

UC Davis

UC Davis Previously Published Works

Title

On the Marčenko–Pastur law for linear time series

Permalink

<https://escholarship.org/uc/item/0894c5bm>

Journal

The Annals of Statistics, 43(2)

ISSN

0090-5364

Authors

Liu, Haoyang
Aue, Alexander
Paul, Debashis

Publication Date

2015

DOI

10.1214/14-aos1294

Peer reviewed

ON THE MARČENKO–PASTUR LAW FOR LINEAR TIME SERIES

BY HAORYANG LIU*, ALEXANDER AUE^{†,1} AND DEBASHIS PAUL^{†,2}

University of California, Berkeley and University of California, Davis[†]*

This paper is concerned with extensions of the classical Marčenko–Pastur law to time series. Specifically, p -dimensional linear processes are considered which are built from innovation vectors with independent, identically distributed (real- or complex-valued) entries possessing zero mean, unit variance and finite fourth moments. The coefficient matrices of the linear process are assumed to be simultaneously diagonalizable. In this setting, the limiting behavior of the empirical spectral distribution of both sample covariance and symmetrized sample autocovariance matrices is determined in the high-dimensional setting $p/n \rightarrow c \in (0, \infty)$ for which dimension p and sample size n diverge to infinity at the same rate. The results extend existing contributions available in the literature for the covariance case and are one of the first of their kind for the autocovariance case.

1. Introduction. One of the exciting developments in statistics during the last decade has been the development of the theory and methodologies for dealing with high-dimensional data. The term *high dimension* is primarily interpreted as meaning that the dimensionality of the observed multivariate data is comparable to the available number of replicates or subjects on which the measurements on the different variables are taken. This is often expressed in the asymptotic framework as $p/n \rightarrow c > 0$, where p denotes the dimension of the observation vectors (forming a triangular array) and n the sample size. Much of this development centered on understanding the behavior of the sample covariance matrix and especially its eigenvalues and eigenvectors, due to their role in dimension reduction, in estimation of population covariances and as building block of numerous inferential procedures for multivariate data. Comprehensive reviews of this topic can be found in Johnstone [22] and Paul and Aue [30].

One most notable high-dimensional phenomena associated with sample covariance matrices is that the sample eigenvalues do not converge to their population counterparts if dimension and sample sizes remain comparable even as the sample size increases. A formal way to express this phenomenon is through the use of the *empirical spectral distribution* (ESD), that is, the empirical distribution of the

Received October 2013; revised November 2014.

¹Supported in part by NSF Grants DMS-12-09226, DMS-13-05858 and DMS-14-07530.

²Supported in part by NSF Grants DMR-10-35468, DMS-11-06690 and DMS-14-07530.

MSC2010 subject classifications. Primary 62H25; secondary 62M10.

Key words and phrases. Autocovariance matrices, empirical spectral distribution, high-dimensional statistics, linear time series, Marčenko–Pastur law, Stieltjes transform.

eigenvalues of the sample covariance matrix. The celebrated work of Marčenko and Pastur [26] shows that if one studies a triangular array of random vectors X_1, \dots, X_n , whose components form independent, identically distributed (i.i.d.) random variables with zero mean, unit variance and finite fourth moment, then as $p, n \rightarrow \infty$ such that $p/n \rightarrow c \in (0, \infty)$, the ESD of $\mathbf{S} = n^{-1} \sum_{j=1}^n X_j X_j^T$ converges almost surely to a nonrandom probability distribution known as the Marčenko–Pastur distribution. Since this highly influential discovery a large body of literature under the banner of random matrix theory (RMT) has been developed to explore the properties of the eigenvalues and eigenvectors of large random matrices. One may refer to Anderson et al. [2], Bai and Silverstein [5] and Tao [34] to study various aspects of this literature.

Many important classes of high-dimensional data, particularly those arising in signal processing, economics and finance, have the feature that in addition to the dimensional correlation, the observations are correlated in time. Classical models for time series often assume a stationary correlation structure and use spectral analysis methods or methods built on the behavior of the sample autocovariance matrices for inference and prediction purposes. In spite of this, to our knowledge, no work exists that analyzes the behavior of the sample autocovariance matrices of a time series from a random matrix perspective, even though Jin et al. [20] have dealt recently covered autocovariance matrices in the independent case. A striking observation is that, in the high-dimensional scenario, the distribution of the eigenvalues of the symmetrized sample autocovariance of a given lag order tends to stabilize to a nondegenerate distribution even in the setting where the observations are i.i.d. This raises questions about the applicability of sample autocovariance matrices as diagnostic tools for determining the nature of temporal dependence in high-dimensional settings. Thus a detailed study of the phenomena associated with the behavior of the ESD of the sample autocovariance matrices when the observations have both dimensional and temporal correlation is of importance to gain a better understanding of the ways in which the dimensionality affects the inference for high-dimensional time series.

All the existing work on high-dimensional time series dealing with the limiting behavior of the ESD focuses on the sample covariance matrix of the data when X_1, \dots, X_n are p -dimensional observations recorded in time and $p, n \rightarrow \infty$ such that $p/n \rightarrow c \in (0, \infty)$. This includes the works of Jin et al. [21], who assume the process $(X_t : t \in \mathbb{Z})$ has i.i.d. rows with each row following a causal ARMA process. Pfaffel and Schlemm [32] and Yao [35] extend this framework to the setting where the rows are arbitrary i.i.d. stationary processes with short-range dependence. Zhang [36], Paul and Silverstein [31] and El Karoui [9], under slightly different assumptions, consider the limiting behavior of the ESD of the sample covariance when the data matrices are of the form $\mathbf{A}^{1/2} \mathbf{Z} \mathbf{B}^{1/2}$ where \mathbf{A} and \mathbf{B} are positive semidefinite matrices, and \mathbf{Z} has i.i.d. entries with zero mean, unit variance and finite fourth moment. This model is known as the separable covariance model, since the covariance of the data matrix is the Kronecker product of \mathbf{A} and \mathbf{B} . If the

rows indicate spatial coordinates and columns indicate time instances, then this model implies that spatial (dimensional) and temporal dependencies in the data are independent of each other. The work of this paper is also partly related to the results of Hachem et al. [16], who prove the existence of the limiting ESD for sample covariance of data matrices that are rectangular slices from a bistationary Gaussian process on \mathbb{Z}^2 .

In this paper, the focus is on a class of time series known as linear processes [or $\text{MA}(\infty)$ processes]. The assumptions to be imposed in Section 2 imply that, up to an unknown rotation, the coordinates of the linear process, say $(X_t : t \in \mathbb{Z})$, are uncorrelated stationary linear processes with short range dependence. Extending the work of Jin et al. [20] to the time series case, the goal is to relate the behavior of the ESD of the lag- τ symmetrized sample autocovariances, defined as $\mathbf{C}_\tau = (2n)^{-1} \sum_{t=1}^{n-\tau} (X_t X_{t+\tau}^* + X_{t+\tau} X_t^*)$, with $*$ denoting complex conjugation, to that of the spectra of the coefficient matrices of the linear process when $p, n \rightarrow \infty$ such that $p/n \rightarrow c \in (0, \infty)$. This requires assuming certain stability conditions on the joint distribution of the eigenvalues of the coefficient matrices which are described later. The class of models under study here includes the class of causal autoregressive moving average (ARMA) processes of finite orders satisfying the requirement that the coefficient matrices are simultaneously diagonalizable and the joint empirical distribution of their eigenvalues (when diagonalized in the common orthogonal or unitary basis), converges to a finite-dimensional distribution. The results are expressed in terms of the *Stieltjes transform* of the ESD of the sample autocovariances. Specifically, it is shown that the ESD of the symmetrized sample autocovariance matrix of any lag order converges to a nonrandom probability distribution on the real line whose Stieltjes transform can be expressed in terms a unique *Stieltjes kernel*. The definition of the Stieltjes kernel involves integration with respect to the limiting joint empirical distribution of the eigenvalues of the coefficient matrices as well as the spectral density functions of the one-dimensional processes that correspond to the coordinates of the process $(X_t : t \in \mathbb{Z})$, after rotation in the common unitary or orthogonal matrix that simultaneously diagonalizes the coefficient matrices. Thus this result neatly ties the dimensional correlation, captured by the eigenvalues of the coefficient matrices, with the temporal correlation, captured by the spectral density of the coordinate processes.

The main contributions of this paper are the following: (i) A framework is provided for analyzing the behavior of symmetrized autocovariance matrices of linear processes; (ii) for linear processes satisfying appropriate regularity conditions, a concrete description of the limiting Stieltjes transform is given in terms of the limiting joint ESD of the coefficient matrices and the spectral density of the coordinate processes after a rotation of the coordinates of the observation. Extensions to these main results are (iii) the characterization of the behavior of the ESD of autocovariances of linear filters applied to the observed process; (iv) the description of the ESDs of a class of tapered estimates of the spectral density operator of the

observed process that can be used to analyze the long-run variance and spectral coherence of the process. These contributions surpass the work in the existing literature dealing with high-dimensionality effects for time series in two different ways. First, the class of time series models that are analyzed in detail encompasses the setting of stationary i.i.d. rows studied by Jin et al. [21], Pfaffel and Schlemm [32] and Yao [35], as well as the setting of separable covariance structure studied by Zhang [36], Paul and Silverstein [31] and El Karoui [9]. The proofs of the main results also require more involved arguments. They are partly related to the constructions in Hachem et al. [16], but additional technical arguments are needed to go beyond Gaussanity. The results are also related to the work of Hachem et al. [17], who studies limiting spectral distributions of covariance matrices for data with a given variance profile. The connection is through the fact that after an approximation of lag operators by circulant shift matrices, and appropriate row and column rotations, the data matrix in our setting can be equivalently expressed as a matrix with independent entries and with a variance profile related to the spectral densities of the different coordinates of the time series. Second, the framework allows for a unified analysis of the ESD of symmetrized autocovariance matrices of all lag orders as well as that of the tapered spectral density operator. None of the existing works deals with the behavior of autocovariances for time series (note again that Jin et al. [20] treat the i.i.d. case), and this analysis requires a nontrivial variation of the arguments used for dealing with the Stieltjes transform of the sample covariance matrix. Moreover, even though we stick to the setting where the coefficient matrices are Hermitian and simultaneously diagonalizable, the main steps in the derivation, especially the construction of a “deterministic equivalent” of the resolvent of the symmetrized autocovariance matrix, is very general and can be applied to linear processes with structures that go beyond the settings studied in this paper, for example, when the simultaneous diagonalizability of the coefficient matrices is replaced by a form of simultaneous block diagonalizability, even though the latter is not pursued in this paper due to lack of clear statistical motivation. The existence and uniqueness of the limits of the resulting equations and their solutions is the key to establishing the existence of limiting ESDs of the autocovariances. This step requires certain regularity conditions on the coefficient matrices and is not pursued beyond the setting described in Section 2. A number of potential applications, for example, to problems in signal processing, and dynamic and static factor models, are discussed in Section 3.

The remaining sections of the paper are organized as follows. Extensions of the main results in Section 2 are discussed in Section 4. The outcomes of a small simulation study are reported in Section 5, while the proofs of the main results are provided in Sections 6–11. Several technical lemmas are collected in the online Supplemental Material (SM) [24].

2. Main results. Let \mathbb{Z} denote the set of integers. A sequence of random vectors $(X_t : t \in \mathbb{Z})$ with values in \mathbb{C}^p is called a linear process or moving average

process of order infinity, abbreviated by the acronym $MA(\infty)$, if it has the representation

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

where $(Z_t : t \in \mathbb{Z})$ denotes a sequence of independent, identically distributed p -dimensional random vectors whose entries are independent and satisfy $\mathbb{E}[Z_{jt}] = 0$, $\mathbb{E}[|Z_{jt}|^2] = 1$ and $\mathbb{E}[|Z_{jt}|^4] < \infty$, where Z_{jt} denotes the j th coordinate of Z_t . In the complex-valued case this is meant as $\mathbb{E}[\text{Re}(Z_{jt})^2] = \mathbb{E}[\text{Im}(Z_{jt})^2] = 1/2$. It is also assumed that real and imaginary parts are independent. Let further $\mathbf{A}_0 = \mathbf{I}$, the identity matrix. To ensure finite fourth moments for $(X_t : t \in \mathbb{Z})$ and a sufficiently fast decaying weak dependence structure, Assumption 2.1 below lists several additional conditions imposed on the coefficient matrices \mathbf{A}_ℓ .

The results presented in this paper are concerned with the behavior of the symmetrized lag- τ sample autocovariances

$$\mathbf{C}_\tau = \frac{1}{2n} \sum_{t=1}^{n-\tau} (X_t X_{t+\tau}^* + X_{t+\tau} X_t^*), \quad \tau \in \mathbb{N}_0,$$

assuming observations for X_1, \dots, X_n are available. For $\tau = 0$, this definition gives the covariance matrix $\mathbf{S} = \mathbf{C}_0$ discussed in the Introduction. Note that in order to make predictions in the linear process setting, it is imperative to understand the second-order dynamics which are captured in the population autocovariance matrices $\mathbf{\Gamma}_\tau = \mathbb{E}[X_{t+\tau} X_t^*]$, $\tau \in \mathbb{N}_0$, as all of the popular prediction algorithms such as the Durbin–Levinson and innovations algorithms are starting from there; see, for example, Lütkepohl [25]. The set-up in (2.1) provides a (strictly) stationary process and consequently the definition of $\mathbf{\Gamma}_\tau$ does not depend on the value of t . The main goal of this paper is to analyze the behavior of the matrices \mathbf{C}_τ , which can be viewed as a special sample counterpart to the corresponding $\mathbf{\Gamma}_\tau$, in the high-dimensional setting for which $p = p(n)$ is a function of the sample size such that

$$(2.2) \quad \lim_{n \rightarrow \infty} \frac{p}{n} = c \in (0, \infty),$$

thereby extending the above mentioned Marčenko–Pastur-type results to more general time series models and to autocovariance matrices. We can weaken requirement (2.2) to “ p/n bounded away from zero and infinity,” in which case, the asymptotic results hold for subsequences (p_k, n_k) satisfying p_k/n_k converging to a positive constant c_k , provided that the structural assumptions on the model continue to hold. Let then \hat{F}_τ denote the empirical spectral distribution (ESD) of \mathbf{C}_τ given by

$$\hat{F}_\tau(\sigma) = \frac{1}{p} \sum_{j=1}^p \mathbb{I}_{\{\sigma_j \leq \sigma\}},$$

where $\sigma_1, \dots, \sigma_p$ are the eigenvalues of \mathbf{C}_τ . The proof techniques for establishing large-sample results about \hat{F}_τ are based on exploiting convergence properties of Stieltjes transforms, which continue to play an important role in verifying theoretical results in RMT; see, for example, Paul and Aue [30] for a recent summary. The Stieltjes transform of a distribution function F on the real line is the function

$$s_F : \mathbb{C}^+ \rightarrow \mathbb{C}^+, \quad z \mapsto s_F(z) = \int \frac{1}{\sigma - z} dF(\sigma),$$

where $\mathbb{C}^+ = \{x + iy : x \in \mathbb{R}, y > 0\}$ denotes the upper complex half plane. It can be shown that s_F is analytic on \mathbb{C}^+ and that the distribution function F can be reconstructed from s_F using an inversion formula; see [30]. In order to make statements about \hat{F}_τ , the following additional assumptions on the coefficient matrices \mathbf{A}_ℓ are needed. Let \mathbb{N} and \mathbb{N}_0 denote the positive and nonnegative integers, respectively.

ASSUMPTION 2.1. (a) The matrices $(\mathbf{A}_\ell : \ell \in \mathbb{N}_0)$ are simultaneously diagonalizable random Hermitian matrices, independent of $(Z_t : t \in \mathbb{Z})$ and satisfying $\|\mathbf{A}_\ell\| \leq \bar{\lambda}_{\mathbf{A}_\ell}$ for all $\ell \in \mathbb{N}_0$ and large p with

$$\sum_{\ell=0}^{\infty} \bar{\lambda}_{\mathbf{A}_\ell} \leq \bar{\lambda}_{\mathbf{A}} < \infty \quad \text{and} \quad \sum_{\ell=0}^{\infty} \ell \bar{\lambda}_{\mathbf{A}_\ell} \leq \bar{\lambda}'_{\mathbf{A}} < \infty.$$

Note that one can set $\bar{\lambda}_{\mathbf{A}_0} = 1$.

(b) There are continuous functions $f_\ell : \mathbb{R}^m \rightarrow \mathbb{R}, \ell \in \mathbb{N}_0$, such that, for every p , there is a set of points $\lambda_1, \dots, \lambda_p \in \mathbb{R}^m$, not necessarily distinct, and a unitary $p \times p$ matrix \mathbf{U} such that

$$\mathbf{U}^* \mathbf{A}_\ell \mathbf{U} = \text{diag}(f_\ell(\lambda_1), \dots, f_\ell(\lambda_p)), \quad \ell \in \mathbb{N},$$

and $f_0(\lambda) = 1$. [Note that the functions f_ℓ are allowed to depend on $p = p(n)$ as long as they converge to continuous functions as $n \rightarrow \infty$ uniformly.]

(c) With probability one, $F_p^{\mathbf{A}}$, the ESD of $\{\lambda_1, \dots, \lambda_p\}$, converges weakly to a nonrandom probability distribution function $F^{\mathbf{A}}$ on \mathbb{R}^m as $p \rightarrow \infty$.

Let $\mathbf{A} = [\mathbf{A}_0 : \mathbf{A}_1 : \dots]$ denote the matrix collecting the coefficient matrices of the linear process $(X_t : t \in \mathbb{Z})$. Define the transfer functions

$$(2.3) \quad \psi(\lambda, \nu) = \sum_{\ell=0}^{\infty} e^{i\ell\nu} f_\ell(\lambda) \quad \text{and} \quad \psi(\mathbf{A}, \nu) = \sum_{\ell=0}^{\infty} e^{i\ell\nu} \mathbf{A}_\ell,$$

as well as the power transfer functions

$$h(\lambda, \nu) = |\psi(\lambda, \nu)|^2 \quad \text{and} \quad \mathcal{H}(\mathbf{A}, \nu) = \psi(\mathbf{A}, \nu)\psi(\mathbf{A}, \nu)^*.$$

Note that the contribution of the temporal dependence of the underlying time series on the asymptotic behavior of \hat{F}_τ is quantified through $h(\lambda, \nu)$. Specifically,

$h(\lambda_j, \nu)$ with λ_j as in part (b) of Assumption 2.1 is (up to normalization) the spectral density of the j th coordinate of the process rotated with the help of the unitary matrix \mathbf{U} . With these definitions, the main results of this paper can be stated as follows.

