The Annals of Statistics
2019, Vol. 47, No. 6, 3470–3503
https://doi.org/10.1214/18-AOS1785

JOINT CONVERGENCE OF SAMPLE AUTOCOVARIANCE MATRICES WHEN $p/n \to 0$ WITH APPLICATION

BY MONIKA BHATTACHARJEE[1] AND ARUP BOSE[2]

University of Florida and Indian Statistical Institute

Consider a high-dimensional linear time series model where the dimension p and the sample size n grow in such a way that $p/n \to 0$. Let $\hat{\Gamma}_u$ be the uth order sample autocovariance matrix. We first show that the LSD of *any* symmetric polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*, u \geq 0\}$ exists under independence and moment assumptions on the driving sequence together with weak assumptions on the coefficient matrices. This LSD result, with some additional effort, implies the asymptotic normality of the trace of any polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*, u \geq 0\}$. We also study similar results for several independent MA processes.

We show applications of the above results to statistical inference problems such as in estimation of the unknown order of a high-dimensional MA process and in graphical and significance tests for hypotheses on coefficient matrices of one or several such independent processes.

1. Introduction.

A general high-dimensional linear time series model is the infinite dimensional moving average process of order q (MA(q)), where q may be finite or infinite. For this process, the sample $\{X_{t.p}^{(n)} : t = 1, 2, \ldots, n\}$ of size n satisfies

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

For all t, $X_{t.p}^{(n)}$ and $\varepsilon_{t.p} = (\varepsilon_{t,1}, \epsilon_{t,2}, \ldots, \epsilon_{t,p})'$ are p-dimensional vectors and $\psi_{j.p}^{(n)}$ are $p \times p$ *coefficient matrices* and $\psi_{0.p}^{(n)} = I_p$. Moreover, $p, n \to \infty$. For $q = \infty$, appropriate assumptions given later guarantee that the sum in (1.1) is meaningful. Precise assumptions on independence and finiteness of moments for $\{\varepsilon_{t,i}\}$, growth of p, n and conditions on the coefficient matrices are discussed later. For convenience, we will write ψ_j, ε_t and X_t, respectively, for $\psi_{j.p}^{(n)}$, $\varepsilon_{t.p}$ and $X_{t.p}^{(n)}$.

Received October 2017; revised July 2018.

[1]Most of this work forms a part of the author's Ph.D. thesis, written at the Theoretical Statistics and Mathematics Unit in the Indian Statistical Institute, Kolkata, India.

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

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

Key words and phrases. Moving average process, sample autocovariance matrices, limiting spectral distribution, trace, asymptotic normality, estimation, testing of hypothesis.

The infinite dimensional vector autoregressive processes (IVAR) and infinite dimensional autoregressive moving average processes (IVARMA), can be expressed as infinite dimensional MA(∞) processes; for details, see Bhattacharjee and Bose (2014). Statistical inference in such time series models when both p and n are large, is relatively underdeveloped.

Key quantities in time series are the *population autocovariance matrices*,

$$\Gamma_{u.p} := E\big(X_{t.p}X^*_{(t+u).p}\big) = \sum_{j=0}^{q-u} \psi_j \psi^*_{j+u}, \qquad u = 0, 1, \dots. \tag{1.2}$$

Their estimators are the *sample autocovariance matrices*

$$\hat{\Gamma}_{u.p} = \frac{1}{n}\sum_{t=1}^{n-u} X_{t.p}X^*_{(t+u).p}, \qquad 0 \le u \le n-1. \tag{1.3}$$

We write Γ_u and $\hat{\Gamma}_u$, respectively, for $\Gamma_{u.p}$ and $\hat{\Gamma}_{u.p}$. Our major goals are:

(1) Establish the Limiting Spectral distribution (LSD) and the asymptotic normality of trace of symmetric polynomials in $\{\hat{\Gamma}_u, \hat{\Gamma}^*_u\}$, preferably under as weak conditions as possible.

(2) Address one and two sample statistical inference problems in high-dimensional time series, using results from (1).

Suppose the dimension $p \to \infty$, and the sample size $n = n(p) \to \infty$ such that $p/n \to y$. There are three possibilities on the growth of p and n:

(I) $y \in (0, \infty)$, that is, p and n grow at the same rate. This case has been well studied in the literature. Some references are provided later in the discussions.

(II) $y = 0$, that is, p grows slower than n. We shall concentrate on this case and it is significantly different from Case I.

(III) $y = \infty$, that is, p grows faster than n. This case will be considered elsewhere.

1.1. *Results on LSD and trace.* Suppose R_p is a $p \times p$ (random) matrix where $p \to \infty$. Its empirical spectral distribution (ESD) is the (random) probability distribution with mass $1/p$ at each eigenvalue of R_p. If it converges weakly (almost surely) to a probability distribution, then the latter is called the limiting spectral distribution (LSD) of R_p. Incidentally, the study of the LSD for non-Hermitian matrices is extremely difficult and very few general results are known. We shall consider LSD of only symmetric matrices. When we have more than one sequence of random matrices of the same order, a most natural way to study their joint convergence is through the LSD of their (symmetric) matrix polynomials.

Recent LSD results on sample autocovariance matrices has attracted the attention of statisticians due to their usefulness for inference in high-dimensional time series models. We shall discuss this aspect in the next section.

For Case I, LSD of symmetric polynomials in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$ is well studied in the literature; see Pfaffel and Schlemm (2011), Liu, Aue and Paul (2015) and Bhattacharjee and Bose (2016a). Consider the simplest case of a $p \times n$ matrix Z, whose entries are i.i.d. with mean 0, variance 1 and have finite 4-th order moment. Then the almost sure LSD of $n^{-1}ZZ^*$ is the Marčenko–Pastur law with parameter y (see Marčenko and Pastur (1967)). The main idea behind one of the available proofs is to embed Z into a Wigner matrix (a symmetric matrix with i.i.d. entries) of order $(p+n) \times (p+n)$ and then use the *asymptotic freeness* of Wigner and deterministic matrices.

Case II is very different. In this case, it is not hard to check that the LSD of $n^{-1}ZZ^*$ is degenerate at 1. However, the almost sure LSD of $\sqrt{np^{-1}}(n^{-1}ZZ^* - I_p)$ is the standard *semicircle law* (see Bai and Yin (1988)). The proof of this result is significantly harder and involved. The embedding technique mentioned above fails as the growth rates of p and n are different. It gets even harder for general polynomials of autocovariance matrices in the general MA(q) model.

For the general MA(q) model in Case II, the only work in the literature is of Wang, Aue and Paul (2017) who established the existence of LSD for the specific symmetric polynomial, $\sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - \Gamma_u - \Gamma_u^*)$ along with a formula (via a pair of nonlinear equations) for the Stieltjes transformation of the limit. They assumed that $\{\psi_j\}$ are Hermitian, norm bounded (bounded largest singular value) and simultaneously diagonalizable; for details, see Assumption (WAP2) and Theorem 2.1 stated later. Their result continues to hold for the MA(∞) process, under an additional summability assumption on the largest singular value of $\{\psi_j\}$ (see (WAP3)). (WAP2) and (WAP3) are strong assumptions on $\{\psi_j\}$ and exclude many interesting moving average processes (see the introduction of Bhattacharjee and Bose (2016a)).

As a corollary to one of our results, we show that the above result continues to hold under significantly weaker assumptions (see Corollary 2.1(a)(ii)). We drop the Hermitian and simultaneously diagonalizable assumption on $\{\psi_j\}$ and instead assume the following minimal condition:

(B1) $\{\psi_j\}$ are norm bounded and jointly converge.

By "joint convergence" of matrices we mean "convergence in $*$-distribution" of noncommutative variables in the sense of Definition 8.13 in Nica and Speicher (2006).

In particular, (WAP2) implies Assumption (B1). For MA(∞) process also, our Assumption (B2) (given later) on $\{\psi_j\}$ is weaker than (WAP3).

Now consider other matrices formed from sample autocovariance matrices. For example, the singular values of $\hat{\Gamma}_u$ are the eigenvalues of the symmetric product $\hat{\Gamma}_u\hat{\Gamma}_u^*$. This is a completely different LSD problem and no LSD result is known. Indeed, one may wish to consider more general symmetrizations that involve several $\hat{\Gamma}_u$, $\hat{\Gamma}_u^*$. A motivation is provided by white noise tests in the one-dimensional case—all such tests are based on quadratic functions of several autocovariances (see, e.g., Shao (2011) and Xiao and Wu (2014)). The analogous objects in our model (1.1) are quadratic polynomials of sample autocovariance matrices of different orders. Further, taking a cue from real valued time series models, it is easy to be convinced that in high-dimensional models, estimation and testing for parameters in one and two sample problems would involve some polynomials of the autocovariance matrices. Thus we are naturally led to the consideration of matrix polynomials of autocovariances. More detailed motivation for studying functions of autocovariance matrices are laid out in Section 1.2 on statistical inference.

Throughout this paper,

$$\Pi_{\text{sym}}(\cdot) \qquad \text{is a finite degree symmetric matrix polynomial.} \tag{1.4}$$

Suppose $\{\varepsilon_{t,i} : t, i \geq 1\}$ are independently distributed with mean 0, variance 1 and their moments of all order exist and are uniformly bounded. Further suppose that (B1) holds. Then in Theorem 2.2 we prove that the LSD of

$$R_{\Pi_{\text{sym}}} = \sqrt{np^{-1}}\big(\Pi_{\text{sym}}(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0) - \Pi_{\text{sym}}(\Gamma_u, \Gamma_u^* : u \geq 0)\big) \tag{1.5}$$

exists almost surely. This is a significant generalization of Wang, Aue and Paul (2017). Moreover, while they express the LSD (in the special case) via a pair of Stieltjes transform equations, we write the LSD in terms of polynomials of some *freely independent* variables. In Section 1.2, we shall see that such descriptions are useful in testing of hypotheses as they provide insightful information on the nature of the LSD under both null and alternative hypotheses. In contrast, Stieltjes transform descriptions do not provide such insight and are also computationally prohibitive to be easily applicable.

Now a word about the uniform boundedness assumption on the moments of $\{\varepsilon_{t,i} : t, i \geq 1\}$ in Theorem 2.2. This is needed because we are dealing with *all* symmetric polynomials in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$ together. For specific choices of polynomials, this assumption can be weakened. In Corollaries 2.1 and 2.2, we respectively state the existence of the LSD of $\sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - \Gamma_u - \Gamma_u^*)$ and of $\sqrt{np^{-1}}(\hat{\Gamma}_u\hat{\Gamma}_u^* - \Gamma_u\Gamma_u^*)$, along with their limiting Stieltjes transformations, under weaker moment assumptions on $\{\varepsilon_{t,i}\}$ (see Assumption (A3)). In particular, Corollary 2.1 implies the result of Wang, Aue and Paul (2017). Corollary 2.2 is a new result.

The results of Wang, Aue and Paul (2017) are for a single MA(q) process and so is Theorem 2.2. However, our method can tackle polynomials in autocovariance matrices of two or more independent MA(q) processes; see Theorem 2.3 and Corollary 2.3. The proof of this theorem uses asymptotic freeness results for independent and deterministic matrices.

As we shall see in the next section, these results are useful to develop graphical tests for different hypotheses in one and two sample problems.

Statistical inference in high-dimensional models is often based on linear spectral statistics defined as $\sum f(\lambda_i)$ where f is a suitable function and $\{\lambda_i\}$ are eigenvalues of a matrix. There is a large literature on the spectral statistics of high-dimensional variance–covariance matrices and its application in statistical inference; see, for example, Diaconis and Evans (2001), Bai and Silverstein (2004), Bai, Wang and Zhou (2009), Zheng (2012), Bai et al. (2013), Bao et al. (2015) and Zheng, Bai and Yao (2015). However, apparently there are no results on linear spectral statistics of general autocovariance matrices. This would be an important topic of independent research. While we do not consider such statistics in general, we consider the behaviour of traces (of matrix polynomials of Γ_u). The trace is a particular linear spectral statistics with $f(x) = x$.

To state the trace result, we need the following notation. Let

$$\Pi(\cdot) \quad \text{be a finite degree polynomial, and} \tag{1.6}$$

$$R_\Pi = \sqrt{np^{-1}}\big(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0) - \Pi(\Gamma_u, \Gamma_u^* : u \geq 0)\big). \tag{1.7}$$

Let

$$\sigma_{R_\Pi}^2 = \lim E\big(\operatorname{Tr}(R_\Pi)\big)^2.$$

Making use of the arguments in the proof of Theorem 2.2, we show (see Theorem 2.4)

$$\operatorname{Tr} R_\Pi \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma_{R_\Pi}^2\big). \tag{1.8}$$

Theorem 2.5 establishes the two sample version of Theorem 2.4.

In Corollary 2.4, we show the asymptotic normality of $\sqrt{np^{-1}}(\operatorname{Tr}(\hat{\Gamma}_u) - \operatorname{Tr}(\Gamma_u))$ and $\sqrt{np^{-1}}(\operatorname{Tr}(\hat{\Gamma}_u\hat{\Gamma}_u^*) - \operatorname{Tr}(\Gamma_u\Gamma_u^*))$, under weaker moment assumptions. Corollary 2.5 is its two sample version. Note that for these results on trace, symmetry of the matrix polynomials is not required. In the next section, we shall discuss how these results can be used to construct asymptotically valid tests for different hypotheses in one and two sample problems.

