Bernoulli **23**(4A), 2017, 2181–2209
DOI: 10.3150/16-BEJ807

Spectral analysis of sample autocovariance matrices of a class of linear time series in moderately high dimensions

LILI WANG[1], ALEXANDER AUE[2,*] and DEBASHIS PAUL[2,**]

Dedicated to the memory of Peter Gavin Hall

[1]*Department of Mathematics, Zhejiang University, Hangzhou 310027, China. E-mail: liliwang@zju.edu.cn*
[2]*Department of Statistics, University of California, Davis, CA 95616, USA.*
E-mail: [*]aaue@ucdavis.edu; [**]debpaul@ucdavis.edu

This article is concerned with the spectral behavior of p-dimensional linear processes in the moderately high-dimensional case when both dimensionality p and sample size n tend to infinity so that $p/n \to 0$. It is shown that, under an appropriate set of assumptions, the empirical spectral distributions of the renormalized and symmetrized sample autocovariance matrices converge almost surely to a nonrandom limit distribution supported on the real line. The key assumption is that the linear process is driven by a sequence of p-dimensional real or complex random vectors with i.i.d. entries possessing zero mean, unit variance and finite fourth moments, and that the $p \times p$ linear process coefficient matrices are Hermitian and simultaneously diagonalizable. Several relaxations of these assumptions are discussed. The results put forth in this paper can help facilitate inference on model parameters, model diagnostics and prediction of future values of the linear process.

Keywords: empirical spectral distribution; high-dimensional statistics; limiting spectral distribution; Stieltjes transform

1. Introduction

In this article, the spectral properties of a class of multivariate linear time series are studied through the bulk behavior of the eigenvalues of renormalized and symmetrized sample autocovariance matrices when both the dimension p and sample size n are large but the dimension increases at a much slower rate compared to the sample size, so that the dimension-to-sample size ratio p/n converges to zero. The latter asymptotic regime will be referred to as moderately high-dimensional scenario. Under this framework, the existence of limiting spectral distributions (LSD) of the matrices $\mathbf{C}_\tau = \sqrt{n/p}(\mathbf{S}_\tau - \boldsymbol{\Sigma}_\tau)$ is proved, where $\mathbf{S}_\tau$ is the symmetrized lag-τ sample autocovariance matrix and $\boldsymbol{\Sigma}_\tau$, the symmetrized lag-τ population autocovariance matrix, for $\tau = 0, 1, \ldots$. The analysis takes into account both temporal and dimensional correlation and the LSD is described in terms of a kernel that is determined by the transfer function of a univariate linear time series. The results derived in this paper are natural extensions of the work of [2], who proved that the empirical spectral distribution of renormalized sample covariance matrices based on i.i.d. observations with zero mean and unit variance converges to the semi-circle law under the same asymptotic regime. It also extends the work of [9] and [15] in two different ways; first, by allowing nontrivial temporal dependence among the observation vectors and, second, by describing the LSDs of renormalized sample autocovariance matrices of all lag orders. A certain structural assumption on the linear process is needed, namely, that its coefficient matrices are symmetric (Hermitian for complex-valued data) and simultaneously diagonalizable. However, various ways to relax the latter assumption are discussed.

The results derived in this paper can be seen as natural counterparts of the works of [8], who proved the existence of LSDs of symmetrized sample autocovariance matrices under the same structural assumptions on the linear process but in the asymptotic regime $p, n \to \infty$ such that $p/n \to c \in (0, \infty)$. Jin *et al.* [6] derived similar results under the assumption of i.i.d. observations, using them for detecting the presence of factors in a class of dynamic factor models. Under the same asymptotic framework, the existence of the LSD of sample covariance matrices when the different coordinates of the observed process are i.i.d. linear processes has been proved by [12] and [16]. In same setting, Pan, Gao and Yang [10] proved a central limit theorem for linear spectral statistics of sample covariance matrices. They also prove the existence of the LSD under finite second moment conditions. In a related development, Li, Pan and Yao [7] studied the distribution of singular values of sample autocovariance matrices in the same setting as [6]. Further results related to limiting spectral distributions of sample covariance matrices and their implications in the $p/n \to c \in (0, \infty)$ asymptotic domain can be found in [1] and [11].

The main results in this paper originally formed a part of the Ph.D. thesis of the first author [13]. Very recently, Bhattacharjee and Bose [3] proved the existence of the LSD of symmetrized and normalized autocovariance matrices for an MA(q) process with fixed q, under weaker assumptions on the coefficient matrices that involves existence of limits of averaged traces of polynomials of these matrices, where the limits satisfy certain regularity requirements associated with a non-commutative $*$-probability space. They use free probability theory for their derivations and therefore their approach is very different from the one presented in this paper, which relies on the characterization of distributional convergence through the convergence of the corresponding Stieltjes transforms. Moreover, the key structural assumption of this paper, namely, the simultaneous digaonalizability of coefficient matrices, has a natural interpretation in terms of the structure of the linear process, as discussed later. On the other hand, beyond this setting, the conditions assumed in [3] do not seem to have a similar finite-sample interpretation.

The main contribution of this paper is the precise description of the bulk behavior of the eigenvalues of the matrices $\mathbf{C}_\tau$. These are natural objects to study if one is interested in understanding the fluctuations of the sample autocovariance matrices from their population counterparts, since the latter provide useful information about the various characteristics of the observed process. Under the asmyptotic regime $p, n \to \infty$ with $p/n \to 0$, and under fairly weak regularity conditions, the symmetrized sample autocovariance matrices converge to the corresponding population autocovariance matrices in operator norm. However, stronger statements about the quality of the estimates are usually not possible without imposing further restrictions on the process. The results stated here provide a way to quantify the fluctuations of the estimated autocovariance matrices from the population versions, and can be seen as analogous to the standard error bounds in univariate problems. Indeed, if the quality of estimates is assessed through the Frobenius norm of $\mathbf{C}_\tau$, or some other measure that can be expressed as a linear functional of the spectral distribution of $\mathbf{C}_\tau$, the results presented in this paper give a precise description about the asymptotic behavior of such a measure in terms of integrals of the LSD of $\mathbf{C}_\tau$. Some specific applications of the results are discussed in Section 4. A further importance of the results derived here is that they form the building block of further investigations on the fluctuations of linear spectral statistics of matrices such as $\mathbf{C}_\tau$, thus raising the possibility of generalizing results such as those obtained recently by [5].

The rest of the paper is organized as follows. Section 2 gives the main results and develops intuition. Section 3 discusses some specific examples to elucidate the main results. Section 4 discusses a number of potential applications. Sections 5–7 are devoted to describing the key steps in the proofs of the main results. Due to space constraints, further technical details are relegated to the Online Supplement [14].

2. Main results

2.1. Assumptions

Let $\mathbb{Z}$, $\mathbb{N}_0$ and $\mathbb{N}$ denote integers, nonnegative integers and positive integers, respectively. In the following, the linear process $(X_t : t \in \mathbb{Z})$ is studied, given by the set of equations

$$X_t = \sum_{\ell=0}^{\infty} \mathbf{A}_\ell Z_{t-\ell}, \qquad t \in \mathbb{Z}, \tag{2.1}$$

where $(\mathbf{A}_\ell : \ell \in \mathbb{N}_0)$ are coefficient matrices with $\mathbf{A}_0 = \mathbf{I}_p$, the p-dimensional identity matrix, and $(Z_t : t \in \mathbb{Z})$ are innovations for which more specific assumptions are given below. If $\mathbf{A}_\ell = \mathbf{0}_p$, the p-dimensional zero matrix, for all $\ell > q$, then one has the qth order moving average, MA(q), process

$$X_t = \sum_{\ell=0}^{q} \mathbf{A}_\ell Z_{t-\ell}, \qquad t \in \mathbb{Z}. \tag{2.2}$$

In the following results will be stated and motivated first for the MA(q) process and then extended to linear processes. Throughout the following set of conditions are assumed to hold.

Assumption 2.1. *The innovations $(Z_t : t \in \mathbb{Z})$ consist of real- or complex-valued entries Z_{jt} which are independent, identically distributed (i.i.d.) across time t and dimension j and satisfy:*

(Z1) $\mathbb{E}[Z_{jt}] = 0$, $\mathbb{E}[|Z_{jt}|^2] = 1$ *and* $\mathbb{E}[|Z_{jt}|^4] < \infty$;

(Z2) *In case of complex-valued innovations, the real and imaginary parts of Z_{jt} are independent with* $\mathbb{E}[\Re(Z_{jt})] = \mathbb{E}[\Im(Z_{jt})] = 0$ *and* $\mathbb{E}[\Re(Z_{jt})^2] = \mathbb{E}[\Im(Z_{jt})^2] = 1/2$.

Assumption 2.2. *Suppose that:*

(A1) $(\mathbf{A}_\ell : \ell \in \mathbb{N})$ *are Hermitian and simultaneously diagonalizable, that is, there exists a unitary matrix* $\mathbf{U}$ *such that* $\mathbf{U}^* \mathbf{A}_\ell \mathbf{U} = \mathbf{\Lambda}_\ell$, *where* $\mathbf{\Lambda}_\ell$ *is a diagonal matrix with real-valued diagonal entries;*

(A2) *The* j*th diagonal entry of* $\mathbf{\Lambda}_\ell$ *is given by* $f_\ell(\boldsymbol{\alpha}_j)$, *where* $\boldsymbol{\alpha}_j \in \mathbb{R}^{m_0}$ *for* $j = 1, \dots, p$, *where* m_0 *is fixed, and* $(f_\ell : \ell \in \mathbb{N})$ *are continuous functions from* $\mathbb{R}^{m_0}$ *to* $\mathbb{R}$;

(A3) *As* $p \to \infty$, *the empirical distribution of* $(\boldsymbol{\alpha}_j : j = 1, \dots, p)$ *converges to a distribution on* $\mathbb{R}^{m_0}$ *denoted by* $F^{\mathcal{A}}$;

(A4) *Let* $\bar{a}_0 = 1$ *and* $\bar{a}_\ell = \|f_\ell\|_\infty$ *for* $\ell \geq 1$. *Assume that, for some* $r_0 \geq 4$, $\sum_{\ell=0}^{\infty} \ell^{r_0} \bar{a}_\ell < \infty$. *In particular, with* $L_{j+1} = \sum_{\ell=0}^{\infty} \ell^j \|f_\ell\|_\infty$, *it holds that* $L_{j+1} < \infty$ *for* $j = 0, \dots, r_0$.

The assumptions on the innovations $(Z_t : t \in \mathbb{Z})$ are standard in time series and high-dimensional statistics contexts. The assumptions on the coefficient matrices $(A_\ell : \ell \in \mathbb{N})$ are similar to the ones imposed in [8] and generalize condition sets previously established in the literature, for example, the ones in [12,16] and [6]. In assumption (A2), the precise value of m_0 is dependent on the description of the process. In Section 3, several examples of processes satisfying (A1)–(A4) are provided that help clarify the description of $(f_\ell : \ell \in \mathbb{N})$ and the value of m_0 in each case. Regarding assumption (A4), the conditions on L_{j+1} for $j > 1$ are needed for the extension of the results for MA(q) processes to linear processes (see Section 2.4).

2.2. Result for MA(q) processes

The objective of this section is to study the spectral behavior of the lag-τ symmetrized sample autocovariance matrices associated with the MA(q) process $(X_t : t \in \mathbb{Z})$ defined in (2.2) in the moderately high dimensional setting

$$p, n \to \infty \qquad \text{such that } \frac{p}{n} \to 0. \tag{2.3}$$

Extensions to the linear process (2.1) are discussed in Section 2.4 below. The symmetrized sample autocovariance matrices are given by the equations

$$\mathbf{S}_\tau = \frac{1}{2(n-\tau)} \sum_{t=\tau+1}^{n} \left(X_t X_{t-\tau}^* + X_{t-\tau} X_t^* \right), \qquad \tau \in \mathbb{N}_0, \tag{2.4}$$

where $*$ signifies complex conjugate transposition of both vectors and matrices. It should be noted that $\mathbf{S}_0$ is simply the sample covariance matrix. Using the defining equations of the MA(q) process, one can show that

$$\mathbf{\Sigma}_\tau = \mathbb{E}[\mathbf{S}_\tau] = \frac{1}{2} \left(\sum_{\ell=0}^{q-\tau} \left[\mathbf{A}_{\ell+\tau} \mathbf{A}_\ell^* + \mathbf{A}_\ell \mathbf{A}_{\ell+\tau}^* \right] \right), \qquad \tau \in \mathbb{N}_0.$$

Since, under (2.3), $\mathbf{S}_\tau$ is a consistent estimator for $\mathbf{\Sigma}_\tau$, one studies appropriately rescaled fluctuations of $\mathbf{S}_\tau$ about its mean $\mathbf{\Sigma}_\tau$. This leads to the renormalized matrices

$$\mathbf{C}_\tau = \sqrt{\frac{n}{p}} (\mathbf{S}_\tau - \mathbf{\Sigma}_\tau), \qquad \tau \in \mathbb{N}_0. \tag{2.5}$$

To study the spectral behavior of $\mathbf{C}_\tau$, introduce its empirical spectral distribution (ESD) $\hat{F}_\tau$ given by

$$\hat{F}_\tau(\lambda) = \frac{1}{p}\sum_{j=1}^{p} \mathbb{I}_{\{\lambda_{\tau,j} \le \lambda\}},$$

where $\lambda_{\tau,1}, \dots, \lambda_{\tau,p}$ are the eigenvalues of $\mathbf{C}_\tau$. In the RMT literature, proofs of large-sample results about $\hat{F}_\tau$ are often based on convergence properties of Stieltjes transforms [1]. The Stieltjes transform of a distribution function F on the real line is the function

$$s_F : \mathbb{C}^+ \to \mathbb{C}^+, \qquad z \mapsto s_F(z) = \int \frac{1}{\lambda - z}\, dF(\lambda),$$

where $\mathbb{C}^+ = \{x + \mathbf{i}y : x \in \mathbb{R}, y > 0\}$ denotes the upper complex half plane. Note that s_F is analytic on $\mathbb{C}^+$ and that the distribution function F can be obtained from s_F using an inversion formula.

Let $f_0 : \mathbb{R}^{m_0} \to \mathbb{R}$ be defined as $f_0(\mathbf{a}) = 1$ for all $\mathbf{a} \in \mathbb{R}^{m_0}$. Define the MA($q$) transfer function

$$g(\mathbf{a}, \nu) = \sum_{\ell=0}^{q} f_\ell(\mathbf{a}) e^{\mathbf{i}\ell\nu}, \qquad \nu \in [0, 2\pi], \mathbf{a} \in \mathbb{R}^{m_0}, \tag{2.6}$$

and the corresponding power transfer function

$$\psi(\mathbf{a}, \nu) = \left|g(\mathbf{a}, \nu)\right|^2, \qquad \nu \in [0, 2\pi], \mathbf{a} \in \mathbb{R}^{m_0}. \tag{2.7}$$

The effect of the temporal dependence on the spectral behavior of $\mathbf{C}_\tau$ is encoded through the power transfer function $\psi(\mathbf{a}, \nu)$. Keeping $\mathbf{a}$ fixed, it can be seen that $\psi(\mathbf{a}, \nu)$ is up to normalization the spectral density of a univariate MA(q) process with coefficients $f_1(\mathbf{a}), \dots, f_q(\mathbf{a})$. This leads to the following result.

Theorem 2.1. *If the* MA(q) *process* $(X_t : t \in \mathbb{Z})$ *satisfies* (Z1), (Z2) *and* (A1)–(A4), *then*, *with probability one and in the moderately high-dimensional setting* (2.3), $\hat{F}_\tau$ *converges in distribution to a nonrandom distribution* F_τ *whose Stieltjes transform* s_τ *is given by*

$$s_\tau(z) = -\int \frac{dF^{\mathcal{A}}(\mathbf{b})}{z + \beta_\tau(z, \mathbf{b})}, \qquad z \in \mathbb{C}^+, \tag{2.8}$$

where

$$\beta_\tau(z, \mathbf{a}) = -\int \frac{\mathcal{R}_\tau(\mathbf{a}, \mathbf{b})\, dF^{\mathcal{A}}(\mathbf{b})}{z + \beta_\tau(z, \mathbf{b})}, \qquad z \in \mathbb{C}^+, \mathbf{a} \in \mathbb{R}^{m_0}, \tag{2.9}$$

and

$$\mathcal{R}_\tau(\mathbf{a}, \mathbf{b}) = \frac{1}{2\pi}\int_0^{2\pi} \cos^2(\tau\theta)\psi(\mathbf{a}, \theta)\psi(\mathbf{b}, \theta)\, d\theta, \qquad \mathbf{a}, \mathbf{b} \in \mathbb{R}^{m_0}. \tag{2.10}$$

Moreover, $\beta_\tau(z, \mathbf{a})$ *is the unique solution to* (2.9) *subject to the condition that it is a Stieltjes kernel*, *that is*, *for each* $\mathbf{a} \in \operatorname{supp}(F^{\mathcal{A}})$, $\beta_\tau(z, \mathbf{a})$ *is the Stieltjes transform of a measure on the real line with mass* $\int \mathcal{R}_\tau(\mathbf{a}, \mathbf{b})\, dF^{\mathcal{A}}(\mathbf{b})$.

