The Annals of Statistics
2016, Vol. 44, No. 2, 598–628
DOI: 10.1214/15-AOS1378

LARGE SAMPLE BEHAVIOUR OF HIGH DIMENSIONAL AUTOCOVARIANCE MATRICES

BY MONIKA BHATTACHARJEE AND ARUP BOSE[1]

Indian Statistical Institute

The existence of limiting spectral distribution (LSD) of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$, the symmetric sum of the sample autocovariance matrix $\hat{\Gamma}_u$ of order u, is known when the observations are from an infinite dimensional vector linear process with appropriate (strong) assumptions on the coefficient matrices. Under significantly weaker conditions, we prove, in a unified way, that the LSD of any symmetric polynomial in these matrices such as $\hat{\Gamma}_u + \hat{\Gamma}_u^*$, $\hat{\Gamma}_u \hat{\Gamma}_u^*$, $\hat{\Gamma}_u \hat{\Gamma}_u^* + \hat{\Gamma}_k \hat{\Gamma}_k^*$ exist. Our approach is through the more intuitive algebraic method of free probability in conjunction with the method of moments. Thus, we are able to provide a general description for the limits in terms of some freely independent variables. All the previous results follow as special cases. We suggest statistical uses of these LSD and related results in order determination and white noise testing.

1. Introduction. Multivariate linear time series models such as the Autoregressive Moving Average (ARMA) processes are fundamental in the theory of econometrics and finance. Moreover, time series data where the dimension grows along with the sample size are becoming increasingly frequent. A key model in these situations is the infinite dimensional moving average process of order infinity, MA(∞), where the sample $\{X_{t.p}^{(n)} : t = 1, 2, \ldots, n\}$ of size n satisfies

$$X_{t.p}^{(n)} = \sum_{j=0}^{\infty} \psi_{j.p}^{(n)} \varepsilon_{t-j.p} \qquad \forall t, n \geq 1 \text{ (almost surely)}. \tag{1.1}$$

For all t, $X_{t.p}^{(n)}$ and $\varepsilon_{t.p}$ are p-dimensional vectors and $\psi_{j.p}^{(n)}$ are $p \times p$ *coefficient matrices* and $\psi_{0.p}^{(n)} = I_p$. Precise assumptions of independence, finiteness of moments and conditions on the matrices are discussed later.

We work in the framework of the particular high dimensional model where the dimension p increases proportionately with the sample size n, so that $p = p(n) \to$

Received June 2015; revised August 2015.

[1]Supported by J. C. Bose National Fellowship, Department of Science and Technology, Government of India.

MSC2010 subject classifications. Primary 62M10; secondary 37M10, 58C40, 46L54, 62J20.

Key words and phrases. Infinite dimensional vector linear process, symmetrized autocovariance matrices, limiting spectral distribution, Wigner matrix, ID matrix, moment method, semi-circle law, asymptotically free, non-crossing partitions, non-commutative probability space, $*$-algebra, free cumulants, compound free Poisson, Stieltjes transformation.

∞ and $\frac{p}{n} \to y \in (0, \infty)$. The infinite sum in (1.1) exists in the almost sure sense under suitable decay conditions on $\{\psi_{j.p}^{(n)}\}$. If $\psi_{j.p}^{(n)} = 0, \forall j > q$, then it will be called an MA(q) process. For convenience, we will write p for $p(n)$ and ψ_j, ε_t and X_t, respectively, for $\psi_{j.p}^{(n)}, \varepsilon_{t.p}$ and $X_{t.p}^{(n)}$. Many researchers have worked on this model recently. See, for example, Forni et al. (2000, 2004), Forni and Lippi (2001), Bhattacharjee and Bose (2014), Jin et al. (2014) and Liu, Aue and Paul (2015).

One of the key quantities in time series analysis is the autocovariance matrix. The *population autocovariance matrices* are defined as

$$\Gamma_{u.p} := E\big(X_{t.p} X_{(t+u).p}^*\big) = \sum_{j=1}^{\infty} \psi_j \psi_{j+u}^*, \qquad u = 0, 1, \ldots.$$

The moment estimator of $\Gamma_{u.p}$ is the *sample autocovariance matrix*,

$$\hat{\Gamma}_{u.p} = \frac{1}{n} \sum_{t=1}^{n-u} X_{t.p} X_{(t+u).p}^*, \qquad 0 \le u \le n-1. \tag{1.2}$$

We often write Γ_u and $\hat{\Gamma}_u$, respectively, for $\Gamma_{u.p}$ and $\hat{\Gamma}_{u.p}$. Our goal is to study the large sample behaviour of the random matrices $\hat{\Gamma}_u$ and use the asymptotic results for statistical inference purposes such as order determination of infinite dimensional moving average or autoregressive processes. Since we are dealing with several matrices of increasing dimension together, we need to give precise meaning to the large sample behaviour in our context.

The most common way to capture the large sample behaviour of a sequence of random matrices is through its spectral distribution. The empirical spectral distribution (ESD) of an $n \times n$ (random) matrix R_n is the (random) probability distribution with mass $1/n$ at each of its eigenvalues. If it converges weakly (almost surely) to a (non-degenerate) probability distribution, then the latter is called the *limiting spectral distribution* (LSD) of R_n. Incidentally, the study of the limit spectrum of non-Hermitian matrices is extremely difficult and very few results are known for general non-Hermitian sequences. Researchers have concentrated on the additive symmetrized version $\hat{\Gamma}_u + \hat{\Gamma}_u^*$, one at a time.

One widely used approach to establish the LSD is that of Stieltjes transformation, which for any finite measure μ on the real line equals

$$m_\mu(z) = \int \frac{1}{x - z} \mu(dx), \qquad z \in \mathbb{C}^+ := \{x + iy : x \in \mathbb{R}, y > 0\}. \tag{1.3}$$

Pointwise convergence of Stieltjes transforms to a Stieltjes transform implies the convergence of the corresponding distributions. In random matrix theory, this convergence is proved by linking the Stieltjes transform of the ESD to the resolvent and showing convergence by martingale convergence methods. See, for example, Silverstein and Bai (1995), Silverstein (1995), Bai and Zhou (2008) and Bai and

Silverstein (2010). All existing works regarding LSD of autocovariance matrices are based on this method.

Let us first discuss briefly these results. Consider the simplest case of (1.1) where $X_t = \varepsilon_t = (\varepsilon_{t,1}, \varepsilon_{t,2}, \dots, \varepsilon_{t,p})^T$ and $\{\varepsilon_{t,j}\}$ are i.i.d. with mean zero and variance one and with enough high moments. Then $\hat{\Gamma}_0$ is nothing but the unadjusted sample variance–covariance matrix, and it is well known that its LSD is the Marčenko–Pastur law. See, for example, Marčenko and Pastur (1967) and Bai and Silverstein (2010). For the same model, Jin et al. (2014) showed that the LSD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ exist and are equal for every $u \geq 1$ and derived its Stieltjes transform.

Pfaffel and Schlemm (2011) and Yao (2012) derived the limiting Stieltjes transformation of $\{\hat{\Gamma}_u + \hat{\Gamma}_u^*\}$, $u \geq 0$, when the components of X_t are independent samples from an identical univariate MA(q) process. Liu, Aue and Paul (2015) appears to be the only work in model (1.1) for arbitrary q. For each u, they established the existence of the LSD of $\{\hat{\Gamma}_u + \hat{\Gamma}_u^*\}$ and derived its Stieltjes transform as a solution of a pair of functional equations. To derive this result, they assumed that $\{\varepsilon_{t,j}\}$ are i.i.d. with finite 4th moment.

Their assumptions on $\{\psi_j\}$ are, however, quite restrictive. They assumed that $\{\psi_j\}$ are Hermitian and simultaneously diagonalizable (the latter assumption can be replaced by the assumption that $\{\psi_j\}$ are Toeplitz matrices with suitable decay conditions on their entries). Even so, this excludes many interesting linear processes (such as model 4 in Section 4.1). To indicate another limitation of this assumption, suppose further that $\varepsilon_t \sim \mathcal{N}(0, I)$. Let U be a unitary matrix such that $U\psi_j U^* =: \Lambda_j$ (say) are diagonal matrices. Since $U\varepsilon_t$ and ε_t are identically distributed and $UU^* = I$, as far as the LSD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ is concerned, (1.1) is equivalent to the model

$$X_{t,i} \text{ (ith component of X_t)} = \sum_{j=0}^{\infty} \psi_{j,(i,i)} \varepsilon_{t-j,i} \qquad \forall i \geq 1, \tag{1.4}$$

where $\Lambda_j = \operatorname{diag}(\psi_{j,(1,1)}, \psi_{j,(2,2)}, \dots, \psi_{j,(p,p)})$ for every j. Hence, this model does not exhibit spatial dependence or dependence among the components.

Our approach differs from the existing approaches in many ways. First, we do away with the Hermitian and simultaneously diagonalizable condition and replace it with a more natural and much weaker joint convergence assumption [assumption (A3) in Section 3]. Second, all the existing works concentrate on $\hat{\Gamma}_u + \hat{\Gamma}_u^*$. If we wish to study the *singular values* of $\hat{\Gamma}_u$, we need to consider the symmetric product $\hat{\Gamma}_u \hat{\Gamma}_u^*$. This gives rise to a completely different LSD problem. Indeed, one may consider more general symmetrizations that involve several $\hat{\Gamma}_u$. As we may recall, in the one-dimensional case, all tests for white noise are based on quadratic functions of autocovariances. See, for example, Hong and Lee (2003), Shao (2011) and Xiao and Wu (2014). The analogous objects in our model are quadratic polynomials in autocovariances. Thus, we are naturally led to the consideration of matrix polynomials of autocovariances. While it is conceivable that the Stieltjes transform method can be potentially used to tackle these cases, it seems to be rather cumbersome and needlessly lengthy to do so and shall at best be a case by case study. We provide a unified method to study the LSD of symmetric polynomials of the autocovariance matrices. We do not use Stieltjes transforms at all except to cross-check our results with the existing results, all of which follow as special cases.

To obtain the LSD, we use the method of moments. The hth order moment of the ESD of an $n \times n$ real symmetric matrix R_n equals $\beta_h(R_n) := \frac{1}{n}\operatorname{Tr}(R_n^h)$. Consider the following conditions:

(M1) For every $h \geq 1$, $E(\beta_h(R_n)) \to \beta_h$,
(M4) $\sum_{n=1}^{\infty} E(\beta_h(R_n) - E(\beta_h(R_n)))^4 < \infty, \forall h \geq 1$, and
(C) The sequence $\{\beta_h\}$ satisfies Carleman's condition, $\sum_{h=1}^{\infty} \beta_{2h}^{-1/(2h)} = \infty$.

If (M1), (M4) and (C) hold, then ESD of R_n converges almost surely to the distribution F determined uniquely by the moments $\{\beta_h\}$. (M1) is the most crucial condition in this method as it identifies the moments of the LSD.

In Theorem 3.1, we claim the existence of the LSD of any symmetric polynomial in $\{\hat{\Gamma}_u\}$ in model (1.1) and describe the limit in terms of a polynomial of some free variables. To establish (M1), we use tools from non-commutative free probability theory (the next section and Sections 3, 5.1, 5.2 and 5.3 contain the necessary background). Free variables in the non-commutative world are the analogue of independent random variables in the commutative world. As matrices are non-commutative objects, appearance of non-commutative probability spaces is not surprising. The reason for the appearance of free variables is more subtle (see the discussion at the beginning of Section 3). In Section 4.1, we provide simulation results for specific choices of the model. These simulations support the conclusion of Theorem 3.1. Based on simulations, we also conjecture that the LSD exists for the non-Hermitian matrices $\hat{\Gamma}_u$.

It is natural to anticipate that the sample autocovariance matrices will play an increasingly crucial role in the statistical analysis of these models. This seems to be at a rudimentary stage currently, but we anticipate further thrust as the limiting structure of these matrices is uncovered. Liu (2013) estimated the spectrum of the coefficient matrices by minimizing some distance between Stieltjes transformations of the ESD and the LSD of $\{\hat{\Gamma}_u + \hat{\Gamma}_u^*\}$ in some appropriately chosen space of distribution functions. In Sections 4.2.1 and 4.2.2, we use the LSD results to provide a graphical method to determine the order of a moving average and an autoregressive process. Following a suggestion by one of the referees, in Section 4.2.3, we discuss the asymptotic distribution of the trace of sample autocovariance matrices. As a by-product of the calculations used in the derivation of the LSD results, we conclude that these traces have asymptotic normal distributions. This can be used to test simple null and alternative hypotheses for model (1.1).

Section 5 contains the outline of the proofs. Further details of the proofs are available in the supplementary file Bhattacharjee and Bose (2015).

2. Some notions from free probability. To aid the reader, we first highlight the essential notions from free probability that are needed to understand our main theorem. Further concepts and facts, as needed later, are discussed at the beginning of Section 3 and in Sections 5.1, 5.2 and 5.3. An excellent reference for all the details is Nica and Speicher (2006).

Commutative random variables are attached to a probability space $(\mathcal{S}, E)$, which consists of a σ-field $\mathcal{S}$ and an expectation operator E. Similarly, non-commutative variables are attached to some non-commutative $*$-probability space (NCP) $(\mathcal{A}, \varphi)$ consisting of a unital $*$-algebra $\mathcal{A}$ and a unital linear functional (called a state) $\varphi : \mathcal{A} \to \mathbb{C}$, $\varphi(1_{\mathcal{A}}) = 1$. Thus, φ is the analogue of the expectation operator. The elements of $\mathcal{A}$ are called (non-commutative random) variables. The canonical example of NCP that we will need is $\mathcal{M}_d$, the space of all $d \times d$ matrices with the state φ as the average trace. If the matrix has random entries, we modify φ by taking its usual expectation.

