

# Local Whittle estimation of high-dimensional long-run variance and precision matrices <sup>\*†‡</sup>

Changryong Baek  
Sungkyunkwan University

Marie-Christine Düker  
Cornell University

Vladas Pipiras  
University of North Carolina

December 28, 2022

## Abstract

This work develops non-asymptotic theory for estimation of the long-run variance matrix and its inverse, the so-called precision matrix, for high-dimensional time series under general assumptions on the dependence structure including long-range dependence. The estimation involves shrinkage techniques which are thresholding and penalizing versions of the classical multivariate local Whittle estimator. The results ensure consistent estimation in a double asymptotic regime where the number of component time series is allowed to grow with the sample size as long as the true model parameters are sparse. The key technical result is a concentration inequality of the local Whittle estimator for the long-run variance matrix around the true model parameters. In particular, it handles simultaneously the estimation of the memory parameters which enter the underlying model. Novel algorithms for the considered procedures are proposed, and a simulation study and a data application are also provided.

## 1 Introduction

Spectral density matrices characterize the component and temporal dependence of multivariate time series, and its estimation is of interest in many areas, including economics and neuroscience. The long-run variance and precision matrices give, respectively, information about correlations and partial correlations between different component series around zero frequency; see Dahlhaus (2000). Their estimation are the frequency domain analogues of covariance and inverse covariance estimation; see Fan, Liao, and Liu (2016) for a survey on large (inverse) covariance matrix estimation. Obtaining an estimate for the spectral density matrix can become particularly challenging in a high-dimensional regime when the number of component series becomes relatively large compared to the length of the time series. In this regime, estimation often employs different shrinkage methods. The development, theoretical verification and application of different shrinkage methods has been an active research area, along with a growing interest in non-asymptotic theory in high-dimensional statistics; see Wainwright (2019) for a survey on non-asymptotic theory.

---

\*AMS subject classification. Primary: 62M15, 62H12. Secondary: 62H20.

†Keywords and phrases: High-dimensional time series, frequency domain, short- and long-range dependence, spectral density estimation, shrinkage estimation, local Whittle estimation.

‡This work was carried out during stays of the first and second authors in the Department of Statistics and Operation Research at the University of North Carolina, Chapel Hill. The first and second authors thank the department, in particular, Vladas Pipiras for their hospitality. The first author was supported by the National Research Foundation of Korea (NRF-2019R1F1A1057104, NRF-2022R1F1A1066209). The second author was supported by the DFG (RTG 2131) and the NSF grant 1934985. The third author was supported in part by the NSF grant DMS-1712966.

This paper develops non-asymptotic theory for estimation of the long-run variance and precision matrices of a stationary multivariate time series around zero frequency while allowing for short- and long-range dependence. The local Whittle estimation is used with thresholding and LASSO-type penalizations. Our non-asymptotic theory allows to infer consistency results on the estimators around the true parameters in a high-dimensional regime where the number of component series can be large compared to the number of observations. We also note that our non-asymptotic results are new even for the one-dimensional case.

Our setting is as follows. 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  $\Gamma_X(k) = \mathbb{E}X_{n+k}X_n'$ ,  $k \in \mathbb{Z}$ . Suppose that its spectral density matrix  $f_X(\lambda)$ ,  $\lambda \in (-\pi, \pi)$ , related to the autocovariance matrix through  $\Gamma_X(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f_X(\lambda) d\lambda$ , satisfies

$$f_X(\lambda) = \lambda^{-D_0} G(\lambda) \lambda^{-D_0}, \quad G(\lambda) \sim G_0, \quad \text{as } \lambda \rightarrow 0^+, \quad (1.1)$$

where  $\sim$  denotes componentwise asymptotic equivalence,  $D_0 = \text{diag}(d_{0,1}, \dots, d_{0,p})$  with  $d_{0,r} \in (-1/2, 1/2)$ ,  $r = 1, \dots, p$ ,  $\lambda^{-D_0} = \text{diag}(\lambda^{-d_{0,1}}, \dots, \lambda^{-d_{0,p}})$  and  $G_0 = (G_{0,rs})_{r,s=1,\dots,p}$  is Hermitian symmetric and positive definite. Each individual component series  $X_{r,n}$ ,  $n \in \mathbb{Z}$ , satisfies (1.1) with memory parameter  $d_{0,r}$ . The case  $d_{0,r} = 0$  is associated with short-range dependence, the case  $d_{0,r} > 0$  with long-range dependence and  $d_{0,r} < 0$  with antipersistence. We refer to Beran, Feng, Ghosh, and Kulik (2013), Pipiras and Taqqu (2017) for more details on univariate short- and long-range dependence and Kechagias and Pipiras (2015a) for a discussion on multivariate long-range dependence. The matrix  $G_0$  is the long-run variance matrix and  $P_0 = G_0^{-1}$  is the precision matrix. They are the focus of this work.

In the presence of long-range dependence, local Whittle estimation is commonly used to estimate the parameters  $(D_0, G_0)$  of the model (1.1). We introduce here the classical multivariate local Whittle estimators for  $(D_0, G_0)$  and refer to Section 2 for a detailed explanation of the used shrinkage techniques. In particular, we aim to utilize a thresholding technique to estimate the long-run variance matrix  $G_0$  sparsely (see (2.2) below) and a LASSO-type estimator to estimate the precision matrix  $P_0$  sparsely (see (2.3) below).

The local Whittle estimators  $(\hat{D}, \hat{G})$  introduced in Robinson (2008) are given by

$$(\hat{D}, \hat{G}) = \underset{(D, G)}{\text{argmin}} \ell(D, G) \quad (1.2)$$

for the negative log-likelihood

$$\ell(D, G) = \frac{1}{m} \sum_{j=1}^m (\log |\lambda_j^{-D} G \lambda_j^{-D}| + \text{tr}(I_X(\lambda_j) \lambda_j^D G^{-1} \lambda_j^D)), \quad (1.3)$$

where  $|\cdot| := \det(\cdot)$  and  $\text{tr}(\cdot)$  denote the determinant and the trace of a matrix,

$$I_X(\lambda) = \frac{1}{2\pi N} \left( \sum_{n=1}^N X_n e^{in\lambda} \right) \left( \sum_{n=1}^N X_n e^{in\lambda} \right)^* \quad (1.4)$$

is the periodogram for sample size  $N$  and  $m$  is the number of Fourier frequencies  $\lambda_j = 2\pi j/N$  used in estimation. The optimization problem (1.3) can be reduced to

$$\hat{D} = \underset{D}{\text{argmin}} R(D) \quad \text{with} \quad R(D) = \frac{1}{m} \sum_{j=1}^m \log |\lambda_j^{-D} \hat{G}(D) \lambda_j^{-D}|, \quad (1.5)$$

where

$$\widehat{G}(D) = \frac{1}{m} \sum_{j=1}^m \lambda_j^D I_X(\lambda_j) \lambda_j^D. \quad (1.6)$$

Local Whittle estimation was studied by multiple authors. Robinson (1995b) showed consistency and asymptotic normality of the univariate local Whittle estimators. In the bivariate case  $p = 2$ , the asymptotic normality of the local Whittle estimators of memory parameters  $d_{0,1}, d_{0,2}$  was established in Robinson (2008), and that of all model parameters in Baek, Kechagias, and Pipiras (2020). Asymptotic normality results in special cases of (1.1) but general fixed dimension  $p$  appear in Shimotsu (2007), Nielsen (2011). In Düker and Pipiras (2019), an asymptotic normality result for the local Whittle estimators (1.2) of all model parameters and general fixed  $p$  was obtained.

The graphical local Whittle estimator is an  $l_1$ -penalized version of the negative log-likelihood in (1.3), as proposed in Baek, Kechagias, and Pipiras (2017) with the focus on its good numerical performance. It can also be written as a function of the precision matrix  $P_0$ . We refer to Section 2 and equation (2.4) below for a detailed description of the penalized objective function. The objective function coincides with that used in estimating covariance matrices sparsely; see Bien and Tibshirani (2011). On the other hand, for a penalization of the respective inverse, it coincides with the graphical LASSO estimator; see Friedman, Hastie, and Tibshirani (2008). Düker and Pipiras (2019) derived asymptotic results for the estimators of the long-run variance matrix and the precision matrix in a “fixed  $p$ , large  $N$ ” regime under the discussed  $l_1$ -penalization.

Sparse covariance and its precision matrix estimation were studied by numerous authors. LASSO-type estimators were investigated by Rothman, Bickel, Levina, and Zhu (2008), Cai, Liu, and Luo (2011) and Shu and Nan (2019). See also Cai, Ren, and Zhou (2016) for a review of recent developments. Thresholding based strategies were pursued by Bickel and Levina (2008a,b), Rothman, Levina, and Zhu (2009) and Cai and Liu (2011). In contrast, there is less research work for high-dimensional spectral density matrix estimation.

The works of Shu and Nan (2019), Sun, Li, Kuceyeski, and Basu (2018) and Fiecas, Leng, Liu, and Yu (2019) are probably the closest to our work. Shu and Nan (2019) considered the estimation of large covariance and its precision matrices from high-dimensional sub-Gaussian or heavier-tailed observations with slowly decaying temporal dependence. Sun et al. (2018) developed a non-asymptotic theory for estimation of the spectral density matrix of multivariate time series under short-range dependence, that is, when  $D_0 \equiv 0$  in (1.1). The work in Fiecas et al. (2019) developed some non-asymptotic theory for estimation of the spectral density matrix and its inverse for a class of time series exhibiting short-range dependence under a mixing condition. In contrast, we allow for a quite general dependence structure including short- and long-range dependence and antipersistence.

From a practical perspective estimating the spectral density matrix and its inverse have applications in many fields including signal processing (Schneider-Luftman and Walden, 2016), neuroscience (Fiecas and Ombao, 2011; Bowyer, 2016; Bordier, Nicolini, and Bifone, 2017) and economics (Granger, 1969; Hansen and Sargent, 1983; Politis, 2011; Plagborg-Møller and Wolf, 2021; Cavicchioli, 2022). The spectral density matrix captures contemporaneous correlation and correlation across different lags. It therefore provides a richer description of the dependence structure in a multivariate time series than the covariance matrix.

The literature review shows that there is a gap in theoretical results concerning high-dimensional spectral density estimation for time series possibly exhibiting long-range dependence. We are the first to provide non-asymptotic theoretical results for thresholding and graphical local Whittle estimation which allow to infer consistency in a possibly double asymptotic regime of large  $p$  and  $N$ . The presence of long-range dependence and the simultaneous estimation of the memory parameters

$D_0$ , make it particularly challenging to derive non-asymptotic results. We overcome those challenges by using a uniform concentration inequality and controlling the difference between the sample and the population version of the matrix  $D_0$  simultaneously. Our theoretical results turn out to be useful not only for thresholding and graphical local Whittle estimators but can be applied to derive consistency for other kinds of penalized estimators. We demonstrate that by deriving consistency results for estimators based on the coherence matrix and a constrained  $l_1$ -minimization (CLIME). We also address the question of consistent model selection by adopting different thresholding procedures to the spectral setting. Additionally, we introduce novel algorithms to compute the thresholded and penalized local Whittle estimators. The results are accompanied by a simulation study which assesses the numerical performance of the suggested algorithms and estimators.

The rest of the paper is organized as follows. In Section 2, we discuss our estimation procedure and present some assumptions required for our theoretical analysis. In Section 3, we present an outline of the proof, our main results and some discussions of those results. Appendix A provides more technical details for the statements of our results, allowing to keep the notation in the paper's main body shorter. In Section 4, we introduce two algorithms to compute the penalized graphical local Whittle estimators. The performance of those algorithms is analyzed in a simulation study conducted in Section 5 with complementary results in Appendix F. An application can be found in Section 6. We conclude with Section 7. The proofs can be found in Appendix B. In Appendices C and D, we provide some technical results and their proofs. Finally, Appendix E provides the proofs for an extension to linear processes.

*Notation:* For the reader's convenience, we give a collection of notation used throughout the paper. We denote the maximum and minimum eigenvalues of a symmetric or Hermitian matrix  $A$  by  $\lambda_{\max}(A)$  and  $\lambda_{\min}(A)$ , respectively. To indicate that a matrix  $A$  is positive (semi-)definite, we write  $A \succ 0$  ( $A \succeq 0$ ). We use a range of different matrix norms, namely, the maximum norm, the spectral norm and the Frobenius norm, defined respectively as  $\|A\|_{\max} = \max_{1 \leq r, s \leq p} |A_{rs}|$ ,  $\|A\| = \sqrt{\lambda_{\max}(A'A)}$  and  $\|A\|_F = \sqrt{\text{tr}(A'A)}$  for a matrix  $A$ . We use  $e_r$  to denote the  $r$ th unit vector in  $\mathbb{R}^p$  for  $r = 1, \dots, p$ . For the vectorized version of a matrix  $A$ , we write  $\text{vec}(A)$ . The  $\text{vec}$  operator transforms a matrix into a vector by stacking its columns one underneath the other. For a  $p \times N$  matrix, composed of  $N$   $p$ -dimensional vectors  $v_1, \dots, v_N$ , we write  $[v_1 : \dots : v_N]$ . We let  $L^2(0, 1)$  be the space of square-integrable functions on  $(0, 1)$  with respect to the Lebesgue measure. If  $A$  is an integral operator on  $L^2(0, 1)$  of the form  $(Af)(x) = \int_0^1 k(x, y)f(y)dy$ , then  $A$  is called Hilbert-Schmidt if and only if

$$\int_0^1 \int_0^1 |k(x, y)|^2 dx dy < \infty, \quad (1.7)$$

where the double integral in (1.7) is denoted as  $\|A\|_2^2$  and called the Hilbert-Schmidt norm. Let further  $A : V \rightarrow W$  be a linear operator with normed spaces  $V, W$ . We write  $\|A\|_{op} = \sup_{x \neq 0} \|Ax\|_W / \|x\|_V$  for the operator norm of  $A$ , where  $\|\cdot\|_W$  denotes the norm on  $W$ . As a further convention we write  $a \gtrsim b$  if there exists a universal constant  $c$  such that  $a \geq cb$ . We further use the notation  $\frac{\partial}{\partial x}$  to denote the partial derivative with respect to  $x$  and  $\nabla$  to denote the gradient  $\nabla f = \sum_{r=1}^p e_r \frac{\partial}{\partial x_r} f$  of a function  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ .

## 2 Estimation methods and assumptions

In this section, we formulate the long-run variance and precision matrix estimation through the thresholding and graphical local Whittle estimators, respectively. Furthermore, we give the required assumptions to prove non-asymptotic bounds which ensure consistency results for the estimators in a double asymptotic regime of large  $p$  and  $N$ .

The thresholding and graphical local Whittle estimators require estimation of the memory parameters  $D_0 = \text{diag}(d_{0,1}, \dots, d_{0,p})$ . We propose here to estimate each  $d_{0,r}, r = 1, \dots, p$ , by the univariate local Whittle estimator; see Remark 2.2 below for a discussion of this. We introduce a notation different from that used for the multivariate local Whittle estimators in (1.2) to emphasize the use of the univariate version of the local Whittle estimator. For a multivariate time series satisfying (1.1), each individual, univariate time series  $\{X_{r,n}\}_{n \in \mathbb{Z}}, r = 1, \dots, p$ , satisfies

$$f_{X,rr}(\lambda) = \lambda^{-2d_{0,r}} g_r(\lambda), \quad g_r(\lambda) \sim g_{0,r}, \quad \text{as } \lambda \rightarrow 0^+,$$

where  $g_r(\lambda) = G_{rr}(\lambda)$ ,  $g_{0,r} = G_{0,rr}$ , and  $f_{X,rs}(\lambda)$  and  $G_{rs}(\lambda)$  denote the  $(r, s)$ th entry of  $f_X(\lambda)$  and  $G(\lambda)$ , respectively. Then, the univariate local Whittle estimator for  $d_{0,r}$  is given by

$$\hat{d}_r = \underset{d \in \Theta}{\text{argmin}} R_r(d) \quad \text{with} \quad R_r(d) = \frac{1}{m} \sum_{j=1}^m \log(\lambda_j^{-2d} \hat{g}_r(d)), \quad (2.1)$$

where the set of admissible estimates is defined as  $\Theta = \{d \mid \Delta_1 \leq d \leq \Delta_2\}$  with  $-\frac{1}{2} < \Delta_1 < \Delta_2 < \frac{1}{2}$  and

$$\hat{g}_r(d) = \frac{1}{m} \sum_{j=1}^m \lambda_j^{2d} I_{X,rr}(\lambda_j),$$

where  $I_{X,rs}(\lambda)$  denotes the  $(r, s)$ th entry of the periodogram  $I_X(\lambda)$  in (1.4). After estimating each individual memory parameter  $d_{0,r}$  by (2.1), we want to estimate the long-run variance matrix  $G_0$  and the precision matrix  $P_0$  sparsely by thresholding and graphical local Whittle estimation, respectively.

*Thresholding local Whittle:* We propose to use hard thresholding to estimate the long-run variance matrix sparsely. (Soft or adaptive thresholding could also be used.) That is,

$$T_\rho(\hat{G}_{rs}(\hat{D})) = \begin{cases} \hat{G}_{rs}(\hat{D}), & \text{if } |\hat{G}_{rs}(\hat{D})| \geq \rho, \\ 0, & \text{if } |\hat{G}_{rs}(\hat{D})| < \rho, \end{cases} \quad (2.2)$$

where  $\rho > 0$  is a threshold and  $T_\rho(\cdot)$  is a thresholding operator applied to  $\hat{G}_{rs}(\hat{D})$ , the  $(r, s)$ th entry of the estimator for the long-run variance matrix (1.6) and the components of  $\hat{D}$  are estimated univariately by (2.1).

*Graphical local Whittle:* The precision matrix  $P_0 = G_0^{-1}$  can be estimated sparsely by the graphical local Whittle estimator, a penalized version of the negative log-likelihood function (1.3). The penalized estimator  $\hat{P}_\rho$  is given by

$$\hat{P}_\rho = \underset{P \succ 0}{\text{argmin}} \ell_\rho(\hat{D}, P), \quad (2.3)$$

where  $\hat{D}$  is estimated univariately by (2.1) and

$$\ell_\rho(D, P) = -\frac{1}{m} \sum_{j=1}^m \log |\lambda_j^D P \lambda_j^D| + \text{tr}(\hat{G}(D)P) + \rho \|P\|_{1,off} \quad (2.4)$$

with a penalty parameter  $\rho > 0$  and the  $l_1$ -norm  $\|\cdot\|_{1,off}$  excluding the diagonal elements.

Next, we give some assumptions, required to establish our theoretical results. Other assumptions appear in the statements of our results. Subsequently, we discuss those assumptions in several remarks.

**Assumption 1.** Suppose that

$$f_X(\lambda) = \lambda^{-D_0} G(\lambda) \lambda^{-D_0}, \quad G(\lambda) \sim G_0, \quad (2.5)$$

where  $\sim$  denotes componentwise asymptotic equivalence,  $G_0 = (G_{0,rs})_{r,s=1,\dots,p}$  is Hermitian symmetric and positive definite and  $D_0 \in \{D \in \mathcal{M}_{\text{diag}} \mid \Delta_1 I_p \preceq D \preceq \Delta_2 I_p\}$ , where  $\mathcal{M}_{\text{diag}}$  denotes the set of all real-valued diagonal matrices. We further suppose that the positive eigenvalues of  $G_0$  can be bounded from below as

$$\lambda_{\min}(G_0) \geq \underline{k} > 0. \quad (2.6)$$

**Assumption 2.** For some  $q \in (0, 1]$ , the spectral density matrix satisfies

$$|f_{X,rs}(\lambda_j) - \lambda_j^{-d_{0,r}-d_{0,s}} G_{0,rs}| \leq c_{G,1} \lambda_j^{2q-d_{0,r}-d_{0,s}} \quad (2.7)$$

for  $j = 1, \dots, m$  and some  $c_{G,1} > 0$ .

**Assumption 3.** The function  $f_{X,rs}(\lambda)$  is differentiable on  $\lambda \in (-\pi, \pi) \setminus \{0\}$  and there is a constant  $c_{G,2} > 0$  such that

$$\left| \frac{\partial}{\partial \lambda} f_{X,rs}(\lambda) \right| \leq c_{G,2} \lambda^{-1-d_{0,r}-d_{0,s}}.$$

Besides assumptions on the spectral density matrix of the underlying process, we will also impose some mild assumptions on the process itself. In particular, our results are valid not only for Gaussian time series but also for a large class of non-Gaussian processes. Our assumption will be formulated in terms of sub-Gaussian random variables, that is, their distribution is dominated by a centered Gaussian distribution. More precisely, we call a random variable  $X$  sub-Gaussian if there is a constant  $c$  such that

$$\mathbb{E}(|X|^r)^{\frac{1}{r}} \leq cr^{\frac{1}{2}} \quad \text{for all } r \geq 1.$$

We further denote  $\|X\|_\phi = \sup_{r \geq 1} r^{-\frac{1}{2}} \mathbb{E}(|X|^r)^{\frac{1}{r}}$ , the sub-Gaussian norm of a real-valued random variable  $X$ . Gaussian random variables belong to the class of sub-Gaussian random variables. We refer to Vershynin (2010) for more details on sub-Gaussian random variables.

**Assumption 4.** The time series  $\{X_n\}$  is assumed to be either Gaussian or to have a linear representation  $X_n = \sum_{j \in \mathbb{Z}} \Psi_j \varepsilon_{n-j}$  with  $\sum_{j \in \mathbb{Z}} \|\Psi_j\|_F^2 < \infty$  and independent mean 0 innovations  $\{\varepsilon_j\}_{j \geq 1}$ , where each component  $\varepsilon_{r,j}$ ,  $r = 1, \dots, p$  of the random vector  $\varepsilon_j$  is assumed to be sub-Gaussian, satisfying

$$\|\varepsilon_{r,j}\|_\phi \leq \gamma \quad (2.8)$$

for some constant  $\gamma \in (0, \infty)$ .

**Assumption 5.** The number of frequencies  $m = m(N)$  used in estimation and the lower bound of the interval of admissible estimates  $\Delta_1$  satisfy

$$m \gtrsim N^{-2\Delta_1}.$$

Our work intends to provide non-asymptotic results. However, we impose some mild assumptions on our choices of the number of frequencies  $m$  and the sample size  $N$  to simplify some of our bounds. Throughout the paper we suppose that the number of frequencies and the sample size satisfy  $m, N > 2$ . Those assumptions allow us to use  $\log(m) + 1 \leq 2 \log(m)$  and  $\log(m) + 1 \leq (\log(m) + 1)^2$ , and the same for  $N$ . Another assumption we impose is  $m \leq \frac{N}{2} - 1$  which ensures that the bound on the bias term of the periodogram is finite.

We use different measures of sparsity for the long-run variance and the precision matrices. Both are commonly used in the respective literatures.

In the context of thresholding, a commonly used measure of sparsity for the long-run variance matrix  $G_0$  is given by

$$\|G_0\|_a^a = \max_{r=1,\dots,p} \sum_{s=1}^p |G_{0,rs}|^a \quad (2.9)$$

for  $a \in [0, 1)$ . This measure was proposed in Bickel and Levina (2008b) and shown to capture a variety of sparsity patterns. It was further applied in the context of spectral density estimation in a non-asymptotic regime in Sun et al. (2018).

For the precision matrix, we define the set

$$S = \{(r, s) \mid P_{0,rs} \neq 0, r \neq s\} \quad (2.10)$$

and bound its cardinality with  $|S| \leq s$ .

We will also use

$$\|G\| = \operatorname{ess\,sup}_{\lambda \in (-\pi, \pi)} \|G(\lambda)\| = \operatorname{ess\,sup}_{\lambda \in (-\pi, \pi)} \|\lambda^{D_0} f_X(\lambda) \lambda^{D_0}\|$$

as a measure of stability of the time series  $\{X_n\}$ . This follows Basu and Michailidis (2015) and Sun et al. (2018) who considered the case  $G(\lambda) \equiv f_X(\lambda)$  and  $D_0 \equiv 0$ , which is associated with short-range dependence of the underlying time series. See also the second paragraph of Section 2 in Sun et al. (2018) for a discussion on how  $\|G\|$  acts as a measure of stability.

The following remarks comment on the model, estimation procedure and on the assumptions above. Remark 2.1 comments on the model and Remark 2.2 concerns estimating the memory parameters univariately, Remark 2.3 is on Assumptions 1 and 2, and Remarks 2.4, 2.5 and 2.6 are on Assumptions 3, 4 and 5, respectively.

**Remark 2.1.** In this work, we assume that the matrix  $G_0$  in (1.1) is possibly complex valued Hermitian symmetric. In order to achieve sparsity, both real and imaginary part need to be zero. Related literature has also studied an alternative way to parametrize the matrix  $G_0$ . Proposed by Robinson (2008) and further studied in Düker and Pipiras (2019) and Baek et al. (2020), one can write  $G_0$  in terms of polar coordinates, that is,

$$G_0 = (\omega_{kl} e^{\operatorname{sign}(k-l)i\phi_{kl}})_{k,l=1,\dots,p}$$

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 uncorrelatedness between component series ( $\omega_{kl} e^{\operatorname{sign}(k-l)i\phi_{kl}} = 0$ ), since the respective phase parameter  $\phi_{kl}$  is not identifiable for  $\omega_{kl} = 0$ ,  $k \neq l$ ; see Düker and Pipiras (2019) and Baek et al. (2020) for a related discussion.

**Remark 2.2.** Our proposed estimation procedure involves estimating the memory parameters  $d_{0,1}, \dots, d_{0,p}$  by the univariate local Whittle estimators (2.1) rather than using the multivariate estimator of the matrix  $D_0 = \operatorname{diag}(d_{0,1}, \dots, d_{0,p})$  in (1.5). The reasons are twofold, one is theoretical, the other computational.

The theoretical reason is that getting a concentration inequality on  $\|\widehat{D} - D_0\|_{\max}$  for the multivariate estimates of  $D_0$  involves a concentration inequality on  $|\log |\widehat{G}(\widehat{D})| - \log |G_0||$ . In the asymptotic regime  $N \rightarrow \infty$  and for fixed dimension  $p$ , a consistency result for  $|\log |\widehat{G}(\widehat{D})| - \log |G_0||$  can be achieved easily by combining the continuous mapping theorem and a consistency result on  $\|\widehat{G}(\widehat{D}) - G_0\|_{\max}$ . However, our non-asymptotic setting involves an inequality of the form  $|\log |\widehat{G}(\widehat{D})| - \log |G_0|| \leq Cp \|\widehat{G}(\widehat{D}) - G_0\|$ , with a generic constant  $C$ . The additional  $p$  weakens

the results in the sense that  $p$  has to grow much slower than  $m$  in order to achieve consistency. Bounding further as  $\|\widehat{G}(\widehat{D}) - G_0\| \leq p\|\widehat{G}(\widehat{D}) - G_0\|_{\max}$  results in an additional  $p$ . A potential way to avoid the second  $p$  one gets through bounding the operator norm might be to impose a sparsity assumption on  $G_0$  and threshold  $\widehat{G}$  in the objective function (1.5). That is, estimating  $D_0$  as in (1.5) involves an estimator for  $G_0$ . However, the estimator  $\widehat{D}$  for  $D_0$  is not based on a thresholded version of  $\widehat{G}$ . One possibility to address this issue is to introduce a shrinkage on  $\widehat{G}$  in (1.5) by using a thresholded version of  $\widehat{G}$ .

On the other hand, computationally, it is faster to minimize  $p$  univariate functions as opposed to optimizing a matrix function over a certain set of diagonal matrices. In a simulation study in Appendix F.3, we show that the difference between the multivariate and univariate estimates is negligible.

**Remark 2.3.** Assumption 1 with (2.5) coincides with the basic model (1.1) and supposes additionally that the true memory parameters  $D_0$  are contained in the interval of admissible estimates. Besides assuming that the matrix  $G_0$  is positive definite, we suppose in (2.6) that the eigenvalues of  $G_0$  are bounded from below. This is a typical assumption in sparse covariance estimation; see Rothman et al. (2008). Assumption 2 is a smoothness condition and controls the second order terms of the spectral density matrix. Usually, the componentwise relation  $f_X(\lambda) = \lambda^{-D_0} G_0 \lambda^{-D_0} (1 + O(\lambda^{2q}))$ , as  $\lambda \rightarrow 0^+$ , is imposed to derive asymptotic results in the context of spectral density estimation. However, we require a slightly stronger assumption (2.7) in order to control the bias terms to derive non-asymptotic results.

**Remark 2.4.** Assumption 3 is required to ensure that the bias term is asymptotically negligible. This kind of assumption appears in the asymptotic literature regarding local Whittle estimation as well; see Assumption A.2 in Robinson (1995b) and Assumption A.1 in Robinson (2008). However, those assumptions typically only require differentiability in an epsilon region around the origin. We need to impose differentiability in a region which includes all frequencies used in estimation and allows for all choices of  $m = 1, \dots, \frac{N}{2}$ .

**Remark 2.5.** Our main assumptions on the underlying process are the parametrization of the spectral density in terms of the matrices  $(D_0, G_0)$  as formalized in Assumption 1, and Assumption 4 which ensures that the series is either Gaussian or follows a linear representation. Though, we require the innovations of the linear representation to be sub-Gaussian, our results can be used to derive statements for sub-exponential innovations or assuming finite fourth moments; see Remark E.1 for a more detailed discussion. Our assumptions allow for quite general long- and short-range dependent linear time series. For long-range dependence, examples are multivariate FARIMA series as defined in Kechagias and Pipliras (2015a). For short-range dependence, examples are multivariate ARMA models.

**Remark 2.6.** Assumption 5 is satisfied, in particular, when  $\Delta_1 = 0$ , that is, when the underlying time series exhibits only short- or long-range dependence. In other words, the assumption is only needed when the true memory parameters  $d_{0,r}$ ,  $r = 1, \dots, p$ , are known to take also values in  $(-\frac{1}{2}, 0)$ . The case  $d_{0,r} < 0$  contributes to the non-asymptotic bounds in our main results in form of two terms  $mN^{2\Delta_1}$  and  $m \left( \sum_{j=1}^m \lambda_j^{4\Delta_1} \right)^{-1}$ . Assumption 5 is necessary to ensure that our non-asymptotic bounds prove consistency, which is the case as long as both quantities go to infinity while the sample size increases. Assumption 5 not only controls  $mN^{2\Delta_1}$  but is also sufficient to control the second quantity, since

$$\sum_{j=1}^m \lambda_j^{4\Delta_1} \leq cN^{-4\Delta_1} \sum_{j=1}^m j^{4\Delta_1} \leq cN^{-4\Delta_1} \begin{cases} \frac{1}{4\Delta_1+1} m^{4\Delta_1+1}, & \text{if } \Delta_1 \in (-\frac{1}{4}, 0), \\ \log(m) + 1, & \text{if } \Delta_1 \leq -\frac{1}{4}. \end{cases}$$



### 3 Main results

In this section, we present our main results. Section 3.1 provides a roadmap for our proofs which reveals what kind of results are necessary to prove consistency for both the thresholding and graphical local Whittle estimation. This includes in particular a consistency result on the maximum norm of  $\widehat{G}(\widehat{D}) - G_0$ . Subsequently, we formally state our main results in Section 3.2, that is, consistency results for the thresholding and graphical local Whittle estimators. Section 3.3 discusses alternative estimators for precision matrix estimation and their convergence rates. We provide results on consistent model selection in Section 3.4. In Section 3.5, we discuss how our results compare to existing results in the literature.

#### 3.1 Proof idea

In contrast to the spectral density estimation under short-range dependence, allowing for long-range dependence and antipersistence requires estimation of two different kinds of model parameters, the matrix  $G_0$  and the memory parameters  $d_{0,r}$ ,  $r = 1, \dots, p$ . For this reason, deriving a concentration inequality becomes particularly challenging. Results for the graphical and thresholding local Whittle estimators require a concentration inequality on the event

$$\{\|\widehat{G}(\widehat{D}) - G_0\|_{\max} > \delta\}. \quad (3.1)$$

For this, our theoretical analysis reveals that the memory parameters  $d_{0,r}$ ,  $r = 1, \dots, p$ , have to be controlled simultaneously, and we propose to derive a concentration inequality on the event (3.1) by incorporating the event  $\{\|\widehat{D} - D_0\|_{\max} \leq \varepsilon\}$ , and then use multiple bounds of the probability of the event of interest (3.1) in terms of events which are representable as quadratic forms of i.i.d. sub-Gaussian random vectors. A key tool in our analysis is a uniform concentration inequality introduced by Dicker and Erdogdu (2017). For completeness, we present a slightly modified version of their result in Appendix D.1.

Before we present the main results, we introduce some further notation. We write the population analogue of  $\widehat{G}(D)$  in (1.6) as

$$\widetilde{G}(D) = \frac{1}{m} \sum_{j=1}^m \lambda_j^{D-D_0} G_0 \lambda_j^{D-D_0} \quad (3.2)$$

and the respective univariate counterpart as  $\widetilde{g}_r(d_r) = \widetilde{G}_{rr}(D)$ . Furthermore, we write

$$L(d) = \int_0^1 x^{2d} dx = \frac{1}{2d+1} \text{ for } d > -1/2. \quad (3.3)$$

We now present the aforementioned inequalities on the probability of the event (3.1), which give insights into what kind of concentration inequalities are required to prove the desired consistency result on  $\|\widehat{G}(\widehat{D}) - G_0\|_{\max}$ . A detailed analysis can be found in the proof of Proposition 3.5 below. For some  $\delta, \varepsilon > 0$ ,

$$\begin{aligned} & \mathbb{P}(\|\widehat{G}(\widehat{D}) - G_0\|_{\max} > \delta) \\ &= \mathbb{P}\left(\{\|\widehat{G}(\widehat{D}) - G_0\|_{\max} > \delta\} \cap \left(\{\|\widehat{D} - D_0\|_{\max} \leq \varepsilon\} \cup \{\|\widehat{D} - D_0\|_{\max} > \varepsilon\}\right)\right) \\ &\leq \mathbb{P}(\{\|\widehat{G}(\widehat{D}) - \widetilde{G}(\widehat{D})\|_{\max} > \delta/2\} \cap \{\|\widehat{D} - D_0\|_{\max} \leq \varepsilon\}) \\ &\quad + \mathbb{P}\left(\left(\{\|\widetilde{G}(\widehat{D}) - G_0\|_{\max} > \delta/2\} \cap \{\|\widehat{D} - D_0\|_{\max} \leq \varepsilon\}\right) \cup \{\|\widehat{D} - D_0\|_{\max} > \varepsilon\}\right) \end{aligned}$$

$$\leq \mathbb{P}(\{\|\widehat{G}(\widehat{D}) - \widetilde{G}(\widehat{D})\|_{\max} > \delta/2\} \cap \{\|\widehat{D} - D_0\|_{\max} \leq \varepsilon\}) + \mathbb{P}(\|\widehat{D} - D_0\|_{\max} > \eta) \quad (3.4)$$

$$\leq \sum_{r,s=1}^p \mathbb{P}(\sup_{D \in \Omega(\varepsilon)} |\widehat{G}_{rs}(D) - \widetilde{G}_{rs}(D)| > \delta/2) + \mathbb{P}(\|\widehat{D} - D_0\|_{\max} > \eta) \quad (3.5)$$

with

$$\Omega(\varepsilon) = \{D \in \mathcal{M}_{\text{diag}} \mid \Delta_1 I_p \preceq D \preceq \Delta_2 I_p \text{ and } \|D - D_0\|_{\max} \leq \varepsilon\}$$

and  $\eta = \min\{\varepsilon, \frac{\delta}{4}(\|G_0\| \log(N) \lambda_m^{-2\varepsilon} L(-\varepsilon))^{-1}\}$ . The relation (3.4) will follow from Lemma C.6. This way, the problem reduces to finding a uniform concentration inequality on  $|\widehat{G}_{rs}(D) - \widetilde{G}_{rs}(D)|$  and a concentration inequality on  $\|\widehat{D} - D_0\|_{\max}$ . Instead of considering the maximum, we bound the probability of  $\{\|\widehat{D} - D_0\|_{\max} > \eta\}$  componentwise for each  $r = 1, \dots, p$  as

$$\begin{aligned} \mathbb{P}(|\widehat{d}_r - d_{0,r}| > \eta) &\leq \mathbb{P}(|\widehat{g}_r(d_{0,r}) - \widetilde{g}_r(d_{0,r})| > \eta_1) \\ &\quad + \mathbb{P}(\sup_{d \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^m \binom{j}{m}^{2d-2d_{0,r}} (\lambda_j^{2d_{0,r}} I_{X,rr}(\lambda_j) - g_{0,r}) \right| > \eta_2) \\ &\quad + \mathbb{P}(\left| \frac{1}{m} \sum_{j=1}^m (l_j - 1) (\lambda_j^{2d_{0,r}} I_{X,rr}(\lambda_j) - g_{0,r}) \right| \mathbb{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} > \eta_3) \end{aligned} \quad (3.6)$$

with  $\Theta_1$  as in (3.22); see Remark 3.4. Furthermore, with  $\ell = \exp(\frac{1}{m} \sum_{j=1}^m \log(j))$  and  $\Delta$  as in Remark 3.4,

$$l_j = \begin{cases} \binom{j}{\ell}^{2(-\frac{1}{2} + \Delta)}, & 1 \leq j \leq \ell, \\ \binom{j}{\ell}^{2(\Delta_1 - d_{0,r})}, & \ell < j \leq m. \end{cases} \quad (3.7)$$

For the sake of simplicity,  $\eta_1, \eta_2, \eta_3 > 0$  in (3.6) and the arguments for (3.6) are not further specified; we refer to the proof of Proposition 3.5 for more details. The inequalities (3.5) and (3.6) reveal that it is enough to prove a uniform concentration inequality for an object of the form

$$\sup_{D \in \Omega} |\widehat{H}_{rs}(D) - \widetilde{H}_{rs}(D)| \quad (3.8)$$

with  $\widehat{H}_{rs}(D), \widetilde{H}_{rs}(D)$  denoting the  $(r, s)$ th elements of

$$\widehat{H}(D) = \frac{1}{m} \sum_{j=1}^m t_j(D) I_X(\lambda_j) t_j(D) \quad \text{and} \quad \widetilde{H}(D) = \frac{1}{m} \sum_{j=1}^m t_j(D) \lambda_j^{-D_0} G_0 \lambda_j^{-D_0} t_j(D), \quad (3.9)$$

where  $t_j(D) = \text{diag}(t_{j,1}(d_1), \dots, t_{j,p}(d_p)) \in \mathcal{M}_{\text{diag}}$  consist of suitable functions of  $d$ 's. The set  $\Omega$  is of the form

$$\Omega = \{D \in \mathcal{M}_{\text{diag}} \mid A \preceq D \preceq B\} \quad (3.10)$$

with

$$A = \text{diag}(a_1, \dots, a_p), \quad B = \text{diag}(b_1, \dots, b_p). \quad (3.11)$$

The functions  $t_{j,r} : [\Delta_1, \Delta_2] \rightarrow [0, \infty)$  are assumed to be differentiable on  $(\Delta_1, \Delta_2)$  with bounded derivatives.

From here on, (3.8) can be separated into a probabilistic and a deterministic part as

$$\sup_{D \in \Omega} |\widehat{H}_{rs}(D) - \widetilde{H}_{rs}(D)| \leq \sup_{D \in \Omega} |\widehat{H}_{rs}(D) - \mathbb{E}\widehat{H}_{rs}(D)| + \sup_{D \in \Omega} |\mathbb{E}\widehat{H}_{rs}(D) - \widetilde{H}_{rs}(D)|. \quad (3.12)$$

We treat both terms separately. For the first summand, we need a high probability upper bound stated in Lemma B.1. On the other hand, the second term in (3.12) is deterministic and an upper bound is given in Lemma B.2. Lemmas B.1 and B.2 are stated in Appendix B. Both are crucial to infer upper bounds on the probabilities in (3.5) and (3.6). Those results are stated in Propositions 3.1–3.4 below and its proofs can be found in Appendix B.

*Technical contributions:* The following points highlight our main technical contributions and give some orientation of how the different appendices contribute.

The statement for a probabilistic bound on the first summand of (3.12) is stated in Lemma B.1 in Appendix B. The key tool to handle the first summand of (3.12) is a uniform concentration bound of Dicker and Erdogdu (2017), slightly reformulated to serve better our needs in Appendix D.1. Our arguments above show how the incorporation of the event  $\{\|\widehat{D} - D_0\|_{\max} \leq \varepsilon\}$  allows one to get to the setting where that bound could potentially be applicable. Making the bound workable for the probabilities in (3.5) and (3.6) was another major challenge and can be found in Appendix C.1. The difficulties involved allowing for general dependence structure (short- and long-range dependence, and antipersistence) and, more importantly, developing non-asymptotic theory in terms of any dimension  $p$  and sample size  $N$ . Though there are certainly many works on local Whittle estimation (and we adapt some of their techniques), non-asymptotic results are not available even for the one-dimensional case  $p = 1$ . Key to our non-asymptotic developments are bounds of independent interest on various autocovariance matrices under general dependence assumptions in Appendix C.3. Those results are derived by replacing the matrices by integral operators. We believe that those inequalities are crucial in order to derive non-asymptotic theory in any context involving high-dimensional long-range dependence.

To deal with the bounds on the second summand of (3.12) and deterministic part of the quantities in (3.5) and (3.6), we state Lemma B.2. The proof involves the derivation of a series of non-asymptotic bounds for the periodogram in Appendix C.2.

## 3.2 Statements

In this section we formally state our main results. In order to keep the statements as simple as possible, we moved the expressions of some quantities to Appendix A.

We further introduce the following quantity which will allow us to express our bounds in a simplified way and emphasize the necessary distinction between different ranges of the memory parameters as will become clearer in the proofs

$$\bar{\Delta}_N = \max_{r=1, \dots, p} (\mathbb{1}_{\{d_{0,r} \leq 0\}} + \log(N) \mathbb{1}_{\{d_{0,r} > 0\}}). \quad (3.13)$$

**Proposition 3.1.** *Let  $\{X_n\}_{n \in \mathbb{Z}}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–5 are satisfied. Then, for any  $\varepsilon \in (0, \min\{\Delta_2, -\Delta_1, q\})$  with  $q$  as in Assumption 2, there are positive constants  $c_1, c_2$  such that for any  $\mathbf{C} \geq 1$ ,*

$$\mathbb{P}\left(\sup_{D \in \Omega(\varepsilon)} |\widehat{G}_{rs}(D) - \widetilde{G}_{rs}(D)| > \nu\right) \leq c_1 p^{-c_2 \mathbf{C}}$$

for

$$\nu = \mathbf{C} \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_1}} + \mathcal{T}_1(\varepsilon), \quad (3.14)$$

where  $\mathcal{R}_1 = \min\{(\log(N)\bar{\Delta}_N)^{-1}\mathcal{R}_{11}, (\log(N)\bar{\Delta}_N)^{-2}\mathcal{R}_{12}\}$  with  $\bar{\Delta}_N$  as in (3.13) and  $\mathcal{R}_{1i}$ ,  $i = 1, 2$  are characterized in Table A.1. A representation of  $\mathcal{T}_1(\varepsilon)$  can be found in Table A.2.

The parameter  $\varepsilon$  controls the deviation of the estimated memory parameters  $\widehat{D}$  around  $D_0$ , and will be chosen appropriately in Proposition 3.5 below. To ensure that the deviation of  $\widehat{G}_{rs}(D)$  around  $\widetilde{G}_{rs}(D)$  can be controlled, the quantities which characterize  $\nu$  in (3.14) have to satisfy  $\mathcal{R}_1 \gtrsim \|G\|^2 \log(p)$  and  $mN^{-2\varepsilon} \gtrsim (1 + \frac{1}{2\varepsilon})Q_m$  with  $Q_m$  defined in (A.3) in Appendix A. The quantities in  $\nu$  can be expected to satisfy those assumptions since our bounds are sharp enough to get  $\mathcal{R}_1 \rightarrow \infty$  and  $\mathcal{T}_1(\varepsilon) \rightarrow 0$  as  $N \rightarrow \infty$ . Those asymptotics are crucial in order to achieve consistency which entails a negligible bias. This observation can be made not only for Proposition 3.1 but as well in the similarly structured Propositions 3.2–3.4 below.

**Remark 3.1.** Proposition 3.1 and subsequent results provide non-asymptotic bounds when estimating quantities of interest. In the considered setting,  $\{X_n\}_{n \in \mathbb{Z}}$  is a stationary series with a spectral density  $f_X(\lambda)$  and observed for  $n = 1, \dots, N$ , and of fixed dimension  $p$ . But note that our non-asymptotic bounds are expressed in terms of  $p, N, \|G\|$  and other quantities. When  $p$  changes, the dependence structure of  $\{X_n\}_{n \in \mathbb{Z}}$  also changes and our bounds adjust through changing  $\|G\|$  and those other quantities (e.g.  $c_{G,1}, c_{G,2}$  in Assumptions 2 and 3). The same with changing  $N$ . We note that because of the term  $\log(p)/\mathcal{R}_1$ , in (3.14), the obtained bounds suggest consistent estimation in a typical high-dimensional regime where  $p$  is much larger than  $N$ , but  $\log(p)$  is much smaller than  $N$  (or the power of  $N$ ). We also note that because of the constants  $c_{G,1}, c_{G,2}$  in Proposition 3.1 and similar subsequent results are absolute in the sense that they do not depend on  $p, N$  and the underlying stationary series; the dependence on the latter is captured through the other quantities in the bounds.

The following three propositions give upper bounds on the probabilities in (3.6). Those are the probabilities required to control the estimates for the memory parameters  $D_0$ .

**Proposition 3.2.** *Let  $\{X_n\}_{n \in \mathbb{Z}}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–5 are satisfied. Then, there are positive constants  $c_1, c_2$  such that for any  $C \geq 1$ ,*

$$\mathbb{P}\left(|\widehat{g}_r(d_{0,r}) - g_{0,r}| > \nu_1\right) \leq c_1 p^{-c_2 C}$$

for

$$\nu_1 = C \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_2}} + \mathcal{T}_2, \quad (3.15)$$

where  $\mathcal{R}_2 = \min\{\bar{\Delta}_N^{-1} \mathcal{R}_{21}, \bar{\Delta}_N^{-2} \mathcal{R}_{22}\}$  with  $\bar{\Delta}_N$  as in (3.13) and  $\mathcal{R}_{2i}$ ,  $i = 1, 2$  are characterized in Table A.1. A representation of  $\mathcal{T}_2$  can be found in Table A.2.

In order to ensure meaningful estimation, the quantities which characterize  $\nu_1$  in (3.15) have to satisfy  $\mathcal{R}_2 \gtrsim \|G\|^2 \log(p)$  and  $m \gtrsim \log(m)Q_m$  with  $Q_m$  defined in (A.3) in Appendix A.

For the following proposition, we use the set  $\Theta_1$  in (3.22), which is characterized by some  $\Delta > 0$ .

**Proposition 3.3.** *Let  $\{X_n\}_{n \in \mathbb{Z}}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–5 are satisfied. Then, there are positive constants  $c_1, c_2$  such that for any  $C \geq 1$ ,*

$$\mathbb{P}\left(\sup_{d \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_{0,r}} \lambda_j^{2d_{0,r}} (I_{X,rr}(\lambda_j) - \lambda_j^{-2d_{0,r}} g_{0,r}) \right| > \nu_2\right) \leq c_1 p^{-c_2 C}$$

for

$$\nu_2 = C \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_3}} + \mathcal{T}_3, \quad (3.16)$$

where  $\mathcal{R}_3 = \min\{(\log(m)\bar{\Delta}_N)^{-1}\mathcal{R}_{31}, (\log(m)\bar{\Delta}_N)^{-2}\mathcal{R}_{32}\}$  with  $\bar{\Delta}_N$  as in (3.13) and  $\mathcal{R}_{3i}$ ,  $i = 1, 2$  are characterized in Table A.1. A representation of  $\mathcal{T}_3$  can be found in Table A.2.

In order to ensure meaningful estimation, the quantities which characterize  $\nu_2$  in (3.16) have to satisfy  $\mathcal{R}_3 \gtrsim \|G\|^2 \log(p)$  and  $m^{2\bar{\Delta}_r} \gtrsim Q_m$  with  $Q_m$  defined in (A.3) in Appendix A.

The next proposition gives a bound on the third probability in (3.6). Recall the definitions of  $\ell$  and  $l_j$  given in (3.7).

**Proposition 3.4.** *Let  $\{X_n\}_{n \in \mathbb{Z}}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–5 are satisfied. Then, there are positive constants  $c_1, c_2$  such that for any  $C \geq 1$ ,*

$$\mathbb{P}\left(\left|\frac{1}{m} \sum_{j=1}^m (l_j - 1) \lambda_j^{2d_{0,r}} (I_{X,rr}(\lambda_j) - \lambda_j^{-2d_{0,r}} g_{0,r})\right| \mathbb{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} > \nu_3\right) \leq c_1 p^{-c_2 C}$$

for

$$\nu_3 = C \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_4}} + \mathcal{T}_4, \quad (3.17)$$

where  $\mathcal{R}_4 = \min\{\bar{\Delta}_N^{-1}\mathcal{R}_{41}, \bar{\Delta}_N^{-2}\mathcal{R}_{42}\}$  with  $\bar{\Delta}_N$  as in (3.13) and  $\mathcal{R}_{4i}$ ,  $i = 1, 2$  are characterized in Table A.1. A representation of  $\mathcal{T}_4$  can be found in Table A.2.

In order to ensure meaningful estimation, the quantities which characterize  $\nu_3$  in (3.17) have to satisfy  $\mathcal{R}_4 \gtrsim \|G\|^2 \log(p)$ ,  $m \gtrsim \ell^{1-2\Delta} Q_m$  and  $m \gtrsim \log(m) Q_m$  with  $Q_m$  defined in (A.3) in Appendix A.

Propositions 3.1–3.4 combined together enable us to obtain a consistency result for (3.1), which is stated in the following proposition. Recall the definition of  $\underline{k}$  given in (2.6) and of the function  $L$  in (3.3).