1.2. *Statistical applications.* Statistical inference in high-dimensional time series is relatively underdeveloped. We believe that spectral properties of autocovariance matrices have a crucial role to play in this development. Wang, Aue and Paul (2017) also mentioned some possible applications of their results. We discuss a few statistical inference problems in one and two sample situations to demonstrate how the results discussed in Section 1.1 can be used. In Section 3, we discuss some problems which cannot be handled if we use the restrictive results of Wang, Aue and Paul (2017). In each case, we provide both graphical and significance tests. Graphical tests are based on LSD results given in Theorems 2.2 and Corollary 2.1 and significance tests are constructed using Theorem 2.4 and Corollary 2.4. Until now, inference for two independent high-dimensional moving average processes has not been considered in the literature. We discuss some testing of hypotheses for two independent MA(q) processes using Theorems 2.3 and 2.5.

First, consider the following simple testing problem:

$$H_{01} : X_t = \varepsilon_t \quad \text{against} \quad H_{11} : X_t = \varepsilon_t + \varepsilon_{t-1}. \tag{1.9}$$

A graphical test is as follows. Using free probability description of the LSD of symmetric polynomials in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$, given in Theorem 3.1 of the Supplementary Material (Bhattacharjee and Bose (2018)), under H_{01} and H_{11}, the LSD of $\frac{1}{2}\sqrt{np^{-1}}(\hat{\Gamma}_2 + \hat{\Gamma}_2^*)$ are semicircle distributions with variances 0.5 and 3, respectively. A semicircle distribution with variance σ^2 has the following density:

$$f(x) = \begin{cases} \dfrac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}, & -2\sigma \le x \le 2\sigma, \\ 0, & \text{otherwise.} \end{cases} \tag{1.10}$$

The support of this distribution is $[-2\sigma, 2\sigma]$. Therefore, we can plot the histogram of the eigenvalues of $\frac{1}{2}\sqrt{np^{-1}}(\hat{\Gamma}_2 + \hat{\Gamma}_2^*)$ and accept H_{01} if the support of the histogram appears to be $[-2\sqrt{0.5}, 2\sqrt{0.5}] = [-1.41, 1.41]$. In Section 3.1, we generalize this idea to graphically test

$$H_{02} : q = q_0 \quad \text{against} \quad H_{12} : q = q_1 \text{ when } X_t = \sum_{t=0}^{q} \varepsilon_{t-q}. \tag{1.11}$$

We can also use our results on traces to construct asymptotically valid test statistics. For example, by Theorem 2.4,

$$\sqrt{np^{-1}}\,\mathrm{Tr}(\hat{\Gamma}_1) \xrightarrow{\mathcal{D}} N(0, 1), \qquad \text{under } H_{01}, \tag{1.12}$$

$$\sqrt{np^{-1}}\big(\mathrm{Tr}(\hat{\Gamma}_1) - 1\big) \xrightarrow{\mathcal{D}} N(0, 7), \qquad \text{under } H_{11}. \tag{1.13}$$

Therefore, we reject H_{02} at $100\alpha\%$ level of significance if $|\sqrt{np^{-1}}\,\mathrm{Tr}(\hat{\Gamma}_1)| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the upper $\alpha/2$-th quantile of the standard normal distribution. Moreover, this test is consistent by (1.13). In the last part of Section 3.1, we generalize the above idea to construct a significance test for (1.11).

In Section 3.2, we compare the empirical power of our test with the well known Ljung–Box test. Table 2 documents the simulated power of the tests and it is observed that our test performs better. Moreover, the performance of the Ljung–Box test deteriorates as p increases. Whereas due to consistency, our test statistic does increasingly better as p increases.

Next, consider the MA(q) process (1.1) and the following testing problem:

$$H_{03} : q = q_0 \quad \text{against} \quad H_{13} : q = q_1. \tag{1.14}$$

Wang, Aue and Paul (2017) suggested to graphically compare the eigenvalue distributions of the appropriately centered and scaled symmetrized sample autocovariance matrices with the LSD of the same under null hypothesis. Unfortunately, their Stieltjes transform formulae are very complicated and are unable to provide any idea about the shape of the LSD under null and alternative hypotheses that would facilitate graphical comparison.

To construct appropriate tests of significance for the hypotheses (1.14), Wang, Aue and Paul (2017) proposed to consider a class of test statistics that equals the squared integral of the difference between the ESD of renormalized sample autocovariance matrices and the corresponding LSD under null. Calculation of the LSD under null hypothesis requires inversion of the Stieltjes transform formula, which is computationally challenging and involves hard calculations in complex analysis. They also mentioned an alternative test statistic based on computing the differences between the Stieltjes transforms of the ESD and the LSD for a finite set of $z \in \mathbb{C}^+$ and then combining them through some norm. The distribution of either statistic under null and alternative hypotheses are not known, and hence it is difficult to calculate their appropriate cut off values and power.

Our method of testing (1.14) is based on a graphical estimate of q. Consider the MA(q) process given in (1.1). A graphical estimate of q is based on the fact that the LSD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ are all degenerate at 0 if $u > q$ and are not so for $u \leq q$; see Section 3.3. Hence, we can plot the histogram or the CDF of the ESD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ for first few $u \geq 1$. Then q is estimated by the smallest $\hat{q}$ for which the eigenvalue distributions appear concentrated at 0 only for all $u > \hat{q}$. We accept H_{03} if $\hat{q} = q_0$. Simulation results in Figure 1 show that this method performs remarkably well.

One of the referees mentioned that often it may not be easy to decide if the ESD is close to being degenerate. Thus one should base a test on some non-degenerate limit. The LSD of $\sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - \Gamma_u - \Gamma_u^*)$ is nondegenerate but unfortunately cannot be used directly as $\{\Gamma_u\}$ are unknown. A way out is to use a suitable consistent estimator C_u of Γ_u. Indeed, appropriately banded version of $\hat{\Gamma}_u$ are consistent for Γ_u; for example, see Bhattacharjee and Bose (2014). Then we can consider the statistic $\sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - C_u - C_u^*)$. Its LSD is nondegenerate and is the same as the LSD of $\sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - \Gamma_u - \Gamma_u^*)$. This can then be used for a graphical test; for details, see Theorem 3.1 and the corresponding discussion.

Now consider hypothesis testing for $\{\psi_j\}$ for the MA(q) process (1.1). In Section 3.4.1, we provide a graphical method and also an asymptotically valid test statistic to test the hypotheses

$$H_{04} : \psi_q = A_q \quad \text{against} \quad H_{14} : \psi_q \neq A_q \quad \text{and} \tag{1.15}$$

$$H_{05} : \psi_j = A_j \quad \forall j \quad \text{against} \quad H_{15} : \psi_j \neq A_j \text{ for at least one } j \tag{1.16}$$

for some $p \times p$ deterministic matrices $\{A_j\}$. Specific nontrivial choice of $\{A_j\}$ is also given in this section. Since all our results refer to eigenvalue distribution, by equality of two matrices we mean that their eigenvalue distributions are identical.

A graphical method to test (1.15) is based on the LSD of $(\hat{\Gamma}_q + \hat{\Gamma}_q^* - A_q - A_q^*)$ and a significance test is based on the (asymptotic) distribution of $\sqrt{np^{-1}}(\mathrm{Tr}(\hat{\Gamma}_q) - \mathrm{Tr}(A))$.

As H_{05} and H_{15} specify all the coefficient matrices $\{\psi_j : 0 \le j \le q\}$, it is natural to base any test for (1.16) on $\{\hat{\Gamma}_u\}$, $1 \le u \le q$. Observe that the results of Wang, Aue and Paul (2017) are then of no use.

Let

$$\hat{G}_q = \sum_{u=1}^{q} \hat{\Gamma}_u \quad \text{and} \quad G_{qH_0} = \sum_{u=1}^{q} \sum_{j=0}^{q} A_j A_{j+u}^*. \tag{1.17}$$

We provide a graphical test based on the LSD of $(\hat{G}_q + \hat{G}_q^* - G_{qH_0} - G_{qH_0}^*)$ and a significance test based on the asymptotic normality of $\sqrt{np^{-1}}(\mathrm{Tr}(\hat{G}_q) - \mathrm{Tr}(G_q))$. These LSD and asymptotic normality are consequences of Theorems 2.2 and 2.4.

Examples 6 and 7 contain statistical application of asymptotic normality of $\sqrt{np^{-1}}(\mathrm{Tr}(\hat{\Gamma}_u \hat{\Gamma}_u^*) - \mathrm{Tr}(\Gamma_u \Gamma_u^*))$ in hypothesis testing. Examples 8–11 deal with some interesting tests on IVAR and IVARMA processes which need asymptotic normality of polynomials in sample autocovariance matrices other than symmetric sums and products.

In Section 3.4.2, we discuss hypothesis testing for the equality of coefficient matrices of two independent MA(q) processes, where the joint convergence results of polynomials in independent autocovariance matrices are used.

The Supplementary Material (Bhattacharjee and Bose (2018)) provides all the technical details. We shall often refer to technical lemmas, corollaries and examples from the Supplementary Material. For example, Lemma 1.2 in the Supplementary Material will be referred here as Lemma S1.2.

2. Results on LSD and trace. In this section, we present our main theoretical results in the general MA(q) setup. These will be put to use in Section 3 for statistical inference.

2.1. *Existing LSD result.* The only existing LSD result in this area is Wang, Aue and Paul (2017), proved under the assumptions (WAP1)–(WAP3):

(WAP1) $\{\varepsilon_{t,i}\}$ are i.i.d. with mean 0, variance 1 and $E|\varepsilon_{t,i}|^4 < \infty$.

(WAP2) $\{\psi_j\}$ are Hermitian and simultaneously diagonalizable, norm bounded matrices. There are continuous functions $f_j : \mathbb{R}^m \to \mathbb{R}$ and a unitary matrix U of order p such that $U\psi_j U^* = \mathrm{diag}(f_j(\alpha_1), f_j(\alpha_2), \dots, f_j(\alpha_p))$, $\alpha_j \in \mathbb{R}^m$ for all j and some positive integer m. ESD of $\{\alpha_1, \alpha_2, \dots, \alpha_p\}$ converges weakly to a compactly supported probability distribution F on $\mathbb{R}^m$.

For any matrix A of order p, let

$$\|\psi_j\|_2 = \sqrt{\text{largest eigenvalue of } A^* A}. \tag{2.1}$$

(WAP3) (when $q = \infty$) $\sum_{j=0}^{\infty} j^4 (\sup_p \|\psi_j\|_2) < \infty$.

Let

$$\hat{\Pi}_{a,u} = \sqrt{np^{-1}} \big(2^{-1}(\hat{\Gamma}_u + \hat{\Gamma}_u^*) - 2^{-1}(\Gamma_u + \Gamma_u^*)\big), \qquad \forall u \geq 0. \tag{2.2}$$

THEOREM 2.1 (Wang, Aue and Paul (2017)). *Consider the* MA(q) *process given in* (1.1) *and let* $0 \leq q < \infty$. *Suppose* (WAP1) *and* (WAP2) *hold, and* $n = n(p)$, $p \to \infty$, $p/n \to 0$. *Then the LSD of* $\hat{\Pi}_{a,u}$ *exists almost surely and its Stieltjes transform can be expressed as the unique solution of the system of equations* S(4.5)–S(4.8). *This continues to hold for the case* $q = \infty$ *under the additional assumption* (WAP3).

Theorem 2.1 is only for the additively symmetrized autocovariance matrix. Also note that (WAP2) is fairly restrictive. It excludes many interesting time series models. For example, consider the MA(2) process

$$X_t = \varepsilon_t + C\varepsilon_{t-1} + D\varepsilon_{t-2}, \qquad \forall t \tag{2.3}$$

where the $p \times p$ matrices $C = ((I(1 \leq i = j \leq [p/2]) - I([p/2] + 1 \leq i = j \leq p)))$ and $D = ((I(i + j = p + 1)))$. Note that C and D are not simultaneously diagonalizable. Hence, Theorem 2.1 is not applicable.

Moreover, Theorem 2.1 deals with one autocovariance matrix at a time in a single moving average process and hence cannot be used to claim any LSD result for any function of several sample autocovariance matrices from one or more independent moving average processes. Such LSD results would be useful in one and two sample inference problems; see Section 3.

We shall impose a much relaxed condition on $\{\psi_j\}$ and shall still be able to establish the LSD for all self-adjoint polynomials of $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$ for one (or more than one) MA(q) processes.

2.2. *LSD results for one and many samples.* We need the following assumptions.

(A1) $\{\varepsilon_{t,i}\}$ are independent with $E(\varepsilon_{i,j}) = 0$, $E|\varepsilon_{i,j}|^2 = 1$, $\sup_{i,j} E|\varepsilon_{i,j}|^4 < \infty$.

(A2) For some $\eta, \delta > 0$,

$$P\big(|\varepsilon_{i,j}| < \eta p^{\frac{1}{2+\delta}}\big) = 1, \qquad \forall 1 \leq i \leq n, 1 \leq j \leq p$$

or, $\sup_{i,j} E|\varepsilon_{i,j}|^k < C_k < \infty$ for all $k \geq 1$.