THEOREM 2.1. *If a complex-valued linear process $(X_t : t \in \mathbb{Z})$ with independent, identically distributed Z_{jt} , $\mathbb{E}[Z_{jt}] = 0$, $\mathbb{E}[\operatorname{Re}(Z_{jt})^2] = \mathbb{E}[\operatorname{Im}(Z_{jt})^2] = 1/2$, $\operatorname{Re}(Z_{jt})$ and $\operatorname{Im}(Z_{jt})$ are independent, and $\mathbb{E}[|Z_{jt}|^4] < \infty$, satisfies Assumption 2.1, then, with probability one and in the high-dimensional setting (2.2), \hat{F}_τ converges to a nonrandom probability distribution F_τ with Stieltjes transform s_τ determined by the equation*

$$(2.4) \quad s_\tau(z) = \int \left[\frac{1}{2\pi} \int_0^{2\pi} \frac{\cos(\tau\nu)h(\lambda, \nu)}{1 + c \cos(\tau\nu)K_\tau(z, \nu)} d\nu - z \right]^{-1} dF^\mathbf{A}(\lambda),$$

where $K_\tau : \mathbb{C}^+ \times [0, 2\pi] \rightarrow \mathbb{C}^+$ is a Stieltjes kernel; that is, $K_\tau(\cdot, \nu)$ is the Stieltjes transform of a measure with total mass $m_\nu = \int h(\lambda, \nu) dF^\mathbf{A}(\lambda)$ for every fixed $\nu \in [0, 2\pi]$, whenever $m_\nu > 0$. Moreover, K_τ is the unique solution of

$$(2.5) \quad \begin{aligned} &K_\tau(z, \nu) \\ &= \int \left[\frac{1}{2\pi} \int_0^{2\pi} \frac{\cos(\tau\nu')h(\lambda, \nu')}{1 + c \cos(\tau\nu')K_\tau(z, \nu')} d\nu' - z \right]^{-1} h(\lambda, \nu) dF^\mathbf{A}(\lambda), \end{aligned}$$

subject to the restriction that K_τ is a Stieltjes kernel. Otherwise, if $m_\nu = 0$, then $K_\tau(z, \nu)$ is identically zero on \mathbb{C}^+ and so still satisfies (2.5).

THEOREM 2.2. *If a real-valued linear process $(X_t : t \in \mathbb{Z})$ with independent, identically distributed real-valued Z_{jt} , $\mathbb{E}[Z_{jt}] = 0$, $\mathbb{E}[Z_{jt}^2] = 1$ and $\mathbb{E}[Z_{jt}^4] < \infty$, satisfies Assumption 2.1 with real symmetric coefficient matrices $(\mathbf{A}_\ell : \ell \in \mathbb{N}_0)$, then the result of Theorem 2.1 is retained.*

REMARK 2.1. When each coefficient matrix \mathbf{A}_ℓ is a multiple of the identity matrix, that is, $\mathbf{A}_\ell = \alpha_\ell I_p$ where $(\alpha_\ell : \ell \in \mathbb{N})$ is a sequence of real numbers satisfying $\sum_{\ell=1}^\infty \ell |\alpha_\ell| < \infty$, the result of Theorem 2.1 reduces to the results obtained in Pfaffel and Schlemm [32] and Yao [35].

REMARK 2.2. One can relax the assumption of simultaneous diagonalizability of the coefficient matrices of the linear process to certain forms of near-simultaneous diagonalizability, so that the conclusions of Theorem 2.1 continue to hold for linear processes where the MA coefficients are Toeplitz matrices whose entries decay away from the diagonal at an appropriate rate. Specifically, if \mathbf{A}_ℓ is the Toeplitz matrix with j th row equaling $(a_{k-j, \ell} : 0 \leq k < p)$, for the bi-infinite sequence $(a_{k, \ell})$ satisfying the condition

$$\sum_{\ell=0}^\infty \sum_{k=-\infty}^\infty |k|^\beta (\ell + 1) |a_{k, \ell}| < \infty$$

for some $\beta > 0$, which in particular implies Assumption 2.1(a) by the Gershgorin theorem, then the existence of the limiting ESD of symmetrized autocovariance matrices can be proved. For brevity, instead of giving a thorough technical argument, we only provide the main idea of proof. First, the $\text{MA}(\infty)$ series is approximated by an $\text{MA}(q_p)$ series with $q_p = O(p^{1/3})$, using arguments along the line of Section 6.3. Second, banding with bandwidth k_p is applied to the coefficient matrices \mathbf{A}_ℓ . It can be shown through an application of norm inequality, that the limiting spectral behavior is unchanged under the banding so long as $k_p \rightarrow \infty$ under (2.2). Third, circulant matrices are constructed from the banded Toeplitz matrices by periodization. The resulting matrices are therefore simultaneously diagonalizable, and the eigenvalues of the ℓ th approximate coefficient matrix approximate the transfer function of the sequence $(a_{k,\ell} : k \in \mathbb{Z})$. The limiting spectral behavior is seen to be unchanged after the use of the rank inequality so long as $k_p/p \rightarrow 0$ under (2.2). The rest of the derivations follow the arguments in the proof of Theorem 2.1. While this particular result is related to the work of Hachem et al. [16], who study the convergence of the empirical distribution of the sample covariance matrix of rectangular slices of bstationary Gaussian random fields, [16] does not cover the transition to non-Gaussian processes or the spectral behavior of sample autocovariance matrices.

Several extensions of Theorem 2.1 are discussed in Section 4 below. The proof steps needed in order to verify the main result are outlined in Section 6, and the details are in Sections 7–11. The online SM [24] contains additional technical lemmas.

3. Examples and applications.

3.1. *An ARMA(1, 1) example.* In this section, let $(X_t : t \in \mathbb{Z})$ be the causal ARMA(1, 1) process given by the stochastic difference equations

$$\Phi(L)X_t = \Theta(L)Z_t, \quad t \in \mathbb{Z},$$

where $\Phi(L) = I - \Phi_1 L$ and $\Theta(L) = I + \Theta_1 L$ are, respectively, the matrix-valued autoregressive and moving average polynomials in the lag operator L for which it is assumed that $\|\Phi_1\| \leq \bar{\phi} < 1$ and $\|\Theta_1\| \leq \bar{\theta} < \infty$. Moreover, $(Z_t : t \in \mathbb{Z}) \sim \text{IID}(0, I)$ with entries possessing finite fourth moments. Under these conditions $(X_t : t \in \mathbb{Z})$ admits the $\text{MA}(\infty)$ representation

$$X_t = \mathbf{A}(L)Z_t, \quad t \in \mathbb{Z},$$

with $\mathbf{A}(L) = \sum_{\ell=0}^{\infty} \mathbf{A}_\ell L^\ell = \Phi^{-1}(L)\Theta(L)$. Assume now further that Φ_1 and Θ_1 are simultaneously diagonalizable. Then $\Phi_1 = \mathbf{U}\mathbf{\Lambda}_\Phi\mathbf{U}^*$ and $\Theta_1 = \mathbf{U}\mathbf{\Lambda}_\Theta\mathbf{U}^*$, where $\mathbf{\Lambda}_\Phi = \text{diag}(\phi_1, \dots, \phi_p)$ and $\mathbf{\Lambda}_\Theta = \text{diag}(\theta_1, \dots, \theta_p)$ such that $\max_j |\phi_j| \leq \bar{\phi}$ and $\max_j |\theta_j| \leq \bar{\theta}$. With regard to Assumption 2.1, let $\lambda_j = (\phi_j, \theta_j)' \in \mathbb{R}^2$. Part (c)

of the assumption then requires almost sure weak convergence of the ESD of $\{\lambda_1, \dots, \lambda_p\}$ to a nonrandom probability distribution function on \mathbb{R}^2 . Moreover, using that for each coordinate,

$$\frac{1 + \theta_j L}{1 - \phi_j L} = (1 + \theta_j L) \sum_{\ell=0}^{\infty} (\phi_j L)^\ell = 1 + (\theta_j + \phi_j) \sum_{\ell=1}^{\infty} \phi_j^{\ell-1} L^\ell,$$

it follows that $\mathbf{A}_\ell = \mathbf{U} \text{diag}(f_\ell(\lambda_1), \dots, f_\ell(\lambda_p)) \mathbf{U}^*$ with $f_0(\lambda_j) = 1$ and $f_\ell(\lambda_j) = (\theta_j + \phi_j) \phi_j^{\ell-1}$ for $\ell \in \mathbb{N}$. This illustrates part (b) of Assumption 2.1. The summability conditions stated in part (a) are clearly satisfied. Generalization to arbitrary causal ARMA models follows in a similar fashion.

3.2. *Time series with independent rows.* In this section, the situation of time series with independent rows is considered. Our results describe the limiting ESD of the symmetrized sample autocovariances in the setting where the j th row of the time series, denoted $\{\xi_{jt} : t \in \mathbb{Z}\}$, is given by

$$(3.1) \quad \xi_{jt} = \sum_{\ell=0}^{\infty} f_\ell(\lambda_j) Z_{j,t-\ell},$$

where: (i) the Z_{jt} 's are independent, identically distributed real- or complex-valued random variables with mean zero, unit variance and finite fourth moments; (ii) the f_ℓ 's are continuous functions from $\mathbb{R}^m \rightarrow \mathbb{R}$ satisfying $f_0(\lambda) \equiv 1$ and the summability condition $\sup_{\lambda \in \text{supp}(F^A)} \sum_{\ell=0}^{\infty} (\ell + 1) |f_\ell(\lambda)| < \infty$; (iii) the λ_j 's are i.i.d. realizations from an m -dimensional probability distribution denoted by F^A . If the supremum in condition (ii) is taken over \mathbb{R}^m , condition (iii) can be weakened to require that the empirical distribution of the λ_j 's converges almost surely to a nonrandom distribution F^A .

Let $h(\lambda, \nu) = |\sum_{\ell=0}^{\infty} e^{i\ell\nu} f_\ell(\lambda)|^2$. Then the empirical distributions of the eigenvalues of the lag- τ symmetrized autocovariance matrices converge almost surely to a nonrandom probability distribution F_τ with Stieltjes transform s_τ determined by equations (2.4) and (2.5), where $K_\tau(z, \nu)$ is as in Theorem 2.1. This is in the spirit of the works of Jin et al. [21], Pfaffel and Schlemm [32] and Yao [35], who studied the sample covariance case with $f_\ell(\lambda) \equiv \tilde{f}_\ell$ for all $\ell \geq 0$, Hachem et al. [16], who considered the sample covariance case for stationary Gaussian fields and Jin et al. [20], who studied the symmetrized sample autocovariance case with $f_1(\lambda) \equiv 1$ and $f_\ell(\lambda) \equiv 0$ for $\ell \geq 1$ (i.e., when the ξ_{jt} 's are i.i.d. with zero mean and unit variance).

3.3. *Signal processing and diagnostic checks.* The results derived here can be useful in dealing with a number of important statistical questions. Signal detection in a noisy background is one of the most important problems in signal processing

and communications theory. Often the observations are taken in time, and the standard assumption is that the noise is i.i.d. in time, referred to as white noise. However, in spatio-temporal signal processing, it is quite apt to formulate the noise as “colored” or correlated in time, as well as in the spatial dimension. The proposed model for the time series is a good prototype for such a noise structure. Thus the problem of detecting a low-dimensional signal embedded in high-dimensional noise, for example, through a factor model framework, can be effectively addressed by making use of the behavior of the ESDs of autocovariances of the noise. Another potential application of the results is in building diagnostic tools for high-dimensional time series. By focusing on the ESDs of the autocovariances for various lag orders, or that of a tapered estimate of the spectral density operator, one can infer about the nature of dependence, provided the model assumptions hold. The proposed model also provides a broad class of alternatives for the hypothesis of independence of observations in settings where those observations are measured in time. Finally, in practical applications, it is of interest whether the spectrum of the coefficient matrices of the linear process can be estimated from the data. The equations for the limiting Stieltjes kernel and its relation to the Stieltjes transform of the autocovariance matrices provide a tool for attacking this problem. This aspect has been explored in the Ph.D. thesis of the first author [23] and the methodology will be reported elsewhere.

3.4. Dynamic factor models. Forni and Lippi [13] describe a class of time series models that captures the subject specific variations of microeconomic activities. This class of models, referred to as Dynamic Factor Models (DFM), has proved immensely popular in the econometrics community and beyond. DFMs have, for example, been used for describing the stock returns in [28], forecasting national accounts in [3], modeling portfolio allocation in [1] and modeling psychological development in [27], as well as in many other applications. Important theoretical and inferential questions regarding DFMs have been investigated in a series of papers by Forni and Lippi [14], Forni et al. [10–12] and Stock and Watson [33], to name a few. DFMs have also shown early promise for applications to other interesting multivariate time series problems such as the study of fMRI data.

A DFM can be described as follows. As in [13], let Y_{jt} be the response corresponding to the j th individual/agent at time t , modeled as

$$(3.2) \quad Y_{jt} = b_{j1}(L)U_{1t} + \cdots + b_{jM}(L)U_{Mt} + \xi_{jt}, \quad j = 1, \dots, p.$$

The model specifies that Y_{jt} is determined by a small, fixed number of underlying common factors U_{kt} and their lags, determined by the polynomials $b_{jk}(L)$ in the lag-operator L , plus an idiosyncratic component ξ_{jt} assumed independent across individuals. Typically, $(\xi_{jt} : t \in \mathbb{Z})$ is taken to be a stationary linear processes, independent across j .

One of the key questions pertaining to DFM is the determination of the number of dynamic factors. This question has been investigated by Bai and Ng [4], Stock

and Watson [33] and Hallin and Liška [18]. Unlike in PCA, here one has to deal with the additional problem of detecting the lag orders of the dynamic factors. This can be approached through the study of the behavior of the extreme eigenvalues of the sample autocovariance matrices as in Jin et al. [20]. The issue becomes even more challenging when the dimensionality of the problem increases. In such settings, one expects that a form of phase transition phenomenon, well known in the context of a high-dimensional static factor model (or spiked covariance model) with i.i.d. observations (see, e.g., Baik and Silverstein [6]), will set in. In particular, as Jin et al. [20] argue, a dynamic factor will be detectable from the data only if the corresponding total signal intensity, as measured, for example, by the sum of the variances of the factor loadings, is above a threshold. Moreover, the number of eigenvalues that lie outside the bulk of the eigenvalues of the symmetrized sample autocovariance of a certain lag order provide information about the lag order of the DFM. Driven by the analogy with the static factor model with i.i.d. observations, it is expected that the detection thresholds will depend on the dimension-to-sample size ratio, as well as the behavior of the bulk spectrum of the autocovariances of the idiosyncratic terms at specific lag orders, including the support of the limiting ESD. Equation (3.1) in Section 3.2 constitutes the “null” model for the DFM in which the dynamic common factors are absent. Therefore follow-up studies on the different aspects of the ESD of the symmetrized sample autocovariances of such processes will be helpful in determining the detection thresholds and estimation characteristics of high-dimensional DFMs.

3.5. *An idealized production model.* Onatski [29] describes a model for production Y_t , at time t , involving p different industries in an economy that is given by the equations

$$(3.3) \quad Y_t = \mathbf{W}^{-1} \left(\sum_{k=1}^M u_{kt} \mathbf{f}_k + \xi_t \right).$$

Model (3.3) is a static factor model in which $p \times 1$ vectors \mathbf{f}_k denote the (unobserved) common static factors, u_{kt} denote the (unobserved) factor scores consisting of independent time series corresponding to different factors k and ξ_t denote the $p \times 1$ vectors of idiosyncratic components. The entries of the matrix \mathbf{W} indicate the interactions among the different industries. In the following an enhanced version of the model is considered where the economy is thought to be divided into a finite number of distinct sectors for which the interaction across the sectors is assumed “weak” in a suitable sense to be described. In addition, the assumption of separable covariance structure of the ξ_t made in [29] is relaxed by requiring instead that the temporal variation in ξ_t for all the industrial units within a sector is the same and is stationary in time. This assumption means that the component of the vector ξ_t corresponding to a particular sector has a separable covariance structure with stationary time variation, and the components corresponding to different

sectors are independent. Specifically, if there are K sectors, we can divide \mathbf{W} into $K \times K$ block matrices

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} & \cdots & \mathbf{W}_{1K} \\ \cdot & \cdot & \cdots & \cdot \\ \mathbf{W}_{K1} & \mathbf{W}_{K2} & \cdots & \mathbf{W}_{KK} \end{bmatrix} = \widetilde{\mathbf{W}} + \Delta,$$

where $\widetilde{\mathbf{W}} = \text{diag}(\mathbf{W}_{11}, \dots, \mathbf{W}_{KK})$. If the sectors have no interaction at all, that is, if $\mathbf{W}_{jk} = 0$ for all $j \neq k$, then the corresponding data model is an instance of a blockwise separable covariance model. ‘‘Weak interaction’’ means that the norms of the off-diagonal blocks in the matrix \mathbf{W} are small. More precisely, if $p^{-1} \|\Delta\|_F^2 \rightarrow 0$ as $p \rightarrow \infty$, then the limiting ESD of the symmetrized autocovariances for the data matrix \mathbf{Y} is the same as that of $\widetilde{\mathbf{Y}}$ obtained by replacing \mathbf{W} by $\widetilde{\mathbf{W}}$ in (3.3). Under the assumption of a linear process structure on the different components of ξ_t , and a natural requirement on the stability of the singular values of $\widetilde{\mathbf{W}}$, the existence and characterization of the limiting ESDs of the symmetrized autocovariances of $\widetilde{\mathbf{Y}}$ can be dealt within the framework studied in Section 2. These limiting ESDs will help in determining the detection thresholds for the static factors, or even dynamic factors, if the model were to be enhanced further.

4. Extensions of the main results. This section discusses three different extensions of the main result. The arguments for the proof are similar to that of the proof of Theorem 2.1 and hence only a brief outline is provided. Moreover, the results stated here apply to both real- and complex-valued cases, the only difference being that, in the former case, the relevant matrices are real symmetric while, in the latter case, they are Hermitian.

The first extension involves a rescaling of the process defined in (2.1). Thus it is assumed that

$$(4.1) \quad X_t = \mathbf{B}^{1/2} \sum_{\ell=0}^{\infty} \mathbf{A}_\ell Z_{t-\ell}, \quad t \in \mathbb{Z},$$

where the processes $(Z_t : t \in \mathbb{Z})$ and matrices $(A_\ell : \ell \in \mathbb{N}_0)$ satisfy Assumption 2.1, and the matrix $\mathbf{B}^{1/2}$ is the square root of a $p \times p$ positive semidefinite Hermitian (or symmetric) matrix \mathbf{B} satisfying the following assumption.

ASSUMPTION 4.1. Let \mathbf{U} be as in Assumption 2.1. Then

$$\mathbf{U}^* \mathbf{B} \mathbf{U} = \text{diag}(g_{\mathbf{B}}(\lambda_1), \dots, g_{\mathbf{B}}(\lambda_p)),$$

where $g_{\mathbf{B}} : \mathbb{R}^m \rightarrow \mathbb{R}_+$ is continuous and bounded on \mathbb{R}^m , and $\{\lambda : g_{\mathbf{B}}(\lambda) > 0\} \cap \text{supp}(F^{\mathbf{A}})$ is nonempty.

