mathcal{G}^{-1} \otimes \mathcal{G}^{-1}) \mathcal{E}'_{p(p-1),p^2}$. We introduce the test statistic

$$\widehat{\xi}_N = m\ \widehat{\vartheta}' \widehat{C}_0^{-1} \widehat{\vartheta}. \tag{16}$$

Under the hypothesis H_0, it satisfies

$$\widehat{\xi}_N \xrightarrow{\mathrm{d}} \chi^2(p(p-1)),$$

as $N \to \infty$, where $\chi^2(K)$ denotes the chi-square distribution with K degrees of freedom. Also, under the alternative H_1, one can show that $\widehat{\xi}_N \xrightarrow{\mathrm{p}} \infty$. When $p = 2$, the test statistic (16) was introduced in [5].

Remark 3.1: As pointed out in Remark 2.1, one can replace the matrix G with the precision matrix P in the asymptotic normality result in Theorem 2.1. This then leads to a similar test for partial uncorrelatedness, written as the hypothesis testing problem $H_0 : p_{1,kl} = p_{2,kl} = 0$ for all $k \neq l$, where $p_{kl} = p_{1,kl} + i p_{2,kl}$.

Remark 3.2: Though the fractal non-connectivity test above is developed on a global level, a similar test can also be introduced at a local level, that is, to test $r_{1,kl} = r_{2,kl} = 0$ for fixed $k \neq l$. But the resulting covariance matrix would not have such a simple form as in (14)-(15).

IV. Asymptotic results for LASSO-type estimators

Motivated by Section III concerning fractal non-connectivity, one might be interested in estimating the matrix G in (1) (or the corresponding precision matrix $P = G^{-1}$) under a sparsity assumption. Sparse estimation has been considered by numerous authors, for example, [16]–[19].

A penalized version of the local Whittle estimation (2) was proposed in [20] with the focus on its good numerical performance. As a consequence of Theorem 2.1, we will establish here theoretically the asymptotic properties of the penalized estimators. The penalized estimators $\widehat{G}_L$ and $\widehat{P}_L = \widehat{G}_L^{-1}$ of [20] are given by

$$\widehat{G}_L^r = \underset{G^r}{\arg\min}\, \ell_{L,r}(\widehat{D}, G), \quad \text{for } r = -1, 1,$$

where $\widehat{D}$ is the local Whittle estimator,

$$\ell_{L,r}(D, G) = \frac{1}{m} \sum_{j=1}^m \log |\lambda_j^{-D} G \lambda_j^{-D}| + \operatorname{tr}(\widehat{G}(D) G^{-1}) + \rho_N \|G^r\|_{1,off} \tag{17}$$

with a penalty parameter $\rho_N > 0$,

$$\widehat{G}(D) = \frac{1}{m} \sum_{j=1}^m \lambda_j^D I_X(\lambda_j) \lambda_j^D \tag{18}$$

and the l_1-norm $\|\cdot\|_{1,off}$ excluding the diagonal elements. For fixed $\widehat{D}$, $\widehat{G}_L$ coincides with estimators used in estimating covariance matrices sparsely; see [19]. On the other hand, $\widehat{P}_L$ coincides with the graphical LASSO estimator; see [18].

The next result gives an asymptotic normality result for $\widehat{G}_L$ and $\widehat{P}_L$ in the "fixed p, large N" asymptotics.

Theorem 4.1: Suppose that the above Assumptions 1-4 are satisfied and $\sqrt{\nu}\rho_N \to \rho_0 \geqslant 0$, as $N \to \infty$ for $\nu = m/(\log(N/m))^2$. Then,

$$\sqrt{\nu}(\widehat{G}_L^r - G^r) \xrightarrow{\mathrm{d}} \underset{U=U^*}{\arg\min}\, V_r(U), \quad \text{for } r = -1, 1, \tag{19}$$

as $N \to \infty$, where

$$V_r(U) = \frac{1}{2}\operatorname{tr}(UGUG) + \operatorname{tr}(UN) + R(U, G^r)$$

and for $U = (u_{kl})_{k,l=1,\dots,p}$ and $V = (v_{kl})_{k,l=1,\dots,p}$,

$$R(U,V) = \rho_0 \sum_{k \neq l} (\operatorname{sign}(v_{kl}) u_{kl} \mathbb{1}_{\{v_{kl} \neq 0\}} + |u_{kl}| \mathbb{1}_{\{v_{kl}=0\}}).$$

The $p \times p$ random matrix N is such that $\operatorname{vec}(N)$ follows $\mathcal{N}(0, RZ_G^{-1}R^*)$.