Since it only differs from the spectral density of an MA(q) process by a constant, it follows that $\psi(\mathbf{a},\theta)$ is strictly positive for all arguments. Consequently, $\mathcal{R}_\tau(\mathbf{a},\mathbf{b})$ and $\int \mathcal{R}_\tau(\mathbf{a},\mathbf{b})\,dF^{\mathcal{A}}(\mathbf{b})$ are always strictly positive as well. The intuition for the proof of Theorem 2.1 is given in the next section and will then be completed in Section 5.

Remark 2.1. It is easily checked, that for an MA(q) process, the kernel $\mathcal{R}_\tau(\mathbf{a},\mathbf{b})$ is the same for all $\tau \geq q+1$. This implies that the Stieltjes transforms $s_\tau(z)$, and hence the LSDs (limiting spectral distributions) of $\sqrt{n/p}(\mathbf{S}_\tau - \Sigma_\tau)$ are the same for $\tau \geq q+1$.

2.3. Intuition for Gaussian MA(q) processes

Assume for now that the innovations $(Z_t\colon t\in\mathbb{Z})$ are complex Gaussian, the extension to general innovations will be established in Section 7. Define the $p\times n$ data matrix $\mathbf{X}=[X_1:\cdots:X_n]$ and the $p\times n$ innovations matrix $\mathbf{Z}=[Z_1:\cdots:Z_n]$. Using the $n\times n$ lag operator matrix $\mathbf{L}=[o:e_1:\cdots:e_{n-1}]$, where o and e_j denote the zero vector and the jth canonical unit vector, respectively, it follows that

$$\mathbf{X}=\sum_{\ell=0}^{q}\mathbf{A}_\ell\mathbf{Z}\mathbf{L}^\ell+\sum_{\ell=1}^{q}\mathbf{A}_\ell\mathbf{Z}_{[-q]}\mathbf{L}^{\ell-q}, \tag{2.11}$$

where $\mathbf{Z}_{[-q]}=[Z_{-q+1}:\cdots:Z_0:0:\cdots:0]$ is a $p\times n$ matrix and $\mathbf{L}^{\ell-q}=(\mathbf{L}^{q-\ell})^{-1}$. In the next step, $\mathbf{L}$ is approximated by the circulant matrix $\tilde{\mathbf{L}}=[e_n:e_1:\cdots:e_{n-1}]$. As in [8], one defines the matrix $\bar{\mathbf{X}}=\sum_{\ell=0}^{q}\mathbf{A}_\ell\mathbf{Z}\tilde{\mathbf{L}}^\ell$ that differs from $\mathbf{X}$ only in the first q columns. Let $\mathbf{F}_n=[e^{\mathbf{i}s\nu_t}]_{s,t=1}^n$, with $\nu_t=2\pi t/n$, be a Fourier rotation matrix and $\tilde{\mathbf{\Lambda}}_n=\operatorname{diag}(e^{\mathbf{i}\nu_1},\dots,e^{\mathbf{i}\nu_n})$. Then

$$\tilde{\mathbf{L}}=\mathbf{F}_n\tilde{\mathbf{\Lambda}}_n\mathbf{F}_n^*. \tag{2.12}$$

Using this and noticing that $\mathbf{X}$ and $\bar{\mathbf{X}}$ differ by a matrix of rank q, it can be seen that as long as q small compared to p, $\mathbf{S}_\tau=(n-\tau)^{-1}\mathbf{X}\mathbf{D}_\tau\mathbf{X}^*$ can be approximated by $\bar{\mathbf{S}}_\tau=(n-\tau)^{-1}\bar{\mathbf{X}}\bar{\mathbf{D}}_\tau\bar{\mathbf{X}}^*$, where $\mathbf{D}_\tau=[\mathbf{L}^\tau+(\mathbf{L}^\tau)^*]/2$ and $\bar{\mathbf{D}}_\tau=[\tilde{\mathbf{L}}^\tau+(\tilde{\mathbf{L}}^\tau)^*]/2$. Notice next that, due to the assumed Gaussianity of the innovations, the entries of $\tilde{\mathbf{Z}}=\mathbf{U}^*\mathbf{Z}\mathbf{F}_n$ are i.i.d. copies of the entries of $\mathbf{Z}$, with $\mathbf{U}$ denoting the matrix diagonalizing the coefficient matrices $(\mathbf{A}_\ell\colon\ell\in\mathbb{N})$. Define then $\tilde{\mathbf{S}}_\tau=\mathbf{U}^*\bar{\mathbf{S}}_\tau\mathbf{U}$ and

$$\tilde{\mathbf{C}}_\tau=\sqrt{\frac{n}{p}}(\tilde{\mathbf{S}}_\tau-\tilde{\mathbf{\Sigma}}_\tau), \tag{2.13}$$

where $\tilde{\mathbf{\Sigma}}_\tau=\mathbb{E}[\tilde{\mathbf{S}}_\tau]$ is a diagonal matrix. It will be shown in Section 5.1 that the LSD of $\mathbf{C}_\tau^U=\mathbf{U}^*\mathbf{C}_\tau\mathbf{U}$ is the same as that of $\tilde{\mathbf{C}}_\tau$.

2.4. Extensions to linear processes

In this section, Theorem 2.1 is extended to cover linear processes as defined in (2.1). To do so, the continuity condition (A2) is strengthened to assumption (A6) below. In order to approximate the linear process with MA(q) models of increasing order, a rate on q is imposed.

Assumption 2.3. *The following conditions are assumed to hold.*

(A5) $(f_\ell : \ell \in \mathbb{N})$ *are Lipschitz functions such that* $|f_\ell(\mathbf{a}) - f_\ell(\mathbf{b})| \le C\ell^{r_1}\|\mathbf{a}-\mathbf{b}\|$ *for* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{m_0}$ *and* $\ell \in \mathbb{N}$, *where* $r_1 \le r_0$ *and* r_0 *is as in* (A4).

Analogously (2.6) and (2.7) are extended to the linear process transfer function and power transfer function

$$g(\mathbf{a}, \nu) = \sum_{\ell=0}^{\infty} f_\ell(\mathbf{a}) e^{\mathbf{i}\ell\nu} \quad \text{and} \quad \psi(\mathbf{a}, \nu) = \left|g(\mathbf{a}, \nu)\right|^2, \qquad \nu \in [0, 2\pi], \mathbf{a} \in \mathbb{R}^{m_0}, \tag{2.14}$$

respectively. Then, the following result holds.

Theorem 2.2. *If the linear process* $(X_t : t \in \mathbb{Z})$ *satisfies* (Z1), (Z2) *and* (A1)–(A5), *then*, *the result of Theorem* 2.1 *is retained if* (2.14) *is used in place of* (2.6) *and* (2.7).

The proof of Theorem 2.2 is based on a truncation argument, that involves approximating the linear process with MA(q) processes of increasing order q. More delicate arguments are needed for this case as the intuitive arguments outlined in the previous section do not carry over to this case. Indeed conditions on the approximating MA(q) ensure that q does not grow too fast or too slow in order for the LSD of the linear process and its truncated version to be the same. The proof details are given in Section 6 below, where it turns out that one can choose $q = O(p^{1/4})$. Consequently, as a byproduct of the proof of Theorem 2.2, an extension of Theorem 2.1 to the setting where the order q of the MA(q) process grows at the rate $O(p^{1/4})$ is obtained.

A slight generalization of the above result can be given if the innovation terms of the linear process are not isotropic, but have a certain covariance structure that commutes with the coefficient matrices of the linear process. This is stated as a corollary to Theorem 2.2 through the following formulation. Let the process $(Y_t : t \in \mathbb{Z})$ be obtained from the linear process $(X_t : t \in \mathbb{Z})$ through

$$Y_t = \mathbf{B}^{1/2} X_t, \qquad t \in \mathbb{Z}, \tag{2.15}$$

where it is assumed that:

(A6) $\mathbf{B}^{1/2}$ is a square root of the nonnegative definite Hermitian matrix $\mathbf{B}$ with $\|\mathbf{B}\| \le \bar{b}_0 < \infty$, and there is a nonnegative measurable function g_B, not identically zero on supp($F^{\mathcal{A}}$), such that for each p, $\mathbf{U}^*\mathbf{B}\mathbf{U} = \operatorname{diag}(g_B(\boldsymbol{\alpha}_1), \ldots, g_B(\boldsymbol{\alpha}_p)) = \mathbf{\Lambda}_B$, with $\mathbf{U}$ and $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p$ as defined in (A1) and (A2).

Then the following result holds.

Corollary 2.1. *If the process* $(Y_t : t \in \mathbb{Z})$ *defined in* (2.15) *satisfies* (Z1), (Z2) *and* (A1)–(A6), *then*, *with probability one and in the setting* (2.3), $\hat{F}_\tau^Y$ *converges in distribution to a nonrandom distribution* F_τ^Y *whose Stieltjes transform* s_τ^Y *is given by*

$$s_\tau^Y(z) = -\int \frac{dF^{\mathcal{A}}(\mathbf{a})}{z + \beta_\tau^Y(z, \mathbf{a})}, \qquad z \in \mathbb{C}^+, \tag{2.16}$$

where

$$\beta_\tau^Y(z,\mathbf{a}) = -\int \frac{g_B(\mathbf{a})g_B(\mathbf{b})\mathcal{R}_\tau(\mathbf{a},\mathbf{b})\,dF^{\mathcal{A}}(\mathbf{b})}{z+\beta_\tau^Y(z,\mathbf{b})}, \qquad z\in\mathbb{C}^+, \mathbf{a}\in\mathbb{R}^{m_0}, \tag{2.17}$$

and $\mathcal{R}_\tau(\mathbf{a},\mathbf{b})$ *is defined in* (2.10). *Moreover,* $\beta_\tau^Y(z,\mathbf{a})$ *is the unique solution to* (2.17) *subject to the condition that it is a Stieltjes kernel, that is, for each* $\mathbf{a}\in\operatorname{supp}(F^{\mathcal{A}})$, $\beta_\tau^Y(z,\mathbf{a})$ *is the Stieltjes transform of a measure on the real line with mass* $g_B(\mathbf{a})\int g_B(\mathbf{b})\mathcal{R}_\tau(\mathbf{a},\mathbf{b})\,dF^{\mathcal{A}}(\mathbf{b})$ *whenever* $g_B(\mathbf{a})>0$.

The proof of Corollary 2.1 is similar to that of Theorem 2.4. The key structural modification is as follows. Observe that the autocovariance matrices of the process $(Y_t\colon t\in\mathbb{Z})$ are given by $\mathbf{S}_\tau^Y=\mathbf{B}^{1/2}\mathbf{S}_\tau\mathbf{B}^{1/2}$ and have expectation $\mathbf{\Sigma}_\tau^Y=\mathbf{B}^{1/2}\mathbf{\Sigma}_\tau\mathbf{B}^{1/2}$. Assumption (A6) implies that the approximating autocovariance matrix obtained from replacing the lag operator matrix with the corresponding circulant matrix takes on the form

$$\tilde{\mathbf{S}}_\tau^Y = \frac{1}{n-\tau}\left(\sum_{\ell=0}^\infty \sqrt{\mathbf{\Lambda}_B}\mathbf{\Lambda}_\ell\tilde{\mathbf{Z}}\tilde{\mathbf{\Lambda}}_n^\ell\right)\left(\frac{\tilde{\mathbf{\Lambda}}_n^\tau+(\tilde{\mathbf{\Lambda}}_n^\tau)^*}{2}\right)\left(\sum_{\ell=0}^\infty \sqrt{\mathbf{\Lambda}_B}\mathbf{\Lambda}_\ell\tilde{\mathbf{Z}}\tilde{\mathbf{\Lambda}}_n^\ell\right)^* \tag{2.18}$$

with expectation $\tilde{\mathbf{\Sigma}}_\tau^Y=\operatorname{diag}(\tilde{\sigma}_{\tau,1}^Y,\dots,\tilde{\sigma}_{\tau,p}^Y)$ and

$$\tilde{\sigma}_{\tau,j}^Y=\frac{1}{n-\tau}\sum_{t=1}^n g_B(\boldsymbol{\alpha}_j)\cos(\tau\nu_t)\psi(\boldsymbol{\alpha}_j,\nu_t),$$

in which $\psi(\boldsymbol{\alpha}_j,\nu_t)$ is defined in (2.14). Following similar arguments as in the finite and infinite order MA cases, it can be shown that the LSD of $\mathbf{C}_\tau^Y=\sqrt{n/p}(\mathbf{S}_\tau^Y-\mathbf{\Sigma}_\tau^Y)$ is the same as that of $\tilde{\mathbf{C}}_\tau^Y=\sqrt{n/p}(\tilde{\mathbf{S}}_\tau^Y-\tilde{\mathbf{\Sigma}}_\tau^Y)$. The rest of the proof is essentially the same as that of Theorem 2.2.

2.5. Relaxation of commutativity condition

The assumption of commutativity or simultaneous diagonalizability of the coefficients (assumption (A1)) indeed restricts the class of linear processes for which the main result of existence and uniqueness of the limiting ESD applies. However, this assumption can be relaxed to a milder one in which the coefficients of the linear processes are only approximately Hermitian and commutative. Two such scenarios are discussed below, which are natural but by no means exhaustive. In both settings, it is assumed that the linear process

$$X_t=\sum_{\ell=0}^\infty \mathbf{B}_\ell Z_{t-\ell}, \qquad t\in\mathbb{Z}, \tag{2.19}$$

is observed with the standard assumptions (Z1) and (Z2) on the sequence $(Z_t\colon t\in\mathbb{Z})$, whereas $\mathbf{B}_0=\mathbf{I}_p$ and the sequence $(\mathbf{B}_\ell\colon \ell\in\mathbb{N})$ satisfies the conditions:

(B1) Let $\bar{b}_0 = 1$, $\bar{b}_\ell = \|\mathbf{B}_\ell\|$ for $\ell \in \mathbb{N}$, and $L'_{j+1} = \sum_{\ell=0}^{\infty} \ell^j \bar{b}_\ell$. There exists an integer $r_0 \geq 1$ such that for $j = 0, \ldots, r_0$ $L'_{j+1} < \infty$;

(B2) There is a sequence of Hermitian matrices $(\mathbf{A}_\ell : \ell \in \mathbb{N})$ approximating the sequence $(\mathbf{B}_\ell : \ell \in \mathbb{N})$ and satisfying (A1)–(A5).

In addition to (B1) and (B2), it is assumed that the sequence $(\mathbf{A}_\ell : \ell \in \mathbb{N})$ satisfies one of the following conditions specifying the approximation property in (B2):

(B3) For some $1 \leq \beta < 4$, $p^{-1} \sum_{\ell=1}^{\lceil p^{1/\beta} \rceil} \operatorname{rank}(\mathbf{B}_\ell - \mathbf{A}_\ell) \to 0$ under (2.3);

(B4) For some $1 \leq \beta < 4$, $\sqrt{n/p} \sum_{\ell=1}^{\lceil p^{1/\beta} \rceil} \|\mathbf{B}_\ell - \mathbf{A}_\ell\| \to 0$ under (2.3).

The importance of these conditions is discussed. First, restricting the sums involving $\mathbf{B}_\ell - \mathbf{A}_\ell$ to the first $p^{1/\beta}$ terms is sufficient in view of (B1) ensuring that the process $(X_t : t \in \mathbb{Z})$ can be approximated by the truncated process given by $X_t^q = \sum_{\ell=0}^{q} \mathbf{B}_\ell Z_{t-\ell}$ with $q = O(p^{1/4})$ without changing the LSD of $\sqrt{n/p}(\mathbf{S}_\tau - \mathbb{E}[\mathbf{S}_\tau])$. This can be verified by following the derivation in Section 2.4. Condition (B3), on the other hand, says that the coefficient matrices $(\mathbf{B}_\ell : \ell \in \mathbb{N})$ can be seen as low-rank perturbations of a sequence of Hermitian and commutative matrices $(\mathbf{A}_\ell : \ell \in \mathbb{N})$. Condition (B4), which bounds the norms of differences between the coefficients and their approximations, is a bit restrictive in the sense that it depends on n. Presence of the factor $\sqrt{n/p}$ suggests that this condition is non-trivial essentially if n is moderately large compared to p.