As before, \widehat{F}_τ is defined to be the ESD of the symmetrized autocovariance \mathbf{C}_τ of lag order τ . If the linear process $(X_t : t \in \mathbb{Z})$ defined through (4.1) satisfies Assumptions 2.1 and 4.1, then the statement of Theorem 2.1 holds with the function $h(\lambda, \nu)$ replaced by $g_{\mathbf{B}}(\lambda)h(\lambda, \nu)$.

The second extension is about the existence and description of the limiting ESD of the autocovariances of linear filters of the process $(X_t : t \in \mathbb{Z})$ defined through (4.1). A linear filter of this process is of the form

$$(4.2) \quad Y_t = \sum_{k=0}^{\infty} b_k X_{t-k}, \quad t \in \mathbb{Z},$$

where $(b_k : k \in \mathbb{N}_0)$ is a sequence of real numbers for which the following summability condition is needed.

ASSUMPTION 4.2. The sequence $(b_k : k \geq 0)$ satisfies $\sum_{k=0}^{\infty} k|b_k| < \infty$.

If the linear process $(X_t : t \in \mathbb{Z})$ defined through (4.1) satisfies Assumptions 2.1 and 4.1, then the statement of Theorem 2.1 holds with the function $h(\lambda, \nu)$ replaced by $\zeta(\nu)g_{\mathbf{B}}(\lambda)h(\lambda, \nu)$, where $\zeta(\nu) = |\sum_{k=0}^{\infty} e^{ik\nu} b_k|^2$, $\nu \in [0, 2\pi]$. This result follows using the properties of convolution and Fourier transform.

The third extension is about estimation of the spectral density operator

$$(4.3) \quad \Gamma(\eta) = \sum_{\tau=-\infty}^{\infty} e^{i\tau\eta} \mathbb{E}[X_t X_{t+\tau}^*], \quad \eta \in [0, 2\pi].$$

It is well known from classical multivariate time series analysis (see, e.g., Chapter 10 of Hamilton [19]) that the “natural” estimator that replaces the population autocovariance $\mathbb{E}[X_t X_{t+\tau}^*]$ by the corresponding sample autocovariance may not be positive definite. In order to obtain positive definite estimators and a better bias-variance trade-off, it is therefore standard in the literature to consider certain tapered estimators with standard choices given, for example, by the Bartlett and Parzen kernels as described in [19]. In the following, the behavior of a class of tapered estimators of $\Gamma(\eta)$, which are given by

$$(4.4) \quad \widehat{\Gamma}_n(\eta) = \sum_{\tau=-\infty}^{\infty} T_n(\tau) e^{i\tau\eta} \frac{1}{n} \sum_{t=1}^{n-\tau} X_t X_{t+\tau}^*, \quad \eta \in [0, 2\pi],$$

is studied, where $T_n(\cdot)$ is a sequence of even functions and the quantities $T_n(\tau)$ are known as tapering weights for which the following restriction is imposed.

ASSUMPTION 4.3. (i) The even functions $T_n(\cdot)$ are such that $T_n(\tau) = 0$ for $\tau \geq n$; (ii) there exists an even function $T(x)$ such that $T_n(x) \rightarrow T(x)$ as $n \rightarrow \infty$ and $|T_n(x)| \leq C|T(x)|$ for some $C > 0$, for all x ; (iii) $\sum_{\tau=0}^{\infty} (1 + |\tau|)|T(\tau)| < \infty$.

An implication of this assumption is that the function $f_T(\eta)$ defined by

$$(4.5) \quad f_T(\eta) = 1 + 2 \sum_{\tau=1}^{\infty} T(\tau) \cos(\tau\eta)$$

is well defined and is uniformly Lipschitz, and $f_{T_n}(\eta) = 1 + \sum_{\tau=1}^{\infty} T_n(\tau) \cos(\tau \eta)$ converges to $f_T(\eta)$ uniformly in η . Examples of kernels T_n are $T_n(x) = (1 + |x|)^{-\alpha} \mathbb{I}(|x| < n)$ for $\alpha > 2$ and $T_n(x) = \beta^{|x|} \mathbb{I}(|x| < n)$ for $\beta \in (0, 1)$. It can be seen from Assumption 4.3 that in the high-dimensional setting under consideration here, standard choices for tapering weights, such as those given by the Bartlett and Parzen kernels, are ruled out. Now the following generalization of Theorem 2.1 is obtained.

THEOREM 4.1. *Suppose that the linear process $(X_t : t \in \mathbb{Z})$ defined through (4.1) satisfies Assumptions 2.1 and 4.1, and that the estimated spectral density operators $(\widehat{\Gamma}_n(\eta) : \eta \in [0, 2\pi])$ are defined by (4.4) with tapering weights satisfying Assumption 4.3. Then, with probability one and in the high-dimensional setting (2.2), for every $\eta \in [0, 2\pi]$, the ESD of $\widehat{\Gamma}_n(\eta)$ converges weakly to a probability distribution $F_{T,\eta}$ with Stieltjes transform $s_{T,\eta}$ determined by the equation*

$$s_{T,\eta}(z) = \int \left[\frac{1}{2\pi} \int_0^{2\pi} \frac{f_T(\eta - \nu) g_{\mathbf{B}}(\lambda) h(\lambda, \nu)}{1 + c f_T(\eta - \nu) K_{T,\eta}(z, \nu)} d\nu - z \right]^{-1} dF^{\mathbf{A}}(\lambda),$$

where $K_{T,\eta} : \mathbb{C}^+ \times [0, 2\pi] \rightarrow \mathbb{C}^+$ is a Stieltjes kernel; that is, $K_{T,\eta}(\cdot, \nu)$ is the Stieltjes transform of a measure with total mass $m_\nu = \int g_{\mathbf{B}}(\lambda) h(\lambda, \nu) dF^{\mathbf{A}}(\lambda)$ for every fixed $\nu \in [0, 2\pi]$, whenever $m_\nu > 0$. Moreover, $K_{T,\eta}$ is the unique solution to

$$\begin{aligned} &K_{T,\eta}(z, \nu) \\ &= \int \left[\frac{1}{2\pi} \int_0^{2\pi} \frac{f_T(\eta - \nu') g_{\mathbf{B}}(\lambda) h(\lambda, \nu')}{1 + c f_T(\eta - \nu') K_{T,\eta}(z, \nu')} d\nu' - z \right]^{-1} \\ &\quad \times g_{\mathbf{B}}(\lambda) h(\lambda, \nu) dF^{\mathbf{A}}(\lambda), \end{aligned}$$

subject to the restriction that $K_{T,\eta}$ is a Stieltjes kernel. Else, if $m_\nu = 0$, $K_{T,\eta}(z, \nu)$ is identically zero on \mathbb{C}^+ and so still satisfies the latter equation.

Section 6.5 outlines the main argument needed to prove Theorem 4.1.

5. Simulations. In this section, a small simulation study is conducted to illustrate the behavior of the LSD (limiting ESD) of the symmetrized sample autocovariances \mathbf{C}_τ for different lag orders τ when the observations are i.i.d. ($X_t = Z_t$) versus when they come from an MA(1) process ($X_t = Z_t + \mathbf{A}_1 Z_{t-1}$) with a symmetric coefficient matrix \mathbf{A}_1 . In this study, two different sets of (n, p) are chosen such that $p/n = 0.5$ and $p/n = 1$, respectively, for $p = 20$ and 50. Also, for the MA(1) case, the ESD of \mathbf{A}_1 is chosen to be $0.5\delta_{0.2} + 0.5\delta_{0.8}$, where δ_b denotes the degenerate distribution at b . The distribution of the Z_{j_t} 's is chosen to be i.i.d. $N(0, 1)$. In Figure 1, the ESD of \mathbf{C}_τ is plotted for the lags $\tau = 0, 1, 2$, for one random realization corresponding to each setting. The theoretical limits

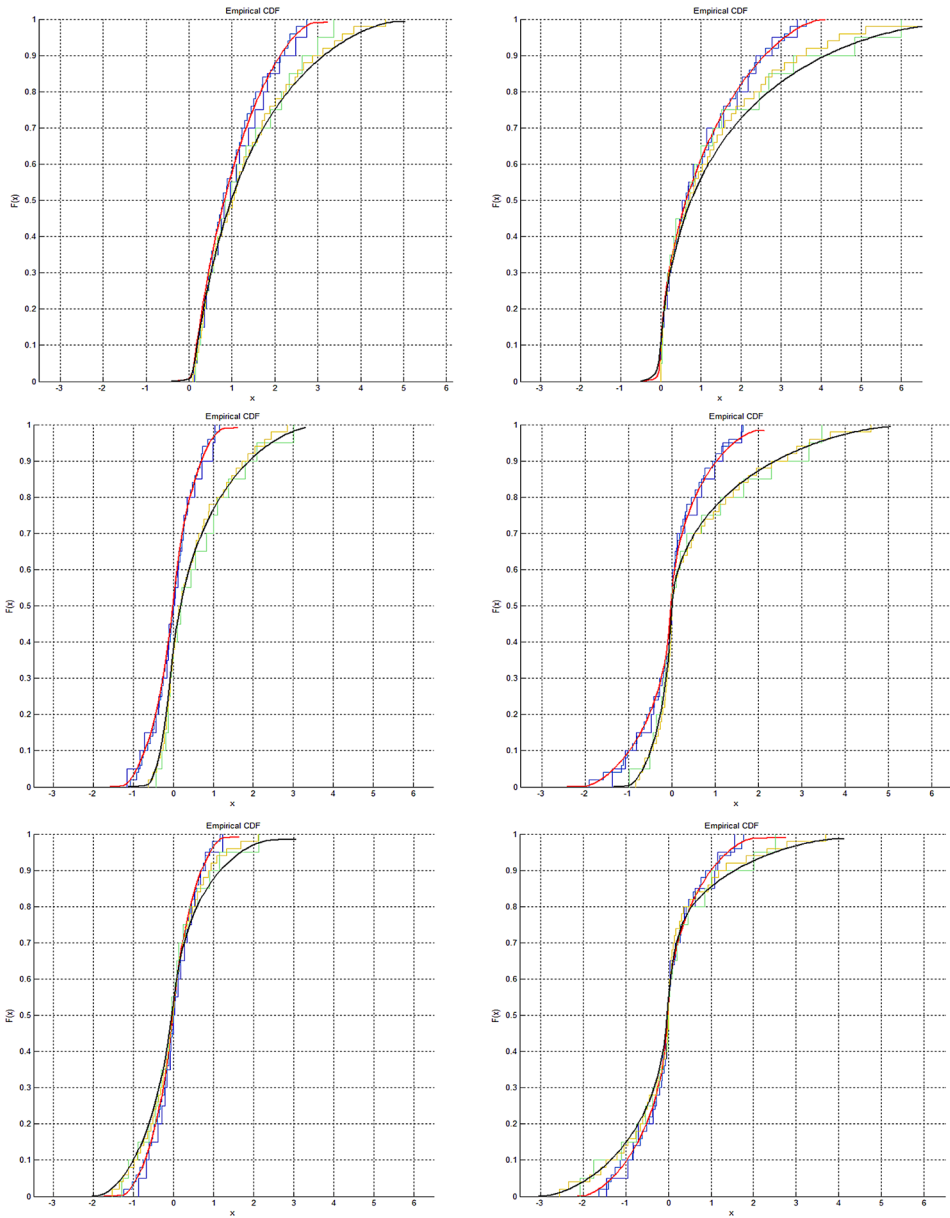


FIG. 1. ESDs and LSDs of the sample autocovariances for i.i.d. and MA(1) observations. Left panel: $p/n = 0.5$; right panel: $p/n = 1$. Top panel: $\tau = 0$; middle panel: $\tau = 1$; bottom panel: $\tau = 2$. Red curve: LSD for i.i.d.; black curve: LSD for MA(1). ESDs are corresponding to $p = 20$ [deep blue for i.i.d., light green for MA(1)] and $p = 50$ [light blue for i.i.d., orange for MA(1)].

(c.d.f.) for all cases are plotted as solid smooth curves [red for i.i.d. and black for MA(1)]. These c.d.f.'s are obtained through numerically inverting the Stieltjes transform s_τ of the LSD of \mathbf{C}_τ using the inversion formula for Stieltjes transforms; cf. [5], [30]. Here s_τ is obtained from equations (2.4) and (2.5) where $c = p/n$ and $h(\lambda, \nu) = 1 + 2 \cos(\nu)\lambda + \lambda^2$. Given $F^{\mathbf{A}} = 0.5\delta_{0,2} + 0.5\delta_{0,8}$, the Stieltjes kernel $K_\tau(z, \nu)$ for the MA(1) case is solved numerically by using calculus of residues. The computational algorithm can be found in the first author's Ph.D. thesis [23].

The graphs clearly show the distinction in behavior of ESD of symmetrized sample autocovariances between the i.i.d. and the MA(1) processes. The LSDs of \mathbf{C}_τ for the i.i.d. case for $\tau = 1$ and 2 are the same, which follows immediately from equations (2.4) and (2.5) that reduce to a single equation since the i.i.d. case corresponds to $F^{\mathbf{A}} = \delta_0$. The behavior of the LSDs in the MA(1) case is distinctly different. It can be shown that LSDs of \mathbf{C}_τ in the MA(1) case converge to the LSD for \mathbf{C}_1 in the i.i.d. case as τ becomes larger. Owing to space constraints, the graphical displays for higher lags are omitted. Another important feature is that the LSDs approximate the ESDs quite well even for p as small as 20, indicating a fast convergence.

6. The structure of the proof.

6.1. Developing intuition for the Gaussian MA(1) case.

OUTLINE. In this section, the overall proof strategy is briefly outlined, and the intuition behind the individual steps is developed for simpler first-order moving average, MA(1), time series. The key ideas in the proof of Theorem 2.1 consist of showing that:

- there is a unique Stieltjes kernel solution K_τ to equation (2.5);
- almost surely, the Stieltjes transform of \hat{F}_τ , say, $s_{p,\tau}$ converges pointwise to a Stieltjes transform s_τ which will be identified with F_τ ;
- \hat{F}_τ is tight.

To achieve the second item, one can argue as follows. First, replace the original linear process observations X_1, \dots, X_n with transformed vectors $\tilde{X}_1, \dots, \tilde{X}_n$ that are serially independent. Second, replace the symmetrized lag- τ autocovariance matrix \mathbf{C}_τ by a transformed version $\tilde{\mathbf{C}}_\tau$ built from $\tilde{X}_1, \dots, \tilde{X}_n$. A heuristic formulation for the simpler Gaussian MA(1) case is given below in some detail. Once these two steps have been achieved, the proof proceeds by verifying some technical conditions with the help of classical RMT results, available, for example, in the monograph Bai and Silverstein [5]. In the following, let $(X_t : t \in \mathbb{Z})$ denote the p -dimensional MA(1) process given by the equations

$$(6.1) \quad X_t = Z_t + \mathbf{A}_1 Z_{t-1}, \quad t \in \mathbb{Z},$$

where $(Z_t : t \in \mathbb{Z})$ are assumed complex Gaussian in addition to the requirements of Section 2. For this time series, the conditions imposed through Assumption 2.1 simplify considerably with part (a) reducing to the condition that the eigenvalues of \mathbf{A}_1 be uniformly bounded and part (b) being satisfied by choosing f_1 as the identity. Moreover, $\mathbf{A} = [\mathbf{I} : \mathbf{A}_1]$, $\psi(\lambda, \nu) = 1 + e^{i\nu}\lambda$, $\psi(\mathbf{A}, \nu) = \mathbf{I} + e^{i\nu}\mathbf{A}_1$, $h(\lambda, \nu) = 1 + 2\cos(\nu)\lambda + \lambda^2$ and $\mathcal{H}(\mathbf{A}, \nu) = \mathbf{I} + 2\cos(\nu)\mathbf{A}_1 + \mathbf{A}_1^2$, implying for each λ , $h(\lambda, \nu)$ is the spectral density (up to normalization) of a univariate MA(1) process with parameter λ .

TRANSFORMATION TO SERIAL INDEPENDENCE. The transformation to serial independence requires two steps, the first consisting of an approximation of the lag operator by a circulant matrix and the second of a rotation using the complex Fourier basis to achieve independence. Accordingly, let

$$\mathbf{L} = [o : e_1 : \dots : e_{n-1}] \quad \text{and} \quad \tilde{\mathbf{L}} = [e_n : e_1 : \dots : e_{n-1}]$$

be the $n \times n$ lag operator and its approximating circulant matrix, respectively, where o denotes the n -dimensional zero vector and e_j the j th canonical unit vector in \mathbb{R}^n taking the value 1 in the j th component and 0 elsewhere. Since $\tilde{\mathbf{L}}$ is a circulant matrix, its spectral decomposition is $\tilde{\mathbf{L}}u_t = \eta_t u_t$, $t = 1, \dots, n$, where $\eta_t = e^{i\nu_t}$, $\nu_t = 2\pi t/n$ and u_t the vector whose j th entry is η_t^j . It follows that $\tilde{\mathbf{L}}$ diagonalizes in the complex Fourier basis with the usual Fourier frequencies. Let $\mathbf{\Lambda}_{\tilde{\mathbf{L}}} = \text{diag}(\eta_1, \dots, \eta_n)$ and $\mathbf{U}_{\tilde{\mathbf{L}}} = [u_1 : \dots : u_n]$ denote the corresponding eigenvalue and eigenvector matrices, so that $\tilde{\mathbf{L}} = \mathbf{U}_{\tilde{\mathbf{L}}}\mathbf{\Lambda}_{\tilde{\mathbf{L}}}\mathbf{U}_{\tilde{\mathbf{L}}}^*$. Using $\mathbf{X} = [X_1 : \dots : X_n]$ and $\mathbf{Z} = [Z_1 : \dots : Z_n]$, the MA(1) process (6.1) can be transformed into

$$\mathbf{X}_1 = \mathbf{Z} + \mathbf{A}_1\mathbf{Z}\tilde{\mathbf{L}},$$

where \mathbf{X}_1 constitutes a redefinition of \mathbf{X} such that the first column is changed to $Z_1 + \mathbf{A}_1 Z_n$, while all other columns are as in the original data matrix \mathbf{X} . Rotating in the complex Fourier basis, the observations are transformed again into the vectors $\tilde{X}_1, \dots, \tilde{X}_n$ given by

$$\tilde{\mathbf{X}} = [\tilde{X}_1 : \dots : \tilde{X}_n] = \mathbf{X}_1\mathbf{U}_{\tilde{\mathbf{L}}}.$$

Observe that $\tilde{\mathbf{X}}$ has independent columns. To see this, note first that $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{U}_{\tilde{\mathbf{L}}}$ possesses the same distribution as \mathbf{Z} , since \mathbf{Z} has (complex) Gaussian entries and $\mathbf{U}_{\tilde{\mathbf{L}}}$ is a unitary matrix. Write then

$$\tilde{\mathbf{X}} = \tilde{\mathbf{Z}} + \mathbf{A}_1\tilde{\mathbf{Z}}\mathbf{\Lambda}_{\tilde{\mathbf{L}}} = [(\mathbf{I} + \eta_1\mathbf{A}_1)\tilde{Z}_1 : \dots : (\mathbf{I} + \eta_n\mathbf{A}_1)\tilde{Z}_n],$$

where $\tilde{\mathbf{Z}} = [\tilde{Z}_1 : \dots : \tilde{Z}_n]$, and independence of the columns (and thus serial independence) follows. Note also that $\mathbf{I} + \eta_t\mathbf{A}_1 = \psi(\mathbf{A}, \nu_t)$ and consequently $\tilde{\mathbf{X}} = \psi(\mathbf{A}, \nu_t)\tilde{\mathbf{Z}}$, using the transfer function ψ defined in (2.3).