**Proposition 3.5.** *Suppose that the assumptions in Propositions 3.1–3.4 hold. Then, there are positive constants  $c_1, c_2$  such that for any  $C \geq 1$ ,*

$$\mathbb{P}(\|\widehat{G}(\widehat{D}) - G_0\|_{\max} > \delta) \leq c_1 p^{2-c_2 C} \quad (3.18)$$

for

$$\delta = \max\{2\nu, \varepsilon 4 \|G_0\| \log(N) \lambda_m^{-2\varepsilon} L(-\varepsilon)\}, \quad \varepsilon = \max_{i=1,2,3} \eta_i, \quad (3.19)$$

with  $\nu$  as in (3.14),

$$\eta_1^2 = 8(\mathcal{V}_1(m)\underline{k})^{-1} \nu_1, \quad \eta_2^2 = 8(\mathcal{V}_1(m)L(\Delta_2 - \Delta_1)\underline{k})^{-1} \nu_2, \quad \eta_3^2 = (\mathcal{V}_2(m)\underline{k})^{-1} \nu_3, \quad (3.20)$$

where  $\nu_i$ ,  $i = 1, 2, 3$  are as in (3.15), (3.16) and (3.17) and

$$\mathcal{V}_1(m) = \frac{1}{3} \frac{1}{m^4} \sum_{i,j=1}^m (i-j)^2, \quad \mathcal{V}_2(m) = \frac{1}{m} \sum_{j=1}^m (l_j - 1), \quad (3.21)$$

and it is assumed that  $\varepsilon \in (0, \frac{1}{2})$ .

In order to ensure meaningful estimation, the term  $\mathcal{V}_2(m)$  needs to be positive. This is proven in Lemma C.7.

**Remark 3.2.** In Corollary A.1, we formally state an analogous result to Proposition 3.5 under the assumption that the underlying process is either short- or long-range dependent. Note that in contrast to Proposition 3.5, Corollary A.1 does not require Assumption 5. In particular, one can infer an asymptotic result without requiring any further assumptions on the relation between  $m$  and  $N$  besides  $\frac{1}{m} + \frac{m}{N} \rightarrow \infty$  for  $N \rightarrow \infty$  which coincides with Assumption 4 in Robinson (1995b). A numeric illustration of our non-asymptotic results is given in Appendix F.2 in terms of Corollary A.1.

The following propositions give non-asymptotic consistency results for the graphical and thresholding local Whittle estimators, respectively.

**Proposition 3.6.** *Let  $\{X_n\}_{n \in \mathbb{Z}}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–5 are satisfied. Then, there are positive constants  $c_1, c_2$ , such that choosing a threshold  $\rho = \delta$  as in (3.19) yields, for any  $\mathbf{C} \geq 1$ ,*

$$\begin{aligned} \mathbb{P}(\|T_\rho(\widehat{G}(\widehat{D})) - G_0\|_F^2 > 13p\|G_0\|_a^a \rho^{2-a}) &\leq c_1 p^{2-c_2 \mathbf{C}}, \\ \mathbb{P}(\|T_\rho(\widehat{G}(\widehat{D})) - G_0\| > 7\|G_0\|_a^a \rho^{1-a}) &\leq c_1 p^{2-c_2 \mathbf{C}} \end{aligned}$$

for any  $a \in [0, 1)$  and  $T_\rho$  as in (2.2).

**Proposition 3.7.** *Let  $\{X_n\}_{n \in \mathbb{Z}}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–5 are satisfied. Then, there are positive constants  $c_1, c_2$ , such that choosing a penalty parameter  $\rho = \delta$  as in (3.19) yields, for any  $\mathbf{C} \geq 1$ ,*

$$\mathbb{P}(\|\widehat{P}_\rho(\widehat{D}) - P_0\|_F^2 > \frac{16^2}{k^4}(p+s)\rho^2) \leq c_1 p^{2-c_2 \mathbf{C}}$$

with  $s$  as in (2.10) and  $\frac{16}{k^2}\sqrt{p+s}\rho \leq \|P_0\|$ .

Propositions 3.6 and 3.7 can be expressed in terms of the quantities in Corollary A.1 when allowing only for long- and short-range dependence. That is,  $\delta$  in Propositions 3.6 and 3.7 can be chosen as in Corollary A.1 defined in terms of (A.4).

In contrast to Proposition 3.6, Proposition 3.7 states only a result for the Frobenius norm. A result on the operator norm can be inferred based on the inequality  $\|\cdot\| \leq \|\cdot\|_F$ . However, the operator norm provides the same convergence rate as the Frobenius norm. In contrast, for thresholding estimators, the convergence rates differ by one  $p$ ; see Proposition 3.6. The literature on covariance estimation has addressed this problem by considering alternative estimators. Section 3.3 below presents a modified graphical local Whittle estimator based on estimating the coherence matrix and a constrained  $l_1$ -minimization for inverse matrix estimation (CLIME) version.

**Remark 3.3.** The probability bounds in Propositions 3.1–3.7 are given in terms of  $p$  which might suggest consistency only in the limit of  $p \rightarrow \infty$ , as long as  $\mathbf{C}$  is large enough. Note, however, that  $\mathbf{C} = \mathbf{C}_m$  may depend on  $m$  and  $N$ , and enters in our choices of  $\nu$  in (3.14) and  $\nu_i$ ,  $i = 1, 2, 3$ , in (3.15), (3.16) and (3.17). Under suitable assumptions, one can in fact have

$$\mathbf{C}_m \sqrt{\frac{\log(p)}{\mathcal{R}_i}} \rightarrow 0 \quad \text{and} \quad \mathbf{C}_m \rightarrow \infty \quad \text{as} \quad m \rightarrow \infty \quad \text{for} \quad i = 1, \dots, 4,$$

in Propositions 3.1–3.4, so that the resulting probability bounds are small even for fixed low-dimensional  $p$ . In the latter case, even when  $p = 1$ , these results also provide new non-asymptotic exponential bounds, for example, on the probability of  $\{\|\widehat{G}(\widehat{D}) - G_0\|_{\max} > \delta\}$ .

**Remark 3.4.** The work by Robinson (1995b) concerns consistency results for the univariate local Whittle estimators for  $G_0$  and  $D_0$ . The consistency of  $\widehat{D}$  is stated in Theorem 1 in Robinson (1995b). The proof in the asymptotic regime mentions that the function  $R(D)$  behaves nonuniformly around  $D = d_0 - \frac{1}{2}$ . For this reason, one has to consider the cases  $d_0 - \frac{1}{2} < \Delta_1$  and  $d_0 - \frac{1}{2} \geq \Delta_1$  to separate the set  $\Theta$  as  $\Theta = \Theta_1 \cup \Theta_2$  with

$$\Theta_1 = \begin{cases} \{d \mid d_0 - \frac{1}{2} + \Delta \leq d \leq \Delta_2\}, & \text{if } d_0 \geq \Delta_1 + \frac{1}{2}, \\ \{d \mid \Delta_1 \leq d \leq \Delta_2\}, & \text{if } d_0 < \Delta_1 + \frac{1}{2}, \end{cases} \quad (3.22)$$

$$\Theta_2 = \begin{cases} \{d \mid \Delta_1 \leq d < d_0 - \frac{1}{2} + \Delta\}, & \text{if } d_0 \geq \Delta_1 + \frac{1}{2}, \\ \emptyset, & \text{if } d_0 < \Delta_1 + \frac{1}{2}, \end{cases} \quad (3.23)$$

where  $\Delta \in (0, \Delta_2)$ . As displayed in (3.22), separating  $\Theta$  is only necessary if  $d_0 < \Delta_1 + \frac{1}{2}$ . The case  $d_0 < \Delta_1 + \frac{1}{2}$  includes  $d_0 \leq 0$  since  $-\frac{1}{2} < \Delta_1$ . Note that  $d_0 \leq 0$  coincides with the prior knowledge that the observed time series is not long-range dependent. The sets  $\Theta_1$  and  $\Theta_2$  are used to determine the range of admissible estimates  $d$  for an individual component series  $X_{r,n}$ . Then,  $\Theta_1$  and  $\Theta_2$  depend on  $d_{0,r}$  instead of  $d_0$ . Throughout the paper we do not reflect the dependence on  $r$  in the notation of  $\Theta_1$  and  $\Theta_2$ .

### 3.3 Alternative estimators for precision matrix

We study here a modified graphical local Whittle estimator based on estimating the coherence matrix and a CLIME version; see Sections 3.3.1 and 3.3.2. The section and in particular the corresponding proofs in Appendix B also emphasize the value of our Proposition 3.5 since it can be used to infer consistency results even for modified versions of penalized local Whittle estimators.

#### 3.3.1 Modified graphical local Whittle

As known for covariance matrices, the rate of convergence can be improved for the operator norm by considering the correlation matrix instead; see Rothman et al. (2008) and Shu and Nan (2019) for temporally correlated data. Analogously, we can consider the coherence matrix. Let  $G_0 = W_0 \Gamma_0 W_0$ , where  $W_0 = \text{diag}(G_{0,11}^{1/2}, \dots, G_{0,pp}^{1/2})$  and  $\Gamma_0$  is the true coherence matrix. Then, the precision matrix satisfies  $P_0 = W_0^{-1} \Gamma_0^{-1} W_0^{-1}$  and therefore  $K_0 := \Gamma_0^{-1} = W_0 P_0 W_0$ . We write  $\widehat{\Gamma} = \widehat{\Gamma}(\widehat{D})$  and  $\widehat{W} = \widehat{W}(\widehat{D})$  for their sample counterparts, and indicate their dependence on the matrix  $D$ . The matrix  $K_0$  can then be estimated as

$$\widehat{K}_\rho = \underset{K \succ 0}{\text{argmin}} \ell_\rho^\Gamma(\widehat{D}, K), \quad (3.24)$$

where  $\widehat{D}$  is estimated univariately by (2.1) and

$$\ell_\rho^\Gamma(D, K) = -\frac{1}{m} \sum_{j=1}^m \log |\lambda_j^D K \lambda_j^D| + \text{tr}(\widehat{\Gamma}(D)K) + \rho \|K\|_{1,off}. \quad (3.25)$$

Then, we can define a modified coherence-based graphical local Whittle estimator

$$\widehat{P}_\rho^M = \widehat{W}^{-1} \widehat{K}_\rho \widehat{W}^{-1}, \quad (3.26)$$

where  $\widehat{W}(\widehat{D}) = \text{diag}(\widehat{G}_{11}^{1/2}(\widehat{D}), \dots, \widehat{G}_{pp}^{1/2}(\widehat{D}))$ . The following statement gives a non-asymptotic consistency result on the spectral norm.

**Proposition 3.8.** *Let  $\{X_n\}_{n \in \mathbb{Z}}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–5 are satisfied. Then, there are positive constants  $c_1, c_2$ , such that choosing a penalty parameter  $\rho = \delta$  as in (3.19) yields, for any  $\mathbf{C} \geq 1$ ,*

$$\mathbb{P}(\|\widehat{P}_\rho^M(\widehat{D}) - P_0\| > 30 \times 48 \max\{1, 1/\underline{k}^9\} \max\{1, \|K_0\|\} \sqrt{s} \rho) \leq c_1 p^{2-c_2 \mathbf{C}}$$

with  $s$  as in (2.10) and  $48 \max\{1, 1/\underline{k}^4\} \sqrt{s} \rho \leq \|K_0\|$ .

In contrast to Proposition 3.7, Proposition 3.8 provides a convergence rate for the spectral norm which shows that the modified graphical local Whittle estimator (3.26) can achieve the same convergence rate as the thresholding local Whittle estimator.

A handy result to prove Proposition 3.8 and of independent interest is the following lemma which gives a consistency result for the coherence matrix estimator in (3.24).

**Lemma 3.1.** *Let  $\{X_n\}_{n \in \mathbb{Z}}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–5 are satisfied. Then, there are positive constants  $c_1, c_2$ , such that choosing a penalty parameter  $\rho = \delta$  as in (3.19) with  $N, p$  such that  $\delta < 1$  yields, for any  $\mathbf{C} \geq 1$ ,*

$$\mathbb{P}(\|\widehat{K}_\rho(\widehat{D}) - K_0\|_F^2 > (48 \max\{1, 1/\underline{k}^4\})^2 s \rho^2) \leq c_1 p^{2-c_2 \mathbf{C}}$$

with  $s$  as in (2.10) and  $48 \max\{1, 1/\underline{k}^4\} \sqrt{s} \rho \leq \|K_0\|$ .

Note that we assume  $N, p$  such that  $\delta < 1$  only to achieve a simplified representation of the result. In general, it is possible to state the result for any fixed  $N, p$ .

### 3.3.2 CLIME estimation

CLIME estimation for i.i.d. samples was introduced in Cai et al. (2011) and further studied in Shu and Nan (2019) allowing for temporal correlation. We adopt their approach to the spectral domain and set  $\widehat{\Theta} = (\widehat{\theta}_{rs})_{r,s=1,\dots,p}$  to be the solution of the minimization problem

$$\min \|\Theta\|_1 \quad \text{subject to} \quad \|\widehat{G}(\widehat{D})\Theta - I_p\|_{\max} \leq \rho,$$

where  $\rho$  is a tuning parameter. Then, the CLIME estimator is defined as

$$\widehat{P}_\rho^C = (\widehat{\theta}_{rs}^C)_{r,s=1,\dots,p} \quad \text{with} \quad \widehat{\theta}_{rs}^C = \widehat{\theta}_{sr}^C = \widehat{\theta}_{rs} \mathbb{1}_{\{|\widehat{\theta}_{rs}| \leq |\widehat{\theta}_{sr}|\}} + \widehat{\theta}_{sr} \mathbb{1}_{\{|\widehat{\theta}_{rs}| > |\widehat{\theta}_{sr}|\}}. \quad (3.27)$$

We impose the sparsity assumption (2.9) on  $P_0$ .

**Proposition 3.9.** *Let  $\{X_n\}_{n=1,\dots,N}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–5 are satisfied. Then, there are positive constants  $c_1, c_2$ , such that choosing a penalty parameter  $\rho = \delta$  as in (3.19) yields, for any  $\mathbf{C} \geq 1$ ,*

$$\mathbb{P}(\|\widehat{P}_\rho^C(\widehat{D}) - P_0\|^2 > 24 \|P_0\|_1^{1-a} \|P_0\|_a^a \rho^{1-a}) \leq c_1 p^{2-c_2 \mathbf{C}}$$

for any  $a \in [0, 1)$ .



### 3.4 Graphical model selection consistency

We give here results on consistent recovery of the sparsity pattern and sign consistency for our estimators. For the long-run variance matrix estimation, we focus on the thresholding local Whittle estimator. For the precision matrix, we consider a thresholded CLIME estimator and conclude with a discussion on consistent graph recovery for other estimators. The proofs of the statements in this section can be found in Appendix B.

The following proposition gives a non-asymptotic result for consistent graph recovery of the thresholding local Whittle. Rothman et al. (2009) consider covariance matrix estimation for i.i.d.  $p$ -dimensional random vectors in a high-dimensional regime. In particular, their Theorem 2 states that the thresholding operator consistently recovers the sparsity pattern. The proof of Theorem 2 in Rothman et al. (2009) is generic and proves, combined with our Proposition 3.5, consistent recovery of the sign and sparsity pattern.

**Proposition 3.10.** *Let  $\{X_n\}_{n=1,\dots,N}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–5 are satisfied. Then, there are positive constants  $c_1, c_2$ , such that choosing a threshold  $\rho = \delta$  as in (3.19) yields, for any  $\mathbf{C} \geq 1$ ,*

$$\mathbb{P}\left(T_\rho(\widehat{G}_{rs}(\widehat{D})) = 0 \text{ for all } r, s \text{ such that } G_{0,rs} = 0\right) \geq 1 - c_1 p^{2-c_2 \mathbf{C}}. \quad (3.28)$$

*If we additionally assume that all non-zero elements of  $G_0$  satisfy  $|G_{0,rs}| > \tau$ , where  $\tau$  is of the same order as  $\rho$ , we have*

$$\mathbb{P}\left(\text{sign}(T_\rho(\widehat{G}_{rs}(\widehat{D}))G_{0,rs}) = 1 \text{ for all } r, s \text{ such that } G_{0,rs} \neq 0\right) \geq 1 - c_1 p^{2-c_2 \mathbf{C}}. \quad (3.29)$$

The CLIME estimator in (3.27) can be modified to recover the support of the precision matrix. More precisely, we conduct an additional thresholding step by applying (2.2) to  $\widehat{P}_\rho^{\mathbf{C}}(\widehat{D})$  in (3.27). This procedure follows Section 4 in Cai et al. (2011) who considered inverse covariance estimation for i.i.d. data. Subsequently, Shu and Nan (2019) used a thresholded CLIME for inverse covariance estimation under temporal dependence; see Theorem 5 in Shu and Nan (2019) for their result on consistent sparsity and sign recovery. Following their arguments, we can state the following proposition.

**Proposition 3.11.** *Let  $\{X_n\}_{n \in \mathbb{Z}}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–5 are satisfied. Then, there are positive constants  $c_1, c_2$ , such that choosing a penalty parameter  $\rho = \delta$  as in (3.19) yields, for any  $\mathbf{C} \geq 1$ ,*

$$\mathbb{P}\left(T_\tau(\widehat{P}_\rho^{\mathbf{C}}(\widehat{D})) = 0 \text{ for all } r, s \text{ such that } P_{0,rs} = 0\right) \geq 1 - c_1 p^{2-c_2 \mathbf{C}}. \quad (3.30)$$

*If we additionally assume that all non-zero elements of  $G_0$  satisfy  $|G_{0,rs}| > \tau$ , where  $\tau$  is of the same order as  $\rho$ , we have*

$$\mathbb{P}\left(\text{sign}(T_\tau(\widehat{\theta}_{rs}^{\mathbf{C}})P_{0,rs}) = 1 \text{ for all } r, s \text{ such that } P_{0,rs} \neq 0\right) \geq 1 - c_1 p^{2-c_2 \mathbf{C}}. \quad (3.31)$$

For the graphical local Whittle estimation, consistent recovery of the sparsity pattern is not expected. This is discussed in Zou (2006) for the classical LASSO in a regression context. There have been several approaches to modify the classical LASSO to achieve consistent graph recovery. For instance the consideration of an adaptive version (Zou, 2006) and a thresholded LASSO (Zhou, 2010; Ravikumar, Wainwright, Raskutti, and Yu, 2011; Wang and Allen, 2021).

An adaptive version is based on a weighted penalty, where the weights are data driven using a preliminary estimator. Though our results are expected to be helpful to prove consistency for an adaptive version, a detailed investigation goes beyond the scope of this work. A related discussion on the difficulties of proving consistent recovery of the sparsity pattern with help of an extended Bayesian information criterion can be found in Remark 4.1.

A thresholding graphical local Whittle estimator is expected to consistently recover the sparsity pattern. It is similar in flavor to the thresholded CLIME and a consistency result can be inferred with help of Proposition 3.5.

### 3.5 Comparison to existing results

In this section, we compare our results to related work. Existing results are either for short-range dependent time series or, if they allow for stronger temporal and spatial dependence, the dependence measure is characterized by an unknown and unestimated quantity. In particular, results for data with stronger dependence structure have only been derived in the time domain.

*Sun et al. (2018)*: We recover recently proven results on spectral density estimation at frequency zero for short-range dependent time series. Sun et al. (2018) supposed that  $D_0 \equiv 0$  and could prove results in a  $\log(p)/m \rightarrow 0$  regime. We get the same result by setting  $\Delta_1 = \Delta_2 = \varepsilon = 0$  in (3.19). Strictly speaking, we do not allow for  $\varepsilon = 0$  since  $\mathcal{T}_1(\varepsilon)$  in Table A.2 involves  $(1 + \frac{1}{2\varepsilon})$ . However, a look into the proof of Proposition 3.1 reveals that for  $\varepsilon = 0$ , one can bound the respective term by  $\log(m)$  instead. We refrained from incorporating the case  $\varepsilon = 0$  explicitly for simplicity. We note also that our result includes an additional  $\log(N)$ . However, this is an artifact of using a slightly simplified notation to make reading easier. More precisely, to prove Proposition 3.5, we apply a uniform concentration inequality in Lemma B.1 which involves the supremum of a partial derivative; see (B.2). The supremum is taken over a closed set. In particular, the set is not empty. For this reason, even when the set contains only one point as in the short-range dependent case ( $D_0 \equiv 0$ ), the derivative is included in the respective bounds. The logarithm appears because of the derivative in our uniform concentration inequality. This can be easily avoided by considering the supremum over a half-open interval. However, it would require careful distinction through all our proofs between whether the set is empty or not. In order to avoid over complicated notation, we refrained from incorporating this case.

*Shu and Nan (2019)*: This related work focusses on the estimation of the covariance matrix and its inverse, with the results in the time domain. However, Shu and Nan (2019) also allow for long-range dependence. They assume that the correlation  $\rho_{kn}^i = \frac{\sigma_{kn}^i}{\sigma_{kk}^i \sigma_{nn}^i}$  with  $\sigma_{kn}^i = \mathbb{E}X_{i,k}X_{i,n}$  in the component series  $X_{i,n}$  satisfies

$$\max_{i=1,\dots,p} |\rho_{kn}^i| \leq C|k-n|^{2\alpha-1} \quad \text{for } k \neq n. \quad (3.32)$$

For  $\alpha \in (0, \frac{1}{2})$  the individual time series can thus be long-range dependent in the sense that the correlation sequences are not absolutely summable. To make a fair comparison, we will assume a known memory parameters  $D_0$  and, for simplicity, further ignore the bounds on the deterministic part and only consider the case when the time series is short- or long-range dependent. Then, based on Proposition 3.1 for  $\varepsilon = 0$  and  $\Delta_1 = 0$ , our convergence rate simply reduces to

$$\nu = \mathbf{C} \|\|G\|\| \sqrt{\frac{\log(p)}{\mathcal{R}_1}}, \quad \mathcal{R}_1 = \min\{(\log(N)\bar{\Delta}_N)^{-1}\mathcal{R}_{11}, (\log(N)\bar{\Delta}_N)^{-2}\mathcal{R}_{12}\} \quad (3.33)$$

with  $\bar{\Delta}_N$  as in (3.13) and

$$\mathcal{R}_{11} = m^{1-2\Delta_2}, \quad \mathcal{R}_{12} = \min \left\{ m^{2-4\bar{\Delta}_u}, m^{\frac{3}{2}-2\bar{\Delta}_u} \right\}.$$

In Shu and Nan (2019), the result analogous to our Proposition 3.1 is Lemma A.2., (i) with rates given in Remark 2. Based on Remark 2 in Shu and Nan (2019), the quantity analogous to  $\mathcal{R}_1$  in (3.33) is given by

$$\mathcal{R} = \begin{cases} N^{1-2\alpha}, & \alpha \in (\frac{1}{4}, \frac{1}{2}), \\ \min \left\{ N^{1-2\alpha}, N^{\frac{1}{2}} \right\}, & \alpha \in (0, \frac{1}{4}), \\ \min \left\{ (\log(N))^{-1}N, N^{\frac{1}{2}} \right\}, & \alpha = \frac{1}{4}. \end{cases} \quad (3.34)$$

Since the results are in the time domain, the convergence rates are in terms of the sample size  $N$  rather than the number of frequencies used in estimation  $m$  as in our results and Sun et al. (2018). See also Remark C.1, for a discussion on why our bounds include  $\log(N)$ . Otherwise, our bounds coincide with those in Shu and Nan (2019) for  $\Delta_2 = \alpha$ . In contrast to our statements, the results in Shu and Nan (2019) are not non-asymptotic. Furthermore, they do not estimate  $\alpha$ .

## 4 The choice of the shrinkage parameters and algorithms

Both thresholding (2.2) and graphical local Whittle estimation (2.3) depend respectively on the threshold and penalization parameter  $\rho$ . In this section, we discuss how to select  $\rho$ . This choice plays a critical role in finite sample performance. We propose to use cross-validation for the thresholding parameter and an extended Bayesian information criterion (eBIC) for graphical local Whittle estimation.

### 4.1 Thresholding local Whittle estimation

Cross-validation is generally suggested to select the penalization parameter for thresholding covariance matrix estimation; see Bickel and Levina (2008b). Sun et al. (2018) modified it to sampling over Fourier frequencies in order to account for temporal dependence in spectral density matrix estimation. Similarly, we propose to select the optimal thresholding parameter  $\rho$  by cross-validating over the periodogram (1.4). More precisely, we split the sequence of the periodogram evaluated at different frequencies into two groups. Then, we apply the thresholding estimator of the long-run variance matrix (2.2) to the first group. The long-run variance matrix estimator from the latter group is used as a reference. The optimal thresholding parameter is selected by minimizing the average Frobenius norm (squared) between the thresholding estimators and the reference estimators of the long-run variance matrix. The detailed procedure can be found in Algorithm 1. As in Sun et al. (2018), this procedure remains to be justified in theory even for short-range dependent series.

### 4.2 Graphical local Whittle estimation

For the penalization parameter in graphical local Whittle estimation, we suggest to use an extended Bayesian information criterion (eBIC). Tuning parameter selection for penalized likelihood estimation by an eBIC has been studied by multiple authors allowing the dimension to grow with the sample size. The eBIC was proposed by Foygel and Drton (2010) for Gaussian graphical models and further used in Foygel and Drton (2011) for model selection in sparse generalized linear models. Later, Gao, Pu, Wu, and Xu (2012) proved that using the eBIC to select the tuning parameter in penalized likelihood estimation with the so-called SCAD penalty (Fan and Li (2001)) can lead to

---

**Algorithm 1:** Threshold selection by cross-validation over periodogram
 

---

**Input:**  $I_X(\lambda_j)$ ,  $j = 1, \dots, m$ ,  $\widehat{D}$ . Range  $\mathcal{L}$  of  $\rho$ . Number  $N_1$  of validation sets.

**Output:** Optimal threshold parameter  $\widehat{\rho} := \underset{\rho \in \mathcal{L}}{\operatorname{argmin}} \widehat{R}_\rho$ .

**for**  $\rho \in \mathcal{L}$  **do**

**for**  $k = 1, \dots, N_1$  **do**

        1. Randomly divide  $\{1, \dots, m\}$  into two sets  $J_1$  and  $J_2$  of sizes  $m_1 = \lceil m/2 \rceil$  and  $m_2 = m - \lceil m/2 \rceil$ , respectively.

        2. Calculate  $\widehat{G}_{1,k} = m_1^{-1} \sum_{j \in J_1} \lambda_j^{\widehat{D}} I_X(\lambda_j) \lambda_j^{\widehat{D}}$  and  $\widehat{G}_{2,k} = m_2^{-1} \sum_{j \in J_2} \lambda_j^{\widehat{D}} I_X(\lambda_j) \lambda_j^{\widehat{D}}$ .

**end**

    Obtain  $\widehat{R}_\rho = N_1^{-1} \sum_{k=1}^{N_1} \|T_\rho(\widehat{G}_{1,k}(\widehat{D})) - \widehat{G}_{2,k}(\widehat{D})\|_F^2$ .

**end**

---

consistent graphical model selection. See also Chen and Chen (2012). To be more specific, we use the following criterion

$$\widehat{P}_{\text{eBIC}} = \underset{P \in \mathbb{G}}{\operatorname{argmin}} \text{eBIC}_\gamma(P), \quad \text{eBIC}_\gamma(P) = \operatorname{tr}(\widehat{G}(\widehat{D})P) - \log |P| + \|P\|_0 \frac{1}{N} (\log N + 4\gamma \log p), \quad (4.1)$$

where  $\mathbb{G}$  is the set of all  $\widehat{P}_\rho$  estimated by (2.3) over a range of  $\rho$  and  $\|P\|_0$  denotes the norm counting the number of non-zero elements in  $P$ . Furthermore, the criterion is indexed by a parameter  $\gamma \in [0, 1]$ ; see Foygel and Drton (2010) and Chen and Chen (2008) for the Bayesian interpretation of  $\gamma$ . Then,  $\widehat{P}_{\text{eBIC}}$  gives us a data-driven estimate of  $\widehat{P}_\rho$  and the associated penalty  $\rho$ .

In order to determine  $\widehat{P}_\rho$  over a range of  $\rho$  and the set  $\mathbb{G}$ , we consider an algorithm to compute the graphical local Whittle estimator for a given penalty parameter  $\rho$ . This algorithm is a natural extension of graphical LASSO algorithms for real symmetric covariance matrices to complex-valued Hermitian matrices. The generalization to complex-valued Hermitian matrices is necessary since the true precision matrix  $P_0$  is possibly complex-valued. We propose a complex-valued alternating linearization method (ALM) which is a variation of alternating direction method of multipliers (ADMM) proposed in Scheinberg, Ma, and Goldfarb (2010). Our limited simulation study shows that the proposed method converges faster than the naïve ADMM algorithm. We also note that the complex-valued ADMM for SRD series was considered by Jung, Hannak, and Goertz (2015).

The ALM (Algorithm 2) solves the problem

$$\underset{P, Y}{\operatorname{argmin}} \left\{ -\log |P| + \operatorname{tr}(\widehat{G}P) + \rho \|Y\|_{1, \text{off}} \right\}$$

subject to  $P = Y$  being positive definite. It invokes the ADMM algorithm to find sparse positive definite estimates of  $P_0$  by introducing augmented Lagrangian. It furthermore carefully selects augmented Lagrangian penalty parameter  $\widetilde{\mu}_k$  so that the positive definiteness is achieved throughout the iterations. For the initial estimator  $\widetilde{P}$  in lower dimensions, we take  $(\widehat{G}(\widehat{D}))^{-1}$ . The shrinkage operator in Step 4 of Algorithm 2 is defined as  $\text{shrink}(M, \nu) = \operatorname{sign}(M_{rs}) \max(|M_{rs}| - \nu, 0)$  for a matrix  $M = (M_{rs})_{r,s=1,\dots,p}$  and some  $\nu \geq 0$ . Step 5 of Algorithm 2 requires to update  $\mu_k$ . It is reduced by a constant factor  $\eta_\mu$  on every  $N_\mu$  iteration till a lower bound is achieved by following the idea of Scheinberg et al. (2010). That is, set  $\mu$  as  $\max\{\mu\eta_\mu, \bar{\mu}\}$  to reduce  $\mu$  by a constant factor  $\eta_\mu$  on every  $N_\mu$  iteration. In this paper, we choose  $\mu_0 = .01$ ,  $N_\mu = 10$ ,  $\bar{\mu} = 10^{-3}$  and  $\eta_\mu = 1/4$ . Finally, we terminate the ALM algorithm following the stopping rules in (20) in Scheinberg et al. (2010) except that the first condition is replaced by stopping after 1000 iterations.

---

**Algorithm 2:** Alternating linearization method algorithm for graphical local Whittle
 

---

**Input:**  $Y^0 = \text{diag}(\tilde{P})$  with initial estimator  $\tilde{P}$  of  $P_0$ ,  $\Lambda^0 = 0$ ,  $\mu_0$ ,  $\rho$ .

**Output:** Sparse estimation of  $P_0$ .

**Repeat until convergence:**

**for**  $k = 0, 1, \dots$  **do**

1. Let  $W^{k+1} = Y^k + \mu_k(\Lambda^k - \hat{G})$  and perform the singular value decomposition  $W = U \text{diag}(\eta_1, \dots, \eta_p) V^*$ .
2.  $X^{k+1} = U \text{diag}(\gamma_1, \dots, \gamma_p) V^*$ , where  $\gamma_i = .5(\eta_i + \sqrt{\eta_i^2 + 4\mu_k})$ ,  $i = 1, \dots, p$ .
3.  $Y^{k+1} = \text{shrink}(X^{k+1} - \mu_k(\hat{G} - (X^{k+1})^{-1}), \mu_k \rho)$ .
4.  $\Lambda^{k+1} = \hat{G} - (X^{k+1})^{-1} + (X^{k+1} - Y^{k+1})/\mu_k$ .
5. Pick  $\mu_{k+1} \leq \mu_k$ .

**end**

---

We conclude with a discussion on the criterion (4.1).

**Remark 4.1.** The criterion (4.1) is adapted from Foygel and Drton (2010). The log-likelihood function in equation (2) in Foygel and Drton (2010) which gives an estimate for the inverse covariance matrix of a Gaussian model is replaced by the negative log-likelihood function  $\ell$  in (1.3) in terms of the matrix  $P_0 = G_0^{-1}$ . The criterion seems to perform well in our simulation study. However, from a theoretical perspective, a consistency result can only be established when the penalization term in (4.1) is chosen in dependence of the rate of convergence of the  $(r, s)$ th component of  $\hat{G}(\hat{D})$  around the true  $G_{0,rs}$ . In their theoretical results, Foygel and Drton (2010) considered independent and identically distributed Gaussian random vectors. In this case, the penalization depends on the convergence rate of the deviation of the sample covariance matrix around the true covariance matrix, that is  $\sqrt{\frac{\log(p)}{N}}$ . Foygel and Drton (2010) proved that the eBIC selects the correct model consistently in a high-dimensional regime  $p, N \rightarrow \infty$ ; see Theorem 5 in Foygel and Drton (2010). To establish an analogous consistency result in our setting, we suggest to replace the penalization  $\frac{\log(p)}{N}$  in the eBIC objective function by  $\delta^2$  in (3.19), since  $\delta$  gives the convergence rate of the maximum norm of  $\hat{G}(\hat{D})$  around the true  $G_0$ . The detailed proof of such a consistency result goes beyond the scope of this work and, from a practical perspective, the usage of (4.1) seems to be more natural, since including  $\delta$  in the penalization would involve a number of unknown parameters.

## 5 Simulation study

In this section, we examine the proposed methods through simulations. Our two-stage approach first estimates the memory parameters  $D_0$  and non-sparse long-run variance matrix  $G_0$  based on the local Whittle estimator (1.6). Then, we apply either thresholding or graphical local Whittle estimation to get sparse estimators. Several tuning parameters need to be selected for our methods. We comment first on the number of frequencies used in local Whittle estimation. Details about the selection of tuning parameters for sparse estimation are provided in the subsequent sections.

The selection of the number of frequencies  $m$  is important in practice and should be balanced: be small enough to capture long-range dependence and large enough to get reliable estimates. In univariate local Whittle estimation, asymptotic theory suggests  $m = O(N^{.8})$ ; see Robinson (1995b). There are several papers studying data dependent bandwidth. In this regard, the most influential paper is Henry (2001). Henry (2001) suggests a bandwidth minimizing the mean squared error of

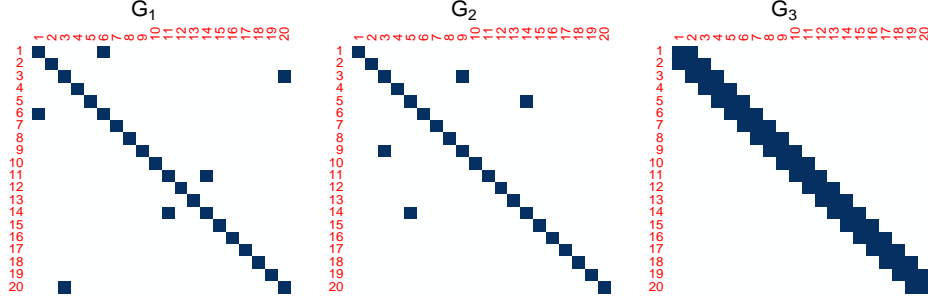


Figure 1: Sparsity patterns of DGPs.

the univariate local Whittle estimator. A visual approach to ensure the balance between capturing long-range dependence and getting reliable estimates is proposed in Baek et al. (2020). Baek et al. (2020) used the so-called local Whittle plots which present estimates of the memory parameters as function of the tuning parameter  $m$  supplemented with confidence intervals.

In our simulation study we base our choice on the asymptotic theory in Robinson (1995b) suggesting  $m = \lfloor N \cdot 8 \rfloor$ , where  $\lfloor x \rfloor$  is the largest integer less than or equal to  $x$ .

### 5.1 Thresholding local Whittle estimation

We consider the following three data generating processes (DGPs) to evaluate the finite sample performance of the thresholding local Whittle method. The one-sided and two-sided VARFIMA(0,  $D$ , 0) (Vector Autoregressive Fractionally Integrated Moving Average) models are used to generate multivariate long-range dependent time series. See Kechagias and Pipiras (2015a,b) for definitions of these models. We consider dimensions  $p = 20, 40, 60$  with sample sizes  $N = 200, 400, 1000$ . Long-range dependent parameters  $D_0$  are selected at random from .1 to .45, if not specified otherwise. We use the notation  $G(r, s)$  to denote the  $(r, s)$ th entry of  $G_0$ . To be more precise, the DGP's are given as follows:

- (thDGP1) One-sided VARFIMA(0,  $D$ , 0) with  $G_0 = (I_{\lfloor p/20 \rfloor} \otimes G_1)$ , where the diagonal entries of  $G_1$  are .159 except  $G_1(1, 1) = G_1(6, 6) = .312$ ,  $G_1(11, 11) = G_1(14, 14) = .212$ ,  $G_1(3, 3) = G_1(20, 20) = .189$ , and  $G_1(1, 6) = -.208 - .064i$ ,  $G_1(3, 20) = -.074 + .015i$ , and  $G_1(11, 14) = -.105 - .013i$ .
- (thDGP2) Two-sided VARFIMA(0,  $D$ , 0) with  $G_0 = (I_{\lfloor p/20 \rfloor} \otimes G_2)$ , where the diagonal entries of  $G_2$  are 1 and  $G_2(3, 9) = .5 + .2i$ ,  $G_2(5, 14) = .4 + .2i$ .
- (thDGP3) Two-sided VARFIMA(0,  $D$ , 0) with banded  $G_0$  matrix given by  $G_0 = (I_{\lfloor p/20 \rfloor} \otimes G_3)$ , where the diagonal entries of  $G_3$  are 1 and  $G_3(r, r + 1) = .4 + .2i$ ,  $r = 1, \dots, 19$ .

For the reader's convenience, the sparsity patterns imposed on  $G_k$ ,  $k = 1, 2, 3$ , are depicted in Figure 1.

The thresholding parameter  $\rho$  is selected based on cross-validation introduced in Section 4.1. We set  $\mathcal{L}$  to be the smallest and the largest value of  $|\hat{G}(\hat{D})|$  in (1.6). We evaluated the performance of the thresholding local Whittle estimator using the mean squared error of the parameters  $\hat{D}$ , total number of misspecified coefficients  $\sum_{r \geq s} (\mathbb{1}_{\{T_\rho(\hat{G}_{rs}(\hat{D}))=0\}} - \mathbb{1}_{\{G_{0,rs}=0\}})^2$ , the Frobenius norm  $\|T_\rho(\hat{G}(\hat{D})) - G_0\|_F$  and the spectral norm  $\|T_\rho(\hat{G}(\hat{D})) - G_0\|$ . The performance measures are calculated based on 1000 iterations.

Figure 2 shows the proportion of times each component is estimated to be non-zero using our thresholding local Whittle approach with  $p = 20$  and  $N = 200$ . The estimation is close to

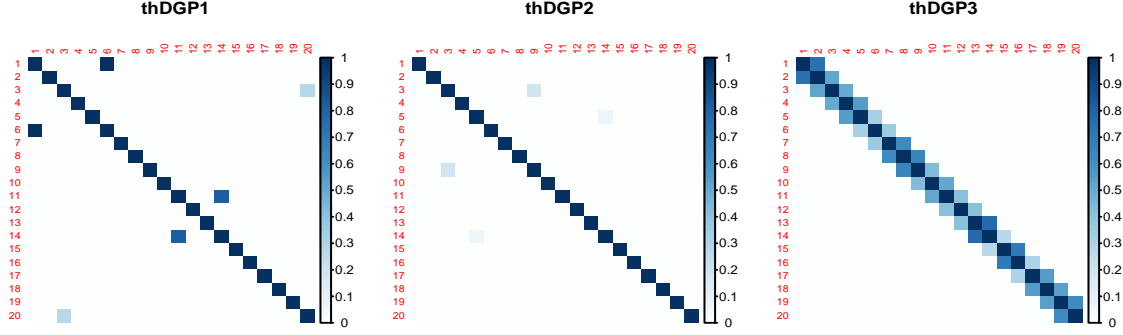


Figure 2: The number of times having non-zero coefficients using thresholding local Whittle estimation with cross-validation tuning parameter selection where  $p = 20$ ,  $N = 200$ .

the true sparsity pattern though some locations are more difficult to be estimated correctly. For example, in thDGP1,  $G_1(3, 20)$  is detected as non-zero less frequently, but this is natural since  $|G_1(3, 20)| = .075$  is smaller than the other coefficients. However, such misspecification vanishes as sample size increases. Table F.3 in Appendix F shows the performance measures calculated for  $G_0$ . It can be observed that all performance measures are decreasing as the sample size increases.

## 5.2 Graphical local Whittle estimation

We consider the following three DGPs to see the performance of the graphical local Whittle estimation. We introduce the matrices  $P_k$ ,  $k = 1, 2, 3$ , to define the sparsity pattern of  $P_0$ . The notation  $P_k(r, s)$ ,  $k = 1, 2, 3$ , denotes the  $(r, s)$ th entry of  $P_k$ .

- (DGP1) One-sided VARFIMA(0,  $D$ , 0) with  $P_0 = (I_{\lfloor p/20 \rfloor} \otimes P_1)$ , where  $P_1(r, r) = 6.28$ ,  $P_1(1, 6) = 4.20 + 1.29i$ ,  $P_1(3, 20) = 2.46 - .51i$  and  $P_1(11, 14) = 3.12 + .39i$ .
- (DGP2) Two-sided VARFIMA(0,  $D$ , 0) with  $P_0 = (I_{\lfloor p/20 \rfloor} \otimes P_2)$ , where the diagonal entries of  $P_2$  are 1 and  $P_2(3, 9) = .5 + .2i$ ,  $P_2(5, 14) = .4 + .2i$ .
- (DGP3) Two-sided VARFIMA(0,  $D$ , 0) with banded  $P_0$  matrix given by  $P_0 = (I_{\lfloor p/20 \rfloor} \otimes P_3)$ , where the diagonal entries of  $P_3$  are 1 and  $P_3(r, r + 1) = .2 + .1i$ ,  $r = 1, \dots, 19$ .

The penalty parameter  $\rho$  is chosen by the eBIC in (4.1) with  $\gamma = 1$ . Also, the ALM algorithm requires the Lagrangian penalty parameter, which as noted above is set to decrease by  $\frac{1}{4}$  on every 10th iteration with  $\mu_0 = 10^{-2}$  but it is taken no smaller than  $10^{-6}$ . The same performance measures are used as those for thresholding local Whittle estimation in Section 5.1, but the inverse of long-run variance  $P_0$  is used instead of  $G_0$ .

Figure 3 shows the proportions of estimated non-zero coefficients in  $P_k$ ,  $k = 1, 2, 3$ , when the dimension is  $p = 20$  and the sample size is  $N = 200$ . Note that the sparsity patterns of  $P_k$ ,  $k = 1, 2, 3$ , are the same as in Figure 1. It can be observed that the graphical local Whittle estimator recovers the sparsity pattern of the underlying model. For example, the non-zero coefficient  $P_1(11, 14)$  is found to be non-zero about 90% of times by our proposed method. More detailed performance measures can be found in Table F.4. Table F.4 suggests that for all considered models our performance measures tend to improve as the sample size increases for fixed dimension. The other way around, for fixed sample size  $N$ , the performance measures are getting worse as the dimension increases. We can conclude that the graphical local Whittle estimator correctly identifies zero coefficients and yields estimates close to the true values as the sample size increases.

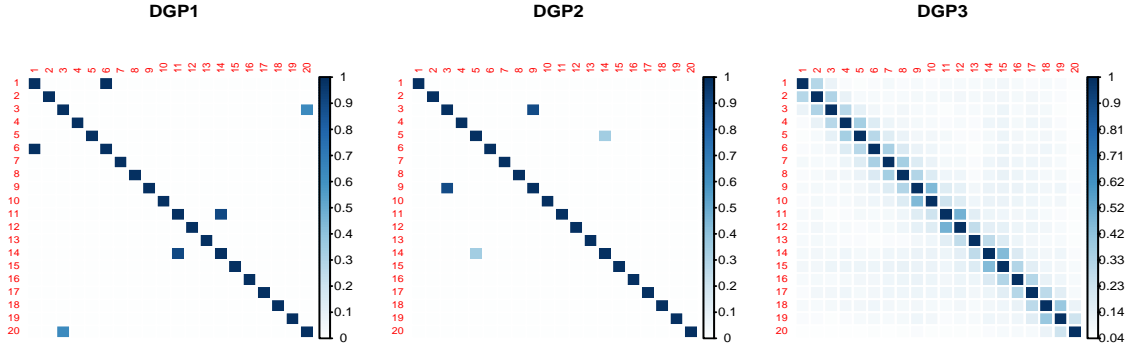


Figure 3: The number of times having non-zero coefficients using graphical local Whittle estimation with eBIC tuning parameter selection where  $p = 20$ ,  $N = 200$ .

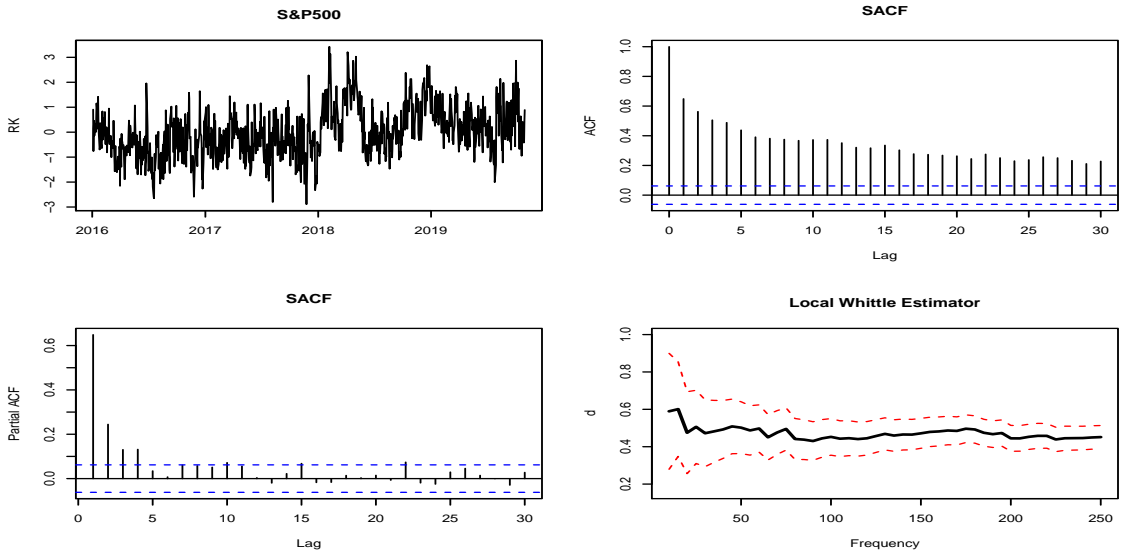


Figure 4: Time plot (top left), sample ACF plot (top right), sample PACF plot (bottom left) and the local Whittle estimators (bottom right) for S&P 500.

## 6 Real data application

In this section, we apply our proposed methods to 31 realized volatilities obtained by aggregating the 5-min within-day returns taken from the Oxford Man Institute of Quantitative Finance (<http://www.oxford-man.ox.ac.uk>). We adjusted the different opening days over the stock markets by applying linear interpolation and log-transforming the data. Furthermore, we removed the possible mean changes in the data by following the proposed procedure in Baek and Pipiras (2014). We also studentized each series to have zero-mean and unit variance in order to focus on volatility linkage. The total number of observations is 1001 dating from Jan 4, 2016 to Oct 31, 2019. The 31 global stock indices are AEX index (AEX), All Ordinaries (AORD), Bell 20 index (BFX), S&P BSE Sensex (BSESN), PSI All Shares Gross Return Index (BVLG), BVSP BOVESPA Index (BVSP), Dow Jones Industrial Average (DJI), CAC 40 (FCHI), FTSE MIB (FTMIB), FTSE 100 (FTSE), DAX (GDAXI), S&P/TSX Composite index (GSPTSE), HANG SENG Index (HSI), IBEX 35 Index (IBEX), Nasdaq 100 (IXIC), Korea Composite Stock Price Index (KS11), Karachi SE 100



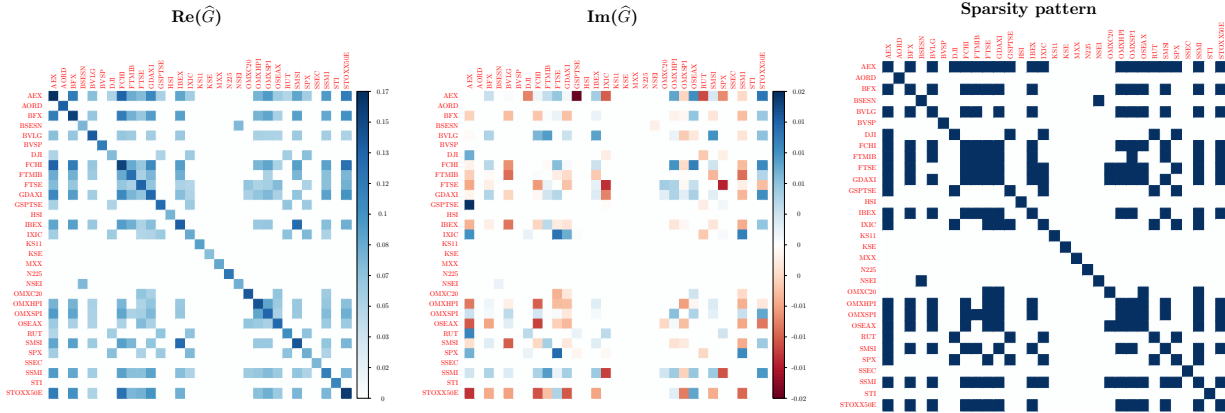


Figure 5: Thresholding local Whittle estimation.

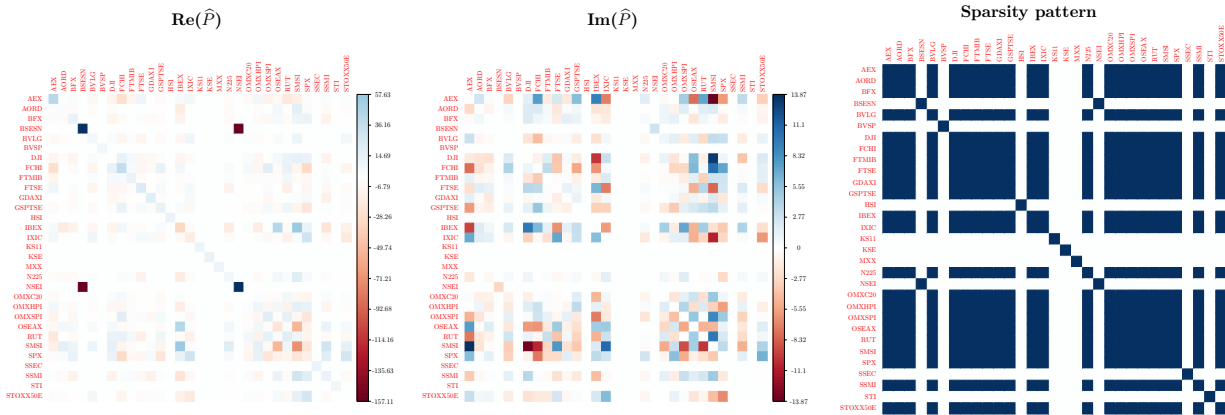


Figure 6: Graphical local Whittle estimation.