The result is stated in the form of the following corollary.

Corollary 2.2. *Suppose that the linear process $(X_t : t \in \mathbb{Z})$ satisfies conditions* (B1), (B2) *and either* (B3) *or* (B4), *and let $\mathbf{S}_\tau$ denote the lag-τ symmetrized sample autocovariance matrix. Then the limiting ESD of the matrix $\sqrt{n/p}(\mathbf{S}_\tau - \mathbb{E}[\mathbf{S}_\tau])$ exists and its Stieltjes transform $s_\tau(z)$ satisfies* (2.8)–(2.10).

The proof of Corollary 2.2 is given in Section S.2 of the Online Supplement [14].

The conditions imposed in Corollary 2.2 can be used to prove that results hold for processes $(X_t : t \in \mathbb{Z})$ satisfying (2.19) and whose coefficient matrices are certain classes of symmetric (Hermitian) Toeplitz matrices. Specifically, if the matrix $\mathbf{B}_\ell$ is determined by the sequence $(b_{\ell k} : k \in \mathbb{Z})$, satisfying the condition $\sup_{\ell \geq 1} \sum_{|k| \geq m} |k|^s |b_{\ell k}| \to 0$ as $m \to \infty$ for some $s \geq 1$, and (B1) holds, then the LSDs of the corresponding normalized sample autocovariance matrices exist under (2.3) provided $n = O(p^{s+1/2})$. In this case, the symmetric (Hermitian) Toeplitz matrices $\mathbf{B}_\ell$ can be approximated by symmetric (Hermitian) circulant matrices whose eigenvalues are precisely the symbols associated with the sequence $(b_{\ell k} : k \in \mathbb{Z})$ evaluated at the discrete Fourier frequencies $2\pi j/p$, $j = 1, \ldots, p$.

3. Examples

In this section, a number of special cases are presented for which the results stated in Section 2 take on an easier form.

Example 3.1. Consider the MA(1) process

$$X_t = Z_t + \mathbf{A} Z_{t-1},$$

with $\mathbf{A} = \operatorname{diag}(\alpha_1, \ldots, \alpha_p)$ for $\alpha_j \in \mathbb{R}$. Thus, in this example, $m_0 = 1$. Suppose further that $f_1(a) = a$. Then, the transfer function (2.6) is given by $g(a,\theta) = 1 + a e^{\mathbf{i}\theta}$ and the power transfer function (2.7) by $\psi(a,\theta) = 1 + a^2 + 2a\cos(\theta)$. This yields the explicit expressions

$$\mathcal{R}_\tau(a,b) = \begin{cases} (1+a^2)(1+b^2) + 2ab, & \tau = 0, \\ (1+a^2)(1+b^2)/2 + 3ab/2, & \tau = 1, \\ (1+a^2)(1+b^2)/2 + ab, & \tau \geq 2. \end{cases}$$

Example 3.2. Consider the special case of an MA(q) process with $\mathbf{A}_\ell = \gamma_\ell \mathbf{I}_p$, $\ell = 1, \ldots, q$. Set here $m_0 = q$, $f_\ell \equiv \gamma_\ell$ and $\boldsymbol{\alpha}_j = \mathbf{1}_q$ for all $j = 1, \ldots, p$. Then $F^{\mathcal{A}}$ is the distribution on $\mathbb{R}^q$ with probability 1 at $\mathbf{1}_q$. Since

$$g(\mathbf{a},\nu) = \sum_{\ell=0}^{\infty} f_\ell(\mathbf{a}) e^{\mathbf{i}\ell\nu} = \sum_{\ell=0}^{\infty} \gamma_\ell e^{\mathbf{i}\ell\nu} = \tilde{g}(\nu)$$

and therefore $\psi(\mathbf{a},\nu) = \tilde{\psi}(\nu)$ does not depend on $\mathbf{a}$, it follows that

$$\mathcal{R}_\tau(\mathbf{a},\mathbf{1}) = \frac{1}{2\pi}\int_0^{2\pi} \cos^2(\tau\nu)\big(\tilde{\psi}(\nu)\big)^2\,d\nu = \bar{\mathcal{R}}_\tau,$$

so that equations (2.9) and (2.8) reduce respectively to $\beta_\tau(z,\mathbf{a}) = \beta_\tau(z) = \bar{\mathcal{R}}_\tau s_\tau(z)$ and

$$s_\tau(z) = -\frac{1}{z + \bar{\mathcal{R}}_\tau s_\tau(z)}.$$

For $\tau = 0$, the latter equation coincides with that for the Stieltjes transform for the case of independent observations with separable covariance structure discussed in [15]. Indeed, taking in their notation $\mathbf{A}_p = \mathbf{I}_p$ and $\mathbf{B}_n^{1/2} = \operatorname{diag}(\tilde{g}(\nu_1), \ldots, \tilde{g}(\nu_n))$, equation (2.1) of Theorem 2.1 in [15] reduces to $s(z) = -[z + \bar{b}_2 s(z)]^{-1}$, where

$$\bar{b}_2 = \lim_{n\to\infty} \frac{1}{n}\operatorname{Tr}\big(\mathbf{B}_n^2\big) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\big|\tilde{g}(\nu_t)\big|^4 = \frac{1}{2\pi}\int_0^{2\pi}\tilde{\psi}(\nu)^2\,d\nu = \bar{\mathcal{R}}_0.$$

Example 3.3. Consider the AR(1) process

$$X_t = \mathbf{A} X_{t-1} + Z_t,$$

with $\mathbf{A} = \operatorname{diag}(\alpha_1, \ldots, \alpha_p)$ for $\alpha_j \in \mathbb{R}$ such that $|\alpha_j| < 1$. The AR(1) process then admits the linear process representation $X_t = \sum_{\ell=0}^{\infty} \mathbf{A}^\ell Z_{t-\ell}$. Thus, here $m_0 = 1$ and $f_\ell(a) = a^\ell$ for $\ell = 0, 1, \ldots,$ and $a \in \mathbb{R}$. The transfer function (2.6) is given by $g(a,\theta) = (1 - a e^{\mathbf{i}\theta})^{-1}$ and the power transfer function (2.7) by $\psi(a,\theta) = (1 + a^2 - 2a\cos\theta)^{-1}$.

Example 3.4. Consider the causal ARMA(1, 1) process

$$\mathbf{\Phi}(L)X_t = \mathbf{\Theta}(L)Z_t,$$

where $\mathbf{\Phi}(L) = \mathbf{I} - \mathbf{\Phi}_1$ and $\mathbf{\Theta}(L) = \mathbf{I} + \mathbf{\Theta}_1$ are matrix-valued autoregressive and moving average polynomials in the lag operator L such that $\|\Phi_1\| < 1$ and $\|\Theta_1\| < \infty$. Then $(X_t : t \in \mathbb{Z})$ can be represented as the linear process

$$X_t = \mathbf{A}(L)Z_t,$$

in which $\mathbf{A}(L) = \sum_{\ell=0}^{\infty} \mathbf{A}_\ell L^\ell = (\mathbf{I} - \mathbf{\Phi}_1 L)^{-1}(\mathbf{I} + \mathbf{\Theta}_1 L)$. Assume further that $\mathbf{\Phi}_1$ and $\mathbf{\Theta}_1$ are simultaneously diagonalizable by $\mathbf{U}$, that is, $\mathbf{U}^*\mathbf{\Phi}_1\mathbf{U} = \mathrm{diag}(\phi_1, \dots, \phi_p)$ and $\mathbf{U}^*\mathbf{\Theta}_1\mathbf{U} = \mathrm{diag}(\theta_1, \dots, \theta_p)$. Let $\boldsymbol{\alpha}_j = (\phi_j, \theta_j)^T \in \mathbb{R}^2$. Thus, here $m_0 = 2$. Assumption (A3) is satisfied if the empirical distribution of $\{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p\}$ converges weakly to a non-random distribution function defined on $\mathbb{R}^2$. Note that

$$\mathbf{U}^*\mathbf{A}_\ell\mathbf{U} = \mathrm{diag}\big(f_\ell(\boldsymbol{\alpha}_1), \dots, f_\ell(\boldsymbol{\alpha}_p)\big),$$

with $f_\ell(\boldsymbol{\alpha}_j) = 1$ and $f_\ell(\boldsymbol{\alpha}_j) = (\theta_j + \phi_j)\phi_j^{\ell-1}$ for $\ell \in \mathbb{N}$. Thus, the transfer function (2.6) is given by

$$g(\boldsymbol{\alpha}_j, \nu) = \sum_{\ell=0}^{\infty} f_\ell(\boldsymbol{\alpha}_j)e^{\mathbf{i}\ell\nu} = 1 + \sum_{\ell=1}^{\infty}(\theta_j + \phi_j)\phi_j^{\ell-1}e^{\mathbf{i}\ell\nu} = \frac{1 + \theta_j e^{\mathbf{i}\nu}}{1 - \phi_j e^{\mathbf{i}\nu}}$$

and the power transfer function (2.7) is the squared modulus of the ratio on right-hand side of the last equation.

Example 3.5. Suppose that for each $\ell \geq 1$, $\mathbf{A}_\ell$ is a block diagonal matrix with B (a fixed number) diagonal blocks such that the bth block of $\mathbf{A}_\ell$ is of the form $a_{\ell b}\mathbf{I}_{p_b}$, for $b = 1, \dots, B$, where $\sum_{b=1}^{B} p_b = p$, and $\sum_{\ell=1}^{\infty} \ell^4 \max_{1\leq b\leq B} |a_{\ell b}| < \infty$. Suppose further that for each b, $p_b/p \to \omega_b$ as $p \to \infty$, where $\omega_b > 0$ for all b. In this example, one can take $\boldsymbol{\alpha}_j = b/(m+1)$ if $\sum_{b'=1}^{b-1} p_{b'} + 1 \leq j \leq \sum_{b'=1}^{b} p_{b'}$, set $m_0 = 1$, and define f_ℓ to be a function on [0, 1] that smoothly interpolates the values $\{(b/(m+1), a_{\ell b}) : b = 1, \dots, B\}$. Then, Theorem 2.2 applies and the Stieltjes transform $s_\tau(z)$ of the LSD of $\sqrt{n/p}(\mathbf{S}_\tau - \mathbf{\Sigma}_\tau)$ is given by

$$s_\tau(z) = -\sum_{b=1}^{B} \omega_b \frac{1}{z + \beta_{\tau,b}(z)}, \qquad z \in \mathbb{C}^+, \tag{3.1}$$

where the functions (Stieltjes transforms) $\beta_{\tau,b}(z)$ are determined by the system of nonlinear equations

$$\beta_{\tau,b}(z) = -\sum_{b'=1}^{B} \omega_{b'} \frac{\bar{R}_{\tau,bb'}}{z + \beta_{\tau,b'}(z)}, \qquad z \in \mathbb{C}^+, b = 1, \dots, B, \tag{3.2}$$

where

$$\bar{R}_{\tau,bb'} = \frac{1}{2\pi}\int_0^{2\pi} \cos^2(\tau\theta)\tilde{\psi}_b(\theta)\tilde{\psi}_{b'}(\theta)\,d\theta \tag{3.3}$$

with $\tilde{\psi}_b(\theta) = |1 + \sum_{\ell=1}^{\infty} a_{\ell b} e^{i\ell\theta}|^2$. Note that, using the notations of Theorem 2.2, $\beta_{\tau,b}(z) \equiv \beta_\tau(z, \mathbf{a})$ for $\mathbf{a} = b/(m+1)$, and $F^{\mathcal{A}}$ is the discrete distribution that associates probability ω_b to the point $b/(m+1)$, for $b = 1, \dots, B$. This example illustrates that often the precise description of f_ℓ's is not necessary in order for the LSDs to exist. Numerical methods, such as a fixed point method, for solving (3.2), while ensuring that $\Im(\beta_{\tau,b}(z)) > 0$ whenever $z \in \mathbb{C}^+$, are easy to implement, and can be used to compute $s_\tau(z)$ for any given z.

4. Applications

The main result (Theorem 2.2) gives a description of the bulk behavior of the eigenvalues of the matrices $\mathbf{C}_\tau = \sqrt{n/p}(\mathbf{S}_\tau - \mathbf{\Sigma}_\tau)$ under the stated assumptions on the process and the asymptotic regime $p/n \to 0$. Thus, this result provides a building block for further investigation of the behavior of spectral statistics of the same matrix. It can also be used to investigate potential departures from a hypothesized model.

An immediate application of Theorem 2.2 is that it provides a way of calculating an error bound on $\mathbf{S}_\tau$ as an estimate of $\mathbf{\Sigma}_\tau$. Indeed, if the quality of estimates is assessed through the Frobenius norm of $\mathbf{C}_\tau$, or some other measure that can be expressed as a linear functional of the spectral distribution of $\mathbf{C}_\tau$, our result gives a precise description about the asymptotic behavior of such a measure in terms of integrals of the LSD of $\mathbf{C}_\tau$. This can be seen as analogous to the standard error bounds in univariate problems.

One potential application is in the context of model diagnostics. Using the results for the LSD of the normalized symmetrized autocovariance matrices, one can check whether the residuals from a time series regression model have i.i.d. realizations. This can be done by graphically comparing the eigenvalue distributions of $\sqrt{n/p}\mathbf{S}_1^e, \sqrt{n/p}\mathbf{S}_2^e, \dots,$ where $\mathbf{S}_\tau^e$ is the lag-τ symmetrized autocovariance matrix of the residuals obtained from fitting a time series regression model, with the LSDs of the renormalized autocovariances of the same orders corresponding to i.i.d. data.

Further, these results can also be used to devise a formal test for the hypothesis H_0: $X_1, \dots, X_n$ are i.i.d. with zero mean and a known covariance, versus H_1: $X_1, \dots, X_n$ follow a stationary linear time series model. If an MA(q_0) process (q_0 can be ∞) is specified, another type of test may be proposed, say, H_0: X_t is the given MA(q_0) process (satisfying the assumptions of Theorem 2.2), versus the alternative that X_t is a different process than the one specified under H_0. This can be done through the construction of a class of test statistics that equal the squared integrals of the differences between the ESDs of observed renormalized sample covariance and autocovariance matrices and the corresponding LSDs under H_0, for certain lag orders. The LSDs under H_0 are computable by using the inversion formula of Stieltjes transforms whenever the Stieltjes transform of the LSDs can be computed numerically. An example of such a setting is given by Example 3.5. The actual numerical calculations of the LSD can be done along the lines of [15]. The test of whether a time series follows a given MA(q_0) model, with a fixed q_0, can be further facilitated by making use of the observation in Remark 2.1 which shows that if the process is indeed MA(q_0), then the LSDs of the renormalized lag-τ symmetrized sample autocovariances will be the same for all $\tau \geq q_0 + 1$.

Calculation of the theoretical LSD under the null model requires inversion of the corresponding Stieltjes transform, which is somewhat challenging due to the need for selection of the correct root, as it is necessary to let the imaginary part of the argument of the Stieltjes transform converge to zero. A simpler alternative is to compute the differences $|s_{\tau,p}(z) - s_\tau(z)|$ between the Stieltjes transforms of the ESD and the LSDs for a finite, pre-specified set of $z \in \mathbb{C}^+$, and then combine them through some norm (like l_∞, l_1 or l_2) and use the latter as a test statistic. The null distribution of this statistic can be simulated from a Gaussian ensemble, which can then be used to determine the critical values of the test.