In the commutative case, random variables (say with bounded support) are independent if and only if all joint moments obey the product rule. It is well known that the cumulants and moments are related via the Möbius transformation on the partially ordered set (POSET) of *all partitions*. Using this, it can be shown that independence is also equivalent to the vanishing of all mixed cumulants.

For a set of non-commutative variables $\{a_i\}_{i \geq 1}$, the set of all joint moments is defined as $\{\varphi(\Pi(a_i, a_i^* : i \geq 1)) : \Pi \text{ polynomials}\}$ and is known as the distribution of $\{a_i\}$. Here, we have the notion of joint cumulants, called *free cumulants*. These can be uniquely obtained from the above moments and vice versa via a different Möbius transformation and its inverse on the POSET of all *non-crossing partitions*. Non-commutative variables are said to be *free* (freely independent) if and only if all their mixed free cumulants vanish.

A consequence of freeness is that all joint moments of free variables are computable in terms of the moments of the individual variables. Of course, the algorithm for computing moments under freeness is different from (and more complicated than) the product rule under usual independence. The notion of freness of variables extends to freeness of sub-algebras in the natural way. Now consider NCPs $\{(\mathcal{A}_u, \varphi_u)\}_{1 \leq u \leq r}$. Then, analogous to the product space in the commutative case, we can have $(\mathcal{A}, \varphi)$, the *free product* of $\{(\mathcal{A}_u, \varphi_u)\}$ so that the restriction of φ on $\mathcal{A}_u$ is φ_u and $\mathcal{A}_u$ are free sub-algebras of $\mathcal{A}$.

While matrices are seldom free, there is a large class of matrices that are free in an asymptotic sense (which is made precise in Section 5.1) as the dimension increases. For example, if W_1 and W_2 are $n \times n$ independent symmetric matrices with all entries i.i.d. whose all moments are finite, then they are asymptotically free. Using such asymptotic freeness, we shall be able to compute the limits of required traces by using tools from free probability. This will help us to establish the (M1) condition and in the bargain also provide us with expressions for the limits in terms of free variables.

3. Main result. Consider the following assumptions on the driving process $\{\varepsilon_t\}$ and the coefficient matrices $\{\psi_j\}$:

(A1) $\{\varepsilon_{t,j}\}$ are independent with $E(\varepsilon_{t,j}) = 0$ and $E|\varepsilon_{t,j}|^2 = 1, \forall i, j$.

(A2) $\sup_{t,j} E(|\varepsilon_{t,j}|^k) < C_k < \infty, \forall k \geq 1$ or, for some sequence $\eta_n \downarrow 0$, $|\varepsilon_{t,j}| < \eta_n \sqrt{n}, \forall i, j$.

(A3) $\{\psi_j\}$ are compactly supported and for any polynomial Π in $\{\psi_j, \psi_j^*\}$, $\lim p^{-1}\operatorname{Tr}(\Pi)$ exists and is finite.

Later we shall relax assumption (A2).

To see how freeness comes into the picture, and hence how it motivates the statement of our main theorem, let us focus on $\hat{\Gamma}_0$ when

$$X_t = \varepsilon_t + \psi_1 \varepsilon_{t-1.}$$

Let $Z = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)_{p \times n}$ be the independent (ID) matrix. For $i \geq 0$, let P_i be the $n \times n$ matrix whose ith upper diagonal is 1 and 0 otherwise. Note that $P_0 = I_n$. For $i < 0$, let $P_i = P_{-i}^T$ be the transpose of P_{-i}. Note that

$$\begin{aligned}
\hat{\Gamma}_0 &= n^{-1} \sum_{t=1}^{n} (\varepsilon_t + \psi_1 \varepsilon_{t-1})(\varepsilon_t + \psi_1 \varepsilon_{t-1})^* \\
&= n^{-1}\big(Z P_0 Z^* + \psi_1 Z P_0 Z^* \psi_1^* + \psi_1 Z P_1 Z^* + Z P_{-1} Z^* \psi_1^*\big) + R_n \\
&= \Delta_0 + R_n \qquad \text{(say)}.
\end{aligned}$$

By Lemma 7.1 of the supplementary file Bhattacharjee and Bose (2015), $\hat{\Gamma}_0$ and Δ_0 have identical LSD. Thus, our primary goal is to show that for all $r \geq 1$, $\lim p^{-1} E \operatorname{Tr}(\Delta_0^r)$ exists. To achieve this, we first define an NCP generated by these matrices. However, the matrices Z, $\{I_p, \psi_1\}$ and $\{P_0, P_1, P_{-1}\}$ are all of different orders. Therefore, we embed these matrices into larger square matrices of order $(n + p)$. We embed Z into a Wigner[2] matrix W of order $(n + p)$. Thus,

$$W = \begin{pmatrix} W^{(1)} & Z \\ Z^* & W^{(2)} \end{pmatrix}, \tag{3.1}$$

where $W^{(1)}$ and $W^{(2)}$ are two independent Wigner matrices of order p and n respectively and also independent of Z and whose entries satisfy assumption (A2). For any matrices B and D of order p and n, respectively, let $\bar{B}$ and $\underline{D}$ of order $(n + p)$ be the matrices

$$\bar{B} = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}, \qquad \underline{D} = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix}. \tag{3.2}$$

[2]A Wigner matrix is a square symmetric random matrix with independent mean 0 variance 1 entries on and above the diagonal.

Note that for any integer r, if the right-hand side limits below exist, then

$$\lim(n+p)^{-1}\operatorname{Tr}(\bar{B}^r) = y(1+y)^{-1}\lim p^{-1}\operatorname{Tr}(B^r), \tag{3.3}$$

$$\lim(n+p)^{-1}\operatorname{Tr}(\underline{D}^r) = (1+y)^{-1}\lim n^{-1}\operatorname{Tr}(D^r) \qquad \text{and} \tag{3.4}$$

$$\lim p^{-1}\operatorname{Tr}(\Delta_0^r) = y^{-1}(1+y)\lim(n+p)^{-1}\operatorname{Tr}(\bar{\Delta}_0^r). \tag{3.5}$$

On the other hand,

$$n\bar{\Delta}_0 = \bar{I}_p W \underline{P}_0 W \bar{I}_p + \bar{\psi}_1 W \underline{P}_0 W \bar{\psi}_1^* + \bar{\psi}_1 W \underline{P}_1 W \bar{I}_p + \bar{I}_p W \underline{P}_{-1} W \bar{\psi}_1^*.$$

Thus, $\bar{\Delta}_0^r$ involves polynomials in these matrices. So it is a question of computing the limiting trace of such polynomials. Now observe that for any monomial m:

(1) $\lim p^{-1}\operatorname{Tr}(m(P_0, P_1, P_{-1}))$ exists and can be computed easily.
(2) under assumption (A3), $\lim n^{-1}\operatorname{Tr}(m(\bar{I}_p, \bar{\psi}_1, \bar{\psi}_1^*))$ exists.
Moreover, from random matrix theory it is well known that
(3) if (A1) and (A2) hold then $\lim E\operatorname{Tr}((n+p)^{-1/2}W)^r = E(s^r)$, where s is a standard semi-circle variable with moments

$$\varphi(s^k) = \begin{cases} \dfrac{k!}{(k/2)!(k/2+1)!}, & \text{if } k \text{ is even,} \\ 0, & \text{if } k \text{ is odd.} \end{cases} \tag{3.6}$$

(4) Finally, results from free probability guarantee that *in the limit*, the matrices $(n+p)^{-1/2}W$, $\{\bar{I}_p, \bar{\psi}_1\}$ and $\{\underline{P}_0, \underline{P}_1, \underline{P}_{-1}\}$ are free variables say $s, \{a_0, a_1\}$ and $\{c_0, c_1, c_{-1}\}$ where $c_1^* = c_{-1}$ in some NCP $(\mathcal{A}, \varphi)$.

Thus, using the above conclusions (1), (2) and (3) in conjunction with equations (3.3), (3.4), (3.5) and (3.6), we can conclude that $\lim p^{-1}\operatorname{Tr}(\Delta_0^r)$ exists and

$$\lim p^{-1}\operatorname{Tr}(\Delta_0^r) = y^{-1}(1+y)\varphi\biggl((1+y)\sum_{j,j'=0,1} a_j s c_{j-j'} s a_j^*\biggr)^r. \tag{3.7}$$

The factor $(1+y)$ within φ is the adjustment needed for the replacement of $Z/\sqrt{n}$ by $W/\sqrt{n+p}$. The right-hand side of the above equation, involving free variables, are then the moments of the LSD of $\hat{\Gamma}_0$.

This is the idea we implement in the general MA(q) process and for general symmetric polynomials of the autocovariances. Now we have q coefficient matrices $\{\psi_j\}$ and $\{P_i : i = 0, \pm 1, \pm 2, \ldots\}$. To describe the limit, consider the NCP $(\mathcal{A}, \varphi)$, the *free product* of the semi-circle variable s, $\{a_j\}$ and $\{c_i\}$ such that $\varphi(s^k)$ is given by (3.6) and for any finite monomial m, we have

$$\varphi\bigl(m(c_i, c_{-i} : i \geq 0)\bigr) = (1+y)^{-1} I(J=0), \tag{3.8}$$

where J is sum of the subscripts of $\{c_i\}$ which appear in m [the right-hand side of (3.8) equals $\lim n^{-1}\operatorname{Tr}(m(P_i, P_{-i} : i \geq 0))$ and can be checked by direct calculation], and

$$\varphi\bigl(m(a_j, a_j^* : i \geq 0)\bigr) = \frac{y}{1+y}\lim\frac{1}{p}\operatorname{Tr}\bigl(m(\psi_j, \psi_j^* : j \geq 0)\bigr) \tag{3.9}$$

[the right-hand side of (3.9) exists by assumption (A3)].

Let us define for all $u = 0, 1, 2, \dots,$

$$\gamma_{uq} = (1+y)\sum_{j,j'=0}^{q} a_j s c_{j'-j+u} s a_{j'}^*,$$
$$\gamma_{uq}^* = (1+y)\sum_{j,j'=0}^{q} a_{j'} s c_{-j'+j-u} s a_j^*. \tag{3.10}$$

Then we have the following theorem, the proof of which is given in Section 5.2.

THEOREM 3.1. *Suppose* $X_t \sim \mathrm{MA}(q)$, $q < \infty$ *and* (A1), (A2), (A3) *and* $pn^{-1} \to y \in (0, \infty)$ *hold. Then the LSD of any symmetric polynomial* $\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)$ *exists and the limit moments are given by*

$$\lim p^{-1} E\,\mathrm{Tr}\big(\Pi\big(\hat{\Gamma}_u, \hat{\Gamma}_u^* : i \geq 0\big)\big) = y^{-1}(1+y)\varphi\big(\Pi\big(\gamma_{uq}, \gamma_{uq}^* : i \geq 0\big)\big). \tag{3.11}$$

For particular symmetric polynomials, the LSD exist under relaxed moment assumptions. In the next remark, we consider the LSD of $\{\hat{\Gamma}_u + \hat{\Gamma}_u^*\}$ and $\{\hat{\Gamma}_u \hat{\Gamma}_u^*\}$. Its proof, given in Section 5 of the supplementary file Bhattacharjee and Bose (2015), is based on the same truncation arguments as in Jin et al. (2014) after some necessary modifications.

REMARK 3.1. Suppose $X_t \sim \mathrm{MA}(q)$, $q < \infty$, and (A1), (A3) and $pn^{-1} \to y \in (0, \infty)$ hold. Then the following hold true:

(a) For each $0 \leq u < \infty$, LSD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ exists if for some $\delta \in (0, 2]$,
 (A4) $\sup_{t,j} E(|\varepsilon_{t,j}|^{2+\delta}) < M < \infty$, and
 (A5) for any $\eta > 0$, $\frac{1}{\eta^{2+\delta} np} \sum_{j=1}^{p} \sum_{t=1}^{n} E(|\varepsilon_{t,j}|^{2+\delta} I(|\varepsilon_{t,j}| > \eta n^{1/(2+\delta)})) \to 0$.

(b) For each $0 \leq u < \infty$, LSD of $\hat{\Gamma}_u \hat{\Gamma}_u^*$ exists if (A5) holds and
 (A6) $\sup_{t,j} E|\varepsilon_{t,j}|^4 < M < \infty$.

(c) Suppose $X_t \sim \mathrm{MA}(0)$ process and assumptions (A1), (A3) hold. Then the existence of the LSD of $\hat{\Gamma}_0$, under assumption
 (A7) For any $\eta > 0$, $\eta^{-2}(np)^{-1} \sum_{j=1}^{p} \sum_{t=1}^{n} E(|\varepsilon_{t,j}|^2 I(|\varepsilon_{t,j}| > \eta\sqrt{n})) \to 0$,
 is well known [see Bai and Silverstein (2010)]. Jin et al. (2014) established the existence of the LSD of any $\hat{\Gamma}_u + \hat{\Gamma}_u^*$, $u \geq 1$, under assumptions (A4) and (A5).

The next remark in particular says that the main result of Liu, Aue and Paul (2015) follows from Theorem 3.1 and Remark 3.1.