Index (KSE), IPC Mexico (MXX), Nikkei 225 (N225), NIFTY 50 (NSEI), OMX Copenhagen 20 Index (OMXC20), OMX Helsinki All Share Index (OMXHPI), OMX Stockholm All Share Index (OMXSPI), Oslo Exchange All-share Index (OSEAX), Russel 2000 (RUT), Madrid General Index (SMSI), S&P 500 Index (SPX), Shanghai Composite Index (SSEC), Swiss Stock Market Index (SSMI), Straits Times Index (STI), EURO STOXX 50 (STOXX50E).

Figure 4 shows some exploratory plots for S&P 500 such as time plot (top left), sample autocorrelation plot (top right), sample partial autocorrelation plot (bottom left) and long-range dependent parameter estimates over the number of frequencies used (bottom right). It shows typical features of long-range dependent time series: non-cyclical trends, slow decay of autocorrelations and memory parameters close to .5. This suggests that multivariate long-range dependence modeling is meaningful and we applied our methods to estimate long-run variance and precision matrices.

Sparse long-run variance matrix estimation by thresholding is presented in Figure 5. The left panel shows the real parts of  $\hat{G}$  and the imaginary parts can be found in the middle panel. The right panel presents the sparsity pattern, that is, the locations of the non-zero coefficients are colored dark blue. Figure 6 follows the same structure as Figure 5 but shows the estimated sparse precision matrix  $\hat{P}$  using graphical local Whittle estimation. The penalty parameter is selected by using eBIC in (4.1) with  $\gamma = 1$ . BSESN and NSEI, both Indian market indices, showed the largest real

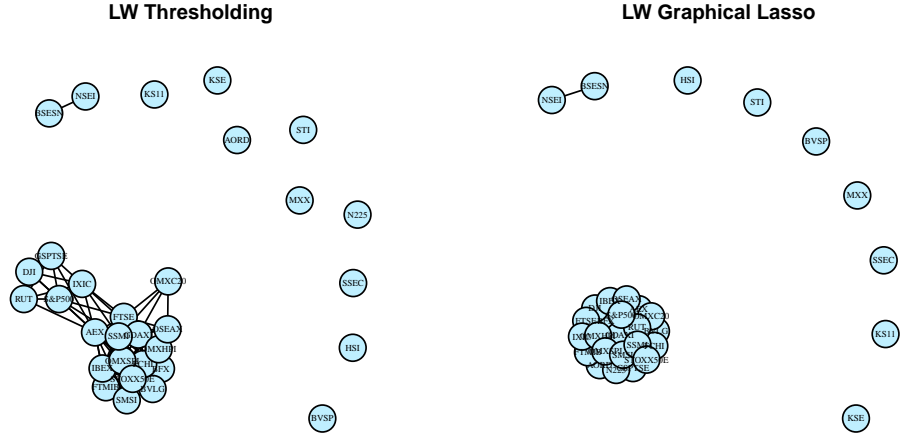


Figure 7: Network graph representation.

coefficients in absolute term.

Observe that the thresholding method gives sparser estimation. We also observe the clustering of stock market indices for both long-run variance and precision matrices. It is seen more clearly from a network representation of linkages as in Figure 7. More interestingly, both methods give similar clusterings. The isolated nodes based on  $\hat{P}$  correspond to India (BSESN/NSEI), China (SSEC), South Korea (KS11), Hong Kong (HSI), Singapore (STI), Mexico (MXN), Portugal (BVSP), Pakistan (KSE). The sparse long-run variance  $\hat{G}$  using thresholding adds Japan (N225) and Australia (AORL). That is, it seems that our empirical analysis tells that stock market indices can be roughly divided into the US-European market and somewhat independent markets from the rest of the world including Asia, India, Australia and Mexico in late 2010s. The thresholding method seems to further distinguish US and European markets. In fact, all US stock indices showed larger values of memory parameter estimates by having more than .4 which is not the case for European market indices. It is particularly interesting to see the grouping of multinational realized volatilities according to regional or spatial dependence.

Finally, we note but do not include the results here, that the analysis of the data assuming short-range dependence led to highly non-sparse patterns for the considered connectivity matrices. These findings were consistent with the scenario where we simulated data from LRD models as in Section 5 but worked as if they were SRD.

## 7 Conclusions

In this work, we derived consistency results for the long-run variance and precision matrices in a non-asymptotic regime allowing the underlying time series to admit a general dependence structure including long-range dependence. The results are derived under mild assumptions on the underlying time series which is allowed to be either Gaussian or have a linear representation. The shrinkage techniques are thresholding and graphical local Whittle estimation. Our non-asymptotic results can be used to infer consistency in a high-dimensional regime where the number of component series can be large compared to the sample size.

The key technical contribution is the incorporation of the memory parameter matrix which carries information about the dependence structure of the underlying time series. Our results allow

estimating those memory parameters simultaneously while estimating the long-run variance and the precision matrices sparsely.

We see the proposed proof techniques as a basis to study other questions concerning high-dimensional long-range dependence. Possible future directions include the use of other shrinkage methods, for example, adaptive penalizations; sparse estimation of fractionally (co)integrated vector autoregressive (VAR) models and sparse estimation of linear regression with long-range dependent errors.

## A Quantities in main results and special cases

In this section we provide the expressions for several quantities appearing in Propositions 3.1–3.4. In addition, we will discuss the case when the underlying time series admits only short- or long-range dependence.

Recall that the non-asymptotic bounds in Propositions 3.1–3.4 are all of the form

$$\mathbf{C} \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_i}} + \mathcal{T}_i \quad \text{with} \quad \mathcal{R}_i = \min\{s_{i,N} \mathcal{R}_{i1}, s_{i,N}^2 \mathcal{R}_{i2}\}. \quad (\text{A.1})$$

Here,  $\mathcal{R}_i$ 's arise in bounds on the probabilistic parts and  $\mathcal{T}_i$ 's on deterministic parts. The sequences  $s_{i,N}$  can be found in Propositions 3.1–3.4. The quantities  $\mathcal{R}_{i1}$ ,  $\mathcal{R}_{i2}$  and  $\mathcal{T}_i$  are respectively given in Tables A.1 and A.2 for Propositions 3.1–3.4.

Recall that  $\Delta_1, \Delta_2$  determine the interval of admissible estimates of the memory parameters and the quantity  $\bar{\Delta}_N$  is defined in (3.13). We further introduce a few quantities which will allow us to express our bounds in a simplified way and emphasize the necessary distinction between different ranges of the memory parameters as will become clearer in the proofs. Let

$$\begin{aligned} \bar{\Delta}_u &= \max_{r=1, \dots, p} \left\{ \Delta_2, \left( \frac{1}{4} + d_{0,r} \right) \mathbb{1}_{\{d_{0,r} < \frac{1}{4}\}} \right\}, \\ \bar{\Delta}_{l,1} &= \max_{r=1, \dots, p} d_{0,r} \mathbb{1}_{\{d_{0,r} < \Delta_1 + \frac{1}{2}\}}, \quad \bar{\Delta}_{l,2} = \max_{r=1, \dots, p} \left( \frac{1}{4} \mathbb{1}_{\{d_{0,r} < \min\{\Delta_1 + \frac{1}{2}, \frac{1}{4}\}\}} + d_{0,r} \mathbb{1}_{\{d_{0,r} < \Delta_1 + \frac{1}{2}\}} \right), \end{aligned} \quad (\text{A.2})$$

where the subscripts  $u$  and  $l$  allude to the dependence on  $\Delta_2$  and  $\Delta_1$ , respectively.

For the bounds on the deterministic parts we will use

$$Q_m = \|G\| \frac{72(\cos(\lambda_m/2))^{-2}}{\pi(1 + 2 \min\{\Delta_1, -\Delta_2\})} + \mathbf{c}_{G,2} 4(2 + \log(m)) \quad (\text{A.3})$$

and  $\tilde{\Delta}_r = (d_{0,r} - \frac{1}{2} + \Delta) \mathbb{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} + \Delta_1 \mathbb{1}_{\{d_{0,r} < \Delta_1 + \frac{1}{2}\}}$ . Recall further that  $q$  appears in Assumption 2.

As a corollary of Proposition 3.5, we give the result where the underlying process is known to admit only short- or long-range dependence, that is, the true memory parameters satisfy  $D_0 \succ 0$ .

**Corollary A.1.** *Let  $\{X_n\}_{n \in \mathbb{Z}}$  be a  $p$ -dimensional, stationary, centered time series with spectral density  $f_X$  and suppose Assumptions 1–4 are satisfied. Then, there are positive constants  $c_1, c_2$  such that for any  $\mathbf{C} \geq 1$ ,*

$$\mathbb{P}(\|\widehat{G}(\widehat{D}) - G_0\|_{\max} > \delta) \leq c_1 p^{2-c_2 \mathbf{C}}$$

with  $\delta$  as in (3.19) and  $\nu, \nu_i$ ,  $i = 1, 2, 3$ , are as in (3.14), (3.15), (3.16) and (3.17) with

$$\begin{aligned} \mathcal{R}_{11} &= m^{1-2\Delta_2+2\epsilon} N^{-2\epsilon}, \quad \mathcal{R}_{12} = \min\{m^{2-4\bar{\Delta}_u+4\epsilon} N^{-4\epsilon}, m^{\frac{3}{2}-2\bar{\Delta}_u+2\epsilon} N^{-2\epsilon}\}, \\ \mathcal{R}_{21} &= m^{1-2\Delta_2}, \quad \mathcal{R}_{22} = m^{2-4\bar{\Delta}_u}, \\ \mathcal{R}_{31} &= \min\{m^{1-2\Delta_2}, m^{2\Delta}, m^{1-2\bar{\Delta}_{l,1}}\}, \quad \mathcal{R}_{32} = \min\{m^{2-4\bar{\Delta}_u}, m^{4\Delta}, m^{2-4\bar{\Delta}_{l,2}}\}, \end{aligned} \quad (\text{A.4})$$

Prop.	$\mathcal{R}_{i1}$ and $\mathcal{R}_{i2}$ in (A.1) (probabilistic parts)
3.1	$\mathcal{R}_{11} = \min \left\{ m^{1-2\Delta_2+2\varepsilon} N^{-2\varepsilon}, m^{1-\Delta_2-\Delta_1+\varepsilon} N^{\Delta_1-\varepsilon}, m N^{\Delta_1-\varepsilon} \right\},$ $\mathcal{R}_{12} = \min \left\{ m^{2-4\bar{\Delta}_u+4\varepsilon} N^{-4\varepsilon}, m^{\frac{3}{2}-2\bar{\Delta}_u-2\Delta_1+2\varepsilon} N^{2\Delta_1-2\varepsilon}, m N^{2\Delta_1-2\varepsilon}, m^2 \left( \sum_{j=1}^m \lambda_j^{4\Delta_1} \right)^{-1} \right\}.$
3.2	$\mathcal{R}_{21} = \min \left\{ m^{1-2\Delta_2}, m N^{2\Delta_1} \right\}, \quad \mathcal{R}_{22} = \min \left\{ m^{2-4\bar{\Delta}_u}, m^2 \left( \sum_{j=1}^m \lambda_j^{4\Delta_1} \right)^{-1} \right\}.$
3.3	$\mathcal{R}_{31} = \min \left\{ m^{1-2\Delta_2}, m^{2\Delta}, m^{1-2\bar{\Delta}_{l,1}+2\Delta_1}, m N^{2\Delta_1} \right\},$ $\mathcal{R}_{32} = \min \left\{ m^{2-4\bar{\Delta}_u}, m^{4\Delta}, m^{4\Delta} N^{1+4\Delta_1}, m^{2-4\bar{\Delta}_{l,2}+4\Delta_1}, m^2 N^{4\Delta_1}, m^2 \left( \sum_{j=1}^m \lambda_j^{4\Delta_1} \right)^{-1} \right\}.$
3.4	$\mathcal{R}_{41} = \min \{ m \ell^{-1+2\Delta}, m^{1-2\Delta_2} \}, \quad \mathcal{R}_{42} = \min \{ m \ell^{-2+4\Delta}, m^{2-4\bar{\Delta}_u} \}.$

Table A.1: Expressions for  $\mathcal{R}_{i1}$  and  $\mathcal{R}_{i2}$  in (A.1).

and it is assumed that  $\varepsilon \in (0, \frac{1}{2})$ .

The corollary is a simple consequence of setting  $\Delta_1 = 0$  in Proposition 3.5.

## B Proofs of the main results

In this section we will give bounds on the probabilistic and deterministic terms in (3.12), stated respectively in Lemmas B.1 and B.2. Lemma B.1 focusses on results under the assumption that the underlying process is Gaussian. Its analogue for linear processes can be found in Appendix E. Up to a constant the bounds are the same as for the Gaussian case. For this reason all proceeding results remain true. Note that Lemma B.2 only relies on assumptions on the spectral density and remains valid as well.

The following Lemma B.1 gives multiple upper bounds on the probabilistic term in (3.12). Those different bounds are used later depending on whether the true parameters  $d_{0,r}$  and  $d_{0,s}$  are positive or negative. Recall in particular the notation  $t_{j,r}(d)$  below (3.9) and  $a_r, b_r$  in (3.10)–(3.11). Let

$$L_{rs,i} = c_{r,i,N} \tilde{c}_{s,i,N} \left( \|T(a_r, a_s)\| + \sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} \|\mathbf{T}(d_r, d_s)\| \right), \quad i = 1, \dots, 4, \quad (\text{B.1})$$

$$L_{rs,5} = N^{\max\{d_{0,r}, 0\} + \max\{d_{0,s}, 0\}} \left( \|T(a_r, a_s)\|_F + \sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} \|\mathbf{T}(d_r, d_s)\|_F \right)$$

with

$$T(d_r, d_s) = \text{diag}(t_{1,r}(d_r)t_{1,s}(d_s), \dots, t_{m,r}(d_r)t_{m,s}(d_s)), \quad (\text{B.2})$$

$$\mathbf{T}(d_r, d_s) = \text{diag}(\|\nabla t_{1,r}(d_r)t_{1,s}(d_s)\|_F, \dots, \|\nabla t_{m,r}(d_r)t_{m,s}(d_s)\|_F)$$

Prop.	$\mathcal{T}_i$ in (A.1) (deterministic parts)
3.1	$\mathcal{T}_1(\varepsilon) = \mathbf{c}_{G,1} \frac{1}{2\pi} \lambda_m^{2q-2\varepsilon} + \frac{1}{2\pi m} \left(1 + \frac{1}{2\varepsilon}\right) \lambda_1^{-2\varepsilon} Q_m$
3.2	$\mathcal{T}_2 = \mathbf{c}_{G,1} \frac{1}{2\pi} \lambda_m^{2q} + \frac{1}{\pi m} \log(m) Q_m$
3.3	$\begin{aligned} \mathcal{T}_3 &= \max_{r=1,\dots,p} \mathcal{T}_3(\tilde{\Delta}_r) \\ &= \max_{r=1,\dots,p} \left( \mathbf{c}_{G,1} \frac{1}{2\pi} \lambda_m^{2q} \frac{1}{2\tilde{\Delta}_r} + \frac{1}{2\pi} m^{-2\tilde{\Delta}_r} \left(1 + \frac{1}{1-2\tilde{\Delta}_r}\right) Q_m \right) \end{aligned}$
3.4	$\mathcal{T}_4 = \mathbf{c}_{G,1} \frac{1}{2\pi} \lambda_m^{2q} \left(\frac{1}{2\Delta} + 1\right) + \frac{1}{2\pi m} \left(\ell^{1-2\Delta} \left(1 + \frac{1}{1-2\Delta}\right) + \log(m)\right) Q_m$

Table A.2: Expressions for  $\mathcal{T}_i$  in (A.1).

and

$$\begin{aligned} c_{r,1,N} &= N^{\max\{d_{0,r},0\}}, \tilde{c}_{s,1,N} = N^{\max\{d_{0,s},0\}}; \quad c_{r,2,N} = N^{\max\{d_{0,r},\frac{1}{4}\}}, \tilde{c}_{s,2,N} = N^{\max\{d_{0,s},\frac{1}{4}\}}; \\ c_{r,3,N} &= N^{\max\{d_{0,r},\frac{1}{4}\}}, \tilde{c}_{s,3,N} = m^{\frac{1}{4}} N^{\max\{d_{0,s},0\}}; \quad c_{r,4,N} = m^{\frac{1}{4}} N^{\max\{d_{0,r},0\}}, \tilde{c}_{s,4,N} = m^{\frac{1}{4}} N^{\max\{d_{0,s},0\}}. \end{aligned}$$

Note that  $c_{r,i,N}$  is different from  $\tilde{c}_{r,i,N}$  only for  $i = 3$ . This notation though will allow writing our arguments in a more unified way.

**Lemma B.1.** *Let  $\{X_n\}_{n \in \mathbb{Z}}$  be a  $p$ -dimensional stationary, centered, Gaussian time series with spectral density  $f_X$  as in (1.1). Then, there are positive constants  $c_1, c_2$  such that*

$$\mathbb{P}\left(\sup_{D \in \Omega} |\hat{H}_{rs}(D) - \mathbb{E}\hat{H}_{rs}(D)| > \|G\| \nu\right) \leq \mathcal{B}(r, s, i), \quad i = 2, \dots, 5, \quad (\text{B.3})$$

for  $\nu^2 \geq \gamma^4 L_{rs,i}^2 / (m^2 c_2)$ , where

$$\mathcal{B}(r, s, i) = c_1 \exp\left(-c_2 \min\left\{\frac{\nu m}{\gamma^2 \bar{\Delta}_N L_{rs,1}}, \frac{\nu^2 m^2}{\gamma^4 \bar{\Delta}_N^2 L_{rs,i}^2}\right\}\right) \quad (\text{B.4})$$

with  $r \neq s$  if  $i = 3$  and  $\Omega$  is given in (3.10). The constant  $\gamma = \frac{\sqrt{2}}{\Gamma(\frac{1}{2})}$  bounds the sub-Gaussian norm of a standard normal random variable as in (D.1) and  $\bar{\Delta}_N$  is defined in (3.13).

*Proof:* To prove the desired concentration inequality, we first introduce some general notation. For a fixed frequency, the periodogram in (1.4) can be written as

$$I_X(\lambda_j) = \frac{1}{2\pi} \left( \mathcal{X}'(C_j C_j' + S_j S_j') \mathcal{X} + i \mathcal{X}'(C_j S_j' - S_j C_j') \mathcal{X} \right) \quad (\text{B.5})$$

with  $\mathcal{X}' = [X_1 : \dots : X_N]$  and

$$\begin{aligned} C_j' &= \frac{1}{\sqrt{N}} (\cos(\lambda_j), \dots, \cos((N-1)\lambda_j), 1), \\ S_j' &= \frac{1}{\sqrt{N}} (\sin(\lambda_j), \dots, \sin((N-1)\lambda_j), 0); \end{aligned} \quad (\text{B.6})$$

see equation (A.6) in Sun et al. (2018). The event of interest can be separated into real and imaginary parts as

$$\begin{aligned}
& \mathbb{P}\left(\sup_{D \in \Omega} |\widehat{H}_{rs}(D) - \mathbb{E}\widehat{H}_{rs}(D)| > \|G\| \nu\right) \\
& \leq \mathbb{P}\left(\sup_{D \in \Omega} |\Re(\widehat{H}_{rs}(D) - \mathbb{E}\widehat{H}_{rs}(D))| > \frac{1}{2} \|G\| \nu\right) \\
& \quad + \mathbb{P}\left(\sup_{D \in \Omega} |\Im(\widehat{H}_{rs}(D) - \mathbb{E}\widehat{H}_{rs}(D))| > \frac{1}{2} \|G\| \nu\right).
\end{aligned} \tag{B.7}$$

Note that the imaginary part of the diagonal elements is zero, that is,  $\Im(e'_r I_X(\lambda_j) e_r) = 0$ . We consider the diagonal elements first ( $r = s$ ) and then distinguish the real and imaginary parts in (B.7) for the off-diagonal elements ( $r \neq s$ ).

*Diagonal elements:* To rewrite the real-valued diagonal elements  $\widehat{H}_{rr}(D)$  as a quadratic form, we define the  $N \times 2m$  matrix

$$R'_m = [C_1 : S_1 : \dots : C_m : S_m] \tag{B.8}$$

and the matrix-valued function

$$T_r(d) = \text{diag}(t_{1,r}(d), t_{1,r}(d), \dots, t_{m,r}(d), t_{m,r}(d)). \tag{B.9}$$

Then, in view of (B.5), (B.8), (B.9) and since  $e'_r t_j(D) = e'_r t_{j,r}(d_r)$ ,  $\widehat{H}_{rr}(D)$  can be written as

$$\begin{aligned}
\frac{1}{m} e'_r \sum_{j=1}^m t_j(D) I_X(\lambda_j) t_j(D) e_r &= \frac{1}{2\pi m} e'_r \sum_{j=1}^m t_j(D) \mathcal{X}'(C_j C'_j + S_j S'_j) \mathcal{X} t_j(D) e_r \\
&= \frac{1}{2\pi m} \sum_{j=1}^m \mathcal{E}'_r \Sigma_{rr}^{\frac{1}{2}} t_{j,r}(d_r) (C_j C'_j + S_j S'_j) t_{j,r}(d_r) \Sigma_{rr}^{\frac{1}{2}} \mathcal{E}_r \\
&= \frac{1}{2\pi m} \mathcal{E}'_r \Sigma_{rr}^{\frac{1}{2}} R'_m T_r^2(d_r) R_m \Sigma_{rr}^{\frac{1}{2}} \mathcal{E}_r
\end{aligned} \tag{B.10}$$

with  $e'_r \mathcal{X}' = \mathcal{E}'_r \Sigma_{rr}^{\frac{1}{2}}$ , where  $\mathcal{E}_r = (\varepsilon_{r,1}, \dots, \varepsilon_{r,N})'$  is Gaussian with  $\mathbb{E}(\mathcal{E}_r \mathcal{E}'_r) = I_N$  and  $\Sigma_{rr} = (\Sigma_{rr}(n-k))_{n,k=1,\dots,N} = \mathbb{E}(\mathcal{X} e_r (\mathcal{X} e_r)')$ . In (B.11) below, we apply Theorem D.1 with  $K = 1$  and  $R = b_r - a_r \leq 1$  to obtain

$$\begin{aligned}
& \mathbb{P}\left(\sup_{D \in \Omega} |\widehat{H}_{rr}(D) - \mathbb{E}\widehat{H}_{rr}(D)| > \|G\| \nu\right) \\
& = \mathbb{P}\left(\sup_{d \in [a_r, b_r]} |\mathcal{E}'_r \Sigma_{rr}^{\frac{1}{2}} R'_m T_r^2(d) R_m \Sigma_{rr}^{\frac{1}{2}} \mathcal{E}_r - \mathbb{E}(\mathcal{E}'_r \Sigma_{rr}^{\frac{1}{2}} R'_m T_r^2(d) R_m \Sigma_{rr}^{\frac{1}{2}} \mathcal{E}_r)| > 2\pi m \|G\| \nu\right) \\
& = \mathbb{P}\left(\sup_{d \in [a_r, b_r]} |\mathcal{E}'_r R(d) \mathcal{E}_r - \mathbb{E}(\mathcal{E}'_r R(d) \mathcal{E}_r)| > 2\pi m \|G\| \nu\right) \\
& \leq c_1 \exp\left(-c_2 \min\left\{\frac{\nu m \|G\|}{\gamma^2 \widetilde{\mathcal{T}}_1}, \frac{\nu^2 m^2 \|G\|^2}{\gamma^4 \widetilde{\mathcal{T}}_i^2}\right\}\right)
\end{aligned} \tag{B.11}$$

$$\leq c_1 \exp\left(-c_2 \min\left\{\frac{\nu m}{\gamma^2 \widetilde{\Delta}_N \widetilde{L}_1}, \frac{\nu^2 m^2}{\gamma^4 \widetilde{\Delta}_N^2 \widetilde{L}_i^2}\right\}\right) \tag{B.12}$$

for  $\nu^2 \geq \gamma^4 \tilde{T}_i^2 / (c_2 m^2 \|G\|^2)$  with  $i = 2, 3, 4$  and  $R(d) = \Sigma_{rr}^{\frac{1}{2}} R'_m T_r^2(d) R_m \Sigma_{rr}^{\frac{1}{2}}$  and

$$\begin{aligned}\tilde{T}_1 &= \|A_m\| \left( \|T_r^2(a_r)\| + \sup_{d \in [a_r, b_r]} \left\| \frac{\partial}{\partial d} T_r^2(d) \right\| \right), \\ \tilde{T}_2 = \tilde{T}_3 &= \|A_m\|_F \left( \|T_r^2(a_r)\| + \sup_{d \in [a_r, b_r]} \left\| \frac{\partial}{\partial d} T_r^2(d) \right\| \right), \\ \tilde{T}_4 &= \|A_m\| \left( \|T_r^2(a_r)\|_F + \sup_{d \in [a_r, b_r]} \left\| \frac{\partial}{\partial d} T_r^2(d) \right\|_F \right),\end{aligned}\tag{B.13}$$

where  $A_m = R_m \Sigma_{rr} R'_m$ . Though  $\tilde{T}_2 = \tilde{T}_3$ , we bound them differently in (B.12). Furthermore,

$$\tilde{L}_1 = L_{rr,1}, \quad \tilde{L}_2 = L_{rr,2}, \quad \tilde{L}_3 = L_{rr,4}, \quad \tilde{L}_4 = L_{rr,5}\tag{B.14}$$

with  $L_{rr,1}, L_{rr,2}, L_{rr,4}$  and  $L_{rr,5}$  as in (B.1). We get (B.12) by bounding the quantities in (B.13). Note that

$$\|A_m\| = \|R_m \Sigma_{rr} R'_m\| \leq \|R_m\|^2 \|\Sigma_{rr}\| \leq c \|G\| N^{\max\{2d_0, r, 0\}} \bar{\Delta}_N,\tag{B.15}$$

$$\|A_m\|_F = \|R_m \Sigma_{rr} R'_m\|_F \leq \|R_m\|^2 \|\Sigma_{rr}\|_F \leq c \|G\| N^{\max\{2d_0, r, \frac{1}{2}\}} \log(N)^{\frac{1}{2}} \bar{\Delta}_N,\tag{B.16}$$

$$\|A_m\|_F \leq \sqrt{2m} \|R_m \Sigma_{rr} R'_m\| \leq c \|G\| \sqrt{2m} N^{\max\{2d_0, r, 0\}} \bar{\Delta}_N,\tag{B.17}$$

where we used the submultiplicativity of the spectral norm, Lemma D.1 and the fact that  $\|A\|_F \leq \sqrt{\text{rk}(A)} \|A\|$  for a matrix  $A$  for the first inequalities in (B.15)–(B.17), respectively. The last inequalities in (B.15) and (B.16) for the spectral and Frobenius norms of  $\Sigma_{rr}$  follow by Lemmas C.12 and C.13. Furthermore, we used the fact that the spectral norm of  $R_m$  can be calculated as

$$\|R_m\| = 1;$$

see Lemma C.4 in Sun et al. (2018).

*Real part (off-diagonal):* The real part of the off-diagonal elements  $\Re(\hat{H}_{rs}(D))$  can be written in terms of (B.5) as

$$\frac{1}{m} e'_r \sum_{j=1}^m t_j(D) \Re(I_X(\lambda_j)) t_j(D) e_s = \frac{1}{2\pi m} e'_r \sum_{j=1}^m t_j(D) \mathcal{X}'(C_j C'_j + S_j S'_j) \mathcal{X} t_j(D) e_s.\tag{B.18}$$

As in (B.10), we will write (B.18) as a quadratic form but now using  $(e'_r \mathcal{X}' \ e'_s \mathcal{X}') = \mathcal{E}' \Sigma^{\frac{1}{2}}$ , where  $\mathcal{E}' = (\mathcal{E}'_r \ \mathcal{E}'_s)$  is a Gaussian vector with  $\mathbb{E}(\mathcal{E} \mathcal{E}') = I_{2N}$  and

$$\Sigma = \begin{pmatrix} \Sigma_{rr} & \Sigma_{rs} \\ \Sigma_{sr} & \Sigma_{ss} \end{pmatrix} \quad \text{with} \quad \Sigma_{rs} = (\Sigma_{rs}(n-k))_{n,k=1,\dots,N} = \mathbb{E}(\mathcal{X} e_r (\mathcal{X} e_s)').\tag{B.19}$$

For this, define the  $4m \times 2N$  matrix

$$\tilde{R}_m = \begin{pmatrix} R_m & 0_{2m \times N} \\ 0_{2m \times N} & R_m \end{pmatrix}$$

with  $R_m$  as in (B.8). Furthermore, define the matrix  $F(d_r, d_s) = \text{diag}(T_r(d_r), T_s(d_s))$  with  $T_r$  as in (B.9) and

$$M_m = \begin{pmatrix} 0_{m,m} & I_m \\ 0_{m,m} & 0_{m,m} \end{pmatrix}.\tag{B.20}$$

Write

$$\begin{aligned} e'_r \sum_{j=1}^m t_j(D) \mathcal{X}'(C_j C'_j + S_j S'_j) \mathcal{X} t_j(D) e_s &= \mathcal{E}' \Sigma^{\frac{1}{2}} \tilde{R}'_m F(d_r, d_s) M_{2m} F(d_r, d_s) \tilde{R}_m \Sigma^{\frac{1}{2}} \mathcal{E} \\ &= \mathcal{E}' R(d_r, d_s) \mathcal{E} \end{aligned}$$

with

$$R(d_r, d_s) = \Sigma^{\frac{1}{2}} \tilde{R}'_m F(d_r, d_s) M_{2m} F(d_r, d_s) \tilde{R}_m \Sigma^{\frac{1}{2}}.$$

In order to apply Theorem D.1, we further write

$$\begin{aligned} &\mathbb{P}\left(\sup_{D \in \Omega} \left| \sum_{j=1}^m e'_r(t_j(D) \mathcal{X}'(C_j C'_j + S_j S'_j) \mathcal{X} t_j(D) - \mathbb{E}(t_j(D) \mathcal{X}'(C_j C'_j + S_j S'_j) \mathcal{X} t_j(D))) e_s \right| > \pi m \| \|G\| \nu\right) \\ &= \mathbb{P}\left(\sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} |\mathcal{E}' R(d_r, d_s) \mathcal{E} - \mathbb{E}(\mathcal{E}' R(d_r, d_s) \mathcal{E})| > \pi m \| \|G\| \nu\right). \end{aligned} \quad (\text{B.21})$$

Note that the matrix  $R(d_r, d_s)$  can be rewritten as

$$R(d_r, d_s) = \Sigma^{\frac{1}{2}} \mathcal{A}_{N,i}^{-1} \tilde{R}'_m \mathcal{A}_{2m,i} F(d_r, d_s) M_{2m} F(d_r, d_s) \mathcal{A}_{2m,i} \tilde{R}_m \mathcal{A}_{N,i}^{-1} \Sigma^{\frac{1}{2}} \quad (\text{B.22})$$

for  $i = 1, \dots, 4$  with

$$\begin{aligned} \mathcal{A}_{m,1} &= \text{diag}(c_{r,1,N} I_m, \tilde{c}_{s,1,N} I_m), \quad \mathcal{A}_{m,2} = \text{diag}(c_{r,2,N} I_m, \tilde{c}_{s,2,N} I_m), \\ \mathcal{A}_{m,3} &= \text{diag}(c_{r,3,N} I_m, \tilde{c}_{s,3,N} I_m), \quad \mathcal{A}_{m,4} = m^{\frac{1}{4}} \mathcal{A}_{m,1}. \end{aligned} \quad (\text{B.23})$$

The matrices  $\mathcal{A}_{N,i}$  in (B.22) are defined by replacing  $I_m$ 's in (B.23) by  $I_N$ . These matrices are to normalize  $\Sigma^{\frac{1}{2}}$  in (B.22).

We continue to bound (B.21) by applying Theorem D.1. In order to verify the applicability of Theorem D.1, note that the matrix  $\mathcal{A}_{2m,i} F(d_r, d_s) M_{2m} F(d_r, d_s) \mathcal{A}_{2m,i}$  in (B.22) is not diagonal but can be represented as a unitary transformation of a diagonal matrix

$$\begin{aligned} &\mathcal{A}_{2m,i} F(d_r, d_s)' M_{2m} F(d_r, d_s) \mathcal{A}_{2m,i} \\ &= \mathcal{A}_{2m,i} \begin{pmatrix} 0_{2m,2m} & T_r(d_r) T_s(d_s) \\ 0_{2m,2m} & 0_{2m,2m} \end{pmatrix} \mathcal{A}_{2m,i} \\ &= \begin{pmatrix} 0_{2m,2m} & c_{r,i,N} \tilde{c}_{s,i,N} T_r(d_r) T_s(d_s) \\ 0_{2m,2m} & 0_{2m,2m} \end{pmatrix} \\ &= \begin{pmatrix} I_{2m} & 0_{2m,2m} \\ 0_{2m,2m} & I_{2m} \end{pmatrix} \begin{pmatrix} c_{r,i,N} \tilde{c}_{s,i,N} T_r(d_r) T_s(d_s) & 0_{2m,2m} \\ 0_{2m,2m} & 0_{2m,2m} \end{pmatrix} \begin{pmatrix} 0_{2m,2m} & I_{2m} \\ I_{2m} & 0_{2m,2m} \end{pmatrix}. \end{aligned} \quad (\text{B.24})$$

For this reason, Theorem D.1 remains applicable due to Remark D.1. In (B.25) below, we apply Theorem D.1 with  $K = 2$  and  $R = b_r - a_r \leq 1$  to obtain

$$\begin{aligned} &\mathbb{P}\left(\sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} |\mathcal{E}' R(d_r, d_s) \mathcal{E} - \mathbb{E}(\mathcal{E}' R(d_r, d_s) \mathcal{E})| > \pi m \| \|G\| \nu\right) \\ &\leq c_1 \exp\left(-c_2 \min\left\{\frac{\nu m \| \|G\|}{\gamma^2 \mathcal{T}_1}, \frac{\nu^2 m^2 \| \|G\|^2}{\gamma^4 \mathcal{T}_i^2}\right\}\right) \end{aligned} \quad (\text{B.25})$$

$$\leq c_1 \exp\left(-c_2 \min\left\{\frac{\nu m}{\gamma^2 \bar{\Delta}_N L_{rs,1}}, \frac{\nu^2 m^2}{\gamma^4 \bar{\Delta}_N^2 L_{rs,i}^2}\right\}\right) \quad (\text{B.26})$$



for  $\nu^2 \geq \gamma^4 \mathcal{T}_i^2 / (c_2 m^2 \|G\|^2)$  and  $i = 2, \dots, 5$  with

$$\mathcal{T}_1 = \|B_{m,1}\|_{L_{rs,1}}, \quad \mathcal{T}_i = \|B_{m,i}\|_F L_{rs,i}, \quad \mathcal{T}_5 = \|B_{m,1}\|_{L_{rs,5}} \quad (\text{B.27})$$

for  $i = 2, \dots, 4$ , where the  $L_{rs,i}$ 's are given in (B.1) and

$$B_{m,i} = \tilde{R}_m \mathcal{A}_{N,i}^{-1} \Sigma \mathcal{A}_{N,i}^{-1} \tilde{R}_m' \quad (\text{B.28})$$

for  $i = 1, \dots, 4$ . The  $L_{rs,i}$ 's in (B.27) can indeed be represented as in (B.1) due to (B.24) and Remark D.1, and since  $\|T_r(d_r)T_s(d_s)\| = \|T(d_r, d_s)\|$  and  $\|T_r(d_r)T_s(d_s)\|_F = \sqrt{2}\|T(d_r, d_s)\|_F$  and similarly with  $\nabla$ , where  $T(d_r, d_s)$  is in (B.2). We will now discuss bounds on the Frobenius and spectral norms of  $B_{m,i}$  to get (B.26).

The relation (B.26) is a consequence of bounding the quantities in (B.27) as follows. Let  $c$  denote a generic constant which might differ from line to line. Then, with the explanations given below,

$$\begin{aligned} \|B_{m,1}\| &= \|\tilde{R}_m \mathcal{A}_{N,1}^{-1} \Sigma \mathcal{A}_{N,1}^{-1} \tilde{R}_m'\| \leq \|\tilde{R}_m\|^2 \|\mathcal{A}_{N,1}^{-1} \Sigma \mathcal{A}_{N,1}^{-1}\| \\ &\leq \|c_{r,1,N}^{-2} \Sigma_{rr}\| + \|\tilde{c}_{s,1,N}^{-2} \Sigma_{ss}\| \\ &= N^{-\max\{2d_{0,r},0\}} \|\Sigma_{rr}\| + N^{-\max\{2d_{0,s},0\}} \|\Sigma_{ss}\| \leq c\bar{\Delta}_N \|G\|, \end{aligned} \quad (\text{B.29})$$

$$\begin{aligned} \|B_{m,2}\|_F &= \|\tilde{R}_m \mathcal{A}_{N,2}^{-1} \Sigma \mathcal{A}_{N,2}^{-1} \tilde{R}_m'\|_F \leq \|\tilde{R}_m\|^2 \|\mathcal{A}_{N,2}^{-1} \Sigma \mathcal{A}_{N,2}^{-1}\|_F \\ &\leq \|c_{r,2,N}^{-2} \Sigma_{rr}\|_F + \|\tilde{c}_{s,2,N}^{-2} \Sigma_{ss}\|_F \\ &= N^{-\max\{2d_{0,r}, \frac{1}{2}\}} \|\Sigma_{rr}\|_F + N^{-\max\{2d_{0,s}, \frac{1}{2}\}} \|\Sigma_{ss}\|_F \leq c\bar{\Delta}_N \|G\|, \end{aligned} \quad (\text{B.30})$$

$$\begin{aligned} \|B_{m,3}\|_F &= \|\tilde{R}_m \mathcal{A}_{N,3}^{-1} \Sigma \mathcal{A}_{N,3}^{-1} \tilde{R}_m'\|_F \\ &\leq \|R_m' c_{r,3,N}^{-1} \Sigma_{rr} c_{r,3,N}^{-1} R_m\|_F + \|R_m' \tilde{c}_{s,3,N}^{-1} \Sigma_{ss} \tilde{c}_{s,3,N}^{-1} R_m\|_F \\ &\leq \|R_m' c_{r,3,N}^{-1} \Sigma_{rr} c_{r,3,N}^{-1} R_m\|_F + \sqrt{2m} \|R_m' \tilde{c}_{s,3,N}^{-1} \Sigma_{ss} \tilde{c}_{s,3,N}^{-1} R_m\|_F \\ &\leq \|R_m\|^2 \|c_{r,3,N}^{-2} \Sigma_{rr}\|_F + \sqrt{2m} \|R_m\|^2 \|\tilde{c}_{s,3,N}^{-2} \Sigma_{ss}\| \\ &= N^{-\max\{2d_{0,r}, \frac{1}{2}\}} \|\Sigma_{rr}\|_F + N^{-\max\{2d_{0,s},0\}} \|\Sigma_{ss}\| \leq c\bar{\Delta}_N \|G\|, \end{aligned} \quad (\text{B.31})$$

$$\begin{aligned} \|B_{m,4}\|_F &= \|\tilde{R}_m \mathcal{A}_{N,4}^{-1} \Sigma \mathcal{A}_{N,4}^{-1} \tilde{R}_m'\|_F \\ &\leq \sqrt{4m} \|\tilde{R}_m' \mathcal{A}_{N,4}^{-1} \Sigma \mathcal{A}_{N,4}^{-1} \tilde{R}_m\| = \sqrt{4} \|\tilde{R}_m' \mathcal{A}_{N,1}^{-1} \Sigma \mathcal{A}_{N,1}^{-1} \tilde{R}_m\| \leq c\bar{\Delta}_N \|G\|. \end{aligned} \quad (\text{B.32})$$

The first bounds in (B.29) and (B.30) follow from the submultiplicativity of the spectral norm and Lemma D.1, respectively. Then, we use Lemma D.2 to eliminate the off-diagonal matrix blocks, and Lemmas C.12 and C.13 are used to bound the spectral and Frobenius norms of  $\Sigma_{rr}$ . In (B.31), we first apply Lemma D.2 and then the fact that  $\|A\|_F \leq \sqrt{\text{rk}(A)}\|A\|$  for a matrix  $A$ . Note also that  $\text{rk}(AB) \leq \min\{\text{rk}(A), \text{rk}(B)\}$  for two matrices  $A, B$ . Similarly, we applied  $\|A\|_F \leq \sqrt{\text{rk}(A)}\|A\|$  in (B.32) to use (B.29). Note also that  $\|\tilde{R}_m\|^2 = \|R_m\|^2 = 1$ , where the first equality is due to the block diagonal structure of  $\tilde{R}_m$  and the second equality follows by Lemma C.4 in Sun et al. (2018).

*Imaginary part (off-diagonal):* The imaginary part of the off-diagonal elements  $\Im(\hat{H}_{rs}(D))$  can be written in terms of (B.5) as

$$\frac{1}{m} e_r' \sum_{j=1}^m t_j(D) \Im(I_X(\lambda_j)) t_j(D) e_s = \frac{1}{2\pi m} e_r' \sum_{j=1}^m t_j(D) \mathcal{X}'(C_j S_j' - S_j C_j') \mathcal{X} t_j(D) e_s.$$

Then,

$$\begin{aligned}
& \mathbb{P}\left(\sup_{D \in \Omega} |\Im(\widehat{H}_{rs}(D) - \mathbb{E}\widehat{H}_{rs}(D))| > \frac{1}{2} \|G\| \nu\right) \\
& \leq \mathbb{P}\left(\sup_{D \in \Omega} \left| \sum_{j=1}^m e'_r(t_j(D) \mathcal{X}' C_j S'_j \mathcal{X} t_j(D) - \mathbb{E}(t_j(D) \mathcal{X}' C_j S'_j \mathcal{X} t_j(D))) e_s \right| > \frac{\pi m}{2} \|G\| \nu\right) \\
& \quad + \mathbb{P}\left(\sup_{D \in \Omega} \left| \sum_{j=1}^m e'_r(t_j(D) \mathcal{X}' S_j C'_j \mathcal{X} t_j(D) - \mathbb{E}(t_j(D) \mathcal{X}' S_j C'_j \mathcal{X} t_j(D))) e_s \right| > \frac{\pi m}{2} \|G\| \nu\right).
\end{aligned} \tag{B.33}$$

We focus on the first probability term in the bound, since the second can be dealt with analogously. Define the  $2m \times 2N$  matrix

$$Q_m = \begin{pmatrix} Q_{1,m} & 0_{m \times N} \\ 0_{m \times N} & Q_{2,m} \end{pmatrix}$$

with

$$Q'_{1,m} = [C_1 : \dots : C_m], \quad Q'_{2,m} = [S_1 : \dots : S_m].$$

Furthermore, define the matrix  $\widetilde{F}(d_r, d_s) = \text{diag}(\widetilde{T}_r(d_r), \widetilde{T}_s(d_s))$  in terms of the function  $\widetilde{T}_r(d) = \text{diag}(t_{1,r}(d), \dots, t_{m,r}(d))$  and recall  $M_m$  in (E.6) to write

$$e'_r \sum_{j=1}^m t_j(D) \mathcal{X}' C_j S'_j \mathcal{X} t_j(D) e_s = \mathcal{E}' \Sigma^{\frac{1}{2}} Q'_m \widetilde{F}(d_r, d_s) M_m \widetilde{F}(d_r, d_s) Q_m \Sigma^{\frac{1}{2}} \mathcal{E} = \mathcal{E}' Q(d_r, d_s) \mathcal{E}$$

with

$$Q(d_r, d_s) = \Sigma^{\frac{1}{2}} Q'_m \widetilde{F}(d_r, d_s) M_m \widetilde{F}(d_r, d_s) Q_m \Sigma^{\frac{1}{2}}$$

and  $\Sigma^{\frac{1}{2}} \mathcal{E}$  characterized as in (B.19). In order to apply Theorem D.1, we further write

$$\begin{aligned}
& \mathbb{P}\left(\sup_{D \in \Omega} \left| \sum_{j=1}^m e'_r(t_j(D) \mathcal{X}' C_j S'_j \mathcal{X} t_j(D) - \mathbb{E}(t_j(D) \mathcal{X}' C_j S'_j \mathcal{X} t_j(D))) e_s \right| > \frac{\pi m}{2} \|G\| \nu\right) \\
& = \mathbb{P}\left(\sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} |\mathcal{E}' Q(d_r, d_s) \mathcal{E} - \mathbb{E}(\mathcal{E}' Q(d_r, d_s) \mathcal{E})| > \frac{\pi m}{2} \|G\| \nu\right).
\end{aligned} \tag{B.34}$$

Note that the matrix  $Q(d_r, d_s)$  can be rewritten as

$$Q(d_r, d_s) = \Sigma^{\frac{1}{2}} \mathcal{A}_{N,i}^{-1} Q'_m \mathcal{A}_{m,i} \widetilde{F}(d_r, d_s) M_m \widetilde{F}(d_r, d_s) \mathcal{A}_{m,i} Q_m \mathcal{A}_{N,i}^{-1} \Sigma^{\frac{1}{2}}$$

for  $i = 1, \dots, 4$  with  $\mathcal{A}_{m,1}$  and  $\mathcal{A}_{N,i}$ ,  $i = 1, \dots, 4$  defined as in (B.23). We continue to bound (B.34). As for the real parts of the off-diagonal elements, note that  $\mathcal{A}_{m,i} \widetilde{F}(d_r, d_s)' M_m \widetilde{F}(d_r, d_s) \mathcal{A}_{m,i}$  is not a diagonal matrix. However, it can be rewritten as a unitary transformation of a diagonal matrix

$$\begin{aligned}
& \mathcal{A}_{m,i} \widetilde{F}(d_r, d_s)' M_m \widetilde{F}(d_r, d_s) \mathcal{A}_{m,i} \\
& = \begin{pmatrix} I_m & 0_{m,m} \\ 0_{m,m} & I_m \end{pmatrix} \begin{pmatrix} c_{r,i,N} \widetilde{c}_{s,i,N} \widetilde{T}_r(d_r) \widetilde{T}_s(d_s) & 0_{m,m} \\ 0_{m,m} & 0_{m,m} \end{pmatrix} \begin{pmatrix} 0_{m,m} & I_m \\ I_m & 0_{m,m} \end{pmatrix},
\end{aligned} \tag{B.35}$$

where we used the same calculations as in (B.24). Following Remark D.1, Theorem D.1 is applicable. In (B.36) below, we apply Theorem D.1 with  $K = 2$  and  $R = b_r - a_r \leq 1$  to obtain

$$\mathbb{P}\left(\sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} |\mathcal{E}' Q(d_r, d_s) \mathcal{E} - \mathbb{E}(\mathcal{E}' Q(d_r, d_s) \mathcal{E})| > \frac{2\pi m}{4} \|G\| \nu\right)$$

$$\leq c_1 \exp \left( -c_2 \min \left\{ \frac{\nu m \|G\|}{\gamma^2 \mathcal{T}_1}, \frac{\nu^2 m^2 \|G\|^2}{\gamma^4 \mathcal{T}_i^2} \right\} \right) \quad (\text{B.36})$$

$$\leq c_1 \exp \left( -c_2 \min \left\{ \frac{\nu m}{\gamma^2 \bar{\Delta}_N L_{rs,1}}, \frac{\nu^2 m^2}{\gamma^4 \bar{\Delta}_N^2 L_{rs,i}^2} \right\} \right) \quad (\text{B.37})$$

for  $\nu^2 \geq \gamma^4 \mathcal{T}_i^2 / (c_2 m^2 \|G\|^2)$  and  $i = 2, \dots, 5$  with

$$\mathcal{T}_1 = \|C_{m,1}\| L_{rs,1}, \quad \mathcal{T}_i = \|C_{m,i}\|_F L_{rs,i}, \quad \mathcal{T}_5 = \|C_{m,1}\| L_{rs,5} \quad (\text{B.38})$$

for  $i = 2, \dots, 4$ , where the  $L_{rs,i}$ 's are given in (B.1) and

$$C_{m,i} = Q_m \mathcal{A}_{N,i}^{-1} \Sigma \mathcal{A}_{N,i}^{-1} Q'_m \quad (\text{B.39})$$

for  $i = 1, \dots, 4$ . Given (B.35), the  $L_{rs,i}$ 's in (B.38) can indeed be represented as in (B.1).

Since the quantities (B.39) are equal to those in (B.28) by replacing  $Q_m$  with  $\tilde{R}_m$ , the norms  $\|C_{m,1}\|$ ,  $\|C_{m,2}\|_F$ ,  $\|C_{m,3}\|_F$  and  $\|C_{m,4}\|_F$  can be bounded as the norms of  $B_{m,i}$ ,  $i = 1, \dots, 4$ , in (B.29)–(B.32). In order to deal with  $Q_m$ , note that  $\|Q_m\|^2 = \max\{\|Q_{1,m}\|^2, \|Q_{2,m}\|^2\} \leq 1$ , where the equality is due to the block diagonal structure of  $Q_m$  and the inequality follows by Lemma C.4 in Sun et al. (2018).

In order to get the statements (E.1)–(E.2) of the lemma, we shall get an upper bound on the probability in (B.7). For the diagonal terms ( $r = s$ ), the imaginary part is zero. As proved in (B.12), for  $r = s$ , (B.7) can be bounded by  $\mathcal{B}(r, r, i)$  for  $i = 2, 4, 5$ ; see (B.14). For the off-diagonal elements, one needs to get a bound on the two probabilities in (B.7). For the real part of the off-diagonal elements, (B.26) gives the upper bound  $\mathcal{B}(r, s, i)$ , for  $i = 2, \dots, 5$ . For the imaginary parts of the off-diagonal elements, which was further bounded in (B.33), (B.37) provides the upper bound  $\mathcal{B}(r, s, i)$ , for  $i = 2, \dots, 5$ . Combining the different bounds leads to the desired result.  $\square$

The next lemma gives a bound on the deterministic term in (3.12).