SKETCH OF PROOF: The result in (19) can be derived as in [17], [21]. The proof requires an asymptotic result for $\widehat{G}(\widehat{D})$ with $\widehat{G}(D)$ as in (18). Note that the optimization problem (2) is equivalent to

$$\widehat{D} = \arg\min_D \frac{1}{m} \sum_{j=1}^m \log|\lambda_j^{-D} \widehat{G}(D) \lambda_j^{-D}|$$

with $\widehat{G}(D)$ as in (18); see Remark 2.9 in [5]. For this reason, Theorem 2.1 gives

$$\sqrt{m}(\log(N/m))^{-1} \operatorname{vec}(\widehat{G}(\widehat{D}) - G) \xrightarrow{\mathrm{d}} \mathcal{N}(0, RZ_G^{-1}R^*). \tag{20}$$

Then, the result follows by adapting the proof of Theorem 1 in [17]. □

Remark 4.1: The asymptotic normality result in (20) can also be used to derive the oracle properties in [22] for an adaptive LASSO-type estimator. The oracle properties state that the estimator identifies the sparsity structure asymptotically correct. An adaptive LASSO estimator adjusts the penalization in (17) by weights depending on the data. In contrast to Theorem 4.1, it requires a consistent pre-estimator for the respective matrix of interest G or P.

V. DATA STUDY

Our numerical study assesses the performance of the fractal non-connectivity test introduced in Section III. A numerical study concerning the LASSO-type estimators of Section IV can be found in [20]. For $p = 2$, the local Whittle estimation on synthetic and real data is examined in [5].

For the fractal non-connectivity test, we shall examine its empirical sizes and powers on the following time series. For the size calculations, we use a fractally non-connected series with the spectral density $f_X(\lambda) = (f_{X,kl}(\lambda))_{k,l=1,\dots,p}$ given by, for $k < l$,

$$\begin{aligned} f_{X,kk}(\lambda) &= \frac{\sigma_{kk}}{2\pi}|1 - e^{-i\lambda}|^{-2d_k}, \\ f_{X,kl}(\lambda) &= \frac{\sigma_{kl}}{2\pi}(1 - e^{-i\lambda})^{-\delta_k}(1 - e^{i\lambda})^{-\delta_l}, \end{aligned} \tag{21}$$

	Number of frequencies m						
	$N^{0.35}$	$N^{0.4}$	$N^{0.45}$	$N^{0.5}$	$N^{0.55}$	$N^{0.65}$	$N^{0.7}$
size	0.089	0.089	0.149	0.228	0.433	0.755	0.958
power	0.497	0.673	0.879	0.976	0.998	1	1

TABLE I
SIZES AND POWERS FOR THE FRACTAL NON-CONNECTIVITY TEST.

where $0 < \delta_k < d_k < 1/2$, $k = 1, \dots, p$, and $\sigma_{kl} \neq 0$, $\sigma_{kk} > 0$. For fractally connected model, we take the same spectral density as in (21) but with $\delta_k = d_k$. The autocovariance functions for these series can be computed explicitly and the Gaussian series can be generated exactly following [23]. For the simulation study, we take $p = 5$, $d = (d_1, \dots, d_5) = (0.1, 0.2, 0.25, 0.3, 0.4)$, $\delta_k = (\delta_1, \dots, \delta_5) = 0.1d$, $\sigma_{kk} = 1$, $\sigma_{kl} \in [0.05, 0.5]$ and the sample size $N = 1000$. For the exact values of σ_{kl} used in the simulation, see [13].

Table I presents the empirical sizes and powers of the fractal non-connectivity test as functions of the tuning parameters m in the local Whittle estimation. As can be seen from the table, the test is even slightly oversized for smaller numbers of frequencies m.

To clarify larger than nominal empirical sizes, we fix the number of frequencies to $m = N^{0.45}$ and consider pairwise testing. It turns out that the test is very much effected by off-diagonal values of σ_{kl}. Applying the test to each possible pair $(X_{k,n}, X_{l,n})$, $k, l = 1, \dots, 5$, one may observe that the test performs well for pairs related to smaller off-diagonal elements σ_{kl}. However, for higher values such as $\sigma_{13} = 0.5$, the pair $(X_{1,n}, X_{3,n})$ yields the empirical size of 0.227.

One may ask further how the pairwise tests interplay. For this, we consider the empirical probabilities for rejecting a pairwise hypothesis given another pairwise hypothesis has already been rejected. This study reveals that rejecting the pairwise hypothesis, which corresponds to the maximal off-diagonal value $\sigma_{13} = 0.5$, increases the probability of rejecting the other pairwise hypotheses. This gets reflected in the larger sizes in Table I. See [13] for more details.