If the linear process $(X_t : t \in \mathbb{Z})$ satisfies all the assumptions of Theorem 2.2 and all the coefficient matrices are determined by a finite dimensional parameter, then under suitable regularity conditions, it may be possible to estimate that parameter with error rate $O_P(1/\sqrt{n})$ through the use of method of moments or maximum likelihood (under the working assumption of Gaussianity). Supposing θ to be the parameter (possibly multi-dimensional), and assuming that $\mathbf{\Sigma}_\tau(\theta)$ is twice continuously differentiable, with $\max_{j,k} \|\frac{\partial^2}{\partial\theta_j \partial\theta_k^T} \mathbf{\Sigma}_\tau(\theta)\|$ uniformly bounded in a neighborhood of the true parameter θ_0, and denoting any $\sqrt{n}$-consistent estimate by $\hat{\theta}$, it can be shown by a simple application of Lemma S.8 that the ESD of $\sqrt{n/p}(\mathbf{S}_\tau - \mathbf{\Sigma}_\tau(\hat{\theta}))$ converges in probability to the same distribution as the LSD of $\sqrt{n/p}(\mathbf{S}_\tau - \mathbf{\Sigma}_\tau(\theta_0))$. Therefore, the hypothesis testing framework described in the previous paragraphs is applicable even if the parameter governing the system is estimated at a suitable precision and plugged into the expressions for the population autocovariances.

Another interesting application is in analyzing the effects of a linear filter applied to the observed time series. Linear filters are commonly used to extract signals from a time series through modulating its spectral characteristics and also for predicting future observations. Suppose that $W_t = \sum_{\ell=0}^\infty c_\ell X_{t-\ell}$ where $(X_t : t \in \mathbb{Z})$ is the MA(q) process defined in Section 2.1 and $(c_\ell : \ell \in \mathbb{N}_0)$ a sequence of filter coefficients satisfying $\sum_{\ell=0}^\infty |c_\ell| < \infty$. Then, the LSDs of the normalized symmetrized autocovariances of the filtered process $(W_t : t \in \mathbb{Z})$ exist and have the same structure as that of the process $(X_t : t \in \mathbb{Z})$, except that in the description of their Stieltjes transforms (equations (2.8) and (2.9)), the spectral density $\psi(\mathbf{a}, \nu)$ is replaced by the function $\tilde{\psi}(\mathbf{a}, \nu; \mathbf{c}) = |\sum_{\ell=0}^\infty c_\ell e^{\mathbf{i}\ell\nu}|^2 |g(\mathbf{a}, \nu)|^2$.

5. Proof of Theorem 2.1

First, make use of a simple observation regarding scaling factors. Since asymptotic spectral properties are unaffected by this change, in all of the proofs, the scaling factor $1/n$ is preferred over $1/(n-\tau)$ for simplicity of exposition. Throughout this section, it is assumed that the Z_{jt} are complex-valued and the $\mathbf{A}_\ell$ Hermitian matrices. If the Z_{jt} are real-valued and the $\mathbf{A}_\ell$ real, symmetric matrices, then the arguments need to be modified very slightly, as indicated in Section 11 of [8]. The key arguments in the proof of the real valued case remain the same, since as in the complex valued case, for Gaussian entries, after appropriate orthogonal transformations, the data matrix can be assumed to have independent Gaussian entries with zero mean and a variance profile determined by the spectrum of the process. We omit the details due to space constraints.

5.1. LSDs of $\mathbf{C}_\tau$ and $\tilde{\mathbf{C}}_\tau$

Recall that $\mathbf{C}_\tau$ defined in (2.5) is the renormalized version of the symmetrized autocovariance matrix $\mathbf{S}_\tau$. In this subsection it is shown that the LSDs of $\mathbf{C}_\tau^U = \mathbf{U}^*\mathbf{C}_\tau\mathbf{U}$ and $\tilde{\mathbf{C}}_\tau$ coincide, where the latter matrix is the renormalized version of $\tilde{\mathbf{S}}_\tau$ and defined in (2.13). Observe that the expectation of $\tilde{\mathbf{S}}_\tau$ is the diagonal matrix $\tilde{\boldsymbol{\Sigma}}_\tau = \operatorname{diag}(\tilde{\sigma}_{\tau,1}, \dots, \tilde{\sigma}_{\tau,p})$ given by

$$\tilde{\sigma}_{\tau,j} = \frac{1}{n}\sum_{t=1}^{n} \cos(\tau\nu_t)\psi(\boldsymbol{\alpha}_j, \nu_t), \qquad j = 1, \dots, p. \tag{5.1}$$

Now write $\mathbf{C}_\tau^U = \sqrt{n/p}(\mathbf{U}^*\mathbf{S}_\tau\mathbf{U} - \boldsymbol{\Sigma}_\tau^U)$, where $\boldsymbol{\Sigma}_\tau^U = \mathbf{U}^*\boldsymbol{\Sigma}_\tau\mathbf{U} = \operatorname{diag}(\sum_{\ell=0}^{q-\tau} f_\ell(\boldsymbol{\alpha}_j) \times f_{\ell+\tau}(\boldsymbol{\alpha}_j))_{j=1}^p$, and define $\mathbf{C}_\tau^{(1)} = \sqrt{n/p}(\mathbf{U}^*\mathbf{S}_\tau\mathbf{U} - \tilde{\boldsymbol{\Sigma}}_\tau)$.

We first show that $\boldsymbol{\Sigma}_\tau^U = \tilde{\boldsymbol{\Sigma}}_\tau$, which implies equality of the ESDs of $\mathbf{C}_\tau^U$ and $\mathbf{C}_\tau^{(1)}$. For each $j = 1, \dots, p$,

$$\begin{aligned}
\tilde{\sigma}_{\tau,j} &= \frac{1}{n}\sum_{t=1}^{n} \cos(\tau\nu_t)\psi(\boldsymbol{\alpha}_j, \nu_t) \\
&= \frac{1}{n}\sum_{t=1}^{n} \cos(\tau\nu_t) \sum_{\ell,\ell'=0}^{q} f_\ell(\boldsymbol{\alpha}_j) f_{\ell'}(\boldsymbol{\alpha}_j) e^{\mathbf{i}(\ell-\ell')\nu_t} \\
&= \frac{1}{2n}\sum_{\ell,\ell'=0}^{q} f_\ell(\boldsymbol{\alpha}_j) f_{\ell'}(\boldsymbol{\alpha}_j)\left(\sum_{t=1}^{n} e^{\mathbf{i}(\ell-\ell'+\tau)\nu_t} + \sum_{t=1}^{n} e^{\mathbf{i}(\ell-\ell'-\tau)\nu_t}\right) = \sum_{\ell=0}^{q-\tau} f_\ell(\boldsymbol{\alpha}_j) f_{\ell+\tau}(\boldsymbol{\alpha}_j),
\end{aligned} \tag{5.2}$$

since $\sum_{t=1}^{n} e^{\mathbf{i}k\nu_t} = n\delta_0(k)$ for $k = 0, 1, \dots, n-1$ where δ_0 denotes the Kronecker's delta function. This proves the assertion.

Lemma 5.1. *If the conditions of Theorem 2.1 are satisfied, then* $\|F^{\mathbf{C}_\tau^U} - F^{\tilde{\mathbf{C}}_\tau}\| \to 0$ *almost surely under* (2.3), *where* $F^{\mathbf{C}_\tau^U}$ *and* $F^{\tilde{\mathbf{C}}_\tau}$ *denote the ESDs of* $\mathbf{C}_\tau^U$ *and* $\tilde{\mathbf{C}}_\tau$, *respectively, and* $\|\cdot\|$ *denotes the sup-norm.*

Proof. Recall that $\tilde{\mathbf{C}}_\tau = \sqrt{n/p}(\mathbf{U}^*(\bar{\mathbf{S}}_\tau - \tilde{\boldsymbol{\Sigma}}_\tau)\mathbf{U})$. Exploiting the relation between $\mathbf{L}$ and $\tilde{\mathbf{L}}$, it can be shown that $\tilde{\mathbf{S}}_\tau = \mathbf{U}^*\bar{\mathbf{S}}_\tau\mathbf{U}$ can be written as at most $4(q+\tau+1)$ rank-one perturbations of $\mathbf{S}_\tau$. Hence, an application of the rank inequality given in Lemma S.6 implies that

$$\left\|F^{\mathbf{C}_\tau^{(1)}} - F^{\tilde{\mathbf{C}}_\tau}\right\| \le \frac{1}{p}\operatorname{rank}(\tilde{\mathbf{S}}_\tau - \mathbf{S}_\tau) \le \frac{4(q+\tau+1)}{p} \to 0 \tag{5.3}$$

under (2.3), which is the assertion since $F^{\mathbf{C}_\tau^{(1)}} = F^{\mathbf{C}_\tau^U}$. □

Define the Stieltjes transforms $s^U_{\tau,p}(z) = p^{-1}\operatorname{Tr}(\mathbf{C}^U_\tau - zI)^{-1}$ and $\tilde{s}_{\tau,p}(z) = p^{-1}\operatorname{Tr}(\tilde{\mathbf{C}}_\tau - zI)^{-1}$. Repeatedly applying Lemma S.1 to each of the rank-one perturbation matrices used in the proof of Lemma 5.1, it follows that, for any fixed $z = w + \mathbf{i}v \in \mathbb{C}^+$, $|s^U_{p,\tau}(z) - \tilde{s}_{p,\tau}(z)| \leq 4(q+\tau+1)/(vp)$ almost surely. It is therefore verified that the LSDs of $\mathbf{C}^U_\tau$ and $\tilde{\mathbf{C}}_\tau$ are almost surely identical.

5.2. Deterministic equation

In this section, a set of deterministic equations is derived that is asymptotically equivalent to the set of equations determining the Stieltjes transform of the limiting ESD of $\tilde{\mathbf{C}}_\tau$. The following decomposition will be useful in the proofs. Using assumptions (A1) and (A2) in combination with (2.12) and some matrix algebra, it can be shown that

$$\tilde{\mathbf{S}}_\tau = \mathbf{U}^*\bar{\mathbf{S}}_\tau\mathbf{U} = \mathbf{V}\boldsymbol{\Delta}_\tau\mathbf{V}^*,$$

where the $p \times n$ matrix $\mathbf{V}$ is defined through its entries

$$v_{jt} = \frac{1}{\sqrt{n}} g(\boldsymbol{\alpha}_j, \nu_t)\tilde{Z}_{jt}, \qquad j = 1, \dots, p, t = 1, \dots, n, \tag{5.4}$$

and $\boldsymbol{\Delta}_\tau = \operatorname{diag}(\cos(\tau\nu_1), \dots, \cos(\tau\nu_n))$. Let $\mathbf{V}_k$ denote the matrix obtained by replacing the kth row of $\mathbf{V}$ with zeros, and let the $n \times 1$ vector v_k be the kth column of the matrix $\mathbf{V}^* = (v_1 : v_2 : \cdots : v_p)$. Let further $\tilde{\boldsymbol{\Sigma}}_{\tau,k}$ be the matrix obtained from $\tilde{\boldsymbol{\Sigma}}_\tau$ by replacing its kth diagonal entry with 0. Denote by $\mathbf{D}_k$, respectively, $\mathbf{D}_{(k)}$ the matrices resulting from $\tilde{\mathbf{C}}_\tau$ from replacing the entries of its kth row, respectively, its kth row and kth column with zeros, that is,

$$\mathbf{D}_k = \sqrt{\frac{n}{p}}\big(\mathbf{V}_k\boldsymbol{\Delta}_\tau\mathbf{V}^* - \tilde{\boldsymbol{\Sigma}}_{\tau,k}\big) \quad \text{and} \quad \mathbf{D}_{(k)} = \sqrt{\frac{n}{p}}\big(\mathbf{V}_k\boldsymbol{\Delta}_\tau\mathbf{V}_k^* - \tilde{\boldsymbol{\Sigma}}_{\tau,k}\big).$$

Then,

$$\tilde{\mathbf{C}}_\tau = \mathbf{D}_k + \mathbf{H}_k = \mathbf{D}_{(k)} + \mathbf{H}_{(k)}, \tag{5.5}$$

where $\mathbf{H}_k = e_k h_k^*$ and $\mathbf{H}_{(k)} = \mathbf{H}_k + w_k e_k^T$ with e_k being the kth canonical unit vector of dimension p, $h_k = w_k + \eta_k e_k$,

$$w_k = \sqrt{\frac{n}{p}}\mathbf{V}_k\boldsymbol{\Delta}_\tau v_k \quad \text{and} \quad \eta_k = \sqrt{\frac{n}{p}}\big(v_k^*\boldsymbol{\Delta}_\tau v_k - \tilde{\sigma}_{\tau,k}\big), \tag{5.6}$$

where $\tilde{\sigma}_{\tau,j}$ is defined in (5.1), thereby ensuring that the kth entry of w_k is zero and collecting the kth diagonal element of $\tilde{\mathbf{C}}_\tau$ in the term η_k. Successively replacing rows of $\tilde{\mathbf{C}}_\tau$ with rows of zeros and noticing that $\tilde{\mathbf{C}}_\tau = \tilde{\mathbf{C}}_\tau^*$ as well as $\mathbf{H}_k^* = (e_k h_k^*)^* = h_k e_k^T$, the same arguments also yield

$$\tilde{\mathbf{C}}_\tau = \sum_{k=1}^{p} e_k h_k^* = \sum_{k=1}^{p} h_k e_k^T. \tag{5.7}$$

Observe next that, since its kth row and column consist of zero entries, e_k is an eigenvector of $\mathbf{D}_{(k)}$ with eigenvalue 0. If now, for $z \in \mathbb{C}^+$, $\mathbf{R}_{(k)}(z) = (\mathbf{D}_{(k)} - z\mathbf{I}_p)^{-1}$ denotes the resolvent of $\mathbf{D}_{(k)}$, then

$$\mathbf{R}_{(k)}(z)e_k = -\frac{1}{z}e_k, \tag{5.8}$$

that is, e_k is an eigenvector of $\mathbf{R}_{(k)}(z)$ with eigenvalue $-z^{-1}$. Let $\mathbf{R}_k(z) = (\mathbf{D}_k - z\mathbf{I}_p)^{-1}$ be the resolvent of $\mathbf{D}_k$. Utilizing (5.5) and Lemma S.1, it follows that

$$\mathbf{R}_k(z)e_k = \mathbf{R}_{(k)}(z)e_k - \frac{\mathbf{R}_{(k)}(z)w_k e_k^T \mathbf{R}_{(k)}(z)e_k}{1 + e_k^T \mathbf{R}_{(k)}(z)w_k} = -\frac{1}{z}e_k + \frac{1}{z}\mathbf{R}_{(k)}(z)w_k,$$

where the second step follows from invoking (5.8), for the denominator part in the middle expression additionally noticing that $\mathbf{R}_{(k)}(z) = \mathbf{R}^*_{(k)}(z)$ and that $e_k^T w_k = 0$ by construction. Now, all preliminary statements are collected that allow for a detailed study the resolvent and the Stieltjes transform of $\tilde{\mathbf{C}}_\tau$.

Lemma 5.2. *Under the assumptions of Theorem 2.1, it follows that the Stieltjes transform $\tilde{s}_{\tau,p}$ of $\tilde{\mathbf{C}}_\tau$ satisfies the equality*

$$\tilde{s}_{\tau,p}(z) = -\frac{1}{p}\sum_{k=1}^{p}\frac{1}{z + w_k^*\mathbf{R}_{(k)}(z)w_k - \eta_k}$$

for any fixed $z \in \mathbb{C}^+$.