**Lemma B.2.** *Suppose Assumptions 1–3. Then, the deterministic term in (3.12) can be bounded as*

$$\begin{aligned} \sup_{D \in \Omega} |\mathbb{E} \hat{H}_{rs}(D) - \tilde{H}_{rs}(D)| &\leq \sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} \left( c_{G,1} \frac{1}{2\pi m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \lambda_j^{2q-d_0, r-d_0, s} \right. \\ &\quad \left. + \frac{1}{2\pi m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) j^{-1} \lambda_j^{-d_0, r-d_0, s} Q_m \right) \end{aligned} \quad (\text{B.40})$$

with

$$Q_m = \|G\| \frac{72(\cos(\lambda_m/2))^{-2}}{\pi(1 + 2 \min\{\Delta_1, -\Delta_2\})} + c_{G,2} 4(2 + \log(m)). \quad (\text{B.41})$$

*Proof:* The deviation of the expected value around the population quantity can be bounded as

$$\begin{aligned} &|\tilde{H}_{rs}(D) - \mathbb{E} \hat{H}_{rs}(D)| \\ &= \left| \frac{1}{m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \lambda_j^{-d_0, r-d_0, s} G_{0,rs} - \mathbb{E} \left( \frac{1}{m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) I_{X,rs}(\lambda_j) \right) \right| \\ &\leq \left| \frac{1}{m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \lambda_j^{-d_0, r-d_0, s} G_{0,rs} - \frac{1}{m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \lambda_j^{-d_0, r-d_0, s} G_{0,rs}(\lambda_j) \right| \end{aligned}$$

$$\begin{aligned}
& + \left| \frac{1}{m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \lambda_j^{-d_{0,r}-d_{0,s}} G_{0,rs}(\lambda_j) - \mathbb{E} \left( \frac{1}{m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) I_{X,rs}(\lambda_j) \right) \right| \\
& = \left| \frac{1}{m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \lambda_j^{-d_{0,r}-d_{0,s}} (G_{0,rs} - G_{0,rs}(\lambda_j)) \right| \\
& \quad + \left| \frac{1}{m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \lambda_j^{-d_{0,r}-d_{0,s}} G_{0,rs}(\lambda_j) - \frac{1}{m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \mathbb{E}(I_{X,rs}(\lambda_j)) \right| \\
& \leq \left| \frac{1}{m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \lambda_j^{-d_{0,r}-d_{0,s}} (G_{0,rs} - G_{0,rs}(\lambda_j)) \right| \\
& \quad + \left| \frac{1}{m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \left( \lambda_j^{-d_{0,r}-d_{0,s}} G_{0,rs}(\lambda_j) - \mathbb{E}(I_{X,rs}(\lambda_j)) \right) \right| \tag{B.42}
\end{aligned}$$

$$\leq c_{G,1} \frac{1}{2\pi m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \lambda_j^{2q-d_{0,r}-d_{0,s}} + \frac{1}{2\pi m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) j^{-1} \lambda_j^{-d_{0,r}-d_{0,s}} Q_m, \tag{B.43}$$

where  $Q_m$  is in (B.41). We consider the two different summands in (B.42) separately to prove (B.43). For the first summand in (B.42), the upper bound is a consequence of Assumption 2.

The second summand in (B.42) can be represented and bounded as

$$\begin{aligned}
& \left| \frac{1}{2\pi m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \left( \lambda_j^{-d_{0,r}-d_{0,s}} G_{0,rs}(\lambda_j) - \mathbb{E}(I_{X,rs}(\lambda_j)) \right) \right| \\
& \leq \frac{1}{2\pi m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) \left| f_{rs}(\lambda_j) - \mathbb{E}(I_{X,rs}(\lambda_j)) \right| \\
& \leq \frac{1}{2\pi m} \sum_{j=1}^m t_{j,r}(d_r) t_{j,s}(d_s) N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} Q_m, \tag{B.44}
\end{aligned}$$

with  $Q_m$  in (B.41) and the last inequality (B.44) is a consequence of Lemma C.8 given Assumptions 1 and 3.  $\square$

We next prove Propositions 3.1–3.4. The proofs are all consequences of Lemmas B.1 and B.2, and structured in the same way. We first choose a function  $t_j(D)$  in (3.8)–(3.9) and then apply Lemma B.1 to the respective probabilistic part and Lemma B.2 to the respective deterministic part in the bound (3.12).

For the probabilistic parts in the proofs of Propositions 3.1–3.4, we note that Lemma B.1 requires  $\tilde{\nu}^2 \geq \gamma^4 L_{rs,i}^2 / (m^2 c_2)$ ,  $i = 2, \dots, 5$  with  $L_{rs,i}$  as in (B.1). We do not verify this condition in the proofs since it is automatically satisfied by the bounds we get on the  $L_{rs,i}$ 's and under the assumptions in Propositions 3.1–3.4. This is due to our choices of  $\nu$ ,  $\nu_1$ ,  $\nu_2$  and  $\nu_3$  in (3.14), (3.15), (3.16) and (3.17), which are always of the form  $\nu = \tilde{\nu} + \mathcal{T}$ , where  $\mathcal{T}$  accounts for the respective bias terms and  $\tilde{\nu}$  is always chosen as  $m^{-1} L_{rs,i} \lesssim \tilde{\nu}$ .

*Proof of Proposition 3.1:* To apply Lemmas B.1 and B.2, we take

$$t_j(D) = \lambda_j^D \quad \text{and} \quad \Omega = \Omega(\varepsilon) = \{D \in \mathcal{M}_{\text{diag}} \mid \Delta_1 I_p \preceq D \preceq \Delta_2 I_p \text{ and } \|D - D_0\|_{\max} \leq \varepsilon\} \tag{B.45}$$

in (3.8)–(3.9). We consider the probabilistic and deterministic parts separately as in (3.12).

*Probabilistic part:* We distinguish three cases depending on whether the true memory parameters  $d_{0,r}, d_{0,s}$  are positive or negative. Therefore, we write

$$\mathbb{P}\left(\sup_{D \in \Omega(\varepsilon)} |\widehat{G}_{rs}(D) - \mathbb{E}(\widehat{G}_{rs}(D))| > \nu\right) \leq \sum_{k=1}^3 \mathbb{P}\left(\sup_{D \in \Omega(\varepsilon)} |\widehat{G}_{rs}(D) - \mathbb{E}(\widehat{G}_{rs}(D))| \mathbf{1}_{A_k} > \nu\right), \quad (\text{B.46})$$

where

$$A_1 = \{d_{0,r} \leq 0, d_{0,s} \leq 0\}, \quad A_2 = \{d_{0,r} > 0, d_{0,s} > 0\}, \quad A_3 = \{d_{0,r} > 0, d_{0,s} \leq 0\}$$

and

$$\Omega(\varepsilon) = ([d_{0,1} - \varepsilon, d_{0,1} + \varepsilon] \cap [\Delta_1, \Delta_2]) \times \cdots \times ([d_{0,p} - \varepsilon, d_{0,p} + \varepsilon] \cap [\Delta_1, \Delta_2]).$$

For each case, we apply Lemma B.1 and bound the respective quantities  $L_{rs,i}$ ,  $i = 1, \dots, 5$  in (B.1) that then yield the desired result. We will show that, for  $k = 1, 2, 3$ ,

$$\mathbb{P}\left(\sup_{D \in \Omega(\varepsilon)} |\widehat{G}_{rs}(D) - \mathbb{E}(\widehat{G}_{rs}(D))| \mathbf{1}_{A_k} > \|G\| \tilde{\nu}\right) \leq c_1 \exp\left(-c_2 \min\left\{\frac{\tilde{\nu}m}{\gamma^2 \mathcal{L}_{k1}}, \frac{\tilde{\nu}^2 m^2}{\gamma^4 \mathcal{L}_{k2}}\right\}\right). \quad (\text{B.47})$$

*Case  $d_{0,r} \leq 0, d_{0,s} \leq 0$ :* We get

$$\begin{aligned} L_{rs,1} &\leq \max_{j=1, \dots, m} |\lambda_j^{2\Delta_1}| + \sup_{(d_r, d_s) \in [\Delta_1, \Delta_2]^2} \max_{j=1, \dots, m} \sqrt{2} |\log(\lambda_j) \lambda_j^{d_r + d_s}| \\ &\leq c(1 + \log(N)) N^{-2\Delta_1} \leq c \log(N) N^{-2\Delta_1} =: \mathcal{L}_{11}, \end{aligned} \quad (\text{B.48})$$

$$\begin{aligned} L_{rs,5} &\leq \left(\sum_{j=1}^m \lambda_j^{4\Delta_1}\right)^{\frac{1}{2}} + \sup_{(d_r, d_s) \in [\Delta_1, \Delta_2]^2} \sqrt{2} \left(\sum_{j=1}^m \log(\lambda_j)^2 \lambda_j^{2d_r + 2d_s}\right)^{\frac{1}{2}} \\ &\leq c(1 + \log(N)) \left(\sum_{j=1}^m \lambda_j^{4\Delta_1}\right)^{\frac{1}{2}} \leq c \log(N) \left(\sum_{j=1}^m \lambda_j^{4\Delta_1}\right)^{\frac{1}{2}} =: \mathcal{L}_{12}. \end{aligned} \quad (\text{B.49})$$

Since  $\mathcal{L}_{11}$  and  $\mathcal{L}_{21}$  do not depend on  $r, s$ , applying Lemma B.1 with  $i = 5$  in (E.2) gives (B.47) with  $k = 1$ .

*Case  $d_{0,r} > 0, d_{0,s} > 0$ :* We get

$$\begin{aligned} L_{rs,1} &= N^{d_{0,r} + d_{0,s}} \left( \max_{j=1, \dots, m} |\lambda_j^{d_{0,r} + d_{0,s} - 2\varepsilon}| + \sup_{D \in \Omega(\varepsilon)} \max_{j=1, \dots, m} \sqrt{2} |\log(\lambda_j) \lambda_j^{d_r + d_s}| \right) \\ &\leq c(1 + \log(N)) N^{d_{0,r} + d_{0,s}} (\lambda_m^{d_{0,r} + d_{0,s} - 2\varepsilon} \mathbf{1}_{\{d_{0,r} + d_{0,s} - 2\varepsilon \geq 0\}} + \lambda_1^{d_{0,r} + d_{0,s} - 2\varepsilon} \mathbf{1}_{\{d_{0,r} + d_{0,s} - 2\varepsilon < 0\}}) \end{aligned} \quad (\text{B.50})$$

$$\begin{aligned} &\leq c \log(N) \max\{m^{d_{0,r} + d_{0,s} - 2\varepsilon} N^{2\varepsilon}, N^{2\varepsilon}\} \\ &\leq c \log(N) m^{2\Delta_2 - 2\varepsilon} N^{2\varepsilon} =: \mathcal{L}_{21}. \end{aligned} \quad (\text{B.51})$$

We pause here to draw the attention to inequality (B.50), since the argument will be used not only here but also in the proofs of Proposition 3.2–3.4. It can be assumed that the frequencies satisfy  $\lambda_j \leq 1$ . Then, the function  $f(d) = \lambda_j^{2d}$  is monotonically decreasing in  $d$  for all  $j = 1, \dots, m$ . For a non-negative exponent  $d > 0$ ,  $f$  reaches its maximum for  $j = m$ , for a negative exponent  $d \leq 0$ , for  $j = 1$ . Due to the monotonicity of  $f$ , the function reaches its supremum for the smallest possible values  $d$  can take.

We proceed with the case  $d_{0,r} > 0, d_{0,s} > 0$ ,

$$\begin{aligned}
L_{rs,i} &= c_{r,i,N} \tilde{c}_{s,i,N} \left( \max_{j=1,\dots,m} |\lambda_j^{d_{0,r}+d_{0,s}-2\varepsilon}| + \sup_{D \in \Omega(\varepsilon)} \max_{j=1,\dots,m} \sqrt{2} |\log(\lambda_j) \lambda_j^{d_r+d_s}| \right) \\
&\leq c(1 + \log(N)) c_{r,i,N} \tilde{c}_{s,i,N} (\lambda_m^{d_{0,r}+d_{0,s}-2\varepsilon} \mathbb{1}_{\{d_{0,r}+d_{0,s}-2\varepsilon \geq 0\}} + \lambda_1^{d_{0,r}+d_{0,s}-2\varepsilon} \mathbb{1}_{\{d_{0,r}+d_{0,s}-2\varepsilon < 0\}}) \\
&\leq c \log(N) \begin{cases} \max\{m^{d_{0,r}+d_{0,s}-2\varepsilon} N^{2\varepsilon}, N^{2\varepsilon}\}, & \text{if } d_{0,r}, d_{0,s} > \frac{1}{4}, i = 2, \\ \max\{m^{\frac{1}{4}+d_{0,r}+d_{0,s}-2\varepsilon} N^{2\varepsilon}, m^{\frac{1}{4}} N^{2\varepsilon}\}, & \text{if } d_{0,r} > \frac{1}{4}, d_{0,s} \leq \frac{1}{4}, i = 3, \\ m^{\frac{1}{2}} N^{d_{0,r}+d_{0,s}} \lambda_m^{d_{0,r}+d_{0,s}-2\varepsilon}, & \text{if } d_{0,r}, d_{0,s} \leq \frac{1}{4}, d_{0,r} + d_{0,s} - 2\varepsilon \geq 0, i = 4, \\ m^{\frac{1}{2}} N^{d_{0,r}+d_{0,s}} \lambda_1^{d_{0,r}+d_{0,s}-2\varepsilon}, & \text{if } d_{0,r}, d_{0,s} \leq \frac{1}{4}, d_{0,r} + d_{0,s} - 2\varepsilon < 0, i = 4, \end{cases} \\
&\leq c \log(N) \begin{cases} m^{2\Delta_2-2\varepsilon} N^{2\varepsilon}, & \text{if } d_{0,r}, d_{0,s} > \frac{1}{4}, i = 2, \\ m^{\frac{1}{4}+\Delta_2+d_{0,s}-2\varepsilon} N^{2\varepsilon}, & \text{if } d_{0,r} > \frac{1}{4}, d_{0,s} \leq \frac{1}{4}, i = 3, \\ m^{\frac{1}{2}+d_{0,r}+d_{0,s}-2\varepsilon} N^{2\varepsilon}, & \text{if } d_{0,r}, d_{0,s} \leq \frac{1}{4}, d_{0,r} + d_{0,s} - 2\varepsilon \geq 0, i = 4, \\ m^{\frac{1}{2}} N^{2\varepsilon}, & \text{if } d_{0,r}, d_{0,s} \leq \frac{1}{4}, d_{0,r} + d_{0,s} - 2\varepsilon < 0, i = 4, \end{cases} \\
&\leq c \log(N) \max\{m^{2\bar{\Delta}_u-2\varepsilon} N^{2\varepsilon}, m^{\frac{1}{2}} N^{2\varepsilon}\} =: \mathcal{L}_{22} \tag{B.52}
\end{aligned}$$

with  $\bar{\Delta}_u$  as in (A.2). Since  $\mathcal{L}_{21}$  and  $\mathcal{L}_{22}$  do not depend on  $r, s$ , applying Lemma B.1 with  $i = 2, 3, 4$ , gives (B.47) with  $k = 2$ .

*Case  $d_{0,r} > 0, d_{0,s} \leq 0$ :* We get

$$\begin{aligned}
L_{rs,1} &= N^{d_{0,r}} \left( \max_{j=1,\dots,m} |\lambda_j^{d_{0,r}-\varepsilon+\max\{d_{0,s}-\varepsilon, \Delta_1\}}| + \sup_{D \in \Omega(\varepsilon)} \max_{j=1,\dots,m} \sqrt{2} |\log(\lambda_j) \lambda_j^{d_r+d_s}| \right) \\
&\leq N^{d_{0,r}} \max_{j=1,\dots,m} |\lambda_j^{d_{0,r}-\varepsilon+\max\{d_{0,s}-\varepsilon, \Delta_1\}}| \\
&\quad + c \log(N) N^{d_{0,r}} (\lambda_m^{d_{0,r}-\varepsilon+\Delta_1} \mathbb{1}_{\{d_{0,r}-\varepsilon+\Delta_1 \geq 0\}} + \lambda_1^{d_{0,r}-\varepsilon+\Delta_1} \mathbb{1}_{\{d_{0,r}-\varepsilon+\Delta_1 < 0\}}) \\
&\leq c(1 + \log(N)) (m^{\Delta_2+\Delta_1-\varepsilon} N^{-\Delta_1+\varepsilon} \mathbb{1}_{\{d_{0,r}-\varepsilon+\Delta_1 \geq 0\}} + N^{-\Delta_1+\varepsilon} \mathbb{1}_{\{d_{0,r}-\varepsilon+\Delta_1 < 0\}}) \\
&\leq c \log(N) \max\{m^{\Delta_2+\Delta_1-\varepsilon}, 1\} N^{-\Delta_1+\varepsilon} =: \mathcal{L}_{31} \tag{B.53}
\end{aligned}$$

since  $\varepsilon < -\Delta_1$ , and

$$\begin{aligned}
L_{rs,i} &= c_{r,i,N} \tilde{c}_{s,i,N} \left( \max_{j=1,\dots,m} |\lambda_j^{d_{0,r}-\varepsilon+\max\{d_{0,s}-\varepsilon, \Delta_1\}}| + \sup_{D \in \Omega(\varepsilon)} \max_{j=1,\dots,m} \sqrt{2} |\log(\lambda_j) \lambda_j^{d_r+d_s}| \right) \\
&\leq c(1 + \log(N)) c_{r,i,N} \tilde{c}_{s,i,N} (\lambda_m^{d_{0,r}-\varepsilon+\Delta_1} \mathbb{1}_{\{d_{0,r}-\varepsilon+\Delta_1 \geq 0\}} + \lambda_1^{d_{0,r}-\varepsilon+\Delta_1} \mathbb{1}_{\{d_{0,r}-\varepsilon+\Delta_1 < 0\}}) \\
&\leq c \log(N) \begin{cases} m^{\frac{1}{4}} N^{d_{0,r}} \max\{m^{d_{0,r}+\Delta_1-\varepsilon}, 1\} N^{-d_{0,r}-\Delta_1+\varepsilon}, & \text{if } d_{0,r} > \frac{1}{4}, d_{0,s} \leq 0, i = 3, \\ m^{\frac{1}{2}} N^{d_{0,r}} \max\{m^{d_{0,r}+\Delta_1-\varepsilon}, 1\} N^{-d_{0,r}-\Delta_1+\varepsilon}, & \text{if } \frac{1}{4} \geq d_{0,r} > 0, d_{0,s} \leq 0, i = 4, \end{cases} \\
&\leq c \log(N) \begin{cases} \max\{m^{\frac{1}{4}+\Delta_2+\Delta_1-\varepsilon}, m^{\frac{1}{4}}\} N^{-\Delta_1+\varepsilon}, & \text{if } d_{0,r} > \frac{1}{4}, d_{0,s} \leq 0, i = 3, \\ \max\{m^{\frac{1}{2}+d_{0,r}+\Delta_1-\varepsilon}, m^{\frac{1}{2}}\} N^{-\Delta_1+\varepsilon}, & \text{if } \frac{1}{4} \geq d_{0,r} > 0, d_{0,s} \leq 0, i = 4, \end{cases} \\
&\leq c \log(N) \max\{m^{\frac{1}{4}+\bar{\Delta}_u+\Delta_1-\varepsilon}, m^{\frac{1}{2}}\} N^{-\Delta_1+\varepsilon} =: \mathcal{L}_{32}. \tag{B.54}
\end{aligned}$$

Applying Lemma B.1 with  $i = 3, 4$  gives (B.47) with  $k = 3$ .

*Deterministic part:* Using Lemma B.2 with (B.45), we get

$$\sup_{D \in \Omega(\varepsilon)} |\mathbb{E} \widehat{G}_{rs}(D) - \widetilde{G}_{rs}(D)| \leq c_{G,1} \frac{1}{2\pi} \lambda_m^{2q-2\varepsilon} + c \frac{1}{2\pi m} \left(1 + \frac{1}{2\varepsilon}\right) N^{2\varepsilon} Q_m = \mathcal{T}_1(\varepsilon), \tag{B.55}$$

where  $\mathcal{T}_1(\varepsilon)$  is in Table A.2 and the inequality can be obtained by bounding the two terms in (B.40) given (B.45) as follows. The first summand in (B.40) can be bounded as

$$\sup_{D \in \Omega(\varepsilon)} \frac{1}{2\pi m} \sum_{j=1}^m \lambda_j^{d_r+d_s} \lambda_j^{2q-d_0, r-d_0, s} \leq \frac{1}{2\pi m} \sum_{j=1}^m \lambda_j^{2q-2\varepsilon} \leq \frac{1}{2\pi} \lambda_m^{2q-2\varepsilon}$$

for  $q - \varepsilon > 0$ , and the second summand as

$$\begin{aligned} \sup_{D \in \Omega(\varepsilon)} \frac{1}{2\pi m} \sum_{j=1}^m \lambda_j^{d_r+d_s} j^{-1} \lambda_j^{-d_0, r-d_0, s} &\leq \frac{1}{2\pi m} \sum_{j=1}^m j^{-1} \lambda_j^{-2\varepsilon} \\ &\leq \frac{1}{2\pi m} \sum_{j=1}^m j^{-1-2\varepsilon} \lambda_1^{-2\varepsilon} \leq \frac{1}{2\pi m} \left(1 + \frac{1}{2\varepsilon}\right) \lambda_1^{-2\varepsilon}. \end{aligned}$$

Finally, we combine our results on the probabilistic and deterministic terms to obtain the statement of the proposition. With our choice of  $\nu$  in (3.14) and for any  $\mathbf{C} \geq 1$ , observe that, with explanations given below,

$$\begin{aligned} &\mathbb{P}\left(\sup_{D \in \Omega(\varepsilon)} |\widehat{G}_{rs}(D) - \widetilde{G}_{rs}(D)| > \nu\right) \\ &\leq \mathbb{P}\left(\sup_{D \in \Omega(\varepsilon)} |\widehat{G}_{rs}(D) - \mathbb{E}(\widehat{G}_{rs}(D))| + \sup_{D \in \Omega(\varepsilon)} |\mathbb{E}(\widehat{G}_{rs}(D)) - \widetilde{G}_{rs}(D)| > \mathbf{C} \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_1}} + \mathcal{T}_1(\varepsilon)\right) \\ &\leq \mathbb{P}\left(\sup_{D \in \Omega(\varepsilon)} |\widehat{G}_{rs}(D) - \mathbb{E}(\widehat{G}_{rs}(D))| > \mathbf{C} \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_1}}\right) \end{aligned} \quad (\text{B.56})$$

$$\leq \sum_{k=1}^3 \mathbb{P}\left(\sup_{D \in \Omega(\varepsilon)} |\widehat{G}_{rs}(D) - \mathbb{E}(\widetilde{G}_{rs}(D))| \mathbf{1}_{A_k} > \mathbf{C} \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_1}}\right) \quad (\text{B.57})$$

$$\leq \sum_{k=1}^3 c_1 \exp\left(-c_2 \mathbf{C} \min\left\{\frac{\sqrt{\frac{\log(p)}{\log(N)^{-1} \bar{\Delta}_N^{-1} \mathcal{R}_{11}}} m}{\gamma^2 \bar{\Delta}_N \mathcal{L}_{k1}}, \frac{\frac{\log(p)}{\log(N)^{-2} \bar{\Delta}_N^{-2} \mathcal{R}_{12}} m^2}{\gamma^4 \bar{\Delta}_N^2 \mathcal{L}_{k2}^2}\right\}\right) \quad (\text{B.58})$$

$$\leq c_1 \exp\left(-c_2 \mathbf{C} \min\left\{\frac{\sqrt{\frac{\log(p)}{\log(N)^{-1} \bar{\Delta}_N^{-1} \mathcal{R}_{11}}} m}{\gamma^2 \bar{\Delta}_N \max_{k=1,2,3} \mathcal{L}_{k1}}, \frac{\frac{\log(p)}{\log(N)^{-2} \bar{\Delta}_N^{-2} \mathcal{R}_{12}} m^2}{\gamma^4 \bar{\Delta}_N^2 \max_{k=1,2,3} \mathcal{L}_{k2}^2}\right\}\right)$$

$$\leq c_1 \exp\left(-c_2 \mathbf{C} \min\left\{\sqrt{\frac{\log(p)}{\bar{\Delta}_N \log(N)} \mathcal{R}_{11}}, \log(p)\right\}\right) \quad (\text{B.59})$$

$$\leq c_1 p^{-c_2 \mathbf{C}} \quad (\text{B.60})$$

with  $\mathcal{L}_{k1}, \mathcal{L}_{k2}, k = 1, 2, 3$  defined in (B.48), (B.49), (B.51), (B.52) and (B.53), (B.54). The constants  $c_1, c_2$  are generic and might differ from line to line. Indeed, the inequality (B.56) is due to (B.55) with  $\mathcal{T}_1(\varepsilon)$  as in Table A.2. In (B.57), the probabilistic part is bounded as in (B.46). Applying (B.47) yields (B.58). For the inequality (B.59), note that  $\mathcal{R}_{11}$  and  $\mathcal{R}_{12}$  are chosen in Table A.1 such that

$$\log(N)^{-1} \mathcal{R}_{11} \lesssim m \left(\max_{k=1,2,3} \mathcal{L}_{k1}\right)^{-1} \quad \text{and} \quad \log(N)^{-2} \mathcal{R}_{12} \lesssim m^2 \left(\max_{k=1,2,3} \mathcal{L}_{k2}\right)^{-2}.$$

The inequality (B.60) follows since we work under the assumption  $\mathcal{R}_{11} \gtrsim \bar{\Delta}_N \log(N) \log(p)$ ; see the discussion following Proposition 3.1.  $\square$

*Proof of Proposition 3.2:* To apply Lemmas B.1 and B.2, we choose

$$t_{j,r}(d_{0,r}) = \lambda_j^{d_{0,r}} \quad (\text{B.61})$$

in (3.8)–(3.9). Here, we do not need a uniform bound. For this reason, it is not necessary to take the supremum over all admissible estimates of  $D$  in Lemmas B.1 and B.2. In particular, Lemma B.1 simplifies, since the derivatives in the respective bounds  $L_{rr,i}$  in (B.1) become zero with the choice (B.61).

*Probabilistic part:* It is enough to distinguish two cases

$$\mathbb{P}\left(|\widehat{g}_r(d_{0,r}) - \mathbb{E}(\widehat{g}_r(d_{0,r}))| > \|G\|\|\tilde{\nu}_1\right) \leq \sum_{k=1}^2 \mathbb{P}\left(|\widehat{g}_r(d_{0,r}) - \mathbb{E}(\widehat{g}_r(d_{0,r}))|\mathbb{1}_{A_k} > \|G\|\|\tilde{\nu}_1\right) \quad (\text{B.62})$$

with  $A_1 = \{d_{0,r} \leq 0\}$  and  $A_2 = \{d_{0,r} > 0\}$ . For both cases, we apply Lemma B.1 and bound the respective quantities  $L_{rr,i}$ ,  $i = 1, \dots, 5$  in (B.1) that then yield the desired result. We will show that, for  $k = 1, 2$ ,

$$\mathbb{P}\left(|\widehat{g}_r(d_{0,r}) - \mathbb{E}(\widehat{g}_r(d_{0,r}))|\mathbb{1}_{A_k} > \|G\|\|\tilde{\nu}_1\right) \leq c_1 \exp\left(-c_2 \min\left\{\frac{\tilde{\nu}_1 m}{\gamma^2 \mathcal{L}_{k1}}, \frac{\tilde{\nu}_1^2 m^2}{\gamma^4 \mathcal{L}_{k2}^2}\right\}\right). \quad (\text{B.63})$$

*Case  $d_{0,r} \leq 0$ :* We get

$$\begin{aligned} L_{rr,1} &= N^{\max\{2d_{0,r}, 0\}} \max_{j=1, \dots, m} \lambda_j^{2d_{0,r}} \leq cN^{-2\Delta_1} =: \mathcal{L}_{11}, \\ L_{rr,5} &= \left(\sum_{j=1}^m \lambda_j^{4d_{0,r}}\right)^{\frac{1}{2}} \leq \left(\sum_{j=1}^m \lambda_j^{4\Delta_1}\right)^{\frac{1}{2}} =: \mathcal{L}_{12}. \end{aligned} \quad (\text{B.64})$$

Then, applying Lemma B.1 with  $i = 5$  gives (B.63) with  $k = 1$ .

*Case  $d_{0,r} > 0$ :* We get

$$\begin{aligned} L_{rr,1} &= N^{\max\{2d_{0,r}, 0\}} \max_{j=1, \dots, m} \lambda_j^{2d_{0,r}} \leq cm^{2\Delta_2} =: \mathcal{L}_{21}, \\ L_{rr,i} &= c_{r,1,N}^2 \lambda_m^{2d_{0,r}} \\ &= \begin{cases} N^{\max\{2d_{0,r}, \frac{1}{2}\}} \lambda_m^{2d_{0,r}}, & \text{if } d_{0,r} > \frac{1}{4}, i = 2, \\ m^{\frac{1}{2}} N^{\max\{2d_{0,r}, 0\}} \lambda_m^{2d_{0,r}}, & \text{if } d_{0,r} \leq \frac{1}{4}, i = 4, \end{cases} \\ &\leq c \begin{cases} m^{2d_{0,r}}, & \text{if } d_{0,r} > \frac{1}{4}, i = 2, \\ m^{\frac{1}{2}+2d_{0,r}}, & \text{if } d_{0,r} \leq \frac{1}{4}, i = 4, \end{cases} \\ &\leq cm^{2\bar{\Delta}_u} =: \mathcal{L}_{22}. \end{aligned} \quad (\text{B.66})$$

Then, applying Lemma B.1 with  $i = 2$  gives (B.63) with  $k = 2$ .

*Deterministic part:* Using Lemma B.2 with (B.61), we get

$$|\mathbb{E}(\widehat{g}_r(d_{0,r})) - g_{0,r}| \leq c_{G,1} \frac{1}{2\pi} \lambda_m^q + \frac{1}{\pi m} \log(m) Q_m = \mathcal{T}_2, \quad (\text{B.67})$$

where  $\mathcal{T}_2$  is in Table A.2 and the inequality can be obtained by bounding the two terms in (B.40) given (B.61) as follows. The first summand in (B.40) can be bounded as

$$\frac{1}{2\pi m} \sum_{j=1}^m \lambda_j^{2d_{0,r}} \lambda_j^{2q-2d_{0,r}} = \frac{1}{2\pi m} \sum_{j=1}^m \lambda_j^{2q} \leq \frac{1}{2\pi} \lambda_m^{2q},$$



the second summand as

$$\frac{1}{2\pi m} \sum_{j=1}^m \lambda_j^{2d_{0,r}} j^{-1} \lambda_j^{-2d_{0,r}} = \frac{1}{2\pi m} \sum_{j=1}^m j^{-1} \leq \frac{1}{2\pi m} (1 + \log(m)) \leq \frac{1}{\pi m} \log(m).$$

Finally, we combine our results on the probabilistic and deterministic terms. With our choice of  $\nu_1$  in (3.15) and for any  $\mathbf{C} \geq 1$ , observe that, with explanations given below,

$$\begin{aligned} & \mathbb{P}\left(|\widehat{g}_r(d_{0,r}) - g_{0,r}| > \nu_1\right) \\ & \leq \mathbb{P}\left(|\widehat{g}_r(d_{0,r}) - \mathbb{E}(\widehat{g}_r(d_{0,r}))| + |\mathbb{E}(\widehat{g}_r(d_{0,r})) - g_{0,r}| > \mathbf{C} \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_2}} + \mathcal{T}_2\right) \\ & \leq \mathbb{P}\left(|\widehat{g}_r(d_{0,r}) - \mathbb{E}(\widehat{g}_r(d_{0,r}))| > \mathbf{C} \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_2}}\right) \end{aligned} \quad (\text{B.68})$$

$$\leq \sum_{k=1}^2 \mathbb{P}\left(|\widehat{g}_r(d_{0,r}) - \mathbb{E}(\widehat{g}_r(d_{0,r}))| \mathbb{1}_{A_k} > \mathbf{C} \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_2}}\right) \quad (\text{B.69})$$

$$\leq \sum_{k=1}^2 c_1 \exp\left(-c_2 \mathbf{C} \min\left\{\frac{\sqrt{\frac{\log(p)}{\Delta_N^{-1} \mathcal{R}_{21}}} m}{\gamma^2 \bar{\Delta}_N \mathcal{L}_{k1}}, \frac{\frac{\log(p)}{\Delta_N^{-2} \mathcal{R}_{22}} m^2}{\gamma^4 \bar{\Delta}_N^2 \mathcal{L}_{k2}^2}\right\}\right) \quad (\text{B.70})$$

$$\begin{aligned} & \leq c_1 \exp\left(-c_2 \mathbf{C} \min\left\{\frac{\sqrt{\frac{\log(p)}{\Delta_N^{-1} \mathcal{R}_{21}}} m}{\gamma^2 \bar{\Delta}_N \max_{k=1,2} \mathcal{L}_{k1}}, \frac{\frac{\log(p)}{\Delta_N^{-2} \mathcal{R}_{22}} m^2}{\gamma^4 \bar{\Delta}_N^2 \max_{k=1,2} \mathcal{L}_{k2}^2}\right\}\right) \\ & \leq c_1 \exp\left(-c_2 \mathbf{C} \min\left\{\sqrt{\frac{\log(p)}{\Delta_N} \mathcal{R}_{21}}, \log(p)\right\}\right) \end{aligned} \quad (\text{B.71})$$

$$\leq c_1 p^{-c_2 \mathbf{C}} \quad (\text{B.72})$$

with  $\mathcal{L}_{k1}, \mathcal{L}_{k2}$ ,  $k = 1, 2$  defined in (B.64), (B.65) and (B.66). The constants  $c_1, c_2$  are generic and might differ from line to line. Indeed, the inequality (B.68) is due to (B.67). In (B.69), the probabilistic part is bounded as in (B.62). Applying (B.63) yields (B.70). For the inequality (B.71), note that  $\mathcal{R}_{21}$  and  $\mathcal{R}_{22}$  are chosen in Table A.1 such that

$$\mathcal{R}_{21} \lesssim m(\max_{k=1,2} \mathcal{L}_{k1})^{-1} \quad \text{and} \quad \mathcal{R}_{22} \lesssim m^2(\max_{k=1,2} \mathcal{L}_{k2})^{-2}.$$

The inequality (B.72) follows since we work under the assumption  $\mathcal{R}_{21} \gtrsim \bar{\Delta}_N \log(p)$ ; see the discussion following Proposition 3.2.  $\square$

*Proof of Proposition 3.3:* To apply Lemma B.1, we choose

$$t_{j,r}(d) = \left(\frac{j}{m}\right)^{d-d_{0,r}} \lambda_j^{d_{0,r}} = \left(\frac{j}{m}\right)^d \lambda_m^{d_{0,r}} \quad (\text{B.73})$$

and

$$\Omega = \Theta_1 = \begin{cases} \{d \mid d_{0,r} - \frac{1}{2} + \Delta \leq d \leq \Delta_2\}, & \text{if } d_{0,r} \geq \Delta_1 + \frac{1}{2}, \\ \{d \mid \Delta_1 \leq d \leq \Delta_2\}, & \text{if } d_{0,r} < \Delta_1 + \frac{1}{2} \end{cases}$$

in (3.8)–(3.9).

*Probabilistic part:* It is enough to distinguish three cases

$$\begin{aligned} & \mathbb{P}\left(\sup_{d_r \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d_r - 2d_{0,r}} \lambda_j^{2d_{0,r}} (I_{X,rr}(\lambda_j) - \mathbb{E}(I_{X,rr}(\lambda_j))) \right| > \|G\| \tilde{\nu}_2\right) \\ & \leq \sum_{k=1}^3 \mathbb{P}\left(\sup_{d_r \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d_r - 2d_{0,r}} \lambda_j^{2d_{0,r}} (I_{X,rr}(\lambda_j) - \mathbb{E}(I_{X,rr}(\lambda_j))) \right| \mathbf{1}_{A_k} > \|G\| \tilde{\nu}_2\right) \end{aligned} \quad (\text{B.74})$$

with

$$\begin{aligned} A_1 &= \{d_{0,r} \geq \Delta_1 + \frac{1}{2}, d_{0,r} - \frac{1}{2} + \Delta > 0\}, \quad A_2 = \{d_{0,r} \geq \Delta_1 + \frac{1}{2}, d_{0,r} - \frac{1}{2} + \Delta \leq 0\}, \\ A_3 &= \{d_{0,r} < \Delta_1 + \frac{1}{2}\}. \end{aligned}$$

For each case, we apply Lemma B.1 and bound the respective quantities  $L_{rr,i}$ ,  $i = 1, \dots, 5$  in (B.1) that then yield the desired result. We will show that, for  $k = 1, 2, 3$ ,

$$\begin{aligned} & \mathbb{P}\left(\sup_{d_r \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d_r - 2d_{0,r}} \lambda_j^{2d_{0,r}} (I_{X,rr}(\lambda_j) - \mathbb{E}(I_{X,rr}(\lambda_j))) \right| \mathbf{1}_{A_k} > \|G\| \tilde{\nu}_2\right) \\ & \leq c_1 \exp\left(-c_2 \min\left\{\frac{\tilde{\nu}_2 m}{\gamma^2 \mathcal{L}_{k1}}, \frac{\tilde{\nu}_2^2 m^2}{\gamma^4 \mathcal{L}_{k2}^2}\right\}\right). \end{aligned} \quad (\text{B.75})$$

*Case  $d_{0,r} \geq \Delta_1 + \frac{1}{2}$  and  $d_{0,r} - \frac{1}{2} + \Delta > 0$ :* We get

$$\begin{aligned} L_{rr,1} &= N^{2d_{0,r}} \left( \max_{j=1, \dots, m} \left(\frac{j}{m}\right)^{2d_{0,r} - 1 + 2\Delta} \lambda_m^{2d_{0,r}} + \sup_{d_r \in \Theta_1} \max_{j=1, \dots, m} 2 \left| \log\left(\frac{j}{m}\right) \right| \left(\frac{j}{m}\right)^{2d_r} \lambda_m^{2d_{0,r}} \right) \\ & \leq c(1 + \log(m)) \max_{j=1, \dots, m} \left(\frac{j}{m}\right)^{2d_{0,r} - 1 + 2\Delta} m^{2d_{0,r}} \\ & \leq c \log(m) m^{2\Delta_2} =: \mathcal{L}_{11}, \end{aligned} \quad (\text{B.76})$$

$$\begin{aligned} L_{rr,i} &= c_{r,i,N}^2 \left( \max_{j=1, \dots, m} \left(\frac{j}{m}\right)^{2d_{0,r} - 1 + 2\Delta} \lambda_m^{2d_{0,r}} + \sup_{d_r \in \Theta_1} \max_{j=1, \dots, m} 2 \left| \log\left(\frac{j}{m}\right) \right| \left(\frac{j}{m}\right)^{2d_r} \lambda_m^{2d_{0,r}} \right) \\ & \leq c(1 + \log(m)) c_{r,i,N}^2 \max_{j=1, \dots, m} \left(\frac{j}{m}\right)^{2d_{0,r} - 1 + 2\Delta} \lambda_m^{2d_{0,r}} \\ & = c \log(m) \begin{cases} N^{\max\{2d_{0,r}, \frac{1}{2}\}} \lambda_m^{2d_{0,r}}, & \text{if } d_{0,r} > \frac{1}{4}, i = 2, \\ N^{\max\{2d_{0,r}, 0\}} m^{\frac{1}{2}} \lambda_m^{2d_{0,r}}, & \text{if } d_{0,r} \leq \frac{1}{4}, i = 4, \end{cases} \\ & \leq c \log(m) m^{2\bar{\Delta}_u} =: \mathcal{L}_{12}. \end{aligned} \quad (\text{B.77})$$

Then, applying Lemma B.1 with  $i = 2$  gives (B.75) with  $k = 1$ .

*Case  $d_{0,r} \geq \Delta_1 + \frac{1}{2}$  and  $d_{0,r} - \frac{1}{2} + \Delta \leq 0$ :* We get

$$\begin{aligned} L_{rr,1} &= N^{2d_{0,r}} \left( \max_{j=1, \dots, m} \left(\frac{j}{m}\right)^{2d_{0,r} - 1 + 2\Delta} \lambda_m^{2d_{0,r}} + \sup_{d_r \in \Theta_1} \max_{j=1, \dots, m} 2 \left| \log\left(\frac{j}{m}\right) \right| \left(\frac{j}{m}\right)^{2d_r} \lambda_m^{2d_{0,r}} \right) \\ & \leq c(1 + \log(m)) \max_{j=1, \dots, m} \left(\frac{j}{m}\right)^{2d_{0,r} - 1 + 2\Delta} m^{2d_{0,r}} \\ & \leq c \log(m) m^{1 - 2\Delta} =: \mathcal{L}_{21} \end{aligned} \quad (\text{B.78})$$

and

$$\begin{aligned}
L_{rr,2} &= \log(N)^{\frac{1}{2}} N^{\max\{2d_{0,r}, \frac{1}{2}\}} \left( \max_{j=1, \dots, m} \left( \frac{j}{m} \right)^{2d_{0,r}-1+2\Delta} \lambda_m^{2d_{0,r}} + \sup_{d_r \in \Theta_1} \max_{j=1, \dots, m} 2 \left| \log \left( \frac{j}{m} \right) \right| \left( \frac{j}{m} \right)^{2d_r} \lambda_m^{2d_{0,r}} \right) \\
&\leq c(1 + \log(m)) N^{\max\{2d_{0,r}, \frac{1}{2}\}} \max_{j=1, \dots, m} \left( \frac{j}{m} \right)^{2d_{0,r}-1+2\Delta} \lambda_m^{2d_{0,r}} \\
&\leq c \log(m) \max\{m^{1-2\Delta}, N^{\frac{1}{2}} m^{1-2\Delta} N^{-2d_{0,r}}\} \\
&\leq c \log(m) \max\{m^{1-2\Delta}, N^{-\frac{1}{2}-2\Delta_1} m^{1-2\Delta}\} =: \mathcal{L}_{22}.
\end{aligned} \tag{B.79}$$

Then, applying Lemma B.1 with  $i = 2, 5$  gives (B.75) with  $k = 2$ .

*Case  $d_{0,r} < \Delta_1 + \frac{1}{2}$ :* In this case,  $\Theta_1 = \{d \mid \Delta_1 \leq d \leq \Delta_2\}$ . Then,

$$\begin{aligned}
L_{1,rr} &= N^{\max\{2d_{0,r}, 0\}} \left( \max_{j=1, \dots, m} \left( \frac{j}{m} \right)^{2\Delta_1} \lambda_m^{2d_{0,r}} + \sup_{d_r \in \Theta_1} \max_{j=1, \dots, m} 2 \left| \log \left( \frac{j}{m} \right) \right| \left( \frac{j}{m} \right)^{2d_r} \lambda_m^{2d_{0,r}} \right) \\
&\leq c(1 + \log(m)) N^{\max\{2d_{0,r}, 0\}} \max_{j=1, \dots, m} \left( \frac{j}{m} \right)^{2\Delta_1} \lambda_m^{2d_{0,r}} \\
&\leq c \log(m) \max\{m^{2d_{0,r}-2\Delta_1}, m^{-2\Delta_1} \lambda_m^{2\Delta_1}\} \\
&\leq c \log(m) \max\{m^{2\bar{\Delta}_{l,1}-2\Delta_1}, N^{-2\Delta_1}\} =: \mathcal{L}_{31}
\end{aligned} \tag{B.80}$$

with  $\bar{\Delta}_{l,1}$  as in (A.2) and we distinguish further two cases: for  $d_{0,r} > 0$ ,

$$\begin{aligned}
L_{rr,i} &= c_{r,i,N}^2 \left( \max_{j=1, \dots, m} \left( \frac{j}{m} \right)^{2\Delta_1} \lambda_m^{2d_{0,r}} + \sup_{d_r \in \Theta_1} \max_{j=1, \dots, m} 2 \left| \log \left( \frac{j}{m} \right) \right| \left( \frac{j}{m} \right)^{2d_r} \lambda_m^{2d_{0,r}} \right) \\
&\leq c(1 + \log(m)) \begin{cases} N^{\max\{2d_{0,r}, \frac{1}{2}\}} m^{-2\Delta_1} \lambda_m^{2d_{0,r}}, & \text{if } d_{0,r} > \frac{1}{4}, i = 2, \\ m^{\frac{1}{2}} N^{\max\{2d_{0,r}, 0\}} m^{-2\Delta_1} \lambda_m^{2d_{0,r}}, & \text{if } d_{0,r} \leq \frac{1}{4}, i = 4, \end{cases} \\
&\leq c \log(m) \begin{cases} m^{2d_{0,r}-2\Delta_1}, & \text{if } d_{0,r} > \frac{1}{4}, i = 2, \\ m^{\frac{1}{2}+2d_{0,r}-2\Delta_1}, & \text{if } d_{0,r} \leq \frac{1}{4}, i = 4, \end{cases} \\
&\leq c \log(m) m^{2\bar{\Delta}_{l,2}-2\Delta_1}
\end{aligned} \tag{B.81}$$

with  $\bar{\Delta}_{l,2}$  as in (A.2) and for  $d_{0,r} \leq 0$ ,

$$\begin{aligned}
L_{rr,5} &= N^{\max\{2d_{0,r}, 0\}} \left( \left( \sum_{j=1}^m \left( \frac{j}{m} \right)^{4\Delta_1} \lambda_m^{4d_{0,r}} \right)^{\frac{1}{2}} + \sup_{d_r \in \Theta_1} 2 \left( \sum_{j=1}^m \left| \log \left( \frac{j}{m} \right) \right| \left( \frac{j}{m} \right)^{4d_r} \lambda_m^{4d_{0,r}} \right)^{\frac{1}{2}} \right) \\
&\leq c(1 + \log(m)) \left( \sum_{j=1}^m \left( \frac{j}{m} \right)^{4\Delta_1} \lambda_m^{4d_{0,r}} \right)^{\frac{1}{2}} \\
&\leq c \log(m) \lambda_m^{2d_{0,r}-2\Delta_1} \left( \sum_{j=1}^m \lambda_j^{4\Delta_1} \right)^{\frac{1}{2}} \\
&\leq c \log(m) \left( \sum_{j=1}^m \lambda_j^{4\Delta_1} \right)^{\frac{1}{2}}.
\end{aligned} \tag{B.82}$$

Finally, given (B.81) and (B.82), we define

$$\mathcal{L}_{32} := c \log(m) \max\{m^{2\bar{\Delta}_{l,2}-2\Delta_1}, \left( \sum_{j=1}^m \lambda_j^{4\Delta_1} \right)^{\frac{1}{2}}\}. \tag{B.83}$$

Then, applying Lemma B.1 gives (B.75) with  $k = 3$ .

*Deterministic part:* Using Lemma B.2 with (B.73), we get

$$\begin{aligned} & \sup_{d_r \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2d_r - 2d_{0,r}} \lambda_j^{2d_{0,r}} (\mathbb{E} I_{X,rr}(\lambda_j) - \lambda_j^{-2d_{0,r}} g_{0,r}) \right| \\ & \leq c_{G,1} \frac{1}{2\pi} \lambda_m^{2q} \frac{1}{2\tilde{\Delta}_r} + \frac{1}{2\pi} m^{-2\tilde{\Delta}_r} \left( 1 + \frac{1}{1 - 2\tilde{\Delta}_r} \right) Q_m \leq \max_{r=1,\dots,p} \mathcal{T}_3(\tilde{\Delta}_r) = \mathcal{T}_3 \end{aligned} \quad (\text{B.84})$$

with  $\tilde{\Delta}_r = (d_{0,r} - \frac{1}{2} + \Delta) \mathbb{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} + \Delta_1 \mathbb{1}_{\{d_{0,r} < \Delta_1 + \frac{1}{2}\}}$  and  $\mathcal{T}_3(\tilde{\Delta}_r)$  is in Table A.2. The inequality can be obtained by bounding the two terms in (B.40) given (B.73) as follows. The first summand in (B.40) can be bounded as

$$\begin{aligned} \sup_{d_r \in \Theta_1} \frac{1}{2\pi m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2d_r} \lambda_m^{2d_{0,r}} \lambda_j^{2q-2d_{0,r}} & \leq \frac{1}{2\pi m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2(\tilde{\Delta}_r - d_{0,r})} \lambda_j^{2q} \\ & \leq \lambda_m^{2q} \frac{1}{2\pi m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2(\tilde{\Delta}_r - d_{0,r})} \\ & \leq \lambda_m^{2q} \frac{1}{2\pi} \frac{1}{2\tilde{\Delta}_r - 2d_{0,r} + 1} \end{aligned}$$

and the second as

$$\begin{aligned} \sup_{d_r \in \Theta_1} \frac{1}{2\pi m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2d_r} \lambda_m^{2d_{0,r}} j^{-1} \lambda_j^{-2d_{0,r}} & \leq \frac{1}{2\pi m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2(\tilde{\Delta}_r - d_{0,r})} \lambda_j^{2d_{0,r}} j^{-1} \lambda_j^{-2d_{0,r}} \\ & = \frac{1}{2\pi m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2(\tilde{\Delta}_r - d_{0,r})} j^{-1} \\ & \leq \frac{1}{2\pi} m^{-1-2\tilde{\Delta}_r+2d_{0,r}} \left( 1 + \frac{1}{2d_{0,r} - 2\tilde{\Delta}_r} \right). \end{aligned}$$

Finally, we combine our results on the probabilistic and deterministic terms to obtain the statement of the proposition. With our choice of  $\nu_2$  in (3.16), observe that, with explanations given below,

$$\begin{aligned} & \mathbb{P} \left( \sup_{d_r \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2d_r - 2d_{0,r}} \lambda_j^{2d_{0,r}} (I_{X,rr}(\lambda_j) - \lambda_j^{-2d_{0,r}} g_{0,r}) \right| > \nu_2 \right) \\ & \leq \mathbb{P} \left( \sup_{d_r \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2d_r - 2d_{0,r}} \lambda_j^{2d_{0,r}} (I_{X,rr}(\lambda_j) - \mathbb{E}(I_{X,rr}(\lambda_j))) \right| \right. \\ & \quad \left. + \sup_{d_r \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2d_r - 2d_{0,r}} \lambda_j^{2d_{0,r}} (\mathbb{E}(I_{X,rr}(\lambda_j)) - \lambda_j^{-2d_{0,r}} g_{0,r}) \right| > C \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_3}} + \mathcal{T}_3 \right) \\ & \leq \mathbb{P} \left( \sup_{d_r \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2d_r - 2d_{0,r}} \lambda_j^{2d_{0,r}} (I_{X,rr}(\lambda_j) - \mathbb{E}(I_{X,rr}(\lambda_j))) \right| > C \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_3}} \right) \quad (\text{B.85}) \end{aligned}$$

$$\leq \sum_{k=1}^3 \mathbb{P} \left( \sup_{d_r \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^m \left( \frac{j}{m} \right)^{2d_r - 2d_{0,r}} \lambda_j^{2d_{0,r}} (I_{X,rr}(\lambda_j) - \mathbb{E}(I_{X,rr}(\lambda_j))) \right| \mathbb{1}_{A_k} > C \|G\| \sqrt{\frac{\log(p)}{\mathcal{R}_3}} \right) \quad (\text{B.86})$$

$$\leq \sum_{k=1}^3 c_1 \exp \left( -c_2 \mathbf{C} \min \left\{ \frac{\sqrt{\frac{\log(p)}{\log(m)^{-1} \bar{\Delta}_N^{-1} \mathcal{R}_{31}} m}}}{\gamma^2 \bar{\Delta}_N \mathcal{L}_{k1}}, \frac{\frac{\log(p)}{\log(m)^{-2} \bar{\Delta}_N^{-2} \mathcal{R}_{32}} m^2}}{\gamma^4 \bar{\Delta}_N^2 \mathcal{L}_{k2}^2} \right\} \right) \quad (\text{B.87})$$

$$\leq c_1 \exp \left( -c_2 \mathbf{C} \min \left\{ \frac{\sqrt{\frac{\log(p)}{\log(m)^{-1} \bar{\Delta}_N^{-1} \mathcal{R}_{31}} m}}}{\gamma^2 \bar{\Delta}_N \max_{k=1,2,3} \mathcal{L}_{k1}}, \frac{\frac{\log(p)}{\log(m)^{-2} \bar{\Delta}_N^{-2} \mathcal{R}_{32}} m^2}}{\gamma^4 \bar{\Delta}_N^2 \max_{k=1,2,3} \mathcal{L}_{k2}^2} \right\} \right) \quad (\text{B.88})$$

$$\leq c_1 p^{-c_2 \mathbf{C}} \quad (\text{B.89})$$

with  $\mathcal{L}_{k1}, \mathcal{L}_{k2}, k = 1, 2, 3$  defined in (B.76), (B.77), (B.78), (B.79) and (B.80), (B.83). The constants  $c_1, c_2$  are generic and might differ from line to line. The inequality (B.85) is due to (B.84). In (B.86), the probabilistic part is bounded as in (B.74). Applying (B.75) yields (B.87). For the inequality (B.88), note that  $\mathcal{R}_{31}$  and  $\mathcal{R}_{32}$  are chosen in Table A.1 such that

$$\log(m)^{-1} \mathcal{R}_{31} \lesssim m \left( \max_{k=1,2,3} \mathcal{L}_{k1} \right)^{-1} \quad \text{and} \quad \log(m)^{-2} \mathcal{R}_{32} \lesssim m^2 \left( \max_{k=1,2,3} \mathcal{L}_{k2} \right)^{-2}.$$

The inequality (B.89) follows since we work under the assumption  $\mathcal{R}_{31} \gtrsim \bar{\Delta}_N \log(m) \log(p)$ ; see the discussion following Proposition 3.3.  $\square$

*Proof of Proposition 3.4:* To apply Lemmas B.1 and B.2, we choose

$$t_{j,r}(d_{0,r}) = |l_j - 1|^{\frac{1}{2}} \lambda_j^{d_{0,r}} \mathbf{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} = |l_j - 1|^{\frac{1}{2}} \left( \frac{j}{\ell} \right)^{d_{0,r}} \lambda_\ell^{d_{0,r}} \mathbf{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} \quad (\text{B.90})$$

in (3.8)–(3.9) with  $\ell$  and  $l_j$  as in (3.7). Here, we do not need a uniform bound. For this reason, it is not necessary to take the supremum over all admissible estimates of  $D_0$  in Lemmas B.1 and B.2. In particular, Lemma B.1 simplifies, since the derivatives in the respective bounds  $L_{rr,i}$  in (B.1) become zero with the choice (B.90). Throughout the proof, we assume  $d_{0,r} \geq \Delta_1 + \frac{1}{2}$ .