TRANSFORMATION OF \mathbf{C}_τ . The vectors $\tilde{X}_1, \dots, \tilde{X}_n$ give rise to approximations $\tilde{\mathbf{C}}_\tau$ to the lag- τ symmetrized autocovariance matrices \mathbf{C}_τ ; in particular $\tilde{\mathbf{C}}_\tau$ and \mathbf{C}_τ will be shown to have the same large-sample spectral behavior, irrespective of the distribution of the entries. So let

$$\tilde{\mathbf{C}}_\tau \equiv \frac{1}{2n} \mathbf{X}_1 (\tilde{\mathbf{L}}^\tau + [\tilde{\mathbf{L}}^*]^\tau) \mathbf{X}_1^* = \frac{1}{n} \sum_{t=1}^n \cos(\tau \nu_t) \tilde{X}_t \tilde{X}_t^*,$$

where the latter equality follows from several small computations using the quantities introduced in the preceding paragraph. Now,

$$\begin{aligned} \mathbf{C}_\tau - \tilde{\mathbf{C}}_\tau &= \left(\mathbf{C}_\tau - \frac{1}{2n} \mathbf{X}_1 (\mathbf{L}^\tau + [\mathbf{L}^*]^\tau) \mathbf{X}_1^* \right) \\ &\quad + \left(\frac{1}{2n} \mathbf{X}_1 (\mathbf{L}^\tau + [\mathbf{L}^*]^\tau) \mathbf{X}_1^* - \frac{1}{2n} \mathbf{X}_1 (\tilde{\mathbf{L}}^\tau + [\tilde{\mathbf{L}}^*]^\tau) \mathbf{X}_1^* \right). \end{aligned}$$

The rank of the first difference on the right-hand side of the last display is at most 2, since \mathbf{X} and \mathbf{X}_1 differ only in the first column. The rank of the second difference is at most 2τ . The rank of $\mathbf{C}_\tau - \tilde{\mathbf{C}}_\tau$ is therefore at most $2(1 + \tau)$. Defining the resolvents

$$\mathbf{R}_\tau(z) = (\mathbf{C}_\tau - z\mathbf{I})^{-1} \quad \text{and} \quad \tilde{\mathbf{R}}_\tau(z) = (\tilde{\mathbf{C}}_\tau - z\mathbf{I})^{-1},$$

the Stieltjes transforms corresponding to the ESDs of \mathbf{C}_τ and $\tilde{\mathbf{C}}_\tau$ are, respectively, given by

$$s_{\tau,p}(z) = \frac{1}{p} \text{tr}[\mathbf{R}_\tau(z)] \quad \text{and} \quad \tilde{s}_{\tau,p}(z) = \frac{1}{p} \text{tr}[\tilde{\mathbf{R}}_\tau(z)].$$

It follows from Lemma S.2 that with probability one, the ESDs of \mathbf{C}_τ and $\tilde{\mathbf{C}}_\tau$ converge to the same limit, provided the limits exist. We conclude from Lemma S.3 that the ESD of $\tilde{\mathbf{C}}_\tau$ converges a.s. to a nonrandom distribution by showing that $\tilde{s}_{\tau,p}$ converge pointwise a.s. to the Stieltjes transform of a probability measure, thus establishing a.s. convergence of the ESD of \mathbf{C}_τ .

APPROXIMATING EQUATIONS FOR THE STIELTJES KERNEL K_τ . In order to derive limiting equation (2.5), a finite sample counterpart is needed. This can be derived as follows. The transformed data $\tilde{\mathbf{X}}$ gives rise to the transformed Stieltjes kernel $\tilde{K}_\tau : \mathbb{C}^+ \times [0, 2\pi] \rightarrow \mathbb{C}^+$ given by

$$\tilde{K}_\tau(z, \nu) = \frac{1}{p} \text{tr}[\tilde{\mathbf{R}}_\tau(z) \mathcal{H}(\mathbf{A}, \nu)].$$

Following arguments typically used to establish the deterministic equivalent of a resolvent matrix, $p \times p$ matrix-valued function solutions $(\mathbf{H}_{\tau,p}(z) : p \in \mathbb{N})$ are needed such that, for sufficiently large p ,

$$(6.2) \quad \frac{1}{p} \text{tr}[(\mathbf{I} + \mathbf{H}_{\tau,p}(z))^{-1} + z\tilde{\mathbf{R}}_\tau(z)] \mathbf{D}_p \approx 0$$

for all $z \in \mathbb{C}^+$ and all $p \times p$ Hermitian matrix sequences $(\mathbf{D}_p : p \in \mathbb{N})$ with uniformly bounded norms $\|\mathbf{D}_p\|$. If one uses $\mathbf{D}_p = \mathcal{H}(\mathbf{A}, \nu)$ and the definition of \tilde{K}_τ , the latter approximate equation becomes

$$\tilde{K}_\tau(z, \nu) \approx -\frac{1}{pz} \operatorname{tr}[(\mathbf{I} + \mathbf{H}_{\tau,p}(z))^{-1} \mathcal{H}(\mathbf{A}, \nu)].$$

Section 7 below is devoted to making precise the use of \approx in the above equations and to showing that choosing

$$(6.3) \quad \mathbf{H}_{\tau,p}(z) = -\frac{1}{zn} \sum_{t=1}^n \frac{\cos(\tau \nu_t) \mathcal{H}(\mathbf{A}, \nu_t)}{1 + c_n \cos(\tau \nu_t) \tilde{K}_\tau(z, \nu_t)},$$

with $\nu_t = 2\pi t/n$ and $c_n = p/n$, is appropriate.

6.2. *Extension to the non-Gaussian case.* In order to verify the statement of Theorem 2.1 for non-Gaussian innovations $(Z_t : t \in \mathbb{Z})$, two key ideas are invoked, namely showing that:

- for any $z \in \mathbb{C}^+$, the Stieltjes transform $s_p(z)$ concentrates around $\mathbb{E}[s_p(z)]$ regardless of the underlying distributional assumption;
- the difference between the expectations $\mathbb{E}[s_p(z)]$ under the Gaussian model and the non-Gaussian model is asymptotically negligible.

To establish the concentration property of the first item, McDiarmid’s inequality is used to bound probabilities of the type $\mathbb{P}(|s_p(z) - \mathbb{E}[s_p(z)]| \geq \varepsilon)$ for arbitrary $\varepsilon > 0$. These probabilities are then shown to converge to zero exponentially fast under (2.2). To establish the second item, the generalized Lindeberg principle of Chatterjee [8] is applied. To this end, the argument z is viewed as a parameter and $s_p(z)$ as a function of the real parts Z_{jt}^R and the imaginary parts Z_{jt}^I of the innovation entries Z_{jt} , for $j = 1, \dots, p$ and $t = 1, \dots, n$. The difference between Gaussian Z_t and non-Gaussian Z_t can then be analyzed by consecutively changing one pair (Z_{jt}^R, Z_{jt}^I) from Gaussian to non-Gaussian, thereby expressing the respective differences in the expected Stieltjes transforms as a sum of these entrywise changes. These differences will be evaluated through a Taylor series expansion, bounding certain third-order partial derivatives of s_p . Details are given in Section 9.

6.3. *Extension to the linear process case.* While the arguments established so far work in the same fashion also for $\text{MA}(q)$ processes, certain difficulties arise when making the transition to the $\text{MA}(\infty)$ or linear process case. First, if one constructs the data matrix \mathbf{X} not from $\text{MA}(1)$ observations as above but from the linear process $X_t = \sum_{\ell=0}^{\infty} \mathbf{A}_\ell Z_{t-\ell}$, then every column of \mathbf{X} is different from the corresponding column in the transformed matrix $\mathbf{X}_\infty = \sum_{\ell=0}^{\infty} \mathbf{A}_\ell \mathbf{Z} \tilde{\mathbf{L}}^\ell$ and not only the first column [or the first q columns for the $\text{MA}(q)$ case]. Second, for the $\text{MA}(1)$ case one can write the Stieltjes transform $s_p(z)$ as a function of $2p(n+1)$ variables

Z_{jt}^R and Z_{jt}^I [or $2p(n + q)$ variables for the MA(q) case], but for linear processes, even for finite p , $s_p(z)$ is a function of infinitely many Z_{jt}^R and Z_{jt}^I . This makes their study substantially harder.

Linear processes are thus, for the purposes of this paper, approached through truncation, that is, by approximation through finite-order MA processes $X_t = \sum_{\ell=0}^{q(p)} \mathbf{A}_\ell Z_{t-\ell}$ whose order $q(p)$ is a function of the dimension p and therefore grows with the sample size under (2.2). Obviously $q(p) \rightarrow \infty$ is a necessary condition to make this approximation work. However, $q(p)$ cannot grow too fast (leading to the same difficulties in transitioning from the Gaussian to the non-Gaussian case as for the linear process itself) or too slow (showing that the LSDs of the linear process and its truncated version are identical becomes an issue). It turns out in Section 10 that $q(p) = \lceil p^{1/3} \rceil$, with $\lceil \cdot \rceil$ denoting the ceiling function, is an appropriate choice.

6.4. *Including the real-valued case.* To address the statements of Theorem 2.2, the arguments presented thus far have to be adjusted for real-valued innovations ($Z_t : t \in \mathbb{Z}$). This is done using the eigen-decomposition of the coefficient matrices in the real Fourier basis, after which arguments already developed for the complex case apply. Detailed steps are given in Section 11.

6.5. *Dealing with the spectral density operator.* The key step toward proving Theorem 4.1 is to express $\widehat{\Gamma}_n(\eta)$ as

$$\begin{aligned} \widehat{\Gamma}_n(\eta) &= \sum_{\tau=-\infty}^{\infty} T_n(\tau) e^{i\tau\eta} \frac{1}{n} \mathbf{X} \mathbf{L}^{-\tau} \mathbf{X}^* \\ &= \frac{1}{n} \mathbf{X} \left(\mathbf{I}_p + \sum_{\tau=1}^{\infty} T_n(\tau) (e^{i\tau\eta} \mathbf{L}^{-\tau} + e^{-i\tau\eta} \mathbf{L}^\tau) \right) \mathbf{X}^*, \end{aligned}$$

and then noticing that the matrix $\mathbf{I}_p + \sum_{\tau=1}^{\infty} T_n(\tau) (e^{i\tau\eta} \widetilde{\mathbf{L}}^{-\tau} + e^{-i\tau\eta} \widetilde{\mathbf{L}}^\tau)$ diagonalizes in the (real or complex) Fourier basis with eigenvalues $\mathfrak{f}_{T_n}(\eta - \nu_t)$, for $t = 1, \dots, n$, so that the ESD of $\widehat{\Gamma}_n(\eta)$ can be approximated by the ESD of the matrix $\widetilde{\Gamma}_n(\eta) = n^{-1} \sum_{t=1}^n \mathfrak{f}_{T_n}(\eta - \nu_t) \widetilde{X}_t \widetilde{X}_t^*$, where \widetilde{X}_t^* is the t th column of the matrix $\mathbf{X} \mathbf{F}_n^*$, and \mathbf{F}_n denotes the $n \times n$ Fourier rotation matrix. We give the main steps of the arguments leading to this result. First, suppose that $m_n \rightarrow \infty$ such that $m_n/n \rightarrow 0$ as $n \rightarrow \infty$. Then we can write

$$(6.4) \quad \widehat{\Gamma}_n(\eta) = \widetilde{\Gamma}_{n,m_n}^{(1)}(\eta) + (\widehat{\Gamma}_{n,m_n}^{(1)}(\eta) - \widetilde{\Gamma}_{n,m_n}^{(1)}(\eta)) + \widehat{\Gamma}_{n,m_n}^{(2)}(\eta),$$

where

$$\begin{aligned} \widetilde{\Gamma}_{n,m_n}^{(1)}(\eta) &= \frac{1}{n} \mathbf{X} \left(\mathbf{I}_p + \sum_{\tau=1}^{m_n} T_n(\tau) (e^{i\tau\eta} \widetilde{\mathbf{L}}^{-\tau} + e^{-i\tau\eta} \widetilde{\mathbf{L}}^\tau) \right) \mathbf{X}^* \\ &= \frac{1}{n} \sum_{t=1}^n \mathfrak{f}_{T_n,m_n}(\eta - \nu_t) \widetilde{X}_t \widetilde{X}_t^*, \end{aligned}$$

with $f_{T_n, m_n}(\theta) = 1 + 2 \sum_{\tau=1}^{m_n} \cos(\tau\theta)$,

$$\widehat{\Gamma}_{n, m_n}^{(1)}(\eta) = \frac{1}{n} \mathbf{X} \left(\mathbf{I}_p + \sum_{\tau=1}^{m_n} T_n(\tau) (e^{i\tau\eta} \mathbf{L}^{-\tau} + e^{-i\tau\eta} \mathbf{L}^{\tau}) \right) \mathbf{X}^*$$

and

$$\widehat{\Gamma}_{n, m_n}^{(2)}(\eta) = \frac{1}{n} \mathbf{X} \left(\sum_{\tau=m_n+1}^n T_n(\tau) (e^{i\tau\eta} \mathbf{L}^{-\tau} + e^{-i\tau\eta} \mathbf{L}^{\tau}) \right) \mathbf{X}^*.$$

Now, the following facts together with representation (6.4) and Theorem A.43 of Bai and Silverstein [5] (rank inequality) and Lemma S.1 (norm inequality) prove the assertion:

- (i) $\sup_{\theta \in [0, 2\pi]} \max\{|f_{T_n, m_n}(\theta) - f_{T_n}(\theta)|, |f_{T_n}(\theta) - f_T(\theta)|\} \rightarrow 0$, as $n \rightarrow \infty$;
- (ii) $\text{rank}(\widehat{\Gamma}_{n, m_n}^{(1)}(\eta) - \widetilde{\Gamma}_{n, m_n}^{(1)}(\eta)) \leq 2m_n = o(p)$;
- (iii)

$$p^{-1} \|\widehat{\Gamma}_{n, m_n}^{(2)}(\eta)\|_F^2 \leq \|\widehat{\Gamma}_{n, m_n}^{(2)}(\eta)\|^2$$

and

$$\|\widehat{\Gamma}_{n, m_n}^{(2)}(\eta)\| \leq \|n^{-1} \mathbf{X} \mathbf{X}^*\| \sum_{\tau=m_n+1}^{\infty} |T_n(\tau)| \rightarrow 0 \quad \text{a.s.}$$

7. Proof for the complex Gaussian MA(q) case. Throughout, $(\mathbf{A}_\ell : \ell \in \mathbb{N})$ is treated as a sequence of nonrandom matrices, and all the arguments are valid conditionally on this sequence. In this section, the result of Theorem 2.1 is first verified for the MA(q) process $X_t = \sum_{\ell=0}^q \mathbf{A}_\ell Z_{t-\ell}$ when Z_{j_t} 's are i.i.d. standard complex Gaussian; that is, real and imaginary parts of Z_{j_t} are independent normals with mean zero and variance one half. Following the outline in Section 6.1, the data matrix $\mathbf{X} = [X_1 : \dots : X_n]$ is transformed into the matrix $\widetilde{\mathbf{X}} = [\widetilde{X}_1 : \dots : \widetilde{X}_n]$, with each column satisfying $\widetilde{X}_t = \psi(\mathbf{A}, \nu_t) \widetilde{Z}_t$. Then, since the rank of $\mathbf{X} - (\mathbf{Z} + \sum_{\ell=1}^q \mathbf{A}_\ell \mathbf{Z} \widetilde{\mathbf{L}}^\ell)$ is at most q , by Lemma S.2, it follows that the ESDs of \mathbf{C}_τ and $\widetilde{\mathbf{C}}_\tau$ have the same limit, if the latter exists. For simplicity, let $\psi_t = \psi(\mathbf{A}, \nu_t)$. To keep notation more compact, the extra subscripts p and τ (indicating the lag of the autocovariance matrix under consideration) are often suppressed when no confusion can arise. For example, in (6.2) the notation $\mathbf{H}(z)$ will be preferred over the more complex $\mathbf{H}_{\tau, p}(z)$. The proof is given in several steps. First, a bound on the approximation error is derived if the Stieltjes kernel K in (2.5) is replaced with its finite sample counterpart \widetilde{K} . Second, existence, convergence and continuity of the solution to (6.2) are verified. Third, tightness of the ESDs \widehat{F}_p and convergence of the corresponding Stieltjes transforms s_p is shown.