REMARK 3.2. (a) Under (A1), (A3), (A4) and (A5), Theorem 3.1 along with Remark 3.1 provides moments of the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$. These moments can be used to get the Stieltjes transform $m(z)$ of this LSD as

$$m(z) = y^{-1}(1+y)\varphi\big(\big(B(\lambda, z) - z\big)^{-1}\big) \quad \text{where} \tag{3.12}$$

$$K(z, \theta) = y^{-1}(1+y)\varphi\big(h(\lambda, \theta)\big(B(\lambda, z) - z\big)^{-1}\big), \tag{3.13}$$

$$h(\lambda, \theta) = \left(\sum_{j=0}^{\infty} e^{ij\theta} a_j\right)\left(\sum_{j=0}^{\infty} e^{-ij\theta} a_j^*\right), \qquad \lambda = \{a_j, a_j^* : j \geq 0\}, \tag{3.14}$$

$$B(\lambda, z) = E_\theta\big(\cos(u\theta) h(\lambda, \theta)\big(1 + y\cos(u\theta) K(z, \theta)\big)^{-1}\big), \tag{3.15}$$

and θ is a $U(0, 2\pi)$ random variable which is commutative with $\{a_j, a_j^*\}$. Details of the arguments is based on a recursion formula for moments and is given in Section 5.5.

(b) As discussed in Section 1, Liu, Aue and Paul (2015) proved the existence of the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ for the model (1.1). Their most crucial assumption was the following.

(B) $\{\psi_j\}$ are Hermitian and simultaneously diagonalizable, norm bounded matrices. There are continuous functions $f_j : \mathbb{R} \to \mathbb{R}$ and a unitary matrix U of order p such that $U\psi_j U^* = \operatorname{diag}(f_j(\alpha_1), f_j(\alpha_2), \ldots, f_j(\alpha_p))$. ESD of $\{\alpha_1, \alpha_2, \ldots, \alpha_p\}$ converges weakly to a compactly supported probability distribution F_a.

Note that assumption (B) implies assumption (A3). The main theorem of Liu, Aue and Paul (2015), under (B), provides the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ with its Stieltjes transform satisfying

$$m(z) = \int \left(E_{\theta'}\left(\frac{\cos(u\theta') h_1(\alpha, \theta')}{1 + y\cos(u\theta') K(z, \theta')}\right) - z\right)^{-1} dF_a(\alpha) \qquad \text{where} \tag{3.16}$$

$$K(z, \theta) = \int \left(E_{\theta'}\left(\frac{\cos(u\theta') h_1(\alpha, \theta')}{1 + y\cos(u\theta') K(z, \theta')}\right) - z\right)^{-1} h_1(\alpha, \theta)\, dF_a(\alpha), \tag{3.17}$$

$$h_1(\alpha, \theta) = \left|\sum_{j=0}^{q} e^{ij\theta} f_j(\alpha)\right|^2. \tag{3.18}$$

It can be shown that under assumption (B), the Stieltjes transform equations (3.12)–(3.15) reduce to equations (3.16)–(3.18). Thus, Theorem 3.1 in conjunction with Remark 3.1 implies the main theorem of Liu, Aue and Paul (2015).

So far, we have assumed $q < \infty$. With some additional assumptions, the results continue to hold for $q = \infty$. The proof of Corollary 3.1 is based on truncation arguments and is given in Section 6 of the supplementary file Bhattacharjee and Bose (2015).

COROLLARY 3.1. *Theorem* 3.1, *Remark* 3.1(a), (b) *and Corollary* 5.2 *hold for* MA(∞) *process also*, *after replacing* q *by* ∞ *provided*

(A8) $\sum_{j=0}^{\infty} \sup_p \|\psi_j\| < \infty$, *where for all* $j \geq 0$, $\|\psi_j\|$ *denotes maximum absolute eigenvalue of* ψ_j.

3.1. *Examples.*

EXAMPLE 1. Consider the MA(0) process, that is, $X_t = \varepsilon_t \; \forall t$ and suppose assumptions (A1), (A2), (A3) and $pn^{-1} \to y \in (0, \infty)$ hold. Then the following results (a)–(c) follow from Theorem 3.1 and Remark 3.2.

(a) *Marčenko–Pastur law*: The LSD of $\hat{\Gamma}_0$ is the Marčenko–Pastur law, whose moment sequence is given by [see, e.g., Marčenko and Pastur (1967) or Bai and Silverstein (2010)]

$$\beta_h = \sum_{k=1}^{h} \frac{1}{k} \binom{h-1}{k-1} \binom{h}{k} y^{k-1}, \qquad h \geq 1. \tag{3.19}$$

(b) *Free Bessel law*: The LSD of $(\frac{n}{p})^2 \hat{\Gamma}_u \hat{\Gamma}_u^*$, $u \geq 1$ is the free Bessel$(2, y^{-1})$ law, characterized by the moment sequence,

$$\beta_h = \sum_{k=1}^{h} \frac{1}{k} \binom{h-1}{k-1} \binom{2h}{k-1} y^{-k}, \qquad h \geq 1. \tag{3.20}$$

(c) The LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ are identical for all $u \geq 1$ and their common Stieltjes transformation $m(z)$ satisfies the bi-quadratic equation (with one valid solution)

$$\big(1 - y^2 m^2(z)\big)\big(yzm(z) + y - 1\big)^2 = 1. \tag{3.21}$$

This is Theorem 2.1 of Liu, Aue and Paul (2015) for the MA(0) case and Theorem 1.1 of Jin et al. (2014).

By Remark 3.1, Example 1(a) continues to hold if we assume (A7) instead of (A2). If we assume (A5) and (A6) instead of (A2), then Example 1(b) continues to hold. Moreover, Example 1(c) holds if we assume (A4) and (A5) instead of (A2). Justification for Example 1 is given in Section 5.6.

EXAMPLE 2. Let $X_t = \varepsilon_t$ where $\{\varepsilon_{t,j}\}$'s are all i.i.d. random variables with mean 0, variance 1 and $E|\varepsilon_{1.1}|^{2+\delta} < \infty$ for some $\delta > 0$. Moreover, suppose $pn^{-1} \to y \in (0, \infty)$ holds. Then, using the same idea as in the proof of Example 1(c), it can be shown that the Stieltjes transform of the LSD of $\Sigma^{1/2} \hat{\Gamma}_0 \Sigma^{1/2}$ is given by

$$m(z) = \int \frac{dF_\Sigma(t)}{z - t(1 - y - yzm(z))}, \tag{3.22}$$

where Σ is a symmetric positive definite matrix with compactly supported LSD F_Σ. This is Theorem 1.1 of Silverstein (1995). If $F_\Sigma = \delta_1$, this reduces to the Stieltjes transform of the Marčenko–Pastur law.

Apart from the expression (3.11) in terms of free variables, in general, there is no further simplified form of the LSD of $(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$. In the special case $\psi_j = \lambda_j I_p$, $\lambda_j \in \mathbb{R}$, for all $j \geq 0$, we can describe the LSD in terms of a compound free Poisson distribution. We need some preparation for this description.

DEFINITION 3.1. A probability measure μ on $\mathbb{R}$ with free cumulants

$$k_n(\mu) = \lambda m_n(\nu) \qquad \forall n \geq 1,$$

for some $\lambda > 0$ and some compactly supported probability measure ν on $\mathbb{R}$ with moments $\{m_n(\nu)\}$, is called a *compound free Poisson distribution* with rate λ and jump distribution ν.

As an example, suppose s is a semi-circular variable, defined by the moment sequence (3.6), and a is another variable free of s. Then the free cumulants of sas are given by [see Proposition 12.18 in Nica and Speicher (2006)]

$$k_n(sas, sas, \dots, sas) = \varphi(a^n) \qquad \forall n \geq 1. \tag{3.23}$$

In particular, if a is self-adjoint with distribution ν, then sas has the compound free Poisson distribution with rate $\lambda = 1$ and jump measure ν.

Let $A_{p\times p}$ be self-adjoint with compactly supported LSD a. Then it can be shown that, under (A1) and (A2), the limiting free cumulants of ZAZ^* are given by

$$\lim_n k_r(ZAZ^*, ZAZ^*, \dots, ZAZ^*) = y^{r-1}\varphi(a^r) \qquad \forall r \geq 1. \tag{3.24}$$

Therefore, asymptotically ZAZ^* is a compound free Poisson variable with rate y^{-1} and jump distribution ya.

Now we are ready to state the next example. Justification for Example 3 and Remark 3.3 are given respectively in Sections 5.7 and 5.8.

EXAMPLE 3. Let $X_t \sim \mathrm{MA}(q)$ process and suppose assumptions (A1), (A2) and $pn^{-1} \to y \in (0, \infty)$ hold. Let $\psi_j = \lambda_j I_p$, $1 \leq j \leq q$. Then the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ is a compound free Poisson whose rth order free cumulant equals

$$k_{ur} = y^{r-1} E_\theta\big(\cos(u\theta)\tilde{h}(\lambda, \theta)\big)^r \qquad \forall i \geq 0, \tag{3.25}$$

where

$$\begin{aligned} &\tilde{h}(\lambda, \theta) = \left|\sum_{j=0}^{q} e^{ij\theta}\lambda_j\right|^2, \qquad \lambda_0 = 1, \qquad \lambda = (\lambda_1, \lambda_2, \dots, \lambda_q) \quad \text{and} \\ &\theta \sim U(0, 2\pi). \end{aligned} \tag{3.26}$$

REMARK 3.3. By Remark 3.1(b), (3.25) continues to hold if we assume (A4) and (A5) instead of (A2). Example 3 together with Remark 3.1(b) justifies Theorem 2.1 in Liu, Aue and Paul (2015) when $\psi_j = \lambda_j$, Theorem 1.2 in Pfaffel and Schlemm (2011) and Theorem 1 in Yao (2012), though none of them had identified the limit as a compound free Poisson.

4. Numerical examples and applications.

4.1. *Numerical examples.* Let I_p and J_p be respectively the identity matrix of order p and the $p \times p$ matrix with all entries 1 and let $\varepsilon_t \sim \mathcal{N}_p(0, I_p), \forall t$. Let $A_p = 0.5I_p$, $B_p = 0.5(I_p + J_p)$. Let $C_p = ((c_{i,j}))$ and $D_p = ((d_{i,j}))$ be two $p \times p$ matrices with $c_{i,i} = I(1 \le i \le [p/2]) - I([p/2] < i \le p)$, $d_{i,p+1-i} = 1$ for all $i \ge 1$ and 0 otherwise. We consider the following models.

Model 1: $X_t = \varepsilon_t$.

Model 2: $X_t = \varepsilon_t + A_p\varepsilon_{t-1}$.

Model 3: $X_t = \varepsilon_t + B_p\varepsilon_{t-1}$.

Model 4: $X_t = \varepsilon_t + C_p\varepsilon_{t-1} + D_p\varepsilon_{t-2}$.

Note that in model 4, $C_pD_p \ne D_pC_p$, and hence they are not simultaneously diagonalizable and the result of Liu, Aue and Paul (2015) is not applicable. For each of these models, we draw one random sample of size n ($n = 300, 500$ and 1000). For each $1 \le u \le 4$, we plot the cumulative distribution function of ESD (ECDF) of $\hat{\Gamma}_u\hat{\Gamma}_u^*$ and $\hat{\Gamma}_u\hat{\Gamma}_u^* + \hat{\Gamma}_{u+1}\hat{\Gamma}_{u+1}^*$. The graphs for $n = 300$ are given in Figures 1 and 2. Figures 1 and 2 in the supplementary file Bhattacharjee and Bose (2015) contain graphs for $n = 500$ and 1000. These graphs support the following points:

(a) For each of the above models, the ECDF are nearly identical for $n = 300, 500$ and 1000, that is, convergence has already occurred at $n = 300$. For smaller values of n, convergence did not occur in our simulation. Some modification may improve the situation for smaller sample sizes. Here, we did not investigate any possible modifications.

(b) ECDF of $\hat{\Gamma}_u\hat{\Gamma}_u^*$ (or $\hat{\Gamma}_u\hat{\Gamma}_u^* + \hat{\Gamma}_{u+1}\hat{\Gamma}_{u+1}^*$) are almost identical—for all $u > 0$ in model 1, for all $u > 1$ in models 2 and 3 and for all $u > 2$ in model 4. Moreover, ECDFs are different—for $u = 1, 2$ in both models 2 and 3, and for $u = 1, 2, 3$ in model 4.

(c) For the MA(1) process, LSD of $\hat{\Gamma}_u\hat{\Gamma}_u^*$ (or $\hat{\Gamma}_u\hat{\Gamma}_u^* + \hat{\Gamma}_{u+1}\hat{\Gamma}_{u+1}^*$) depends on ψ_1 only through its LSD. Since LSD of A_p and B_p are identical (both have mass 1 at 0.5), the ECDF for models 2 and 3 are almost identical.

(d) As noted above, the result of Liu, Aue and Paul (2015) is not applicable for model 4. However, by Theorem 3.1, the LSD of any symmetric polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$ for model 4 exists and this is supported by row 2 right panel of Figures 1 and 2.