Later we shall replace (A2) by weaker moment assumptions while dealing with specific polynomials, including $\hat{\Pi}_{a,u}$.

(B1) $\{\psi_j\}$ are norm bounded and converge jointly.

By "joint convergence" of matrices we mean "convergence in $*$-distribution" of noncommutative variables in the sense of Definition 8.13 in Nica and Speicher (2006) (see Section S2 for more details).

It is easy to see that (WAP2) implies (B1).

(B2) (when $q = \infty$) $\sum_{j=0}^{\infty} \|\psi_j\|_2 < \infty$.

Note that (WAP3) implies (B2).

The proof of the following theorem is given in Section S1. This, in conjunction with Corollary 2.1(a)(ii) and (b), is a significantly more general version of Theorem 2.1 and does not use the restrictive (WAP2) or (WAP3). Recall $R_{\Pi_{\mathrm{sym}}}$ defined in (1.5).

THEOREM 2.2. *In the model* (1.1), *suppose* $0 \le q < \infty$, (A1), (A2), (B1) *hold and* $n = n(p)$, $p \to \infty$, $p/n \to 0$. *When* $q = \infty$ *assume further that* (B2) *holds. Then the almost sure LSD of* $R_{\Pi_{\mathrm{sym}}}$ *exists.*

The above LSD can be described in terms of freely independent variables; for details, see Sections S2 and S3.

REMARK 2.1. LSD results for symmetric polynomials in sample autocovariance matrices when $p/n \to y > 0$ are discussed in Bhattacharjee and Bose (2016a). These results for $p/n \to 0$ and $p/n \to y > 0$ are *very* different. For example, consider the model $X_t = \varepsilon_t + \varepsilon_{t-1}$. By the results given in Bhattacharjee and Bose (2016a), we can say that the LSD of $\hat{\Gamma}_1 + \hat{\Gamma}_1^*$ exists and the moment sequence of the LSD is given by $\beta_k = 2^k + O(y)$ for all $k \ge 1$. Thus when $y = 0$, this LSD will be degenerate at 2. Bhattacharjee and Bose (2016a) cannot say anything further. Theorem 2.2 proves that the LSD of $0.5\sqrt{np^{-1}}(\hat{\Gamma}_1 + \hat{\Gamma}_1 - 2I)$ is the semicircle variable with variance given in S(10.1) (with $q = 1$, $u = 1$).

The proofs of results for $p/n \to 0$ are harder and more involved compared to $p/n \to y > 0$. In the latter case, we construct a $p \times n$ independent matrix $Z = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$, embed Z into a larger Wigner matrix of order $(p + n)$ and then use asymptotic freeness of Wigner and deterministic matrices. This embedding technique fails for $p/n \to 0$ as the growth rates of p and n are different and makes this case challenging. It gets even harder for general polynomials of autocovariance matrices in the general MA(q) model. Finally, it is not at all easy to guess the nature of the limit for $y = 0$ case from the $y > 0$ case.

In the next two corollaries, we relax (A2) and consider weaker moment assumption for two specific polynomials. We need the following assumption:

(A3) For some $\delta > 0$,

$$\lim_{p \to \infty} \frac{1}{p^2} \sum_{t=1}^{n} \sum_{j=1}^{p} E\left(|\varepsilon_{t,j}|^{2+\delta} I\left(|\varepsilon_{t,j}| > \eta p^{\frac{1}{2+\delta}}\right)\right) = 0.$$

COROLLARY 2.1. (a) *Consider the model* (1.1) *and* $0 \le q < \infty$.

(i) *Suppose Assumptions* (A1), (A2), (B1) *hold and* $n(p), p \to \infty$, $p/n \to 0$. *Then the LSD of* $\hat{\Pi}_{a,u}$ *exists almost surely. The Stieltjes transform of the LSD is given in Theorem* S4.1(a).

(ii) *The result in* (i) *holds if instead of* (A2) *we assume* (WAP1) *or* (A3).

(iii) *The almost sure LSD of* (i) (*or* (ii)) *are identical whenever* $u > q$.

(b) *The results in* (i) *and* (ii) *continue to hold for an* MA(∞) *process if in addition* (B2) *holds*.

Corollary 2.1(a)(i) is immediate from Theorem 2.2 except the Stieltjes transform, which is derived in Section S4. Part (a)(ii) is established in Section S5 via truncation. Proof of (a)(iii) appears in Section S7. Proof of (b) is similar to the proof of the same for the case $p/n \to y \in (0, \infty)$, which is given in Theorem 7.3.4 in Bhattacharjee (March, 2016). Hence we omit it.

Corollary 2.1(a)(ii) and (b) imply Theorem 2.1 of Wang, Aue and Paul (2017). The argument is similar to Corollary 2 in Bhattacharjee and Bose (2016b) and the discussion after that. We omit the details.

The next corollary is apparently new. Let

$$\hat{\Pi}_{m,u} = \sqrt{np^{-1}}(\hat{\Gamma}_u \hat{\Gamma}_u^* - \Gamma_u \Gamma_u^*). \tag{2.4}$$

COROLLARY 2.2. (a) *Consider the model* (1.1) *and* $0 \le q < \infty$.

(i) *Suppose Assumptions* (A1), (A2), (B1) *hold and* $n(p), p \to \infty$, $p/n \to 0$. *Then the LSD of* $\hat{\Pi}_{m,u}$ *exists almost surely. The Stieltjes transform of the LSD is given in Theorem* S4.1(b).

(ii) *The result in* (i) *holds if instead of* (A2) *we assume* (WAP1) *and* $E|\varepsilon_{i,j}|^8 < \infty$ or, (A3) *and* $\sup_{i,j} E|\varepsilon_{i,j}|^8 < \infty$.

(iii) *The almost sure LSD of* (i) (*or* (ii)) *are identical whenever* $u > q$.

(b) *The results in* (i) *and* (ii) *continue to hold for an* MA(∞) *process if in addition* (B2) *holds*.

Corollary 2.2(a)(i) is immediate from Theorem 2.2. The Stieltjes transform of the LSD of $\hat{\Pi}_{m,u}$ is derived in the proof of Theorem S4.1(b). Proof of Corollary 2.2(a)(ii) and (iii) are respectively given in Sections S6 and S7. We omit the proof of Corollary 2.2(b) for the same reasons that were given for Corollary 2.1(b).

A multisample version of Theorem 2.2 is true. We state the two-sample version for simplicity. Its proof is given in Section S8. So, consider another MA(q) process,

$$Y_t = \sum_{j=0}^{q} \phi_j \eta_{t-j}, \tag{2.5}$$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are independent. By the statement '(A1)–(A3), (B1), (B2) hold for $\{Y_t\}$', we mean they hold after replacing $\{\varepsilon_t\}$ and $\{\psi_j\}$, respectively, by $\{\eta_t\}$ and $\{\phi_j\}$. Let $\{\Gamma_{uX}, \hat{\Gamma}_{uX}\}$ and $\{\Gamma_{uY}, \hat{\Gamma}_{uY}\}$ be the population and sample autocovariance matrices respectively for the processes $\{X_t\}$ and $\{Y_t\}$. Also let an arbitrary symmetric polynomial in population and sample autocovariance matrices of $\{X_t\}$ and $\{Y_t\}$ be denoted by

$$\Pi_{\mathrm{sym},XY} = \Pi(\Gamma_{u,X}, \Gamma^*_{u,X}, \Gamma_{u,Y}, \Gamma^*_{u,Y} : u \geq 0), \tag{2.6}$$

$$\hat{\Pi}_{\mathrm{sym},XY} = \Pi(\hat{\Gamma}_{u,X}, \hat{\Gamma}^*_{u,X}, \hat{\Gamma}_{u,Y}, \hat{\Gamma}^*_{u,Y} : u \geq 0). \tag{2.7}$$

THEOREM 2.3. *Suppose all the assumptions of Theorem 2.2 hold for both $\{X_t\}$ and $\{Y_t\}$ and $n = n(p)$, $p \to \infty$, $p/n \to 0$. Then the almost sure LSD of $\sqrt{np^{-1}}(\hat{\Pi}_{\mathrm{sym},XY} - \Pi_{\mathrm{sym},XY})$ exists.*

As in Corollaries 2.1 and 2.2, for specific polynomials, reduced moment conditions suffice. Recall $\{\hat{\Pi}_{a,u}, \hat{\Pi}_{m,u}\}$ from (2.2) and (2.4). Let $\{\hat{\Pi}_{a,u,X}, \hat{\Pi}_{m,u,X}\}$ and $\{\hat{\Pi}_{a,u,Y}, \hat{\Pi}_{m,u,Y}\}$ have the obvious meaning. We state the following corollary for the difference $(\hat{\Pi}_{a,u,X} - \hat{\Pi}_{a,u,Y})$ and $(\hat{\Pi}_{m,u,X} - \hat{\Pi}_{m,u,Y})$. These are useful to construct graphical test for the hypotheses comparing coefficient matrices of two independent MA(q) processes; for details, see Example 12. We omit the proof.

COROLLARY 2.3. *Consider the process* (1.1). *Suppose* (A1), (A3) *and* (B1) *hold and $n = n(p)$, $p \to \infty$, $p/n \to 0$.*

(a) *Suppose $0 \leq q < \infty$.*

- (i) *Then the LSD of $(\hat{\Pi}_{a,u,X} - \hat{\Pi}_{a,u,Y})$ exists almost surely for $u \geq 0$.*
- (ii) *Moreover, if $\sup_{i,j}(E|\varepsilon_{i,j}|^8 + E|\eta_{i,j}|^8) < \infty$, the almost sure LSD of $(\hat{\Pi}_{m,u,X} - \hat{\Pi}_{m,u,Y})$, $u \geq 0$, exists.*

(b) *Above results continue to hold for $q = \infty$ if in addition we assume* (B2).

2.3. *Asymptotic normality of traces.* For any random matrix M with eigenvalues $\{\lambda_i\}$ and a "suitable" function f, the linear spectral statistic is given by $\frac{1}{n}\sum_{i=1}^n f(\lambda_i)$. Asymptotic normality of these statistics is extremely useful in statistical inference. While such results are known for many random matrix models, no results are known for general sample autocovariance matrices. We deal with a specific class of linear spectral statistics namely traces of polynomials. This will be useful later in Section 3 to test hypotheses in high-dimensional time series.

Recall R_Π from (1.7). Let

$$\sigma^2_{R_\Pi} = \lim E(\mathrm{Tr}(R_\Pi))^2.$$

Then we have the following theorem. Its proof is given in Section S9. In particular, the above limit exists and is finite under our assumptions.

THEOREM 2.4. *Consider the model* (1.1) *and* $0 \le q < \infty$. *Suppose* (A1), (A2), (B1) *hold and* $n = n(p)$, $p \to \infty$, $p/n \to 0$. *Then* $\mathrm{Tr}(R_\Pi) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{R_\Pi})$. *This also holds for an* MA(∞) *process if we additionally assume* (B2).

The theorem as stated is for one polynomial at a time. If we have a collection of polynomials, then by the Cramér–Wold technique, their joint asymptotic normality follows easily by an application of the above result. We omit the detailed proof. The same comment is true for all subsequent results of this kind. Also note that Π is not assumed to be symmetric for this result.

The following corollary is in the spirit of Corollaries 2.1 and 2.2. We omit its proof. Let

$$\sigma^2_{a,u} = \lim E\big(\mathrm{Tr}(\hat{\Pi}_{a,u})\big)^2 \quad \text{and} \quad \sigma^2_{m,u} = \lim E\big(\mathrm{Tr}(\hat{\Pi}_{m,u})\big)^2.$$

COROLLARY 2.4. *Consider the process* (1.1). *Suppose* (A1), (A3) *and* (B1) *hold and* $n = n(p)$, $p \to \infty$, $p/n \to 0$.

(a) *Suppose* $0 \le q < \infty$.

(i) *Then* $\mathrm{Tr}(\hat{\Pi}_{a,u}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{a,u})$ *for* $u \ge 0$.

(ii) *Moreover, if* $\sup_{i,j} E|\varepsilon_{i,j}|^8 < \infty$, *then* $\mathrm{Tr}(\hat{\Pi}_{m,u}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{m,u})$, $\forall u \ge 0$.

(b) *These continue to hold for* $q = \infty$ *if in addition we assume* (B2).

Proof of the following two-sample version of Theorem 2.4 is omitted. Let Π_{XY} and $\hat{\Pi}_{XY}$ be arbitrary polynomials respectively in population and sample autocovariance matrices of $\{X_t\}$ and $\{Y_t\}$. Let

$$R_{\Pi_{XY}} = \sqrt{np^{-1}}(\hat{\Pi}_{XY} - \Pi_{XY}) \quad \text{and} \quad \sigma^2_{R_{XY}} = \lim E\big(\mathrm{Tr}(R_{\Pi_{XY}})\big)^2.$$

THEOREM 2.5. *Consider the models* (1.1) *and* (2.5) *and* $0 \le q < \infty$. *Suppose* (A1), (A2) *and* (B1) *hold for both* (1.1) *and* (2.5). *Let* $n = n(p)$, $p \to \infty$, $p/n \to 0$. *Then* $\mathrm{Tr}(R_{\Pi_{XY}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{R_{XY}})$. *Also this continues to hold for two* MA(∞) *processes if in addition we assume* (B2) *for both processes*.