Proof. Writing $\mathbf{I}_p + z(\tilde{\mathbf{C}}_\tau - z\mathbf{I}_p)^{-1} = (\tilde{\mathbf{C}}_\tau - z\mathbf{I}_p)^{-1}\tilde{\mathbf{C}}_\tau$, invoking (5.7) and Lemma S.1 implies that

$$\begin{aligned} \mathbf{I}_p + z(\tilde{\mathbf{C}}_\tau - z\mathbf{I}_p)^{-1} &= \sum_{k=1}^{p}(\tilde{\mathbf{C}}_\tau - z\mathbf{I}_p)^{-1}e_k h_k^* \\ &= \sum_{k=1}^{p}\mathbf{R}_k(z)e_k\left(1 - \frac{h_k^*\mathbf{R}_k(z)e_k}{1 + h_k^*\mathbf{R}_k(z)e_k}\right)h_k^* \\ &= \sum_{k=1}^{p}\frac{\mathbf{R}_k(z)e_k h_k^*}{1 + h_k^*\mathbf{R}_k(z)e_k}. \end{aligned} \tag{5.9}$$

Recall that the Stieltjes transform of $\tilde{\mathbf{C}}_\tau$ is given by $p^{-1}\operatorname{Tr}((\tilde{\mathbf{C}}_\tau - zI_p)^{-1})$. Therefore, taking trace on both sides of (5.9) and dividing by p leads to

$$\tilde{s}_{\tau,p}(z) = \frac{1}{zp}\sum_{k=1}^{p}\left(\frac{h_k^*\mathbf{R}_k(z)e_k}{1 + h_k^*\mathbf{R}_k(z)e_k} - 1\right) = -\frac{1}{zp}\sum_{k=1}^{p}\frac{1}{1 + h_k^*\mathbf{R}_k(z)e_k}. \tag{5.10}$$

In order to complete the proof of the lemma, it remains to study $h_k^*\mathbf{R}_k(z)e_k$. Using Lemma S.1 on $\mathbf{R}_k(z)$ and subsequently first utilizing (5.8) and then inserting the definition of w_k given in (5.6), it follows that

$$\begin{aligned} h_k^*\mathbf{R}_k(z)e_k &= h_k^*\mathbf{R}_{(k)}(z)e_k - h_k^*\frac{\mathbf{R}_{(k)}(z)w_k e_k^T\mathbf{R}_{(k)}(z)e_k}{1+e_k^T\mathbf{R}_{(k)}(z)w_k} \\ &= -\frac{1}{z}h_k^*e_k + \frac{1}{z}h_k^*\mathbf{R}_{(k)}(z)w_k \\ &= -\frac{1}{z}\eta_k + \frac{1}{z}w_k^*\mathbf{R}_{(k)}(z)w_k, \end{aligned} \tag{5.11}$$

where the third step also makes use of $e_k^T w_k = 0$. Plugging (5.11) into (5.10) finishes the proof. □

In the next auxiliary lemma, the expected value of the Stieltjes transform of $\tilde{\mathbf{C}}_\tau$ is determined. More generally, equations for the kernel

$$\tilde{\beta}_{\tau,p}(z,\mathbf{a}) = \frac{1}{p}\operatorname{Tr}\big((\tilde{\mathbf{C}}_\tau - zI_p)^{-1}\mathbf{\Gamma}_\tau(\mathbf{a})\big) \tag{5.12}$$

are introduced, where $\mathbf{\Gamma}_\tau(\mathbf{a}) = \operatorname{diag}(\mathcal{R}_\tau(\mathbf{a},\boldsymbol{\alpha}_k): k=1,\dots,p)$ with $\mathcal{R}_\tau(\mathbf{a},\boldsymbol{\alpha}_k)$ defined in (2.10). It is a central object of this study and the (approximate) finite-sample companion of the Stieltjes kernel $\beta_\tau(z,\mathbf{a})$ appearing in the statement of Theorem 2.1. Its properties will be further scrutinized in Sections 5.3 and 5.4.

Lemma 5.3. *Under the assumptions of Theorem 2.1, it follows that the expected value of the Stieltjes transform $\tilde{s}_{\tau,p}$ of $\tilde{\mathbf{C}}_\tau$ satisfies the equality*

$$\mathbb{E}\big[\tilde{s}_{\tau,p}(z)\big] = -\frac{1}{p}\sum_{k=1}^{p}\frac{1}{z+\mathbb{E}[\tilde{\beta}_{\tau,p}(z,\boldsymbol{\alpha}_k)]} + \tilde{\delta}_n \tag{5.13}$$

for any fixed $z\in\mathbb{C}^+$, where the remainder term $\tilde{\delta}_n$ converges to zero under (2.3). Moreover,

$$\mathbb{E}\big[\tilde{\beta}_{\tau,p}(z,\mathbf{a})\big] = -\frac{1}{p}\sum_{k=1}^{p}\frac{\mathcal{R}_\tau(\mathbf{a},\boldsymbol{\alpha}_k)}{z+\mathbb{E}[\tilde{\beta}_{\tau,p}(z,\boldsymbol{\alpha}_k)]} + \delta_n \tag{5.14}$$

for any fixed $z\in\mathbb{C}^+$, where the remainder term δ_n converges to zero under (2.3).

Proof. The proof of the lemma is given in three parts. In view of the expression for $\tilde{s}_{\tau,p}$ derived in Lemma 5.2, $\mathbb{E}[w_k^*\mathbf{R}_{(k)}(z)w_k]$ is estimated first and in the second step related to $\tilde{\beta}_{\tau,p}(z,\mathbf{a})$. The third step is concerned with the estimation of remainder terms δ_n and $\tilde{\delta}_n$.

Step 1: For $k = 1, \dots, p$, let $\mathbf{\Sigma}_{k,v} = \mathrm{Var}(v_k) = n^{-1}\,\mathrm{diag}(\psi(\boldsymbol{\alpha}_k, \nu_t) \colon t = 1, \dots, n)$ and further $\mathbf{\Xi}_{\tau,k} = \mathbf{\Delta}_\tau \mathbf{\Sigma}_{k,v} \mathbf{\Delta}_\tau = n^{-1}\,\mathrm{diag}(\cos^2(\tau\nu_t)\psi(\boldsymbol{\alpha}_k, \nu_t) \colon t = 1, \dots, n)$. Define

$$\gamma_{\tau,j}(\mathbf{a}) := \frac{1}{n}\sum_{t=1}^{n} \cos^2(\tau\nu_t)\psi(\mathbf{a}, \nu_t)\psi(\boldsymbol{\alpha}_j, \nu_t),$$

and observe that $\gamma_{\tau,j}(\mathbf{a}) = \mathcal{R}(\mathbf{a}, \boldsymbol{\alpha}_j)$ for all $j = 1, \dots, p$. This follows calculations similar to those leading to 5.2. Define the matrix $\Gamma_{\tau,k}(\mathbf{a})$ as the one obtained from $\Gamma_\tau(\mathbf{a})$ by replacing its kth diagonal entry with zero. Observe next that the definition of w_k in (5.6) implies that it suffices to estimate the following expectation, for which it holds that

$$\begin{aligned}
&\frac{n}{p}\mathbb{E}\big[v_k^* \mathbf{\Delta}_\tau \mathbf{V}_k^* \mathbf{R}_{(k)}(z) \mathbf{V}_k \Delta_\tau v_k\big] \\
&= \frac{n}{p}\mathbb{E}\big[\mathrm{Tr}\big(\mathbf{\Delta}_\tau v_k v_k^* \mathbf{\Delta}_\tau \mathbf{V}_k^* \mathbf{R}_{(k)}(z) \mathbf{V}_k\big)\big] \\
&= \frac{n}{p}\mathrm{Tr}\big(\mathbf{\Delta}_\tau \mathbf{\Sigma}_k \mathbf{\Delta}_\tau \mathbb{E}\big[\mathbf{V}_k^* \mathbf{R}_{(k)}(z) \mathbf{V}_k\big]\big) = \frac{n}{p}\mathbb{E}\big[\mathrm{Tr}\big(\mathbf{V}_k \mathbf{\Xi}_{\tau,k} \mathbf{V}_k^* \mathbf{R}_{(k)}(z)\big)\big] \qquad (5.15) \\
&= \frac{n}{p}\sum_{j\neq k}\mathbb{E}\big[v_j^* \mathbf{\Xi}_{\tau,k} v_j \big(\mathbf{R}_{(k)}(z)\big)_{jj}\big] = \frac{1}{p}\sum_{j\neq k}\mathbb{E}\big[\gamma_{\tau,j}(\boldsymbol{\alpha}_k)\big(\mathbf{R}_{(k)}(z)\big)_{jj}\big] + d_k^{(0)} \\
&= \frac{1}{p}\mathbb{E}\big[\mathrm{Tr}\big(\mathbf{R}_{(k)}(z)\mathbf{\Gamma}_{\tau,k}(\boldsymbol{\alpha}_k)\big)\big] + d_k^{(0)},
\end{aligned}$$

where independence between v_k and $\mathbf{V}_k$ was used to obtain the second equality and

$$d_k^{(0)} = \frac{1}{p}\sum_{j\neq k}\mathbb{E}\big[\big(n v_j^* \mathbf{\Xi}_{\tau,k} v_j - \gamma_{\tau,j}(\boldsymbol{\alpha}_k)\big)\big(\mathbf{R}_{(k)}(z)\big)_{jj}\big]. \qquad (5.16)$$

An application of the Cauchy–Schwarz inequality to the expectation on the right-hand side of (5.16), subsequently using the fact that $\max_j |(\mathbf{R}_{(k)}(z))_{jj}| \le \Im(z)^{-1}$ and squaring the resulting estimate, yields that

$$\big|d_k^{(0)}\big|^2 \le \frac{1}{p\Im(z)^2}\sum_{j\neq k}\mathbb{E}\big[\big|n v_j^* \mathbf{\Xi}_{\tau,k} v_j - \gamma_{\tau,j}(\boldsymbol{\alpha}_k)\big|^2\big] = \frac{1}{p\Im(z)^2}\sum_{j\neq k}\mathrm{Var}\big(n v_j^* \mathbf{\Xi}_{\tau,k} v_j\big) \le \frac{C^2}{p\Im(z)^2},$$

where the equality follows from recognizing that $\mathbb{E}[n v_j^* \mathbf{\Xi}_{\tau,k} v_j] = \gamma_{\tau,j}(\boldsymbol{\alpha}_k)$ and the inequality from observing that each $n v_j^* \mathbf{\Xi}_{\tau,k} v_j$ is a quadratic form in the i.i.d. standard Gaussians $\tilde{Z}_{j1}, \dots \tilde{Z}_{jn}$ and has bounded variance. Taking the square root gives

$$\big|d_k^{(0)}\big| \le \frac{C}{\sqrt{p}\Im(z)} \qquad (5.17)$$

for some constant $C > 0$.

Step 2: Multiplying $\boldsymbol{\Gamma}_\tau(\mathbf{a})$ to both sides of the equation $\mathbf{I}_p + z(\tilde{\mathbf{C}}_\tau - z\mathbf{I}_p)^{-1} = \tilde{\mathbf{C}}_\tau(\tilde{\mathbf{C}}_\tau - z\mathbf{I}_p)^{-1}$, then following the arguments that led to (5.9), and making use of $\boldsymbol{\Gamma}_\tau(\mathbf{a})e_k = \mathcal{R}(\mathbf{a}, \boldsymbol{\alpha}_k)e_k$ gives

$$\boldsymbol{\Gamma}_\tau(\mathbf{a}) + z\boldsymbol{\Gamma}_\tau(\mathbf{a})(\tilde{\mathbf{C}}_\tau - z\mathbf{I}_p)^{-1} = \sum_{k=1}^p \mathcal{R}_\tau(\mathbf{a}, \boldsymbol{\alpha}_k)e_k h_k^*(\tilde{\mathbf{C}}_\tau - z\mathbf{I}_p)^{-1} = \sum_{k=1}^p \frac{\mathcal{R}_\tau(\mathbf{a}, \boldsymbol{\alpha}_k)e_k h_k^* \mathbf{R}_k(z)}{1 + h_k^* \mathbf{R}_k(z)e_k}.$$

Further taking trace on both sides and invoking (5.11) yields

$$\begin{aligned}\tilde{\beta}_{\tau,p}(z, \mathbf{a}) &= -\frac{1}{p}\sum_{k=1}^p \frac{\mathcal{R}_\tau(\mathbf{a}, \boldsymbol{\alpha}_k)}{z + w_k^* \mathbf{R}_{(k)}(z)w_k - \eta_k} \\ &= -\frac{1}{p}\sum_{k=1}^p \frac{\mathcal{R}_\tau(\mathbf{a}, \boldsymbol{\alpha}_k)}{z + \mathbb{E}[\tilde{\beta}_{\tau,p}(z, \boldsymbol{\alpha}_k)] - \varepsilon_k},\end{aligned} \tag{5.18}$$

where $\varepsilon_k = \mathbb{E}[\tilde{\beta}_{\tau,p}(z, \boldsymbol{\alpha}_k)] - w_k^* \mathbf{R}_{(k)}(z)w_k + \eta_k$. Taking expectation on the left- and right-hand side of (5.18) leads to equation (5.14) with the remainder term having the explicit form

$$\begin{aligned}\delta_n &= -\frac{1}{p}\sum_{k=1}^p \frac{\mathcal{R}_\tau(\mathbf{a}, \boldsymbol{\alpha}_k)\mathbb{E}[\varepsilon_k]}{(z + \mathbb{E}\tilde{\beta}_{\tau,p}(z, \boldsymbol{\alpha}_k))^2} - \frac{1}{p}\sum_{k=1}^p \mathbb{E}\biggl(\frac{\mathcal{R}_\tau(\mathbf{a}, \boldsymbol{\alpha}_k)\varepsilon_k^2}{(z + \mathbb{E}[\tilde{\beta}_{\tau,p}(z, \boldsymbol{\alpha}_k)])^2(z + \mathbb{E}[\tilde{\beta}_{\tau,p}(z, \boldsymbol{\alpha}_k)] - \varepsilon_k)}\biggr) \\ &= \delta_{n,1} + \delta_{n,2}.\end{aligned}$$

It remains to show that $\delta_n \to 0$ under (2.3). This will be done in the next step.

Step 3: To show that $\delta_n \to 0$, it suffices to verify that $\delta_{n,1} \to 0$ and $\delta_{n,2} \to 0$. Note that, since $\tilde{\beta}_{\tau,p}(z, \boldsymbol{\alpha}_k)$ is a Stieltjes transform of a measure,

$$\bigl|z + \mathbb{E}\bigl[\tilde{\beta}_{\tau,p}(z, \boldsymbol{\alpha}_k)\bigr]\bigr| \geq \Im\bigl(z + \mathbb{E}\bigl[\tilde{\beta}_{\tau,p}(z, \boldsymbol{\alpha}_k)\bigr]\bigr) \geq \Im(z) + \mathbb{E}\bigl[\Im\bigl(\tilde{\beta}_{\tau,p}(z, \boldsymbol{\alpha}_k)\bigr)\bigr] \geq \Im(z)$$

and since $\eta_k \in \mathbb{R}$, and $\mathbf{w}_k^* \mathbf{R}_{(k)}(z)\mathbf{w}_k$ is a Stieltjes transform of a measure,

$$\bigl|z + \mathbb{E}\bigl[\tilde{\beta}_{\tau,p}(z, \boldsymbol{\alpha}_k)\bigr] - \varepsilon_k\bigr| = \bigl|z + \mathbf{w}_k^* \mathbf{R}_{(k)}(z)\mathbf{w}_k - \eta_k\bigr| \geq \Im(z) + \Im\bigl(\mathbf{w}_k^* \mathbf{R}_{(k)}(z)\mathbf{w}_k\bigr) \geq \Im(z).$$

Thus, since moreover $|\mathcal{R}_\tau(\mathbf{a}, \mathbf{b})| \leq L_1^2$ with L_1 as in (A4), it only needs to be shown that $\max_k |\mathbb{E}[\varepsilon_k]| \to 0$ and $\max_k \mathbb{E}[|\varepsilon_k - \mathbb{E}[\varepsilon_k]|^2] \to 0$.

Let $\tilde{\mathbf{R}}(z) = (\tilde{\mathbf{C}}_\tau - z\mathbf{I})^{-1}$. Since $\mathbb{E}[\eta_k] = 0$, it follows from (5.15) and (5.12) that

$$\begin{aligned}\bigl|\mathbb{E}[\varepsilon_k]\bigr| &= \biggl|\frac{1}{p}\mathbb{E}\bigl[\operatorname{Tr}\bigl(\tilde{\mathbf{R}}(z)\boldsymbol{\Gamma}_\tau(\boldsymbol{\alpha}_k)\bigr)\bigr] - \frac{1}{p}\mathbb{E}\bigl[\operatorname{Tr}\bigl(\mathbf{R}_{(k)}(z)\boldsymbol{\Gamma}_{\tau,k}(\boldsymbol{\alpha}_k)\bigr)\bigr] - d_k^{(0)}\biggr| \\ &\leq \frac{1}{p}\bigl|\mathbb{E}\bigl[\operatorname{Tr}\bigl(\tilde{\mathbf{R}}(z)\boldsymbol{\Gamma}_\tau(\boldsymbol{\alpha}_k)\bigr)\bigr] - \mathbb{E}\bigl[\operatorname{Tr}\bigl(\mathbf{R}_{(k)}(z)\boldsymbol{\Gamma}_\tau(\boldsymbol{\alpha}_k)\bigr)\bigr]\bigr| \\ &\quad + \frac{1}{p}\bigl|\mathbb{E}\bigl[\operatorname{Tr}\bigl(\mathbf{R}_{(k)}(z)\bigl\{\boldsymbol{\Gamma}_\tau(\boldsymbol{\alpha}_k) - \boldsymbol{\Gamma}_{\tau,k}(\boldsymbol{\alpha}_k)\bigr\}\bigr)\bigr]\bigr| + \bigl|d_k^{(0)}\bigr| \\ &= d_k^{1,1} + d_k^{1,2} + \bigl|d_k^{(0)}\bigr|,\end{aligned} \tag{5.19}$$

where $\mathbf{\Gamma}_{\tau,k}(\boldsymbol{\alpha}_k) = \mathbf{\Gamma}_\tau(\boldsymbol{\alpha}_k) - \mathcal{R}_\tau(\boldsymbol{\alpha}_k, \boldsymbol{\alpha}_k)e_k e_k^T$. Arguments as the more general ones leading to (5.21), imply that $\max_k d_k^{1,1} \le 6qL_1^2(p\Im(z))^{-1}$. Since $\|\mathbf{R}_{(k)}(z)\| \le (\Im(z))^{-1}$ and $\mathcal{R}_\tau(\boldsymbol{\alpha}_k, \boldsymbol{\alpha}_k)$ is uniformly bounded, it follows that $\max_k d_k^{1,2} \le L_1^2(p\Im(z))^{-1}$. Together with (5.17) and (5.19), these guarantee that $\max_k |\mathbb{E}[\varepsilon_k]| \to 0$ and thus $|\delta_{n,1}| \le L_1^2(\Im(z))^{-2} \max_k |\mathbb{E}[\varepsilon_k]| \to 0$.