In Table 1, we have recorded the mean and variance of the ESD of $\hat{\Gamma}_u\hat{\Gamma}_u^*$ and $\hat{\Gamma}_u\hat{\Gamma}_u^* + \hat{\Gamma}_{u+1}\hat{\Gamma}_{u+1}^*$, $1 \le u \le 4$, for model 4 and $n = p = 300$ along with the mean

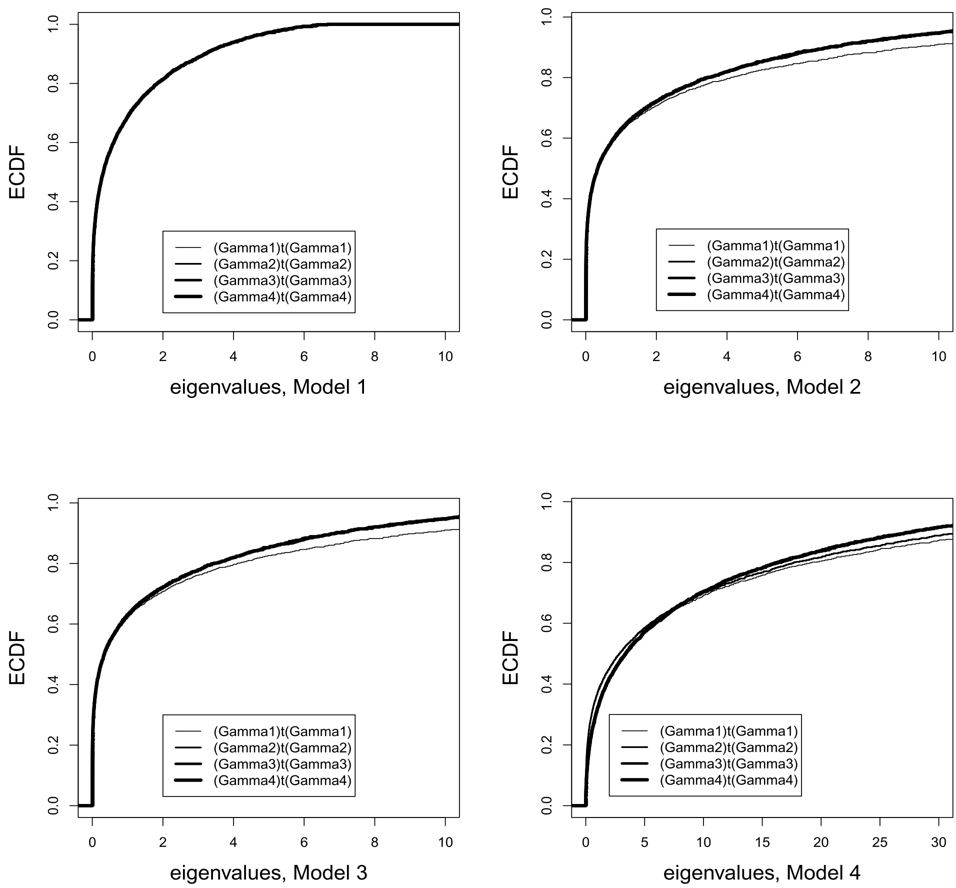

FIG. 1. *ECDF of* $\hat{\Gamma}_u \hat{\Gamma}_u^*$, $1 \le u \le 4$ *for* $n = p = 300$.

and the variance of their LSD using the description of the limits, given in Theorem 3.1, in terms of free variables and limits of coefficient matrices C_p and D_p. The empirical results agree with the theoretical results.

Incidentally, the autocovariance matrices $\{\hat{\Gamma}_u\}$ themselves are not symmetric for $u \ge 1$ and Theorem 3.1 does not apply. Nevertheless, their ESD should also converge. Figure 3 supports this for $\hat{\Gamma}_1$ of the MA(0) process. These non-symmetric matrices are under investigation.

4.2. *Applications*.

4.2.1. *Order determination of a moving average process*. A method to determine the order q of a moving average process in the univariate case is to plot the correlogram (lag vs. sample autocorrelation graph) and $\hat{q}$ is taken to be an estimate of q, if the sample autocorrelations of order greater than $\hat{q}$ are small. In the

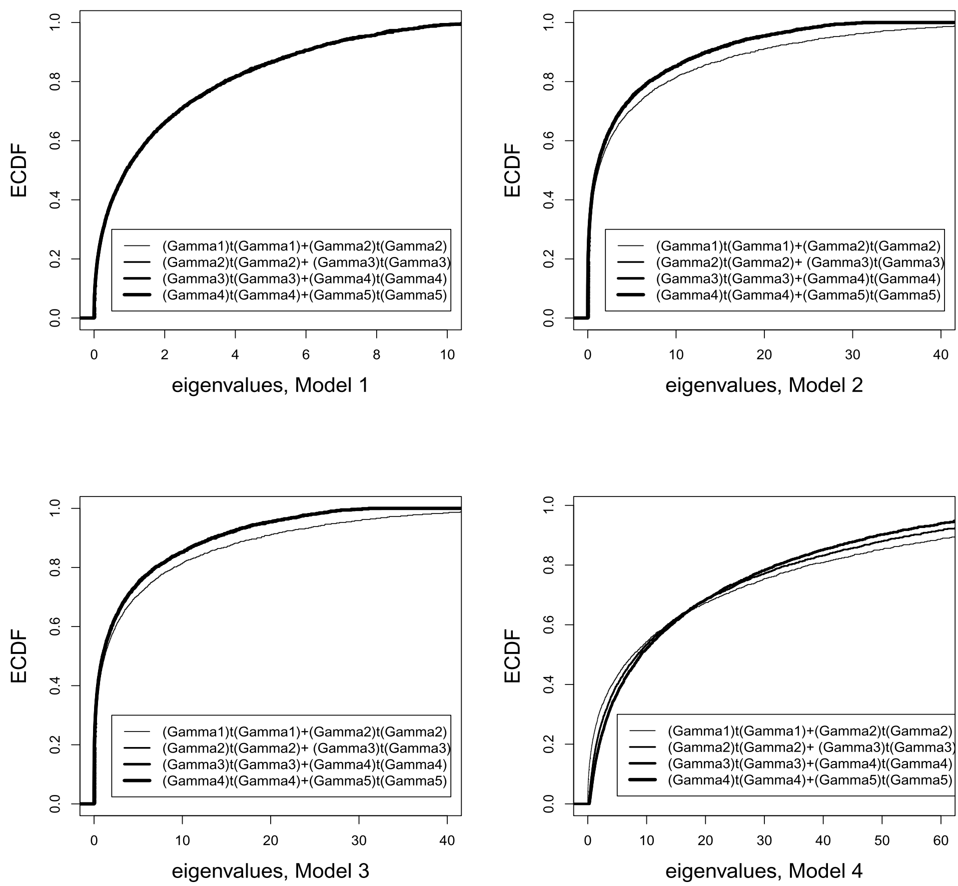

FIG. 2. *ECDF of* $\hat{\Gamma}_u\hat{\Gamma}_u^* + \hat{\Gamma}_{u+1}\hat{\Gamma}_{u+1}^*$, $1 \le u \le 4$ *for* $n = p = 300$.

TABLE 1

Means and variances for model 4, $n = p = 300$

Matrix	**Sample mean**	**Mean of LSD**	**Sample variance**	**Variance of LSD**
$\hat{\Gamma}_1\hat{\Gamma}_1^*$	10.92	11	277.8869	278
$\hat{\Gamma}_2\hat{\Gamma}_2^*$	9.93	10	214.437	215
$\hat{\Gamma}_3\hat{\Gamma}_3^*$	8.91	9	143.2908	143
$\hat{\Gamma}_4\hat{\Gamma}_4^*$	8.89	9	143.2524	143
$\hat{\Gamma}_1\hat{\Gamma}_1^* + \hat{\Gamma}_2\hat{\Gamma}_2^*$	20.85	21	802.6798	805
$\hat{\Gamma}_2\hat{\Gamma}_2^* + \hat{\Gamma}_3\hat{\Gamma}_3^*$	18.85	19	547.4531	546
$\hat{\Gamma}_3\hat{\Gamma}_3^* + \hat{\Gamma}_4\hat{\Gamma}_4^*$	17.81	18	433.1116	434
$\hat{\Gamma}_4\hat{\Gamma}_4^* + \hat{\Gamma}_5\hat{\Gamma}_5^*$	17.76	18	433.507	434

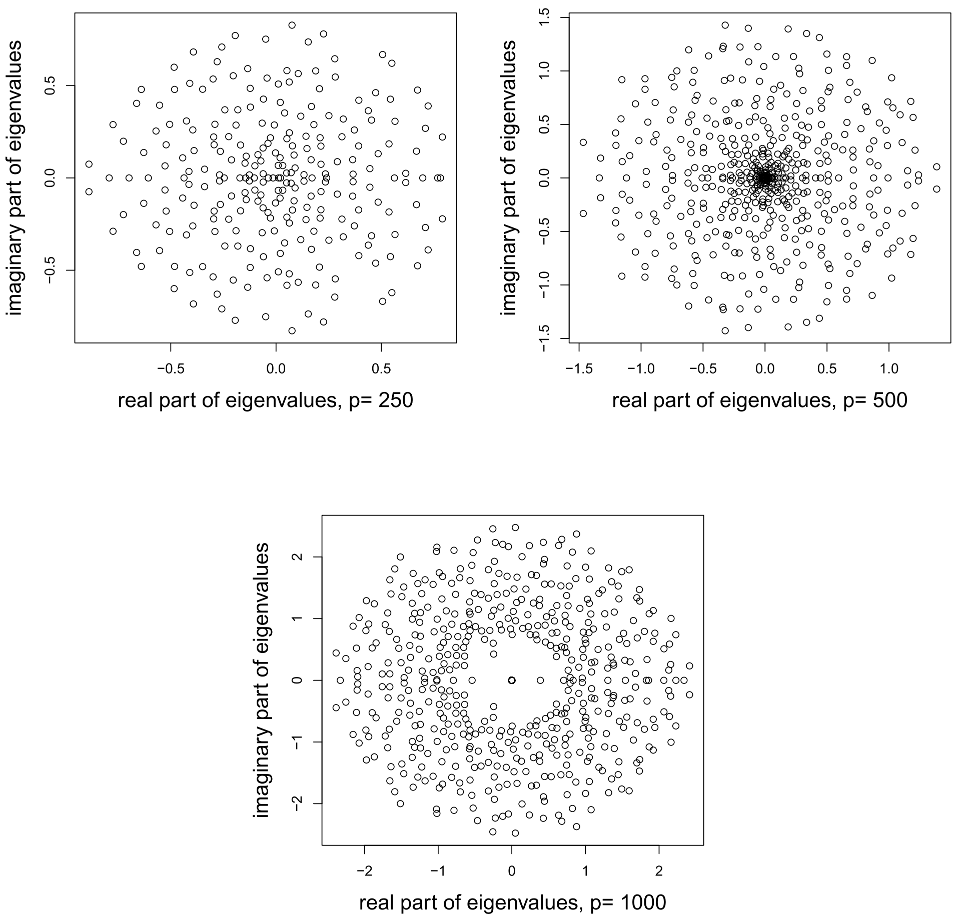

FIG. 3. *ESD of* $\hat{\Gamma}_1$ *for* MA(0) *standard Gaussian process for* $n = 500$ (*Multiple eigenvalues are plotted only once*).

high-dimensional case, as far as we know, there is no method in the literature for estimating q.

We use Theorem 3.1 to propose an analogous graphical method of determining q. First, a look at Theorem 3.1 reveals that the LSD of $\hat{\Gamma}_u \hat{\Gamma}_u^*$, for different u, can differ only due to the distribution of $\mathcal{C}_u = \{c_{j-j'+u} : 0 \le j, j' \le q\}$. However, by applying (3.8), it is not hard to see that the joint distribution of $\mathcal{C}_u$ are identical for all $u > q$ and are different for all $0 \le u \le q$.

Therefore, when X_t is a MA(q) process, the LSD of $\hat{\Gamma}_u \hat{\Gamma}_u^*$ are identical for all $u > q$ and are different for all $0 \le u \le q$. These observations also hold true for any symmetric polynomial Π_u in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$, for $u \ge 0$.

Let, for all $u \ge 0$, Π_u be a symmetric polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$. Note that the lower the order of the polynomials, the lesser would be the moment conditions required for the LSD to be valid. As an analogue of the correlogram, we propose to plot the ECDF of some chosen Π_u for first few sample autocovariance matrices in the same graph. We say that $\hat{q}$ is an estimate of q, if the ECDF of Π_u with order $u > \hat{q}$ empirically coincide with each other. For example, consider the discussions in part (b) of Section 4.1 and Figures 1 and 2. There q is determined quite accurately in the simulated data.

4.2.2. *Order determination of an autoregressive processes.* Another important problem is to determine the order of an infinite dimensional vector Autoregressive (IVAR) process

$$X_t = \varepsilon_t + A_1 X_{t-1} + A_2 X_{t-1} + \cdots + A_k X_{t-k}, \tag{4.1}$$

where k is unknown. Under suitable assumptions on the $p \times p$ parameter matrices $\{A_i\}$, one can show that X_t satisfies (1.1) [see Bhattacharjee and Bose (2014)]. Suppose $\{\varepsilon_t\}$ satisfies assumptions (A1) and (A2). Suppose the unknown parameter matrices $\{A_i\}$ are such that (4.1) is stationary, and consistent estimators $\{\hat{A}_i\}$ for $\{A_i\}$ are available. By consistency, here we mean that the limit of the spectral norm of $(\hat{A}_i - A_i)$ is zero (in probability). Such estimates are often available [see the end of Section 3 and discussions after Theorem 4.2 in Section 4 of Bhattacharjee and Bose (2014)]. Also suppose that $\{A_i\}$ are compactly supported and for any finite symmetric polynomial Π of $\{A_i\}$ $\lim p^{-1} E \operatorname{Tr}(\Pi) < \infty$ so that assumption (A3) is satisfied.