The next corollary is the two sample version of Corollary 2.4(a). Part (a)(i), (ii) are immediate from Corollary 2.4(a) as traces for $\{X_t\}$ and $\{Y_t\}$ are independent. We omit the proof.

Let

$$\hat{\Pi}_{m,(u,v),(X,Y)} = \sqrt{np^{-1}}\big(\hat{\Gamma}_{u,X}\hat{\Gamma}^*_{v,Y} - \Gamma_{u,X}\Gamma^*_{v,Y}\big),$$
$$\sigma^2_{m,(u,v),(X,Y)} = \lim E\big(\mathrm{Tr}(\hat{\Pi}_{m,(u,v),(X,Y)})\big)^2.$$

COROLLARY 2.5. *Consider the models* (1.1) *and* (2.5). *Suppose* (A1), (A3) *and* (B1) *hold and* $n = n(p)$, $p \to \infty$, $p/n \to 0$.

(a) *Suppose* $0 \le q < \infty$.

(i) *Then for all* $u \ge 0$,

$$\mathrm{Tr}(\hat{\Pi}_{a,u,X} - \hat{\Pi}_{a,u,Y}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{a,u,X} + \sigma^2_{a,u,Y}).$$

(ii) *If* $\sup_{i,j}(E|\varepsilon_{i,j}|^8 + E|\eta_{i,j}|^8) < \infty$, *then for all* $u \ge 0$,

$$\mathrm{Tr}(\hat{\Pi}_{m,u,X} - \hat{\Pi}_{m,u,Y}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{m,u,X} + \sigma^2_{m,u,Y}).$$

(iii) *Under the same assumptions as in* (ii), *for all* $u, v \ge 0$,

$$\mathrm{Tr}(\hat{\Pi}_{m,(u,v),(X,Y)}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{m,(u,v),(X,Y)}).$$

(b) *These continue to hold for* $q = \infty$ *if in addition we assume* (B2).

REMARK 2.2. Under appropriate causality conditions, IVAR and IVARMA processes can be expressed as MA(∞) processes which satisfy (B2); for details on these causality conditions, see Bhattacharjee and Bose (2014). Therefore, all of the above results on the MA(∞) process also hold for such IVAR and IVARMA processes.

3. Statistical inference. As indicated earlier, there are not too many procedures, especially those based on LSD results that have been developed in high-dimensional time series. We shall now indicate how the results obtained in Section 2 can be substantially used. This includes estimation of the unknown order of the moving average processes and testing of different hypotheses on the coefficient matrices—these are done by graphical methods or consistent tests of significance which rely on the LSD results.

Throughout this section, we assume (A1) and (A3) for both the processes $\{\varepsilon_t\}$ and $\{\eta_t\}$. Moreover, $0 \le q < \infty$ and $n = n(p)$, $p \to \infty$, $p/n \to 0$.

3.1. *Hypothesis testing for* q: *A simple model.* Consider a random sample $\{X_t : 1 \le t \le n\}$ from the process

$$X_t = \sum_{j=0}^{q} \varepsilon_{t-j} \qquad \forall t \ge 0. \tag{3.1}$$

Based on the above sample, suppose we wish to test

$$H_0 : q = q_0 \quad \text{against} \quad H_1 : q = q_1. \tag{3.2}$$

The LSD of $\hat{\Pi}_{a,u}$ given in Corollary 2.1 and its free probability description discussed in Corollary S3.1(a), are useful in this context. Recall the semi-circle distribution defined in (1.10). In Lemma S10.1, we show that the LSD of $\hat{\Pi}_{a,u}$, for $u \geq q$, are all semicircle distributions with variance $\sigma^2_{q,u}$. The expression for $\sigma^2_{q,u}$ for any arbitrary q and u, is cumbersome and is given in S(10.1). However, by Corollary S10.1, whenever $u > q$, $\sigma^2_{q,u}$ depends on q only; for example, $\sigma^2_{1,u} = 3\ \forall u > 1$ and $\sigma^2_{2,u} = 9.5\ \forall u > 2$. For more details, see Examples S1 and S2.

For $u > q$, let σ^2_q denote the common value of $\sigma^2_{q,u}$, which can be computed using (S10.8). An easy graphical way to test the hypotheses in (3.2) is to plot the ESD of $\hat{\Pi}_{a,u}$ for $u = \max(q_0, q_1) + 1$. If the support of the distribution is $[-2\sigma_{q_i}, 2\sigma_{q_i}]$, then we accept H_i.

Significant tests can be designed based on the asymptotic normality of the traces. Let us assume $q_0 < q_1$. A similar method will work for the reverse case. In Corollary S11.1, using Theorem 2.4, we prove that under H_i,

$$\sqrt{np^{-1}}\big(\operatorname{Tr}(\hat{\Gamma}_u) - a_{u,q_i} p\big)/\tau_{u,q_i} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \qquad \text{where}$$

$$\tau^2_{u,q_i} = \frac{1}{2} \sum_{v=-\infty}^{\infty} (a_{v-u,q_i} + a_{v+u,q_i})^2,$$

$$a_{u,q_i} = \big((q_i + 1) - |u|\big) I\big(|u| \leq q_i\big). \tag{3.3}$$

For $0 \leq u \leq q_1$, we can use any of the $T_u = \sqrt{np^{-1}}(\operatorname{Tr}(\hat{\Gamma}_u) - a_{u,q_0} p)/\tau_{u,q_0}$ as a test statistic. The test will reject H_0 at $100\alpha\%$ level of significance if $T_u > z_\alpha$, where z_α is the upper α point of the standard normal distribution.

All of the above tests are consistent. For large p, by asymptotic normality, the power of T_u at level α is approximately

$$1 - \Phi\big(z_\alpha \tau_{u,q_0} \tau^{-1}_{u,q_1} - (a_{u,q_1} - a_{u,q_0}) p \sqrt{np^{-1}} \tau^{-1}_{u,q_1}\big), \qquad 0 \leq u \leq q_1, \tag{3.4}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal variable. It is easy to see that as a_{u,q_1} is nonnegative and a nonincreasing function of $u > 0$, τ_{u,q_1} is smaller and consequently (3.4) is larger for $u = q_1$. Thus for large p, the power of T_{q_1} dominates the power of all other T_u mentioned above.

Following simulation results provide numerical support. Consider the null hypothesis $H_0 : q_0 = 1$ against the two alternatives $H_1 : q_1 = 2$ and $H_1 : q_1 = 3$ for model (3.1). We simulate from model (3.1) for $q_1 = 2, 3$, $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, I_p)$, $n = p^{1.8}$, $p = 300, 500, 700, 1000$ and compute $\{T_u : 0 \leq u \leq q_1\}$ for $q_0 = 1$. Table 1 provides the empirical power (E.P.) of these test statistics computed for 300 replications. Note that E.P. of T_{q_1} dominates E.P. of $\{T_u : 0 \leq u < q_1\}$.

3.2. *Comparison between the above test and the Ljung–Box test.* One of the referees raised the issue of using the well-known multivariate Ljung–Box test for the hypothesis (3.2) in the model (3.1) when $q_0 = 0$. The test statistic is given by

$$T_{\mathrm{LB}} = n \operatorname{Vec}(\hat{C})' \big(I_s \otimes \hat{\Gamma}_0^{-1} \otimes \hat{\Gamma}_0^{-1}\big) \operatorname{Vec}(\hat{C}) \tag{3.5}$$

TABLE 1
Comparison of E.P. of $\{T_u : 0 \le u \le q_1\}$

Model (3.1)	p	T_0	T_1	T_2	T_3
$q_0 = 1, q_1 = 2$	300	0.89	0.90	0.92	
$q_0 = 1, q_1 = 2$	500	0.91	0.93	0.94	
$q_0 = 1, q_1 = 2$	700	0.92	0.95	0.97	
$q_0 = 1, q_1 = 2$	1000	0.94	0.96	0.99	
$q_0 = 1, q_1 = 3$	300	0.85	0.88	0.92	0.94
$q_0 = 1, q_1 = 3$	500	0.90	0.92	0.95	0.97
$q_0 = 1, q_1 = 3$	700	0.94	0.95	0.97	0.98
$q_0 = 1, q_1 = 3$	1000	0.95	0.97	0.98	0.99

where $\hat{C} = (\hat{\Gamma}_1 \hat{\Gamma}_2 \dots \hat{\Gamma}_s)$ (a $p \times ps$ matrix), $\mathrm{Vec}(\hat{C})$ is the vector of dimension $p^2 s$ which is built by stacking the columns of $\hat{C}$ beneath one another, I_s is the identity matrix of order s and $\otimes$ is the Kronecker product of matrices. Under $H_0 : q = 0$, T_{LB} follows asymptotically $\chi^2_{p^2 s}$ and we reject H_0 at $100\alpha\%$ level of significance if the observed value of T_{LB} is greater than the upper α point of $\chi^2_{p^2 s}$ distribution; for more details, see Hosking (1980).

Since the degrees of freedom is large, we can also consider the test statistic $T^*_{\mathrm{LB}} = \frac{T_{\mathrm{LB}} - p^2 s}{\sqrt{2p^2 s}}$. Under H_0, T^*_{LB} is asymptotic normal. Thus we can reject H_0 at $100\alpha\%$ level of significance if the observed $|T^*_{\mathrm{LB}}| > z_{\alpha/2}$ where $z_{\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the standard normal distribution.

For the same hypothesis, results of Section 3.1 provides our test statistic. Substituting $q = q_0 = 0$ in (3.3), we have $a_{u,q_0} = I(u = 0)$ and $\tau_{u,q_0} = I(|u| \ge 1)$. Hence this test statistic reduces to $T = \sqrt{np^{-1}}\,\mathrm{Tr}(\hat{\Gamma}_{q_1})$ for testing against the alternative $H_1 : q = q_1$. Under H_0, T is asymptotically $\mathcal{N}(0, 1)$ and we reject H_0 at $100\alpha\%$ level if observed $|T| > z_{\alpha/2}$ where $z_{\alpha/2}$ is the upper $\alpha/2$ point of $\mathcal{N}(0, 1)$.

Let us compare the empirical power of T_{LB}, T^*_{LB} and T for the two models: $X_t = \varepsilon_t + \varepsilon_{t-1}$ and $X_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2}$ where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}_p(0, I_p)$. We simulate a sample of size $n = p^{1.2}$ where $p = 300, 500, 700, 1000$, compute the test statistic T_{LB}, T^*_{LB} (for $s = 10$) and T, and replicate this a 100 times.

Table 2 provides the empirical powers. Clearly, the empirical power of T dominates those of T_{LB} and T^*_{LB}. Moreover, the latter decrease as p increases whereas, as discussed in Section 3.1, the opposite happens for T.

REMARK 3.1. In Section 3.1, we used test statistics based on $\{\hat{\Gamma}_u : 0 \le u \le q_1\}$ for testing (3.2) in the model (3.1) with $q_0 < q_1$. This is possible because, in the model (3.1), all the elements of $\{\hat{\Gamma}_u : 0 \le u \le q\}$ vary with q. In general, we cannot use $\{\hat{\Gamma}_u : 0 \le u \le q_0\}$ to construct a valid test statistic for testing (3.2) with

TABLE 2
Comparison of empirical power (E.P.) of T_{LB}, T^*_{LB} and T for $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}_p(0, I_p)$, $n = p^{1.2}$ and 100 replications

Model	p	E. P. of T_{LB}	E.P. of T^*_{LB}	E.P. of T
$X_t = \varepsilon_t + \varepsilon_{t-1}$	300	0.83	0.84	0.87
$X_t = \varepsilon_t + \varepsilon_{t-1}$	500	0.76	0.79	0.89
$X_t = \varepsilon_t + \varepsilon_{t-1}$	700	0.69	0.65	0.93
$X_t = \varepsilon_t + \varepsilon_{t-1}$	1000	0.52	0.55	0.96
$X_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2}$	300	0.8	0.8	0.88
$X_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2}$	500	0.73	0.78	0.91
$X_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2}$	700	0.58	0.63	0.94
$X_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2}$	1000	0.52	0.59	0.97

$q_0 < q_1$. For example, consider the following hypotheses:

$$H_0 : X_t = \varepsilon_t + \varepsilon_{t-1} \quad \text{against} \quad H_1 : X_t = \varepsilon_t + \varepsilon_{t-2},$$

$$H_0 : X_t = \varepsilon_t + \varepsilon_{t-1} \quad \text{against} \quad H_2 : X_t = \varepsilon_t + 0.5\varepsilon_{t-1} + \varepsilon_{t-2}.$$

Note that $(\text{Tr}(\Gamma_0), \text{Tr}(\Gamma_1), \text{Tr}(\Gamma_2))$ equals $(2, 1, 0)p$, $(2, 0, 1)p$ and $(2.25, 1, 1)p$, respectively, under H_0, H_1 and H_2. Note that we cannot use $\hat{\Gamma}_0$ to test H_0 against H_1 as asymptotic distribution of $Tr(\Gamma_0)$ under H_0 and H_1 are identical. Similarly, $\hat{\Gamma}_1$ cannot not be used to test H_0 against H_2. It is better to use $\hat{\Gamma}_2$ in both cases.