Observe next that, by (5.15),

$$\begin{aligned}
\mathbb{E}\big[\big|\varepsilon_k - \mathbb{E}[\varepsilon_k]\big|^2\big] &= \mathbb{E}\bigg[\bigg|-w_k^*\mathbf{R}_{(k)}(z)w_k + \frac{1}{p}\mathbb{E}\big[\mathrm{Tr}\big(\mathbf{R}_{(k)}(z)\mathbf{\Gamma}_{\tau,k}(\boldsymbol{\alpha}_k)\big)\big] + d_k^{(0)} + \eta_k\bigg|^2\bigg] \\
&\le 3\mathbb{E}\bigg[\bigg|-w_k^*\mathbf{R}_{(k)}(z)w_k + \frac{1}{p}\mathrm{Tr}\big(\mathbf{R}_{(k)}(z)\mathbf{\Gamma}_{\tau,k}(\boldsymbol{\alpha}_k)\big) + \eta_k\bigg|^2\bigg] \\
&\quad + 3\mathbb{E}\bigg[\bigg|\frac{1}{p}\mathrm{Tr}\big(\mathbf{R}_{(k)}(z)\mathbf{\Gamma}_{\tau,k}(\boldsymbol{\alpha}_k)\big) - \mathbb{E}\bigg[\frac{1}{p}\mathrm{Tr}\big(\mathbf{R}_{(k)}(z)\mathbf{\Gamma}_{\tau,k}(\boldsymbol{\alpha}_k)\big)\bigg]\bigg|^2\bigg] + 3\big|d_k^{(0)}\big|^2 \\
&= d_k^{2,1} + d_k^{2,2} + 3\big|d_k^{(0)}\big|^2,
\end{aligned}$$

where

$$\begin{aligned}
d_k^{2,1} &\le 6\mathbb{E}\bigg[\bigg|-w_k^*\mathbf{R}_{(k)}(z)w_k + \frac{1}{p}\mathrm{Tr}\big(\mathbf{R}_{(k)}(z)\mathbf{\Gamma}_{\tau,k}(\boldsymbol{\alpha}_k)\big)\bigg|^2\bigg] + 6\mathbb{E}\big[|\eta_k|^2\big] \\
&= 6d_k^{2,3} + 6\mathbb{E}\big[|\eta_k|^2\big].
\end{aligned}$$

Now, $\max_k \mathbb{E}[|\eta_k|^2] < Cp^{-1}$ for some $C > 0$ as proved in Section S.4.1. It is shown in Sections S.4.2 and S.4.3 that $\max_k d_k^{2,2} \to 0$ and $\max_k d_k^{2,3} \to 0$, respectively. Consequently, $\max_k \mathbb{E}[|\varepsilon_k - \mathbb{E}[\varepsilon_k]|^2] \to 0$ and hence also $\delta_{n,2} \to 0$.

Step 4: Using the expression for $\tilde{s}_{\tau,p}(z)$ derived in Lemma 5.2, relation (5.13) can be obtained from similar arguments as in Steps 1–3 of this proof. In particular, it can be shown that $\tilde{\delta}_n \to 0$. □

5.3. Convergence of random part

In this section, it is shown that, almost surely $s_{\tau,p}(z) - \mathbb{E}[s_{\tau,p}(z)] \to 0$ and $\beta_{\tau,p}(z, \mathbf{a}) - \mathbb{E}[\beta_{\tau,p}(z, \mathbf{a})] \to 0$ for any $z \in \mathbb{C}^+$ when the entries of $\mathbf{Z}$ are i.i.d. standardized random variables with arbitrary distributions. The concentration inequalities on $s_{\tau,p}(z)$ and $\beta_{\tau,p}(z, \mathbf{a})$ are derived by using the McDiarmid's inequality given in Lemma S.2 and the proof of almost sure convergence is obtained through the use of the Borel–Cantelli lemma. To apply the McDiarmid inequality, treat $\mathbf{C}_\tau$ as a function of the independent rows of $\mathbf{Z}$, say, $\mathbf{z}_1^*, \dots, \mathbf{z}_p^*$. Let

$$\mathbf{Z}_{(j)} = \mathbf{Z} - e_j e_j^T\mathbf{Z} = \mathbf{Z} - e_j\mathbf{z}_j^*, \qquad j = 1, \dots, p,$$

where $\mathbf{Z} = [\mathbf{z}_1^* : \cdots : \mathbf{z}_p^*]^*$. Let further $\mathbf{X}_{(j)}$ be the $p \times n$ matrix obtained from the original data matrix $\mathbf{X}$ with the jth row removed, that is,

$$\mathbf{X}_{(j)} = \sum_{\ell=0}^{q} \mathbf{A}_\ell \mathbf{Z}_{(j)} \mathbf{L}^\ell.$$

Define $\mathbf{S}_\tau^{(j)} = n^{-1}\mathbf{X}_{(j)}\mathbf{D}_\tau\mathbf{X}_{(j)}^*$ and $\mathbf{C}_\tau^{(j)} = \sqrt{n/p}(\mathbf{S}_\tau^{(j)} - \mathbf{\Sigma}_\tau)$, where $\mathbf{D}_\tau = [\mathbf{L}^\tau + (\mathbf{L}^\tau)^*]/2$. It follows then from the relation

$$\begin{aligned}\mathbf{S}_\tau &= \frac{1}{n}\left(\sum_{\ell=0}^{q} \mathbf{A}_\ell\big(\mathbf{Z}_{(j)} + e_j\mathbf{z}_j^*\big)\mathbf{L}^\ell\right)\mathbf{D}_\tau\left(\sum_{\ell=0}^{q} \mathbf{A}_\ell\big(\mathbf{Z}_{(j)} + e_j\mathbf{z}_j^*\big)\mathbf{L}^\ell\right)^* \\ &= \mathbf{S}_\tau^{(j)} + \frac{1}{n}\left(\sum_{\ell=0}^{q} a_{j\ell}y_{j\ell}^*\mathbf{D}_\tau\mathbf{X}_{(j)}^* + \sum_{\ell=0}^{q}\mathbf{X}_{(j)}\mathbf{D}_\tau y_{j\ell}a_{j\ell}^* + \sum_{\ell,\ell'=0}^{q} a_{j\ell}y_{j\ell}^*\mathbf{D}_\tau y_{j\ell'}a_{j\ell'}^*\right),\end{aligned}$$

where $a_{j\ell} = \mathbf{A}_\ell e_j$, $y_{j\ell}^* = \mathbf{z}_j^*\mathbf{L}^\ell$, that

$$\mathbf{C}_\tau = \mathbf{C}_\tau^{(j)} + \sum_{\ell=0}^{q} a_{j\ell}\zeta_{j\ell}^* + \sum_{\ell=0}^{q}\zeta_{j\ell}a_{j\ell}^* + \sum_{\ell,\ell'=0}^{q}\omega_{\ell,\ell'}^{j}a_{j\ell}a_{j\ell'}^*, \tag{5.20}$$

making use of the notations $\zeta_{j\ell} = (np)^{-1/2}y_{j\ell}^*\mathbf{\Delta}\mathbf{X}_{(j)}^*$ and $\omega_{\ell,\ell'}^j = (pn)^{-1/2}y_{j\ell}^*\mathbf{\Delta}y_{j\ell'}$. The following lemma will be instrumental in determining the convergence of the random part.

Lemma 5.4. *Under the assumptions of Theorem 2.1, it follows that*

$$\operatorname{diff}_{\tau,j}(\mathbf{H}) = \frac{1}{p}\left|\operatorname{Tr}\big((\mathbf{C}_\tau - z\mathbf{I})^{-1}\mathbf{H}\big) - \frac{1}{p}\operatorname{Tr}\big(\big(\mathbf{C}_\tau^{(j)} - z\mathbf{I}\big)^{-1}\mathbf{H}\big)\right| \le \frac{3(q+1)\|\mathbf{H}\|}{p\Im(z)},$$

where $\mathbf{H}$ *is an arbitrary* $p \times p$ *Hermitian matrix with* $\|\mathbf{H}\|$ *bounded.*

Proof. First observe that $\sum_{\ell,\ell'=0}^{q}\omega_{\ell,\ell'}^{j}a_{j\ell}a_{j\ell'}^*$ is a Hermitian matrix of rank $q+1$ and hence we can write it as $\sum_{\ell=0}^{q}\tilde{\omega}_{j\ell}b_{j\ell}b_{j\ell}^*$, where each $\tilde{\omega}_{j\ell} \in \{-1, +1\}$ and observe that $a_{j\ell}\zeta_{j\ell}^* + \zeta_{j\ell}a_{j\ell}^* = u_{j\ell}u_{j\ell}^* - v_{j\ell}v_{j\ell}^*$ where $u_{j\ell} = 2^{-1/2}(\zeta_{j\ell} + a_{j\ell})$ and $v_{j\ell} = 2^{-1/2}(\zeta_{j\ell} - a_{j\ell})$. Define the matrices $\mathbf{D}_{1j} = \mathbf{C}_\tau^{(j)} + \sum_{\ell=0}^{q}u_{j\ell}u_{j\ell}^*$ and $\mathbf{D}_{2j} = \mathbf{D}_{1j} - \sum_{\ell=0}^{q}v_{j\ell}v_{j\ell}^*$, and notice that it then follows from (5.20) that $\mathbf{C}_\tau = \mathbf{D}_{2j} + \sum_{\ell=0}^{q}\tilde{\omega}_{j\ell}b_{j\ell}b_{j\ell}^*$. Therefore,

$$\begin{aligned}\operatorname{diff}_{\tau,j}(\mathbf{H}) &\le \frac{1}{p}\big|\operatorname{Tr}\big((\mathbf{C}_\tau - z\mathbf{I})^{-1}\mathbf{H}\big) - \operatorname{Tr}\big((\mathbf{D}_{2j} - z\mathbf{I})^{-1}\mathbf{H}\big)\big| \\ &\quad + \frac{1}{p}\big|\operatorname{Tr}\big((\mathbf{D}_{2j} - z\mathbf{I})^{-1}\mathbf{H}\big) - \operatorname{Tr}\big((\mathbf{D}_{1j} - z\mathbf{I})^{-1}\mathbf{H}\big)\big|\end{aligned}$$

$$+\frac{1}{p}\big|\mathrm{Tr}\big((\mathbf{D}_{1j}-z\mathbf{I})^{-1}\mathbf{H}\big)-\mathrm{Tr}\big(\big(\mathbf{C}_\tau^{(j)}-z\mathbf{I}\big)^{-1}\mathbf{H}\big)\big|$$
$$=K_{j1}+K_{j2}+K_{j3}.$$

In the following an estimate for K_{j2} is given. For $1\le k\le q+1$, let then $\mathbf{T}_j^{(k)}=\mathbf{D}_{2j}+\sum_{\ell=0}^{k-1}v_{j\ell}v_{j\ell}^*$, so that $\mathbf{T}_j^{(0)}=\mathbf{D}_{2j}$ and $\mathbf{T}_j^{(q+1)}=\mathbf{D}_{1j}$. An application of Lemmas S.1 and S.3 implies that

$$\begin{aligned}K_{j2}&=\frac{1}{p}\sum_{k=1}^{q+1}\big|\mathrm{Tr}\big(\big(\mathbf{T}_j^{(k)}-z\mathbf{I}\big)^{-1}\mathbf{H}\big)-\mathrm{Tr}\big(\big(\mathbf{T}_j^{(k-1)}-z\mathbf{I}\big)^{-1}\mathbf{H}\big)\big|\\&\le\frac{1}{p}\sum_{k=1}^{q+1}\left|\frac{v_{jk}^*(\mathbf{T}_j^{(k-1)}-z\mathbf{I})^{-1}\mathbf{H}(\mathbf{T}_j^{(k-1)}-z\mathbf{I})^{-1}v_{jk}}{1+v_{jk}^*(\mathbf{T}_j^{(k-1)}-z\mathbf{I})^{-1}v_{jk}}\right|\le\frac{(q+1)\|\mathbf{H}\|}{p\Im(z)}.\end{aligned}$$

Estimates for K_1 and K_3 can be obtained in a similar way, leading to the bound $(q+1)(p\Im(z))^{-1}\|\mathbf{H}\|$ in each case. This proves the lemma. □

Lemma 5.4 gives the bound $\mathrm{diff}_{\tau,j}(\mathbf{I}_p)\le 3(q+1)(p\Im(z))^{-1}$ and $\mathrm{diff}_{\tau,j}(\boldsymbol{\Gamma}_\tau(\mathbf{a}))\le 3(q+1)(p\Im(z))^{-1}L_1^2$. Let $\mathrm{diff}'_{\tau,j}$ be defined as $\mathrm{diff}_{\tau,j}$ with $\mathbf{C}_\tau$ replaced with $\mathbf{C}'_\tau$, where the latter matrix in turn is obtained from the former replacing its jth's row $\mathbf{z}_j^*$ with an independent copy $(\mathbf{z}'_j)^*$. From Lemma 5.4, it follows then that

$$\frac{1}{p}\big|\mathrm{Tr}\big((\mathbf{C}_\tau-z\mathbf{I})^{-1}\big)-\mathrm{Tr}\big(\big(\mathbf{C}'_\tau-z\mathbf{I}\big)^{-1}\big)\big|\le\frac{6(q+1)}{p\Im(z)}$$

and

$$\frac{1}{p}\big|\mathrm{Tr}\big((\mathbf{C}_\tau-z\mathbf{I})^{-1}\Gamma_\tau(\mathbf{a})\big)-\mathrm{Tr}\big(\big(\mathbf{C}'_\tau-z\mathbf{I}\big)^{-1}\Gamma_\tau(\mathbf{a})\big)\big|\le\frac{6(q+1){L_1}^2}{p\Im(z)}.\tag{5.21}$$

Recognizing that $s_{\tau,p}(z)=p^{-1}\,\mathrm{Tr}((\mathbf{C}_\tau-z\mathbf{I})^{-1})$ and $\beta_{\tau,p}(z,\mathbf{a})=p^{-1}\,\mathrm{Tr}((\mathbf{C}_\tau-z\mathbf{I})^{-1}\Gamma_\tau(\mathbf{a}))$ and applying the McDiarmid's inequality (Lemma S.2) yields that, for any $\varepsilon>0$,

$$\mathbb{P}\big(\big|s_{\tau,p}(z)-\mathbb{E}\big[s_{\tau,p}(z)\big]\big|>\varepsilon\big)\le 4\exp\left(-\frac{p\Im(z)\varepsilon^2}{18(q+1)^2}\right)\tag{5.22}$$

and

$$\mathbb{P}\big(\big|\beta_{\tau,p}(z,\mathbf{a})-\mathbb{E}\big[\beta_{\tau,p}(z,\mathbf{a})\big]\big|>\varepsilon\big)\le 4\exp\left(-\frac{p\Im(z)\varepsilon^2}{18(q+1)^2{L_1}^2}\right).\tag{5.23}$$

Now the Borel–Cantelli lemma implies that $|s_{\tau,p}(z)-\mathbb{E}[s_{\tau,p}(z)]|\to 0$ and $|\beta_{\tau,p}(z,\mathbf{a})-\mathbb{E}[\beta_{\tau,p}(z,\mathbf{a})]|\to 0$ almost surely under (2.3). Moreover, it can be readily seen that these almost sure convergence results also hold for $\tilde{s}_{\tau,p}$ and $\tilde{\beta}_{\tau,p}$.