Then it is easy to see that, for each $u \geq 0$, the LSD of $\hat{\Gamma}_u \hat{\Gamma}_u^*$ for the process $\{\varepsilon_t\}$ [i.e., for the MA(0) process], coincides with the LSD (in probability) for $\{\hat{\varepsilon}_t^{(k)} = X_t - \sum_{i=1}^{k} \hat{A}_i X_{t-i}\}$. See Section 8 of the supplementary file Bhattacharjee and Bose (2015) for the proof. Instead of k, if we use any other positive integer $s < k$, then the residual process $\{\hat{\varepsilon}_t^{(s)}\}$ does not behave like the MA(0) process. As ECDF of $\hat{\Gamma}_u \hat{\Gamma}_u^*$ for $u = 1, 2$ coincide (almost surely) under MA(0) process, to determine the order of the IVAR process, it is enough to check whether the ECDF of $\hat{\Gamma}_u \hat{\Gamma}_u^*$ of $\{\hat{\varepsilon}_t^{(k)}\}$ for $u = 1, 2$ coincides or not. Therefore, if we plot the ECDF of $\hat{\Gamma}_u \hat{\Gamma}_u^*$, $u = 1, 2$ for the residual process $\{\hat{\varepsilon}_t^{(s)}\}$ in the same graph, the two distribution functions coincide only when $s = k$. Hence, we may successively fit an IVAR(s) process for $s = 0, 1, 2, \ldots$ and for each s, plot the ECDF of $\hat{\Gamma}_u \hat{\Gamma}_u^*$, $u = 1, 2$ for residuals $\{\hat{\varepsilon}_t^{(s)}\}$ in the same graph. We say that m is an estimate of order k of the IVAR process, if the ECDF of $\hat{\Gamma}_u \hat{\Gamma}_u^*$, $u = 1, 2$ do not coincide for all $s < m$ and coincide for $s = m$.

For illustration, consider the following IVAR processes. Let $\varepsilon_t \sim \mathcal{N}_p(0, I_p)$, $\forall t$.

Model 5: $X_t = \varepsilon_t + 0.5 X_{t-1}$.

Model 6: $X_t = \varepsilon_t + 0.5 X_{t-1} + 0.2 X_{t-2}$.

We let $p = n$ and draw a sample of size $n = 500$. Assuming that we do not know the parameter matrices, we use their banded estimator from Bhattacharjee and Bose (2014).

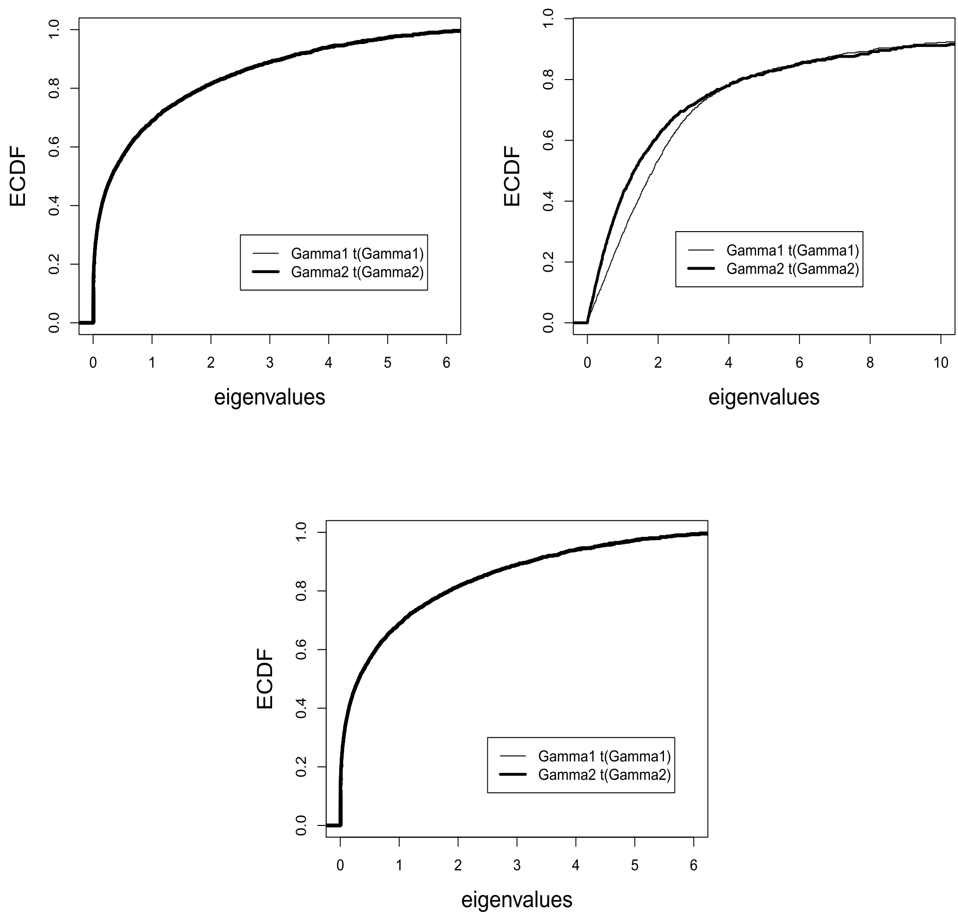

FIG. 4. $n = p = 500$. *Row* 1 *left*: *ECDF of* $\hat{\Gamma}_1\hat{\Gamma}_1^*$ *and* $\hat{\Gamma}_2\hat{\Gamma}_2^*$ *for residuals after fitting* AR(1) *in model* 5. *Row* 1 *right*: *same after fitting* AR(1) *in model* 6. *Row* 2: *same after fitting* AR(2) *in model* 6.

For model 5, we plot the two ECDFs of $\hat{\Gamma}_u\hat{\Gamma}_u^*$, $u = 1, 2$ for the residual process $\{\hat{\varepsilon}_t^{(1)}\}$ in the same graph and observe that they coincide. See row 1, left panel in Figure 4. Therefore, 1 is an estimate of the order of model 5. For model 6, we do the same but the two ECDFs do not coincide (see row 1, right panel in Figure 4). In row 2 of Figure 4, the same two ECDFs are plotted for $\{\hat{\varepsilon}_t^{(2)}\}$ and they coincide and hence 2 is an estimate of the order for model 6.

4.2.3. *Asymptotic distribution of traces and an application in testing*. One of the referees raised the issue of convergence in distribution of the trace of any autocovariance matrix and if such a result could be possibly used for testing problems. Let $\Pi := \Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)$ be a symmetric polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0\}$ and

$\sigma_\Pi^2 = \lim E(\mathrm{Tr}(\Pi) - E\,\mathrm{Tr}(\Pi))^2$. Then, for $d \geq 1$, using some combinatorial calculations [see Lemma 2.1 in the supplementary file Bhattacharjee and Bose (2015)], we have

$$\lim E\big(\mathrm{Tr}(\Pi) - E\,\mathrm{Tr}(\Pi)\big)^T = \begin{cases} 0, & \text{if } T = 2d-1, \\ \left(\prod_{k=1}^{d}(2d-2k+1)\right)\sigma_\Pi^{2d}, & \text{if } T = 2d. \end{cases}$$

Therefore,

$$(\mathrm{Tr}(\Pi) - E\,\mathrm{Tr}(\Pi)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_\Pi^2). \tag{4.2}$$

The following are some examples and simulations to support (4.2). We consider $n = p$ and $\varepsilon_t \sim \mathcal{N}_p(0, I_p)$, where ε_t's are independent.

EXAMPLE 4. Let $X_t = \varepsilon_t, \forall t$. Then $E(\mathrm{Tr}\,\hat{\Gamma}_0) = n$, $E(\mathrm{Tr}(\hat{\Gamma}_1\hat{\Gamma}_1^*)) = n-1$, $E(\mathrm{Tr}(\hat{\Gamma}_1 + \hat{\Gamma}_1^*)) = 0$ and $\lim E(\mathrm{Tr}(\hat{\Gamma}_0) - E\,\mathrm{Tr}(\hat{\Gamma}_0))^2 = 2$, $\lim E(\mathrm{Tr}(\hat{\Gamma}_1\hat{\Gamma}_1^*) - E\,\mathrm{Tr}(\hat{\Gamma}_1\hat{\Gamma}_1^*))^2 = 10$, $\lim E(\mathrm{Tr}(\hat{\Gamma}_1 + \hat{\Gamma}_1^*) - E\,\mathrm{Tr}(\hat{\Gamma}_1 + \hat{\Gamma}_1^*))^2 = 4$. We omit the detailed calculations which are simpler than the calculations in the next example. Hence, $(\mathrm{Tr}(\hat{\Gamma}_0) - n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2)$, $(\mathrm{Tr}(\hat{\Gamma}_1\hat{\Gamma}_1^*) - n + 1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 10)$, and $\mathrm{Tr}(\hat{\Gamma}_1 + \hat{\Gamma}_1^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 4)$. Simulation results given in rows 1 and 2, left panel, Figure 5 support the above convergences.

EXAMPLE 5. Let $X_t = \varepsilon_t + \varepsilon_{t-1}$. Then $E(\mathrm{Tr}(\hat{\Gamma}_0)) = 2(n-1)$, $\lim E(\mathrm{Tr}(\hat{\Gamma}_0) - E\,\mathrm{Tr}(\hat{\Gamma}_0))^2 = 8$. In Section 5.9, we show details of this calculation. Hence, $(\mathrm{Tr}(\hat{\Gamma}_0) - 2(n-1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 8)$. The simulation result given in row 2, right panel, Figure 5 supports this convergence.

These results can be used for testing. For example, suppose we wish to test

$$H_0 : X_t = \varepsilon_t \qquad \forall t \quad \text{against} \quad H_1 : X_t = \varepsilon_t + \varepsilon_{t-1} \qquad \forall t.$$

Then $(\mathrm{Tr}(\hat{\Gamma}_0) - n)$ can be used as a test statistic and large value of the test statistic will imply rejection of H_0. Clearly, this idea can be extended to test other pairs of simple null and alternative hypotheses for model (1.1).

5. Proofs. We first prove Theorem 3.1. For this purpose, we need the following notions and results.

5.1. *Convergence of NCPs and assymptotic freeness.* A sequence of NCPs $(\mathrm{Span}\{b_i^{(n)}\}, \varphi_n)$ is said to converge to an NCP $(\mathrm{Span}\{b_i\}, \varphi)$, if for any polyno-

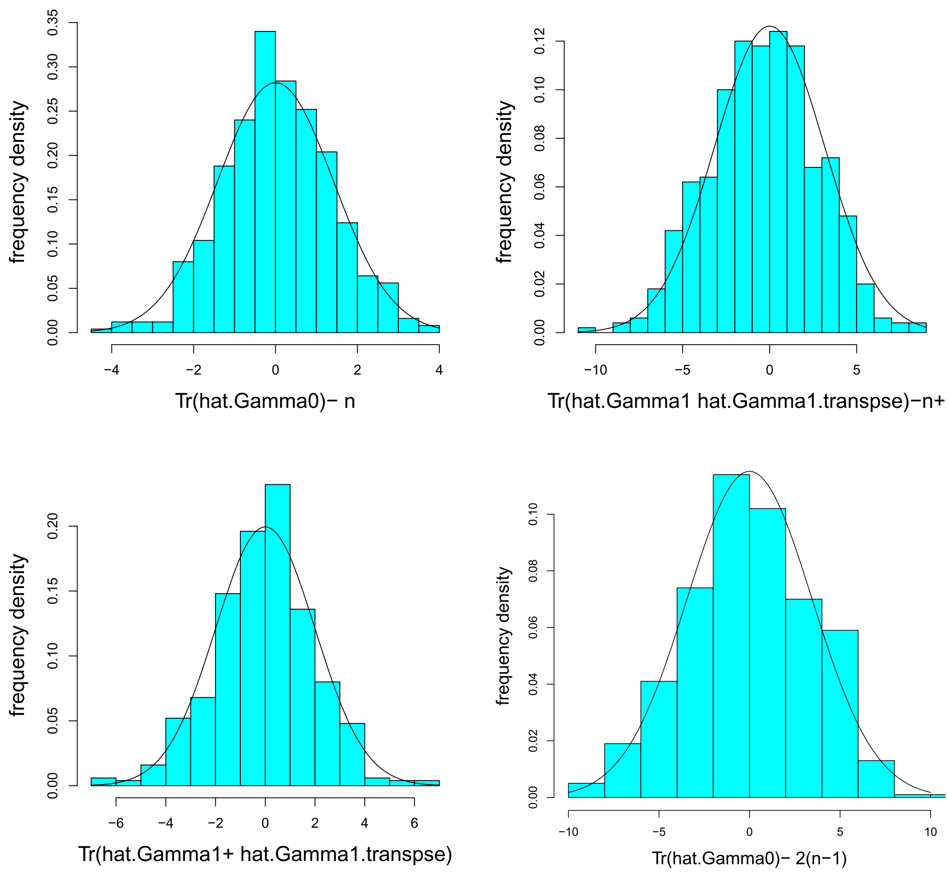

FIG. 5. $n = p = 500$ *and* 500 *replications. Row* (1) *left, rows* (1) *right and* (2) *left represent respectively the histogram of* $(\mathrm{Tr}(\hat{\Gamma}_0) - n)$, $(\mathrm{Tr}(\hat{\Gamma}_1\hat{\Gamma}_1^*) - n + 1)$ *and* $\mathrm{Tr}(\hat{\Gamma}_1 + \hat{\Gamma}_1^*)$, *when* $X_t = \varepsilon_t$. *Row* (2) *right represents the histogram of* $(\mathrm{Tr}(\hat{\Gamma}_0) - 2n)$, *when* $X_t = \varepsilon_t + \varepsilon_{t-1}$.

mial π,

$$\lim_n \varphi_n\bigl(\pi\bigl(b_i^{(n)} : i \geq 0\bigr)\bigr) = \varphi\bigl(\pi(b_i : i \geq 0)\bigr). \tag{5.1}$$

Suppose $(\mathrm{Span}\{b_{ij}^{(n)} : i \geq 0, 1 \leq j \leq k\}, \varphi_n)$ converges to $(\mathrm{Span}\{b_{ij} : i \geq 0, 1 \leq j \leq k\}, \varphi)$. Then $\mathrm{Span}\{b_{ij}^{(n)} : i \geq 0\}$, $1 \leq j \leq k$, are said to be *asymptotically free* if $\mathrm{Span}\{b_{ij} : i \geq 0\}$ are free across $1 \leq j \leq k$.