3.3. *Estimation of q.* Consider the MA(q) process (1.1) with unknown q and unknown coefficient matrices. Suppose $\{X_t : 1 \leq t \leq n\}$ is a sample from the process (1.1). We now discuss graphical estimation of q.

Applying Corollary 2.1(a)(ii), it is immediate that, under (B1), the LSD of $0.5(\hat{\Gamma}_u + \hat{\Gamma}^*_u)$ and $0.5(\Gamma_u + \Gamma^*_u)$ are identical for all $u \geq 0$. Moreover, this LSD is degenerate at 0 when $u > q$. Hence, we propose to plot the histogram of the ESD of $0.5(\hat{\Gamma}_u + \hat{\Gamma}^*_u)$ for $u = 0, 1, 2, \ldots$. We say that $\hat{q}$ is an estimate of q, if the ESD of these matrices appear degenerate at 0 for order $u > \hat{q}$. However, as pointed out by one of the referees, often it may not be easy to decide if the ESD is close to being degenerate.

There is an alternative approach which uses a significance test based on the trace. By Corollary 2.1(a)(i)), $0.5\sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}^*_u - \Gamma_u - \Gamma^*_u)$ have non-degenerate LSD which are described in Corollary S3.1. However, $\{\Gamma_u\}$ are unknown so this cannot be used. A way out is to use appropriate consistent estimator of Γ_u. Fortunately, such an estimate is available in Bhattacharjee and Bose (2014) for certain combinations of n and p.

Let us first define the appropriate parameter spaces for $\{\psi_j\}$ where the above mentioned consistency can be achieved. To impose restrictions on the parameter space, define for any nested sequence of matrices $\{M_p = ((m_{ij}))_{p\times p}\}$,

$$\|M\|_{(1,1)} = \max_{j\geq 1} \sum_{i\geq 1} |m_{ij}| \quad \text{and} \quad T(M,t) = \max_{j\geq 1} \sum_{i:|i-j|>t} |m_{ij}|.$$

Let $\max(\|\psi_j\|_{(1,1)}, \|\psi_j'\|_{(1,1)}) = r_j,\ j \geq 0$. We define the following class of $\{\psi_j\}_{j=0}^{\infty}$ for some $0 < \beta < 1$ and $\lambda \geq 0$:

$$\Im(\beta,\lambda) = \left\{\{\psi_j\}_{j=0}^{\infty} : \sum_{j=0}^{\infty} r_j^{\beta} < \infty, \sum_{j=0}^{\infty} r_j^{2(1-\beta)} j^{\lambda} < \infty\right\}$$

which ensures that the dependence among X_t and $X_{t+\tau}$ decreases with the lag τ. Note that the summability implies that the decay rate of r_j cannot be slower than a polynomial rate. In case of a finite order moving average process, as we have a finite number of norm bounded parameter matrices, $\{\psi_j\}$ will automatically belong to $\Im(\beta,\lambda)$ for all $0 < \beta < 1$ and $\lambda \geq 0$.

For any $1 \leq i \leq p$, let $X_{t,i.p}$ be the ith component of the vector $X_{t.p}$. Here, we ensure that for any $t_1 < t$ and $k > 0$, the dependence between $X_{t_1,(i\pm k).p}$ and $X_{t,i.p}$ grows weaker as the lag k increases. We achieve this by putting restrictions over $\{T(\psi_j,t) : j = 0,1,2,\dots\}$ for all $t > 0$. Consider the following class for some $C,\alpha,\nu > 0$ and $0 < \eta < 1$ as

$$\mathcal{G}(C,\alpha,\eta,\nu) = \left\{\{\psi_j\} : T\left(\psi_j, t\sum_{u=0}^{j} \eta^u\right) < Ct^{-\alpha} r_j j^{\nu} \sum_{u=0}^{j} \eta^{-u\alpha}, \text{ and}\right.$$

$$\left. \sum_{j=k}^{\infty} \frac{r_j r_{j-k} j^{\nu}}{\eta^{\alpha j}} < \infty \right\}. \tag{3.6}$$

Consider the following assumptions:

(C1) $\{\psi_j\} \in \Im(\beta,\lambda) \cap \mathcal{G}(C,\alpha,\eta,\nu)$ for some $0 < \beta, \eta < 1$ and $C,\lambda,\alpha,\nu > 0$.

Let $\varepsilon_{t,j.p}$ be the jth component of $\varepsilon_{t.p}$.

(C2) For some $\lambda_0 > 0$,

$$\sup_{j\geq 1} E\left(e^{\lambda \varepsilon_{t,j.p}}\right) < \infty \quad \text{for all } |\lambda| < \lambda_0. \tag{3.7}$$

For any matrix M of order p and $k > 0$, the *banded* version of M is

$$B_k(M) = \left(\left(m_{ij} I\left(|i-j| \leq k\right)\right)\right).$$

By Theorem 4.1 of Bhattacharjee and Bose (2014), if (C1) and (C2) hold then for $k_n = (n^{-1}\log p)^{-\frac{1}{2(\alpha+1)}}$, $\|B_{k_n}(\hat{\Gamma}_u) - \Gamma_u\|_2 = O_p(k_n^{-\alpha})$ for all u.

Let $\hat{\Pi}_{a,u,B} = \sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - B_{k_n}(\hat{\Gamma}_u) - B_{k_n}(\hat{\Gamma}_u^*))$. We then have the following theorem.

THEOREM 3.1. *Consider the model* (1.1) *with* $0 \le q < \infty$. *Suppose* (A1), (A2), (B1), (C1) *and* (C2) *hold, and* $n(p), p \to \infty$ *such that* $p/n \to 0$ *and* $np^{-(\alpha+1)}(\log p)^{\alpha} \to 0$. *Then for* $k_n = (n^{-1}\log p)^{-\frac{1}{2(\alpha+1)}}$, *the LSD* (*in probability*) *of* $\hat{\Pi}_{a,u,B}$ *and* $\hat{\Pi}_{a,u}$ *are identical. This conclusion also holds for* $q = \infty$ *if we further assume* (B2).

PROOF. Observe that by Corollary A.41 of Bai and Silverstein (2010) and Corollary 2.1 we have

$$\begin{aligned}
&\frac{1}{p}\operatorname{Tr}\big(\sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - B_{k_n}(\hat{\Gamma}_u) - B_{k_n}(\hat{\Gamma}_u^*)) - \sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - \hat{\Gamma}_u - \hat{\Gamma}_u^*)\big)^2 \\
&\quad = \frac{1}{p}\operatorname{Tr}\big(\hat{\Gamma}_u + \hat{\Gamma}_u^* - B_{k_n}(\hat{\Gamma}_u) - B_{k_n}(\hat{\Gamma}_u^*)\big)\big)^2 \\
&\quad \le 4np^{-1}\|B_{k_n}(\hat{\Gamma}_u) - \Gamma_u\|_2^2 = o_{\mathrm{P}}\bigg(\frac{n}{p}\bigg(\frac{\log p}{n}\bigg)^{\frac{\alpha}{(\alpha+1)}}\bigg) \to 0. \qquad \square
\end{aligned}$$

Now by Corollary 2.1(a)(iii) and Theorem 3.1, the LSD (in probability) of $\hat{\Pi}_{a,u,B}$ are identical for $u > q$ and are different for $u \le q$. Thus we can plot the CDF of the ESD of $\hat{\Pi}_{a,u,B}$ for the first few sample autocovariance matrices in the same graph. We say that $\hat{q}$ is an estimate of q, if the ESD of $\hat{\Pi}_{a,u,B}$ with order $u > \hat{q}$ empirically coincide with each other.

EXAMPLES (SIMULATIONS). 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 $A = 0.5I_p$, $B = 0.5(I_p + J_p)$, $C = ((I(1 \le i = j \le [p/2]) - I([p/2] < i = j \le p)))$, $D = ((I(i + j = p + 1)))$ and $E = (((|i - j| + 1)^{-1}I(i + j = p + 1)))$ be five $p \times p$ matrices. We consider the following models:

Model 1 $X_t = \varepsilon_t$.
Model 2 $X_t = \varepsilon_t + A\varepsilon_{t-1}$.
Model 3 $X_t = \varepsilon_t + B\varepsilon_{t-1}$.
Model 4 $X_t = \varepsilon_t + C\varepsilon_{t-1} + D\varepsilon_{t-2}$.
Model 5 $X_t = \varepsilon_t + C\varepsilon_{t-1} + E\varepsilon_{t-2}$.

Suppose we have samples of size n respectively from Models 1–5. Now suppose that we do not know the models and wish to identify the unknown order of the MA processes. This will be done in two ways: first, by using the fact that the LSD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ is degenerate when $u > q$; second by using the fact that the nondegenerate LSD of $\hat{\Pi}_{a,u,B}$ are identical for all $u > q$.

Let $p = 500$, $n = p^{1.5}$. For each of the five models, we plot the CDF of the above two ESD for $1 \le u \le 3$ and $1 \le u \le 4$, respectively; see Figures 1 and 2.

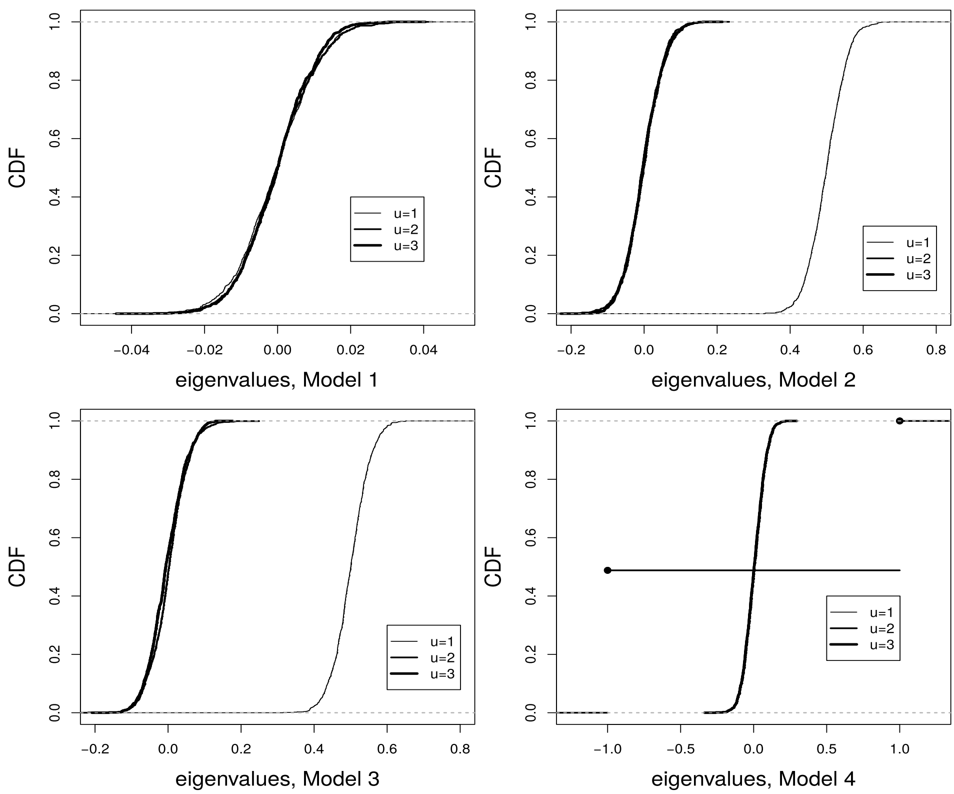

FIG. 1. *CDF of ESD of* $0.5(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$, $1 \leq u \leq 3$ *with* $p = 500$ *and* $n = p^{1.5}$.

From the figures, we see that the ESD of $0.5(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ are almost degenerate at 0 for $u \geq 1$ in Model 1, for $u \geq 2$ in Models 2, 3 and for $u \geq 3$ in Models 4, 5. In Figure 1, we have not displayed the ESD of Model 5 as the LSD of $0.5(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ for Models 4 and 5 are identical.

Similarly, the ESD of $\hat{\Pi}_{a,u,B}$ are observed to be identical in the above cases except for Model 4. The matrix D in Model 4 does not belong to the class $\mathcal{G}$ defined in (3.6) and that explains why the corresponding ESDs are not performing well in the simulation.

From the plots, the estimated value of q is 0, 1, 1, 2 and 2, respectively, for Models 1−5. This shows that our method is performing very well.

Incidentally, it follows by simple algebra and application of Corollary 2.1(a)(i) that the LSDs in Model 4 are $2\text{Ber}(0.5) - 1$ for $u = 1, 2$. This is supported by the graph in the bottom right panel of Figure 1.