*Probabilistic part:* Note that

$$\begin{aligned} \max_{j=1,\dots,m} |l_j - 1| \left( \frac{j}{\ell} \right)^{2d_{0,r}} &= \max_{j=1,\dots,m} \begin{cases} \left| \left( \frac{j}{\ell} \right)^{2(-\frac{1}{2}+\Delta)} - 1 \right| \left( \frac{j}{\ell} \right)^{2d_{0,r}}, & 1 \leq j \leq \ell, \\ \left| \left( \frac{j}{\ell} \right)^{2(\Delta_1 - d_{0,r})} - 1 \right| \left( \frac{j}{\ell} \right)^{2d_{0,r}}, & \ell < j \leq m, \end{cases} \\ &= \max_{j=1,\dots,m} \begin{cases} \left| \left( \frac{j}{\ell} \right)^{2(d_{0,r} - \frac{1}{2} + \Delta)} - \left( \frac{j}{\ell} \right)^{2d_{0,r}} \right|, & 1 \leq j \leq \ell, \\ \left| \left( \frac{j}{\ell} \right)^{2\Delta_1} - \left( \frac{j}{\ell} \right)^{2d_{0,r}} \right|, & \ell < j \leq m, \end{cases} \\ &\leq \max \left\{ \ell^{-\min\{2(d_{0,r} - \frac{1}{2} + \Delta), 0\}}, \left( \frac{m}{\ell} \right)^{2d_{0,r}} \right\}. \end{aligned}$$

Then,

$$\begin{aligned} L_{rr,1} &= N^{\max\{2d_{0,r}, 0\}} \max_{j=1,\dots,m} |l_j - 1| \left( \frac{j}{\ell} \right)^{2d_{0,r}} \lambda_\ell^{2d_{0,r}} \\ &\leq c N^{2d_{0,r}} \max \left\{ \ell^{-\min\{2(d_{0,r} - \frac{1}{2} + \Delta), 0\}}, \left( \frac{m}{\ell} \right)^{2d_{0,r}} \right\} \left( \frac{\ell}{N} \right)^{2d_{0,r}} \\ &\leq c \max\{\ell^{1-2\Delta}, m^{2\Delta_2}\} =: \mathcal{L}_1. \end{aligned} \quad (\text{B.91})$$

Furthermore,

$$\begin{aligned}
L_{rr,i} &= c_{r,i,N}^2 \max_{j=1,\dots,m} |\ell_j - 1| \left(\frac{j}{\ell}\right)^{2d_{0,r}} \lambda_\ell^{2d_{0,r}} \\
&\leq c \begin{cases} N^{2d_{0,r}} \max \left\{ \ell^{-\min\{2(d_{0,r}-\frac{1}{2}+\Delta),0\}}, \left(\frac{m}{\ell}\right)^{2d_{0,r}} \right\} \left(\frac{\ell}{N}\right)^{2d_{0,r}}, & \text{if } d_{0,r} > \frac{1}{4}, i = 2, \\ m^{\frac{1}{2}} N^{2d_{0,r}} \max \left\{ \ell^{-\min\{2(d_{0,r}-\frac{1}{2}+\Delta),0\}}, \left(\frac{m}{\ell}\right)^{2d_{0,r}} \right\} \left(\frac{\ell}{N}\right)^{2d_{0,r}}, & \text{if } d_{0,r} \leq \frac{1}{4}, i = 4, \end{cases} \\
&\leq c \begin{cases} \max\{\ell^{1-2\Delta}, m^{2\Delta_2}\}, & \text{if } d_{0,r} > \frac{1}{4}, i = 2, \\ \max\{m^{\frac{1}{2}}\ell^{1-2\Delta}, m^{\frac{1}{2}+2d_{0,r}}\}, & \text{if } d_{0,r} \leq \frac{1}{4}, i = 4, \end{cases} \\
&\leq c \max\{m^{\frac{1}{2}}\ell^{1-2\Delta}, m^{2\bar{\Delta}_u}\} =: \mathcal{L}_2.
\end{aligned} \tag{B.92}$$

*Deterministic part:* Using Lemma B.2 with (B.90), we get

$$\begin{aligned}
& \left| \frac{1}{m} \sum_{j=1}^m (l_j - 1) \lambda_j^{2d_{0,r}} (\mathbb{E} I_{X,rr}(\lambda_j) - \lambda_j^{-2d_{0,r}} g_{0,r}) \right| \\
& \leq c_{G,1} \frac{1}{2\pi} \lambda_m^{2q} \left( \frac{1}{2\Delta} + 1 \right) + \frac{1}{2\pi m} \left( \ell^{1-2\Delta} \left( 1 + \frac{1}{1-2\Delta} \right) + \log(m) \right) Q_m = \mathcal{T}_4,
\end{aligned} \tag{B.93}$$

where  $\mathcal{T}_4$  is in Table A.2 and the inequality can be obtained by bounding the two terms in (B.40) given (B.90). as follows. For  $d_{0,r} \geq \Delta_1 + \frac{1}{2}$ , the first summand in (B.40) can be bounded as

$$\begin{aligned}
& \frac{1}{2\pi m} \sum_{j=1}^m |l_j - 1| \lambda_j^{2d_{0,r}} \lambda_j^{2q-2d_{0,r}} \\
& = \frac{1}{2\pi m} \left( \sum_{j=1}^{\ell} \left| \left(\frac{j}{\ell}\right)^{-1+2\Delta} - 1 \right| \lambda_j^{2q} + \sum_{j=\ell+1}^m \left| \left(\frac{j}{\ell}\right)^{2\Delta_1-2d_{0,r}} - 1 \right| \lambda_j^{2q} \right) \\
& \leq \frac{1}{2\pi m} \lambda_m^{2q} \left( \int_0^\ell x^{-1+2\Delta} dx \ell^{1-2\Delta} + (m - \ell) \right) \\
& \leq \frac{1}{2\pi m} \lambda_m^{2q} \left( \frac{1}{2\Delta} \ell^{1-2\Delta} + (m - \ell) \right) \\
& \leq \frac{1}{2\pi} \lambda_m^{2q} \left( \frac{1}{2\Delta} + 1 \right)
\end{aligned}$$

and the second as

$$\begin{aligned}
& \frac{1}{2\pi m} \sum_{j=1}^m |l_j - 1| \lambda_j^{2d_{0,r}} j^{-1} \lambda_j^{-2d_{0,r}} = \frac{1}{2\pi m} \sum_{j=1}^m |l_j - 1| j^{-1} \\
& = \frac{1}{2\pi m} \left( \sum_{j=1}^{\ell} j^{-1} \left| \left(\frac{j}{\ell}\right)^{-1+2\Delta} - 1 \right| + \sum_{j=\ell+1}^m j^{-1} \left| \left(\frac{j}{\ell}\right)^{2\Delta_1-2d_{0,r}} - 1 \right| \right) \\
& \leq \frac{1}{2\pi m} \left( \sum_{j=1}^{\ell} j^{-1} \left(\frac{j}{\ell}\right)^{-1+2\Delta} + \log(m) \right) \\
& \leq \frac{1}{2\pi m} \left( \ell^{1-2\Delta} \left( 1 + \frac{1}{1-2\Delta} \right) + \log(m) \right).
\end{aligned}$$

Finally, we combine our results on the probabilistic and deterministic terms to obtain the statement of the proposition. With our choice of  $\nu_3$  in (3.17) and for any  $\mathbf{C} \geq 1$ , observe that

$$\begin{aligned}
& \mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^m(l_j-1)\lambda_j^{2d_{0,r}}(I_{X,rr}(\lambda_j)-\lambda_j^{-2d_{0,r}}g_{0,r})\right|\mathbb{1}_{\{d_{0,r}\geq\Delta_1+\frac{1}{2}\}}>\nu_3\right) \\
& \leq \mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^m(l_j-1)\lambda_j^{2d_{0,r}}(I_{X,rr}(\lambda_j)-\mathbb{E}I_{X,rr}(\lambda_j))\right|\mathbb{1}_{\{d_{0,r}\geq\Delta_1+\frac{1}{2}\}}\right. \\
& \quad \left.+\left|\frac{1}{m}\sum_{j=1}^m(l_j-1)\lambda_j^{2d_{0,r}}(\mathbb{E}I_{X,rr}(\lambda_j)-\lambda_j^{-2d_{0,r}}g_{0,r})\right|\mathbb{1}_{\{d_{0,r}\geq\Delta_1+\frac{1}{2}\}}>\mathbf{C}\|G\|\sqrt{\frac{\log(p)}{\mathcal{R}_4}}+\mathcal{T}_4\right) \\
& \leq \mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^m(l_j-1)\lambda_j^{2d_{0,r}}(I_{X,rr}(\lambda_j)-\lambda_j^{-2d_{0,r}}g_{0,r})\right|\mathbb{1}_{\{d_{0,r}\geq\Delta_1+\frac{1}{2}\}}>\mathbf{C}\|G\|\sqrt{\frac{\log(p)}{\mathcal{R}_4}}\right) \tag{B.94}
\end{aligned}$$

$$\leq \exp\left(-c_2\mathbf{C}\min\left\{\sqrt{\frac{\frac{\log(p)}{\Delta_N^{-1}\mathcal{R}_{41}}m}{\gamma^2\bar{\Delta}_N\mathcal{L}_1}},\frac{\frac{\log(p)}{\Delta_N^{-2}\mathcal{R}_{42}}m^2}{\gamma^4\bar{\Delta}_N^2\mathcal{L}_2^2}\right\}\right) \tag{B.95}$$

$$\leq c_1\exp\left(-c_2\mathbf{C}\min\left\{\sqrt{\frac{\log(p)}{\bar{\Delta}_N}\mathcal{R}_{41}},\log(p)\right\}\right) \tag{B.96}$$

$$\leq c_1p^{-c_2\mathbf{C}} \tag{B.97}$$

with  $\mathcal{L}_1, \mathcal{L}_2$  defined in (B.91) and (B.92). The constants  $c_1, c_2$  are generic and might differ from line to line. The inequality (B.94) is due to (B.93). Applying Lemma B.1 yields (B.95). For the inequality (B.96), note that  $\mathcal{R}_{41}$  and  $\mathcal{R}_{42}$  are chosen in Table A.1 such that  $\mathcal{R}_{41} \lesssim m\mathcal{L}_1^{-1}$  and  $\mathcal{R}_{42} \lesssim m^2\mathcal{L}_2^{-2}$ . The inequality (B.97) follows since we work under the assumption  $\mathcal{R}_{41} \lesssim \bar{\Delta}_N \log(p)$ ; see the discussion following Proposition 3.4.  $\square$

*Proof of Proposition 3.5:* To prove the consistency result in (3.18), we use the inequalities on the probability of  $\{\|\hat{G}(\hat{D}) - G_0\|_{\max} > \delta\}$  discussed in Section 3 and combine the concentration inequalities in Propositions 3.1–3.4. We have

$$\begin{aligned}
& \mathbb{P}(\|\hat{G}(\hat{D}) - G_0\|_{\max} > \delta) \\
& = \mathbb{P}\left(\{\|\hat{G}(\hat{D}) - G_0\|_{\max} > \delta\} \cap \left(\{\|\hat{D} - D_0\|_{\max} \leq \varepsilon\} \cup \{\|\hat{D} - D_0\|_{\max} > \varepsilon\}\right)\right) \\
& \leq \mathbb{P}(\{\|\hat{G}(\hat{D}) - \tilde{G}(\hat{D})\|_{\max} > \delta/2\} \cap \{\|\hat{D} - D_0\|_{\max} \leq \varepsilon\}) \\
& \quad + \mathbb{P}\left(\left(\{\|\tilde{G}(\hat{D}) - G_0\|_{\max} > \delta/2\} \cap \{\|\hat{D} - D_0\|_{\max} \leq \varepsilon\}\right) \cup \{\|\hat{D} - D_0\|_{\max} > \varepsilon\}\right) \\
& \leq \mathbb{P}(\{\|\hat{G}(\hat{D}) - \tilde{G}(\hat{D})\|_{\max} > \delta/2\} \cap \{\|\hat{D} - D_0\|_{\max} \leq \varepsilon\}) + \mathbb{P}(\|\hat{D} - D_0\|_{\max} > \eta) \tag{B.98}
\end{aligned}$$

$$\leq \sum_{r,s=1}^p \mathbb{P}(\sup_{D \in \Omega(\varepsilon)} |\hat{G}_{rs}(D) - \tilde{G}_{rs}(D)| > \delta/2) + \mathbb{P}(\|\hat{D} - D_0\|_{\max} > \eta) \tag{B.99}$$

with  $\eta = \min\{\varepsilon, \frac{\delta}{4}(\|G_0\| \log(N)\lambda_m^{-2\varepsilon}L(-\varepsilon))^{-1}\}$ . The inequality (B.98) follows by Lemma C.6. We proceed with bounding the second probability in (B.98) before we conclude with utilizing Propositions 3.1–3.4 to get the statement of this proposition.

To find an upper bound on the second probability in (B.98), we follow the Robinson (1995b) approach used to prove the classical consistency result in the univariate case. Thus, we write the

set  $\Theta = [\Delta_1, \Delta_2]$  as a union  $\Theta = \Theta_1 \cup \Theta_2$  with  $\Theta_1$  as in (3.22) and  $\Theta_2$  as in (3.23); see Remark 3.4. Denote further  $S_r(d) = R_r(d) - R_r(d_{0,r})$  with  $R_r$  in (2.1). Then, with explanations given below,

$$\begin{aligned} & \mathbb{P}(|\widehat{d}_r - d_{0,r}| > \eta) \\ & \leq \mathbb{P}(\inf_{d \in N_\eta^c \cap \Theta_1} S_r(d) < 0) + \mathbb{P}(\inf_{d \in \Theta_2} S_r(d) < 0) \end{aligned} \quad (\text{B.100})$$

$$\begin{aligned} & \leq \mathbb{P}(|\widehat{g}_r(d_{0,r}) - g_{0,r}| > \frac{1}{8}\eta^2\mathcal{V}_1(m)g_{0,r}) + \mathbb{P}(\sup_{d \in \Theta_1} |\widehat{h}_r(d) - \widetilde{h}_r(d)| > \frac{1}{4}\eta^2\mathcal{V}_1(m)L(\Delta_2 - \Delta_1)g_{0,r}) \\ & \quad + \mathbb{P}(|\frac{1}{m} \sum_{j=1}^m (l_j - 1)(\lambda_j^{2d_{0,r}} I_X(\lambda_j) - g_{0,r})| \mathbf{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} > \mathcal{V}_2(m)g_{0,r}), \end{aligned} \quad (\text{B.101})$$

where  $N_\eta = \{d \mid |d - d_{0,r}| \leq \eta\}$  and  $N_\eta^c$  denotes the complement of  $N_\eta$ . The inequality in (B.100) follows from (3.2) in Robinson (1995b). Lemmas C.1 and C.5 give (B.101).

Finally, combining (B.99) and (B.101) and choosing  $\delta$  and  $\varepsilon$  according to (3.19), we can infer that there exist constants  $c_1, c_2$  such that for any  $\mathbf{C} \geq 1$ ,

$$\begin{aligned} & \mathbb{P}(\|\widehat{G}(\widehat{D}) - G_0\|_{\max} > \delta) \\ & \leq \sum_{r,s=1}^p \mathbb{P}(\sup_{D \in \Omega(\varepsilon)} |\widehat{G}_{rs}(D) - \widetilde{G}_{rs}(D)| > \delta/2) + \sum_{r=1}^p \mathbb{P}(|\widehat{g}_r(d_{0,r}) - g_{0,r}| > \frac{1}{8}\eta^2\mathcal{V}_1(m)g_{0,r}) \\ & \quad + \sum_{r=1}^p \mathbb{P}(\sup_{d \in \Theta_1} |\widehat{h}_r(d) - \widetilde{h}_r(d)| > \frac{1}{4}\eta^2\mathcal{V}_1(m)L(\Delta_2 - \Delta_1)g_{0,r}) \\ & \quad + \sum_{r=1}^p \mathbb{P}(|\frac{1}{m} \sum_{j=1}^m (l_j - 1)(\lambda_j^{2d_{0,r}} I_X(\lambda_j) - g_{0,r})| \mathbf{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} > \varepsilon^2\mathcal{V}_2(m)g_{0,r}), \end{aligned} \quad (\text{B.102})$$

$$\begin{aligned} & \leq \sum_{r,s=1}^p \mathbb{P}(\sup_{D \in \Omega(\varepsilon)} |\widehat{G}_{rs}(D) - \widetilde{G}_{rs}(D)| > \nu) + \sum_{r=1}^p \mathbb{P}(|\widehat{g}_r(d_{0,r}) - g_{0,r}| > \nu_1) \\ & \quad + \sum_{r=1}^p \mathbb{P}(\sup_{d \in \Theta_1} |\widehat{h}_r(d) - \widetilde{h}_r(d)| > \nu_2) \\ & \quad + \sum_{r=1}^p \mathbb{P}(|\frac{1}{m} \sum_{j=1}^m (l_j - 1)(\lambda_j^{2d_{0,r}} I_X(\lambda_j) - g_{0,r})| \mathbf{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} > \nu_3) \end{aligned} \quad (\text{B.103})$$

$$\leq c_1 p^{2-c_2\mathbf{C}}, \quad (\text{B.104})$$

where (B.102) follows by (B.101) and since  $\varepsilon \in (0, \frac{1}{2})$  is assumed. For (B.103), recall our choice of  $\delta = \max\{2\nu, \varepsilon 4\|G_0\| \log(N)\lambda_m^{-2\varepsilon}L(-\varepsilon)\}$  in (3.19). Then,  $\delta/2$  in the first probability of (B.102) can be bounded from below by  $\nu$ . For the remaining three probabilities in (B.102), note that  $\eta$  satisfies

$$\eta = \min\{\varepsilon, \frac{\delta}{4}(\|G_0\| \log(N)\lambda_m^{-2\varepsilon}L(-\varepsilon))^{-1}\} \geq \varepsilon,$$

since  $\delta \geq \varepsilon 4\|G_0\| \log(N)\lambda_m^{-2\varepsilon}L(-\varepsilon)$ . Note also that  $\varepsilon = \max_{i=1,2,3} \eta_i$  in (3.19) can be bounded from below by  $\eta_1, \eta_2$  or  $\eta_3$ . Then, for example, the lower bound in the second probability in (B.102) satisfies

$$\frac{1}{8}\eta^2\mathcal{V}_1(m)g_{0,r} \geq \frac{1}{8}\varepsilon^2\mathcal{V}_1(m)g_{0,r} \geq \frac{1}{8}\eta_1^2\mathcal{V}_1(m)g_{0,r} \geq \nu_1,$$

where the last inequality follows by our choice of  $\eta_1$  in (3.20). This implies the second term as a bound in (B.103). The other two probabilities can be dealt with similarly. Furthermore, (B.104) is a consequence of applying Propositions 3.1–3.4 to the individual probabilities in (B.103).  $\square$



*Proof of Proposition 3.6:* The key in proving Proposition 3.6 is finding a concentration inequality on  $\|\widehat{G}(\widehat{D}) - G_0\|_{\max}$ . Such a concentration inequality is provided in Proposition 3.5. Then, it is left to follow exactly the proof of Proposition 3.6 in Sun et al. (2018).  $\square$

*Proof of Proposition 3.7:* The proof follows the ideas used in the proof of Theorem 1 in Rothman et al. (2008). However, we need to modify their proof in order to achieve a non-asymptotic result. With the objective function  $\ell_\rho$  as in (2.3), let

$$\begin{aligned} Q(\widehat{D}, P) &= \ell_\rho(\widehat{D}, P) - \ell_\rho(\widehat{D}, P_0) \\ &= -\frac{1}{m} \sum_{j=1}^m \log |\lambda_j^{\widehat{D}} P \lambda_j^{\widehat{D}}| + \text{tr}(\widehat{G}(\widehat{D})P) + \rho\|P\|_{1,off} \\ &\quad + \frac{1}{m} \sum_{j=1}^m \log |\lambda_j^{\widehat{D}} P_0 \lambda_j^{\widehat{D}}| - \text{tr}(\widehat{G}(\widehat{D})P_0) - \rho\|P_0\|_{1,off} \\ &= \text{tr}(\widehat{G}(\widehat{D})(P - P_0)) - (\log|P| - \log|P_0|) + \rho(\|P\|_{1,off} - \|P_0\|_{1,off}) \end{aligned}$$

and set  $Z(\Delta) = Q(\widehat{D}, P_0 + \Delta)$ . We further define the set  $R_\lambda = \{P = P^* \mid \|P - P_0\|_F > \lambda\}$ . Then,

$$\begin{aligned} \mathbb{P}(\|\widehat{P}_\rho(\widehat{D}) - P_0\|_F > \lambda) &= \mathbb{P}(\widehat{P}_\rho(\widehat{D}) \in R_\lambda) \leq \mathbb{P}(\inf_{P \in R_\lambda} \ell_\rho(\widehat{D}, P) \leq \inf_{P \in R_\lambda^c} \ell_\rho(\widehat{D}, P)) \\ &= \mathbb{P}(\inf_{P \in R_\lambda} \ell_\rho(\widehat{D}, P) - \ell_\rho(\widehat{D}, P_0) \leq \inf_{P \in R_\lambda^c} \ell_\rho(\widehat{D}, P) - \ell_\rho(\widehat{D}, P_0)) \\ &= \mathbb{P}(\inf_{P \in R_\lambda} Q(\widehat{D}, P) \leq \inf_{P \in R_\lambda^c} Q(\widehat{D}, P)) \\ &\leq \mathbb{P}(\inf_{P \in R_\lambda} Q(\widehat{D}, P) \leq 0) \tag{B.105} \end{aligned}$$

$$\leq \mathbb{P}(\inf_{\Delta \in \Theta_N} Z(\Delta) \leq 0), \tag{B.106}$$

where (B.105) follows since  $P_0 \in R_\lambda^c$  and (B.106) with

$$\Theta_N = \{\Delta = \Delta^* \mid \|\Delta\|_F = \lambda\} \quad \text{and} \quad \lambda = \frac{16}{k^2} \sqrt{p+s} \delta \tag{B.107}$$

follows since  $Z(\cdot)$  is a convex function. Then, it remains to prove that  $\inf_{\Delta \in \Theta_N} Z(\Delta) > 0$  with high probability.

From the objective function in (2.3) recall that the  $l_1$ -norm  $\|\cdot\|_{1,off}$  excludes the diagonal elements and introduce its counterpart  $\|\cdot\|_{1,on}$ . Analogously, we write  $\|\cdot\|_{F,off}$  and  $\|\cdot\|_{F,on}$  for the Frobenius norm of the off and on diagonal elements of a matrix, respectively. Recall the definition of the index set  $S$  in (2.10) and define further  $M_S = (m_{rs} \mathbf{1}_{\{(r,s) \in S\}})_{r,s=1,\dots,p}$  for a matrix  $M = (m_{rs})_{r,s=1,\dots,p}$  and an index set  $S$ . Then, with further explanations given below,

$$\begin{aligned} Z(\Delta) &= \text{tr}(\widehat{G}(\widehat{D})\Delta) - (\log|\Delta + P_0| - \log|P_0|) + \rho(\|\Delta + P_0\|_{1,off} - \|P_0\|_{1,off}) \\ &= \text{tr}((\widehat{G}(\widehat{D}) - G_0)\Delta) + \text{vec}(\Delta)' \int_0^1 (1-v)(P_0 + v\Delta)^{-1} \otimes (P_0 + v\Delta)^{-1} dv \text{vec}(\Delta) \\ &\quad + \rho(\|\Delta + P_0\|_{1,off} - \|P_0\|_{1,off}) \tag{B.108} \end{aligned}$$

$$\geq \text{tr}((\widehat{G}(\widehat{D}) - G_0)\Delta) + \|\Delta\|_F^2 \frac{1}{8} k^2 + \rho(\|\Delta + P_0\|_{1,off} - \|P_0\|_{1,off}) \tag{B.109}$$

$$\geq \|\Delta\|_F^2 \frac{1}{8} k^2 - \delta \|\Delta\|_1 + \rho(\|\Delta + P_0\|_{1,off} - \|P_0\|_{1,off}) \tag{B.110}$$

$$\geq \|\Delta\|_F^2 \frac{1}{8} k^2 - \delta \|\Delta\|_1 + \rho(-\|\Delta_S\|_{1,off} + \|\Delta_{S^c}\|_{1,off}) \quad (\text{B.111})$$

$$\begin{aligned} &\geq \|\Delta\|_F^2 \frac{1}{8} k^2 - \delta(\|\Delta\|_{1,on} + \|\Delta_S\|_{1,off} + \|\Delta_{S^c}\|_{1,off}) \\ &\quad + \rho(-\|\Delta_S\|_{1,off} + \|\Delta_{S^c}\|_{1,off}) \end{aligned} \quad (\text{B.112})$$

$$\geq \|\Delta\|_{F,on}^2 \left( \frac{1}{8} k^2 - \delta \sqrt{p} \|\Delta\|_{F,on}^{-1} \right) + \|\Delta\|_{F,off}^2 \left( \frac{1}{8} k^2 - 2\delta \sqrt{s} \|\Delta\|_{F,off}^{-1} \right) \quad (\text{B.113})$$

$$\begin{aligned} &\geq \|\Delta\|_{F,on}^2 \left( \frac{1}{8} k^2 - 2\delta \sqrt{p+s} \|\Delta\|_{F,on}^{-1} \right) + \|\Delta\|_{F,off}^2 \left( \frac{1}{8} k^2 - 2\delta \sqrt{p+s} \|\Delta\|_{F,off}^{-1} \right) \\ &\geq \|\Delta\|_F^2 \left( \frac{1}{8} k^2 - \sqrt{2} \delta \sqrt{p+s} \|\Delta\|_F^{-1} \right) > 0. \end{aligned} \quad (\text{B.114})$$

The equality (B.108) follows by the Taylor expansion of the function  $f(t) = \log|P_0 + t\Delta|$ , which gives

$$\begin{aligned} &\log|\Delta + P_0| - \log|P_0| \\ &= \text{tr}(G_0\Delta) - \text{vec}(\Delta)' \int_0^1 (1-v)(P_0 + v\Delta)^{-1} \otimes (P_0 + v\Delta)^{-1} dv \text{vec}(\Delta). \end{aligned}$$

In (B.109), following Rothman et al. (2008), we further bound the integral part from below by its minimal eigenvalue

$$\begin{aligned} &\lambda_{\min} \left( \int_0^1 (1-v)(P_0 + v\Delta)^{-1} \otimes (P_0 + v\Delta)^{-1} dv \right) \\ &\geq \frac{1}{2} \min_{0 \leq v \leq 1} \lambda_{\min}^2(P_0 + v\Delta)^{-1} \geq \frac{1}{2} \min_{\|\Delta\|_F \leq \lambda} \lambda_{\min}^2(P_0 + \Delta)^{-1} \geq \frac{1}{8} k^2, \end{aligned}$$

since, for  $\|\Delta\|_F \leq \lambda$ ,

$$\begin{aligned} \lambda_{\min}^2(P_0 + \Delta)^{-1} &= \lambda_{\max}^{-2}(P_0 + \Delta) \geq (\|P_0\| + \|\Delta\|)^{-2} \geq (\|P_0\| + \|\Delta\|_F)^{-2} \\ &\geq (\|P_0\| + \lambda)^{-2} \geq (2\|P_0\|)^{-2} = \frac{1}{4} \lambda_{\min}^2(G_0) \geq \frac{1}{4} k^2, \end{aligned}$$

since  $\lambda = \frac{16}{k^2} \sqrt{p+s} \leq \|P_0\|$  by the assumption of Proposition 3.7. Applying Proposition 3.5 gives inequality (B.110) with probability at least  $1 - c_1 p^{2-c_2} \mathcal{C}$ . The inequality (B.111) follows since

$$\begin{aligned} \|\Delta + P_0\|_{1,off} - \|P_0\|_{1,off} &= \|\Delta_S + P_{0,S}\|_{1,off} + \|\Delta_{S^c}\|_{1,off} - \|P_0\|_{1,off} \\ &\geq -\|\Delta_S\|_{1,off} + \|\Delta_{S^c}\|_{1,off}. \end{aligned}$$

Line (B.112) follows since  $\rho = \delta$ . In (B.113), we used the Cauchy-Schwarz inequality to get  $\|\Delta\|_{1,on} \leq \sqrt{p} \|\Delta\|_{F,on}$  and  $\|\Delta_S\|_{1,off} \leq \sqrt{s} \|\Delta\|_{F,off}$ . Then, in (B.114) it remains to verify that the expression is positive, which is a consequence of the choice of  $\lambda$  in (B.107).  $\square$

*Proof of Lemma 3.1:* Lemma 3.1 is the analogue of Corollary 1 in Rothman et al. (2008). The proof follows similar arguments as the proof of Proposition 3.7, aiming to use Proposition 3.5.

Recall that  $\Gamma_0 = W_0^{-1} G_0 W_0^{-1}$  and  $\widehat{\Gamma}(\widehat{D}) = \widehat{W}^{-1}(\widehat{D}) \widehat{G}(\widehat{D}) \widehat{W}^{-1}(\widehat{D})$ . The first step is to prove an analogue of Proposition 3.5 in terms of the coherence matrix  $\Gamma_0$ , that is,

$$\mathbb{P}(\|\widehat{\Gamma}(\widehat{D}) - \Gamma_0\|_{\max} > \widetilde{\delta}) \leq c_1 p^{2-c_2} \mathcal{C} \quad (\text{B.115})$$

with  $\tilde{\delta} = 3\delta \max\left\{1, \frac{1}{\underline{k}^2}\right\}$  and  $\bar{\delta}$  as in Proposition 3.5. In order to prove (B.115), note that

$$\begin{aligned}
\|\widehat{\Gamma}(\widehat{D}) - \Gamma_0\|_{\max} &= \max_{r,s=1,\dots,p} \left| \frac{\widehat{G}_{rs}(\widehat{D})}{\sqrt{\widehat{G}_{rr}(\widehat{D})\widehat{G}_{ss}(\widehat{D})}} - \frac{G_{0,rs}}{\sqrt{G_{0,rr}G_{0,ss}}} \right| \\
&\leq \max_{r,s=1,\dots,p} \left| \frac{\widehat{G}_{rs}(\widehat{D})}{\sqrt{\widehat{G}_{rr}(\widehat{D})\widehat{G}_{ss}(\widehat{D})}} - \frac{\widehat{G}_{rs}(\widehat{D})}{\sqrt{G_{0,rr}G_{0,ss}}} \right| + \max_{r,s=1,\dots,p} \left| \frac{\widehat{G}_{rs}(\widehat{D}) - G_{0,rs}}{\sqrt{G_{0,rr}G_{0,ss}}} \right| \\
&\leq \max_{r,s=1,\dots,p} \left| 1 - \left( \frac{\widehat{G}_{rr}(\widehat{D})\widehat{G}_{ss}(\widehat{D})}{G_{0,rr}G_{0,ss}} \right)^{\frac{1}{2}} \right| + \max_{r,s=1,\dots,p} \left| \frac{\widehat{G}_{rs}(\widehat{D}) - G_{0,rs}}{\sqrt{G_{0,rr}G_{0,ss}}} \right| \\
&\leq \max_{r,s=1,\dots,p} \left| 1 - \frac{\widehat{G}_{rr}(\widehat{D})\widehat{G}_{ss}(\widehat{D})}{G_{0,rr}G_{0,ss}} \right| + \delta \frac{1}{\underline{k}} \tag{B.116} \\
&= \max_{r,s=1,\dots,p} (G_{0,rr}G_{0,ss})^{-1} \left| \widehat{G}_{rr}(\widehat{D})\widehat{G}_{ss}(\widehat{D}) - G_{0,rr}G_{0,ss} \right| + \delta \frac{1}{\underline{k}} \\
&\leq \max_{r,s=1,\dots,p} \frac{1}{\underline{k}^2} \left| (\widehat{G}_{rr}(\widehat{D}) - G_{0,rr} + G_{0,rr})(\widehat{G}_{ss}(\widehat{D}) - G_{0,ss} + G_{0,ss}) - G_{0,rr}G_{0,ss} \right| + \delta \frac{1}{\underline{k}} \\
&\leq \frac{1}{\underline{k}^2} (\|\widehat{G}(\widehat{D}) - G_0\|_{\max}^2 + 2 \max_{r=1,\dots,p} G_{0,rr} \|\widehat{G}(\widehat{D}) - G_0\|_{\max}) + \delta \frac{1}{\underline{k}} \\
&\leq \delta^2 \frac{1}{\underline{k}^2} + 2\delta \frac{1}{\underline{k}} \leq 3\delta \max\left\{1, \frac{1}{\underline{k}^2}\right\}, \tag{B.117}
\end{aligned}$$

where (B.116) follows since

$$\max_{r,s=1,\dots,p} \left| \frac{\widehat{G}_{rs}(\widehat{D}) - G_{0,rs}}{\sqrt{G_{0,rr}G_{0,ss}}} \right| \leq \|W_0^{-1}\|^2 \|\widehat{G}(\widehat{D}) - G_0\|_{\max} \leq \delta \frac{1}{\underline{k}} \tag{B.118}$$

with probability at least  $1 - c_1 p^{2-c_2 C}$  by Proposition 3.5 and since  $\|W_0^{-1}\| \leq 1/\lambda_{\min}^{\frac{1}{2}}(G_0) \leq 1/\underline{k}^{\frac{1}{2}}$  by Assumption 1. The inequality (B.117) is also due to Proposition 3.5. This concludes the proof of (B.115).

Moving on to proving the actual statement of Lemma 3.1 we follow Rothman et al. (2008) and Shu and Nan (2019). The proof requires a slight modification of the proof of Proposition 3.7. We conduct the same steps but use the objective function  $\ell_\rho^\Gamma$  in (3.25); set  $\Delta = \widehat{K} - K_0$  and replace (B.107) by

$$\Theta_N = \{\Delta = \Delta^* \mid \|\Delta\|_F = \lambda\} \quad \text{and} \quad \lambda = 48 \max\left\{1, \frac{1}{\underline{k}^4}\right\} \sqrt{s} \delta. \tag{B.119}$$

We omit all steps up to (B.109). Then,

$$\begin{aligned}
Z(\Delta) &\geq \text{tr}((\widehat{\Gamma}(\widehat{D}) - \Gamma_0)\Delta) + \|\Delta\|_F^2 \frac{1}{8} \underline{k}^2 + \rho(\|\Delta + K_0\|_{1,off} - \|K_0\|_{1,off}) \\
&\geq \|\Delta\|_F^2 \frac{1}{8} \underline{k}^2 - \tilde{\delta} \|\Delta\|_1 + \rho(\|\Delta + K_0\|_{1,off} - \|K_0\|_{1,off}) \tag{B.120}
\end{aligned}$$

$$\geq \|\Delta\|_F^2 \frac{1}{8} \underline{k}^2 - \tilde{\delta} \|\Delta\|_1 + \rho(-\|\Delta_S\|_{1,off} + \|\Delta_{S^c}\|_{1,off}) \tag{B.121}$$

$$\geq \|\Delta\|_F^2 \frac{1}{8} \underline{k}^2 - \tilde{\delta} (\|\Delta\|_{1,on} + \|\Delta_S\|_{1,off} + \|\Delta_{S^c}\|_{1,off})$$

$$\begin{aligned}
& + \rho(-\|\Delta_S\|_{1,off} + \|\Delta_{S^c}\|_{1,off}) \\
& \geq \|\Delta\|_{F,on}^2 \frac{1}{8} \underline{k}^2 - \tilde{\delta} \|\Delta\|_{1,on} + \|\Delta\|_{F,off}^2 \frac{1}{8} \underline{k}^2 - 2\tilde{\delta} \|\Delta_S\|_{1,off}
\end{aligned} \tag{B.122}$$

$$\geq \|\Delta\|_{F,on}^2 \left( \frac{1}{8} \underline{k}^2 - \tilde{\delta} \sqrt{s} \|\Delta\|_{F,on}^{-1} \right) + \|\Delta\|_{F,off}^2 \left( \frac{1}{8} \underline{k}^2 - 2\tilde{\delta} \sqrt{s} \|\Delta\|_{F,off}^{-1} \right) \tag{B.123}$$

$$\begin{aligned}
& \geq \|\Delta\|_{F,on}^2 \left( \frac{1}{8} \underline{k}^2 - 2\tilde{\delta} \sqrt{s} \|\Delta\|_{F,on}^{-1} \right) + \|\Delta\|_{F,off}^2 \left( \frac{1}{8} \underline{k}^2 - 2\tilde{\delta} \sqrt{s} \|\Delta\|_{F,off}^{-1} \right) \\
& \geq \|\Delta\|_{F,\rho}^2 \left( \frac{1}{8} \underline{k}^2 - \sqrt{2} \tilde{\delta} \sqrt{s} \|\Delta\|_{F,\rho}^{-1} \right) > 0.
\end{aligned} \tag{B.124}$$

Applying (B.115) gives inequality (B.120) with probability at least  $1 - c_1 p^{2-c_2 C}$ , and (B.121) follows by the same arguments as (B.111). Line (B.122) follows since  $\rho = \tilde{\delta} = 3\delta \max\left\{1, \frac{1}{\underline{k}^2}\right\}$ . In (B.123), we used the Cauchy-Schwarz inequality to get  $\|\Delta\|_{1,on} \leq \sqrt{s} \|\Delta\|_{F,on}$  and  $\|\Delta_S\|_{1,off} \leq \sqrt{s} \|\Delta\|_{F,off}$ . Then, in (B.124) it remains to verify that the expression is positive, which is a consequence of the choice of  $\lambda$  in (B.119).  $\square$

*Proof of Proposition 3.8:* The modified graphical local Whittle estimator (3.26) requires a consistency result for the estimator of  $W_0^{-1}$  which can be derived, with explanations given below, as follows

$$\begin{aligned}
\|\widehat{W}(\widehat{D})^{-1} - W_0^{-1}\| &= \max_{r=1,\dots,p} \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) - G_{0,rr}^{-\frac{1}{2}} \right| \\
&= \max_{r=1,\dots,p} \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) - G_{0,rr}^{-\frac{1}{2}} \right| \mathbf{1}_{\{|\widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) - G_{0,rr}^{-\frac{1}{2}}| \geq 1\}} + \max_{r=1,\dots,p} \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) - G_{0,rr}^{-\frac{1}{2}} \right| \mathbf{1}_{\{|\widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) - G_{0,rr}^{-\frac{1}{2}}| < 1\}} \\
&\leq \max_{r=1,\dots,p} \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) - G_{0,rr}^{-\frac{1}{2}} \right|^2 + \max_{r=1,\dots,p} \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) - G_{0,rr}^{-\frac{1}{2}} \right| \mathbf{1}_{\{|\widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) - G_{0,rr}^{-\frac{1}{2}}| < 1\}} \\
&\leq \max_{r=1,\dots,p} \left| \frac{\widehat{G}_{rr}(\widehat{D}) - G_{0,rr}}{G_{0,rr}} \right|^2 \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) \right|^2 \\
&\quad + \max_{r=1,\dots,p} \left| \frac{\widehat{G}_{rr}(\widehat{D}) - G_{0,rr}}{G_{0,rr}} \right| \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) \right| \mathbf{1}_{\{|\widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) - G_{0,rr}^{-\frac{1}{2}}| < 1\}}
\end{aligned} \tag{B.125}$$

$$\begin{aligned}
&= \max_{r=1,\dots,p} \left| \frac{\widehat{G}_{rr}(\widehat{D}) - G_{0,rr}}{G_{0,rr}} \right|^2 \left| \widehat{G}_{rr}(\widehat{D}) - G_{0,rr} + G_{0,rr} \right|^{-1} \\
&\quad + \max_{r=1,\dots,p} \left| \frac{\widehat{G}_{rr}(\widehat{D}) - G_{0,rr}}{G_{0,rr}} \right| \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) \right| \mathbf{1}_{\{|\widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) - G_{0,rr}^{-\frac{1}{2}}| < 1\}} \\
&\leq \max_{r=1,\dots,p} \left| \frac{\widehat{G}_{rr}(\widehat{D}) - G_{0,rr}}{G_{0,rr}} \right|^2 \left| \widehat{G}_{rr}(\widehat{D}) - G_{0,rr} \right|^{-1} + \max_{r=1,\dots,p} \left| \frac{\widehat{G}_{rr}(\widehat{D}) - G_{0,rr}}{G_{0,rr}^{\frac{3}{2}}} \right|^2 \\
&\quad + \max_{r=1,\dots,p} \left| \frac{\widehat{G}_{rr}(\widehat{D}) - G_{0,rr}}{G_{0,rr}} \right| \left| G_{0,rr}^{-\frac{1}{2}} + 1 \right| \\
&\leq \left( \frac{1}{\underline{k}^2} + \frac{1}{\underline{k}^3} \right) \max_{r=1,\dots,p} \left| \widehat{G}_{rr}(\widehat{D}) - G_{0,rr} \right|^2 + \left( \frac{1}{\underline{k}^{\frac{3}{2}}} + \frac{1}{\underline{k}} \right) \max_{r=1,\dots,p} \left| \widehat{G}_{rr}(\widehat{D}) - G_{0,rr} \right|
\end{aligned} \tag{B.126}$$

$$\leq \delta 4 \max \left\{ \frac{1}{\underline{k}^3}, 1 \right\}, \tag{B.127}$$

where (B.125) is due to

$$\begin{aligned} \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) - G_{0,rr}^{-\frac{1}{2}} \right| &= \left| 1 - \left( \frac{G_{0,rr}}{\widehat{G}_{rr}} \right)^{-\frac{1}{2}} \right| \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) \right| \leq \left| 1 - \left( \frac{G_{0,rr}}{\widehat{G}_{rr}} \right)^{-\frac{1}{2}} \right| \left| 1 + \left( \frac{G_{0,rr}}{\widehat{G}_{rr}} \right)^{-\frac{1}{2}} \right| \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) \right| \\ &= \left| \frac{\widehat{G}_{rr}(\widehat{D}) - G_{0,rr}}{G_{0,rr}} \right| \left| \widehat{G}_{rr}^{-\frac{1}{2}}(\widehat{D}) \right|, \end{aligned}$$

and (B.126) follows since  $\lambda_{\min}(G_0) \geq \underline{k}$  by Assumption 1. Finally, (B.127) is an application of Proposition 3.5. We can further infer

$$\|\widehat{W}(\widehat{D})^{-1}\| \leq \|\widehat{W}(\widehat{D})^{-1} - W_0^{-1}\| + \|W_0^{-1}\| \leq \delta 4 \max \left\{ \frac{1}{\underline{k}^3}, 1 \right\} + \frac{1}{\underline{k}^{\frac{1}{2}}} \leq 5 \max \left\{ \frac{1}{\underline{k}^3}, 1 \right\}. \quad (\text{B.128})$$

Finally, with further details given below,

$$\begin{aligned} &\|\widehat{P}_\rho^M(\widehat{D}) - P_0\| \\ &\leq \|\widehat{K}(\widehat{D}) - K_0\|_F (\|\widehat{W}(\widehat{D})^{-1} - W_0^{-1}\|^2 + \|W_0^{-1}\| \|\widehat{W}(\widehat{D})^{-1}\|) \\ &\quad + \|\widehat{W}(\widehat{D})^{-1} - W_0^{-1}\| \left( (\|\widehat{K}(\widehat{D}) - K_0\|_F + \|K_0\|) \|W_0^{-1}\| + \|K_0\| \|\widehat{W}(\widehat{D})^{-1}\| \right) \end{aligned} \quad (\text{B.129})$$

$$\begin{aligned} &\leq 48 \max \left\{ \frac{1}{\underline{k}^3}, 1 \right\} \sqrt{s} \delta \left( 16 \max \left\{ \frac{1}{\underline{k}^6}, 1 \right\} + 5 \max \left\{ \frac{1}{\underline{k}^4}, 1 \right\} \right) \\ &\quad + 4\delta \max \left\{ \frac{1}{\underline{k}^3}, 1 \right\} \left( \left( 48 \max \left\{ \frac{1}{\underline{k}^3}, 1 \right\} \sqrt{s} \delta + \|K_0\| \right) \frac{1}{\underline{k}^{\frac{1}{2}}} + \|K_0\| 5 \max \left\{ \frac{1}{\underline{k}^3}, 1 \right\} \right) \end{aligned} \quad (\text{B.130})$$

$$\leq 30 * 48 \max \left\{ \frac{1}{\underline{k}^9}, 1 \right\} \max\{1, \|K_0\|\} \sqrt{s} \delta$$

with probability at least  $1 - c_1 p^{2-c_2 C}$ . The relation (B.129) is due to (S.52) in Shu and Nan (2019). Then, combining Lemma 3.1, (B.127) and (B.128), we get (B.130).  $\square$

*Proof of Proposition 3.9:* We follow the proof of Theorem 6 in Cai et al. (2011). Note that Theorem 6 in Cai et al. (2011) only requires that the maximum norm of the difference between covariance matrix and its estimator is controlled. Given that our Proposition 3.5 provides an analogue result in the spectral domain, namely

$$\|\widehat{G}(\widehat{D}) - G_0\|_{\max} \leq \delta \quad (\text{B.131})$$

with probability at least  $1 - c_1 p^{2-c_2 C}$ , we will only provide some of the key steps in the proof and otherwise refer to Theorem 6 in Cai et al. (2011). Following the proof of Theorem 6 in Cai et al. (2011) up to the point where equation (13) is established, we get

$$\|\widehat{\Theta}_\rho - P_0\|_{\max} \leq 4\delta \|P_0\|_1 \quad (\text{B.132})$$

by setting  $\Sigma_0 = G_0$  and  $\Omega_0 = P_0$  in Cai et al. (2011). Then, one can further infer that

$$\|\widehat{P}_\rho^C(\widehat{D}) - P_0\| \leq (1 + 2^{1-a} + 3^{1-a}) \|\widehat{\Theta}_\rho - P_0\|_{\max}^{1-a} \|P_0\|_a^a \quad (\text{B.133})$$

$$\leq 6(4\delta \|P_0\|_1)^{1-a} \|P_0\|_a^a, \quad (\text{B.134})$$

where (B.133) follows by (27) in Cai et al. (2011) and (B.134) by applying (B.132).  $\square$

*Proof of Proposition 3.10:* We prove (3.28) and (3.29) separately.

*Proof of (3.28):* By the definition of the thresholding operator (2.2), we get

$$\begin{aligned} \{(r, s) \mid T_\rho(\widehat{G}_{rs}(\widehat{D})) \neq 0, G_{0,rs} = 0\} &= \{(r, s) \mid |\widehat{G}_{rs}(\widehat{D})| \geq \rho, G_{0,rs} = 0\} \\ &\subseteq \{(r, s) \mid |\widehat{G}_{rs}(\widehat{D}) - G_{0,rs}| \geq \rho\}. \end{aligned}$$

Therefore,

$$\mathbb{P} \left( \sum_{r,s=1}^d \mathbb{1}_{\{T_\rho(\widehat{G}_{rs}(\widehat{D})) \neq 0, G_{0,rs} = 0\}} > 0 \right) \leq \mathbb{P} \left( \max_{r,s=1,\dots,d} |\widehat{G}_{rs}(\widehat{D}) - G_{0,rs}| > \rho \right) \leq c_1 p^{2-c_2 \mathcal{C}} \quad (\text{B.135})$$

by Proposition 3.5.