Let $W_{n\times n}$ be a *Wigner* matrix. Let $\varphi_n = n^{-1} E\,\mathrm{Tr}$. Let $\{B_{i,n}\}$ and $\{D_{i,n}\}$, $1 \leq i \leq J$ be sequences of non-random, compactly supported, square matrices of order n each of which converges in the above sense. Then, under assumption (A2), the following facts are true. For (a) and (b), see Zeitouni, Anderson and Guionnet (2010). (c) follows from (a), (b) and Theorem 11.12, page 180 of Nica and

Speicher (2006). (d) is immediate from (a), (b) and (c). We drop the suffix n for clarity.

(a) $W/\sqrt{n}$ converges to the semi-circle law with moment sequence (3.6).
(b) $W/\sqrt{n}$ and $\{B_i, D_i, 1 \le i \le J\}$ are asymptotically free.
(c) $\{n^{-1}WB_iW, 1 \le i \le J\}$ and $\{D_i, 1 \le i \le J\}$ are asymptotically free.
(d) To compute $\lim n^{-1} E\,\mathrm{Tr}(n^{-k} D_1 W B_1 W D_2 W B_2 W D_3 \cdots W B_k W D_{k+1})$, one can assume that $W/\sqrt{n}$, $\{B_i\}$ and $\{D_i\}$ are asymptotically free.

5.2. *Proof of Theorem* 3.1. To prove Theorem 3.1, as discussed in Section 1, we have to show (M1), (M4) and (C) are satisfied. Here, we shall only establish (M1). Proof of (M4) and (C) are given respectively in Sections 3 and 4 of the supplementary file Bhattacharjee and Bose (2015).

To establish (M1), we have to essentially show (3.11). Let $\hat{\Gamma}_u(\varepsilon)$ be the uth order sample autocovariance matrix of the process $\{\varepsilon_t\}$. Let

$$\Delta_u = \sum_{j=0}^{q}\sum_{j'=0}^{q} \psi_j \hat{\Gamma}_{j'-j+u}(\varepsilon)\psi_{j'}^* \qquad \forall u \ge 0. \tag{5.2}$$

Then by Lemma 1.1 of the supplementary file Bhattacharjee and Bose (2015), it is enough to show (3.11) after we replace $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$ by $\{\Delta_u, \Delta_u^*\}$. Now

$$n(n+p)^{-1}\bar{\Delta}_i = (n+p)^{-1}\sum_{j=0}^{q}\sum_{j'=0}^{q} \bar{\psi}_j W \underline{\mathrm{P}}_{j'-j+i} W \bar{\psi}_{j'}^* \qquad \forall i \ge 0. \tag{5.3}$$

Note that by (a) of Section 5.1, $W/\sqrt{n+p}$ converges to the semi-circle law with moment sequence (3.6). Moreover, by (3.8) and (3.9), $\{\bar{\psi}_j\}$ and $\{\underline{\mathrm{P}}_j\}$ converge respectively to $\{a_j\}$ and $\{c_j\}$. Also, by (b), (c) and (d) of Section 5.1, s, $\{a_j\}$ and $\{c_j\}$ are freely independent. Therefore, by (5.1), (3.11) holds and (M1) is verified. Hence, proof of Theorem 3.1 is complete.

Next, we need an algorithm for computing moments of a particular type of polynomials of free variables.

5.3. *Algorithm to compute moments of free variables.* As we have discussed in Section 2, all joint moments of free variables are computable in terms of the moments of the individual variables. The algorithm for computing moments under freeness is different from the product rule under usual independence. Note that, for our purpose, a typical term in the moment calculations [see, e.g., (3.7)] is

$$\varphi(d_0 s b_1 s d_1 s b_2 s d_2 \cdots s b_n s d_n) \qquad \text{where } \{b_i\}, \{d_i\} \text{ and } s \text{ are free.} \tag{5.4}$$

Note that in our case, since $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$, our φ satisfies $\varphi(ab) = \varphi(ba)$, $\forall a, b$. In this section, we shall discuss the algorithm for computing (5.4) in terms of the moments of $\{b_i\}$, $\{d_i\}$ and s.

Let $NC(n)$ be the set of all non-crossing partitions of $\{1, 2, \ldots, n\}$. Define recursively a family of multiplicative, multi-linear functionals φ_π $(n \geq 1, \pi \in NC(n))$ by the following formula. If $\pi = \{V_1, V_2, \ldots, V_r\} \in NC(n)$, then

$$\varphi_\pi[a_1, a_2, \ldots, a_n] := \varphi(V_1)[a_1, a_2, \ldots, a_n] \cdots \varphi(V_r)[a_1, a_2, \ldots, a_n], \tag{5.5}$$

where

$$\varphi(V)[a_1, a_2, \ldots, a_n] := \varphi_s(a_{i_1} a_{i_2} \cdots a_{i_s}) \qquad \text{for } V = (i_1 < i_2 < \cdots < i_s). \tag{5.6}$$

Let $NC_2(2n)$ be the set of all non-crossing pair partitions of $\{1, 2, \ldots, 2n\}$ and $K(\pi) \in NC(n)$ be the Kreweras complement of the partition π [see Definition 9.21 in Nica and Speicher (2006)]. Then we have the following lemma. Relation (5.9) will be useful to justify Example 1(a) and (b). Relations (5.7) and (5.10) will be useful to prove Lemma 5.2.

LEMMA 5.1. (a)

$$\begin{aligned} &\varphi(d_0 s b_1 s d_1 s b_2 \cdots s d_n) \\ &\quad = \sum_{\pi \in NC_2(2n)} \varphi_{K(\pi)}[b_1, d_1, b_2, d_2, \ldots, b_n, d_n d_0] &(5.7)\\ &\quad = \sum_{\pi \in NC(n)} \varphi_\pi[b_1, b_2, \ldots, b_n] \varphi_{K(\pi)}[d_1, d_2, \ldots, d_n d_0] &(5.8)\\ &\quad = \sum_{\pi \in NC(n)} \varphi_\pi[d_1, d_2, \ldots, d_n d_0] \varphi_{K(\pi)}[b_1, b_2, \ldots, b_n]. &(5.9)\end{aligned}$$

(b) *Fix* $1 = k_0 < k_1 < \cdots < k_t \leq n$ *and the following subset of* $NC_2(2n)$ *as*

$$\mathcal{S} = \{\pi \in NC_2(2n) : \{2k_i, 2k_{i+1} - 1\} \in \pi, 0 \leq i \leq t, k_{t+1} = k_0\}.$$

Then

$$\begin{aligned} &\sum_{\pi \in \mathcal{S}} \varphi_{K(\pi)}[b_1, d_1, b_2, d_2, \ldots, b_n, d_n d_0] \\ &\quad = \varphi\left(\prod_{s=0}^{t} b_{k_s}\right) \prod_{s=1}^{t+1} \varphi(d_{k_{s-1}} s b_{k_{s-1}+1} s d_{k_{s-1}+1} \cdots s d_{k_s - 1}), \end{aligned} \tag{5.10}$$

where $k_0 = 1$, $d_{k_{t+1}-1} = d_n d_0$.

Relation (5.7) follows by (22.10) of Nica and Speicher (2006). By freeness of $\{b_i\}$ and $\{d_i\}$, and by properties of the Kreweras complement [Exercises 9.41(1), 9.42(1) and (2) in Nica and Speicher (2006)], (5.8) and (5.9) follow from (5.7). Relation (5.10) follows from the multiplicative property (5.5) and (5.6) of partitions and from certain properties of Kreweras complement. A detailed proof of (5.10) is given in Section 9 of the supplementary file Bhattacharjee and Bose (2015).

5.4. *A recursion formula for moments and its proof.* In this section, we shall prove a lemma that provides a recursion formula for the moments of the LSD of $2^{-1}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$, which will be used in the proof of Remark 3.2 in the next section.

Let

$$D = 2^{-1}(\hat{\Gamma}_u + \hat{\Gamma}_u^*) \quad \text{and} \quad d_{uq} = 2^{-1}(\gamma_{uq} + \gamma_{uq}^*),$$

where γ_{uq} is as in (3.10). Suppose θ is a $U(0, 2\pi)$ variable, which is (classical) independent and commutative with $\{a_j\}$ and d_{uq}. Recall $h(\lambda, \theta)$ in (3.14). For any polynomial $\Pi = \Pi(\psi_j, \psi_j^* : j \geq 0)$, let $\bar{\Pi} = \Pi(a_j, a_j^* : j \geq 0)$. For all $j \geq 0$, let

$$R_{uj}(\theta) = y^{-1}(1+y)\varphi(h(\lambda, \theta) d_{uq}^{j-1}), \tag{5.11}$$

$$S_{uj}(\theta, \Pi) = y^{-1}(1+y)\varphi(\bar{\Pi} h(\lambda, \theta) d_{uq}^{j-1}). \tag{5.12}$$

LEMMA 5.2. *Let $X_t \sim \mathrm{MA}(q)$ process and suppose assumptions* (A1), (A2), (A3) *and $pn^{-1} \to y \in (0, \infty)$ hold. Then for any polynomial $\Pi = \Pi(\psi_j, \psi_j^* : j \geq 0)$, we have*

$$\begin{aligned} &\lim p^{-1} E \operatorname{Tr}(\Pi D^r) \\ &\quad = \frac{1}{y} \sum_{t=1}^{r} E_\theta \Bigg[\big(y \cos(u\theta)\big)^t \sum_{\substack{1 \leq i_1, i_2, \ldots, i_t \leq r \\ \sum_{j=1}^t i_j = r}} S_{u i_1}(\theta, \Pi) \Bigg(\prod_{k=2}^{t} R_{u i_k}(\theta) \Bigg) \Bigg]. \end{aligned} \tag{5.13}$$

PROOF. From the proof of (3.11), it is immediate that

$$\begin{aligned} &\frac{y}{1+y} \lim p^{-1} E \operatorname{Tr}(\Pi D^r) \\ &\quad = \varphi(\bar{\Pi} d_{uq}^r) \\ &\quad = \sum_{\substack{j_k, j_k' = 1 \\ 1 \leq k \leq r}}^{q} \sum_{\substack{v_k = u, -u \\ 1 \leq k \leq r}} \varphi\Bigg(\bar{\Pi} \prod_{k=1}^{r} a_{j_k} s c_{j_k - j_k' + v_k} s a_{j_k'}^* \Bigg) \qquad \text{by (3.10)} \\ &\quad = \sum_{\sigma \in NC_2(2r)} \tau_\sigma \qquad \text{by (5.7)}, \end{aligned} \tag{5.14}$$

where

$$\tau_\sigma = \sum_{\substack{j_k, j_k' = 1 \\ 1 \leq k \leq r}}^{q} \sum_{\substack{v_k = u, -u \\ 1 \leq k \leq r}} \varphi_{K(\sigma)}[c_{j_1 - j_1' + v_1}, a_{j_1'}^* a_{j_2}, c_{j_2 - j_2' + v_2}, a_{j_2'}^* a_{j_3}, \ldots, a_{j_r'}^* a_{j_1} \bar{\Pi}],$$

and $K(\sigma)$ is the Kreweras complement of σ. Now to compute (5.14), we consider the decomposition of $NC_2(2r) = \bigcup_{t=1}^{r} \mathcal{P}_t^{2r}$, where $\mathcal{P}_1^{2r} = \{\sigma \in NC_2(2r) : \{1, 2\} \in$

$\sigma\}$ and for all $2 \le t \le r$,

$$\mathcal{P}_t^{2r} = \big\{\sigma \in NC_2(2r) : \{2k_0 - 1, 2k_t\}, \{2k_0, 2k_1 - 1\}, \{2k_1, 2k_2 - 1\}, \dots, \{2k_{t-2}, 2k_{t-1} - 1\} \in \sigma, 1 = k_0 < k_1 < k_2 < \dots < k_{t-1} \le r\big\}.$$

Hence, (5.14) is equivalent to

$$y(1+y)^{-1} \lim p^{-1} E\, \mathrm{Tr}\big(\Pi D^r\big) = \sum_{t=1}^{r} \mathcal{T}_t, \tag{5.15}$$

where for all $1 \le t \le r$,

$$\mathcal{T}_t = \sum_{\sigma \in \mathcal{P}_t^{2r}} \tau_\sigma = \sum_{1 = k_0 < k_1 < k_2 < \dots < k_{t-1} \le r} g(t, k_1, k_2, \dots, k_{t-1}), \tag{5.16}$$

and

$$\begin{aligned} &2^{t+1}(1+y)^{-t-1} g(t+1, k_1, k_2, \dots, k_t) \\ &\quad = \sum_{\substack{1 \le j_{k_s}, j'_{k_s} \le q \\ v_{k_s} = u, -u}} \varphi\Bigg(\prod_{s=0}^{t} c_{j_{k_s} - j'_{k_s} + v_{k_s}}\Bigg) \prod_{s=0}^{t} \varphi\big(a^*_{j'_{k_s}} d_{uq}^{k_{s+1} - k_s - 1} a_{j_{k_{(s+1)}}}\big) \end{aligned} \tag{5.17}$$

$$\big[\text{by (5.10) and where } k_{t+1} = r + 1 \text{ and } a_{j_{k_{t+1}}} = a_{j_{k_0}} \bar{\Pi}\big].$$

Now, by (3.8),

$$\begin{aligned} \varphi\Bigg(\prod_{s=0}^{t} c_{j_{k_s} - j'_{k_s} + v_{k_s}}\Bigg) &= \begin{cases} 1, & \text{if } \sum_{s=0}^{t} (j_{k_s} - j'_{k_s} + v_{k_s}) = 0, \\ 0, & \text{otherwise} \end{cases} \\ &= E_\theta\big(e^{i\theta \sum_{s=0}^{t} (j_{k_s} - j'_{k_s} + v_{k_s})}\big) \\ &\qquad \text{where } \theta \sim U(0, 2\pi). \end{aligned}$$