5.4. Existence, uniqueness and continuity of the solution

This section provides a proof of the existence of a unique solution $s_\tau(z)$ and $\beta_\tau(z, \mathbf{a})$, for $\mathbf{a} \in \operatorname{supp}(F^{\mathcal{A}})$ and $z \in \mathbb{C}^+$, to the set of equations (2.8)–(2.10). Assuming that these solutions exist, it can be shown that $\tilde{s}_{\tau,p}(z) \xrightarrow{\text{a.s.}} s_\tau(z)$ and $\tilde{\beta}_{\tau,p}(z, \mathbf{a}) \xrightarrow{\text{a.s.}} \beta_\tau(z, \mathbf{a})$ for any $\mathbf{a} \in \operatorname{supp}(F^{\mathcal{A}})$ and $z \in \mathbb{C}^+$. In view of the results derived in Section 5.3 and Lemma 5.3, it suffices to show that for every sequence $\{p_j : j \in \mathbb{N}\}$ there exists a further subsequence $\{\tilde{p}_j : j \in \mathbb{N}\}$ such that $\mathbb{E}(\tilde{\beta}_{\tau,\tilde{p}_j}(z, \mathbf{a}))$ converges to a limit $\beta_\tau(z, \mathbf{a})$ satisfying (2.8)–(2.10). The verification is based on a diagonal subsequence argument and the Arzelà–Ascoli theorem.

Lemma 5.5. *Let $\{p_j : j \in \mathbb{N}\}$ denote a subsequence of the integers $\mathbb{N}$ and define $\rho_{\tau,p_j}(z, \mathbf{a}) = \mathbb{E}[\tilde{\beta}_{\tau,p_j}(z, \mathbf{a})]$. Then the following statements hold.*

(a) *There is a further subsequence $\{\tilde{p}_j : j \in \mathbb{N}\}$ such that $\rho_{\tau,\tilde{p}_j}(z, \mathbf{a})$ convergences uniformly in $\mathbf{a} \in \operatorname{supp}(F^{\mathcal{A}})$ and pointwise in $z \in \mathbb{C}^+$ to a limit $\rho_\tau(z, \mathbf{a})$ which is analytic in z and continuous in* $\mathbf{a}$;

(b) *The limit $\rho_\tau(z, \mathbf{a})$ in (a) coincides with $\beta_\tau(z, \mathbf{a})$ and is the Stieltjes transform of a measure on the real line with mass $\int \mathcal{R}_\tau(\mathbf{a}, \mathbf{b})\, dF^{\mathcal{A}}(\mathbf{b})$ satisfying (2.9).*

Proof. *Step* 1: Define $\mathcal{F} = \{\rho_{\tau,p_j(\mathbf{a})}(\cdot, \mathbf{a}) : \mathbf{a} \in \operatorname{supp}(F^{\mathcal{A}})\}$. For any compact set $K \subset \mathbb{C}^+$,

$$\left|\rho_{\tau,p_j(\mathbf{a})}(z, \mathbf{a})\right| \le L_1^2 \Big/ \min_{z \in K} \Im(z) = M(K).$$

Let $\{\mathbf{a}_1, \mathbf{a}_2, \ldots\}$ be an enumeration of the dense subset $\operatorname{supp}(F^{\mathcal{A}}) \cap \mathbb{Q}^m$ of $\operatorname{supp}(F^{\mathcal{A}})$. An application of Lemma S.9 yields that for any $\mathbf{a}_\ell$ there exists a further subsequence $\{p_j(\mathbf{a}_\ell) : j \in \mathbb{N}\}$ such that $\cdots \subset \{p_j(\mathbf{a}_\ell)\} \subset \{p_j(\mathbf{a}_{\ell-1})\} \subset \cdots \subset \{p_j(\mathbf{a}_1)\}$ such that $\rho_{\tau,p_j(\mathbf{a}_\ell)}(z, \mathbf{a}_\ell)$ converges uniformly on compact subsets of $\mathbb{C}^+$ to a limit denoted by $\rho_\tau(z, \mathbf{a}_\ell)$, which is an analytic function of $z \in \mathbb{C}^+$ for each $\ell \in \mathbb{N}$. Choosing the diagonal subsequence $\{p_j(\mathbf{a}_j) : \mathbb{N}\}$, it follows that

$$\rho_{\tau,p_j(\mathbf{a}_j)}(z, \mathbf{a}_\ell) \to \rho_\tau(z, \mathbf{a}_\ell) \qquad (j \to \infty)$$

for all $\ell \in \mathbb{N}$ uniformly on compact subsets of $\mathbb{C}^+$. Note that the limit is defined on $\mathbb{C}^+ \times (\operatorname{supp}(F^{\mathcal{A}}) \cap \mathbb{Q}^m)$.

Step 2: It is shown in Section S.5 of the Supplementary Material that, for any fixed $z \in \mathbb{C}^+$ and subsequence $\{p_j\}$, $\{\rho_{\tau,p_j}(z, \mathbf{a})\}$ are equicontinuous functions. Since $\rho_{\tau,p_\ell(\mathbf{a}_\ell)}(z, \mathbf{a})$ converges pointwise to $\rho_\tau(z, \mathbf{a})$ on the dense subset $\operatorname{supp}(F^{\mathcal{A}}) \cap \mathbb{Q}^m$ of $\operatorname{supp}(F^{\mathcal{A}})$, the Arzelà–Ascoli theorem (Lemma S.10) implies that $\rho_{\tau,p_\ell(\mathbf{a}_\ell)}(z, \mathbf{a})$ uniformly converges to a limit, a continuous function of $\mathbf{a} \in \operatorname{supp}(F^{\mathcal{A}})$, that coincides with $\rho_\tau(z, \mathbf{a})$ for $\mathbf{a} \in \operatorname{supp}(F^{\mathcal{A}}) \cap \mathbb{Q}^m$. Thus, the limit $\rho_\tau(z, \mathbf{a})$ is now defined on $\mathbb{C}^+ \times \operatorname{supp}(F^{\mathcal{A}})$ and is analytic in $z \in \mathbb{C}^+$. From (5.14), it follows that the limit $\rho_\tau(z, \mathbf{a})$ coincides with $\beta_\tau(z, \mathbf{a})$ for $\mathbf{a} \in \operatorname{supp}(F^{\mathcal{A}})$.

Step 3: It remains to show that $\beta_\tau(z, \mathbf{a})$ is the Stieltjes transform of a measure on the real line with mass $m_\tau(\mathbf{a}) := \int \mathcal{R}_\tau(\mathbf{a}, \mathbf{b})\, dF^{\mathcal{A}}(\mathbf{b})$. This is equivalent to showing that $(m_\tau(\mathbf{a}))^{-1}\beta_\tau(z, \mathbf{a})$ is the Stieltjes transform of a Borel probability measure. The proof relies on the Lemma 5.6, stated below. From the definition of $\tilde{\beta}_{\tau,p}(z,\mathbf{a})$ and the fact that $\Gamma_\tau(\mathbf{a})$ is a positive definite matrix with bounded norm, it follows that $(m_{\tau,p}(\mathbf{a}))^{-1}\tilde{\beta}_{\tau,p}(z,\mathbf{a})$ is the Stieltjes transform of a probability measure $\mu_{p,\mathbf{a}}$ where $m_{\tau,p}(\mathbf{a}) = p^{-1}\operatorname{Tr}(\Gamma_\tau(\mathbf{a}))$. The measure $\mu_{p,\mathbf{a}}$ is such that $\mu_{p,\mathbf{a}}((x,\infty)) \le \|\Gamma_\tau(\mathbf{a})\|(m_{\tau,p}(\mathbf{a}))^{-1}F^{\tilde{\mathbf{C}}_\tau}((x,\infty))$ for all x. Now, by the tightness of the sequence $\{F^{\tilde{\mathbf{C}}_\tau}\}$ (by Lemma 5.6), it follows that $\{\mu_{p,\mathbf{a}}\}$ is a tight sequence of probability measures. Now, by Step 2 and the conclusion in Section 5.3, it follows there is a subsequence $\{p_\ell\}$ such that the Stieltjes transform of $(m_{\tau,p_\ell}(\mathbf{a}))^{-1}\tilde{\beta}_{\tau,p_\ell}(z,\mathbf{a})$ converges almost surely to $(m_\tau(\mathbf{a}))^{-1}\beta_\tau(z,\mathbf{a})$ for each $z \in \mathbb{C}^+$. The conclusion that $(m_\tau(\mathbf{a}))^{-1}\beta_\tau(z,\mathbf{a})$ is the Stieltjes transform of a Borel probability measure then follows from Lemma S.11. □

Lemma 5.6. *Under the conditions of Theorem 2.2, $F^{\mathbf{C}_\tau}$ is a tight sequence.*

It should be noted that Lemma 5.6, together with $s_{\tau,p}(z) \xrightarrow{\text{a.s.}} s_\tau(z)$ for $z \in \mathbb{C}^+$, proves the existence of the LSD of $\mathbf{C}_\tau$. The proof of Lemma 5.6 is given in Section S.3 of the Supplementary Material.

Next, we prove the uniqueness of the solutions $\beta(z,\mathbf{a})$ under the constraint that the solutions belong to the class of Stieltjes kernels that are analytic on $\mathbb{C}^+$ for all $\mathbf{a} \in \operatorname{supp}(F^{\mathcal{A}})$. First, we verify the uniqueness of the solution for $z \in \mathbb{C}^+(v_0) = \{z \in \mathbb{C}^+ : \Im(z) > v_0\}$ for sufficiently large $v_0 > 0$. At the same time, continuity of the solution with respect to $F^{\mathcal{A}}$ is verified. Accordingly, let $\beta_\tau(z,\mathbf{a})$ satisfy (2.9) for any $\mathbf{a} \in \operatorname{supp}(F^{\mathcal{A}})$. In view of establishing the continuous dependence of $\beta_\tau(z,\mathbf{a})$, and hence $s_\tau(z)$, on $F^{\mathcal{A}}$, on $F^{\mathcal{A}}$ and the kernel $\mathcal{R}_\tau$, suppose that there is a possibly different distribution $F^{\bar{\mathcal{A}}}$ and a possibly different kernel $\bar{\mathcal{R}}_\tau$ (but having the same properties as $\mathcal{R}_\tau$) such that $\bar{\beta}_\tau(z,\mathbf{a})$ satisfies

$$\bar{\beta}_\tau(z,\mathbf{a}) = -\int \frac{\bar{\mathcal{R}}_\tau(\mathbf{a},\mathbf{b})\,dF^{\bar{\mathcal{A}}}(\mathbf{b})}{z + \bar{\beta}_\tau(z,\mathbf{b})}, \qquad \mathbf{a} \in \mathbb{R}^{m_0},$$

and is a Stieltjes transform of a measure for all $\mathbf{a} \in \operatorname{supp}(F^{\bar{\mathcal{A}}})$. Note that, by the defining equations and the continuity of $\mathcal{R}_\tau(\mathbf{a},\mathbf{b})$, and $\bar{\mathcal{R}}_\tau(\mathbf{a},\mathbf{b})$, the functions $\beta(z,\mathbf{a})$ and $\bar{\beta}(z,\mathbf{a})$ are continuous in $\mathbf{a}$ for all $z \in \mathbb{C}^+$. Also,

$$\begin{aligned} &\beta_\tau(z,\mathbf{a}) - \bar{\beta}_\tau(z,\mathbf{a}) \\ &\quad= \int \frac{\mathcal{R}_\tau(\mathbf{a},\mathbf{b})(\beta_\tau(z,\mathbf{b}) - \bar{\beta}_\tau(z,\mathbf{b}))\,dF^{\mathcal{A}}(\mathbf{b})}{(z + \beta_\tau(z,\mathbf{b}))(z + \bar{\beta}_\tau(z,\mathbf{b}))} - \int \frac{(\mathcal{R}_\tau(\mathbf{a},\mathbf{b}) - \bar{\mathcal{R}}_\tau(\mathbf{a},\mathbf{b}))\,dF^{\mathcal{A}}(\mathbf{b})}{z + \bar{\beta}_\tau(z,\mathbf{b})} \\ &\qquad - \int \frac{\bar{\mathcal{R}}_\tau(\mathbf{a},\mathbf{b})\,d(F^{\mathcal{A}}(\mathbf{b}) - F^{\bar{\mathcal{A}}})}{z + \bar{\beta}_\tau(z,\mathbf{b})}. \end{aligned} \tag{5.24}$$

Define

$$\|\beta_\tau(z,\cdot) - \bar{\beta}_\tau(z,\cdot)\|_{\mathcal{A}}^2 = \int |\beta(z,\mathbf{a}) - \bar{\beta}_\tau(z,\mathbf{a})|^2\,dF^{\mathcal{A}}(\mathbf{a}). \tag{5.25}$$

Then, by Cauchy–Schwarz inequality,

$$
\begin{aligned}
&\left|\beta_\tau(z,\mathbf{a})-\bar\beta_\tau(z,\mathbf{a})\right|^2\\
&\quad\le 3\left|\int \frac{\mathcal{R}_\tau(\mathbf{a},\mathbf{b})(\beta_\tau(z,\mathbf{b})-\bar\beta_\tau(z,\mathbf{b}))\,dF^{\mathcal{A}}(\mathbf{b})}{(z+\beta_\tau(z,\mathbf{b}))(z+\bar\beta_\tau(z,\mathbf{b}))}\right|^2 + r_\tau^{(1)}(\mathbf{a})+r_\tau^{(2)}(\mathbf{a})\\
&\quad\le 3\left[\int \left|\beta_\tau(z,\mathbf{b})-\bar\beta_\tau(z,\mathbf{b})\right|^2 dF^{\mathcal{A}}(\mathbf{b})\right]\left[\int \frac{\mathcal{R}_\tau^2(\mathbf{a},\mathbf{b})\,dF^{\mathcal{A}}(\mathbf{b})}{|z+\beta_\tau(z,\mathbf{b})|^2|z+\bar\beta_\tau(z,\mathbf{b})|^2}\right]\\
&\qquad + r_\tau^{(1)}(\mathbf{a})+r_\tau^{(2)}(\mathbf{a}),
\end{aligned}
\tag{5.26}
$$

where

$$
r_\tau^{(1)}(\mathbf{a}) = 3\left|\int \frac{(\mathcal{R}_\tau(\mathbf{a},\mathbf{b})-\bar{\mathcal{R}}_\tau(\mathbf{a},\mathbf{b}))\,dF^{\mathcal{A}}(\mathbf{b})}{z+\bar\beta_\tau(z,\mathbf{b})}\right|^2 \le \frac{3}{v^2}\|\mathcal{R}_\tau-\bar{\mathcal{R}}_\tau\|_\infty^2,
$$

where $\|\mathcal{R}_\tau-\bar{\mathcal{R}}_\tau\|_\infty = \sup_{\mathbf{a},\mathbf{b}\in\mathbb{R}^{m_0}}|\mathcal{R}_\tau(\mathbf{a},\mathbf{b})-\bar{\mathcal{R}}_\tau(\mathbf{a},\mathbf{b})|$, and

$$
r_\tau^{(2)}(\mathbf{a}) = 3\left|\int \frac{\bar{\mathcal{R}}_\tau(\mathbf{a},\mathbf{b})\,d(F^{\bar{\mathcal{A}}}-F^{\mathcal{A}})(\mathbf{b})}{z+\bar\beta_\tau(z,\mathbf{b})}\right|^2 \le \frac{6(L_1^4+\|\mathcal{R}_\tau-\bar{\mathcal{R}}_\tau\|_\infty^2)}{v^2}\|F^{\mathcal{A}}-F^{\bar{\mathcal{A}}}\|_{\mathrm{TV}}^2,
$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation distance. Taking $v_0=\max\{1,\sqrt{2}L_1\}$, if follows for $v>v_0$ that

$$
\int \frac{\mathcal{R}_\tau^2(\mathbf{a},\mathbf{b})\,dF^{\mathcal{A}}(\mathbf{b})}{|z+\beta_\tau(z,\mathbf{b})|^2|z+\bar\beta_\tau(z,\mathbf{b})|^2} \le \frac{L_1^4}{v^4} < \frac{1}{4}.
$$

Therefore, by (5.26), for $v>v_0$,

$$
\begin{aligned}
\|\beta_\tau(z,\cdot)-\bar\beta_\tau(z,\cdot)\|_{\mathcal{A}}^2 &\le 4\int \bigl(r_\tau^{(1)}(\mathbf{a})+r_\tau^{(2)}(\mathbf{a})\bigr)\,dF^{\mathcal{A}}(\mathbf{a})\\
&\le \frac{12}{v^2}\bigl(\|\mathcal{R}_\tau-\bar{\mathcal{R}}_\tau\|_\infty^2 + 2\bigl(L_1^4+\|\mathcal{R}_\tau-\bar{\mathcal{R}}_\tau\|_\infty^2\bigr)\|F^{\mathcal{A}}-F^{\bar{\mathcal{A}}}\|_{\mathrm{TV}}^2\bigr).
\end{aligned}
\tag{5.27}
$$

If $F^{\mathcal{A}}=F^{\bar{\mathcal{A}}}$, and $\mathcal{R}_\tau=\bar{\mathcal{R}}_\tau$, (5.27) and the continuity of $\beta_\tau(z,\mathbf{a})$ and $\bar\beta_\tau(z,\mathbf{a})$ in $\mathbf{a}$ imply that $\beta_\tau(z,\mathbf{a})=\bar\beta_\tau(z,\mathbf{a})$ for $z\in\mathbb{C}^+(v_0)$ and $\mathbf{a}\in\mathrm{supp}(F^{\mathcal{A}})$. Then, since both are analytic functions on $\mathbb{C}^+$ for every fixed $\mathbf{a}\in\mathrm{supp}(F^{\mathcal{A}})$, the uniqueness of the solution in $z\in\mathbb{C}^+$ follows. Moreover, (5.27) proves the continuous dependence of the solution $\beta_\tau(z,\cdot)$ on $\mathcal{R}_\tau$ and $F^{\mathcal{A}}$, with respect to the topology of uniform convergence and that of total variation norm, respectively. From this, similar properties for s_τ are easily deduced.