*Proof of (3.29):* By the definition of the thresholding operator (2.2), we get

$$\begin{aligned} \{(r, s) \mid T_\rho(\widehat{G}_{rs}(\widehat{D})) \leq 0, G_{0,rs} > 0 \text{ or } T_\rho(\widehat{G}_{rs}(\widehat{D})) \geq 0, G_{0,rs} < 0\} \\ \subseteq \{(r, s) \mid -\widehat{G}_{rs}(\widehat{D}) > \rho, G_{0,rs} > \tau \text{ or } \widehat{G}_{rs}(\widehat{D}) > \rho, -G_{0,rs} > \tau\} \\ \subseteq \{(r, s) \mid |\widehat{G}_{rs}(\widehat{D}) - G_{0,rs}| > \tau + \rho\}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{P} \left( \sum_{r,s=1}^d \mathbb{1}_{\{T_\rho(\widehat{G}_{rs}(\widehat{D})) \leq 0, G_{0,rs} > 0 \text{ or } T_\rho(\widehat{G}_{rs}(\widehat{D})) \geq 0, G_{0,rs} < 0\}} > 0 \right) \\ \leq \mathbb{P} \left( \max_{r,s=1,\dots,d} |\widehat{G}_{rs}(\widehat{D}) - G_{0,rs}| > \tau + \rho \right) \leq c_1 p^{2-c_2 \mathcal{C}} \end{aligned}$$

by Proposition 3.5. □

*Proof of Proposition 3.11:* We omit the proof and as it is similar to the proof of Proposition 3.10. □

## C Some technical results and their proofs

Section C.1 provides probabilistic bounds which are used to show that Lemmas B.1 and B.2 are sufficient to prove a concentration inequality on the deviation between  $\widehat{G}(\widehat{D})$  and the true  $G_0$ . Section C.2 gives some non-asymptotic results on the bias of the periodogram used to prove Lemma B.2. Finally, Section C.3 concerns results on the matrix norms of the covariance matrix used to prove Lemma B.1.

### C.1 Probabilistic bounds

We prove here some probabilistic bounds used to show that the proof of Proposition 3.5 can be reduced to proving Propositions 3.1–3.4. We continue using the notation of the proof of Proposition 3.5.

**Lemma C.1.** *For  $\eta \in (0, \frac{1}{2})$ , let  $N_\eta = \{d \mid |d - d_{0,r}| \leq \eta\}$  and  $\Theta_1$  as in (3.22). Then,*

$$\begin{aligned} \mathbb{P} \left( \inf_{d \in N_\eta^c \cap \Theta_1} S_r(d) < 0 \right) &\leq \mathbb{P} \left( \frac{1}{2} L(\Delta_2 - \Delta_1) g_{0,r} \eta^2 \mathcal{V}_1(m) < \sup_{d \in \Theta_1} |\widehat{h}_r(d) - \widetilde{h}_r(d)| \right) \\ &\quad + \mathbb{P} \left( \frac{1}{2} g_{0,r} \eta^2 \mathcal{V}_1(m) < |\widehat{g}_r(d_{0,r}) - g_{0,r}| \right), \end{aligned}$$

where  $S_r(d) = R_r(d) - R_r(d_{0,r})$  with  $R_r$  in (2.1),  $L(\cdot)$  in (3.3),  $\mathcal{V}_1(m)$  in (3.21) and

$$\widehat{h}_r(d) = \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_{0,r}} \lambda_j^{2d_{0,r}} I_{X,rr}(\lambda_j), \quad \widetilde{h}_r(d) = \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_{0,r}} g_{0,r}. \quad (\text{C.1})$$

*Proof:* Let  $\widetilde{S}_r(d) = \widetilde{R}_r(d) - \widetilde{R}_r(d_{0,r})$  with  $\widetilde{R}_r(d) = \frac{1}{m} \sum_{j=1}^m \log(\lambda_j^{-2d} \widetilde{g}_r(d))$  and  $\widetilde{g}_r(d)$  defined below (3.2). Then,

$$\begin{aligned} \mathbb{P}\left(\inf_{d \in N_\eta^c \cap \Theta_1} S_r(d) < 0\right) &\leq \mathbb{P}\left(\inf_{N_\eta^c \cap \Theta_1} \widetilde{S}_r(d) < \sup_{d \in \Theta_1} |\widetilde{S}_r(d) - S_r(d)|\right) \\ &\leq \mathbb{P}\left(\eta^2 \mathcal{V}_1(m) < \sup_{d \in \Theta_1} |\widetilde{S}_r(d) - S_r(d)|\right) \end{aligned} \quad (\text{C.2})$$

$$\begin{aligned} &\leq \mathbb{P}\left(\frac{1}{2} L(\Delta_2 - \Delta_1) g_{0,r} \eta^2 \mathcal{V}_1(m) < \sup_{d \in \Theta_1} |\widehat{h}_r(d) - \widetilde{h}_r(d)|\right) \\ &\quad + \mathbb{P}\left(\frac{1}{2} g_{0,r} \eta^2 \mathcal{V}_1(m) < |\widehat{g}_r(d_{0,r}) - g_{0,r}|\right), \end{aligned} \quad (\text{C.3})$$

where (C.2) is proved in Lemma C.2 and (C.3) in Lemma C.4 which is applicable since  $\eta^2 \mathcal{V}_1(m) < \frac{1}{4} \mathcal{V}_1(m) = \frac{1}{4} \frac{1}{18} (1 - \frac{1}{m^2}) < 1$ .  $\square$

**Lemma C.2.** *For  $d \in N_\eta^c \cap \Theta_1$ ,  $\eta \in (0, \frac{1}{2})$ , the quantity  $\widetilde{S}_r(d) = \widetilde{R}_r(d) - \widetilde{R}_r(d_{0,r})$  can be bounded from below as*

$$\widetilde{S}_r(d) \geq \eta^2 \mathcal{V}_1(m)$$

with

$$\mathcal{V}_1(m) = \frac{1}{3} \frac{1}{m^4} \sum_{i,j=1}^m (i-j)^2$$

satisfying  $\mathcal{V}_1(m) = \frac{1}{18} (1 - \frac{1}{m^2}) \sim \frac{1}{18}$  as  $m \rightarrow \infty$ .

*Proof:* The expression  $\widetilde{S}_r(d)$  can be rewritten as

$$\begin{aligned} \widetilde{S}_r(d) &= \frac{1}{m} \sum_{j=1}^m \log(\lambda_j^{-2d}) + \log(\widetilde{g}_r(d)) - \frac{1}{m} \sum_{j=1}^m \log(\lambda_j^{-2d_{0,r}}) - \log(g_{0,r}) \\ &= \log\left(\frac{1}{m} \sum_{j=1}^m \lambda_j^{2d-2d_{0,r}}\right) - \frac{1}{m} \sum_{j=1}^m \log(\lambda_j^{2d-2d_{0,r}}) \\ &= \log\left(\frac{1}{m} \sum_{j=1}^m j^{2d-2d_{0,r}}\right) - \frac{1}{m} \sum_{j=1}^m \log(j^{2d-2d_{0,r}}) =: f_m(d - d_{0,r}). \end{aligned}$$

By Lemma C.3 below, we know that

$$f_m(d - d_{0,r}) \geq \min\{f_m(-\eta), f_m(\eta)\}.$$

We further prove lower bounds for  $f_m(\eta)$  and  $f_m(-\eta)$ . The proof is based on a result due to Mercer (1999). For the readers' convenience, we shortly repeat this result: Let  $J(x)$  be the smallest closed interval that contains some  $x_j$ ,  $j = 1, \dots, m$  and  $g_1, g_2$  be two twice differentiable functions on  $J(x)$  with continuous second derivatives  $g_1'', g_2''$ . Then,

$$\frac{g_1\left(\frac{1}{m} \sum_{j=1}^m x_j\right) - \frac{1}{m} \sum_{j=1}^m g_1(x_j)}{g_2\left(\frac{1}{m} \sum_{j=1}^m x_j\right) - \frac{1}{m} \sum_{j=1}^m g_2(x_j)} = \frac{g_1''(\xi)}{g_2''(\xi)} \text{ for some } \xi \in J(x), \quad (\text{C.4})$$

given  $g_2''(\xi) \neq 0$ ; see p. 678 in Mercer (1999)<sup>1</sup>. From here on, we consider  $f_m(\eta)$  and  $f_m(-\eta)$  separately in order to find lower bounds.

*Bounding  $f_m(\eta)$ :* In (C.5) below, we apply (C.4) with  $(x_1, \dots, x_m) = (1, 2^{2\eta}, \dots, m^{2\eta})$  and  $g_1(x) = \log(x)$ ,  $g_2(x) = x^2$ . The second derivatives are respectively  $g_1''(x) = -x^{-2}$  and  $g_2''(x) = 2$  such that  $\frac{g_1''(x)}{g_2''(x)} = -\frac{x^{-2}}{2}$ . Then, there is a  $\xi \in [1, m^{2\eta}]$  such that

$$\begin{aligned} f_m(\eta) &= \log\left(\frac{1}{m} \sum_{j=1}^m j^{2\eta}\right) - \frac{1}{m} \sum_{i=1}^m \log(i^{2\eta}) \\ &= -\frac{1}{2\xi^2} \left( \left(\frac{1}{m} \sum_{j=1}^m j^{2\eta}\right)^2 - \frac{1}{m} \sum_{j=1}^m j^{4\eta} \right) \end{aligned} \quad (\text{C.5})$$

$$\begin{aligned} &\geq \frac{1}{2m^{4\eta}} \left( \frac{1}{m} \sum_{j=1}^m j^{4\eta} - \left(\frac{1}{m} \sum_{j=1}^m j^{2\eta}\right)^2 \right) \\ &\geq \frac{1}{m^4} \sum_{i,j=1}^m (i-j)^2 \eta^2, \end{aligned} \quad (\text{C.6})$$

where the last inequality (C.6) is due to (C.10) and proved below.

*Bounding  $f_m(-\eta)$ :* In (C.7) below, we apply (C.4) with  $(x_1, \dots, x_m) = (1, 2^{-2\eta}, \dots, m^{-2\eta})$  and  $g_1(x) = \log(x)$ ,  $g_2(x) = x^{-2}$ . The second derivatives are respectively  $g_1''(x) = -x^{-2}$  and  $g_2''(x) = 6x^{-4}$  such that  $\frac{g_1''(x)}{g_2''(x)} = -\frac{x^2}{6}$ . Then, there is a  $\xi \in [m^{-2\eta}, 1]$  such that

$$\begin{aligned} f_m(-\eta) &= \log\left(\frac{1}{m} \sum_{j=1}^m j^{-2\eta}\right) - \frac{1}{m} \sum_{i=1}^m \log(i^{-2\eta}) \\ &= -\frac{\xi^2}{6} \left( \left(\frac{1}{m} \sum_{j=1}^m j^{-2\eta}\right)^{-2} - \frac{1}{m} \sum_{j=1}^m j^{4\eta} \right) \end{aligned} \quad (\text{C.7})$$

$$\begin{aligned} &\geq \frac{m^{-4\eta}}{6} \left( \frac{1}{m} \sum_{j=1}^m j^{4\eta} - \left(\frac{1}{m} \sum_{j=1}^m j^{-2\eta}\right)^{-2} \right) \\ &\geq \frac{1}{6m^{4\eta}} \left( \frac{1}{m} \sum_{j=1}^m j^{4\eta} - \left(\frac{1}{m} \sum_{j=1}^m j^{2\eta}\right)^2 \right) \end{aligned} \quad (\text{C.8})$$

$$\geq \frac{1}{3} \frac{1}{m^4} \sum_{i,j=1}^m (i-j)^2 \eta^2, \quad (\text{C.9})$$

where (C.8) is due to Jensen's inequality and (C.9) is proved in (C.10) below.

For the inequality in (C.6) and (C.9), note that for  $d \in N_\eta^c \cap \Theta_1$  with  $\eta \in (0, \frac{1}{2})$ ,

$$\begin{aligned} \frac{1}{m^{4\eta}} \left( \frac{1}{m} \sum_{j=1}^m j^{4\eta} - \left(\frac{1}{m} \sum_{j=1}^m j^{2\eta}\right)^2 \right) &= \frac{1}{2m^{4\eta}} \frac{1}{m^2} \sum_{i,j=1}^m (i^{2\eta} - j^{2\eta})^2 \\ &= \frac{1}{2m^{4\eta}} \frac{1}{m^2} \sum_{i,j=1}^m ((i-j)(2\eta)x_{ij}^{2\eta-1})^2 \end{aligned} \quad (\text{C.10})$$

---

<sup>1</sup>The relation (C.4) is also akin to Cauchy's mean value theorem.



$$\geq 2 \frac{1}{m^4} \sum_{i,j=1}^m (i-j)^2 \eta^2,$$

where  $x_{ij} \in (i, j)$  in (C.10) after applying the mean value theorem.  $\square$

**Lemma C.3.** For  $d \in N_\eta^c \cap \Theta_1$ , the following relation holds,

$$f_m(d - d_{0,r}) \geq \min\{f_m(-\eta), f_m(\eta)\}$$

with

$$f_m(x) = \log\left(\frac{1}{m} \sum_{j=1}^m j^{2x}\right) - \frac{1}{m} \sum_{j=1}^m \log(j^{2x}).$$

*Proof:* The first and second derivatives of the function  $f_m$  are given by

$$\begin{aligned} \frac{\partial}{\partial x} f_m(x) &= \left(\frac{1}{m} \sum_{j=1}^m j^{2x}\right)^{-1} 2 \frac{1}{m} \sum_{j=1}^m \log(j) j^{2x} - 2 \frac{1}{m} \sum_{j=1}^m \log(j), \\ \frac{\partial^2}{\partial x^2} f_m(x) &= -\left(\frac{1}{m} \sum_{j=1}^m j^{2x}\right)^{-2} \left(2 \frac{1}{m} \sum_{j=1}^m \log(j) j^{2x}\right)^2 + \left(\frac{1}{m} \sum_{j=1}^m j^{2x}\right)^{-1} 4 \frac{1}{m} \sum_{j=1}^m (\log(j))^2 j^{2x}. \end{aligned}$$

Furthermore, the second derivative satisfies  $\frac{\partial^2}{\partial x^2} f_m(x) > 0$  for all  $x$  since

$$\left(\frac{1}{m} \sum_{j=1}^m \log(j) j^{2x}\right)^2 < \frac{1}{m} \sum_{j=1}^m (\log(j))^2 j^{2x} \frac{1}{m} \sum_{j=1}^m j^{2x} \quad (\text{C.11})$$

using Cauchy-Schwarz inequality. The two sides in (C.11) cannot be equal since the vectors  $(0, \log(2)2^x, \dots, \log(m)m^x)$  and  $(1, 2^x, \dots, m^x)$  are linearly independent. The facts that  $f_m$  is non-negative (by Jensen's inequality), is zero at  $x = 0$  and has a positive second derivative prove our claim.  $\square$

**Lemma C.4.** For  $d \in \Theta_1$  and all  $\nu \in (0, 1)$ , the inclusion

$$\begin{aligned} &\{|\tilde{S}_r(d) - S_r(d)| > \nu\} \\ &\subseteq \left\{|\hat{h}_r(d) - \tilde{h}_r(d)| > \nu \frac{1}{2} L(\Delta_2 - \Delta_1) g_{0,r}\right\} \cup \left\{|\hat{g}_r(d_{0,r}) - g_{0,r}| > \nu \frac{1}{2} g_{0,r}\right\} \end{aligned}$$

is satisfied with  $\hat{h}_r(d), \tilde{h}_r(d)$  in (C.1) and  $L(\cdot)$  in (3.3).

*Proof:* The distance between  $\tilde{S}_r(d)$  and  $S_r(d)$  can be written as

$$\begin{aligned} |\tilde{S}_r(d) - S_r(d)| &= \left| \frac{1}{m} \sum_{j=1}^m \log(\lambda_j^{-2d}) + \log(\tilde{g}_r(d)) - \frac{1}{m} \sum_{j=1}^m \log(\lambda_j^{-2d_{0,r}}) - \log(g_{0,r}) \right. \\ &\quad \left. - \left( \frac{1}{m} \sum_{j=1}^m \log(\lambda_j^{-2d}) + \log(\hat{g}_r(d)) - \frac{1}{m} \sum_{j=1}^m \log(\lambda_j^{-2d_{0,r}}) - \log(\hat{g}_r(d_{0,r})) \right) \right| \\ &= |\log(\tilde{g}_r(d)) - \log(g_{0,r}) - \log(\hat{g}_r(d)) + \log(\hat{g}_r(d_{0,r}))| \\ &\leq |\log(\hat{g}_r(d)(\tilde{g}_r(d))^{-1})| + |\log(\hat{g}_r(d_{0,r})g_{0,r}^{-1})|. \end{aligned} \quad (\text{C.12})$$

We consider only  $|\log(\widehat{g}_r(d)(\widetilde{g}_r(d))^{-1})|$  in (C.12), since  $|\log(\widehat{g}_r(d_{0,r})g_{0,r}^{-1})|$  can be treated analogously. Then, with explanation given below,

$$\begin{aligned} & \{|\log(\widehat{g}_r(d)(\widetilde{g}_r(d))^{-1})| > \nu\} \\ &= \{|\log(\widehat{g}_r(d)) - \log(\widetilde{g}_r(d))| > \nu\} \\ &= \{\log(\widehat{g}_r(d)) - \log(\widetilde{g}_r(d)) > \nu\} \cup \{\log(\widetilde{g}_r(d)) - \log(\widehat{g}_r(d)) > \nu\} \\ &\subseteq \left\{ \frac{\widehat{g}_r(d) - \widetilde{g}_r(d)}{\widetilde{g}_r(d)} > \nu \right\} \cup \left\{ \frac{\widetilde{g}_r(d) - \widehat{g}_r(d)}{\widehat{g}_r(d)} > \nu \right\} \end{aligned} \quad (\text{C.13})$$

$$= \left\{ \frac{\widehat{h}_r(d) - \widetilde{h}_r(d)}{\widetilde{h}_r(d)} > \nu \right\} \cup \left\{ \frac{\widetilde{h}_r(d) - \widehat{h}_r(d)}{\widehat{h}_r(d)} > \nu \right\} \quad (\text{C.14})$$

$$\begin{aligned} &\subseteq \{\widehat{h}_r(d) - \widetilde{h}_r(d) > \nu L(\Delta_2 - \Delta_1)g_{0,r}\} \cup \left( \{\widetilde{h}_r(d) - \widehat{h}_r(d) > \nu \widehat{h}_r(d)\} \right. \\ &\quad \left. \cap \left( \{\widehat{h}_r(d) \geq \frac{1}{2}L(\Delta_2 - \Delta_1)g_{0,r}\} \cup \{\widehat{h}_r(d) < \frac{1}{2}L(\Delta_2 - \Delta_1)g_{0,r}\} \right) \right) \end{aligned} \quad (\text{C.15})$$

$$\begin{aligned} &= \{\widehat{h}_r(d) - \widetilde{h}_r(d) > \nu L(\Delta_2 - \Delta_1)g_{0,r}\} \cup \{\widetilde{h}_r(d) - \widehat{h}_r(d) > \frac{1}{2}\nu L(\Delta_2 - \Delta_1)g_{0,r}\} \\ &\quad \cup \left( \{\widetilde{h}_r(d) - \widehat{h}_r(d) \geq \nu \widehat{h}_r(d)\} \cap \{\widehat{h}_r(d) < \frac{1}{2}L(\Delta_2 - \Delta_1)g_{0,r}\} \right) \\ &\subseteq \{\widehat{h}_r(d) - \widetilde{h}_r(d) > \nu L(\Delta_2 - \Delta_1)g_{0,r}\} \cup \{\widetilde{h}_r(d) - \widehat{h}_r(d) > \frac{1}{2}\nu L(\Delta_2 - \Delta_1)g_{0,r}\} \\ &\quad \cup \{\widetilde{h}_r(d) < \frac{1}{2}L(\Delta_2 - \Delta_1)g_{0,r} + \widetilde{h}_r(d) - \widehat{h}_r(d)\} \\ &\subseteq \{|\widehat{h}_r(d) - \widetilde{h}_r(d)| > \frac{1}{2}\nu L(\Delta_2 - \Delta_1)g_{0,r}\} \cup \{|\widehat{h}_r(d) - \widetilde{h}_r(d)| > \frac{1}{2}L(\Delta_2 - \Delta_1)g_{0,r}\} \\ &\subseteq \{|\widehat{h}_r(d) - \widetilde{h}_r(d)| > \frac{1}{2}L(\Delta_2 - \Delta_1)g_{0,r}\}, \end{aligned}$$

where (C.13) follows by the mean value theorem. The equality in (C.14) can be seen by the definitions of  $\widehat{h}_r$  and  $\widetilde{h}_r$  in (C.1) and noting that  $\lambda_m^{2d-2d_{0,r}}\widehat{h}_r(d) = \widehat{g}_r(d)$  and  $\lambda_m^{2d-2d_{0,r}}\widetilde{h}_r(d) = \widetilde{g}_r(d)$ . The relation (C.15) is a consequence of the lower bound

$$\widetilde{h}_r(d) = \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_{0,r}} g_{0,r} \geq \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2(\Delta_2-\Delta_1)} g_{0,r} \geq \int_0^1 x^{2(\Delta_2-\Delta_1)} dx g_{0,r} =: L(\Delta_2 - \Delta_1)g_{0,r}$$

since  $x^{2d}$  is monotonically increasing. □

**Lemma C.5.** *The second probability in (B.100) can be bounded as*

$$\mathbb{P}(\inf_{d \in \Theta_2} S_r(d) < 0) \leq \mathbb{P}\left(\left|\frac{1}{m} \sum_{j=1}^m (l_j - 1) \left(\left(\frac{j}{\ell}\right)^{2d_{0,r}} I_{X,rr}(\lambda_j) - g_{0,r}\right)\right| \mathbf{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} > g_{0,r} \mathcal{V}_2(m)\right) \quad (\text{C.16})$$

with  $l_j$  and  $\ell$  as in (3.7) and

$$\mathcal{V}_2(m) = \frac{1}{m} \sum_{j=1}^m (l_j - 1),$$

satisfying  $\mathcal{V}_2(m) > 0$ .

*Proof:* The proof of relation (C.16) is part of the proof of Theorem 1 in Robinson (1995b). More precisely, one can infer the inequality (C.16) from equation (3.21) in Robinson (1995b). For completeness, we sketch the proof. Recall the definitions of  $\ell$  and  $l_j$  given in (3.7). Then,

$$\begin{aligned} & \mathbb{P}(\inf_{d \in \Theta_2} S_r(d) < 0) \\ & \leq \mathbb{P}\left(\frac{1}{m} \sum_{j=1}^m (l_j - 1) j^{2d_{0,r}} I_{X,rr}(\lambda_j) \mathbf{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} > 0\right) \end{aligned} \quad (\text{C.17})$$

$$\leq \mathbb{P}\left(\left|\frac{1}{m} \sum_{j=1}^m (l_j - 1) \left(\left(\frac{j}{\ell}\right)^{2d_{0,r}} I_{X,rr}(\lambda_j) - g_{0,r}\right)\right| \mathbf{1}_{\{d_{0,r} \geq \Delta_1 + \frac{1}{2}\}} > \frac{1}{m} \sum_{j=1}^m (l_j - 1) g_{0,r}\right), \quad (\text{C.18})$$

where (C.17) is due to (3.21) in Robinson (1995b) and (C.18) results from division with  $\ell^{2d_{0,r}}$  and subtracting  $\frac{1}{m} \sum_{j=1}^m (l_j - 1) g_{0,r}$  on both sides in (C.17). Finally,  $\mathcal{V}_2(m) > 0$  is proved in Lemma C.7.  $\square$

**Lemma C.6.** *Given  $\|\widehat{D} - D_0\|_{\max} \leq \varepsilon$ , the population quantity (3.2) evaluated at  $\widehat{D}$  can be bounded in terms of  $\widehat{D}$  by*

$$\|\widetilde{G}(\widehat{D}) - G_0\|_{\max} \leq 2\|\widehat{D} - D_0\|_{\max} \|G_0\| \log(N) \lambda_m^{-2\varepsilon} L(-\varepsilon),$$

where  $L(\cdot)$  is defined in (3.3).

*Proof:* Note that  $\widehat{d}_r - d_{0,r} \in [-\varepsilon, \varepsilon]$  for all  $r = 1, \dots, p$ . Then,

$$\begin{aligned} \|\widetilde{G}(\widehat{D}) - G_0\|_{\max} &= \max_{r,s=1,\dots,p} \left| \frac{1}{m} \sum_{j=1}^m (\lambda_j^{(\widehat{d}_r - d_{0,r}) + (\widehat{d}_s - d_{0,s})} - 1) G_{0,rs} \right| \\ &\leq 2 \max_{r=1,\dots,p} |\widehat{d}_r - d_{0,r}| \|G_0\| \frac{1}{m} \sum_{j=1}^m |\log(\lambda_j)| \lambda_j^{-2\varepsilon} \end{aligned} \quad (\text{C.19})$$

$$\begin{aligned} &\leq 2\|\widehat{D} - D_0\|_{\max} \|G_0\| \log(N) \lambda_m^{-2\varepsilon} \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{-2\varepsilon} \\ &\leq 2\|\widehat{D} - D_0\|_{\max} \|G_0\| \log(N) \lambda_m^{-2\varepsilon} \int_0^1 x^{-2\varepsilon} dx. \end{aligned} \quad (\text{C.20})$$

The relation (C.19) is a consequence of applying the mean value theorem while (C.20) follows since  $x^{-2\varepsilon}$  is monotonically decreasing.  $\square$

**Lemma C.7.** *The quantity  $\mathcal{V}_2(m)$  in (3.21) satisfies  $\mathcal{V}_2(m) > 0$ .*

*Proof:* Recall  $\mathcal{V}_2(m)$  in (3.21) and note that, with  $l_j$  as in (3.7), it can be written as

$$\mathcal{V}_2(m) = \frac{1}{m} \sum_{j=1}^m (l_j - 1) = \frac{1}{m} \sum_{j=1}^{\ell} \left( \left(\frac{j}{\ell}\right)^{2(-\frac{1}{2} + \Delta)} - 1 \right) + \frac{1}{m} \sum_{j=\ell+1}^m \left( \left(\frac{j}{\ell}\right)^{2(\Delta_1 - d_{0,r})} - 1 \right). \quad (\text{C.21})$$

We will prove that both summands are positive. The first one satisfies

$$\frac{1}{m} \sum_{j=1}^{\ell} \left( \left(\frac{j}{\ell}\right)^{2(-\frac{1}{2} + \Delta)} - 1 \right) > 0 \quad \text{if} \quad \frac{1}{\ell} \sum_{j=1}^{\ell} j^{2(-\frac{1}{2} + \Delta)} - \ell^{2(-\frac{1}{2} + \Delta)} > 0.$$

Set  $\alpha = 2(-\frac{1}{2} + \Delta)$  and note that  $\alpha < 0$ . Then,

$$\frac{1}{\ell} \sum_{j=1}^{\ell} j^{\alpha} - \ell^{\alpha} \geq \frac{1}{\ell} \sum_{j=1}^{\ell} j^{\alpha} - \left( \frac{1}{\ell} \sum_{j=1}^{\ell} j \right)^{\alpha} \quad (\text{C.22})$$

$$\begin{aligned} &= \frac{\alpha(\alpha-1)\xi^{\alpha-2}}{2} \left( \frac{1}{\ell} \sum_{j=1}^{\ell} j^2 - \left( \frac{1}{\ell} \sum_{j=1}^{\ell} j \right)^2 \right) \\ &\geq \frac{\alpha(\alpha-1)\ell^{\alpha-2}}{4} \frac{1}{\ell^2} \sum_{i,j=1}^{\ell} (i-j)^2 > 0, \end{aligned} \quad (\text{C.23})$$

where (C.22) follows since  $\ell \geq \frac{\ell+1}{2}$  and (C.23) is due to the result stated in (C.4) with  $g_1(x) = x^{\alpha}$ ,  $g_2(x) = x^2$  and  $\xi \in [1, \ell]$ . The second summand in (C.21) satisfies

$$\frac{1}{m} \sum_{j=\ell+1}^m \left( \left( \frac{j}{\ell} \right)^{2(\Delta_1 - d_{0,r})} - 1 \right) > 0 \quad \text{if} \quad \frac{1}{m-\ell} \sum_{j=\ell+1}^m j^{2(\Delta_1 - d_{0,r})} - \ell^{2(\Delta_1 - d_{0,r})} > 0.$$

Set  $\beta = 2(\Delta_1 - d_{0,r})$  and note that  $\beta < 0$ . Then,

$$\frac{1}{m-\ell} \sum_{j=\ell+1}^m j^{\beta} - \ell^{\beta} \geq \frac{1}{m-\ell} \sum_{j=\ell+1}^m j^{\beta} - \left( \frac{1}{m-\ell} \sum_{j=\ell+1}^m j \right)^{\beta} \quad (\text{C.24})$$

$$\begin{aligned} &= \frac{\alpha(\beta-1)\xi^{\beta-2}}{2} \left( \frac{1}{m-\ell} \sum_{j=\ell+1}^m j^2 - \left( \frac{1}{m-\ell} \sum_{j=\ell+1}^m j \right)^2 \right) \\ &\geq \frac{\beta(\beta-1)m^{\beta-2}}{4} \frac{1}{(m-\ell)^2} \sum_{j=\ell+1}^m (i-j)^2 > 0, \end{aligned} \quad (\text{C.25})$$

where (C.24) follows since  $3\ell > m+1$  which can be shown by induction principal. The inequality (C.25) follows by applying (C.4) with  $g_1(x) = x^{\beta}$ ,  $g_2(x) = x^2$  and  $\xi \in [\ell+1, m]$ .  $\square$

## C.2 Bound for periodogram bias

We prove here the results used in the proof of Lemma B.2.

**Lemma C.8.** *Suppose Assumptions 1 and 3. Then,*

$$\left| f_{rs}(\lambda_j) - \mathbb{E}(I_{X,rs}(\lambda_j)) \right| \leq N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} \left( \left\| G \right\| \frac{72(\cos(\lambda_m/2))^{-2}}{\pi(1+2\min\{\Delta_1, -\Delta_2\})} + \mathbf{c}_{G,2} 4(2+\log(m)) \right).$$

*Proof:* In order to bound the bias of the periodogram, we follow the ideas in the proof of Theorem 2 in Robinson (1995a). We adapt its arguments of asymptotic nature to find non-asymptotic bounds which still ensure that the bias is negligible.

The bias term of the periodogram can be written as

$$\left| f_{rs}(\lambda_j) - \mathbb{E}(I_{X,rs}(\lambda_j)) \right| = \left| \int_{-\pi}^{\pi} \left( f_{rs}(\lambda) - f_{rs}(\lambda_j) \right) K_N(\lambda - \lambda_j) d\lambda \right|, \quad (\text{C.26})$$

where

$$K_N(\lambda) = \frac{1}{2\pi N} \left( \frac{\sin(N\lambda/2)}{\sin(\lambda/2)} \right)^2$$

is the Fejér kernel (see equation (4.3) in Robinson (1995a)). The Fejér kernel can be expressed in terms of the Dirichlet kernel as

$$K_N(\lambda) = \frac{1}{2\pi N} |D_N(\lambda)|^2 \quad \text{with} \quad D_N(\lambda) = \sum_{n=1}^N e^{in\lambda} = \frac{\sin(N\lambda/2)}{\sin(\lambda/2)}, \quad (\text{C.27})$$

which satisfies

$$|D_N(\lambda)| \leq 2|\lambda|^{-1} \quad \text{for} \quad \lambda \in (-\pi, \pi) \setminus \{0\}; \quad (\text{C.28})$$

see equation (4.7) in Robinson (1995a).

In the following analysis, we focus on the integral in (C.26) and separate the integration range as follows

$$\int_{-\pi}^{\pi} = \int_{-\pi}^{-\frac{\lambda_j}{2}} + \int_{-\frac{\lambda_j}{2}}^{\frac{\lambda_j}{2}} + \int_{\frac{\lambda_j}{2}}^{2\lambda_j} + \int_{2\lambda_j}^{\pi}. \quad (\text{C.29})$$

Note that the positive range of  $(-\pi, \pi)$  is separated into one more interval than the negative range. The additional interval  $(\frac{\lambda_j}{2}, 2\lambda_j]$  takes care of a potentially zero argument in  $K_N$ . Handling the intervals on the right-hand side of (C.29) from left to right, we get

$$\begin{aligned} & \left| \int_{-\pi}^{-\frac{\lambda_j}{2}} (f_{rs}(\lambda) - f_{rs}(\lambda_j)) K_N(\lambda - \lambda_j) d\lambda \right| \\ & \leq \int_{\frac{\lambda_j}{2}}^{\pi} (|f_{rs}(\lambda)| + |f_{rs}(\lambda_j)|) K_N(\lambda + \lambda_j) d\lambda \\ & \leq \|G\| \int_{\frac{\lambda_j}{2}}^{\pi} (\lambda^{-d_{0,r}-d_{0,s}} + \lambda_j^{-d_{0,r}-d_{0,s}}) K_N(\lambda + \lambda_j) d\lambda \\ & = \|G\| \frac{1}{2\pi N} \left( \int_{\frac{\lambda_j}{2}}^{\pi} \lambda^{-d_{0,r}-d_{0,s}} |D_N(\lambda + \lambda_j)|^2 d\lambda + \lambda_j^{-d_{0,r}-d_{0,s}} \int_{\frac{\lambda_j}{2}}^{\pi} |D_N(\lambda + \lambda_j)|^2 d\lambda \right) \\ & \leq \frac{1}{2\pi} \|G\| N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} (\cos(\lambda_m/2))^{-2} \left( 16 \frac{1}{1+2\Delta_1} + 8 \right), \end{aligned} \quad (\text{C.30})$$

where (C.30) follows by Lemmas C.9 and C.10.

The integral centered around zero can be handled as

$$\begin{aligned} & \left| \int_{-\frac{\lambda_j}{2}}^{\frac{\lambda_j}{2}} (f_{rs}(\lambda) - f_{rs}(\lambda_j)) K_N(\lambda - \lambda_j) d\lambda \right| \\ & \leq \max_{|\lambda| \leq \lambda_j/2} K_N(\lambda - \lambda_j) \int_{-\frac{\lambda_j}{2}}^{\frac{\lambda_j}{2}} |f_{rs}(\lambda) - f_{rs}(\lambda_j)| d\lambda \\ & \leq \max_{|\lambda| \leq \lambda_j/2} \frac{1}{2\pi N} 4|\lambda - \lambda_j|^{-2} \int_{-\frac{\lambda_j}{2}}^{\frac{\lambda_j}{2}} (|f_{rs}(\lambda)| + |f_{rs}(\lambda_j)|) d\lambda \end{aligned} \quad (\text{C.31})$$

$$\begin{aligned} & \leq \frac{16}{2\pi N} |\lambda_j|^{-2} \|G\| \left( \int_{-\frac{\lambda_j}{2}}^{\frac{\lambda_j}{2}} |\lambda|^{-d_{0,r}-d_{0,s}} d\lambda + \lambda_j^{1-d_{0,r}-d_{0,s}} \right) \\ & \leq \frac{16}{2\pi} \|G\| N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} \left( 4 \frac{1}{1-2\Delta_2} + 1 \right), \end{aligned} \quad (\text{C.32})$$

where (C.31) follows by applying (C.27)–(C.28).

For the next integral, applying the mean value theorem gives

$$\begin{aligned}
& \left| \int_{\frac{\lambda_j}{2}}^{2\lambda_j} \left( f_{rs}(\lambda) - f_{rs}(\lambda_j) \right) K_N(\lambda - \lambda_j) d\lambda \right| \\
& \leq \max_{\frac{\lambda_j}{2} < \lambda \leq 2\lambda_j} \left| \frac{\partial}{\partial \lambda} f_{rs}(\lambda) \right| \int_{\frac{\lambda_j}{2}}^{2\lambda_j} |\lambda - \lambda_j| K_N(\lambda - \lambda_j) d\lambda \\
& \leq \mathbf{c}_{G,2} \lambda_j^{-1-d_{0,r}-d_{0,s}} \int_{\frac{\lambda_j}{2}}^{2\lambda_j} |\lambda - \lambda_j| \frac{1}{2\pi N} |D_N(\lambda - \lambda_j)|^2 d\lambda \tag{C.33}
\end{aligned}$$

$$\begin{aligned}
& \leq \mathbf{c}_{G,2} \lambda_j^{-1-d_{0,r}-d_{0,s}} \frac{1}{\pi N} \int_{\frac{\lambda_j}{2}}^{2\lambda_j} |D_N(\lambda - \lambda_j)| d\lambda \\
& \leq \mathbf{c}_{G,2} \lambda_j^{-1-d_{0,r}-d_{0,s}} \frac{1}{\pi N} (4(\pi + 2) + 4 \log(j)) \tag{C.34}
\end{aligned}$$

$$\leq \mathbf{c}_{G,2} \lambda_j^{-1-d_{0,r}-d_{0,s}} N^{-1} 4(2 + \log(j)), \tag{C.35}$$

where (C.33) is a consequence of Assumption 3 and (C.34) follows by Lemma D.3.

The remaining integral can be bounded as

$$\begin{aligned}
& \left| \int_{2\lambda_j}^{\pi} \left( f_{rs}(\lambda) - f_{rs}(\lambda_j) \right) K_N(\lambda - \lambda_j) d\lambda \right| \\
& \leq \int_{2\lambda_j}^{\pi} \left( |f_{rs}(\lambda)| + |f_{rs}(\lambda_j)| \right) K_N(\lambda - \lambda_j) d\lambda \\
& \leq \|G\| \int_{2\lambda_j}^{\pi} \left( \lambda^{-d_{0,r}-d_{0,s}} + \lambda_j^{-d_{0,r}-d_{0,s}} \right) K_N(\lambda - \lambda_j) d\lambda \\
& = \frac{1}{2\pi N} \|G\| \left( \int_{2\lambda_j}^{\pi} \lambda^{-d_{0,r}-d_{0,s}} |D_N(\lambda - \lambda_j)|^2 d\lambda + \lambda_j^{-d_{0,r}-d_{0,s}} \int_{2\lambda_j}^{\pi} |D_N(\lambda - \lambda_j)|^2 d\lambda \right) \\
& \leq \frac{1}{2\pi} \|G\| N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} \left( 16 \frac{1}{1+2\Delta_1} + 8 \right), \tag{C.36}
\end{aligned}$$

where (C.36) follows by Lemmas C.9 and C.10.

Finally, using the integral representation (C.26) and combining the inequalities (C.30), (C.32), (C.35) and (C.36) for the individual integrals in (C.29) gives

$$\begin{aligned}
& \left| \int_{-\pi}^{\pi} \left( f_{rs}(\lambda) - f_{rs}(\lambda_j) \right) K_N(\lambda - \lambda_j) d\lambda \right| \\
& \leq \frac{1}{2\pi} \|G\| N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} (\cos(\lambda_m/2))^{-2} \left( 16 \frac{1}{1+2\Delta_1} + 8 \right) \\
& \quad + \frac{16}{2\pi} \|G\| N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} \left( 4 \frac{1}{1-2\Delta_2} + 1 \right) \\
& \quad + \mathbf{c}_{G,2} N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} 4(2 + \log(j)) + \frac{1}{2\pi} \|G\| N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} \left( 16 \frac{1}{1+2\Delta_1} + 8 \right) \\
& \leq \frac{1}{\pi} \|G\| N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} \left( ((\cos(\lambda_m/2))^{-2} + 1) \left( 8 \frac{1}{1+2\Delta_1} + 4 \right) + 4 \left( 8 \frac{1}{1-2\Delta_2} + 2 \right) \right) \\
& \quad + \mathbf{c}_{G,2} N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} 4(2 + \log(j)) \\
& \leq \frac{1}{\pi} \|G\| N^{-1} \lambda_j^{-1-d_{0,r}-d_{0,s}} \left( (\cos(\lambda_m/2))^{-2} 24 \frac{1}{1+2\Delta_1} + 48 \frac{1}{1-2\Delta_2} \right)
\end{aligned}$$

$$\begin{aligned}
& + \mathbf{c}_{G,2} N^{-1} \lambda_j^{-1-d_0,r-d_0,s} 4(2 + \log(j)) \\
\leq & \frac{1}{\pi} \|G\| \|N^{-1} \lambda_j^{-1-d_0,r-d_0,s} ((\cos(\lambda_m/2))^{-2} + 2) 24 \frac{1}{1 + 2 \min\{\Delta_1, -\Delta_2\}} \\
& + \mathbf{c}_{G,2} N^{-1} \lambda_j^{-1-d_0,r-d_0,s} 4(2 + \log(j)) \\
\leq & N^{-1} \lambda_j^{-1-d_0,r-d_0,s} \left( \|G\| \frac{72(\cos(\lambda_m/2))^{-2}}{\pi(1 + 2 \min\{\Delta_1, -\Delta_2\})} + \mathbf{c}_{G,2} 4(2 + \log(m)) \right).
\end{aligned}$$

□

**Lemma C.9.** *With the Dirichlet kernel defined in (C.27),*

$$\int_{\frac{\lambda_j}{2}}^{\pi} \lambda^{-d_0,r-d_0,s} |D_N(\lambda + \lambda_j)|^2 d\lambda \leq 16 \frac{1}{1 + 2\Delta_1} \lambda_j^{-1-d_0,r-d_0,s} (\cos(\lambda_m/2))^{-2}, \quad (\text{C.37})$$

$$\int_{2\lambda_j}^{\pi} \lambda^{-d_0,r-d_0,s} |D_N(\lambda - \lambda_j)|^2 d\lambda \leq 16 \frac{1}{1 + 2\Delta_1} \lambda_j^{-1-d_0,r-d_0,s}. \quad (\text{C.38})$$

*Proof:* We prove the inequalities (C.37) and (C.38) separately. For (C.37),

$$\begin{aligned}
\int_{\frac{\lambda_j}{2}}^{\pi} \lambda^{-d_0,r-d_0,s} |D_N(\lambda + \lambda_j)|^2 d\lambda &= \int_{\frac{\lambda_j}{2}}^{\pi} \lambda^{-d_0,r-d_0,s} |D_N(\lambda)|^2 \left( \frac{|D_N(\lambda + \lambda_j)|}{|D_N(\lambda)|} \right)^2 d\lambda \\
&= \sup_{\frac{\lambda_j}{2} \leq \lambda \leq \pi} \left( \frac{|D_N(\lambda + \lambda_j)|}{|D_N(\lambda)|} \right)^2 \int_{\frac{\lambda_j}{2}}^{\pi} \lambda^{-d_0,r-d_0,s} |D_N(\lambda)|^2 d\lambda \\
&\leq (\cos(\lambda_m/2))^{-2} \int_{\frac{\lambda_j}{2}}^{\pi} \lambda^{-d_0,r-d_0,s} |D_N(\lambda)|^2 d\lambda \quad (\text{C.39})
\end{aligned}$$

$$\leq 4(\cos(\lambda_m/2))^{-2} \int_{\frac{\lambda_j}{2}}^{\infty} \lambda^{-2-d_0,r-d_0,s} d\lambda \quad (\text{C.40})$$

$$\begin{aligned}
&\leq 4(\cos(\lambda_m/2))^{-2} \frac{1}{1 + 2\Delta_1} 2^{1+d_0,r+d_0,s} \lambda_j^{-1-d_0,r-d_0,s} \\
&\leq 16(\cos(\lambda_m/2))^{-2} \frac{1}{1 + 2\Delta_1} \lambda_j^{-1-d_0,r-d_0,s},
\end{aligned}$$

where (C.39) is proved in Lemma D.4 and (C.40) follows by (C.28). Similarly,

$$\begin{aligned}
\int_{2\lambda_j}^{\pi} \lambda^{-d_0,r-d_0,s} |D_N(\lambda - \lambda_j)|^2 d\lambda &= \int_{2\lambda_j}^{\pi} \lambda^{-d_0,r-d_0,s} |D_N(\lambda)|^2 \left( \frac{|D_N(\lambda - \lambda_j)|}{|D_N(\lambda)|} \right)^2 d\lambda \\
&\leq \sup_{2\lambda_j \leq \lambda \leq \pi} \left( \frac{|D_N(\lambda - \lambda_j)|}{|D_N(\lambda)|} \right)^2 \int_{2\lambda_j}^{\pi} \lambda^{-d_0,r-d_0,s} |D_N(\lambda)|^2 d\lambda \\
&\leq 4 \int_{2\lambda_j}^{\pi} \lambda^{-d_0,r-d_0,s} |D_N(\lambda)|^2 d\lambda \quad (\text{C.41})
\end{aligned}$$

$$\leq 16 \int_{2\lambda_j}^{\infty} \lambda^{-2-d_0,r-d_0,s} d\lambda \quad (\text{C.42})$$

$$\leq 16 \frac{1}{1 + 2\Delta_1} 2^{-1-d_0,r-d_0,s} \lambda_j^{-1-d_0,r-d_0,s}$$

$$\leq 16 \frac{1}{1 + 2\Delta_1} \lambda_j^{-1-d_0,r-d_0,s},$$

where (C.41) is proved in Lemma D.4 and (C.42) follows by (C.28). □

**Lemma C.10.** *With the Dirichlet kernel defined in (C.27),*

$$\int_{\frac{\lambda_j}{2}}^{\pi} |D_N(\lambda + \lambda_j)|^2 d\lambda \leq 8\lambda_j^{-1} (\cos(\lambda_m/2))^{-2} \quad \text{and} \quad \int_{2\lambda_j}^{\pi} |D_N(\lambda - \lambda_j)|^2 d\lambda \leq 8\lambda_j^{-1}. \quad (\text{C.43})$$

*Proof:* The first integral in (C.43) can be bounded as

$$\begin{aligned} \int_{\frac{\lambda_j}{2}}^{\pi} |D_N(\lambda + \lambda_j)|^2 d\lambda &= \int_{\frac{\lambda_j}{2}}^{\pi} |D_N(\lambda)|^2 \left( \frac{|D_N(\lambda + \lambda_j)|}{|D_N(\lambda)|} \right)^2 d\lambda \\ &\leq \sup_{\frac{\lambda_j}{2} \leq \lambda \leq \pi} \left( \frac{|D_N(\lambda + \lambda_j)|}{|D_N(\lambda)|} \right)^2 \int_{\frac{\lambda_j}{2}}^{\pi} |D_N(\lambda)|^2 d\lambda \\ &\leq (\cos(\lambda_m/2))^{-2} \int_{\frac{\lambda_j}{2}}^{\pi} |D_N(\lambda)|^2 d\lambda \end{aligned} \quad (\text{C.44})$$

$$\leq 4(\cos(\lambda_m/2))^{-2} \int_{\frac{\lambda_j}{2}}^{\infty} |\lambda|^{-2} d\lambda \leq 8(\cos(\lambda_m/2))^{-2} |\lambda_j|^{-1}, \quad (\text{C.45})$$

where (C.44) is proved in Lemma D.4 and the inequality (C.45) follows by (C.28). Similarly,

$$\begin{aligned} \int_{2\lambda_j}^{\pi} |D_N(\lambda - \lambda_j)|^2 d\lambda &= \int_{2\lambda_j}^{\pi} |D_N(\lambda)|^2 \left( \frac{|D_N(\lambda - \lambda_j)|}{|D_N(\lambda)|} \right)^2 d\lambda \\ &\leq \sup_{2\lambda_j \leq \lambda \leq \pi} \left( \frac{|D_N(\lambda - \lambda_j)|}{|D_N(\lambda)|} \right)^2 \int_{2\lambda_j}^{\pi} |D_N(\lambda)|^2 d\lambda \end{aligned} \quad (\text{C.46})$$

$$\begin{aligned} &\leq 4 \int_{2\lambda_j}^{\pi} |D_N(\lambda)|^2 d\lambda \\ &\leq 16 \int_{2\lambda_j}^{\infty} |\lambda|^{-2} d\lambda \leq 8|\lambda_j|^{-1}, \end{aligned} \quad (\text{C.47})$$

where (C.46) is proved in Lemma D.4 and the inequality (C.47) follows by (C.28).  $\square$

### C.3 Bounds on covariance matrices norms

This section collects our results with bounds on different matrix norms of the covariance matrix and their proofs.

**Lemma C.11.** *Let  $\Sigma_{rr} = (\Sigma_{rr}(n-k))_{n,k=1,\dots,N} = \mathbb{E}(\mathcal{X}e_r(\mathcal{X}e_r)')$  and  $d_{0,r} \in [\Delta_1, \Delta_2]$  in (1.1). Then, there exist constants  $c_1, c_2$  such that the spectral and Frobenius norms of  $\Sigma_{rr}$  can be bounded as*

$$\begin{aligned} \|\Sigma_{rr}\| &\leq c_1 \|G\| N^{\max\{2d_{0,r}, 0\}} \bar{\Delta}_N, \\ \|\Sigma_{rr}\|_F &\leq c_2 \|G\| N^{\max\{2d_{0,r}, \frac{1}{2}\}} \bar{\Delta}_N \end{aligned}$$

with  $\bar{\Delta}_N$  as in (3.13).

*Proof:* The bounds follow from Lemmas C.12 and C.13 below.  $\square$

**Lemma C.12.** *Let  $\Sigma_{rr} = (\Sigma_{rr}(n-k))_{n,k=1,\dots,N} = \mathbb{E}(\mathcal{X}e_r(\mathcal{X}e_r)')$  and  $d_{0,r} \leq 0$  in (1.1). Then, there exists a constant  $c$  such that the spectral and Frobenius norms of  $\Sigma_{rr}$  can be bounded as*

$$\|\Sigma_{rr}\| \leq c \|G\|, \quad (\text{C.48})$$

$$\|\Sigma_{rr}\|_F \leq c \|G\| N^{\frac{1}{2}}. \quad (\text{C.49})$$



*Proof:* See Lemma C.5. in Sun et al. (2018) for the inequality (C.48). The second inequality (C.49) is a simple consequence of using the fact that  $\|A\|_F \leq \sqrt{\text{rk}(A)}\|A\|$  for a matrix  $A$  and applying (C.48).  $\square$

**Lemma C.13.** *Let  $\Sigma_{rr} = (\Sigma_{rr}(n-k))_{n,k=1,\dots,N} = \mathbb{E}(\mathcal{X}e_r(\mathcal{X}e_r)')$  and  $d_{0,r} > 0$  in (1.1). Then, there exist constants  $c_1, c_2$  such that the spectral and Frobenius norms of  $\Sigma_{rr}$  can be bounded as*

$$\|\Sigma_{rr}\| \leq c_1 \|G\| N^{2d_{0,r}} \log(N), \quad (\text{C.50})$$

$$\|\Sigma_{rr}\|_F \leq c_2 \|G\| \begin{cases} N^{\frac{1}{2}} \log(N)^{\frac{1}{2}}, & \text{if } d_{0,r} \leq \frac{1}{4}, \\ N^{2d_{0,r}} \log(N)^{\frac{1}{2}}, & \text{if } d_{0,r} > \frac{1}{4}. \end{cases} \quad (\text{C.51})$$

**Remark C.1.** We pause here to comment on Lemma C.13. In the light of asymptotic results on the autocovariances under long-range dependence, it might be surprising to the reader to find the logarithm in all our bounds opposed to only in the case  $d_{0,r} = \frac{1}{4}$ . This could certainly be avoided by bounding certain sums in our proofs by integrals depending on the memory parameter  $d_{0,r}$ . Then, one can replace the logarithm respectively by  $\frac{1}{d_{0,r}}$  for the bound in (C.50) and  $\frac{1}{\sqrt{4d_{0,r}-1}} \mathbb{1}_{\{d_{0,r} \neq \frac{1}{4}\}}$  in (C.51). However, the memory parameter  $d_{0,r}$  approaching either zero or  $\frac{1}{4}$  results in potentially high constants. In contrast, the logarithm can be controlled in our high probability bounds.