7.1. *Bound on the approximation error.* The goal is to provide a rigorous formulation of (6.2) and a bound on the resulting approximation error. The first step consists of giving a heuristic argument for the definition of $\mathbf{H}(z)$ in (6.3). To this

end, note that

$$\begin{aligned}
 & (\mathbf{I} + \mathbf{H}(z))^{-1} + z\tilde{\mathbf{R}}(z) \\
 (7.1) \quad & = \tilde{\mathbf{R}}(z)\tilde{\mathbf{R}}(z)^{-1}(\mathbf{I} + \mathbf{H}(z))^{-1} - \tilde{\mathbf{R}}(z)(\mathbf{I} + \mathbf{H}(z))(\mathbf{I} + \mathbf{H}(z))^{-1} \\
 & = \tilde{\mathbf{R}}(z)(\tilde{\mathbf{C}}_\tau + z\mathbf{H}(z))(\mathbf{I} + \mathbf{H}(z))^{-1}.
 \end{aligned}$$

It follows that, to achieve (6.2), it is sufficient to solve $\tilde{\mathbf{R}}(z)\tilde{\mathbf{C}}_\tau \approx -z\tilde{\mathbf{R}}(z)\mathbf{H}(z)$. (The use of \approx will be clarified below.) Let

$$\gamma_{\tau,t} = \cos(\tau v_t),$$

and define the rank-one perturbation and its corresponding resolvent, respectively, given by

$$\tilde{\mathbf{C}}_{\tau,t} = \tilde{\mathbf{C}}_\tau - \frac{1}{n}\gamma_{\tau,t}\tilde{X}_t\tilde{X}_t^* \quad \text{and} \quad \tilde{\mathbf{R}}_t(z) = (\tilde{\mathbf{C}}_{\tau,t} - z\mathbf{I})^{-1}.$$

Using $\tilde{\mathbf{R}}(z) = ([\tilde{\mathbf{C}}_{\tau,t} - z\mathbf{I}] + \frac{1}{n}\gamma_{\tau,t}X_tX_t^*)^{-1}$ and defining

$$\mathcal{H}_t = \mathcal{H}(\mathbf{A}, v_t),$$

the Sherman–Morrison formula and some matrix algebra lead to

$$\begin{aligned}
 (7.2) \quad \tilde{\mathbf{R}}(z)\tilde{\mathbf{C}}_\tau &= \frac{1}{n} \sum_{t=1}^n \left(\tilde{\mathbf{R}}_t(z) - \frac{(1/n)\gamma_{\tau,t}\tilde{\mathbf{R}}_t(z)\tilde{X}_t\tilde{X}_t^*\tilde{\mathbf{R}}_t(z)}{1 + (1/n)\gamma_{\tau,t}\tilde{X}_t^*\tilde{\mathbf{R}}_t(z)\tilde{X}_t} \right) \gamma_{\tau,t}\tilde{X}_t\tilde{X}_t^* \\
 &= \frac{1}{n} \sum_{t=1}^n \tilde{\mathbf{R}}_t(z) \frac{\gamma_{\tau,t}\tilde{X}_t\tilde{X}_t^*}{1 + (1/n)\gamma_{\tau,t}\tilde{X}_t^*\tilde{\mathbf{R}}_t(z)\tilde{X}_t} \\
 &= \frac{1}{n} \sum_{t=1}^n \tilde{\mathbf{R}}_t(z) \frac{\gamma_{\tau,t}\boldsymbol{\psi}_t\tilde{Z}_t\tilde{Z}_t^*\boldsymbol{\psi}_t^*}{1 + (1/n)\gamma_{\tau,t}\tilde{Z}_t^*\boldsymbol{\psi}_t^*\tilde{\mathbf{R}}_t(z)\boldsymbol{\psi}_t\tilde{Z}_t} \\
 &\approx \tilde{\mathbf{R}}(z) \frac{1}{n} \sum_{t=1}^n \frac{\gamma_{\tau,t}\mathcal{H}_t}{1 + (1/n)\gamma_{\tau,t} \text{tr}[\tilde{\mathbf{R}}(z)\mathcal{H}_t]} \\
 &= \tilde{\mathbf{R}}(z) \frac{1}{n} \sum_{t=1}^n \frac{\gamma_{\tau,t}\mathcal{H}_t}{1 + c_n\gamma_{\tau,t}\tilde{K}(z, v_t)},
 \end{aligned}$$

thus validating the choice of $\mathbf{H}(z)$ as given in (6.3). The \approx sign is due to substituting $\tilde{\mathbf{R}}_t(z)$, $\tilde{Z}_t\tilde{Z}_t^*$, $\tilde{Z}_t^*\boldsymbol{\psi}_t^*\tilde{\mathbf{R}}_t(z)\boldsymbol{\psi}_t\tilde{Z}_t$ with $\tilde{\mathbf{R}}(z)$, \mathbf{I} , $\text{tr}[\boldsymbol{\psi}_t^*\tilde{\mathbf{R}}(z)\boldsymbol{\psi}_t]$, respectively. The arguments are made precise in the following theorem.

THEOREM 7.1. *Let \mathbf{D} be a Hermitian matrix such that $\|\mathbf{D}\| \leq \bar{\lambda}_{\mathbf{D}}$ and $z \in \mathbb{C}^+$. If the assumptions of Theorem 2.1 are satisfied, then*

$$(7.3) \quad \frac{1}{zp} \text{tr}[(\mathbf{I} + \mathbf{H}(z))^{-1} + z\tilde{\mathbf{R}}(z)\mathbf{D}] \rightarrow 0 \quad \text{a.s.}$$

under (2.2), where \mathbf{H} is defined in (6.3), and $\tilde{\mathbf{R}}(z) = (\tilde{\mathbf{C}}_\tau - z\mathbf{I})^{-1}$ is the resolvent of the symmetrized autocovariance matrix $\tilde{\mathbf{C}}_\tau = \frac{1}{n} \sum_{t=1}^n \gamma_{\tau,t} \tilde{X}_t \tilde{X}_t^*$ with $\gamma_{\tau,t} = \cos(\tau \nu_t)$.

PROOF. Observe that using (7.1), (7.2) and the definition of \mathbf{H} in (6.3), the a.s. convergence in (7.3) is shown to be equivalent to

$$\bar{d}_\tau^{(n)} = \frac{1}{n} \sum_{t=1}^n d_{\tau,t} \rightarrow 0 \quad \text{a.s.,}$$

where

$$d_{\tau,t} = \frac{1}{zp} \operatorname{tr}[\gamma_{\tau,t}(\tilde{\beta}_{\tau,t} \tilde{\mathbf{R}}_t(z) \tilde{\mathcal{H}}_t - \beta_{\tau,t} \tilde{\mathbf{R}}(z) \mathcal{H}_t)(\mathbf{I} + \mathbf{H}(z))^{-1} \mathbf{D}],$$

with $\tilde{\mathcal{H}}_t = \boldsymbol{\psi}_t \tilde{Z}_t \tilde{Z}_t^* \boldsymbol{\psi}_t^* = \tilde{X}_t \tilde{X}_t^*$,

$$(7.4) \quad \begin{aligned} \tilde{\beta}_{\tau,t} &= \left(1 + \frac{1}{n} \gamma_{\tau,t} \tilde{X}_t^* \tilde{\mathbf{R}}_t(z) \tilde{X}_t\right)^{-1}, \\ \beta_{\tau,t} &= \left(1 + \frac{1}{n} \gamma_{\tau,t} \operatorname{tr}[\tilde{\mathbf{R}}(z) \mathcal{H}_t]\right)^{-1}. \end{aligned}$$

Decomposing further, write next $d_{\tau,t} = d_{\tau,t}^{(1)} + \dots + d_{\tau,t}^{(5)}$, where

$$\begin{aligned} d_{\tau,t}^{(1)} &= \frac{1}{zp} \operatorname{tr}[\gamma_{\tau,t} \tilde{\beta}_{\tau,t} \tilde{\mathbf{R}}_t(z) \tilde{\mathcal{H}}_t ((\mathbf{I} + \mathbf{H}(z))^{-1} - (\mathbf{I} + \mathbf{H}_t(z))^{-1}) \mathbf{D}], \\ d_{\tau,t}^{(2)} &= \frac{1}{zp} \operatorname{tr}[\gamma_{\tau,t} \tilde{\beta}_{\tau,t} \tilde{\mathbf{R}}_t(z) (\tilde{\mathcal{H}}_t - \mathcal{H}_t) (\mathbf{I} + \mathbf{H}_t(z))^{-1} \mathbf{D}], \\ d_{\tau,t}^{(3)} &= \frac{1}{zp} \operatorname{tr}[\gamma_{\tau,t} \tilde{\beta}_{\tau,t} \tilde{\mathbf{R}}_t(z) \mathcal{H}_t ((\mathbf{I} + \mathbf{H}_t(z))^{-1} - (\mathbf{I} + \mathbf{H}(z))^{-1}) \mathbf{D}], \\ d_{\tau,t}^{(4)} &= \frac{1}{zp} \operatorname{tr}[\gamma_{\tau,t} \tilde{\beta}_{\tau,t} (\tilde{\mathbf{R}}_t(z) - \tilde{\mathbf{R}}(z)) \mathcal{H}_t (\mathbf{I} + \mathbf{H}(z))^{-1} \mathbf{D}], \\ d_{\tau,t}^{(5)} &= \frac{1}{zp} \operatorname{tr}[\gamma_{\tau,t} (\tilde{\beta}_{\tau,t} - \beta_{\tau,t}) \tilde{\mathbf{R}}(z) \mathcal{H}_t (\mathbf{I} + \mathbf{H}(z))^{-1} \mathbf{D}], \end{aligned}$$

with

$$(7.5) \quad \mathbf{H}_t = -\frac{1}{zn} \sum_{t=1}^n \tilde{\beta}_{\tau,t} \gamma_{\tau,t} \mathcal{H}_t, \quad \bar{\beta}_{\tau,t} = \left(1 + \frac{1}{n} \gamma_{\tau,t} \operatorname{tr}[\tilde{\mathbf{R}}_t(z) \mathcal{H}_t]\right)^{-1},$$

thus exhibiting the various approximations being made in the proof.

The Borel–Cantelli lemma provides that $\bar{d}_\tau^{(n)} \rightarrow 0$ almost surely is implied if $\mathbb{P}(\max_{t \leq n} |d_{\tau,t}| > \varepsilon) \rightarrow 0$ faster than $\frac{1}{n}$ for all $\varepsilon > 0$. Since $\mathbb{P}(\max_{t \leq n} |d_{\tau,t}| > \varepsilon) \leq \sum_{\ell=1}^5 \mathbb{P}(\max_{t \leq n} |d_{\tau,t}^{(\ell)}| > \frac{\varepsilon}{5})$, in order to verify (7.3), it is sufficient to show that

$\mathbb{P}(\max_{t \leq n} |d_{\tau,t}^{(\ell)}| > \varepsilon)$ goes to zero faster than $\frac{1}{n}$ for all $\varepsilon > 0$ and $\ell = 1, \dots, 5$. The corresponding arguments are detailed in Section S1.1 of the online SM. \square

7.2. Existence and uniqueness of the solution. In this section, the proof of Theorem 2.1 is completed for the complex Gaussian innovation model. In what follows, \mathbf{A} is without loss of generality assumed to be nonrandom, thereby restricting randomness to the innovations $\mathbf{Z} = [Z_1 : \dots : Z_n]$. For a fixed ω in the underlying sample space Ω , notation such as $\mathbf{Z}(\omega)$ will be utilized to indicate realizations of the respective random quantities.

Noticing first that Theorem 2.1 makes an almost sure convergence statement, a suitable subset $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$ is determined. This subset is used for all subsequent arguments. To this end, observe that since the matrix $\tilde{\mathbf{Z}} = [\tilde{Z}_1 : \dots : \tilde{Z}_n]$ has i.i.d. entries with zero mean and unit variance the norm of $n^{-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^*$ converges almost surely to a number not exceeding $(1 + \sqrt{c})^2 + 1$. Let $F_{\tilde{\mathbf{Z}}}$ denote the ESD of $n^{-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^*$, and define

$$\Omega_1 = \{\omega \in \Omega : F_{\tilde{\mathbf{Z}}(\omega)} < (1 + \sqrt{c})^2 + 1 \text{ for } p \geq p_0(\omega)\},$$

with suitably chosen $p_0(\omega)$. Then $\mathbb{P}(\Omega_1) = 1$. Define next

$$\bar{d}(z, \nu, \omega) = \frac{1}{zp} \operatorname{tr}[(\mathbf{I} + \mathbf{H}(z))^{-1} + z\tilde{\mathbf{R}}(z, \omega)]\mathcal{H}(\mathbf{A}, \nu),$$

$$\bar{d}_{\mathbf{I}}(z, \omega) = \frac{1}{zp} \operatorname{tr}[(\mathbf{I} + \mathbf{H}(z))^{-1} + z\tilde{\mathbf{R}}(z, \omega)].$$

Let $\mathbb{C}_{\mathbb{Q}}^+$ denote the set of complex numbers with rational real part and positive rational imaginary part and $[0, 2\pi]_{\mathbb{Q}} = [0, 2\pi] \cap \mathbb{Q}$. Define the set

$$\Omega_2 = \{\omega \in \Omega : \bar{d}(z, \nu, \omega) \rightarrow 0, \bar{d}_{\mathbf{I}}(z, \omega) \rightarrow 0, z \in \mathbb{C}_{\mathbb{Q}}^+, \nu \in [0, 2\pi]_{\mathbb{Q}}\}.$$

In view of Theorem 7.1, it follows that $\bar{d}(z, \nu, \cdot) \rightarrow 0$ a.s. and $\bar{d}_{\mathbf{I}}(z, \cdot) \rightarrow 0$ a.s. for all fixed $z \in \mathbb{C}_{\mathbb{Q}}^+$ and $\nu \in [0, 2\pi]_{\mathbb{Q}}$. Thus $\mathbb{P}(\Omega_2) = 1$. Henceforth only $\omega \in \Omega_0 = \Omega_1 \cap \Omega_2$, so that $\mathbb{P}(\Omega_0) = 1$, are considered.

Recall that $\tilde{K}_{\tau}(z, \nu, \omega) = p^{-1} \operatorname{tr}[\tilde{\mathbf{R}}(z, \omega)\mathcal{H}(\mathbf{A}, \nu)]$. The following theorem establishes existence of a Stieltjes kernel solution to the equations in (2.5) along a subsequence.

THEOREM 7.2 (Existence). *Suppose that the assumptions of Theorem 2.1 are satisfied:*

(a) *For all $\omega \in \Omega_0$ and for all subsequences of $\{p\}$, there exists another subsequence $\{p_{\ell}\}$ along which $\tilde{K}_{\tau}(z, \nu, \omega)$ converges pointwise in $z \in \mathbb{C}^+$ and uniformly in $\nu \in [0, 2\pi]$ to a limit $\mathfrak{K}_{\tau}(z, \nu, \omega)$ analytic in z and continuous in ν .*

(b) *For every subsequence $\{p_{\ell}\}$ satisfying (a), $\mathfrak{K}_{\tau}(z, \nu, \omega)$ satisfies (2.5) for any $z \in \mathbb{C}^+$. Moreover, $\mathfrak{K}_{\tau}(z, \nu, \omega)$ is the Stieltjes transform of a measure with mass $m_{\nu} = \int h(\lambda, \nu) dF^{\mathbf{A}}(\lambda)$, provided that $m_{\nu} > 0$.*

PROOF. (a) Let $z = w + iv$. Then for a compact subset $S \subset \mathbb{C}^+$, $\tilde{K}_\tau(z, \nu, \omega) \leq \bar{\lambda}_A^2 (\min_{z \in S} \nu)^{-1}$ for all $\nu \in [0, 2\pi]$ and $z \in S$. Enumerate $[0, 2\pi]_{\mathbb{Q}} = \{\nu_\ell : \ell \in \mathbb{N}\}$. Let $\{p_\ell^j(\omega)\} \subset \{p_\ell^{j-1}(\omega)\}$ mean that $\{p_\ell^j(\omega)\}$ is a further subsequence of $\{p_\ell^{j-1}(\omega)\}$, and let $\{p\}$ denote the original sequence. An application of Lemma 3 in Geronimo and Hill [15] yields that, for any fixed $\omega \in \Omega_0$, there is a sequence of subsequences

$$\{p_\ell^j(\omega)\} \subset \{p_\ell^{j-1}(\omega)\} \subset \dots \subset \{p_\ell^1(\omega)\} \subset \{p\},$$

so that $K_\tau^{(p_\ell^j)}(z, \nu, \omega)$ converges to an analytic function of z on $\{\nu_\ell : \ell = 1, \dots, \ell\}$. A standard diagonal argument implies that $K_\tau^{(p_\ell^j)}(z, \nu, \omega)$ converges to an analytic function of z on $[0, 2\pi]_{\mathbb{Q}}$. To simplify notation, write $K_\tau^{(p_\ell)}(z, \nu, \omega)$ in place of $K_\tau^{(p_\ell^j)}(z, \nu, \omega)$. Observe that the thus obtained limit, which will be denoted by $\mathfrak{K}_\tau(z, \nu, \omega)$, is so far defined only on $\mathbb{C}^+ \times [0, 2\pi]_{\mathbb{Q}}$. It remains to obtain the extension of the limit to $\mathbb{C}^+ \times [0, 2\pi]$. Note that Lemma S.10 implies that, for any $z \in \mathbb{C}^+$, $K_\tau^{(p_\ell)}(z, \nu, \omega)$ are equicontinuous in ν and converge pointwise to $\mathfrak{K}_\tau(z, \nu, \omega)$ on the dense subset $[0, 2\pi]_{\mathbb{Q}}$ of $[0, 2\pi]$. By the Arzela–Ascoli theorem, $K_\tau^{(p_\ell)}(z, \nu, \omega)$ converges therefore uniformly to a continuous function of ν on $\mathbb{C}^+ \times [0, 2\pi]$. This limit, denoted again by $\mathfrak{K}_\tau(z, \nu, \omega)$, is also analytic on \mathbb{C}^+ . To see this, pick $\nu_0 \in [0, 2\pi] \setminus \mathbb{Q}$ and a sequence $\{\nu_p\} \subset [0, 2\pi]_{\mathbb{Q}}$ such that $\nu_p \rightarrow \nu_0$. Then $\{\mathfrak{K}_\tau(z, \nu_p, \omega)\}$ satisfies the conditions of Lemma 3 in [15], and consequently there is a subsequence of $\{\mathfrak{K}_\tau(z, \nu_p, \omega)\}$ that converges to an analytic function. The limit of this subsequence has to coincide with $\mathfrak{K}_\tau(z, \nu_0, \omega)$ by continuity on $[0, 2\pi]$. It follows that $\mathfrak{K}_\tau(z, \nu, \omega)$ analytic.

(b) By Lemma S.11 and the definition of Ω_0 , for all $\omega \in \Omega_0$, $\mathfrak{K}_\tau(z, \nu, \omega)$ satisfies (2.5) for all $z \in \mathbb{C}_{\mathbb{Q}}^+$. So, by analyticity in z , it holds for all \mathbb{C}^+ .

Suppose first that $m_\nu = 0$. Then $|K_\tau(z, \nu, \omega)| \leq \nu^{-1} \int h(\lambda, \nu) dF_p^A(\lambda) \rightarrow \nu^{-1} \int h(\lambda, \nu) dF^A(\lambda) = 0$. Thus $\mathfrak{K}(z, \nu, \omega) = 0$ for all $z \in \mathbb{C}^+$, and the claim is verified.