6. Proof of Theorem 2.2

In this section, the results are extended to the setting that q is not fixed, but tends to infinity at certain rate. In fact, $q=O(p^{1/4})$ is an appropriate choice. This rate plays a crucial role in two places of the derivations. First in verifying properties (such as continuity) of the solution and then in transitioning from the Gaussian to the non-Gaussian case. The latter situation requires the $1/4$ power, while the former can be worked out under the weaker assumption that $q = o(p^{1/2})$. It is shown here that the LSD of the truncated process is the same as that of the linear process almost surely. Denote then by

$$\mathbf{S}_\tau^{\mathrm{tr}} = \frac{1}{2n}\left(\sum_{t=\tau+1}^{n} X_t^{\mathrm{tr}} X_{t-\tau}^{\mathrm{tr}*} + \sum_{t=\tau+1}^{n} X_{t-\tau}^{\mathrm{tr}} X_t^{\mathrm{tr}*}\right) \tag{6.1}$$

the symmetrized auto-covariance matrix for the truncated process $X_t^{\mathrm{tr}} = \sum_{\ell=0}^{q} \mathbf{A}_\ell Z_{t-\ell}$, $t \in \mathbb{Z}$. Let $L(F, G)$ denote the Levy distance between distribution function F and G, defined by

$$L(F, G) = \inf\{\varepsilon > 0: F(x-\varepsilon) - \varepsilon \le G(x) \le F(x+\varepsilon) + \varepsilon\}.$$

In view of Lemma S.7, the aim is to show that

$$L^3\big(F^{\mathbf{C}_\tau}, F^{\mathbf{C}_\tau^{\mathrm{tr}}}\big) \le \frac{1}{p}\operatorname{Tr}\big(\mathbf{C}_\tau - \mathbf{C}_\tau^{\mathrm{tr}}\big)^2 \to 0 \qquad \text{a.s.} \tag{6.2}$$

To this end, define $\bar{X}_t = X_t - X_t^{\mathrm{tr}} = \sum_{\ell=q+1}^{\infty} \mathbf{A}_\ell Z_{t-\ell}$ and notice that

$$\begin{aligned}
\mathbf{S}_\tau - \mathbf{S}_\tau^{\mathrm{tr}} &= \frac{1}{2n}\sum_{t=1}^{n-\tau}\big(X_t X_{t+\tau}^* + X_{t+\tau} X_t^*\big) - \frac{1}{2n}\sum_{t=1}^{n-\tau}\big(X_t^{\mathrm{tr}} X_{t+\tau}^{\mathrm{tr}*} + X_{t+\tau}^{\mathrm{tr}} X_t^{\mathrm{tr}*}\big) \\
&= \frac{1}{2n}\sum_{t=1}^{n-\tau}\big(\bar{X}_t X_{t+\tau}^{\mathrm{tr}*} + X_{t+\tau}^{\mathrm{tr}} \bar{X}_t^*\big) + \frac{1}{2n}\sum_{t=1}^{n-\tau}\big(X_t^{\mathrm{tr}} \bar{X}_{t+\tau}^* + \bar{X}_{t+\tau} X_t^{\mathrm{tr}*}\big) \\
&\quad + \frac{1}{2n}\sum_{t=1}^{n-\tau}\big(\bar{X}_t \bar{X}_{t+\tau}^* + \bar{X}_{t+\tau} \bar{X}_t^*\big) \\
&= \mathbf{S}_{\tau,1} + \mathbf{S}_{\tau,2} + \mathbf{S}_{\tau,3}.
\end{aligned}$$

Therefore,

$$\big\|\mathbf{C}_\tau - \mathbf{C}_\tau^{\mathrm{tr}}\big\|_F^2 \le 3\left(\frac{n}{p}\big\|\mathbf{S}_{\tau,1} - \mathbb{E}[\mathbf{S}_{\tau,1}]\big\|_F^2 + \frac{n}{p}\big\|\mathbf{S}_{\tau,2} - \mathbb{E}[\mathbf{S}_{\tau,2}]\big\|_F^2 + \frac{n}{p}\big\|\mathbf{S}_{\tau,3} - \mathbb{E}[\mathbf{S}_{\tau,3}]\big\|_F^2\right). \tag{6.3}$$

Hence, to prove that (6.2) holds, it suffices to show that

$$\sum_{p=1}^{\infty} \frac{n}{p^2}\mathbb{E}\big[\big\|\mathbf{S}_{\tau,i} - \mathbb{E}[\mathbf{S}_{\tau,i}]\big\|_F^2\big] < \infty, \qquad i = 1, 2, 3, \tag{6.4}$$

due to the Borel–Cantelli lemma. The corresponding detailed calculations are performed in Section S.6 of the Supplementary Material.

7. Extension to non-Gaussian settings

In this section, it is shown that Theorem 2.1 extends beyond the Gaussian setting. In order to show this, Lindeberg's replacement strategy as developed in [4] is applied to a process consisting of truncated, centered and rescaled versions of the original innovation entries Z_{tj}. To formally define this transformation, let $\varepsilon_p > 0$ be such that $\varepsilon_p \to 0$, $p^{1/4}\varepsilon_p \to \infty$ and $\mathbb{P}(|Z_{11}| \geq n^{1/4}\varepsilon_p) \leq n^{-1}\varepsilon_p$. The existence of such an ε_p follows from (Z1) and (Z2). Let then $\breve{Z}^c_{tj} = Z^c_{tj} I_{\{|Z^c_{tj}| \leq n^{1/4}\varepsilon_p\}}$ denote the truncated innovations and $\hat{Z}^c_{tj} = (\breve{Z}^c_{tj} - \mathbb{E}[\breve{Z}^c_{tj}])/(2\mathrm{sd}(\breve{Z}^c_{ij}))$ the standardized versions where $c \in \{\mathbf{R}, \mathbf{I}\}$ with the superscripts $\mathbf{R}$ and $\mathbf{I}$ denoting the real and imaginary parts. Let further $\hat{X}_t = \sum_{\ell=0}^q \mathbf{A}_\ell \hat{Z}_{t-\ell}$, $t \in \mathbb{Z}$, and define the autocovariance matrix of $(\hat{X}_t : t \in \mathbb{Z})$ be defined by

$$\hat{\mathbf{C}}_\tau := \sqrt{\frac{n}{p}}\big(\hat{\mathbf{S}}_\tau - \mathbb{E}[\hat{\mathbf{S}}_\tau]\big),$$

where

$$\hat{\mathbf{S}}_\tau = \frac{1}{2(n-\tau)}\left(\sum_{t=\tau+1}^{n} \hat{X}_t \hat{X}^*_{t-\tau} + \sum_{t=\tau+1}^{n} \hat{X}_{t-\tau} \hat{X}^*_t\right). \tag{7.1}$$

The LSD of the auto-covariance matrix of $\mathbf{C}_\tau$ is the same as that of $\hat{\mathbf{C}}_\tau$, since, according to [2] and [8], an application of a rank inequality and Bernstein's inequality implies that

$$\sup_x \big|F^{\mathbf{C}_\tau}(x) - F^{\hat{\mathbf{C}}_\tau}(x)\big| \to 0 \qquad \text{a.s.}$$

For notational simplicity, the truncated, centered and rescaled variables are therefore henceforth still denoted by Z_{jt} (correspondingly, X_{jt}) and it is assumed that they are i.i.d. with $|Z_{11}| \leq n^{1/4}\varepsilon_p$, $\mathbb{E}[Z_{11}] = 0$, $\mathbb{E}[|Z_{11}|^2] = 1$, the real and imaginary parts are independent with equal variance, and $\mathbb{E}[|Z_{11}|^4] = \mu_4$ for some finite constant μ_4.

Consider now the process $(X'_t : t \in \mathbb{Z})$ given by

$$X'_t = \sum_{\ell=0}^{q} \mathbf{A}_\ell W_{t-\ell}, \qquad t \in \mathbb{Z}, \tag{7.2}$$

with the innovations $(W_t : t \in \mathbb{Z})$ consisting of i.i.d. real- or complex-valued (not necessarily Gaussian) entries W_{jt} satisfying:

(T1) $\mathbb{E}[W_{jt}] = 0$, $\mathbb{E}[|W_{jt}|^2] = 1$ and $\mathbb{E}[|W_{jt}|^4] \leq C$ for some finite constant $C > 0$;

(T2) In case of complex-valued innovations, the real and imaginary parts of W_{jt} are independent with $\mathbb{E}[\Re(W_{jt})] = \mathbb{E}[\Im(W_{jt})] = 0$ and $\mathbb{E}[\Re(W_{jt})^2] = \mathbb{E}[\Im(W_{jt})^2] = 1/2$;

(T3) $|W_{jt}| \leq n^{1/4}\varepsilon_p$ with $\varepsilon_p > 0$ such that $\varepsilon_p \to 0$ and $p^{1/4}\varepsilon_p \to \infty$;

(T4) The W_{jt} are independent of the Z_{tj} defined in Theorem 2.1.

It is assumed that the coefficient matrices $(\mathbf{A}_\ell : \ell \in \mathbb{N})$ satisfy conditions (A1)–(A4). Define the lag-τ auto-covariance matrix of $(X'_t : t \in \mathbb{Z})$ by

$$\mathbf{S}'_\tau = \frac{1}{2(n-\tau)} \left(\sum_{t=\tau+1}^{n} X'_t \left(X'_{t-\tau}\right)^* + \sum_{t=\tau+1}^{n} X'_{t-\tau} \left(X'_t\right)^* \right), \tag{7.3}$$

so that the corresponding renormalized lag-τ auto-covariance matrix is given by

$$\mathbf{C}'_\tau = \sqrt{\frac{n}{p}} \left(\mathbf{S}'_\tau - \mathbb{E}[\mathbf{S}'_\tau]\right)$$

and the lag-τ Stieltjes transform by $s'_{\tau,p}(z) = \frac{1}{p} \operatorname{Tr}(\mathbf{C}'_\tau - zI)^{-1}$, $z \in \mathbb{C}^+$. We denote the Stieltjes transform of $\mathbf{C}_\tau$, defined in terms of the bounded (after trunctation and normalization) Z_{jt}'s, by $s_{\tau,p}$. Since we have proved the existence and uniqueness of LSD in the case where Z_{jt}'s are i.i.d. standard Gaussian, it follows that for all $z \in \mathbb{C}^+$, $s_{\tau,p}(z)$ converges a.s. to the Stieltjes transform of the LSD determined by (2.8) and (2.9). Thus, proving that the results hold for non-Gaussian innovations means showing that (i) $s'_{\tau,p}(z) - \mathbb{E}[s'_{\tau,p}(z)] \to 0$ a.s. and (ii) $\mathbb{E}[s_{\tau,p}(z) - s'_{\tau,p}(z)] \to 0$ for all $z \in \mathbb{C}^+$ under (2.3). Since (5.22) has been derived without invoking Gaussianity of the innovations, (i) follows readily. To show that (ii) holds requires an application of the Linderberg principle developed in [4]. This task is equivalent to verifying that the difference

$$\mathbb{E}\left(\frac{1}{p} \operatorname{Tr}(\mathbf{C}_\tau - zI)^{-1}\right) - \mathbb{E}\left(\frac{1}{p} \operatorname{Tr}\left(\mathbf{C}'_\tau - zI\right)^{-1}\right) \tag{7.4}$$

tends to zero. The arguments for (ii) to hold are provided in Section S.7 of the Supplementary Material.

Acknowledgements

The authors thank the Editor, the Associate Editor and two anonymous referees for their insightful comments and suggestions. The research of Wang was partially supported by the NSF China Grant 11371317 and a grant from the Australian Research Council. She gratefully acknowledges the hospitality of the Department of Statistics, University of California, Davis and Department of Mathematics and Statistics, University of Melbourne. Aue's research was partially supported by US NSF Grants DMS-12-09226, DMS-13-05858 and DMS-14-07530, and by a grant from the California Department of Pesticide Regulation. Paul's research was partially supported by US NSF Grants DMR 1035468, DMS-11-06690 and DMS-14-07530.

Supplementary Material

Supplement to "Spectral analysis of sample autocovariance matrices of a class of linear time series in moderately high dimensions" (DOI: 10.3150/16-BEJ807SUPP; .pdf). The online supplement to the paper contains auxiliary technical results needed in order to complete the proofs of the main theorems in this paper.

References

[1] Bai, Z. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2567175

[2] Bai, Z.D. and Yin, Y.Q. (1988). Convergence to the semicircle law. *Ann. Probab.* **16** 863–875. MR0929083

[3] Bhattacharjee, M. and Bose, A. (2015). Matrix polynomial generalizations of the sample variance-covariance matrix when $pn^{-1} \to 0$. Technical report, Indian Statistical Institute, Kolkata.

[4] Chatterjee, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. MR2294976

[5] Chen, B. and Pan, G. (2015). CLT for linear spectral statistics of normalized sample covariance matrices with the dimension much larger than the sample size. *Bernoulli* **21** 1089–1133. MR3338658

[6] Jin, B., Wang, C., Bai, Z.D., Nair, K.K. and Harding, M. (2014). Limiting spectral distribution of a symmetrized auto-cross covariance matrix. *Ann. Appl. Probab.* **24** 1199–1225. MR3199984

[7] Li, Z., Pan, G. and Yao, J. (2015). On singular value distribution of large-dimensional autocovariance matrices. *J. Multivariate Anal.* **137** 119–140. MR3332802

[8] Liu, H., Aue, A. and Paul, D. (2015). On the Marčenko–Pastur law for linear time series. *Ann. Statist.* **43** 675–712. MR3319140

[9] Pan, G. and Gao, J. (2009). Asymptotic theory for sample covariance matrix under cross-sectional dependence. Unpublished manuscript.

[10] Pan, G., Gao, J. and Yang, Y. (2014). Testing independence among a large number of high-dimensional random vectors. *J. Amer. Statist. Assoc.* **109** 600–612. MR3223736

[11] Paul, D. and Aue, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718

[12] Pfaffel, O. and Schlemm, E. (2011). Eigenvalue distribution of large sample covariance matrices of linear processes. *Probab. Math. Statist.* **31** 313–329. MR2853681

[13] Wang, L. (2014). Topics on spectral analysis of covariance matrices and its application. Ph.D. thesis, Zhejiang Univ.

[14] Wang, L., Aue, A. and Paul, D. (2016). Supplement to "Spectral analysis of sample autocovariance matrices of a class of linear time series in moderately high dimensions." DOI:10.3150/16-BEJ807SUPP.

[15] Wang, L. and Paul, D. (2014). Limiting spectral distribution of renormalized separable sample covariance matrices when $p/n \to 0$. *J. Multivariate Anal.* **126** 25–52. MR3173080

[16] Yao, J. (2012). A note on a Marčenko–Pastur type theorem for time series. *Statist. Probab. Lett.* **82** 22–28. MR2863018

Received April 2015 and revised December 2015