*Proof of Lemma C.13:* The autocovariance matrix  $\Sigma_{rr} = (\Sigma_{rr}(n-k))_{n,k=1,\dots,N}$  of the component series  $\{X_{r,n}\}_{n \in \mathbb{Z}}$  can be represented in terms of the spectral density  $f_{rr}(\omega)$  of the respective component series as

$$\begin{aligned} \Sigma_{rr} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{rr}(\omega) (e^{i(n-k)\omega})_{n,k=1,\dots,N} d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |\omega|^{-2d_{0,r}} G_{rr}(\omega) (e^{i(n-k)\omega})_{n,k=1,\dots,N} d\omega. \end{aligned} \quad (\text{C.52})$$

In the following, we prove the statements for the spectral and the Frobenius norm separately.

*Spectral norm:* Using (C.52), we can bound the spectral norm of  $\Sigma_{rr}$  as

$$\|\Sigma_{rr}\| \leq \text{ess sup}_{\omega \in (-\pi, \pi)} |G_{rr}(\omega)| \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} |\omega|^{-2d_{0,r}} (e^{i(n-k)\omega})_{n,k=1,\dots,N} d\omega \right\| \quad (\text{C.53})$$

$$\leq \|G\| \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} |\exp(i\omega) - 1|^{-2d_{0,r}} (e^{i(n-k)\omega})_{n,k=1,\dots,N} d\omega \right\|, \quad (\text{C.54})$$

where (C.53) follows since  $(e^{i(n-k)\omega})_{n,k=1,\dots,N}$  is positive semidefinite and because of  $|e^{ix} - e^{iy}| \leq |x - y|$ . In order to bound the spectral norm in (C.54), we replace the matrix by an integral operator with piecewise constant kernel. More specifically, define the integral operator  $K_k : L^2(0, 1) \rightarrow L^2(0, 1)$  as

$$(K_k g)(x) = \int_0^1 g(y) k_N(x, y) dy \quad (\text{C.55})$$

with kernel function

$$k_N(x, y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |e^{i\omega} - 1|^{-2d_{0,r}} e^{i([Nx] - [Ny])\omega} d\omega.$$

Then,

$$\left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} |e^{i\omega} - 1|^{-2d_{0,r}} (e^{i(n-k)\omega})_{n,k=1,\dots,N} d\omega \right\| = N \|K_k\|_{op} \leq c N^{2d_{0,r}} \log(N); \quad (\text{C.56})$$

see Lemma 4.1 in Böttcher and Dörfler (2010) for the equality in (C.56). The inequality stated in (C.56) is proved in Lemma C.14. Combining (C.54) and (C.56) yields (C.50).

*Frobenius norm:* We deal with the Frobenius norm similarly as with the spectral norm as

$$\begin{aligned}
\|\Sigma_{rr}\|_F &= \left( \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} G_{rr}(\omega_1) G_{rr}(\omega_2) |\omega_1 \omega_2|^{-2d_{0,r}} \sum_{n,k=1}^N e^{i(n-k)(\omega_1-\omega_2)} d\omega_1 d\omega_2 \right)^{\frac{1}{2}} \\
&= \left( \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} G_{rr}(\omega_1) G_{rr}(\omega_2) |\omega_1 \omega_2|^{-2d_{0,r}} \left| \sum_{n=1}^N e^{in(\omega_1-\omega_2)} \right|^2 d\omega_1 d\omega_2 \right)^{\frac{1}{2}} \\
&\leq \operatorname{ess\,sup}_{\omega \in (-\pi, \pi)} |G_{rr}(\omega)| \left( \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |\omega_1 \omega_2|^{-2d_{0,r}} \left| \sum_{n=1}^N e^{in(\omega_1-\omega_2)} \right|^2 d\omega_1 d\omega_2 \right)^{\frac{1}{2}} \\
&\leq \|G\| \left( \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |\exp(i\omega_1) - 1|^{-2d_{0,r}} |\exp(i\omega_2) - 1|^{-2d_{0,r}} \left| \sum_{n=1}^N e^{in(\omega_1-\omega_2)} \right|^2 d\omega_1 d\omega_2 \right)^{\frac{1}{2}} \tag{C.57}
\end{aligned}$$

$$= \|G\| \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} |\exp(i\omega) - 1|^{-2d_{0,r}} (e^{i(n-k)\omega})_{n,k=1,\dots,N} d\omega \right\|_F, \tag{C.58}$$

where (C.57) follows since  $|e^{ix} - e^{iy}| \leq |x - y|$ . Again, we apply Lemma 4.1 in Böttcher and Dörfler (2010) to replace the matrix in (C.58) by an integral operator

$$\left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} |\exp(i\omega) - 1|^{-2d_{0,r}} (\exp(i(n-k)\omega))_{n,k=1,\dots,N} d\omega \right\|_F = N \|K_k\|_2 \tag{C.59}$$

with integral operator  $K_k$  as in (C.55). The equality follows by Lemma 4.1 in Böttcher and Dörfler (2010). Combining (C.58) and (C.59) with Lemma C.15 below gives

$$\|\Sigma_{rr}\|_F \leq c \begin{cases} N^{\frac{1}{2}} \log(N)^{\frac{1}{2}}, & \text{if } d_{0,r} \leq \frac{1}{4}, \\ N^{2d_{0,r}} \log(N)^{\frac{1}{2}}, & \text{if } d_{0,r} > \frac{1}{4}. \end{cases}$$

□

**Lemma C.14.** For  $d_{0,r} > 0$ , the operator norm of the integral operator  $K_k$  defined in (C.55) can be bounded as

$$N \|K_k\|_{op} \leq c N^{2d_{0,r}} \log(N). \tag{C.60}$$

*Proof:* In view of the definition of the integral operator norm (see the end of Section 1),

$$\begin{aligned}
\int_0^1 \left| \int_0^1 g(y) k_N(x, y) dy \right|^2 dx &\leq \int_0^1 \left( \int_0^1 |g(y)| |k_N(x, y)|^{\frac{1}{2}} |k_N(x, y)|^{\frac{1}{2}} dy \right)^2 dx \\
&\leq \int_0^1 \int_0^1 |g(y)|^2 |k_N(x, y)| dy \int_0^1 |k_N(x, y)| dy dx \\
&\leq \sup_{x \in (0,1)} \int_0^1 |k_N(x, y)| dy \int_0^1 \int_0^1 |g(y)|^2 |k_N(x, y)| dy dx \\
&\leq c N^{2d_{0,r}-1} \log(N) \sup_{y \in (0,1)} \int_0^1 |k_N(x, y)| dx \int_0^1 |g(y)|^2 dy \tag{C.61} \\
&\leq c (N^{2d_{0,r}-1} \log(N))^2 \int_0^1 |g(y)|^2 dy, \tag{C.62}
\end{aligned}$$

where (C.61) and (C.62) follow since, with explanations given below,

$$\sup_{x \in (0,1)} \int_0^1 |k_N(x,y)| dy = \sup_{y \in (0,1)} \int_0^1 |k_N(x,y)| dx \leq cN^{2d_0,r-1}(\log(N) + 1). \quad (\text{C.63})$$

In order to prove (C.63), we consider  $z_N \neq 0$  and  $z_N = 0$  separately, with  $z_N = [Nx] - [Ny]$ . By Lemma C.16, we get

$$\begin{aligned} \int_0^1 |k_N(x,y)| \mathbf{1}_{\{z_N \neq 0\}} dx &\leq c \int_0^1 |[Nx] - [Ny]|^{2d_0,r-1} \mathbf{1}_{\{[Nx] - [Ny] \neq 0\}} dx \\ &= c \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} |i-1 - [Ny]|^{2d_0,r-1} \mathbf{1}_{\{i-1 - [Ny] \neq 0\}} dx \\ &= c \frac{1}{N} \sum_{i=1}^N |i-1 - [Ny]|^{2d_0,r-1} \mathbf{1}_{\{i-1 - [Ny] \neq 0\}} \\ &\leq c \frac{1}{N} \sum_{i=1}^N \left| i-1 - \left[ \frac{N-1}{2} \right] \right|^{2d_0,r-1} \mathbf{1}_{\{i-1 - [\frac{N-1}{2}] \neq 0\}} \\ &= c \frac{1}{N} \sum_{i=-[\frac{N-1}{2}]}^{N-1-[\frac{N-1}{2}]} |i|^{2d_0,r-1} \mathbf{1}_{\{i \neq 0\}} \\ &\leq cN^{2d_0,r-1} \frac{1}{N} \sum_{i=-[\frac{N-1}{2}]}^{N-1-[\frac{N-1}{2}]} \left| \frac{i}{N} \right|^{2d_0,r-1} \mathbf{1}_{\{i \neq 0\}} \\ &\leq cN^{2d_0,r-1} \sum_{i=-[\frac{N-1}{2}]}^{N-1-[\frac{N-1}{2}]} |i|^{-1} \mathbf{1}_{\{i \neq 0\}} \\ &\leq cN^{2d_0,r-1} \sum_{i=1}^N i^{-1} \leq cN^{2d_0,r-1}(\log(N) + 1). \end{aligned}$$

Using (C.68) in Lemma C.16, for  $z_N = 0$ ,

$$\begin{aligned} \int_0^1 |k_N(x,y)|^2 \mathbf{1}_{\{z_N=0\}} dx &\leq \left( \frac{\Gamma(1-2\Delta_2)}{\Gamma^2(1-\Delta_2)} \right)^2 \int_0^1 \mathbf{1}_{\{[Nx] - [Ny] = 0\}} dx \\ &= \left( \frac{\Gamma(1-2\Delta_2)}{\Gamma^2(1-\Delta_2)} \right)^2 \int_{\frac{[Ny]}{N}}^{\frac{[Ny]+1}{N}} dx \\ &= \left( \frac{\Gamma(1-2\Delta_2)}{\Gamma^2(1-\Delta_2)} \right)^2 \frac{1}{N}. \end{aligned} \quad (\text{C.64})$$

□

**Lemma C.15.** For  $d_{0,r} > 0$ , the  $L^2$  norm of the integral operator  $K_k$  defined in (C.55) can be bounded as

$$N \|K_k\|_2 \leq c \begin{cases} N^{\frac{1}{2}} \log(N)^{\frac{1}{2}}, & \text{if } d_{0,r} \leq \frac{1}{4}, \\ N^{2d_{0,r}} \log(N)^{\frac{1}{2}}, & \text{if } d_{0,r} > \frac{1}{4}. \end{cases} \quad (\text{C.65})$$

*Proof:* For  $z_N = [Nx] - [Ny]$ , we consider  $z_N \neq 0$  and  $z_N = 0$  separately. By Lemma C.16, we get

$$\begin{aligned}
& \int_0^1 \int_0^1 |k_N(x, y)|^2 \mathbf{1}_{\{z_N \neq 0\}} dx dy \\
& \leq c \int_0^1 \int_0^1 |[Nx] - [Ny]|^{4d_{0,r}-2} \mathbf{1}_{\{[Nx] - [Ny] \neq 0\}} dx dy \\
& = c \sum_{i=1}^N \sum_{j=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \int_{\frac{j-1}{N}}^{\frac{j}{N}} |[Nx] - [Ny]|^{4d_{0,r}-2} \mathbf{1}_{\{[Nx] - [Ny] \neq 0\}} dx dy \\
& = c \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N |i - j|^{4d_{0,r}-2} \mathbf{1}_{\{i \neq j\}} \\
& = c \frac{1}{N^2} \sum_{k=1}^N (N - k) k^{4d_{0,r}-2} \\
& \leq c \begin{cases} \frac{1}{N} \sum_{k=1}^N (1 - \frac{k}{N}) k^{-1}, & \text{if } d_{0,r} \leq \frac{1}{4}, \\ N^{4d_{0,r}-2} \left( \frac{1}{N} \sum_{k=1}^N (\frac{k}{N})^{4d_{0,r}-2} - \frac{1}{N} \sum_{k=1}^N (\frac{k}{N})^{4d_{0,r}-1} \right), & \text{if } d_{0,r} > \frac{1}{4}, \end{cases} \\
& \leq c \begin{cases} \frac{1}{N} \sum_{k=1}^N k^{-1}, & \text{if } d_{0,r} \leq \frac{1}{4}, \\ N^{4d_{0,r}-2} \sum_{k=1}^N k^{-1}, & \text{if } d_{0,r} > \frac{1}{4}, \end{cases} \\
& \leq c \begin{cases} \frac{1}{N} (\log(N) + 1), & \text{if } d_{0,r} \leq \frac{1}{4}, \\ N^{4d_{0,r}-2} (\log(N) + 1), & \text{if } d_{0,r} > \frac{1}{4}. \end{cases}
\end{aligned}$$

Using (C.68) in Lemma C.16 and the same arguments for  $z_N = 0$ ,

$$\begin{aligned}
\int_0^1 \int_0^1 |k_N(x, y)|^2 \mathbf{1}_{\{z_N = 0\}} dx dy & \leq \left( \frac{\Gamma(1 - 2\Delta_2)}{\Gamma^2(1 - \Delta_2)} \right)^2 \int_0^1 \int_0^1 \mathbf{1}_{\{[Nx] - [Ny] = 0\}} dx dy \\
& = \left( \frac{\Gamma(1 - 2\Delta_2)}{\Gamma^2(1 - \Delta_2)} \right)^2 \frac{1}{N}
\end{aligned}$$

following (C.64). □

**Lemma C.16.** For  $d_{0,r} > 0$ , the function

$$k_N(x, y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |e^{i\omega} - 1|^{-2d_{0,r}} e^{i([Nx] - [Ny])\omega} d\omega \quad (\text{C.66})$$

can be bounded as

$$0 < k_N(x, y) \leq c |z_N|^{2d_{0,r}-1} \quad \text{with } z_N = [Nx] - [Ny]. \quad (\text{C.67})$$

Furthermore,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |e^{i\omega} - 1|^{-2d_{0,r}} d\omega \leq \frac{\Gamma(1 - 2\Delta_2)}{\Gamma^2(1 - \Delta_2)}. \quad (\text{C.68})$$

*Proof:* Set  $z_N = [Nx] - [Ny]$ . Then, the kernel (C.66) satisfies

$$k_N(x, y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |e^{i\omega} - 1|^{-2d_{0,r}} e^{i([Nx] - [Ny])\omega} d\omega$$

$$\begin{aligned}
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} |\exp(i\omega) - 1|^{-2d_{0,r}} \exp(iz_N\omega) d\omega \\
&= (-1)^{|z_N|} \frac{\Gamma(1 - 2d_{0,r})}{\Gamma(1 - |z_N| - d_{0,r})\Gamma(1 + |z_N| - d_{0,r})} \tag{C.69}
\end{aligned}$$

$$= \Gamma(1 - 2d_{0,r}) \frac{\sin(\pi d_{0,r})}{\pi(|z_N| + d_{0,r})} \frac{\Gamma(|z_N| + 1 + d_{0,r})}{\Gamma(|z_N| + 1 - d_{0,r})}. \tag{C.70}$$

See p. 665 in Böttcher and Virtanen (2007) for equations (C.69) and (C.70). The last relation shows  $k_N(x, y) > 0$ . We further have

$$k_N(x, y) < \Gamma(1 - 2d_{0,r}) \frac{\sin(\pi d_{0,r})}{\pi(|z_N| + d_{0,r})} \frac{(|z_N| + d_{0,r})^{2d_{0,r}}}{d_{0,r}^{2d_{0,r}}} \frac{\Gamma(d_{0,r} + 1)}{\Gamma(d_{0,r} + 1 - 2d_{0,r})} \tag{C.71}$$

$$\leq c(|z_N| + d_{0,r})^{2d_{0,r}-1} \left( \frac{d_{0,r} + 1 - 2d_{0,r}}{d_{0,r}} \right)^{2d_{0,r}} \tag{C.72}$$

$$\leq c|z_N|^{2d_{0,r}-1} 2. \tag{C.73}$$

The inequality (C.71) follows by using

$$\frac{\Gamma(x + a)}{\Gamma(x + b)} < \frac{x^{a-b} \Gamma(x_0 + a)}{x_0^{a-b} \Gamma(x_0 + b)} \tag{C.74}$$

for  $a > b \geq 0$  and  $a + b \geq 1$  on  $[x_0, \infty)$  for any  $x_0 > 0$ ; see (3.75) in Qi and Luo (2012). The inequality (C.74) is applied with  $x = |z_N| + d_{0,r}$ ,  $x_0 = d_{0,r}$ ,  $a = 1$ ,  $b = 1 - d_{0,r}$ . Furthermore, (C.72) follows from the inequality

$$\frac{\Gamma(x + a)}{x^a \Gamma(x)} \leq 1 \tag{C.75}$$

for  $x > 0$  and  $a \in (0, 1)$ ; see (2.2) in Qi and Luo (2012). The inequality (C.75) is applied with  $x = 1 - d_{0,r}$  and  $a = 2d_{0,r}$ .

For the integral in (C.68), we get

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\exp(i\omega) - 1|^{-2d_{0,r}} d\omega = \frac{\Gamma(1 - 2d_{0,r})}{\Gamma^2(1 - d_{0,r})} =: g(d_{0,r}) \leq \frac{\Gamma(1 - 2\Delta_2)}{\Gamma^2(1 - \Delta_2)},$$

where the equality follows by p. 665 in Böttcher and Virtanen (2007) as in (C.69) and the inequality since the function  $g$  can be checked to be monotonically increasing in  $d_{0,r}$ .  $\square$

## D Some additional technical results

This appendix concerns three different kinds of technical results, required to prove our main results. Section D.1 presents a uniform concentration inequality. The two remaining Sections D.2 and D.3 give some inequalities on matrix norms and the Dirichlet kernel, respectively.

### D.1 Uniform concentration inequality

For completeness, we give a slightly modified version of Theorem 1 in Dicker and Erdogdu (2017) and explain the differences from the original formulation.

**Theorem D.1.** Let  $0 < R < \infty$  and  $t_1(u), \dots, t_m(u)$  be real-valued on  $[0, R]^K \subseteq \mathbb{R}^K$  and differentiable on  $(0, R)^K \subseteq \mathbb{R}^K$  with bounded derivative. Define  $T(u) = \text{diag}(t_1(u), \dots, t_m(u))$  and  $Q(u) = VT(u)V'$ , where  $V$  is a  $p \times m$  matrix. Let  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)'$ , where  $\varepsilon_1, \dots, \varepsilon_N$  are independent mean 0 sub-Gaussian random variables satisfying

$$\max_{i=1, \dots, N} \|\varepsilon_i\|_\phi \leq \gamma \quad (\text{D.1})$$

for some constant  $\gamma \in (0, \infty)$ . Then, there exists a constant  $C \in (0, \infty)$  such that

$$\mathbb{P}\left(\sup_{u \in [0, R]^K} |\varepsilon' Q(u) \varepsilon - \mathbb{E}(\varepsilon' Q(u) \varepsilon)| > \nu\right) \leq C \exp\left(-\frac{1}{C} \min\left\{\frac{\nu}{\gamma^2 \mathcal{T}}, \frac{\nu^2}{\gamma^4 \mathcal{T}_i^2}\right\}\right), \quad (\text{D.2})$$

for  $\nu^2 \geq C\gamma^4 \mathcal{T}_i^2 K^2$ ,  $i = 1, 2$ , where

$$\mathcal{T} = \|V'V\|_{L_1}, \quad \mathcal{T}_1 = \|V'V\|_{L_2}, \quad \mathcal{T}_2 = \|V'V\|_F L_1$$

with

$$L_1 = \|T(0)\| + RK^{\frac{1}{2}} \sup_{u \in [0, R]^K} \|\mathbf{T}(u)\| \quad \text{and} \quad L_2 = \|T(0)\|_F + RK^{\frac{1}{2}} \sup_{u \in [0, R]^K} \|\mathbf{T}(u)\|_F, \quad (\text{D.3})$$

where  $\mathbf{T}(u) = \text{diag}(\|\nabla t_1(u)\|_F, \dots, \|\nabla t_m(u)\|_F)$ .

The statement in Theorem D.1 differs in two points from the original Theorem 1 stated in Dicker and Erdogdu (2017). First, the condition that the functions  $t_i(u), i = 1 \dots, m$ , are differentiable with bounded derivatives replaces a Lipschitz condition on the functions  $t_i(u), i = 1 \dots, m$ ; see (1) in Dicker and Erdogdu (2017). This assumption makes it more convenient to write our results. Second, Dicker and Erdogdu (2017) get  $\mathcal{T}_1$  in the bound, which uses the fact that  $\|Q(u)\|_F = \|VT(u)V'\|_F \leq \|VV'\| \|T(u)\|_F$ ; see Lemma D.1. In some situations, it turns out to be helpful to consider  $\mathcal{T}_2$ , which is a consequence of the inequality

$$\|Q(u)\|_F = \|VT(u)V'\|_F \leq \|VV'\|_F \|T(u)\|;$$

see again Lemma D.1. Following the proof of Theorem 1 in Dicker and Erdogdu (2017), those changes yield the concentration inequality in (D.2) with  $i = 2$ .

We conclude with a remark which comments on a possibility to slightly generalize the results in Theorem D.1.

**Remark D.1.** Theorem D.1 is stated in terms of the matrix  $Q(u) = VT(u)V'$  under the assumption that  $T(u)$  is a diagonal matrix. The proof of Theorem D.1 relies on a chaining technique with a subsequent application of the Hanson-Wright inequality (Rudelson and Vershynin, 2013, Theorem 1.1). The Hanson-Wright inequality is applicable for arbitrary matrices. In particular, it does not require symmetry or diagonality. However, in order to bound the spectral and Frobenius norms of  $T(u) - T(u')$ ,  $u, u' \in [0, R]^K$  in the proof of Theorem D.1 in terms of the quantities in (D.3), we need to impose some structural assumptions. As stated in Theorem D.1, Dicker and Erdogdu (2017) required  $T(u)$  to be diagonal. This assumption can be slightly relaxed by supposing that there exists a representation  $T(u) = A' \tilde{T}(u) B$  with  $A, B \in \mathbb{R}^{m \times m}$  unitary and independent of  $u$  and  $\tilde{T}(u)$  diagonal since the Frobenius norm and the spectral norm are invariant under unitary transformation.

## D.2 Matrix norm inequalities

For completeness, we collect here some results on matrix norms.

**Lemma D.1.** *Let  $B$  be a positive semidefinite matrix. Then,*

$$\|ABA'\|_F \leq \|B\| \|A'A\|_F \quad \text{and} \quad \|ABA'\|_F \leq \|A'A\| \|B\|_F.$$

*Proof:* The Frobenius norm can be rewritten as

$$\begin{aligned} \|ABA'\|_F^2 &= \text{tr}((ABA')'ABA') = \text{tr}(BA'ABA'A) \\ &\leq \lambda_{\max}(B) \text{tr}(BA'AA'A) \\ &\leq \lambda_{\max}(B)^2 \|A'A\|_F^2 = \|B\|^2 \|A'A\|_F^2; \end{aligned}$$

see Theorem 1 in Fang, Loparo, and Feng (1994) for the eigenvalue-trace inequality in the second line. Similarly,

$$\begin{aligned} \|ABA'\|_F^2 &= \text{tr}((ABA')'ABA') = \text{tr}(A'ABA'AB) \\ &\leq \lambda_{\max}(A'A) \text{tr}(A'AB^2) \\ &\leq \lambda_{\max}(A'A)^2 \|B\|_F^2 = \|A'A\|^2 \|B\|_F^2. \end{aligned}$$

□

**Lemma D.2.** *Let  $M$  be a positive semidefinite block-matrix*

$$M = \begin{pmatrix} A & X \\ X' & B \end{pmatrix}.$$

*Then,*

$$\|M\| \leq \|A\| + \|B\| \quad \text{and} \quad \|M\|_F \leq \|A\|_F + \|B\|_F.$$

*Proof:* The inequalities are consequences of Lemma 1.1 in Bourin, Lee, and Lin (2012) and the unitary invariance of the spectral and the Frobenius norm. □

## D.3 Dirichlet kernel

In this section, we present some results regarding the Dirichlet kernel in (C.27).

**Lemma D.3.** *The Dirichlet kernel in (C.27) satisfies*

$$\int_{\frac{\lambda_j}{2}}^{2\lambda_j} |D_N(\lambda - \lambda_j)| d\lambda \leq 4(\pi + 2) + 4 \log(j).$$

*Proof:* Using the representation (C.27) of the Dirichlet kernel, a series of inequalities lead to

$$\begin{aligned} \int_{\frac{\lambda_j}{2}}^{2\lambda_j} |D_N(\lambda - \lambda_j)| d\lambda &= \int_{\frac{\lambda_j}{2}}^{2\lambda_j} \left| \frac{\sin(N(\lambda - \lambda_j)/2)}{\sin((\lambda - \lambda_j)/2)} \right| d\lambda = \int_{-\frac{\lambda_j}{2}}^{\lambda_j} \left| \frac{\sin(N\lambda/2)}{\sin(\lambda/2)} \right| d\lambda \\ &= 2 \int_0^{\lambda_j} \left| \frac{\lambda}{\sin(\lambda/2)} \frac{\sin(N\lambda/2)}{\lambda} \right| d\lambda \leq 2 \frac{\lambda_j}{\sin(\lambda_j/2)} \int_0^{\lambda_j} \left| \frac{\sin(N\lambda/2)}{\lambda} \right| d\lambda \\ &= 2 \frac{\lambda_j}{\sin(\lambda_j/2)} \sum_{i=0}^{j-1} \int_{\lambda_i}^{\lambda_{i+1}} \left| \frac{\sin(N\lambda/2)}{\lambda} \right| d\lambda \leq 2\pi \sum_{i=0}^{j-1} \int_0^{\lambda_1} \left| \frac{\sin(N(\lambda + \lambda_i)/2)}{\lambda + \lambda_i} \right| d\lambda \end{aligned}$$

$$\begin{aligned}
&= 2\pi \left( \int_0^{\lambda_1} \left| \frac{\sin(N\lambda/2)}{\lambda} \right| d\lambda + \sum_{i=1}^{j-1} \int_0^{\lambda_1} \left| \frac{\sin(N(\lambda + \lambda_i)/2)}{\lambda + \lambda_i} \right| d\lambda \right) \\
&= 2\pi \left( \int_0^\pi \left| \frac{\sin(\lambda)}{\lambda} \right| d\lambda + \int_0^{\lambda_1} \sin(N\lambda/2) \sum_{i=1}^{j-1} \frac{1}{\lambda + \lambda_i} d\lambda \right) \\
&\leq 4\pi + 2\pi \int_0^{\lambda_1} \sin(N\lambda/2) d\lambda \sum_{i=1}^{j-1} \frac{1}{\lambda_i} = 4\pi + 2\pi \frac{4}{N} \sum_{i=1}^{j-1} \frac{1}{\lambda_i} \\
&\leq 4(\pi + 1) + 4 \sum_{i=2}^j \frac{1}{i} \leq 4(\pi + 1) + 4 \int_1^j \frac{1}{x} dx \\
&\leq 4(\pi + 1) + 4(\log(j) + 1) \\
&= 4(\pi + 2) + 4 \log(j).
\end{aligned}$$

□

**Lemma D.4.** *The Dirichlet kernel (C.27) satisfies*

$$\sup_{\frac{\lambda_j}{2} \leq \lambda \leq \pi} \left( \frac{|D_N(\lambda + \lambda_j)|}{|D_N(\lambda)|} \right)^2 \leq (\cos(\lambda_m/2))^{-2} \quad \text{and} \quad \sup_{2\lambda_j \leq \lambda \leq \pi} \left( \frac{|D_N(\lambda - \lambda_j)|}{|D_N(\lambda)|} \right)^2 \leq 4. \quad (\text{D.4})$$

*Proof:* Though very similar, we prove the two inequalities in (D.4) separately. Using the representation (C.27) of the Dirichlet kernel leads to

$$\begin{aligned}
\sup_{\frac{\lambda_j}{2} \leq \lambda \leq \pi} \left( \frac{|D_N(\lambda + \lambda_j)|}{|D_N(\lambda)|} \right)^2 &= \sup_{\frac{\lambda_j}{2} \leq \lambda \leq \pi} \left( \frac{|\sin(N(\lambda + \lambda_j)/2) \sin(\lambda/2)|}{|\sin(N\lambda/2) \sin((\lambda + \lambda_j)/2)|} \right)^2 \\
&= \sup_{\frac{\lambda_j}{2} \leq \lambda \leq \pi} \left( \frac{|\sin(N\lambda/2) \sin(\lambda/2)|}{|\sin(N\lambda/2) \sin((\lambda + \lambda_j)/2)|} \right)^2 \quad (\text{D.5})
\end{aligned}$$

$$\begin{aligned}
&= \sup_{\frac{\lambda_j}{2} \leq \lambda \leq \pi} \left( \frac{|\sin(\lambda/2)|}{|\sin((\lambda + \lambda_j)/2)|} \right)^2 =: \sup_{\frac{\lambda_j}{2} \leq \lambda \leq \pi} f_1(\lambda) \\
&= \left( \frac{1}{|\sin(\pi/2 + \lambda_j/2)|} \right)^2 \leq \left( \frac{1}{|\sin((\pi + \lambda_m)/2)|} \right)^2 \\
&= (\cos(\lambda_m/2))^{-2}, \quad (\text{D.6})
\end{aligned}$$

where (D.5) follows since  $|\sin(\lambda)|$  is  $\pi$ -periodic. The inequality (D.6) follows since the the function  $f_1(\lambda)$  can be checked to be monotonically increasing on the interval  $(\frac{\lambda_j}{2}, \pi)$

Similarly,

$$\begin{aligned}
\sup_{2\lambda_j \leq \lambda \leq \pi} \left( \frac{|D_N(\lambda - \lambda_j)|}{|D_N(\lambda)|} \right)^2 &= \sup_{2\lambda_j \leq \lambda \leq \pi} \left( \frac{|\sin(N(\lambda - \lambda_j)/2) \sin(\lambda/2)|}{|\sin(N\lambda/2) \sin((\lambda - \lambda_j)/2)|} \right)^2 \\
&= \sup_{2\lambda_j \leq \lambda \leq \pi} \left( \frac{|\sin(N\lambda/2) \sin(\lambda/2)|}{|\sin(N\lambda/2) \sin((\lambda - \lambda_j)/2)|} \right)^2 \\
&= \sup_{2\lambda_j \leq \lambda \leq \pi} \left( \frac{|\sin(\lambda/2)|}{|\sin((\lambda - \lambda_j)/2)|} \right)^2 =: \sup_{2\lambda_j \leq \lambda \leq \pi} f_2(\lambda)
\end{aligned}$$



$$= \left( \frac{|\sin(\lambda_j)|}{|\sin(\lambda_j/2)|} \right)^2 \leq 4, \quad (\text{D.7})$$

where the inequality (D.7) follows since the function  $f_2(\lambda)$  is monotonically decreasing on the interval  $(2\lambda_j, \pi)$ .  $\square$

## E Results and proofs for linear processes

In this section, we provide an extension of our results for linear processes  $X_n = \sum_{j \in \mathbb{Z}} \Psi_j \varepsilon_{n-j}$  with  $\sum_{j \in \mathbb{Z}} \|\Psi_j\|_F^2 < \infty$  and sub-Gaussian innovations.

**Remark E.1.** It is certainly conceivable to consider other innovations such as sub-exponential or with finite fourth moments. Concentration inequalities for i.i.d. random vectors which are either sub-Gaussian, sub-exponential or have finite fourth moments are respectively available in Section 5.2.3 in Rudelson and Vershynin (2013), Lemma 8.3 in Erdős, Yau, and Yin (2012) and Lemma 4.1 in Sun et al. (2018). However, our proofs require a uniform concentration inequality stated in Theorem D.1. The original proof of Theorem D.1 in Dicker and Erdogdu (2017) involves chaining techniques and subsequent application of the Hanson-Wright inequality for sub-Gaussian processes. Replacing the Hanson-Wright inequality by the respective inequalities for sub-exponential variables or data with finite fourth moments, would lead to uniform concentration inequalities as well.

For completeness, we restate Lemma B.1 here for linear processes. The result only differs from the one for Gaussian time series by a general constant  $\gamma$  which is determined by the sub-Gaussian innovations of the linear process in (D.1).

**Lemma E.1.** *Suppose  $\{X_n\}_{n \in \mathbb{Z}}$  can be represented as a  $p$ -dimensional linear process with spectral density  $f_X$  as in (1.1). Then, there are positive constants  $c_1, c_2$  such that*

$$\mathbb{P} \left( \sup_{D \in \Omega} |\widehat{H}_{rs}(D) - \mathbb{E} \widehat{H}_{rs}(D)| > \|G\| \nu \right) \leq \mathcal{B}(r, s, i), \quad i = 2, \dots, 5, \quad (\text{E.1})$$

for  $\nu^2 \geq \gamma^4 L_{rs,i}^2 / (m^2 c_2)$ , where

$$\mathcal{B}(r, s, i) = c_1 \exp \left( -c_2 \min \left\{ \frac{\nu m}{\gamma^2 \bar{\Delta}_N L_{rs,1}}, \frac{\nu^2 m^2}{\gamma^4 \bar{\Delta}_N^2 L_{rs,i}^2} \right\} \right) \quad (\text{E.2})$$

with  $r \neq s$  if  $i = 3$  and  $\Omega$  is given in (3.10). The constant  $\gamma$  bounds the sub-Gaussian norm as in (D.1),  $\bar{\Delta}_N$  is defined in (3.13) and  $L_{rs,i}$ 's as in (D.1).

*Proof:* The main idea is to consider a truncated version of the linear process and to rewrite it as a linear mapping of an i.i.d. vector. The truncated version allows us then to write the local Whittle estimator as a quadratic form in sub-Gaussian random variables.

Introduce a truncated version of the linear process as  $X_n^L = \sum_{j=-L}^L \Psi_j \varepsilon_{n-j}$ . The truncated version can be expressed as a linear transformation of an i.i.d. vector. Let  $\psi_{r,j}$  and  $\psi_{r,j}$  denote the  $r$ th column and row of  $\Psi_j$  and write the data matrix of the truncated process as  $\mathcal{X}_L = [X_1^L : \dots : X_N^L]$  with  $X_n^L = (X_{1,n}^L, \dots, X_{p,n}^L)'$ . Then,

$$e'_r \mathcal{X}'_L = (X_{r,1}^L, \dots, X_{r,N}^L) = \left( e'_r \sum_{j=-L}^L \Psi_j \varepsilon_{1-j}, \dots, e'_r \sum_{j=-L}^L \Psi_j \varepsilon_{N-j} \right)$$

$$\begin{aligned}
&= \left( \sum_{j=-L}^L \psi_{r,j} \varepsilon_{1-j}, \dots, \sum_{j=-L}^L \psi_{r,j} \varepsilon_{N-j} \right) \\
&= (\varepsilon'_{N+L}, \dots, \varepsilon'_1, \dots, \varepsilon'_{1-L}) \begin{pmatrix} 0 & 0 & \cdots & \psi'_{r,-L} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \psi'_{r,-L} & & \vdots \\ \psi'_{r,-L} & \vdots & & \psi'_{r,L} \\ \vdots & \vdots & \ddots & 0 \\ \vdots & \psi'_{r,L} & & \vdots \\ \psi'_{r,L} & 0 & \cdots & 0 \end{pmatrix} =: \mathcal{E}'_L A'_{L,r},
\end{aligned} \tag{E.3}$$

where  $A'_{L,r}$  is a  $(N+2L)p \times N$  matrix. As in (B.7) in the proof for Gaussian random variables in Lemma B.1, we separate  $|\widehat{H}_{rs}(D) - \mathbb{E}\widehat{H}_{rs}(D)|$  into diagonal elements and real and imaginary parts of the off-diagonal elements. In order to avoid replicating all steps from the proof of Lemma B.1, we focus here on the real part of the off-diagonal terms. The proofs for the diagonal elements and the imaginary part of the off-diagonal elements can be adapted analogously and are omitted.

*Real part (off-diagonal):* For the truncated linear process, the real part of the off-diagonal elements  $\Re(\widehat{H}_{rs}(D))$  can be written in terms of (B.5) evaluated at the periodogram of the truncated process  $I_{X_L}(\lambda_j)$  as

$$\frac{1}{m} e'_r \sum_{j=1}^m t_j(D) \Re(I_{X_L}(\lambda_j)) t_j(D) e_s = \frac{1}{2\pi m} e'_r \sum_{j=1}^m t_j(D) \mathcal{X}'_L (C_j C'_j + S_j S'_j) \mathcal{X}_L t_j(D) e_s. \tag{E.4}$$

As in (B.10), we write (E.4) as a quadratic form but now using  $(e'_r \mathcal{X}'_L \ e'_s \mathcal{X}'_L) = \mathcal{E}'_L A'_L$ , where  $A'_L = (A'_{L,r} \ A'_{L,s})$  and set

$$\Sigma_L = A_L A'_L = \begin{pmatrix} A_{L,r} \\ A_{L,s} \end{pmatrix} (A'_{L,r} \ A'_{L,s}). \tag{E.5}$$

Due to Lemma E.2,  $\Sigma_L$  can also be written as

$$\Sigma_L = \begin{pmatrix} \Sigma_{L,rr} & \Sigma_{L,rs} \\ \Sigma_{L,sr} & \Sigma_{L,ss} \end{pmatrix} \quad \text{with} \quad \Sigma_{L,rs} = (\Sigma_{L,rs}(n-k))_{n,k=1,\dots,N} = (\mathbb{E} X_{r,k}^L X_{s,n}^L)_{n,k=1,\dots,N}.$$

We further recall from Lemma B.1 the  $4m \times 2N$  matrix

$$\widetilde{R}_m = \begin{pmatrix} R_m & 0_{2m \times N} \\ 0_{2m \times N} & R_m \end{pmatrix}$$

with  $R_m$  as in (B.8) and the matrix  $F(d_r, d_s) = \text{diag}(T_r(d_r), T_s(d_s))$  with  $T_r$  as in (B.9) and

$$M_m = \begin{pmatrix} 0_{m,m} & I_m \\ 0_{m,m} & 0_{m,m} \end{pmatrix}. \tag{E.6}$$

Write

$$\begin{aligned}
e'_r \mathcal{X}'_L \sum_{j=1}^m t_{j,r}(d_r)(C_j C'_j + S_j S'_j) t_{j,r}(d_s) \mathcal{X}_L e_s &= (e'_r \mathcal{X}'_L \ e'_s \mathcal{X}'_L) \tilde{R}'_m F(d_r, d_s) M_{2m} F(d_r, d_s) \tilde{R}_m \begin{pmatrix} \mathcal{X}_L e_r \\ \mathcal{X}_L e_s \end{pmatrix} \\
&= (e'_r \mathcal{X}'_L \ e'_s \mathcal{X}'_L) \tilde{F}(d_r, d_s) \begin{pmatrix} \mathcal{X}_L e_r \\ \mathcal{X}_L e_s \end{pmatrix} \\
&= \mathcal{E}'_L A'_L \tilde{F}(d_r, d_s) A_L \mathcal{E}_L \\
&= \mathcal{E}'_L R_L(d_r, d_s) \mathcal{E}_L
\end{aligned} \tag{E.7}$$

with

$$R_L(d_r, d_s) = A'_L \tilde{F}(d_r, d_s) A_L \quad \text{and} \quad \tilde{F}(d_r, d_s) = \tilde{R}'_m F(d_r, d_s) M_{2m} F(d_r, d_s) \tilde{R}_m.$$

In order to apply Theorem D.1, we further write

$$\begin{aligned}
&\mathbb{P}\left(\sup_{D \in \Omega} \left| \sum_{j=1}^m e'_r(t_j(D) \mathcal{X}'_L (C_j C'_j + S_j S'_j) \mathcal{X}_L t_j(D) - \mathbb{E}(t_j(D) \mathcal{X}'_L (C_j C'_j + S_j S'_j) \mathcal{X}_L t_j(D))) e_s \right| > \pi m \|G\| \nu \right) \\
&= \mathbb{P}\left(\sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} \left| \mathcal{E}'_L R_L(d_r, d_s) \mathcal{E}_L - \mathbb{E}(\mathcal{E}'_L R_L(d_r, d_s) \mathcal{E}_L) \right| > \pi m \|G\| \nu \right).
\end{aligned} \tag{E.8}$$

Note that the matrix  $R_L(d_r, d_s)$  can be rewritten as

$$R_L(d_r, d_s) = A'_L \mathcal{A}_{N,i}^{-1} \tilde{R}'_m \mathcal{A}_{2m,i} F(d_r, d_s) M_{2m} F(d_r, d_s) \mathcal{A}_{2m,i} \tilde{R}_m \mathcal{A}_{N,i}^{-1} A_L \tag{E.9}$$

for  $i = 1, \dots, 4$  with

$$\begin{aligned}
\mathcal{A}_{m,1} &= \text{diag}(c_{r,1,N} I_m, \tilde{c}_{s,1,N} I_m), \quad \mathcal{A}_{m,2} = \text{diag}(c_{r,2,N} I_m, \tilde{c}_{s,2,N} I_m), \\
\mathcal{A}_{m,3} &= \text{diag}(c_{r,3,N} I_m, \tilde{c}_{s,3,N} I_m), \quad \mathcal{A}_{m,4} = m^{\frac{1}{4}} \mathcal{A}_{m,1}.
\end{aligned} \tag{E.10}$$

The matrices  $\mathcal{A}_{N,i}$  in (E.9) are defined by replacing  $I_m$ 's in (E.10) by  $I_N$ .

We continue to bound (E.8) by applying Theorem D.1. In order to verify the applicability of Theorem D.1, note that the matrix  $\mathcal{A}_{2m,i} F(d_r, d_s) M_{2m} F(d_r, d_s) \mathcal{A}_{2m,i}$  in (E.9) is not diagonal but can be represented as a unitary transformation of a diagonal matrix as described in (B.24). For this reason, Theorem D.1 remains applicable due to Remark D.1. In (E.11) below, we apply Theorem D.1 with  $K = 2$  and  $R = b_r - a_r \leq 1$  to obtain

$$\begin{aligned}
&\mathbb{P}\left(\sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} \left| \mathcal{E}'_L R_L(d_r, d_s) \mathcal{E}_L - \mathbb{E}(\mathcal{E}'_L R_L(d_r, d_s) \mathcal{E}_L) \right| > \pi m \|G\| \nu \right) \\
&\leq c_1 \exp\left(-c_2 \min\left\{\frac{\nu m \|G\|}{\gamma^2 \mathcal{T}_1^L}, \frac{\nu^2 m^2 \|G\|^2}{\gamma^4 (\mathcal{T}_i^L)^2}\right\}\right)
\end{aligned} \tag{E.11}$$

for  $\nu^2 \geq \gamma^4 (\mathcal{T}_i^L)^2 / (c_2 m^2 \|G\|^2)$  and  $i = 2, \dots, 5$  with

$$\mathcal{T}_1^L = \|B_{m,1}^L\|_{L_{rs,1}}, \quad \mathcal{T}_i^L = \|B_{m,i}^L\|_{L_{rs,i}}, \quad \mathcal{T}_5^L = \|B_{m,1}^L\|_{L_{rs,5}} \tag{E.12}$$

for  $i = 2, \dots, 4$ , where the  $L_{rs,i}$ 's are given in (B.1) and

$$B_{m,i}^L = \tilde{R}_m \mathcal{A}_{N,i}^{-1} \Sigma_L \mathcal{A}_{N,i}^{-1} \tilde{R}'_m \tag{E.13}$$

for  $i = 1, \dots, 4$ , which is due to the relation (E.5). Based on (E.7), an equivalent formulation of (E.11) can be given as

$$\begin{aligned} & \mathbb{P}\left(\sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} |e'_r \mathcal{X}'_L \tilde{F}(d_r, d_s) \mathcal{X}_L e_s - \mathbb{E}(e'_r \mathcal{X}'_L \tilde{F}(d_r, d_s) \mathcal{X}_L e_s)| > \pi m \|G\| \nu\right) \\ & \leq c_1 \exp\left(-c_2 \min\left\{\frac{\nu m \|G\|}{\gamma^2 \mathcal{T}_1^L}, \frac{\nu^2 m^2 \|G\|^2}{\gamma^4 (\mathcal{T}_i^L)^2}\right\}\right). \end{aligned}$$

Then, letting  $L$  go to infinity on both sides, Lemmas E.5 and E.3 lead to

$$\begin{aligned} & \mathbb{P}\left(\sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} |e'_r \mathcal{X}' \tilde{F}(d_r, d_s) \mathcal{X} e_s - \mathbb{E}(e'_r \mathcal{X}' \tilde{F}(d_r, d_s) \mathcal{X} e_s)| > \pi m \|G\| \nu\right) \\ & \leq c_1 \exp\left(-c_2 \min\left\{\frac{\nu m \|G\|}{\gamma^2 \mathcal{T}_1}, \frac{\nu^2 m^2 \|G\|^2}{\gamma^4 \mathcal{T}_i^2}\right\}\right), \end{aligned}$$

which coincides with (B.25). Then, one can use the same arguments as in Lemma B.1 following (B.25).  $\square$

The following lemma gathers some relationships between the matrices  $A_{L,r}$  in (E.3) and the autocovariance matrices of the truncated linear process. For completeness, we also state the representation of the autocovariance function for the original linear process.

**Lemma E.2.** *The autocovariance matrix and the coefficient matrices of the linear process relate as follows. For  $k \leq n$ ,*

$$\begin{aligned} \Sigma_{L,rs}(n-k) &= \sum_{j=-L}^{L-(n-k)} \psi_{r, j+(n-k)} \psi'_{s,j}, \quad \Sigma_{rs}(n-k) = \sum_{j \in \mathbb{Z}} \psi_{r, j+(n-k)} \psi'_{s,j}, \\ A_{L,s} A'_{L,r} &= \Sigma_{L,rs}(k-n). \end{aligned} \tag{E.14}$$

Furthermore,  $\Sigma_{L,rs}(k-n) = (\Sigma_{L,sr}(n-k))'$  and  $\Sigma_{rs}(k-n) = (\Sigma_{sr}(n-k))'$ .

*Proof:* Recall  $X_r^L = (X_{r,1}^L, \dots, X_{r,N}^L)'$  and suppose  $k \leq n$ . Then,

$$\begin{aligned} \Sigma_{L,rs}(k-n) &= \mathbb{E} X_{r,k}^L X_{s,n}^{L'} = \mathbb{E} \left( \sum_{j=-L}^L \psi_{r, j \varepsilon_{n-j}} \sum_{j=-L}^L \psi_{s, j \varepsilon_{k-j}} \right) \\ &= \sum_{j_1=n-L}^{n+L} \sum_{j_2=k-L}^{k+L} \psi_{r, n-j_1} \mathbb{E}(\varepsilon_{j_1} \varepsilon'_{j_2}) \psi'_{s, k-j_2} \\ &= \sum_{j=\max\{k,n\}-L}^{\min\{k,n\}+L} \psi_{r, n-j} \psi'_{s, k-j} \\ &= \sum_{j=-L}^{L-(n-k)} \psi_{r, j+(n-k)} \psi'_{s,j}. \end{aligned}$$

Analogously, for the non-truncated version of the linear process leading to the second relation in (E.14). Finally, the  $kn$ th element of  $A_{L,r} A'_{L,s}$  satisfies

$$(A_{L,r} A'_{L,s})_{kn} = \sum_{j=-L}^{L-(n-k)} \psi_{r, j+(n-k)} \psi'_{s,j} = \Sigma_{L,rs}(n-k).$$

□

The next three lemmas follow the ideas in Sun et al. (2018), but also allow for long-range dependence and address that we need a uniform concentration inequality.

**Lemma E.3.** Recall  $B_{m,i}^L = \tilde{R}_m \mathcal{A}_{N,i}^{-1} \Sigma_L \mathcal{A}_{N,i}^{-1} \tilde{R}'_m$  in (E.13) with  $\Sigma_L$  as in (E.5). Then,

$$\|B_{m,i}^L\| \rightarrow \|B_{m,i}\| \quad \text{and} \quad \|B_{m,i}^L\|_F \rightarrow \|B_{m,i}\|_F, \quad \text{as } L \rightarrow \infty, \quad (\text{E.15})$$

with  $B_{m,i} = \tilde{R}_m \mathcal{A}_{N,i}^{-1} \Sigma \mathcal{A}_{N,i}^{-1} \tilde{R}'_m$  as in (B.28).