3.4. *Hypothesis testing for coefficient matrices*. Since an MA(q) process (1.1) is characterized by its coefficient matrices $\{\psi_j\}$, it is relevant to make statistical inference on them. To the best of our knowledge, there are no results in the literature

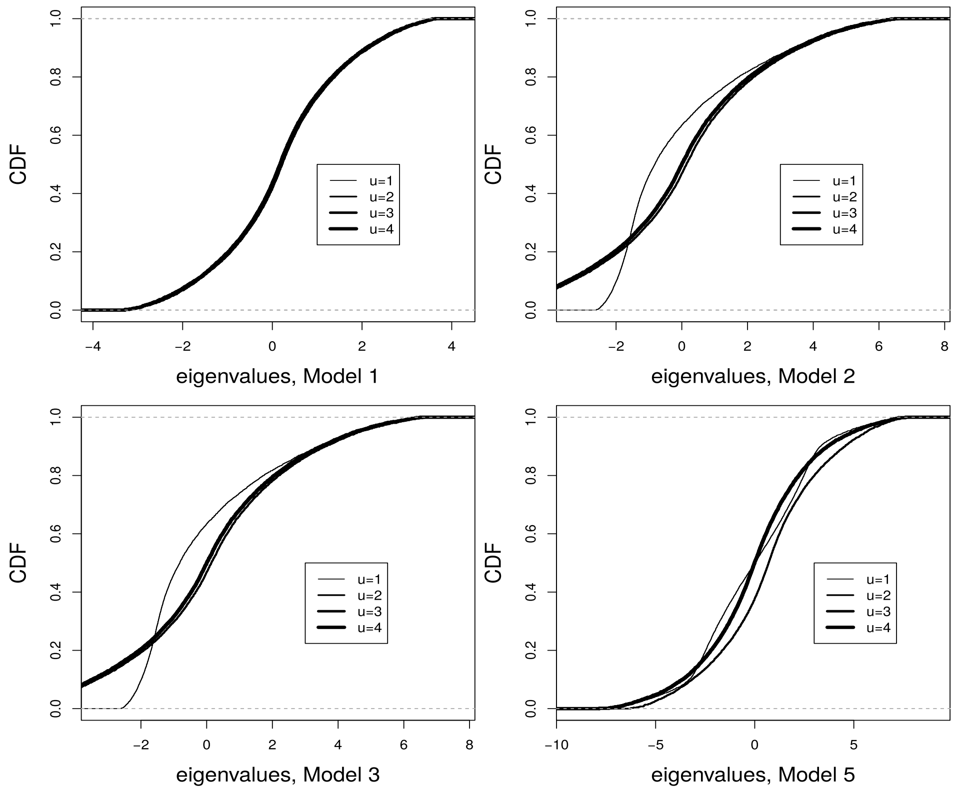

FIG. 2. *CDF of ESD of* $0.5\sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - B_{k_n}(\hat{\Gamma}_u) - B_{k_n}(\hat{\Gamma}_u^*)$, $1 \le u \le 4$ *with* $p = 500$, $n = p^{1.5}$ *and* $k_n = (n^{-1}\log p)^{-\frac{1}{4}}$.

which provide consistent estimators of $\{\psi_j\}$ or deals with testing of hypotheses for $\{\psi_j\}$. In this section, we shall discuss testing of $\{\psi_j\}$ for some simple null and alternative hypotheses.

The entries of ψ_j provide spatial correlations (correlation between components) at time lag j. Hence to understand these correlations, one may be interested to test different hypothesis regarding ψ_j. For instance, $H_0 : \psi_q = 0$ or $H_0 : \psi_j = 0\ \forall j$—these are equivalent to testing for the unknown order of the process and have already been discussed in Sections 3.1–3.3. We may also be interested to test $H_0 : \psi_q = A_q$ and $H_0 : \psi_j = A_j\ \forall j$ for some known nonnull $\{A_j\}$. Moreover it is known that, as $p \to \infty$, the sample autocovariance matrices $\{\hat{\Gamma}_u\}$ are not consistent for $\{\Gamma_u\}$ unless we have appropriate restrictions on the parameter space of coefficient matrices $\{\psi_j\}$; for example, see Bhattacharjee and Bose (2014) and Basu and Michailidis (2015). Therefore, an important problem is to check whether $\{\psi_j\}$ satisfies these restrictions. This also leads us to test for hypotheses on $\{\psi_j\}$.

In this section, we demonstrate tests using the LSD results and the asymptotic normality of traces for some specific hypotheses.

In Examples 1 and 3, we provide tests of significance for $H_0 : \psi_q = A_q$ and $H_0 : \psi_j = A_j\ \forall j$, respectively, for some known $\{A_j\}$. Specific choices of $\{A_j\}$ are also discussed in Examples 4–7. Unfortunately, we do not know how to test hypothesis on ψ_j, $1 \le j \le q-1$; for details, see Example 2. Examples 8–11 deal with some interesting testing for IVAR and IVARMA processes. We also discuss two-sample hypothesis testing in Section 3.4.2. Much more work needs to be done to generalize our prescription to deal with more general null and alternative hypotheses.

In our formulation nonrandom matrices with the same ESD or LSD are treated to be asymptotically equal. Hence by $A = B$ for two nonrandom matrices, we mean the eigenvalue distributions of A and B are identical. Throughout this section, we assume that all the deterministic (nonrandom) matrices converge jointly.

We provide both graphical and significance tests. Graphical tests are based on the LSD of nonscaled but centered sample autocovariance matrices. These are easy to derive using Theorem 2.2 and Corollary 2.1. Significance tests are based on the asymptotic normality of appropriately centered and scaled trace of sample autocovariance matrices. These can be derived using Theorem 2.4 and Corollary 2.4.

3.4.1. *One sample case.* Consider the MA(q) process given in (1.1).

EXAMPLE 1. Let A be a square matrix of order p with a nondegenerate LSD. Suppose we wish to test

$$H_0 : \psi_q = A \quad \text{against} \quad H_1 : \psi_q \neq A. \tag{3.8}$$

As $\Gamma_q = \psi_q^*$, this is equivalent to testing

$$H_0 : \Gamma_q = A^* \quad \text{against} \quad H_1 : \Gamma_q \neq A^*. \tag{3.9}$$

Using Corollary 2.1(a)(ii), it is easy to establish that, under H_0 the LSD of $(\hat{\Gamma}_q + \hat{\Gamma}_q^* - A - A^*)$ is degenerate at 0. While under H_1 this LSD is identical with the LSD of $(\Gamma_q + \Gamma_q^* - A - A^*)$ which is nondegenerate. Thus to test (3.8) graphically, we plot the eigenvalue distribution of $(\hat{\Gamma}_q + \hat{\Gamma}_q^* - A - A^*)$. If it appears degenerate at 0, then H_0 is accepted; else we reject H_0.

One of the referees raised the issue that it may be difficult to decide whether the ESD is close to being degenerate. Note that the LSD are all almost sure results. It is known from the experience of random matrix theory that the convergences to the LSD for many matrix models are very fast. Even though no convergence rate results are known in the present cases at the moment, we believe the convergence in our theorems are also quite fast.

A significant test can be derived using the asymptotic normality of $\mathrm{Tr}(\hat{\Gamma}_q)$. In Corollary S11.3, using Theorem 2.4 and some additional arguments, we prove that under H_0

$$\sqrt{np^{-1}}(\mathrm{Tr}(\hat{\Gamma}_q) - \mathrm{Tr}(A))/\hat{\lambda}_q \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \tag{3.10}$$

where $\hat{\lambda}_q^2$ is a consistent estimator of the asymptotic variance of $\sqrt{np^{-1}}(\text{Tr}\,(\hat{\Gamma}_q) - \text{Tr}(A))$ and is a function of $\{\lim p^{-1}\,\text{Tr}(\hat{\Gamma}_u\hat{\Gamma}_v) : -q < u, v < q\}$ only. Explicit expression for $\hat{\lambda}_q^2$ is given in S(11.21).

In Corollary S11.3, we also establish that under H_1, for some $0 < \lambda < \infty$,

$$\sqrt{np^{-1}}(\text{Tr}(\hat{\Gamma}_q) - \text{Tr}(\Gamma_q))/\hat{\lambda}_q \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda^2). \tag{3.11}$$

Therefore, we can use the statistic on the left side of (3.10) and reject H_0 if it is large.

One of the referees pointed out that the above test is essentially for $H_0 : \text{Tr}(\psi_q) = \text{Tr}(A)$ against $H_1 : \text{Tr}(\psi_q) \neq \text{Tr}(A)$. While that is true, note that in general, we can also test for $H_0 : \text{Tr}(\psi_q^k) = \text{Tr}(A^k)$ for any $k \geq 1$ against $H_1 : \text{Tr}(\psi_q^k) \neq \text{Tr}(A^k)$ using asymptotic normality of $\sqrt{np^{-1}}(\text{Tr}(\hat{\Gamma}_q^k) - \text{Tr}(\Gamma_q^k))$. Further investigation is needed to propose a combined test for a set of values of k. This would alleviate the problem partially.

EXAMPLE 2. Consider the model (1.1). In general, we do not have any ready-made answers for testing $H_0 : \psi_j = A \text{against} H_1 : \psi_j \neq A$ for some fixed $0 \leq j \leq (q-1)$ and $p \times p$ matrix A. To explain why, consider the MA(3) process with coefficient matrices $\{\psi_0 = I, \psi_1, \psi_2, \psi_3\}$. In this case,

$$\begin{aligned} &\Gamma_1 = \psi_1^* + \psi_1\psi_2^* + \psi_2\psi_3^*, \qquad \Gamma_2 = \psi_2^* + \psi_1\psi_3^*, \\ &\Gamma_3 = \psi_3^*, \qquad \Gamma_u = 0\ \forall u \geq 4. \end{aligned} \tag{3.12}$$

Suppose we wish to test

$$H_0 : \psi_2 = A \quad \text{against} \quad H_1 : \psi_2 \neq A. \tag{3.13}$$

Γ_1, Γ_2 and Γ_3 under H_0 are given by

$$\Gamma_{1H_0} = \psi_1^* + \psi_1 A^* + A\psi_3^*, \qquad \Gamma_{2H_0} = A^* + \psi_1\psi_3^*, \qquad \Gamma_{3H_0} = \psi_3^*. \tag{3.14}$$

If ψ_1 and ψ_3 are known, then by Corollary 2.1(a)(ii), it is easy to see that under H_0, the LSD of $(\hat{\Gamma}_1 + \hat{\Gamma}_1^* - \Gamma_{1H_0} - \Gamma_{1H_0}^*)$ is degenerate at 0, whereas under H_1, this LSD is identical with the LSD of $(\Gamma_1 + \Gamma_1^* - \Gamma_{1H_0} - \Gamma_{1H_0}^*)$. Also, by Theorem 2.4, for some $0 < a, b < \infty$,

$$\text{under } H_0, \qquad \sqrt{np^{-1}}(\text{Tr}(\hat{\Gamma}_1) - \text{Tr}(\Gamma_{1H_0})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, a^2) \quad \text{and}$$

$$\text{under } H_1, \qquad \sqrt{np^{-1}}(\text{Tr}(\hat{\Gamma}_1) - \text{Tr}(\Gamma_1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, b^2).$$

Here, a and b are functions of $\{\lim p^{-1}\text{Tr}(\Gamma_{uH_0}\Gamma_{uH_0}) : -3 \leq u, v \leq 3\}$ and $\{\lim p^{-1}\text{Tr}(\Gamma_u\Gamma_v) : -3 < u, v < 3\}$, respectively. Just as in Example 1, we can make use of these results to test (3.13). One can use autocovariance of order 2 also. $\hat{\Gamma}_3$ cannot be used to test (3.13) as Γ_3 is not a function of ψ_2, and hence it makes no distinction between H_0 and H_1.

Clearly, the above method does not work when ψ_1 and ψ_3 are unknown. Moreover, it is hard to estimate these coefficient matrices. If we consider the method of moments, we get the system of equations given in (3.12) after replacing the population autocovariance matrices by the sample autocovariance matrices. These equations are not easy to solve. Moreover, appropriate regularization could be needed before consistency is achieved. This needs further investigation.

EXAMPLE 3. Next, consider the hypotheses

$$H_0 : \psi_j = A_j \quad \forall j \quad \text{against} \quad H_1 : \psi_j \neq A_j \quad \text{for at least one } j, \tag{3.15}$$

for some known $p \times p$ matrices $\{A_j\}$. As H_0 specifies all the coefficient matrices $\{\psi_j : 0 \le j \le q\}$, a natural testing method should be based on $\hat{\Gamma}_u$ for all $1 \le u \le q$. Let

$$\hat{G}_q = \sum_{u=1}^{q} \hat{\Gamma}_u, \qquad G_q = \sum_{u=1}^{q} \Gamma_u = \sum_{u=0}^{q} \sum_{j=0}^{q-u} \psi_j \psi_{j+u}^* \quad \text{and} \tag{3.16}$$

$$G_{qH_0} = \sum_{u=0}^{q} \sum_{j=0}^{q-u} A_j A_{j+u}^* \quad (\text{i.e., } G_q \text{ under } H_0). \tag{3.17}$$

By Theorem 2.2 and using truncation arguments, it can easily be proved that under H_0, the LSD of $(\hat{G}_q + \hat{G}_q^* - G_{qH_0} - G_{qH_0}^*)$ is degenerate at 0 whereas under H_1, this LSD is identical with the LSD of $(G_q + G_q^* - G_{qH_0} - G_{qH_0}^*)$. Thus to test (3.15) graphically, we plot the eigenvalue distribution of $(\hat{G}_q + \hat{G}_q^* - G_{qH_0} - G_{qH_0}^*)$. We accept H_0 if this ESD appears degenerate at 0.