Therefore, (5.17) is equivalent to

$$\begin{aligned} &= \frac{1}{1+y} \sum_{\substack{1 \le j_{k_s}, j'_{k_s} \le q \\ v_{k_s} = u, -u}} E_\theta\big(e^{\sum_{s=0}^{t} i\theta (j_{k_s} - j'_{k_s} + v_{k_s})}\big) \prod_{s=0}^{t} \varphi\big(a^*_{j'_{k_s}} d_{uq}^{k_{s+1} - k_s - 1} a_{j_{k_{(s+1)}}}\big) \\ &= \frac{1}{1+y} E_\theta\Bigg(\prod_{s=0}^{t} \sum_{j'_{k_s}, j_{k_{s+1}}, v_{k_s}} e^{i\theta(-j'_{k_s} + j_{k_{(s+1)}} + v_{k_s})} \varphi\big(a^*_{j'_{k_s}} d_{uq}^{k_{s+1} - k_s - 1} a_{j_{k_{(s+1)}}}\big)\Bigg) \\ &\qquad \text{where } j_{k_{t+1}} = j_{k_0}. \end{aligned}$$

Note that for a fix $0 \le s \le t$, since $v_{k_s} = u, -u$,

$$\sum_{j'_{k_s}, j_{k_{s+1}}, v_{k_s}} e^{i\theta(-j'_{k_s} + j_{k(s+1)} + v_{k_s})} \varphi(a^*_{j'_{k_s}} d_{uq}^{k_{s+1}-k_s-1} a_{j_{k(s+1)}})$$

$$= 2\cos(u\theta)\varphi\biggl(\Bigl(\sum e^{i\theta j_{k(s+1)}} a_{j_{k(s+1)}}\Bigr)\Bigl(\sum_{j_{k_s}} a^*_{j'_{k_s}} e^{-i\theta j'_{k_s}}\Bigr) d_{uq}^{k_{s+1}-k_s-1}\biggr)$$

$$= \begin{cases} 2\cos(u\theta)\dfrac{y}{1+y} R_{u(k_{s+1}-k_s)}(\theta), & 1 \le s \le t-1, \\ 2\cos(u\theta)\dfrac{y}{1+y} S_{u(r+1-k_t)}(\theta, \Pi), & s = t. \end{cases}$$

Hence, for all $1 = k_0 < k_1 < k_2 < \cdots < k_t \le r$, we have

$$g(t+1, k_1, k_2, \dots, k_t)$$
$$= y^{t+1}(1+y)^{-1} E_\theta\biggl(\cos(u\theta) S_{u(r+1-k_t)}(\theta, \Pi) \prod_{s=0}^{t-1} \cos(u\theta) R_{u(k_{s+1}-k_s)}(\theta)\biggr).$$

Therefore, by (5.16), for all $1 \le t \le r$,

$$\mathcal{T}_t = \frac{y^t}{(1+y)} \sum_{\substack{1=k_0<k_1<\cdots<k_t\le r \\ k_{t+1}=r+1}} E_\theta\biggl[\bigl(\cos(u\theta)\bigr)^t S_{u(r+1-k_t)}(\theta, \Pi) \prod_{s=0}^{t-1} R_{u(k_{s+1}-k_s)}(\theta)\biggr]$$

$$= y^t(1+y)^{-1} \sum_{\substack{1\le i_1, i_2, \dots, i_t \le r \\ i_1+i_2+\cdots+i_t = r}} E_\theta\biggl(\bigl(\cos(u\theta)\bigr)^t S_{ui_1}(\theta, \Pi) \prod_{s=2}^{t} R_{ui_s}(\theta)\biggr).$$

Hence, by using (5.15) and (5.16), relation (5.13) follows. □

5.5. *Proof of Remark* 3.2. (a) We prove (3.12)–(3.15) stated in Remark 3.2 under assumptions (A1)–(A3). By Remark 3.1, assumption (A2) can be replaced by assumptions (A4) and (A5).

Let us define

$$K(z, \theta) = \sum_{i=1}^{\infty} z^{-i} R_{ui}(\theta), \qquad D = \sum_{i=1}^{\infty} z^{-i} d_{uq}^i,$$

$$B(\lambda, z) = E_\theta\bigl(\cos(u\theta) h(\lambda, \theta)\bigl(1 + y\cos(u\theta) K(z, \theta)\bigr)^{-1}\bigr),$$

where $\{R_{ui}\}$ is defined in (5.11). Now note that

$$(5.18)\qquad \begin{aligned} y(1+y)^{-1} K(z, \theta) &= \sum_{i=1}^{\infty} z^{-i} \varphi\bigl(h(\lambda, \theta) d_{uq}^{i-1}\bigr) \qquad \text{by (5.11)} \\ &= -z^{-1}\varphi\bigl(h(\lambda, \theta)\bigr) - z^{-1}\varphi\bigl(h(\lambda, \theta) D\bigr), \end{aligned}$$

where

$$
\begin{aligned}
&\varphi\big(h(\lambda,\theta)D\big)\\
&\quad=\sum_{r=1}^{\infty} z^{-r} y^{-1}\sum_{t=1}^{r} E_{\theta'}\Bigg[\big(y\cos(u\theta')\big)^t\\
&\qquad\times\Bigg(\sum_{\substack{1\le i_1,i_2,\dots,i_t\le r\\ i_1+i_2+\cdots+i_t=r}} S_{ui_1}\big(\theta',h(\lambda,\theta)\big)\prod_{s=2}^{t} R_{ui_s}(\theta')\Bigg)\Bigg] \qquad \text{(by Lemma 5.2)}\\
&\quad=y^{-1}\sum_{t=1}^{\infty} E_{\theta'}\Bigg[\big(y\cos(u\theta')\big)^t\\
&\qquad\times\sum_{r=t}^{\infty} z^{-r}\Bigg(\sum_{\substack{1\le i_1,i_2,\dots,i_t\le r\\ i_1+i_2+\cdots+i_t=r}} S_{ui_1}\big(\theta',h(\lambda,\theta)\big)\prod_{s=2}^{t} R_{ui_s}(\theta')\Bigg)\Bigg]\\
&\quad=y^{-1}\sum_{t=1}^{\infty} E_{\theta'}\Bigg[\big(y\cos(u\theta')\big)^t\big(K(z,\theta')\big)^{t-1}\Bigg(\sum_{r=1}^{\infty} z^{-r} S_{ur}\big(\theta',h(\lambda,\theta)\big)\Bigg)\Bigg]\\
&\quad=E_{\theta'}\Bigg[\cos(u\theta')\big(1+y\cos(u\theta')K(z,\theta')\big)^{-1}\Bigg(\sum_{r=1}^{\infty} z^{-r} S_{ur}\big(\theta',h(\lambda,\theta)\big)\Bigg)\Bigg]\\
&\quad=\varphi\Bigg(\sum_{r=1}^{\infty} z^{-r} d_{uq}^{r-1} h(\lambda,\theta) E_{\theta'}\big[h(\lambda,\theta')\cos(u\theta')\big(1+y\cos(u\theta')K(z,\theta')\big)^{-1}\big]\Bigg)\\
&\quad=z^{-1}\varphi\big(h(\lambda,\theta)B(\lambda,z)\big)+z^{-1}\varphi\big(Dh(\lambda,\theta)B(\lambda,z)\big).
\end{aligned}
$$

In a similar fashion, using $h(\lambda,\theta)B(\lambda,z)$ instead of $h(\lambda,\theta)$ in the above steps,

$$\varphi\big(Dh(\lambda,\theta)B(\lambda,z)\big)=z^{-1}\varphi\big(h(\lambda,\theta)B^2(\lambda,z)\big)+z^{-1}\varphi\big(Dh(\lambda,\theta)B^2(\lambda,z)\big).$$

Finally iterating we have

$$\varphi\big(h(\lambda,\theta)D\big)=\sum_{r=1}^{\infty} z^{-r}\varphi\big(h(\lambda,\theta)B^r(\lambda,z)\big). \tag{5.19}$$

We now need to show only (3.12) and (3.13). Using (5.19) and (5.18),

$$
\begin{aligned}
&y(1+y)^{-1}K(z,\theta)\\
&\quad=-z^{-1}\varphi\Bigg(\sum_{r=0}^{\infty} h(\lambda,\theta) z^{-r} B^r(\lambda,z)\Bigg)=\varphi\big(h(\lambda,\theta)\big(B(\lambda,z)-z\big)^{-1}\big),
\end{aligned}
\tag{5.20}
$$

which is (3.13) in Remark 3.2.

Note that the above steps from (5.18) leading to (5.20) remain valid if we replace $h(\lambda,\theta)$ by 1 in (5.18). This yields [instead of (5.20)],

$$y(1+y)^{-1}m(z)=\varphi\big(\big(B(\lambda,z)-z\big)^{-1}\big), \tag{5.21}$$

which is (3.12) in Remark 3.2. Hence, the proof of Remark 3.2(a) is complete.

(b) We now prove that under assumption (B), the Stieltjes transform equations (3.12)–(3.15) reduce to equations (3.16)–(3.18).

Note that under assumption (B),

$$\begin{aligned}
h(\lambda,\theta)&=\Bigg(\sum_{j=0}^{\infty}e^{ij\theta}a_j\Bigg)\Bigg(\sum_{j=0}^{\infty}e^{-ij\theta}a_j^*\Bigg),\qquad \lambda=\{a_j,a_j^*:j\ge 0\}\\
&=y(1+y)^{-1}\Bigg(\sum_{j=0}^{\infty}e^{ij\theta}f_j(\alpha)\Bigg)\Bigg(\sum_{j=0}^{\infty}e^{-ij\theta}f_j^*(\alpha)\Bigg)\\
&\quad+(1+y)^{-1}\delta_0,\qquad \alpha\sim F_a \qquad\qquad (5.22)\\
&=y(1+y)^{-1}\Bigg|\sum_{j=0}^{\infty}e^{ij\theta}f_j(\alpha)\Bigg|^2+(1+y)^{-1}\delta_0\\
&=y(1+y)^{-1}h_1(\alpha,\theta)+(1+y)^{-1}\delta_0,\qquad \alpha\sim F_a,
\end{aligned}$$

where δ_0 is the degenerate random variable at 0. Therefore, after substituting (3.15) in (3.13), we have

$$\begin{aligned}
K(z,\theta)&=y^{-1}(1+y)\varphi\big(h(\lambda,\theta)\big(B(\lambda,z)-z\big)^{-1}\big)\\
&=y^{-1}(1+y)\varphi\Bigg(\Bigg[E_{\theta'}\Bigg(\frac{\cos(u\theta')h(\lambda,\theta')}{(1+y\cos(u\theta')K(z,\theta'))}\Bigg)-z\Bigg]^{-1}h(\lambda,\theta)\Bigg)\\
&=\int\Bigg[E_{\theta'}\Bigg(\frac{\cos(u\theta')h_1(\alpha,\theta')}{(1+y\cos(u\theta')K(z,\theta'))}\Bigg)-z\Bigg]^{-1}h_1(\alpha,\theta)\,dF_a\qquad \text{by (5.22)}.
\end{aligned}$$

Hence, (3.13) reduces to (3.17). Similarly, one can show (3.12) reduces to (3.16). Therefore, proof of Remark 3.2(b) is complete.

5.6. *Justification for Example* 1. (a) Observe that, by (3.11),

$$\lim p^{-1}E\,\mathrm{Tr}(\hat{\Gamma}_0)^h=y^{-1}(1+y)\varphi(\gamma_{00}^h)\qquad \forall h\ge 1,$$

where by (3.10), $\gamma_{00}=\gamma_{00}^*=(1+y)a_0sc_0sa_0$. By (3.8) and (3.9), a_0 and c_0 are both Bernoulli random variables with success probabilities $y(1+y)^{-1}$ and $(1+y)^{-1}$, respectively. Hence, by (5.9), the hth moment of the LSD of $\hat{\Gamma}_0$ is given by

$$\frac{(1+y)^{h+1}}{y}\sum_{\pi\in NC(h)}\varphi_\pi[a_0^2,a_0^2,\ldots,a_0^2]\varphi_{K(\pi)}[c_0,c_0,\ldots,c_0]. \tag{5.23}$$

Note that if $\pi \in NC(h)$ has k blocks, then

$$\varphi_\pi[a_0^2, a_0^2, \dots, a_0^2] = \varphi_\pi[a_0, a_0, \dots, a_0] = y^k(1+y)^{-k},$$
$$\varphi_\pi[c_0, c_0, \dots, c_0] = (1+y)^{-k}.$$

By (9.18) on page 148 in Nica and Speicher (2006), if $\pi \in NC(h)$ has k blocks then $K(\pi)$ has $(h-k+1)$ many blocks. Therefore, (5.23) equals

$$\sum_{k=1}^{h} \#\{\pi \in NC(h) : \pi \text{ has } k \text{ blocks}\} y^{k-1} = \sum_{k=1}^{h} \frac{1}{k} \binom{h-1}{k-1} \binom{h}{k-1} y^{k-1},$$

which is the hth moment of the Marčenko–Pastur law [see (3.19)]. For the last equality, see page 144 of Nica and Speicher (2006). This proves (a).