For the remainder, suppose that $m_\nu > 0$. Showing that $\mathfrak{K}_\tau(z, \nu, \omega)$ is a Stieltjes transform of a measure with mass $m_\nu = \int h(\lambda, \nu) dF^A(\lambda)$ is equivalent to showing that $m_\nu^{-1} \mathfrak{K}_\tau(z, \nu, \omega)$ is Stieltjes transform of a Borel probability measure. Let $\mathbf{A}_{\tilde{C}}$ and $\mathbf{U}_{\tilde{C}}$ denote the eigenvalue and eigenvector matrices of \tilde{C}_τ . Then $\tilde{K}_\tau^{(p_\ell)}(z, \nu, \omega) = p_\ell^{-1} \text{tr}[(\mathbf{A}_{\tilde{C}} - z\mathbf{I})^{-1} \mathbf{U}_{\tilde{C}}^* h(\mathbf{A}, \nu) \mathbf{U}_{\tilde{C}}]$ is the Stieltjes transform of a measure with mass $m_\nu^{(p_\ell)} = p_\ell^{-1} \text{tr}[h(\mathbf{A}, \nu)]$. By the weak convergence of $F_{p_\ell}^A$ to F^A , $m_\nu^{(p_\ell)} \rightarrow m_\nu > 0$ as $p_\ell \rightarrow \infty$. This shows that $m_\nu^{(p_\ell)} \geq \bar{m}_\nu > 0$ for all p_ℓ . Since the diagonal entries of $\mathbf{U}_{\tilde{C}}^* h(\mathbf{A}, \nu) \mathbf{U}_{\tilde{C}}$ are bounded from above by $\bar{\lambda}_A^2$, it follows that $(m_\nu^{(p_\ell)})^{-1} \tilde{K}_\tau^{(p_\ell)}(z, \nu, \omega)$ is the Stieltjes transform of a measure μ_{p_ℓ} ,

say, such that, for all real x , $\mu_{p_\ell}((x, \infty)) \leq \bar{\lambda}_A^2 \bar{m}_v^{-1} F^{\tilde{C}_\tau}((x, \infty))$, where $F^{\tilde{C}_\tau}$ denotes the ESD of \tilde{C}_τ . It follows from Lemma S.12 that $\{F^{\tilde{C}_\tau}\}$ is a tight sequence. Therefore $(m_v^{(p_\ell)})^{-1} \tilde{K}_\tau^{(p_\ell)}(z, v, \omega)$ are the Stieltjes transforms of a tight sequence of Borel measures. An application of Lemma S.13 yields that $K_\tau(z, v, \omega)$ is the Stieltjes transform of a measure with mass m_v , completing the proof. \square

THEOREM 7.3 (Uniqueness). *If the assumptions of Theorem 2.1 are satisfied, then there is a unique solution $K_\tau(z, v)$ to (2.5) that is analytic in $z \in \mathbb{C}^+$ and continuous in $v \in [0, 2\pi]$ with $K_\tau(z, v)$ being a Stieltjes transform of a measure with mass $\int h(\lambda, v) dF^A(\lambda)$.*

PROOF. Suppose there are two solutions $K_{\tau,1}(z, v)$ and $K_{\tau,2}(z, v)$ to (2.5). Let $\gamma_\tau(v) = \cos(\tau v)$ and $\beta_{\tau,j}(z, v) = (1 + c\gamma_\tau(v)K_{\tau,j}(z, v))^{-1}$. Define then $U_{\tau,j}(z, \lambda) = (2\pi)^{-1} \int_0^{2\pi} \beta_{\tau,j}(z, \lambda)\gamma_\tau(v)h(\lambda, v) dv$, $j = 1, 2$. Note that $U_{\tau,1}(z, \lambda)$ and $U_{\tau,2}(z, \lambda)$ have nonpositive imaginary parts. Now

$$\begin{aligned} &U_{\tau,1}(z, \lambda) - U_{\tau,2}(z, \lambda) \\ &= \frac{c}{2\pi} \int_0^{2\pi} \beta_{\tau,1}(z, v)\beta_{\tau,2}(z, v)\gamma_\tau^2(v)(K_{\tau,1}(z, v) - K_{\tau,2}(z, v))h(\lambda, v) dv \end{aligned}$$

and thus

$$\begin{aligned} &K_{\tau,1}(z, v) - K_{\tau,2}(z, v) \\ &= - \int \frac{(U_{\tau,1}(z, \lambda) - U_{\tau,2}(z, \lambda))h(\lambda, v)}{(U_{\tau,1}(z, \lambda) - z)(U_{\tau,2}(z, \lambda) - z)} dF^A(\lambda) \\ &= - \int \frac{c}{2\pi} \int_0^{2\pi} \prod_{j=1}^2 \frac{\beta_{\tau,j}(z, v')}{U_{\tau,j}(z, \lambda) - z} \gamma_\tau^2(v') \\ &\quad \times (K_{\tau,1}(z, v') - K_{\tau,2}(z, v'))h(\lambda, v') dv' h(\lambda, v) dF^A(\lambda). \end{aligned}$$

Using the fact that $K_{\tau,j}(z, v)$ is a Stieltjes transform with mass bounded from above by $\bar{\lambda}_A^2$, it follows that $\prod_{j=1}^2 \beta_{\tau,j}(z, v')(U_{\tau,j}(z, \lambda) - z)^{-1}$ is bounded by $C(v) = \max\{64c^2\bar{\lambda}_A^4 v^{-4}, 4v^{-2}\}$. Thus

$$\begin{aligned} &\int_0^{2\pi} |K_{\tau,1}(z, v) - K_{\tau,2}(z, v)| dv \\ &\leq 4C(v)c\bar{\lambda}_A^4 \int_0^{2\pi} |K_{\tau,1}(z, v') - K_{\tau,2}(z, v')| dv'. \end{aligned}$$

If $v > 4\bar{\lambda}_A^2 \max\{c^{3/4}, \sqrt{c}\}$, then $4C(v)c\bar{\lambda}_A^4 < 1$ and thus

$$\int_0^{2\pi} |K_{\tau,1}(z, v) - K_{\tau,2}(z, v)| dv = 0,$$

which by continuity in ν implies that $K_{\tau,1}(z, \nu) = K_{\tau,2}(z, \nu)$ for any fixed $z \in \mathbb{C}^+$. Since both solutions are analytic, the equality holds indeed for all $z \in \mathbb{C}^+$. This proves uniqueness. \square

In the remainder of this section, the proof of Theorem 2.1 is completed for the Gaussian MA(q) case. This is done by establishing that (a) the convergence along subsequences as stated in Theorem 7.2 holds indeed for the whole sequence and (b) the relevant ESDs converge.

Toward (a), it is necessary to prove that, for every $\omega \in \Omega_0$, $\tilde{K}_\tau(z, \nu, \omega)$ converges to $K_\tau(z, \nu)$ pointwise in $z \in \mathbb{C}^+$ and uniformly for $\nu \in [0, 2\pi]$ under (2.2). Assume the contrary, and suppose that there are $z_0 \in \mathbb{C}^+$, $\nu_0 \in [0, 2\pi]$ and $\omega_0 \in \Omega_0$ such that $\tilde{K}_\tau(z_0, \nu_0, \omega_0)$ does not converge to $K_\tau(z_0, \nu_0)$. By boundedness of $\tilde{K}_\tau(z_0, \nu_0, \omega_0)$, there is a subsequence $\{p_\ell\}$ along which $\tilde{K}_\tau(z_0, \nu_0, \omega_0)$ converges to a limit different from $K_\tau(z_0, \nu_0)$. Invoking Theorems 7.2 and 7.3, there is a further subsequence $\{p_{\ell'}\}$ of $\{p_\ell\}$ along which $\tilde{K}_\tau(z_0, \nu, \omega_0)$ converges to $K_\tau(z_0, \nu)$ uniformly in $\nu \in [0, 2\pi]$. This is a contradiction. It follows that for every $\omega \in \Omega_0$, $\tilde{K}_\tau(z, \nu, \omega)$ converges to $K_\tau(z, \nu)$ pointwise in $z \in \mathbb{C}^+$ and $\nu \in [0, 2\pi]$. An application of Theorem 7.2 and the Arzela–Ascoli theorem shows that the convergence is uniform on $[0, 2\pi]$. Note that, for any $z \in \mathbb{C}^+$, $K_{\tau,p}(z, \nu, \omega) = p^{-1} \text{tr}[\mathbf{R}(z)\mathcal{H}(\mathbf{A}, \nu)]$ converges to $K_\tau(z, \nu)$ uniformly on $[0, 2\pi]$. Since we have $|\tilde{K}_\tau(z, \nu, \omega) - K_{\tau,p}(z, \nu, \omega)| < \bar{\lambda}_\mathbf{A}^2 (vp)^{-1}$, assertion (a) follows.

Toward (b), let $\omega \in \Omega_0$. It needs to be shown that $s_{\tau,p}(z, \omega) \rightarrow s_\tau(z)$ on \mathbb{C}^+ . By arguments as in the proof of Lemma S.11, it is already established that $\tilde{s}_{\tau,p}(z, \omega) \rightarrow s_\tau(z)$ on $\mathbb{C}_\mathbb{Q}^+$. Now, for any compact $S \subset \mathbb{C}^+$ and $z_1, z_2 \in S$,

$$\begin{aligned} |\tilde{s}_{\tau,p}(z_1, \omega) - \tilde{s}_{\tau,p}(z_2, \omega)| &= \frac{1}{p} |\text{tr}[\tilde{\mathbf{R}}(z_1, \omega) - \tilde{\mathbf{R}}(z_2, \omega)]| \\ &= \frac{1}{p} |\text{tr}[(z_1 - z_2)\tilde{\mathbf{R}}(z_1, \omega)\tilde{\mathbf{R}}(z_2, \omega)]| \\ &\leq |z_1 - z_2| \left(\min_{z \in S} \nu \right)^2. \end{aligned}$$

Thus $\{\tilde{s}_{\tau,p}(z, \omega)\}$ are equicontinuous in z (with p and ω as parameters) on S . By Arzela–Ascoli, $\tilde{s}_{\tau,p}(z, \omega)$ thus converges uniformly to $s_\tau(z)$ on S . Consequently $\tilde{s}_{\tau,p}(z, \omega) \rightarrow s_\tau(z)$ on \mathbb{C}^+ . Since \tilde{F}_τ , the ESD of $\tilde{\mathbf{C}}_\tau$, is tight (by Lemma S.12), it follows from Lemmas S.13 and S.3 that $s_\tau(z)$ is a Stieltjes transform of a (nonrandom) probability measure, and \tilde{F}_τ converges a.s. to the distribution whose Stieltjes transform is given by s_τ . Since, by Lemma S.2, $\sup_\sigma |\tilde{F}_\tau(\sigma) - \hat{F}_\tau(\sigma)| \rightarrow 0$ a.s., it follows that \hat{F}_τ converges a.s. to the same limit, and hence $s_{\tau,p}(z)$ converges a.s. to $s_\tau(z)$. The proof for the Gaussian MA(q) case is complete. It can be checked that all the statements remain valid even if $q = q(p) \rightarrow \infty$ sufficiently slowly under (2.2), for example, if $q(p) = o(p^{1/2})$.

8. Truncation, centering and rescaling. The extension of the result to non-Gaussian innovations requires in its first step, a truncation argument, followed by a centering and rescaling of the innovations. This section justifies that the symmetrized autocovariance matrices obtained from Gaussian innovations and from their truncated, centered and rescaled counterparts have the same LSD. The extension to the non-Gaussian case is then completed in Section 9.

Since the underlying process is an MA(q) series, the observations X_1, \dots, X_n are functions of the innovations Z_{1-q}, \dots, Z_n . For $j = 1, \dots, p$ and $t = 1 - q, \dots, n$, define the quantities

$$\begin{aligned} \bar{Z}_{jt}^R &= Z_{jt}^R \mathbb{I}(|Z_{jt}^R| < \sqrt{p}), & \check{Z}_{jt}^R &= \bar{Z}_{jt}^R - \mathbb{E}[\bar{Z}_{jt}^R], \\ \bar{Z}_{jt}^I &= Z_{jt}^I \mathbb{I}(|Z_{jt}^I| < \sqrt{p}), & \check{Z}_{jt}^I &= \bar{Z}_{jt}^I - \mathbb{E}[\bar{Z}_{jt}^I], \end{aligned}$$

where $Z_{jt} = Z_{jt}^R + iZ_{jt}^I$ and \mathbb{I} the indicator function. Correspondingly define $\bar{\mathbf{C}}_\tau$ and $\check{\mathbf{C}}_\tau$ to be the autocovariance matrices obtained from $\bar{Z}_{jt}^R, \bar{Z}_{jt}^I$ and $\check{Z}_{jt}^R, \check{Z}_{jt}^I$, respectively.

PROPOSITION 8.1. *If the assumptions of Theorem 2.1 are satisfied, then a.s. under (2.2),*

$$\sup_{\lambda} |F^{\mathbf{C}_\tau} - F^{\bar{\mathbf{C}}_\tau}| \rightarrow 0 \quad \text{and} \quad \sup_{\lambda} |F^{\bar{\mathbf{C}}_\tau} - F^{\check{\mathbf{C}}_\tau}| \rightarrow 0,$$

where $F^{\mathbf{C}_\tau}, F^{\bar{\mathbf{C}}_\tau}$ and $F^{\check{\mathbf{C}}_\tau}$ denote the ESDs of $\mathbf{C}_\tau, \bar{\mathbf{C}}_\tau$ and $\check{\mathbf{C}}_\tau$, respectively.

PROOF. Let $\eta_{jt}^R = 1 - \mathbb{I}(|Z_{jt}^R| > \sqrt{p})$ and $\eta_{jt}^I = 1 - \mathbb{I}(|Z_{jt}^I| > \sqrt{p})$. Let further $\delta^R = \mathbb{P}(|Z_{jt}^R| > \sqrt{p})$ and $\delta^I = \mathbb{P}(|Z_{jt}^I| > \sqrt{p})$, and note that these quantities are independent due to the assumed i.i.d. structure on Z_{jt} . Since the fourth moments of the latter random variables are assumed finite, it also follows that $\delta^R < p^{-2} \mathbb{E}[|Z_{11}^R|^4] < \infty$ and $\delta^I < p^{-2} \mathbb{E}[|Z_{11}^I|^4] < \infty$.

Observe next that the rank of a matrix does not exceed the number of its nonzero columns and that each nonzero Z_{jt} causes at most $2(q + 1)$ nonzero columns. Recalling that $\mathbf{C}_\tau = (2n)^{-1} \mathbf{X}(\mathbf{L} + \mathbf{L}^*)\mathbf{X}^*$, Theorem A.44 of [5] (using $\mathbf{F} = \mathbf{0}$ and $\mathbf{D} = \mathbf{L} + \mathbf{L}^*$) implies that, for any $\varepsilon > 0$,

$$\begin{aligned} &\mathbb{P}\left(\sup_{\lambda} |F^{\mathbf{C}_\tau}(\lambda) - F^{\bar{\mathbf{C}}_\tau}(\lambda)| > \varepsilon\right) \\ &\leq \mathbb{P}\left(\frac{2(q + 1)}{p} \sum_{jt} (\eta_{jt}^R + \eta_{jt}^I) > \varepsilon\right) \\ &\leq \mathbb{P}\left(\sum_{jt} \eta_{jt}^R > \frac{\varepsilon p}{4(q + 1)}\right) + \mathbb{P}\left(\sum_{jt} \eta_{jt}^I > \frac{\varepsilon p}{4(q + 1)}\right) \end{aligned}$$

$$\begin{aligned}
 &= \mathbb{P}\left(\sum_{jt} \eta_{jt}^R - p(q+n)\delta^R > \frac{\varepsilon p}{4(q+1)} - p(q+n)\delta^R\right) \\
 &\quad + \mathbb{P}\left(\sum_{jt} \eta_{jt}^I - p(q+n)\delta^R > \frac{\varepsilon p}{4(q+1)} - p(q+n)\delta^I\right).
 \end{aligned}$$

Let $\varepsilon(p, q) = \varepsilon p[4(q+1)]^{-1}$. For p large enough so that $2p(q+n) \min\{\delta^R, \delta^I\} < \varepsilon(p, q)$, Hoeffding’s inequality yields

$$\begin{aligned}
 &\mathbb{P}\left(\sum_{jt} \eta_{jt}^R - p(q+n)\delta^R > \varepsilon(p, q) - p(q+n)\delta^R\right) \\
 &\leq 2 \exp(-(\varepsilon(p, q) - p(q+n)\delta^R)^2 (p(q+n)\delta^R + \varepsilon(p, q))^{-1}) \\
 &\leq 2 \exp\left(-\left(\frac{\varepsilon(p, q)}{2}\right)^2 \left(\frac{3\varepsilon(p, q)}{2}\right)^{-1}\right) \\
 &= 2 \exp\left(-\frac{\varepsilon p}{24(q+1)}\right)
 \end{aligned}$$

as well as

$$\mathbb{P}\left(\sum_{jt} \eta_{jt}^I - p(q+n)\delta^I > \varepsilon(p, q) - p(q+n)\delta^R\right) \leq 2 \exp\left(-\frac{\varepsilon p}{24(q+1)}\right).$$

The Borel–Cantelli lemma now implies that $\sup_{\sigma} |F^{C_{\tau}}(\sigma) - F^{\bar{C}_{\tau}}(\sigma)| \rightarrow 0$ a.s., which is the first claim of the proposition.

To verify the second, note that the equality

$$\sum_{\ell=0}^q \mathbf{A}_{\ell} \bar{Z}_{t-\ell} - \sum_{\ell=0}^q \mathbf{A}_{\ell} \check{Z}_{t-\ell} = \sum_{\ell=0}^q \mathbf{A}_{\ell} (\mathbb{E}[\bar{Z}_{11}^R] + i\mathbb{E}[\bar{Z}_{11}^I]) \mathbf{1},$$

with $\mathbf{1}$ being the vector with all entries equal to 1, shows that $\sum_{\ell=0}^q \mathbf{A}_{\ell} \bar{Z}_{t-\ell} - \sum_{\ell=0}^q \mathbf{A}_{\ell} \check{Z}_{t-\ell}$ is independent of t . Thus an application of Lemma S.1 leads to

$$\sup_{\sigma} |F^{\bar{C}_{\tau}}(\sigma) - F^{\check{C}_{\tau}}(\sigma)| \leq \frac{1}{p},$$

which converges to 0 a.s. under (2.2). This is the second assertion. \square

After truncation and centering, it does not necessarily follow that $\mathbb{E}[|\check{Z}_{11}|^2] = \mathbb{E}[|\check{Z}_{11}^R + i\check{Z}_{11}^I|^2]$ is equal to 1. However, rescaling $\check{Z}_{11}^R + i\check{Z}_{11}^I$ by dividing with $(\mathbb{E}[|\check{Z}_{11}^R + i\check{Z}_{11}^I|^2])^{1/2}$ (in order to obtain unit variance) does not affect the LSD because $\mathbb{E}[|\check{Z}_{11}|^2] \rightarrow 1$ under (2.2). The detailed arguments follow as in Section 3.1.1 of Bai and Silverstein [5]. This shows that the symmetrized autocovariances from Z_{jt} and their truncated, centered and rescaled counterparts have the same LSD. Thus to simplify the argument, it can be assumed that the recentered process has variance one.

9. Extension to the non-Gaussian case. In this section, the results for the Gaussian MA(q) case are extended to general innovation sequences $(W_t : t \in \mathbb{Z})$ satisfying the same moment conditions as their Gaussian counterparts $(Z_t : t \in \mathbb{Z})$. The processes of interest are then the two MA(q) processes

$$X_t = \sum_{\ell=0}^q \mathbf{A}_\ell Z_{t-\ell} \quad \text{and} \quad \bar{X}_t = \sum_{\ell=0}^q \mathbf{A}_\ell W_{t-\ell}.$$

Define the symmetrized autocovariance matrix $\bar{\mathbf{C}}_\tau = (2n)^{-1} \sum_{t=1}^{n-\tau} (\bar{X}_t \bar{X}_{t+\tau}^* + \bar{X}_{t+\tau} \bar{X}_t^*)$, the resolvent $\bar{\mathbf{R}}(z) = (\bar{\mathbf{C}}_\tau - z\mathbf{I})^{-1}$ and the Stieltjes transform $\bar{s}_p(z) = p^{-1} \text{tr}[\bar{\mathbf{R}}(z)]$. In the following, it will be shown that the ESDs of \mathbf{C}_τ and $\bar{\mathbf{C}}_\tau$ converge to the same limit. This is done via verifying that, for all $z \in \mathbb{C}^+$, $\bar{s}_p(z)$ and $s_p(z)$ converge to the same limit $s(z)$ under (2.2), which in turn requires us to show that:

- (a) $\mathbb{E}[\bar{s}_p(z) - s_p(z)] \rightarrow 0$ under (2.2) for all $z \in \mathbb{C}^+$;
- (b) $\mathbb{P}[|\bar{s}_p(z) - \mathbb{E}[\bar{s}_p(z)]| \geq \varepsilon] \rightarrow 0$ under (2.2) for all $z \in \mathbb{C}^+$ and $\varepsilon > 0$.