A test statistic can be proposed using $\operatorname{Tr}(\hat{G}_q)$. In Corollary S11.4, we show that under H_0,

$$\sqrt{np^{-1}}\bigl(\operatorname{Tr}(\hat{G}_q) - \operatorname{Tr}(G_{qH_0})\bigr)/\hat{\delta}_q \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \tag{3.18}$$

where $\hat{\delta}_q^2$ is a consistent estimator of the asymptotic variance of $\sqrt{np^{-1}}(\operatorname{Tr}(\hat{G}_q) - \operatorname{Tr}(G_{qH_0}))$ and is a function of $\{\lim p^{-1} \operatorname{Tr}(\hat{\Gamma}_u \hat{\Gamma}_v) : -q < u, v < q\}$ only. An explicit expression for $\hat{\delta}_q^2$ is given in S(11.23).

In Corollary S11.4, we also establish that under H_1, for some $0 < \delta < \infty$,

$$\sqrt{np^{-1}}\bigl(\operatorname{Tr}(\hat{G}_q) - \operatorname{Tr}(G_q)\bigr)/\hat{\delta}_q \xrightarrow{\mathcal{D}} \mathcal{N}(0, \delta^2). \tag{3.19}$$

Therefore, we can use $T_3 = \sqrt{np^{-1}}(\operatorname{Tr}(\hat{G}_q) - \operatorname{Tr}(G_{qH_0}))/\hat{\delta}_q$ as a test statistic and reject H_0 if it is large in absolute value.

EXAMPLE 4. Consider the MA(q) model (1.1) with $\psi_j = \alpha_j I_p$ for all $0 \le j \le q$, considered by Pfaffel and Schlemm (2011). They established the LSD of

$(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ when $p/n \to y \in (0, \infty)$. Consider two sequences of real numbers $\{a_{0j}\}$ and $\{a_{1j}\}$. Suppose we wish to test

$$H_0 : \alpha_j = a_{0j} \quad \forall 0 \le j \le q \quad \text{against} \quad H_1 : \alpha_j = a_{1j} \quad \forall 0 \le j \le q.$$

Define $\gamma_{iu} := \sum_{j=0}^{q-u} a_{ij} a_{i(j+u)}$ for all $i = 0, 1$ and $u \ge 0$. We assume $\sum_{u=0}^{q}(\gamma_{0u} - \gamma_{1u}) \neq 0$ so that $\sum_{u=0}^{q} \operatorname{Tr}(\Gamma_u)$ under H_0 and H_1 are different. Further suppose $p/n \to 0$. By Theorem S11.1, under H_i, we have

$$T_i := \sqrt{np^{-1}}\left(\operatorname{Tr}\left(\sum_{u=0}^{q} \hat{\Gamma}_u\right) - \left(\sum_{u=0}^{q} \gamma_{iu}\right)p\right)\Big/\bar{\sigma}_i \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \qquad \text{where}$$

$$\bar{\sigma}_i^2 := 0.5 \sum_{v=-\infty}^{\infty}\left(\sum_{u=-q}^{q} \gamma_{i(v+u)}\right)^2.$$

We can use T_0 as test statistic and can reject H_0 when $|T_0|$ is large.

EXAMPLE 5. Consider the MA(q) process (1.1) with $\psi_j = \alpha_j A$. Note that this is a more general model than in Example 4. Suppose $\{\alpha_j\}$ is known and A is unknown. We wish to test

$$H_0 : A = I_p \quad \text{against} \quad H_1 : A = \operatorname{Diag}(a_1, a_2, \ldots, a_p), \tag{3.20}$$

where $\sum_{j=1}^{p} a_j^2 \neq p$ and $a^2 := \lim p^{-1} \sum_{j=1}^{p} a_j^4$ exists. This can be done by testing either of the following $(q+2)$ hypotheses:

$$\begin{aligned}
H_{0u} &: \Gamma_u = \gamma_u I_p \quad \text{against} \\
H_{1u} &: \Gamma_u = \gamma_u \operatorname{Diag}(a_1^2, a_2^2, \ldots, a_p^2), \qquad 0 \le u \le q \\
H_{u(q+1)} &: \psi_j = \alpha_j I_p \qquad \forall j \quad \text{against} \\
H_{1(q+1)} &: \psi_j = \alpha_j \operatorname{Diag}(a_1, a_2, \ldots, a_p) \qquad \forall j
\end{aligned}$$

where $\gamma_u = \sum_{j=0}^{\infty} \alpha_j \alpha_{j+u}$ and $\alpha_j = 0 \; \forall j > q$. Define

$$\sigma_u^2 = 0.5 \sum_{v=-\infty}^{\infty} (\gamma_{v+u} + \gamma_{v-u})^2, \qquad 0 \le u \le q \quad \text{and}$$

$$\sigma_{q+1}^2 = 0.5 \sum_{v=-\infty}^{\infty}\left(\sum_{u=-q}^{q} \gamma_{v+u}\right)^2.$$

By Theorem S11.1, it is easy to see that under H_0,

$$T_{uH_0} = \sqrt{np^{-1}}(\operatorname{Tr}(\hat{\Gamma}_u) - \gamma_u p)/\sigma_u \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \qquad 0 \le u \le q,$$

$$T_{(q+1)H_0} = \sqrt{np^{-1}}\left[\operatorname{Tr}\left(\sum_{u=0}^{q} \hat{\Gamma}_u\right) - \left(\sum_{u=0}^{q} \gamma_u\right)p\right]\Big/\sigma_{q+1} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

and under H_1,

$$T_{uH_1} = \sqrt{np^{-1}}\left(\mathrm{Tr}(\hat{\Gamma}_u) - \gamma_u\left(\sum_{j=1}^{p} a_j^2\right)\right)\Big/ a\sigma_u \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \qquad 0 \le u \le q,$$

$$T_{(q+1)H_1} = \sqrt{np^{-1}}\left[\mathrm{Tr}\left(\sum_{u=0}^{q} \hat{\Gamma}_u\right) - \left(\sum_{u=0}^{q} \gamma_u\right)\left(\sum_{j=1}^{p} a_j^2\right)\right]\Big/ a\sigma_{q+1} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Therefore, we can use any T_{uH_0}, $0 \le u \le q+1$, as a test statistic and can reject H_0 for their larger absolute values.

All of the above tests are consistent. It is not possible to compare the power of these tests in general but we can do it in certain specific cases. Note that for large p and $0 \le u \le q$, the power of T_{uH_0} (at level α) approximately equals

$$1 - \Phi\left(z_{\alpha/2} a^{-1} + \sqrt{np^{-1}}\gamma_u\left(\sum_{j=0}^{q} a_j^2 - p\right)(a\sigma_u)^{-1}\right)$$
$$+ \Phi\left(-z_{\alpha/2} a^{-1} + \sqrt{np^{-1}}\gamma_u\left(\sum_{j=0}^{q} a_j^2 - p\right)(a\sigma_u)^{-1}\right),$$

where Φ and $z_{\alpha/2}$ are the distribution function and the upper $\alpha/2$ point of the standard normal variable. Moreover, if $\{\gamma_u\}$ is nonnegative and non-increasing for $u \ge 0$, then we have $\sigma_q^2 = \min\{\sigma_u^2 : 0 \le u \le q+1\}$. Thus T_{qH_0} provides more power than $\{T_{uH_0} : 0 \le u \le q-1, u = q+1\}$.

Simulations: Here, we simulate from the six MA(q) processes with $q = 1, 2$ and $\alpha_j = \theta^j I(0 \le j \le q)$ with $\theta = 0.5, 1, 2$. For all models, we take $a_j^4 = p/(q+1)\ \forall j$. Thus $\sum_{j=0}^{q} a_j^2 < p$ and $a^2 = 1$. Moreover, we consider $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}_p(0, I_p)$, $n = p^{1.2}$ and $p = 300, 500, 1000$. We compute $\{T_{uH_0} : 0 \le u \le q+1\}$ and their empirical power (E.P.) with 300 replications for each case. These E.P.s are recorded in Table 3. It shows that all of the above tests are consistent and T_{qH_0} dominates others. This is because $\gamma_u = \theta^{|u|}(\sum_{j=0}^{q-|u|} \theta^{2j}) I(|u| \le q)$ is nonnegative and is a nonincreasing function of $u \ge 0$.

EXAMPLE 6. Consider the same model as in Example 5 and hypotheses (3.20). Moreover, assume $\sum_{j=1}^{p} a_j^2 = p$, $\sum_{j=1}^{p} a_j^4 \ne p$ and $\tilde{a}^2 := \lim p^{-1} \sum_{j=1}^{p} a_j^8$ exists. The test statistics given in Example 5 will not work since $\sum_{j=1}^{p} a_j^2 = p$

TABLE 3
Comparison of E.P. of $\{T_{uH_0} : 0 \le u \le (q+1)\}$ in the context of Example 5

$q=1$		T_{0H_0}	T_{1H_0}	T_{2H_0}
$\theta = 0.5$	$p = 300$	0.82	0.85	0.78
$\theta = 0.5$	$p = 500$	0.85	0.89	0.83
$\theta = 0.5$	$p = 1000$	0.94	0.97	0.92
$\theta = 1$	$p = 300$	0.81	0.82	0.8
$\theta = 1$	$p = 500$	0.89	0.9	0.87
$\theta = 1$	$p = 1000$	0.94	0.95	0.93
$\theta = 2$	$p = 300$	0.81	0.85	0.75
$\theta = 2$	$p = 500$	0.9	0.93	0.85
$\theta = 2$	$p = 1000$	0.95	0.98	0.93

$q=2$	•	T_{0H_0}	T_{1H_0}	T_{2H_0}	T_{3H_0}
$\theta = 0.5$	$p = 300$	0.82	0.88	0.94	0.75
$\theta = 0.5$	$p = 500$	0.89	0.92	0.97	0.82
$\theta = 0.5$	$p = 1000$	0.92	0.94	0.98	0.89
$\theta = 1$	$p = 300$	0.85	0.89	0.92	0.78
$\theta = 1$	$p = 500$	0.91	0.93	0.95	0.83
$\theta = 1$	$p = 1000$	0.93	0.94	0.98	0.89
$\theta = 2$	$p = 300$	0.81	0.88	0.94	0.75
$\theta = 2$	$p = 500$	0.9	0.93	0.97	0.86
$\theta = 2$	$p = 1000$	0.92	0.95	0.98	0.9

implies $\mathrm{Tr}(\Gamma_u)$ under H_0 and H_1 are equal. On the other hand, by Corollary S11.5,

$$\text{under } H_0, \qquad \tilde{T}_{qH_0} = \sqrt{np^{-1}}\big(\mathrm{Tr}(\hat{\Gamma}_q\hat{\Gamma}_q^*) - \gamma_q^2 p\big)/\tilde{\sigma}_q \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

$$\text{under } H_1, \qquad \tilde{T}_{qH_1} = \sqrt{np^{-1}}\left(\mathrm{Tr}(\hat{\Gamma}_q\hat{\Gamma}_q^*) - \gamma_q^2\left(\sum_{j=1}^{p} a_j^4\right)^2\right)/\tilde{a}\tilde{\sigma}_q \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

$$\text{where } \tilde{\sigma}_q^2 = 2\left[\sum_{v=-\infty}^{\infty} \gamma_q^2(\gamma_{v+q} + \gamma_{v-q})^2\right].$$

Thus T_{qH_0} can be used to test (3.20) and we can reject H_0 for its large absolute values. As in Example 5, one can also use test statistics based on $\hat{\Gamma}_u\hat{\Gamma}_u^*$ for $0 \le u \le q-1$. We omit the details.

EXAMPLE 7. A matrix $M = ((m_{ij}))$ is said to be K*-banded* if $m_{ij} = 0$ for $|i-j| > K$. It is known that $\{\Gamma_u\}$ are consistently estimable if $\{\psi_j\}$ (or A) are K-banded; for example, see Bhattacharjee and Bose (2014). Consider the MA(q) process (1.1) with $\psi_j = \alpha_j A$ and $A = ((t_{|i-j|} I(|i-j| \le K)))_{p\times p}$ for some integer

$K \geq 0$ and known sequence of real numbers $\{t_j\}$. We want to test

$$H_0 : K = K_0 \quad \text{against} \quad H_1 : K = K_1 > K_0.$$

This cannot be tested using $\mathrm{Tr}(\hat{\Gamma}_u)$ as it is identical under H_0 and H_1. Instead we use $\mathrm{Tr}(\hat{\Gamma}_u \hat{\Gamma}_u^*)$. Let $T_K = \mathrm{Tr}(((t_{|j-l|} I(|j-l| > K_i)))^4)$. Suppose $\tilde{a}_{H_i}^2 := \lim p^{-1} \mathrm{Tr}(((t_{|j-l|} I(|j-l| > K_i)))^8)$, $i = 0, 1$, exist. Then by Corollary S11.5, under H_i,

$$\bar{T}_{uH_i} = \sqrt{np^{-1}} \big(\mathrm{Tr}\big(\hat{\Gamma}_u \hat{\Gamma}_u^*\big) - \gamma_u^2 T_{K_i}\big) / \tilde{\sigma}_u \tilde{a}_{H_i} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$
$$\text{where } \tilde{\sigma}_u^2 = 2\Bigg[\sum_{v=-\infty}^{\infty} \gamma_u^2 (\gamma_{v+u} + \gamma_{v-u})^2\Bigg], 0 \leq u \leq q.$$

Then we can use any of $\bar{T}_{uH_0}$, $0 \leq u \leq q$, as a test statistic and reject H_0 when its value is large. One can also use appropriately centered and scaled $\sum_{u=0}^{q} \mathrm{Tr}(\hat{\Gamma}_u \hat{\Gamma}_u^*)$ as a test statistic. As in Example 5, here also $\bar{T}_{qH_0}$ will have more power if $\{\gamma_u\}$ are nonnegative and nonincreasing.