(b) Observe that, by (3.11),

$$\lim p^{-1} E\, \mathrm{Tr}\big((np^{-1})^2 \hat{\Gamma}_u(\varepsilon)\hat{\Gamma}_u^*(\varepsilon)\big)^h = y^{-(2h+1)}(1+y)\varphi\big(\gamma_{u0}\gamma_{u0}^*\big)^h \qquad \forall h \geq 1,$$

where, by (3.10), $\gamma_{u0} = (1+y)a_0 s c_u s a_0$ and $\gamma_{u0}^* = (1+y)a_0 s c_u^* s a_0$. Since the marginal distribution of all the c_i's are same for $i \geq 1$, using free independence of Span$\{s\}$, $\{a_i, a_i^*\}$ and $\{c_i, c_i^*\}$, the LSD of $(\frac{n}{p})^2 \hat{\Gamma}_u(\varepsilon)\hat{\Gamma}_u^*(\varepsilon)$ are identical for all $u \geq 1$.

Now we show that the LSD is the free Bessel law. Let

$$NCE(2n) = \{\pi \in NC(2n) : \text{ every block of } \pi \text{ has even cardinality}\}.$$

By (5.9), the hth order moment of the LSD of $(\frac{n}{p})^2 \hat{\Gamma}_u(\varepsilon)\hat{\Gamma}_u^*(\varepsilon)$ is given by

$$\text{(5.24)} \quad \frac{(1+y)^{2h+1}}{y^{2h+1}} \sum_{\pi \in NC(2h)} \varphi_{K(\pi)}[a_0, a_0, \dots, a_0]\varphi_\pi[c_i, c_i^*, c_i, c_i^*, \dots, c_i, c_i^*].$$

Note that $\varphi_\pi[c_i, c_i^*, c_i, c_i^*, \dots, c_i, c_i^*] = 0$ if $\pi \in NC(2h) - NCE(2h)$. If $\pi \in NCE(2h)$ has k many blocks, then $\varphi_\pi[c_i, c_i^*, c_i, c_i^*, \dots, c_i, c_i^*] = (1+y)^k$. Note that by (9.18) on page 148 in Nica and Speicher (2006), $K(\pi)$ has $2h+1-k$ blocks, and hence $\varphi_{K(\pi)}[a_0, a_0, \dots, a_0] = y^{2h+1-k}(1+y)^{2h+1-k}$. Therefore, (5.24) equals

$$y^{-2h} \sum_{k=1}^{h} \#\{\pi \in NCE(2h) : \pi \text{ has } k \text{ blocks}\} y^{2h+1-k-1}$$

[by (3.8) and (3.9)]

$$= \sum_{k=1}^{h} \frac{1}{k} \binom{h-1}{k-1} \binom{2h}{k-1} y^{-k},$$

where the last equality follows from Lemma 4.1 of Edelman (1980). The final expression is indeed the hth moment of the free Bessel$(2, y^{-1})$ law. This proves (b) under (A2).

(c) By (5.20) and (5.21), the Stieltjes transform of LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ satisfies,

$$m(z) = -\left(z - \frac{1}{2\pi}\int_0^{2\pi} \frac{\cos\theta\, d\theta}{1 + ym(z)\cos\theta}\right)^{-1}. \tag{5.25}$$

Now by contour integration, it can be shown that

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{\cos\theta\, d\theta}{1 + ym(z)\cos\theta} = \frac{1}{ym(z)} - \frac{2}{y^2m^2(z)}\frac{1}{\omega_1 - \omega_2},$$

where ω_1 and ω_2 are two roots of $\omega^2 + 2(ym(z))^{-1}\omega + 1 = 0$ with $|\omega_1| > 1$, $|\omega_2| < 1$ and $(\omega_1 - \omega_2)^{-2} = \frac{y^2m^2(z)}{4(1-y^2m^2(z))}$. Therefore, by (5.25), we have

$$\begin{aligned} \frac{-1}{m(z)} &= z - \frac{1}{ym(z)} + \frac{2(\omega_1 - \omega_2)^{-1}}{y^2m^2(z)} \\ &\rightarrow \big((1-y) - zym(z)\big)^2\big(1 - y^2m^2(z)\big) = 1. \end{aligned}$$

Hence, Example 1(c) follows.

5.7. *Justification for Example* 3. By Remark 3.1(a) it is enough to work under assumption (A2). Note that the LSD of ψ_j is δ_{λ_j}. We can write

$$n\Delta_u = Z\left(\sum_{j=0}^{q}\sum_{j'=0}^{q} \lambda_j\lambda_{j'}P_{j-j'+u}\right)Z^*, \qquad n\Delta_u^* = Z\left(\sum_{j=0}^{q}\sum_{j'=0}^{q} \lambda_j\lambda_{j'}P_{j-j'+u}^*\right)Z^*.$$

By Lemma 7.1 of the supplementary file Bhattacharjee and Bose (2015), LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ and $\frac{1}{2}(\Delta_u + \Delta_u^*)$ are identical. Moreover,

$$\frac{1}{2}(\Delta_u + \Delta_u^*) = n^{-1}Z\left(\frac{1}{2}\sum_{j=0}^{q}\sum_{j'=0}^{q} \lambda_j\lambda_{j'}\big(P_{j-j'+u} + P_{j-j'+u}^*\big)\right)Z^*,$$

whose LSD is a compound free Poisson [see discussions around (3.24)] and by (3.24), its rth order free cumulant is given by

$$\begin{aligned} & y^{r-1}\lim\frac{1}{n}\operatorname{Tr}\left(\frac{1}{2}\sum_{j,j'=0}^{q} \lambda_j\lambda_{j'}\big(P_{j-j'+u} + P_{j-j'+u}^*\big)\right)^r \\ &\quad = y^{r-1}E_\theta\big(\cos(u\theta)\tilde{h}(\lambda,\theta)\big)^r, \end{aligned}$$

where $\tilde{h}$ is as given in (3.26) and $\theta \sim U(0, 2\pi)$. Hence, Example 3 is justified.

5.8. *Proof of Remark* 3.3. It is enough to show that (3.25) justifies Theorem 2.1 in Liu, Aue and Paul (2015) when $\psi_j = \lambda_j$.

For a variable a, define the free cumulant generating function

$$C(z) = 1 + \sum_{r=1}^{\infty} k_r z^r, \tag{5.26}$$

where k_r is the rth order free cumulant of a. One can show that

$$-C\bigl(-m(z)\bigr) = zm(z), \tag{5.27}$$

[see (12.5) on page 198 of Nica and Speicher (2006)] where $m(z)$ is the Stieltjes transformation of a.

Here, we consider a to be distributed as LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$. Our goal is to find the power series expansion of $C(-m(z))$ from $m(z)$ in (3.16) and show that the coefficient of $(-m(z))^r$ in that expansion is same as k_{ur} in (3.25).

As $\psi_j = \lambda_j I_p$, (3.18) reduces to (3.26), and hence (3.16) and (3.17) reduce to

$$\frac{1}{m(z)} + z = E_\theta\left(\frac{\cos(u\theta)\tilde{h}(\lambda,\theta)}{1 + y\cos(u\theta)K(z,\theta)}\right) \qquad \text{where} \tag{5.28}$$

$$K(z,\theta) = m(z)\tilde{h}(\lambda,\theta). \tag{5.29}$$

Hence, by (5.27),

$$\begin{aligned} C\bigl(-m(z)\bigr) &= -zm(z) = -m(z)\left(\frac{1}{m(z)} + z\right) + 1 \\ &= -E_\theta\left(\frac{\cos(u\theta)m(z)\tilde{h}(\lambda,\theta)}{1 + y\cos(u\theta)K(z,\theta)}\right) + 1 \qquad \text{by (5.28)} \\ &= -E_\theta\left(\frac{\cos(u\theta)m(z)\tilde{h}(\lambda,\theta)}{1 + y\cos(u\theta)m(z)\tilde{h}(\lambda,\theta)}\right) + 1 \qquad \text{by (5.29)} \\ &= 1 + \sum_{r=1}^{\infty} y^{r-1} E_\theta\bigl(\cos(u\theta)\tilde{h}(\lambda,\theta)\bigr)^r \bigl(-m(z)\bigr)^r. \end{aligned} \tag{5.30}$$

Therefore, for all $r \geq 1$, the coefficient of $(-m(z))^r$ in (5.30) agrees with k_{ur} in (3.25). This completes the proof.

5.9. *Detailed calculation for Example 5 in Section 4.2.3.* Note that $X_t = \varepsilon_t + \varepsilon_{t-1}$, where $\varepsilon_t \sim \mathcal{N}(0, I_p)$ and I_p is the identity matrix of order p. Suppose $n = p$. Therefore, for every $i \geq 1$, $(\varepsilon_{t,i} + \varepsilon_{t-1,+i}) \sim \mathcal{N}(0, 2)$, and hence

$$E(\varepsilon_{t,i} + \varepsilon_{t-1,+i})^2 = 2, \qquad E(\varepsilon_{t,i} + \varepsilon_{t-1,+i})^4 = 12, \tag{5.31}$$

$$\begin{aligned} E\bigl(\mathrm{Tr}(\hat{\Gamma}_0)\bigr) &= n^{-1}\sum_{t_1,i_1} E\bigl(X_{t_1,i_1}^2\bigr) = n^{-1}\sum_{t_1,i_1} E(\varepsilon_{t_1,i_1} + \varepsilon_{t_1-1,i_1})^2 \\ &= 2(n-1) \end{aligned} \tag{5.32}$$

and

$$
\begin{aligned}
&E\big(\mathrm{Tr}(\hat{\Gamma}_0) - E\,\mathrm{Tr}(\hat{\Gamma}_0)\big)^2 \\
&\quad = E\big(\mathrm{Tr}(\hat{\Gamma}_0)\big)^2 - \big(E\,\mathrm{Tr}(\hat{\Gamma}_0)\big)^2 \\
&\quad = n^{-2} \sum_{t_1,t_2,i_1,i_2} E\big(X_{t_1,i_1}^2 X_{t_2,i_2}^2\big) - 4(n-1)^2 \qquad \big[\text{by (5.32)}\big] \\
&\quad = n^{-2} \sum_{t_1 \neq t_2, i_1, i_2} E\big(X_{t_1,i_1}^2\big) E\big(X_{t_2,i_2}^2\big) + n^{-2} \sum_{t_1 = t_2, i_1 = i_2} E\big(X_{t_1,i_1}^4\big) \\
&\qquad + n^{-2} \sum_{t_1 = t_2, i_1 \neq i_2} E\big(X_{t_1,i_1}^2\big) E\big(X_{t_2,i_2}^2\big) - 4(n-1)^2 \\
&\quad = 4(n-1)(n-2) + 12 + 4\frac{(n-1)^2}{n} - 4(n-1)^2 \qquad \big[\text{by (5.31)}\big] \\
&\quad = 12 + 4\frac{n-1}{n}\big(n^2 - 2n + n - 1 - n^2 + n\big) = 12 - 4\frac{n-1}{n} \\
&\quad \to 8.
\end{aligned}
$$

Acknowledgements. We thank Octavio Arizmendi and Carlos Vargas Obieta for interesting and helpful discussions on free probability. We thank Debashis Paul for interesting discussions on their article. We thank all the three referees for their extremely constructive comments which have led to a very significant improvement in the article in terms of both substance and presentation.

SUPPLEMENTARY MATERIAL

Supplement to "Large sample behaviour of high dimensional autocovariance matrices" (DOI: 10.1214/15-AOS1378SUPP; .pdf). In this supplement, we provide additional technical details and simulations.

REFERENCES

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

BAI, Z. and ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613

BHATTACHARJEE, M. and BOSE, A. (2014). Estimation of autocovariance matrices for infinite dimensional vector linear process. *J. Time Series Anal.* **35** 262–281. MR3194965

BHATTACHARJEE, M. and BOSE, A. (2015). Supplement to "Large sample behaviour of high dimensional autocovariance matrices." DOI:10.1214/15-AOS1378SUPP.

EDELMAN, P. H. (1980). Chain enumeration and noncrossing partitions. *Discrete Math.* **31** 171–180. MR0583216

FORNI, M. and LIPPI, M. (2001). The generalized dynamic factor model: Representation theory. *Econometric Theory* **17** 1113–1141. MR1867540

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000). The generalized dynamic factor model: Identification and estimation. *Rev. Econom. Statist.* **82** 540–554.

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2004). The generalized dynamic factor model: Consistency and rates. *J. Econometrics* **119** 231–235.

HONG, Y. and LEE, Y. J. (2003). Consistent testing for serial uncorrelation of unknown form under general conditional heteroscedasticity. Dept. Economics, Cornell Univ. Preprint.

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

LIU, H. (2013). Spectral analysis of high dimensional time series. Ph.D. thesis. Univ. California, Davis.

LIU, H., AUE, A. and PAUL, D. (2015). On the Marčenko–Pastur law for linear time series. *Ann. Statist.* **43** 675–712. MR3319140

MARČENKO, V. and PASTUR, L. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik* **1** 457–483.

NICA, A. and SPEICHER, R. (2006). *Lectures on the Combinatorics of Free Probability*. Cambridge Univ. Press, Cambridge.

PFAFFEL, O. and SCHLEMM, E. (2011). Eigenvalue distribution of large sample covariance matrices of linear processes. *Probab. Math. Statist.* **31** 313–329.

SHAO, X. (2011). Testing for white noise under unknown dependence and its application to diagnostic checking for time series models. *Econometric Theory* **27** 312–343.

SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339.

SILVERSTEIN, J. W. and BAI, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54** 175–192.

XIAO, H. and WU, W. B. (2014). Portmanteau test and simultaneous inference for serial covariances. *Statist. Sinica* **24** 577–599.

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

ZEITOUNI, O., ANDERSON, G. W. and GUIONNET, A. (2010). *An Introduction to Random Matrices*. Cambridge Univ. Press, New York.

STATISTICS AND MATHEMATICS UNIT
INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD, KOLKATA 700108
INDIA
E-MAIL: monaiidexp.gamma@gmail.com
bosearu@gmail.com