Part (a) requires the use of the Lindeberg principle, and part (b) is achieved via an application of McDiarmid’s inequality.

9.1. *Showing that $\mathbb{E}[\bar{s}_p(z) - s_p(z)] \rightarrow 0$.* For the use in this section, redefine $\mathbf{Z} = [Z_{1-q} : \dots : Z_n]$, define $\mathbf{W} = [W_{1-q} : \dots : W_n]$ and let $\mathbf{Z}^R, \mathbf{W}^R$ and $\mathbf{Z}^I, \mathbf{W}^I$ be the corresponding matrices of real and imaginary parts. Claim (a) will be verified via the Lindeberg principle developed in Chatterjee [8]. This involves successive replacements of Gaussian variables with non-Gaussian counterparts in a telescoping sum. To this end, define an order on the index set $\{(j, t) : j = 1, \dots, p, t = 1 - q, \dots, n\}$ by letting $(j, t) < (j', t')$ if either (1) $t < t'$ or (2) $t = t'$ and $j < j'$, so that one successively moves columnwise through the entries of a matrix. Moreover, let

$$(j, t) - 1 = \begin{cases} (j - 1, t), & \text{if } j > 1 \text{ and } t \geq 1 - q, \\ (p, t - 1), & \text{if } j = 1 \text{ and } t > 1 - q, \\ (0, -q), & \text{if } j = 1 \text{ and } t = 1 - q. \end{cases}$$

Let $\mathbf{V}_{j,t}^R$ denote the $p \times (q + n)$ matrix given by the entries

$$(\mathbf{V}_{jt}^R)_{j't'} = \begin{cases} Z_{j't'}^R, & \text{if } (j', t') \leq (j, t), \\ W_{j't'}^R, & \text{if } (j', t') > (j, t). \end{cases}$$

Further let the $p \times (q + n)$ matrix $\bar{\mathbf{V}}_{jt}^R$ be equal to \mathbf{V}_{jt}^R for all entries but the (j, t) th one, which is set to equal 0, and define analogously the matrices \mathbf{V}_{jt}^I and $\bar{\mathbf{V}}_{jt}^I$. These matrices determine how many of the original Gaussian Z_{jt} ’s have been

replaced by the non-Gaussian W_{jt} . In the following, $s_p(z)$ will be viewed as a function of \mathbf{Z}^R and \mathbf{Z}^I , fixing z and p as parameters, that is, $s_p^z(\mathbf{Z}^R, \mathbf{Z}^I) = s_p(z)$. Similarly, let $s_p^z(\mathbf{W}^R, \mathbf{W}^I) = \bar{s}_p(z)$. Utilizing this notation, the quantity to be bounded in expectation can be written as

$$\begin{aligned} & s_p^z(\mathbf{Z}^R, \mathbf{Z}^I) - s_p^z(\mathbf{W}^R, \mathbf{W}^I) \\ &= s_p^z(\mathbf{V}_{pn}^R, \mathbf{V}_{pn}^I) - s_p^z(\mathbf{V}_{0,-q}^R, \mathbf{V}_{0,-q}^I) \\ &= \sum_{(j,t)=(1,1-q)}^{(p,n)} [s_p^z(\mathbf{V}_{jt}^R, \mathbf{V}_{pn}^I) - s_p^z(\mathbf{V}_{(j,t)-1}^R, \mathbf{V}_{pn}^I)] \\ &\quad + \sum_{(j,t)=(1,1-q)}^{(p,n)} [s_p^z(\mathbf{V}_{0,-q}^R, \mathbf{V}_{jt}^I) - s_p^z(\mathbf{V}_{0,-q}^R, \mathbf{V}_{(j,t)-1}^I)] \\ &= \Delta_1 + \Delta_2, \end{aligned}$$

where $\Delta_1 = \sum_{(j,t)=(1,1-q)}^{(p,n)} \Delta_{j,t}^{(1)}$ and $\Delta_2 = \sum_{(j,t)=(1,1-q)}^{(p,n)} \Delta_{j,t}^{(2)}$ are the real and imaginary part of the difference with $\Delta_{j,t}^{(1)} = s_p^z(\mathbf{V}_{jt}^R, \mathbf{V}_{pn}^I) - s_p^z(\mathbf{V}_{(j,t)-1}^R, \mathbf{V}_{pn}^I)$ and $\Delta_{j,t}^{(2)} = s_p^z(\mathbf{V}_{0,-q}^R, \mathbf{V}_{jt}^I) - s_p^z(\mathbf{V}_{0,-q}^R, \mathbf{V}_{(j,t)-1}^I)$. In the following only the telescoping real parts will be discussed, as the imaginary parts can be estimated along the same lines. Inserting $\bar{\mathbf{V}}_{jt}^R$, one obtains

$$\begin{aligned} \Delta_{j,t}^{(1)} &= [s_p^z(\mathbf{V}_{jt}^R, \mathbf{V}_{pn}^I) - s_p^z(\bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I)] + [s_p^z(\bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I) - s_p^z(\mathbf{V}_{(j,t)-1}^R, \mathbf{V}_{pn}^I)] \\ &= \Delta_{j,t}^{(1,1)} + \Delta_{j,t}^{(1,2)}. \end{aligned}$$

Let $\partial_{j,t,1}^{(k)} s_p^z$ denote the k th-order partial derivative of s_p^z with respect to Z_{jt}^R . A Taylor series expansion gives

$$\begin{aligned} & \left| \Delta_{j,t}^{(1,1)} - Z_{jt}^R \partial_{j,t,1}^{(1)} s_p^z(\bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I) - \frac{1}{2} (Z_{jt}^R)^2 \partial_{j,t,1}^{(2)} s_p^z(\bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I) \right| \\ & \leq \frac{1}{6} |Z_{jt}^R|^3 \max_{\alpha \in [0,1]} |\partial_{j,t,1}^{(3)} s_p^z(\alpha \mathbf{V}_{jt}^R + (1-\alpha) \bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I)| \end{aligned}$$

and

$$\begin{aligned} & \left| \Delta_{j,t}^{(1,2)} + W_{jt}^R \partial_{j,t,1}^{(1)} s_p^z(\bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I) + \frac{1}{2} (W_{jt}^R)^2 \partial_{j,t,1}^{(2)} s_p^z(\bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I) \right| \\ & \leq \frac{1}{6} |W_{jt}^R|^3 \max_{\alpha \in [0,1]} |\partial_{j,t,1}^{(3)} s_p^z(\alpha \mathbf{V}_{(j,t)-1}^R + (1-\alpha) \bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I)|. \end{aligned}$$

The entries of the matrices $\mathbf{Z}^R, \mathbf{Z}^I, \mathbf{W}^R$ and \mathbf{W}^I are all independent of each other and the first and second moments of the various real parts (and imaginary parts)

coincide, so that the bound in the last two inequalities also hold for higher-order terms (HOT). This leads to

$$\begin{aligned}
 (9.1) \quad |\mathbb{E}[\Delta_{j,t}^{(1)}]| &= \left| \mathbb{E} \left[(Z_{jt}^R - W_{jt}^R) \partial_{j,t,1}^{(1)} s_p^z(\bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I) \right. \right. \\
 &\quad \left. \left. + \frac{1}{2} [(Z_{jt}^R)^2 - (W_{jt}^R)^2] \partial_{j,t,1}^{(2)} s_p^z(\bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I) \pm \text{HOT} \right] \right| \\
 &\leq \frac{1}{6} \mathbb{E} \left[|Z_{jt}^R|^3 \max_{\alpha \in [0,1]} |\partial_{j,t,1}^{(3)} s_p^z(\alpha \mathbf{V}_{jt}^R + (1-\alpha) \bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I)| \right] \\
 &\quad + \frac{1}{6} \mathbb{E} \left[|W_{jt}^R|^3 \max_{\alpha \in [0,1]} |\partial_{j,t,1}^{(3)} s_p^z(\alpha \mathbf{V}_{(j,t)-1}^R + (1-\alpha) \bar{\mathbf{V}}_{jt}^R, \mathbf{V}_{pn}^I)| \right].
 \end{aligned}$$

Dealing with the right-hand side of (9.1) requires the computation and estimation of the third-order derivatives $\partial_{j,t,1}^{(3)} s_p^z(\mathbf{Z}^R, \mathbf{Z}^I) = p^{-1} \text{tr}[\partial_{j,t,1}^{(3)} \mathbf{R}(z)]$. Focusing only on the first term of the right-hand side of (9.1) (the second can be handled similarly), Lemma S.14 shows that this term converges to zero under (2.2) if, almost surely under (2.2),

$$\begin{aligned}
 (9.2) \quad \frac{q+1}{n^3 p} \sum_{(j,t)=(1,1-q)}^{(p,n)} \mathbb{E} \left[|Z_{t,j}^R|^3 \max_{\alpha \in [0,1]} \left(\sum_{\ell \in \mathcal{I}_+(t)} \|\mathbf{A}_\ell\| \|\tilde{\mathbf{X}}_{t+\tau+\ell}^\alpha\| \right. \right. \\
 \left. \left. + \sum_{\ell \in \mathcal{I}_-(t)} \|\tilde{\mathbf{X}}_{t-\tau+\ell}^\alpha\| \|\mathbf{A}_\ell\| \right)^3 \right] \rightarrow 0,
 \end{aligned}$$

$$\begin{aligned}
 (9.3) \quad \frac{q+1}{n^2 p} \sum_{(j,t)=(1,1-q)}^{(p,n)} \mathbb{E} \left[|Z_{t,j}^R|^3 \max_{\alpha \in [0,1]} \left(\sum_{\ell \in \mathcal{I}_+(t)} \|\mathbf{A}_\ell\| \|\tilde{\mathbf{X}}_{t+\tau+\ell}^\alpha\| \right. \right. \\
 \left. \left. + \sum_{\ell \in \mathcal{I}_-(t)} \|\tilde{\mathbf{X}}_{t-\tau+\ell}^\alpha\| \|\mathbf{A}_\ell\| \right) \right] \rightarrow 0,
 \end{aligned}$$

where $\mathcal{I}_+(t) = \{\ell : \max\{0, 1-t-\tau\} \leq \ell \leq \min\{q, n-t-\tau\}\}$, $\mathcal{I}_-(t) = \{\ell : \max\{0, 1-t+\tau\} \leq \ell \leq \min\{q, n-t\}\}$, $\tilde{\mathbf{X}}^\alpha = [\tilde{X}_1^\alpha : \dots : \tilde{X}_n^\alpha] = \sum_{\ell=0}^q \mathbf{A}_\ell \mathbf{V}^\alpha \mathbf{L}^\ell \mathbf{F}$, $\mathbf{V}^\alpha = \alpha \mathbf{V}_{jt}^R + (1-\alpha) \bar{\mathbf{V}}_{jt}^R + i \mathbf{Z}^I$ and $\mathbf{F} = [\mathbf{0}_{n \times q} : \mathbf{I}_n]'$ being an $(n+q) \times n$ matrix. The choice of $\alpha \in [0, 1]$ only affects the value of the j th entry of V_t^α . For $t \neq t'$, the notation $V_{t'} = V_{t'}^\alpha$ is therefore preferred. By definition, V_t^0 is the vector whose j th entry has a real part of zero. Let $\mathcal{J}_\pm(t, \ell) = \{\ell' \in \{0, \dots, q\} : \ell' \neq \tau \pm \ell, t \pm \tau + \ell - \ell' \geq 1\}$. Then

$$\begin{aligned}
 &\max_{\alpha \in [0,1]} \|\tilde{\mathbf{X}}_{t \pm \tau + \ell}^\alpha\| \\
 &\leq \sum_{\ell' \in \mathcal{J}_\pm(t, \ell)} \bar{\lambda}_{\mathbf{A}_{\ell'}} \|Z_{t \pm \tau + \ell - \ell'}\| + \bar{\lambda}_{\mathbf{A}_{\ell \pm \tau}} (\|V_t^0\| + |Z_{j,t}^R|).
 \end{aligned}$$

Hence, setting $\bar{\lambda}_\ell^{\text{rem}} = (\sum_{\ell'=|\ell|}^q \bar{\lambda}_{\mathbf{A}_{\ell'}}^2)^{1/2}$,

$$\begin{aligned} & \max_{\alpha \in [0,1]} \sum_{\ell \in \mathcal{J}_\pm(t)} \|\mathbf{A}_\ell\| \|\tilde{X}_{t \pm \tau + \ell}^\alpha\| \\ & \leq \bar{\lambda}_\mathbf{A} \sum_{\ell = -q, \ell \neq \mp \tau}^q \bar{\lambda}_\ell^{\text{rem}} \|Z_{t \pm \tau + \ell}\| + \bar{\lambda}_\mathbf{A} \bar{\lambda}_{\pm \tau}^{\text{rem}} (\|V_t^0\| + |Z_{j,t}^R|). \end{aligned}$$

Using this, it follows from Lemma S.15 that the left-hand sides of (9.2) and (9.3) converge to zero a.s., thus establishing that $\mathbb{E}[\bar{s}_p(z) - s_p(z)] \rightarrow 0$.

9.2. *Showing that $\mathbb{P}(|\bar{s}_p(z) - \mathbb{E}[\bar{s}_p(z)]| \geq \varepsilon) \rightarrow 0$.* For a fixed z , $\bar{s}_p(z)$ is a function of the $n + q$ vectors W_{1-q}, \dots, W_n . Letting $m = \lceil q^{-1}(n + q) \rceil$, these are now segmented into the groups $W_{(k-1)q+1}, \dots, W_{kq}$, $k = 0, 1, \dots, m$, possibly adding additional vectors to the last group to ensure all groups have the same length [even though the value $\bar{s}_p(z)$ does not depend on the additions]. To satisfy the conditions needed in order to apply McDiarmid’s inequality, note that a change of the values $W_{(k-1)q+1}, \dots, W_{kq}$ in one group to, say, $W'_{(k-1)q+1}, \dots, W'_{kq}$, causes the values of $\bar{X}_{(k-1)q+1}, \dots, \bar{X}_{(k+1)q}$ to change to, say, $\bar{X}'_{(k-1)q+1}, \dots, \bar{X}'_{(k+1)q}$. In the following, the focus is on changes applied to the first group of innovations W_1, \dots, W_q . Consider the case $\tau \leq 2q$ and let

$$\bar{\mathbf{C}}'_\tau = \frac{1}{2n} \sum_{t=1}^{2q} (\bar{X}'_t \bar{X}_{t+\tau}^* + \bar{X}_{t+\tau} \bar{X}'_{t,*}) + \frac{1}{2n} \sum_{t=2q+1}^{n-\tau} (\bar{X}_t \bar{X}_{t+\tau}^* + \bar{X}_{t+\tau} \bar{X}_t^*),$$

$\bar{\mathbf{R}}'_\tau(z) = (\bar{\mathbf{C}}'_\tau - z\mathbf{I})^{-1}$ and $\bar{s}'_p(z) = p^{-1} \text{tr}[\bar{\mathbf{R}}'_\tau(z)]$. The goal is now to represent $\bar{\mathbf{C}}'_\tau$ as a finite rank perturbation of $\bar{\mathbf{C}}_\tau$ in the form $\sum_{j=1}^J r_j r_j^*$ with appropriate r_j and J . Write

$$\begin{aligned} \bar{\mathbf{C}}'_\tau - \bar{\mathbf{C}}_\tau &= \frac{1}{2n} \sum_{t=1}^{2q} (\bar{X}'_t \bar{X}_{t+\tau}^* + \bar{X}_{t+\tau} \bar{X}'_{t,*} - \bar{X}_t \bar{X}_{t+\tau}^* - \bar{X}_{t+\tau} \bar{X}_t^*) \\ &= \frac{1}{2n} \sum_{t=1}^{2q} [(\bar{X}'_t + \bar{X}_{t+\tau})(\bar{X}'_t + \bar{X}_{t+\tau})^* \\ & \quad - (\bar{X}_t + \bar{X}_{t+\tau})(\bar{X}_t + \bar{X}_{t+\tau})^* - \bar{X}'_t \bar{X}'_{t,*} + \bar{X}_t \bar{X}_t^*]. \end{aligned}$$

Choosing $J = 8q$ and repeatedly utilizing (S.3) with \mathcal{H}_t replaced by \mathbf{I}_p , it follows that $|\bar{s}_p(z) - \bar{s}'_p(z)| \leq C_1 q (vp)^{-1}$ for some appropriately chosen constant $C_1 > 0$. This bound holds for any of the m groups of innovations. McDiarmid’s inequality consequently implies, for any $\varepsilon > 0$ and a suitable constant $C_2 > 0$,

$$\mathbb{P}(|\bar{s}_p(z) - \mathbb{E}[\bar{s}_p(z)]| \geq \varepsilon) \leq 4 \exp\left(-C_2 \frac{\varepsilon^2 v^2 p^2}{mq^2}\right) \sim 4 \exp\left(-C_2 \frac{\varepsilon^2 v^2 c}{q} p\right).$$

The right-hand side converges to zero at a rate faster than qp^{-1} and concentration of the Stieltjes transform around its mean is established, since the case $\tau > 2q$ can be handled in a similar fashion. Note that the last argument remains valid if $q = q(p) \rightarrow \infty$ as $p \rightarrow \infty$ at a sufficiently slow rate, for example, if $q = o(p^{1/2})$.

10. Extension to the linear process case. Let $(X_t : t \in \mathbb{Z})$ now denote a linear process. To complete the proof of Theorem 2.1, a truncation argument is invoked. Let $(X_t^{\text{tr}} : \mathbb{Z})$ denote the truncated process given by $X_t^{\text{tr}} = \sum_{\ell=0}^{q(p)} \mathbf{A}_\ell Z_{t-\ell}$, $t \in \mathbb{Z}$, where $q(p)$ depends on the dimension. Let $(\check{X}_t : t \in \mathbb{Z})$ be the process given by $\check{X}_t = X_t - X_t^{\text{tr}}$, $t \in \mathbb{Z}$, and denote by $L(F, G)$ the Lévy distance between distribution functions F and G .