*Proof:* It is sufficient to prove  $\|B_{m,i}^L - B_{m,i}\|_F \rightarrow 0$  as  $L \rightarrow \infty$  since by triangle inequality

$$\begin{aligned} \left| \|B_{m,i}^L\|_F - \|B_{m,i}\|_F \right| &\leq \|B_{m,i}^L - B_{m,i}\|_F, \\ \left| \|B_{m,i}^L\| - \|B_{m,i}\| \right| &\leq \|B_{m,i}^L - B_{m,i}\| \leq \|B_{m,i}^L - B_{m,i}\|_F. \end{aligned}$$

We can further reduce the problem as follows:

$$\|B_{m,i}^L - B_{m,i}\|_F = \|\tilde{R}_m \mathcal{A}_{N,i}^{-1} (\Sigma_L - \Sigma) \mathcal{A}_{N,i}^{-1} \tilde{R}'_m\|_F \leq \|\mathcal{A}_{N,i}^{-1} (\Sigma_L - \Sigma) \mathcal{A}_{N,i}^{-1}\|_F \quad (\text{E.16})$$

$$\leq \|\mathcal{A}_{N,i}^{-2}\| \|\Sigma_L - \Sigma\|_F \leq \|\mathcal{A}_{N,i}^{-2}\| (\|\Sigma_{L,rr} - \Sigma_{rr}\|_F + \|\Sigma_{L,ss} - \Sigma_{ss}\|_F). \quad (\text{E.17})$$

We used Lemma D.1 and  $\|\tilde{R}_m\|^2 = 1$  in (E.16). The first inequality in (E.17) also follows by Lemma D.1. Note that  $\mathcal{A}_{N,i}$  are diagonal matrices which do not depend on  $L$ . The second inequality in (E.17) is due to Lemma D.2. Finally,

$$\begin{aligned} &\|\Sigma_{L,rr} - \Sigma_{rr}\|_F^2 \\ &= \left\| \left( \sum_{j=L-(n-k)+1}^{\infty} \psi_{r,j+(n-k)} \psi'_{r,j} + \sum_{j=-\infty}^{-L-1} \psi_{r,j+(n-k)} \psi'_{r,j} \right)_{k,n=1,\dots,N} \right\|_F^2 \\ &\leq N^2 \max_{k,n=1,\dots,N} \left| \sum_{i=L+1}^{\infty} \psi_{r,i} \psi'_{r,i-(n-k)} \right|^2 + N^2 \max_{k,n=1,\dots,N} \left| \sum_{i=L+1}^{\infty} \psi_{r,-i+(n-k)} \psi'_{r,-i} \right|^2 \\ &\leq N^2 \max_{k,n=1,\dots,N} \left( \left| \sum_{i=L+1}^{\infty} \|\psi_{r,i}\|_F \|\psi_{r,i-(n-k)}\|_F \right|^2 + \left| \sum_{i=L+1}^{\infty} \|\psi_{r,-i+(n-k)}\|_F \|\psi_{r,-i}\|_F \right|^2 \right) \quad (\text{E.18}) \end{aligned}$$

$$\begin{aligned} &\leq N^2 \max_{k,n=1,\dots,N} \left( \sum_{i=L+1}^{\infty} \|\psi_{r,i}\|_F^2 \sum_{i=L+1}^{\infty} \|\psi_{r,i-(n-k)}\|_F^2 \right. \\ &\quad \left. + \sum_{i=L+1}^{\infty} \|\psi_{r,-i+(n-k)}\|_F^2 \sum_{i=L+1}^{\infty} \|\psi_{r,-i}\|_F^2 \right) \rightarrow 0 \text{ as } L \rightarrow \infty, \quad (\text{E.19}) \end{aligned}$$

where both (E.18) and (E.19) follow by the Cauchy-Schwarz inequality. □

**Lemma E.4.** Recall the data matrices  $\mathcal{X}_L = [X_1^L : \dots : X_N^L]$  and  $\mathcal{X} = [X_1 : \dots : X_N]$ . Then,

$$\mathbb{E} \|e'_r \mathcal{X}'_L - e'_r \mathcal{X}'\|_F^2 \rightarrow 0, \quad \text{as } L \rightarrow \infty.$$

*Proof:* The expected value can be explicitly calculated as

$$\begin{aligned}
\mathbb{E}\|e'_r \mathcal{X}'_L - e'_r \mathcal{X}'\|_F^2 &= \sum_{n=1}^N \mathbb{E}|X_{r,n}^L - X_{r,n}|^2 = \sum_{n=1}^N \mathbb{E} \left| \sum_{j=L+1}^{\infty} \psi_{r,j} \varepsilon_{n-j} + \sum_{j=-\infty}^{-L-1} \psi_{r,j} \varepsilon_{n-j} \right|^2 \\
&\leq 2 \sum_{n=1}^N \left( \mathbb{E} \left| \sum_{j=L+1}^{\infty} \psi_{r,j} \varepsilon_{n-j} \right|^2 + \mathbb{E} \left| \sum_{j=-\infty}^{-L-1} \psi_{r,j} \varepsilon_{n-j} \right|^2 \right) \\
&= 2 \sum_{n=1}^N \sum_{j_1, j_2=L+1}^{\infty} (\psi_{r,j_1} \mathbb{E}(\varepsilon_{n-j_1} \varepsilon'_{n-j_2}) \psi'_{r,j_2} + \psi_{r,-j_1} \mathbb{E}(\varepsilon_{n+j_1} \varepsilon'_{n+j_2}) \psi'_{r,-j_2}) \\
&= 2N \sum_{j=L+1}^{\infty} (\psi_{r,j} \psi'_{r,j} + \psi_{r,-j} \psi'_{r,-j}) < \infty
\end{aligned}$$

since  $\sum_{j=0}^{\infty} \|\Psi_j\|_F^2 < \infty$ . Due to the summability of the last expression, letting  $L \rightarrow \infty$  yields the desired convergence result.  $\square$

**Lemma E.5.** *The following convergence result holds:*

$$\begin{aligned}
&\sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} |e'_r \mathcal{X}'_L \tilde{F}(d_r, d_s) \mathcal{X}_L e_s - \mathbb{E}(e'_r \mathcal{X}'_L \tilde{F}(d_r, d_s) \mathcal{X}_L e_s)| \\
&\xrightarrow{d} \sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} |e'_r \mathcal{X}' \tilde{F}(d_r, d_s) \mathcal{X} e_r - \mathbb{E}(e'_r \mathcal{X}' \tilde{F}(d_r, d_s) \mathcal{X} e_r)|, \quad \text{as } L \rightarrow \infty.
\end{aligned}$$

*Proof:* Note that Lemma E.4 implies the convergence in probability,

$$e'_r \mathcal{X}'_L \xrightarrow{P} e'_r \mathcal{X}'.$$

We write  $C_b([a_r, b_r] \times [a_s, b_s], \mathbb{R}^{N \times N})$  for the space of  $\mathbb{R}^{N \times N}$ -valued, bounded and continuous functions on  $[a_r, b_r] \times [a_s, b_s]$ . The space is equipped with the uniform metric

$$\|A - B\|_{\infty} = \sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} \|A(d_r, d_s) - B(d_r, d_s)\|_F.$$

Since the  $\mathbb{R}^N \rightarrow C_b([a_r, b_r] \times [a_s, b_s], \mathbb{R})$  function  $x \mapsto x' A x$  is continuous, the continuous mapping theorem implies

$$e'_r \mathcal{X}'_L \tilde{F}(d_r, d_s) \mathcal{X}_L e_r \xrightarrow{P} e'_r \mathcal{X}' \tilde{F}(d_r, d_s) \mathcal{X} e_r. \quad (\text{E.20})$$

Lemma E.4 also allows us to infer

$$\mathbb{E}(e'_r \mathcal{X}'_L \tilde{F}(d_r, d_s) \mathcal{X}_L e_r) \rightarrow \mathbb{E}(e'_r \mathcal{X}' \tilde{F}(d_r, d_s) \mathcal{X} e_r) \quad (\text{E.21})$$

since for any matrix function  $A \in C_b([a_r, b_r] \times [a_s, b_s], \mathbb{R}^{N \times N})$

$$\begin{aligned}
&\|\mathbb{E}(e'_r \mathcal{X}'_L A \mathcal{X}_L e_r) - \mathbb{E}(e'_r \mathcal{X}' A \mathcal{X} e_r)\|_{\infty} \\
&\leq \|\mathbb{E}(e'_r \mathcal{X}'_L A (\mathcal{X}_L e_r - \mathcal{X} e_r))\|_{\infty} + \|\mathbb{E}((e'_r \mathcal{X}' - e'_r \mathcal{X}'_L) A \mathcal{X} e_r)\|_{\infty}
\end{aligned}$$

and

$$\|\mathbb{E}(e'_r \mathcal{X}'_L A (\mathcal{X}_L e_r - \mathcal{X} e_r))\|_{\infty}^2 \leq \|A\|_{\infty}^2 \mathbb{E}\|e'_r \mathcal{X}'_L\|_F^2 \mathbb{E}\|e'_r \mathcal{X}'_L - e'_r \mathcal{X}'\|_F^2 \rightarrow 0.$$

Combining (E.20) and (E.21) yields

$$\begin{aligned} & e_r' \mathcal{X}'_L \tilde{F}(d_r, d_s) \mathcal{X}_L e_r - \mathbb{E}(e_r' \mathcal{X}'_L \tilde{F}(d_r, d_s) \mathcal{X}_L e_r) \\ & \xrightarrow{P} e_r' \mathcal{X}' \tilde{F}(d_r, d_s) \mathcal{X} e_r - \mathbb{E}(e_r' \mathcal{X}' \tilde{F}(d_r, d_s) \mathcal{X} e_r). \end{aligned}$$

Convergence in probability implies convergence in distribution. Finally, the continuous mapping theorem with  $C_b([a_r, b_r] \times [a_s, b_s], \mathbb{R}^{N \times N}) \rightarrow \mathbb{R}$  and  $A(d_r, d_s) \mapsto \sup_{(d_r, d_s) \in [a_r, b_r] \times [a_s, b_s]} |A(d_r, d_s)|$  proves the claim of the lemma.  $\square$

## F Complementary simulation results

### F.1 Performance measures

This section presents our simulation results for the thresholding and graphical local Whittle estimators discussed in Section 5. A number of different measures are used to assess the performance of our estimators.

The first five measures in Tables F.3 and F.4 concern whether the matrix entries are correctly estimated as zero. We use the standard true positive (TP), false negative (FN), false positive (FP) and true negative (TN) measures. For instance, TP quantifies how often the respective estimator correctly detected a nonzero matrix entry as nonzero. For the respective rates, we write TPR, FNR, TNR and FPR. The fifth metric we consider is the so-called precision calculated as  $\text{TP}/(\text{TP}+\text{FP})$  which measures the estimator's accuracy in classifying entries as nonzero. The introduced measures perform well, even for small sample size  $N$  and dimension  $p$  and improve with increasing  $N$  and  $p$ .

The last five measures in Tables F.3 and F.4 consider different distance measures between our estimators and the true quantities. Besides the mean squared error for the memory parameters ( $\text{MSE}(D)$ ), we consider the distance measures (Frobenius and spectral norm) used in our main results to prove consistency of our estimators (Propositions 3.6 and 3.7). The results have been discussed in more detail in Section 5. Finally, Rel-Frobenius and -spectral denote the ratio between the distance of our sparse estimator and the distance between the associated nonsparse estimator. For example, Rel-Frobenius for Tables F.3 and F.4 are respectively defined as

$$\frac{\|T_\rho(\hat{G}(\hat{D})) - G_0\|_F}{\|\hat{G}(\hat{D}) - G_0\|_F}, \quad \frac{\|\hat{P}_\rho(\hat{D}) - P_0\|_F}{\|\hat{P}(\hat{D}) - P_0\|_F}, \quad (\text{F.1})$$

where  $\hat{G}(\hat{D})$  is a nonsparse estimator of  $G_0$  as defined in (1.6) which is the same as setting the threshold  $\rho = 0$  and analogously  $P(\hat{D}) := \hat{P}_0(\hat{D})$ . Rel-spectral can be defined analogously by replacing the Frobenius norm with the spectral norm in (F.1). In Tables F.3 and F.4, Rel-Frobenius and -spectral take values smaller than one, which indicates that the denominators in (F.1) take values larger than the numerators, that is, the nonsparse estimators perform worse for our sparse DGPs.

Models	Measures	$p = 20$			$p = 40$			$p = 60$		
		$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$
thDGP1	TPR	0.991	0.999	1.000	0.976	0.999	1.000	0.975	0.999	1.000
	FNR	0.009	0.001	0.000	0.024	0.001	0.000	0.025	0.001	0.000
	TNR	0.991	0.999	0.996	0.999	0.999	0.996	0.999	0.998	0.999
	FPR	0.009	0.001	0.004	0.001	0.001	0.004	0.001	0.002	0.001
	Precision	0.899	0.984	0.949	0.973	0.964	0.900	0.960	0.936	0.963
	MSE( $D$ )	0.095	0.050	0.022	0.193	0.102	0.045	0.295	0.155	0.069
	Frobenius	0.174	0.110	0.088	0.228	0.171	0.139	0.294	0.228	0.146
	Spectral	0.108	0.066	0.055	0.115	0.092	0.075	0.139	0.113	0.070
	Rel-Frobenius	0.379	0.318	0.367	0.248	0.247	0.293	0.213	0.219	0.204
	Rel-spectral	0.450	0.378	0.452	0.294	0.321	0.401	0.268	0.299	0.282
thDGP2	TPR	0.968	0.996	1.000	0.947	0.994	1.000	0.936	0.990	1.000
	FNR	0.032	0.004	0.000	0.053	0.007	0.000	0.064	0.010	0.000
	TNR	0.997	0.998	0.999	0.999	0.999	1.000	1.000	1.000	1.000
	FPR	0.003	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000
	Precision	0.958	0.972	0.987	0.973	0.978	0.993	0.977	0.985	0.992
	MSE( $D$ )	0.095	0.052	0.023	0.197	0.104	0.047	0.293	0.157	0.073
	Frobenius	2.569	1.949	1.374	5.047	3.837	2.671	7.512	5.692	3.960
	Spectral	1.253	0.911	0.606	1.997	1.430	0.926	2.629	1.846	1.191
	Rel-Frobenius	0.358	0.364	0.421	0.266	0.265	0.311	0.223	0.220	0.259
	Rel-spectral	0.403	0.404	0.497	0.303	0.297	0.365	0.247	0.250	0.298
thDGP3	TPR	0.939	0.995	1.000	0.877	0.990	1.000	0.835	0.982	1.000
	FNR	0.061	0.005	0.000	0.123	0.010	0.000	0.165	0.018	0.000
	TNR	0.981	0.980	0.990	0.995	0.992	0.997	0.997	0.997	0.998
	FPR	0.019	0.020	0.010	0.005	0.008	0.003	0.003	0.003	0.002
	Precision	0.900	0.900	0.945	0.930	0.913	0.969	0.941	0.943	0.962
	MSE( $D$ )	0.106	0.064	0.034	0.215	0.128	0.068	0.323	0.192	0.100
	Frobenius	2.666	2.211	1.817	4.943	3.965	3.073	7.231	5.695	4.302
	Spectral	1.170	0.935	0.769	1.848	1.342	0.952	2.471	1.758	1.155
	Rel-Frobenius	0.690	0.725	0.803	0.554	0.576	0.665	0.484	0.489	0.585
	Rel-spectral	0.665	0.701	0.750	0.473	0.520	0.619	0.378	0.416	0.524

Table F.3: Simulation results for the thresholding local Whittle estimator.



Models	Measures	$p = 20$			$p = 40$			$p = 60$		
		$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$
DGP1	TPR	0.957	0.994	1.000	0.943	0.980	1.000	0.940	0.973	1.000
	FNR	0.043	0.006	0.000	0.057	0.020	0.000	0.060	0.027	0.000
	TNR	1.000	1.000	1.000	0.912	1.000	1.000	0.887	0.942	1.000
	FPR	0.000	0.000	0.000	0.088	0.000	0.000	0.113	0.058	0.000
	Precision	1.000	0.998	1.000	0.910	0.998	0.999	0.776	0.941	0.998
	MSE( $D$ )	0.095	0.050	0.022	0.193	0.102	0.045	0.292	0.157	0.070
	Frobenius	4.911	3.197	2.323	7.490	5.066	3.367	10.365	6.760	4.080
	Spectral	3.006	1.592	1.137	3.892	2.411	1.335	4.498	3.446	1.407
	Rel-Frobenius	0.184	0.209	0.258	0.065	0.112	0.159	0.012	0.064	0.108
	Rel-spectral	0.167	0.177	0.238	0.053	0.103	0.139	0.007	0.063	0.089
DGP2	TPR	0.935	0.989	1.000	0.956	0.988	1.000	0.969	0.985	1.000
	FNR	0.065	0.011	0.000	0.044	0.012	0.000	0.031	0.015	0.000
	TNR	1.000	1.000	1.000	0.782	1.000	1.000	0.556	0.958	1.000
	FPR	0.000	0.000	0.000	0.218	0.000	0.000	0.444	0.042	0.000
	Precision	1.000	1.000	0.999	0.787	1.000	0.999	0.550	0.959	1.000
	MSE( $D$ )	0.095	0.053	0.024	0.196	0.103	0.047	0.292	0.159	0.071
	Frobenius	0.913	0.670	0.574	1.305	0.948	0.800	1.735	1.192	0.975
	Spectral	0.586	0.366	0.325	0.660	0.414	0.354	0.775	0.495	0.370
	Rel-Frobenius	0.219	0.273	0.386	0.070	0.134	0.236	0.014	0.072	0.163
	Rel-spectral	0.217	0.259	0.426	0.058	0.114	0.236	0.009	0.059	0.152
DGP3	TPR	0.350	0.380	0.659	0.534	0.361	0.599	0.648	0.410	0.539
	FNR	0.650	0.620	0.341	0.466	0.639	0.401	0.352	0.590	0.461
	TNR	1.000	1.000	0.950	0.714	1.000	0.987	0.540	0.915	0.996
	FPR	0.000	0.000	0.050	0.286	0.000	0.013	0.460	0.085	0.004
	Precision	0.999	0.997	0.757	0.735	0.999	0.815	0.559	0.919	0.876
	MSE( $D$ )	0.093	0.048	0.022	0.188	0.098	0.044	0.286	0.149	0.067
	Frobenius	1.547	1.489	1.201	2.204	2.122	1.772	2.727	2.611	2.281
	Spectral	0.608	0.582	0.489	0.638	0.600	0.523	0.668	0.612	0.545
	Rel-Frobenius	0.380	0.615	0.790	0.126	0.306	0.523	0.024	0.161	0.383
	Rel-spectral	0.235	0.421	0.649	0.060	0.166	0.352	0.008	0.075	0.223

Table F.4: Simulation results for the graphical local Whittle estimator.

## F.2 Illustration: Rate of convergence

To illustrate our non-asymptotic results numerically, consider the result in Corollary A.1. Corollary A.1 is a consequence of our main result Proposition 3.5 for only short- or long-range dependence, that is, the true memory parameters satisfy  $D_0 \succcurlyeq 0$ . Then, using (A.4) with  $\mathcal{R}_1 \equiv \mathcal{R}_{11}$  and  $m = N^8$ , we get

$$\|\widehat{G}(\widehat{D}) - G_0\|_{\max} \approx \sqrt{\frac{\log p}{N \cdot 8 \times (1 - 2\Delta_2)}}$$

with high probability. Taking the log-transformation gives

$$\log(\|\widehat{G}(\widehat{D}) - G_0\|_{\max}) \approx .5(\log(\log(p)) - .8(1 - 2\Delta_2)\log N). \quad (\text{F.2})$$

We checked this relationship by considering the model thDGP3 with sample sizes  $N = 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000$  and dimensions  $p = 20, 40, 60, 80, 100, 120$ .

In Figure 8 we fixed the sample size as  $N = 200, 400, 600$  and consider (F.2) as a function in  $\log(\log(p))$ . The expected slope is  $.5$  in the log-log plot. Figure 8 shows that the estimated slope is close to  $.5$ .

In Figure 9 we fixed the dimension  $p = 20, 30, 40$ . In this setting, the expected slope for  $\log N$  is  $-.104$  calculated as

$$.5 \times .8(1 - 2\Delta_2) = .4 \times (1 - 2 \max_{r=1, \dots, r} \widehat{d}_r) = .104,$$

where we chose  $\Delta_2$  as the largest LRD parameter across all dimensions. In Figure 9, it can be seen that the slope is close to  $-.104$ , as expected.

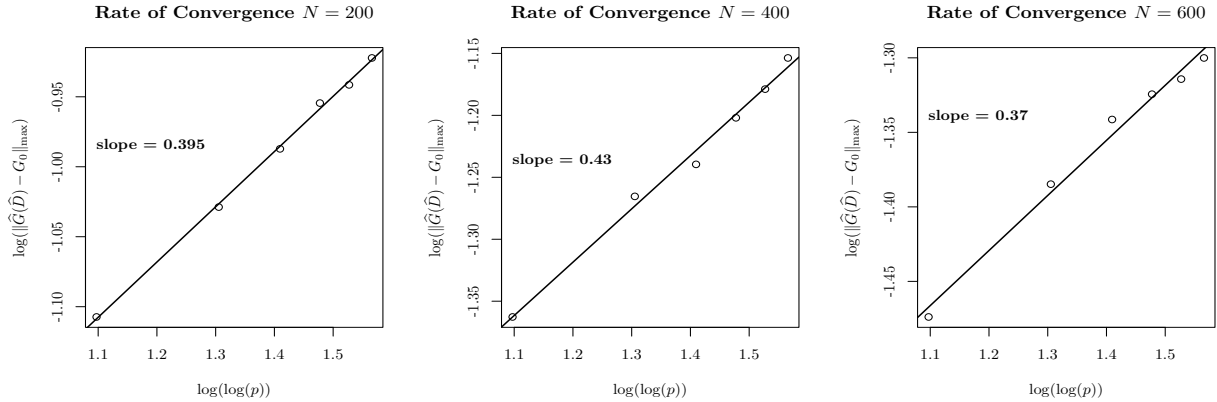


Figure 8: The rate of convergence with fixed sample size  $N$ . Theoretically expected slope is  $.5$ .

## F.3 Univariate versus multivariate estimation of the memory parameters

From a theoretical perspective, Remark 2.2 argues that our proofs remain valid for multivariate estimation of  $D_0$ . However, an additional  $p$  would appear in the bounds and significantly weaken the results.

In Table F.5 below, we report on a small simulation study evaluating the efficiency of our proposed estimator compared to multivariate estimation and also a shrinkage estimator. The shrinkage estimator is obtained by plugging the thresholded estimator  $T_\rho(\widehat{G}(D))$  in (2.2) into the

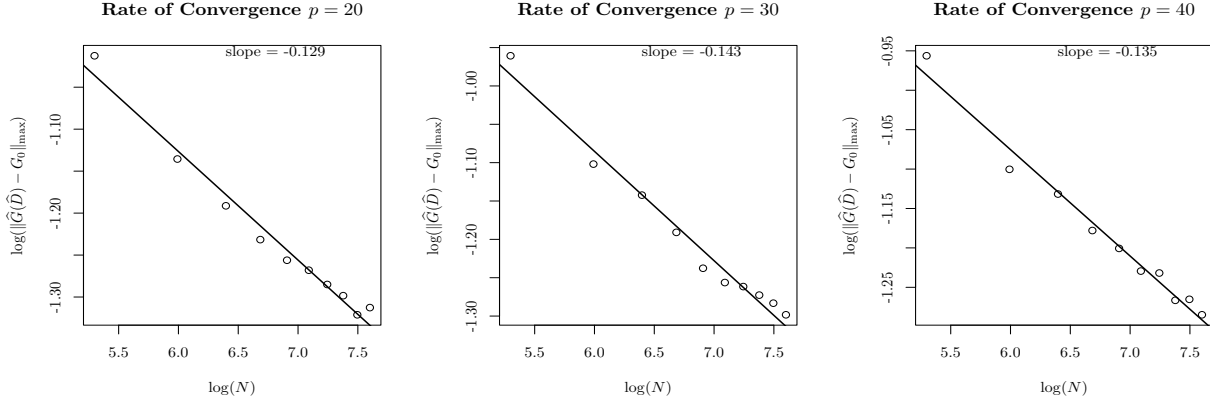


Figure 9: The rate of convergence with fixed dimension  $p$ . Theoretically expected slope is  $-0.104$ .

local Whittle estimation for  $D_0$  in (1.5) and is labeled as “Threshold” in Table F.5. We use the same DGPs as introduced in Section 5 with sample size  $N = 200$  and dimension  $p = 40$ . For comparison we use the mean squared error of the estimated  $D_0$  ( $\text{MSE}(D)$ ) and study the impact on  $\hat{P}(\hat{D})$  of the different estimation procedures by calculating the Frobenius and spectral distances to the true  $P_0$ . The computational time is measured as the duration time for five repetitions.

Models	Methods	MSE( $D$ )	Frobenius	Spectral	Time
DGP1	Univariate	0.190	7.760	3.977	2 sec
	Multivariate	0.190	7.745	3.964	8 sec
	Threshold	0.189	7.744	3.964	20 sec
DGP2	Univariate	0.196	1.413	0.707	2 sec
	Multivariate	0.287	1.478	0.751	146 sec
	Threshold	0.196	1.449	0.717	270 sec
DGP3	Univariate	0.186	2.206	0.642	2 sec
	Multivariate	0.298	2.218	0.646	102 sec
	Threshold	0.194	2.201	0.639	240 sec

Table F.5: Performance measures for different estimation methods of the memory parameters  $D_0$ .

As can be seen in the table, estimating the memory parameters  $D_0$  univariately performs well compared to the other methods (Multivariate, Threshold). Furthermore, the univariate estimation reduces computational time dramatically compared to multivariate or thresholded estimation. Note also that the mean squared error appears to be larger for multivariate estimation than for univariate estimation. One would expect the asymptotic variance of the multivariate estimators for  $D_0$  to be smaller compared to the univariate case which is not reflected in the simulation results. We suspect that this may be due to numerical optimization issues with multivariate estimation. The Threshold method seems to perform slightly better than multivariate estimation. Since all our DGPs are highly sparse (see Figure 1), it may not come as a surprise that the univariate and thresholded estimators perform well.

#### F.4 Modified precision matrix estimators and CLIME

In this section, we compare the performance of our graphical local Whittle estimator for the precision matrix with the alternative estimators presented in Section 3.3. The modified precision matrix

estimators in Section 3.3 are an estimator based on the coherence matrix (Section 3.3.1) and a local Whittle CLIME estimator (Section 3.3.2).

For the coherence-based graphical local Whittle, we use Algorithm 2 as introduced in Section 4.2. Similar to the graphical local Whittle estimator, the local Whittle CLIME estimator can be computed by using an ADMM algorithm. Wang, Banerjee, Hsieh, Ravikumar, and Dhillon (2013) proposed an algorithm for sparse inverse covariance matrix estimation. We modify their Algorithm 1 to our setting, using it for estimation in the spectral domain. The details can be found in Algorithm 3. For the algorithm, recall the function  $\text{shrink}(M, \nu) = \text{sign}(M_{rs}) \max(|M_{rs}| - \nu, 0)$  for a matrix  $M = (M_{rs})_{r,s=1,\dots,p}$  and some  $\nu \geq 0$ .

---

**Algorithm 3:** Alternating direction method of multipliers (ADMM) algorithm for local Whittle CLIME estimator

---

**Input:** Initial estimator  $P^0 = \text{diag}((\widehat{G}(\widehat{D}) + .1I_p)^{-1})$ ,  $V^0 = Y^0 = 0$ ,  $\mu$ ,  $\rho$ .

**Output:** Sparse estimation of  $P$ .

**Repeat until convergence:**

**for**  $k = 0, 1, \dots$  **do**

1.  $P^{k+1} = \text{shrink}(P^k - V^k, \mu)$ .
2.  $U^{k+1} = \widehat{G}(\widehat{D})P^{k+1} + Y^k$ .
3.  $Z^{k+1} = Y^k + \text{shrink}(I_p - U^{k+1}, \lambda)$ .
4.  $Y^{k+1} = U^{k+1} - Z^{k+1}$ .
5.  $V^{k+1} = \widehat{G}(\widehat{D})(2Y^{k+1} - Y^k)(\rho\mu)$ .

**end**

---

Comparison results are summarized in Tables F.6–F.8. Note that we also present the simulation results for our graphical local Whittle estimator (LW-GLASSO). All methods’ performances improve with increasing sample size. The modified precision matrix estimator, labeled as LW-GLASSO (modified) in Tables F.6–F.8 performs very similar to the classical LW-GLASSO. Our study does not reveal any significant differences between the two estimators. On the other hand, the local Whittle CLIME estimator tends to result in higher TNR while the LW-GLASSO gives higher TPR. For example, in Table F.8 when  $p = 60$  and  $N = 200$ , LW-GLASSO gives .54 for TNR while LW-CLIME gives .999. It means that LW-GLASSO is good at finding a nonzero coefficient as nonzero, while LW-CLIME finds a zero coefficient as zero.

Methods	Measures	$p = 20$			$p = 40$			$p = 60$			
		$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$	
LW-GLASSO	TPR	0.957	0.994	1.000	0.943	0.980	1.000	0.940	0.973	1.000	
	FNR	0.043	0.006	0.000	0.057	0.020	0.000	0.060	0.027	0.000	
	TNR	1.000	1.000	1.000	0.912	1.000	1.000	0.887	0.942	1.000	
	FPR	0.000	0.000	0.000	0.088	0.000	0.000	0.113	0.058	0.000	
	Precision	1.000	0.998	1.000	0.910	0.998	0.999	0.776	0.941	0.998	
	Frobenius	4.911	3.197	2.323	7.490	5.066	3.367	10.365	6.760	4.080	
	Spectral	3.006	1.592	1.137	3.892	2.411	1.335	4.498	3.446	1.407	
	Rel-Frobenius	0.184	0.209	0.258	0.065	0.112	0.159	0.012	0.064	0.108	
	Rel-spectral	0.167	0.177	0.238	0.053	0.103	0.139	0.007	0.063	0.089	
		TPR	0.960	0.998	1.000	0.964	0.996	1.000	0.965	0.995	1.000
LW-GLASSO (modified)	FNR	0.040	0.002	0.000	0.036	0.004	0.000	0.035	0.005	0.000	
	TNR	1.000	1.000	1.000	0.632	0.999	1.000	0.547	0.788	1.000	
	FPR	0.000	0.000	0.000	0.368	0.001	0.000	0.453	0.212	0.000	
	Precision	1.000	1.000	0.999	0.639	0.999	1.000	0.539	0.793	1.000	
	Frobenius	4.870	3.137	2.323	8.101	4.575	3.365	11.115	5.857	4.076	
	Spectral	2.973	1.559	1.139	4.352	1.847	1.335	5.189	2.269	1.401	
	Rel-Frobenius	0.183	0.207	0.258	0.072	0.103	0.159	0.014	0.056	0.108	
	Rel-spectral	0.166	0.174	0.238	0.062	0.080	0.139	0.009	0.042	0.089	
		TPR	0.975	0.998	1.000	0.864	0.995	1.000	0.864	0.984	1.000
		FNR	0.025	0.002	0.000	0.136	0.005	0.000	0.136	0.016	0.000
LW-CLIME	TNR	0.996	0.994	0.998	1.000	0.995	0.997	1.000	0.996	0.997	
	FPR	0.004	0.006	0.002	0.000	0.005	0.003	0.000	0.004	0.003	
	Precision	0.956	0.935	0.972	0.998	0.887	0.930	1.000	0.875	0.899	
	Frobenius	4.595	3.348	2.388	10.145	5.056	3.535	12.358	6.494	4.400	
	Spectral	2.561	1.728	1.183	4.963	2.242	1.465	5.044	2.775	1.648	
	Rel-Frobenius	0.176	0.219	0.265	0.085	0.114	0.167	0.013	0.061	0.116	
	Rel-spectral	0.151	0.192	0.248	0.066	0.095	0.153	0.007	0.051	0.105	

Table F.6: Simulation results for the modified graphical local Whittle and CLIME estimators with DGPI.

Methods	Measures	$p = 20$			$p = 40$			$p = 60$			
		$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$	
LW-GLASSO	TPR	0.935	0.989	1.000	0.956	0.988	1.000	0.969	0.985	1.000	
	FNR	0.065	0.011	0.000	0.044	0.012	0.000	0.031	0.015	0.000	
	TNR	1.000	1.000	1.000	0.782	1.000	1.000	0.556	0.958	1.000	
	FPR	0.000	0.000	0.000	0.218	0.000	0.000	0.444	0.042	0.000	
	Precision	1.000	1.000	0.999	0.787	1.000	0.999	0.550	0.959	1.000	
	Frobenius	0.913	0.670	0.574	1.305	0.948	0.800	1.735	1.192	0.975	
	Spectral	0.586	0.366	0.325	0.660	0.414	0.354	0.775	0.495	0.370	
	Rel-Frobenius	0.219	0.273	0.386	0.070	0.134	0.236	0.014	0.072	0.163	
	Rel-spectral	0.217	0.259	0.426	0.058	0.114	0.236	0.009	0.059	0.152	
		TPR	0.928	0.989	1.000	0.938	0.988	1.000	0.940	0.985	1.000
LW-GLASSO (modified)	FNR	0.073	0.011	0.000	0.062	0.012	0.000	0.060	0.015	0.000	
	TNR	1.000	1.000	1.000	0.796	1.000	1.000	0.661	0.953	1.000	
	FPR	0.000	0.000	0.000	0.204	0.000	0.000	0.339	0.047	0.000	
	Precision	1.000	1.000	0.999	0.802	1.000	0.999	0.667	0.954	1.000	
	Frobenius	0.926	0.670	0.574	1.383	0.948	0.800	1.824	1.203	0.975	
	Spectral	0.604	0.366	0.326	0.728	0.415	0.354	0.832	0.511	0.370	
	Rel-Frobenius	0.224	0.273	0.386	0.075	0.134	0.236	0.014	0.073	0.163	
	Rel-spectral	0.224	0.259	0.426	0.065	0.114	0.236	0.008	0.060	0.152	
		TPR	0.951	0.991	1.000	0.957	0.997	0.999	0.954	0.996	1.000
		FNR	0.049	0.009	0.000	0.043	0.003	0.001	0.046	0.004	0.000
LW-CLIME	TNR	1.000	1.000	0.999	0.999	0.998	1.000	0.999	0.999	1.000	
	FPR	0.000	0.000	0.001	0.001	0.002	0.000	0.001	0.001	0.000	
	Precision	0.995	0.998	0.986	0.966	0.953	0.997	0.969	0.944	0.997	
	Frobenius	0.870	0.673	0.580	1.280	0.976	0.804	1.588	1.238	0.978	
	Spectral	0.526	0.368	0.328	0.660	0.419	0.357	0.701	0.470	0.371	
	Rel-Frobenius	0.209	0.273	0.388	0.068	0.138	0.237	0.011	0.075	0.164	
	Rel-spectral	0.194	0.259	0.428	0.055	0.115	0.238	0.006	0.056	0.152	

Table F.7: Simulation results for the modified graphical local Whittle and CLIME estimators with DGP2.

Methods	Measures	$p = 20$			$p = 40$			$p = 60$			
		$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$	
LW-GLASSO	TPR	0.350	0.380	0.659	0.534	0.361	0.599	0.648	0.410	0.539	
	FNR	0.650	0.620	0.341	0.466	0.639	0.401	0.352	0.590	0.461	
	TNR	1.000	1.000	0.950	0.714	1.000	0.987	0.540	0.915	0.996	
	FPR	0.000	0.000	0.050	0.286	0.000	0.013	0.460	0.085	0.004	
	Precision	0.999	0.997	0.757	0.735	0.999	0.815	0.559	0.919	0.876	
	Frobenius	1.547	1.489	1.201	2.204	2.122	1.772	2.727	2.611	2.281	
	Spectral	0.608	0.582	0.489	0.638	0.600	0.523	0.668	0.612	0.545	
	Rel-Frobenius	0.380	0.615	0.790	0.126	0.306	0.523	0.024	0.161	0.383	
	Rel-spectral	0.235	0.421	0.649	0.060	0.166	0.352	0.008	0.075	0.223	
			0.349	0.380	0.645	0.511	0.361	0.618	0.591	0.411	0.523
LW-GLASSO (modified)	FNR	0.651	0.620	0.355	0.489	0.639	0.382	0.409	0.589	0.477	
	TNR	1.000	1.000	0.961	0.748	1.000	0.986	0.625	0.915	0.997	
	FPR	0.000	0.000	0.039	0.252	0.000	0.014	0.375	0.085	0.003	
	Precision	1.000	0.996	0.786	0.766	0.999	0.804	0.642	0.919	0.900	
	Frobenius	1.548	1.491	1.222	2.202	2.124	1.752	2.710	2.611	2.311	
	Spectral	0.609	0.584	0.501	0.638	0.601	0.525	0.654	0.612	0.556	
	Rel-Frobenius	0.380	0.615	0.801	0.125	0.306	0.520	0.022	0.161	0.388	
	Rel-spectral	0.236	0.423	0.669	0.059	0.166	0.354	0.007	0.075	0.229	
			0.373	0.353	0.350	0.384	0.379	0.370	0.388	0.391	0.377
			0.627	0.647	0.650	0.616	0.621	0.630	0.612	0.609	0.623
LW-CLIME	TNR	0.999	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	
	FPR	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.001	0.000	
	Precision	0.992	0.999	1.000	0.962	0.999	0.998	0.938	0.980	0.999	
	Frobenius	1.540	1.509	1.484	2.197	2.107	2.074	2.714	2.578	2.528	
	Spectral	0.607	0.592	0.574	0.636	0.600	0.580	0.661	0.608	0.584	
	Rel-Frobenius	0.379	0.622	0.979	0.120	0.303	0.614	0.019	0.158	0.425	
	Rel-spectral	0.235	0.428	0.771	0.056	0.166	0.391	0.006	0.073	0.241	

Table F.8: Simulation results for the modified graphical local Whittle and CLIME estimators with DGP3.

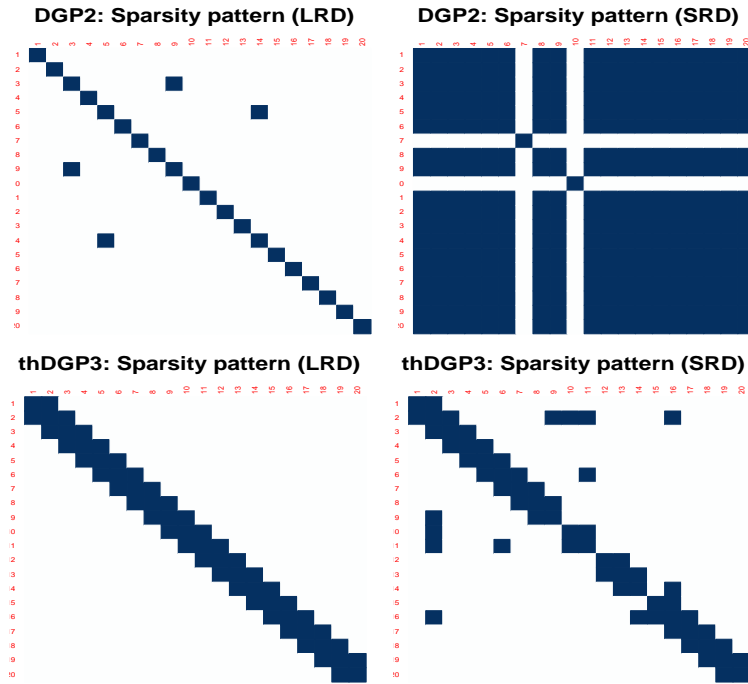


Figure 10: Sparsity pattern when SRD model is used for LRD simulated data.

### F.5 Comparison to existing methods

In this section we emphasize the relevance of considering estimators which account for strong temporal correlation beyond short-range dependence. Sun et al. (2018) consider possibly high-dimensional time series under short-range dependence ( $D_0 \equiv 0$ ) and study estimation of the spectral density and its inverse. See also Section 3.5 for a detailed comparison. We conduct a simulation study with synthetic, long-range dependent data and apply the estimators proposed in Sun et al. (2018). More precisely, we simulated long-range dependent data based on thDGP3 and DGP2 in Sections 5.1 and 5.2 following the sparsity pattern in Figure 1. We then applied the graphical LASSO and thresholded long-run variance estimators in Sun et al. (2018) to recover the sparsity patterns. As it can be seen from Figure 10, the estimators for short-range dependent models (right column) perform poorly and fail to find the true underlying zero coefficients. More detailed performance measures are provided in Table F.9 indicating that the estimators for short-range dependent models perform poorly under long-range dependence.



Models	Methods	Measures	$p = 20$			$p = 40$			$p = 60$		
			$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$	$N = 200$	$N = 400$	$N = 1000$
DGP2	LW-GLASSO (LRD)	TPR	0.969	0.998	1.000	0.980	0.998	1.000	0.989	0.996	1.000
		FNR	0.031	0.002	0.000	0.020	0.003	0.000	0.011	0.004	0.000
		TNR	1.000	0.999	0.998	0.654	1.000	0.999	0.325	0.940	1.000
		FPR	0.000	0.001	0.002	0.346	0.000	0.001	0.675	0.060	0.000
		Precision	0.997	0.990	0.980	0.661	0.991	0.985	0.316	0.936	0.990
	GLASSO (SRD)	TPR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		FNR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		TNR	0.117	0.284	0.559	0.006	0.108	0.348	0.001	0.013	0.259
		FPR	0.883	0.716	0.441	0.994	0.892	0.652	0.999	0.987	0.741
		Precision	0.068	0.085	0.138	0.030	0.034	0.046	0.020	0.020	0.027
thDGP3	Thresholding (LRD)	TPR	0.938	0.996	1.000	0.884	0.992	1.000	0.843	0.987	1.000
		FNR	0.062	0.004	0.000	0.116	0.008	0.000	0.157	0.013	0.000
		TNR	0.981	0.981	0.989	0.994	0.993	0.997	0.997	0.996	0.998
		FPR	0.019	0.019	0.011	0.006	0.007	0.003	0.003	0.004	0.002
		Precision	0.897	0.906	0.943	0.927	0.921	0.969	0.942	0.935	0.956
	Thresholding (SRD)	TPR	0.586	0.783	0.979	0.453	0.584	0.882	0.412	0.495	0.802
		FNR	0.414	0.217	0.021	0.547	0.416	0.118	0.588	0.505	0.198
		TNR	0.989	0.983	0.971	0.997	0.995	0.990	0.998	0.998	0.995
		FPR	0.011	0.017	0.029	0.003	0.005	0.010	0.002	0.002	0.005
		Precision	0.914	0.898	0.858	0.924	0.916	0.882	0.935	0.940	0.896

Table F.9: Performance measures when SRD models are used for LRD models.

## References

- Baek, C. and Pipiras, V. On distinguishing multiple changes in mean and long-range dependence using local Whittle estimation. *Electronic Journal of Statistics*, 8(1):931–964, 2014.
- Baek, C., Kechagias, S., and Pipiras, V. Asymptotics of bivariate local Whittle estimators with applications to fractal connectivity. *Journal of Statistical Planning and Inference*, 205:245–268, 2020.
- Baek, C., Kechagias, S., and Pipiras, V. Semiparametric, parametric, and possibly sparse models for multivariate long-range dependence. In *Wavelets and Sparsity XVII*, volume 10394, page 103941S. International Society for Optics and Photonics, 2017.
- Basu, S. and Michailidis, G. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567, 2015.
- Beran, J., Feng, Y., Ghosh, S., and Kulik, R. *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer-Verlag Berlin Heidelberg, 2013.
- Bickel, P. J. and Levina, E. Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227, 2008a.
- Bickel, P. J. and Levina, E. Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604, 2008b.
- Bien, J. and Tibshirani, R. J. Sparse estimation of a covariance matrix. *Biometrika*, 98(4):807–820, 2011.
- Bordier, C., Nicolini, C., and Bifone, A. Graph analysis and modularity of brain functional connectivity networks: searching for the optimal threshold. *Frontiers in Neuroscience*, 11:441, 2017.
- Böttcher, A. and Dörfler, P. Weighted Markov-type inequalities, norms of Volterra operators, and zeros of Bessel functions. *Mathematische Nachrichten*, 283(1):40–57, 2010.
- Böttcher, A. and Virtanen, J. Norms of Toeplitz matrices with Fisher–Hartwig symbols. *SIAM Journal on Matrix Analysis and Applications*, 29(2):660–671, 2007.
- Bourin, J.-C., Lee, E., and Lin, M. On a decomposition lemma for positive semi-definite block-matrices. *Linear Algebra and its Applications*, 437(7):1906–1912, 2012.
- Bowyer, S. M. Coherence a measure of the brain networks: past and present. *Neuropsychiatric Electrophysiology*, 2(1):1–12, 2016.
- Cai, T. T. and Liu, W. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684, 2011.
- Cai, T. T., Liu, W., and Luo, X. A constrained  $\ell_1$ -minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.
- Cai, T. T., Ren, Z., and Zhou, H. H. Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electronic Journal of Statistics*, 10(1):1–59, 2016.
- Cavicchioli, M. Goodness-of-fit tests for Markov switching VAR models using spectral analysis. *Journal of Statistical Planning and Inference*, 219:189–203, 2022.
- Chen, J. and Chen, Z. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3):759–771, 2008.
- Chen, J. and Chen, Z. Extended BIC for small- $n$ -large- $P$  sparse GLM. *Statistica Sinica*, 22(2):555–574, 2012.

- Dahlhaus, R. Graphical interaction models for multivariate time series. *Metrika*, 51(2):157–172, 2000.
- Dicker, L. H. and Erdogdu, M. A. Flexible results for quadratic forms with applications to variance components estimation. *The Annals of Statistics*, 45(1):386–414, 2017.
- Düker, M.-C. and Pipiras, V. Asymptotic results for multivariate local Whittle estimation with applications. In *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 584–588. IEEE, 2019.
- Erdős, L., Yau, H.-T., and Yin, J. Bulk universality for generalized Wigner matrices. *Probability Theory and Related Fields*, 154(1-2):341–407, 2012.
- Fan, J. and Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.
- Fan, J., Liao, Y., and Liu, H. An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal*, 19(1):C1–C32, 2016.
- Fang, F., Loparo, K. A., and Feng, X. Inequalities for the trace of matrix product. *IEEE Transactions on Automatic Control*, 39(12):2489–2490, 1994.
- Fiecas, M., Leng, C., Liu, W., and Yu, Y. Spectral analysis of high-dimensional time series. *Electronic Journal of Statistics*, 13(2):4079–4101, 2019.
- Fiecas, M. and Ombao, H. The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *The Annals of Applied Statistics*, 5(2A):1102–1125, 2011.
- Foygel, R. and Drton, M. Extended Bayesian information criteria for Gaussian graphical models. In *Advances in Neural Information Processing Systems*, pages 604–612, 2010.
- Foygel, R. and Drton, M. Bayesian model choice and information criteria in sparse generalized linear models. *arXiv preprint arXiv:1112.5635*, 2011.
- Friedman, J., Hastie, T., and Tibshirani, R. J. Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics*, 9(3):432–441, 2008.
- Gao, X., Pu, D. Q., Wu, Y., and Xu, H. Tuning parameter selection for penalized likelihood estimation of Gaussian graphical model. *Statistica Sinica*, 22(3):1123–1146, 2012.
- Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.
- Hansen, L. P. and Sargent, T. J. The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica: Journal of the Econometric Society*, pages 377–387, 1983.
- Henry, M. Robust automatic bandwidth for long memory. *Journal of Time Series Analysis*, 22(3):293–316, 2001.
- Jung, A., Hannak, G., and Goertz, N. Graphical lasso based model selection for time series. *IEEE Signal Processing Letters*, 22(10):1781–1785, 2015.
- Kechagias, S. and Pipiras, V. Definitions and representations of multivariate long-range dependent time series. *Journal of Time Series Analysis*, 36(1):1–25, 2015a.
- Kechagias, S. and Pipiras, V. Identification, estimation and applications of a bivariate long-range dependent times series model with general phase. *preprint*, pages 1–25, 2015b.
- Mercer, A. M. Some new inequalities involving elementary mean values. *Journal of Mathematical Analysis and Applications*, 229(2):677–681, 1999.

- Nielsen, F. S. Local Whittle estimation of multi-variate fractionally integrated processes. *Journal of Time Series Analysis*, 32(3):317–335, 2011.
- Pipiras, V. and Taqqu, M. *Long-Range Dependence and Self-Similarity*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017.
- Plagborg-Møller, M. and Wolf, C. K. Local projections and VARs estimate the same impulse responses. *Econometrica*, 89(2):955–980, 2021.
- Politis, D. N. Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory*, 27(4):703–744, 2011.
- Qi, F. and Luo, Q. Bounds for the ratio of two gamma functions—From Wendel’s and related inequalities to logarithmically completely monotonic functions. *Banach Journal of Mathematical Analysis*, 6(2):132–158, 2012.
- Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. High-dimensional covariance estimation by minimizing  $\ell_1$ -penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980, 2011.
- Robinson, P. M. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23(3):1048–1072, 1995a.
- Robinson, P. M. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23(5):1630–1661, 1995b.
- Robinson, P. M. Multiple local Whittle estimation in stationary systems. *The Annals of Statistics*, 36(5):2508–2530, 2008.
- Rothman, A. J., Bickel, P. J., Levina, E., and Zhu, J. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2:494–515, 2008.
- Rothman, A. J., Levina, E., and Zhu, J. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104(485):177–186, 2009.
- Rudelson, M. and Vershynin, R. Hanson-Wright inequality and sub-Gaussian concentration. *Electronic Communications in Probability*, 18, 2013.
- Scheinberg, K., Ma, S., and Goldfarb, D. Sparse inverse covariance selection via alternating linearization methods. In *Advances in Neural Information Processing Systems*, pages 2101–2109, 2010.
- Schneider-Luftman, D. and Walden, A. T. Partial coherence estimation via spectral matrix shrinkage under quadratic loss. *IEEE Transactions on Signal Processing*, 64(22):5767–5777, 2016.
- Shimotsu, K. Gaussian semiparametric estimation of multivariate fractionally integrated processes. *Journal of Econometrics*, 137(2):277–310, 2007.
- Shu, H. and Nan, B. Estimation of large covariance and precision matrices from temporally dependent observations. *The Annals of Statistics*, 47(3):1321–1350, 2019.
- Sun, Y., Li, Y., Kuceyeski, A., and Basu, S. Large spectral density matrix estimation by thresholding. *arXiv preprint arXiv:1812.00532*, 2018.
- Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.
- Wainwright, M. J. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.
- Wang, H., Banerjee, A., Hsieh, C.-J., Ravikumar, P. K., and Dhillon, I. S. Large scale distributed sparse precision estimation. *Advances in Neural Information Processing Systems*, 26, 2013.

Wang, M. and Allen, G. I. Thresholded graphical lasso adjusts for latent variables: Application to functional neural connectivity. *arXiv preprint arXiv:2104.06389*, 2021.

Zhou, S. Thresholded lasso for high dimensional variable selection and statistical estimation. *arXiv preprint arXiv:1002.1583*, 2010.

Zou, H. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101 (476):1418–1429, 2006.

Changryong Baek  
Dept. of Statistics  
Sungkyunkwan University  
25-2, Sungkyunkwan-ro, Jongno-gu  
Seoul, 110-745, Korea  
*crbaek@skku.edu*

Marie-Christine Düker  
Dept. of Statistics and Data Science  
Cornell University  
129 Garden Ave, Comstock Hall  
Ithaca, NY 14850, USA  
*duker@cornell.edu*

Vladas Pipiras  
Dept. of Statistics and Operations Research  
UNC at Chapel Hill  
CB#3260, Hanes Hall  
Chapel Hill, NC 27599, USA  
*pipiras@email.unc.edu*