VI. CONCLUSIONS

In this work, we established the asymptotic normality result for the local Whittle estimators for general dimension p and formulation of short/long memory. The results were applied in connection to fractal non-connectivity and sparse estimation.

As possible future directions, a model which allows for fractional cointegration could be studied through the lens of this work, a suitable set of local Whittle plots to examine in practice could be decided upon (in the spirit of analogous suggestions in [5] when $p = 2$), the high-dimensional setting of $p, N \to \infty$ simultaneously could be studied, etc.

ACKNOWLEDGMENT

This work was carried out during a stay of the first author in the Department of Statistics and Operation Research at the University of North Carolina, Chapel Hill. The first author thanks the department for its hospitality, and also the DFG (RTG 2131) for financial support. The second author was supported in part by the NSF grant DMS-1712966.

REFERENCES

[1] M. Ø. Nielsen, “Local Whittle analysis of stationary fractional cointegration and the implied realized volatility relation,” *Journal of Business & Economic Statistics*, vol. 25, no. 4, pp. 427–446, 2007.

[2] M. Ø. Nielsen and K. Shimotsu, “Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach,” *Journal of Econometrics*, vol. 141, no. 2, pp. 574–596, 2007.

[3] P. M. Robinson, “Multiple local Whittle estimation in stationary systems,” *The Annals of Statistics*, vol. 36, no. 5, pp. 2508–2530, 10 2008.

[4] F. S. Nielsen, “Local Whittle estimation of multi-variate fractionally integrated processes,” *Journal of Time Series Analysis*, vol. 32, no. 3, pp. 317–335, 2011.

[5] C. Baek, S. Kechagias, and V. Pipiras, “Asymptotics of bivariate local Whittle estimators with applications to fractal connectivity,” Preprint, 2019. Available at http://pipiras.web.unc.edu.

[6] J. Beran, Y. Feng, S. Ghosh, and R. Kulik, *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer-Verlag Berlin Heidelberg, 2013.

[7] V. Pipiras and M. Taqqu, *Long-Range Dependence and Self-Similarity*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017.

[8] S. Kechagias and V. Pipiras, “Definitions and representations of multivariate long-range dependent time series,” *Journal of Time Series Analysis*, vol. 36, no. 1, pp. 1–25, 2015.

[9] K. Shimotsu, “Gaussian semiparametric estimation of multivariate fractionally integrated processes,” *Journal of Econometrics*, vol. 137, no. 2, pp. 277–310, 2007.

[10] S. Achard, D. S. Bassett, A. Meyer-Lindenberg, and E. Bullmore, “Fractal connectivity of long-memory networks,” *Physical Review E*, vol. 77, p. 036104, 2008.

[11] H. Wendt, A. Scherrer, P. Abry, and S. Achard, “Testing fractal connectivity in multivariate long memory processes,” in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 2913–2916.

[12] H. Wendt, G. Didier, S. Combrexelle, and P. Abry, “Multivariate Hadamard self-similarity: testing fractal connectivity,” *Physica D*, vol. 356–357, pp. 1–36, 2017.

[13] M.-C. Düker and V. Pipiras, “Technical appendix for “Asymptotic results for multivariate local Whittle estimation with applications”,” Preprint, 2019. Available at http://pipiras.web.unc.edu.

[14] J. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, ser. Probabilistics and Statistics. Wiley, 2007.

[15] R. A. Horn and C. R. Johnson, Eds., *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 2013.

[16] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[17] M. Yuan and Y. Lin, “Model selection and estimation in the Gaussian graphical model,” *Biometrika*, vol. 94, no. 1, pp. 19–35, 2007.

[18] J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covariance estimation with the graphical lasso,” *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[19] J. Bien and R. J. Tibshirani, “Sparse estimation of a covariance matrix,” *Biometrika*, vol. 98, no. 4, pp. 807–820, 2011.

[20] C. Baek, S. Kechagias, and V. Pipiras, “Semiparametric, parametric, and possibly sparse models for multivariate long-range dependence,” in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 10394, 2017, p. 103941S.

[21] K. Knight and W. Fu, “Asymptotics for lasso-type estimators,” *The Annals of Statistics*, vol. 28, no. 5, pp. 1356–1378, 2000.

[22] H. Zou, “The adaptive lasso and its oracle properties,” *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.

[23] H. Helgason, V. Pipiras, and P. Abry, “Fast and exact synthesis of stationary multivariate gaussian time series using circulant embedding,” *Signal Processing*, vol. 91, no. 5, pp. 1123–1133, 2011.