LEMMA 10.1. *If the assumptions of Theorem 2.1 are satisfied and if $q(p) = [p]^{1/3}$, then*

$$L(F^{\mathbf{C}_\tau}, F^{\mathbf{C}_\tau^{\text{tr}}}) \rightarrow 0 \quad \text{a.s.}$$

under (2.2), where $\mathbf{C}_\tau^{\text{tr}} = (2n)^{-1} \sum_{\tau=1}^{n-\tau} (X_t^{\text{tr}}(X_{t+\tau}^{\text{tr}})^* + X_{t+\tau}^{\text{tr}}(X_t^{\text{tr}})^*)$.

PROOF. By Lemma S.1, it suffices to show that $p^{-1} \text{tr}[(\tilde{\mathbf{C}}_\tau - \tilde{\mathbf{C}}_\tau^{\text{tr}})^2] \rightarrow 0$ a.s. Write $\mathbf{X} = \mathbf{X}^{\text{tr}} + \check{\mathbf{X}}$ and $\mathbf{D}_\tau = (\mathbf{L}_\tau + \mathbf{L}_\tau^*)/2$. Then

$$\tilde{\mathbf{C}}_\tau - \tilde{\mathbf{C}}_\tau^{\text{tr}} = \frac{1}{n} \mathbf{X}^{\text{tr}} \mathbf{D}_\tau \check{\mathbf{X}}^* + \frac{1}{n} \check{\mathbf{X}} \mathbf{D}_\tau (\mathbf{X}^{\text{tr}})^* + \frac{1}{n} \check{\mathbf{X}} \mathbf{D}_\tau \check{\mathbf{X}}^* = \mathbf{P} + \mathbf{P}^* + \mathbf{Q},$$

say. From repeated applications of the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \text{tr}[(\tilde{\mathbf{C}}_\tau - \tilde{\mathbf{C}}_\tau^{\text{tr}})^2] &= 2 \text{tr}(\mathbf{P}^2) + \text{tr}(\mathbf{Q}^2) + 2 \text{tr}(\mathbf{P}\mathbf{P}^*) + 4 \text{tr}(\mathbf{P}\mathbf{Q}) \\ &\leq 4 \text{tr}(\mathbf{P}\mathbf{P}^*) + \text{tr}(\mathbf{Q}^2) + 4\sqrt{\text{tr}(\mathbf{P}\mathbf{P}^*)}\sqrt{\text{tr}(\mathbf{Q}^2)}. \end{aligned}$$

Since $\|\mathbf{D}_\tau\| \leq 1$, $\text{tr}(\mathbf{Q}^2) \leq n^{-2} \text{tr}[(\check{\mathbf{X}}\check{\mathbf{X}}^*)^2]$, and by another application of Cauchy–Schwarz,

$$\text{tr}(\mathbf{P}\mathbf{P}^*) \leq \left(\frac{1}{n^2} \text{tr}[(\mathbf{X}^{\text{tr}}(\mathbf{X}^{\text{tr}})^*)^2]\right)^{1/2} \left(\frac{1}{n^2} \text{tr}[(\check{\mathbf{X}}\check{\mathbf{X}}^*)^2]\right)^{1/2}.$$

Since it is easy to see that $(pn^2)^{-1} \text{tr}[(\mathbf{X}^{\text{tr}}(\mathbf{X}^{\text{tr}})^*)^2]$ is stochastically bounded (e.g., by showing that the expectation is finite), it is enough to show that $(pn^2)^{-1} \text{tr}[(\check{\mathbf{X}}\check{\mathbf{X}}^*)^2] \rightarrow 0$ a.s. This is established by showing that the sum $\sum_{p=1}^\infty (pn^2)^{-1} \mathbb{E}[\text{tr}[(\check{\mathbf{X}}\check{\mathbf{X}}^*)^2]] < \infty$, and then applying the Borel–Cantelli lemma. To this end, note that

$$\begin{aligned} &\mathbb{E}[\text{tr}[(\check{\mathbf{X}}\check{\mathbf{X}}^*)^2]] \\ &= \sum_{t=1}^n \sum_{s=1}^n \mathbb{E}[|\check{X}_t^* \check{X}_s|^2] \end{aligned}$$

$$\begin{aligned}
 &= \sum_{t=1}^n \sum_{s=1}^n \sum_{\ell=q+1}^{\infty} \sum_{\ell'=q+1}^{\infty} \sum_{m=q+1}^{\infty} \sum_{m'=q+1}^{\infty} \mathbb{E}[\text{tr}(Z_{t-\ell'} Z_{t-\ell}^* \mathbf{A}_\ell \mathbf{A}_m \\
 &\qquad \qquad \qquad \times Z_{s-m} Z_{s-m'} \mathbf{A}_{\ell'} \mathbf{A}_{m'})].
 \end{aligned}$$

It is clear from the independence of Z_t 's that the summands are nonzero only if the indices of Z 's pair up. Direct calculation shows that the total contribution of all four types of pairings: (i) $t - \ell = t - \ell' \neq s - m = s - m'$; (ii) $t - \ell = s - m \neq s - m' = t - \ell'$; (iii) $t - \ell = s - m' \neq s - m = t - \ell'$ and (iv) $t - \ell = t - \ell' = s - m = s - m'$ can be bounded by $C(p^2 n^2) (\sum_{\ell=q}^{\infty} \bar{\lambda}_{\mathbf{A}_\ell}^2)^2$ for some $C > 0$, using the fact that p and n are of the same order. Thus since $q = q(p) = \lceil p \rceil^{1/3}$,

$$\begin{aligned}
 \sum_{p=1}^{\infty} p \left(\sum_{\ell=q}^{\infty} \bar{\lambda}_{\mathbf{A}_\ell}^2 \right)^2 &\leq 2 \sum_{p=1}^{\infty} p \sum_{\ell=q}^{\infty} \left(\bar{\lambda}_{\mathbf{A}_\ell}^2 \sum_{j \geq \ell} \bar{\lambda}_{\mathbf{A}_j}^2 \right) \\
 &= 2 \sum_{\ell=1}^{\infty} \left(\bar{\lambda}_{\mathbf{A}_\ell}^2 \sum_{j \geq \ell} \bar{\lambda}_{\mathbf{A}_j}^2 \right) \sum_{p=1}^{\infty} p \mathbf{1}\{p: \lceil p \rceil^{1/3} \leq \ell\} \\
 &\leq 2 \sum_{\ell=1}^{\infty} \left(\bar{\lambda}_{\mathbf{A}_\ell}^2 \sum_{j \geq \ell} \bar{\lambda}_{\mathbf{A}_j}^2 \right) \sum_{p=1}^{\ell^3} p \\
 &\leq 2 \sum_{\ell=1}^{\infty} \ell^4 \bar{\lambda}_{\mathbf{A}_\ell}^2 \sum_{j \geq \ell} \bar{\lambda}_{\mathbf{A}_j}^2 \leq 2 \sum_{\ell=1}^{\infty} \ell^2 \bar{\lambda}_{\mathbf{A}_\ell}^2 \sum_{j \geq \ell} j^2 \bar{\lambda}_{\mathbf{A}_j}^2 \\
 &\leq 2 \left(\sum_{\ell=1}^{\infty} \ell \bar{\lambda}_{\mathbf{A}_\ell} \right)^4.
 \end{aligned}$$

This proves the result. \square

Using Gaussian innovations $(Z_t : t \in \mathbb{Z})$, let $\tilde{\mathbf{X}}^{\text{tr}} = [X_1^{\text{tr}} : \dots : X_n^{\text{tr}}] = \sum_{\ell=0}^q \mathbf{A}_\ell \mathbf{Z} \tilde{\mathbf{L}}^\ell \mathbf{U}_{\tilde{\mathbf{L}}}$ be the corresponding transformed data matrix. Then define $\psi^{\text{tr}}(\lambda, \nu) = \sum_{\ell=0}^{q(p)} e^{i\ell\nu} f_\ell(\lambda)$, $\psi^{\text{tr}}(\mathbf{A}, \nu) = \sum_{\ell=0}^{q(p)} e^{i\ell\nu} \mathbf{A}_\ell$, $\tilde{\mathbf{C}}_\tau^{\text{tr}} = \gamma_{\tau,t} \tilde{\mathbf{X}}_t^{\text{tr}} (\tilde{\mathbf{X}}_t^{\text{tr}})^*$, $\tilde{\mathbf{R}}_\tau^{\text{tr}} = (\tilde{\mathbf{C}}_\tau^{\text{tr}} - z\mathbf{I})^{-1}$, $\tilde{s}^{\text{tr}}(z) = p^{-1} \text{tr}[\tilde{\mathbf{R}}_\tau^{\text{tr}}(z)]$, $\tilde{K}_p^{\text{tr}}(z, \nu) = p^{-1} \text{tr}[\tilde{\mathbf{R}}_\tau^{\text{tr}}(z) \psi^{\text{tr}}(\mathbf{A}, \nu)]$, $h^{\text{tr}}(\lambda, \nu) = |\psi^{\text{tr}}(\lambda, \nu)|^2$ and $\mathcal{H}^{\text{tr}}(\mathbf{A}, \nu) = \psi^{\text{tr}}(\mathbf{A}, \nu) \psi^{\text{tr}}(\mathbf{A}, \nu)^*$. Then one verifies in a similar vein, as in the proofs of Theorems 7.2 and 7.3, that for all $\omega \in \Omega_0$ with Ω_0 defined in the beginning of Section 7.2, under (2.2),

$$\tilde{K}_{p_\ell}^{\text{tr}}(z, \nu, \omega) - \int [U_{p_\ell}^{\text{tr}}(z, \lambda) - z]^{-1} h^{\text{tr}}(\lambda, \nu) dF_{p_\ell}^{\mathbf{A}}(\lambda) \rightarrow 0,$$

where $U_{p_\ell}^{\text{tr}}(z, \lambda) = (n(p_\ell))^{-1} \sum_{t=1}^{n(p_\ell)} [1 + c_{p_\ell} \gamma_{\tau,t} \tilde{K}_{p_\ell}^{\text{tr}}(z, \nu_{tj}, \omega)]^{-1} h^{\text{tr}}(\lambda, \nu_{tj})$. This is done by exploiting the convergence $h^{\text{tr}}(\lambda, \nu) \rightarrow h(\lambda, \nu)$, which is uniform in ν

and λ , and $U_{p_\ell}^{\text{tr}}(z, \lambda) \rightarrow U(z, \lambda, \omega) = (2\pi)^{-1} \int_0^{2\pi} \beta_\tau(z, \lambda) \gamma_\tau(v) h(\lambda, v) dv$, which is uniform in λ . Therein, $\beta_\tau(z, v) = [1 + c\gamma_\tau(v)K_\tau(z, v)]^{-1}$. From these facts it follows that the limiting version of the truncated version satisfies the defining equations for the Stieltjes kernel (2.5). Therefore the results for the complex Gaussian innovation model with fixed order q are, subject to minor modifications, still applicable when orders $q(p)$ grow at a suitable rate.

11. The real-valued case. The idea of the proof of Theorem 2.2 is motivated by focusing on the MA(1) case. The derivation for the MA(q) and finally MA(∞) cases follows the from the corresponding transformations and subsequent constructions analogous to the complex case. So, let $\mathbf{X} = \mathbf{Z} + \mathbf{A}_1\mathbf{ZL}$ be the data matrix obtained from a Gaussian MA(1) time series, and suppose that \mathbf{A}_1 possesses an eigendecomposition $\mathbf{A}_1 = \mathbf{U}_{\mathbf{A}_1} \mathbf{\Lambda}_{\mathbf{A}_1} \mathbf{U}_{\mathbf{A}_1}^*$ with $\mathbf{U}_{\mathbf{A}_1}$ orthogonal. Let \mathbf{U} denote the real Fourier basis (see, e.g., Chapter 10 of [7]), and let for the real case $\tilde{\mathbf{X}} = \mathbf{U}_{\mathbf{A}_1}(\mathbf{Z} + \mathbf{A}_1\mathbf{Z}\tilde{\mathbf{L}})\mathbf{U}$. Since \mathbf{ZU} and $\mathbf{Z}\tilde{\mathbf{L}}\mathbf{U}$ have independent columns, it follows that $\tilde{\mathbf{X}}$ has independent columns. Moreover, $\tilde{\mathbf{X}}$ has also independent rows. To see this, note that the transpose of the j th row of $\tilde{\mathbf{X}}$ is

$$(\mathbf{U}_{\mathbf{A}_1}(\mathbf{Z} + \mathbf{A}_1\mathbf{Z}\tilde{\mathbf{L}})\mathbf{U})^T e_j = \mathbf{U}^T (\mathbf{I} + \lambda_j \tilde{\mathbf{L}}) Z_j,$$

where Z_j is the j th column of $\mathbf{U}_{\mathbf{A}_1}\mathbf{ZU}$ and λ_j the j th eigenvalue of \mathbf{A}_1 . The covariance of the j th column is

$$\mathbb{E}[\mathbf{U}^T (\mathbf{I} + \lambda_j \tilde{\mathbf{L}}) Z_j Z_j^T (\mathbf{I} + \lambda_j \tilde{\mathbf{L}})^T \mathbf{U}] = \mathbf{U}^T (\mathbf{I} + \lambda_j (\tilde{\mathbf{L}} + \tilde{\mathbf{L}}^T) + \lambda_j^2 \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T) \mathbf{U}.$$

Since $\mathbf{I} + \lambda_j (\tilde{\mathbf{L}} + \tilde{\mathbf{L}}^T) + \lambda_j^2 \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T = (1 + \lambda_j^2) \mathbf{I} + \lambda_j (\tilde{\mathbf{L}} + \tilde{\mathbf{L}}^T)$ is a symmetric circulant matrix, it diagonalizes in the real Fourier basis, and hence the covariance matrix of the last display is diagonal. From the same display it follows also that the variance of the (j, t) th entry is $h(\lambda_j, v_t)$, so that the rest of the proof follows as in the complex case (Theorem 2.1).

SUPPLEMENTARY MATERIAL

Supplement to “On the Marčenko–Pastur law for linear time series” (DOI: 10.1214/14-AOS1294SUPP; .pdf). The supplementary material provides additional technical lemmas and their proofs.

REFERENCES

[1] AGUILAR, O. and WEST, M. (2000). Bayesian dynamic factor models and portfolio allocation. *J. Bus. Econom. Statist.* **18** 338–357.
 [2] ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices*. Cambridge Univ. Press, Cambridge. MR2760897
 [3] ANGELINI, M., BAÑBURA, M. and RÜNSTLER, G. (2010). Estimating and forecasting the Euro area monthly national accounts from a dynamic factor model. *J. Bus. Cycle Meas. Anal.* **2010** 1–22.

- [4] BAI, J. and NG, S. (2007). Determining the number of primitive shocks in factor models. *J. Bus. Econom. Statist.* **25** 52–60. [MR2338870](#)
- [5] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. [MR2567175](#)
- [6] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. [MR2279680](#)
- [7] BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York. [MR1093459](#)
- [8] CHATTERJEE, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. [MR2294976](#)
- [9] EL KAROUI, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19** 2362–2405. [MR2588248](#)
- [10] FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000). The generalized factor model: Identification and estimation. *Rev. Econ. Stat.* **82** 540–554.
- [11] FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2004). The generalized dynamic factor model: Consistency and rates. *J. Econometrics* **119** 231–255. [MR2057100](#)
- [12] FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *J. Amer. Statist. Assoc.* **100** 830–840. [MR2201012](#)
- [13] FORNI, M. and LIPPI, M. (1999). Aggregation of linear dynamic microeconomic models. *J. Math. Econom.* **31** 131–158. [MR1676417](#)
- [14] FORNI, M. and LIPPI, M. (2001). The generalized dynamic factor model: Representation theory. *Econometric Theory* **17** 1113–1141. [MR1867540](#)
- [15] GERONIMO, J. S. and HILL, T. P. (2003). Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform. *J. Approx. Theory* **121** 54–60. [MR1962995](#)
- [16] HACHEM, W., LOUBATON, P. and NAJIM, J. (2005). The empirical eigenvalue distribution of a Gram matrix: From independence to stationarity. *Markov Process. Related Fields* **11** 629–648. [MR2191967](#)
- [17] HACHEM, W., LOUBATON, P. and NAJIM, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. [MR2326235](#)
- [18] HALLIN, M. and LIŠKA, R. (2007). Determining the number of factors in the general dynamic factor model. *J. Amer. Statist. Assoc.* **102** 603–617. [MR2325115](#)
- [19] HAMILTON, J. D. (1994). *Time Series Analysis*. Princeton Univ. Press, Princeton, NJ. [MR1278033](#)
- [20] 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](#)
- [21] JIN, B., WANG, C., MIAO, B. and LO HUANG, M.-N. (2009). Limiting spectral distribution of large-dimensional sample covariance matrices generated by VARMA. *J. Multivariate Anal.* **100** 2112–2125. [MR2543090](#)
- [22] JOHNSTONE, I. M. (2007). High dimensional statistical inference and random matrices. In *International Congress of Mathematicians. Vol. I* 307–333. Eur. Math. Soc., Zürich. [MR2334195](#)
- [23] LIU, H. (2013). Spectral analysis of high dimensional time series. Ph.D. thesis, Univ. California, Davis. [MR3232212](#)
- [24] LIU, H., AUE, A. and PAUL, D. (2014). Supplement to “On the Marčenko–Pastur law for linear time series.” DOI:10.1214/14-AOS1294SUPP.
- [25] LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin. [MR2172368](#)

- [26] MARČENKO, V. and PASTUR, L. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR. Sbornik* **1** 457–483.
- [27] MOLENAAR, P. C. M. (1994). A manifesto on psycholog as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement* **2** 201–218.
- [28] NG, V., ENGEL, R. F. and ROTHSCILD, M. (1992). A multi-dynamic-factor model for stock returns. *J. Econometrics* **52** 245–266.
- [29] ONATSKI, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *J. Econometrics* **168** 244–258. [MR2923766](#)
- [30] PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. [MR3206718](#)
- [31] PAUL, D. and SILVERSTEIN, J. W. (2009). No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix. *J. Multivariate Anal.* **100** 37–57. [MR2460475](#)
- [32] PFAFFEL, O. and SCHLEMM, E. (2011). Eigenvalue distribution of large sample covariance matrices of linear processes. *Probab. Math. Statist.* **31** 313–329. [MR2853681](#)
- [33] STOCK, J. H. and WATSON, M. W. (2005). Implications of dynamic factor models for VAR analysis. NBER Working Paper No. 11467.
- [34] TAO, T. (2012). *Topics in Random Matrix Theory. Graduate Studies in Mathematics* **132**. Amer. Math. Soc., Providence, RI. [MR2906465](#)
- [35] YAO, J. (2012). A note on a Marčenko–Pastur type theorem for time series. *Statist. Probab. Lett.* **82** 22–28. [MR2863018](#)
- [36] ZHANG, L. (2006). Spectral analysis of large dimensional random matrices. Ph.D. thesis, National Univ. Singapore.

H. LIU
HAAS SCHOOL OF BUSINESS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: liuhy@berkeley.edu

A. AUE
D. PAUL
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95656
USA
E-MAIL: aaue@ucdavis.edu
debpaul@ucdavis.edu