Testing for IVAR processes.

EXAMPLE 8. Consider the following IVAR(1) process:

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

where A is symmetric, $\|A\|_2 < 1$, the LSD of A exists and $\{\varepsilon_t\}$ satisfies (A1) and (A3). It is easy to see that $\Gamma_u = (I - A^2)^{-1} A^u$. Suppose $p/n \to 0$. We wish to test

$$H_0 : A = 0 \quad \text{against} \quad H_1 : A \neq 0. \tag{3.21}$$

Note that by Theorem S11.1, under H_0,

$$\sqrt{np^{-1}} (\mathrm{Tr}(\hat{\Gamma}_0) - p) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2), \qquad \sqrt{np^{-1}} (\mathrm{Tr}(\hat{\Gamma}_u)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \qquad \forall u \geq 1$$

and under H_1,

$$\sqrt{np^{-1}} \big(\mathrm{Tr}(\hat{\Gamma}_u) - \mathrm{Tr}\big(\big(I - A^2\big)^{-1} A^u\big)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma_u^2\big) \qquad \text{where}$$
$$\sigma_u^2 = 0.5 \sum_{v=-\infty}^{\infty} \lim \frac{1}{p} \mathrm{Tr}\big(\big(I - A^2\big)^{-2} \big(A^{|v-u|} + A^{|v+u|}\big)^2\big).$$

Therefore, we use either $\sqrt{np^{-1}} (\mathrm{Tr}(\hat{\Gamma}_0) - p)$ or $\sqrt{np^{-1}} (\mathrm{Tr}(\hat{\Gamma}_u))$ for any $u \geq 1$ as the test statistic and reject H_0 for their larger absolute value. As discussed in the previous examples, power of the tests increase with u.

Now consider higher order IVAR processes. Recall that tests of significance in higher order univariate AR processes are based on partial autocovariances. Similarly, possible tests of significance in IVAR processes would be based on partial autocovariance matrices. Since these matrices are functions of autocovariance matrices and their inverses, potentially their joint limits can be derived. However, since inverses are involved, this is technically not an easy task—the moment method is not directly applicable, identification of the LSD is nontrivial and results on the trace are not obvious. These need further research. Here, we concentrate on two easy scenarios where inverses do not come into the picture and we can fall back on the results already derived for autocovariances; see examples (Examples 9 and 11).

Consider the following polynomials in $\{a_i, a_i^* : i \geq 0\}$:

$$\Pi_u(\{a_i, a_i^* : i \geq 0\}) = \det \begin{pmatrix} a_0 & a_1^* & \cdots & a_{u-2}^* & a_1 \\ a_1 & a_0 & \cdots & a_{u-3}^* & a_2 \\ a_2 & a_1 & \cdots & a_{u-4}^* & a_3 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{u-1} & a_{u-2} & \cdots & a_1 & a_u \end{pmatrix}. \tag{3.22}$$

Let $\Pi_{\text{Partial},u} = \Pi_u(\{\Gamma_i, \Gamma_i^* : i \geq 0\})$. It is not hard to see that $\Pi_{\text{Partial},u}$ is *proportional* to the uth order population partial autocovariance matrix of model (1.1) with commutative and symmetric $\{\psi_j : j \geq 0\}$; for details, see Brockwell and Davis (2006). Its sample counterpart, the uth order sample partial autocovariance matrix, is denoted by $\hat{\Pi}_{\text{Partial},u}$. As $\hat{\Pi}_{\text{Partial},u}$ is a polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$, its asymptotic normality follows from Theorem 2.4.

EXAMPLE 9. Consider the following IVAR(2) process:

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + \varepsilon_t \tag{3.23}$$

where A_1 and A_2 are symmetric, commutative, norm bounded and converge jointly. Moreover, we assume that $\{\varepsilon_t\}$ satisfies (A1) and (A3). Suppose A_1 is known and $p/n \to 0$. We wish to test

$$H_0 : A_2 = 0 \quad \text{against} \quad H_1 : A_2 \neq 0. \tag{3.24}$$

Note that $\Pi_{\text{Partial},2} = \Gamma_2\Gamma_0 - \Gamma_1^2$ for the model (3.23). Therefore, it equals 0 under H_0 but not under H_1. By Theorem 2.4, there are σ_{H_0} (depends only on A_1) and σ_{H_1} such that under H_0,

$$\sqrt{np^{-1}}\operatorname{Tr}(\hat{\Pi}_{\text{Partial},2}\hat{\Pi}^*_{\text{Partial},2})/\sigma_{H_0} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \tag{3.25}$$

and under H_1

$$\sqrt{np^{-1}}\big(\operatorname{Tr}(\hat{\Pi}_{\text{Partial},2}\hat{\Pi}^*_{\text{Partial},2}) - \operatorname{Tr}(\Pi_{\text{Partial},2}\Pi^*_{\text{Partial},2})\big)/\sigma_{H_1}$$
$$\xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \tag{3.26}$$

Hence we can use $\sqrt{np^{-1}}\text{Tr}(\hat{\Pi}_{\text{Partial},2}\hat{\Pi}^*_{\text{Partial},2})/\sigma_{H_0}$ as our test statistic and reject H_0 when it is large in magnitude.

Testing for IVARMA processes.

EXAMPLE 10. Consider the following stationary IVARMA(r, q) process:

$$X_t = \sum_{j=1}^{r} A_j X_{t-j} + \sum_{j=0}^{q} \psi_j \varepsilon_{t-j}, \tag{3.27}$$

where $\{A_i, \psi_j : 1 \le i \le r, 0 \le j \le q\}$ converge jointly and are norm bounded. $\{\varepsilon_t\}$ satisfies (A1) and (A3). Suppose q and $\{\psi_j : 0 \le j \le q\}$ are known and $p/n \to 0$. We wish to test

$$H_0 : r = 0 \quad \text{against} \quad H_1 : r \neq 0. \tag{3.28}$$

This is equivalent to testing

$$H_0 : A_j = 0 \qquad \forall j \quad \text{against} \quad H_1 : A_j \neq 0 \qquad \text{for at least one } j. \tag{3.29}$$

Note that $\Gamma_{q+1} = 0$ under H_0 but not under H_1. By Theorem S11.2, there are σ_{H_0} (depends only on q and $\{\psi_j : 0 \le j \le q\}$), σ_{H_1}, $\mu_{H_1} \neq 0$ such that, under H_0,

$$\sqrt{np^{-1}}\text{Tr}(\hat{\Gamma}_{q+1}\hat{\Gamma}^*_{q+1})/\sigma_{H_0} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

and under H_1,

$$\sqrt{np^{-1}}(\text{Tr}(\hat{\Gamma}_{q+1}\hat{\Gamma}^*_{q+1}) - \mu_{H_1})/\sigma_{H_1} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Therefore, we can use $\sqrt{np^{-1}}\text{Tr}(\hat{\Gamma}_{q+1}\hat{\Gamma}^*_{q+1})/\sigma_{H_0}$ as our test statistic and reject H_0 when it is large in magnitude.

EXAMPLE 11. Now consider the higher order IVARMA(r, q) processes in (3.27) where $\{A_i, \psi_j : 1 \le i \le r, 0 \le j \le q\}$ converge jointly and are norm bounded. Moreover, we assume that $\{\varepsilon_t\}$ satisfies (A1) and (A3). Suppose r and $\{A_j : 1 \le j \le r\}$ are known and $p/n \to 0$. Further suppose $\{A_i, \psi_j : 1 \le i \le r, 0 \le j \le q\}$ are commutative and symmetric. We wish to test

$$H_0 : q = 0 \quad \text{against} \quad H_1 : q \neq 0. \tag{3.30}$$

It is easy to see that $\Pi_{\text{partial},r+1} = 0$ under H_0 but not under H_1. By Theorem 2.4, there are σ_{H_0} (depends only on r and $\{A_j : 1 \le j \le r\}$), σ_{H_1}, $\mu_{H_1} \neq 0$ such that, under H_0,

$$\sqrt{np^{-1}}\mathrm{Tr}\big(\hat{\Pi}_{\mathrm{Partial},r+1}\hat{\Pi}^*_{\mathrm{Partial},r+1}\big)/\sigma_{H_0} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

and under H_1,

$$\sqrt{np^{-1}}\big(\mathrm{Tr}\big(\hat{\Pi}_{\mathrm{Partial},r+1}\hat{\Pi}^*_{\mathrm{Partial},r+1}\big) - \mu_{H_1}\big)/\sigma_{H_1} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Therefore, we can use $\sqrt{np^{-1}}\mathrm{Tr}(\hat{\Pi}_{\mathrm{Partial},r+1}\hat{\Pi}^*_{\mathrm{Partial},r+1})/\sigma_{H_0}$ as our test statistic and reject H_0 when it is large in magnitude.

3.4.2. *Two samples case.* Two sample version of all tests given in Examples 1–11 can be developed based on Theorems 2.3 and 2.4. Here, we illustrate the analogue of Examples 1 and 3 only. Consider the two MA(q) processes (1.1) and (2.5).

EXAMPLE 12. Suppose we wish to test

$$H_0: \psi_q = \phi_q \quad \text{against} \quad H_1: \psi_q \neq \phi_q. \tag{3.31}$$

Recall $\{\hat{\Pi}_{a,u,X}, \hat{\Pi}_{a,u,Y}\}$ from the discussion after Theorem 2.3. It is immediate from Corollary 2.3(a)(ii) that under H_0, the LSD of $(\hat{\Pi}_{a,q,X} - \hat{\Pi}_{a,q,Y})$ is degenerate at 0 and under H_1, this LSD is identical with the LSD of $0.5(\psi_q + \psi_q^* - \phi_q - \phi_q^*)$, which is nondegenerate. Thus a graphical method to test (3.31) is to plot the eigenvalue distribution of $(\hat{\Pi}_{a,q,X} - \hat{\Pi}_{a,q,Y})$. If it appears degenerate at 0, we accept H_0.

Similar to the one sample cases, we can devise test statistics. Recall $\hat{\lambda}_q$ from (3.10). Let $\hat{\lambda}_{q,X}$ and $\hat{\lambda}_{q,Y}$ be $\hat{\lambda}_q$ respectively for the processes $\{X_t\}$ and $\{Y_t\}$. Define $\hat{\lambda}^2_{q,XY} = \hat{\lambda}^2_{q,X} + \hat{\lambda}^2_{q,Y}$. By the independence of these processes and Corollary S11.3, it is easy to see that under H_0,

$$T_4 \equiv \sqrt{np^{-1}}\big(\mathrm{Tr}(\hat{\Gamma}_{q,X}) - \mathrm{Tr}(\hat{\Gamma}_{q,Y})\big)/\hat{\lambda}_{q,XY} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

and under H_1, for some $0 < \tilde{\lambda} < \infty$,

$$\sqrt{np^{-1}}\big(\big(\mathrm{Tr}(\hat{\Gamma}_{q,X}) - \mathrm{Tr}(\hat{\Gamma}_{q,Y})\big) - \big(\mathrm{Tr}(\psi_q) - \mathrm{Tr}(\phi_q)\big)\big)/\hat{\lambda}_{q,XY} \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \tilde{\lambda}^2\big).$$

Therefore, we can use T_4 as our test statistic and reject H_0 if it is large in absolute value.

EXAMPLE 13. Consider the hypotheses

$$H_0: \psi_j = \phi_j \quad \forall j \quad \text{against} \quad H_1: \psi_j \neq \phi_j \quad \text{for at least one } j. \tag{3.32}$$

Recall the definition of $\{G_q, \hat{G}_q\}$ in (3.16). $\{G_{q,X}, \hat{G}_{q,X}\}$ and $\{G_{q,Y}, \hat{G}_{q,Y}\}$ have obvious meaning. By Theorem 2.3 and using truncation arguments, under H_0 the LSD of $(\hat{G}_{q,X} + \hat{G}^*_{q,X} - \hat{G}_{q,Y} - \hat{G}^*_{q,Y})$ is degenerate at 0 whereas under H_1, this LSD is identical with the LSD of $(G_{q,X} + G^*_{q,X} - G_{q,Y} - G^*_{q,Y})$. Therefore, to test (3.32), we can plot the eigenvalue distribution of $(\hat{G}_{q,X} + \hat{G}^*_{q,X} - \hat{G}_{q,Y} - \hat{G}^*_{q,Y})$ and accept H_0 if the distribution appears to be degenerate at 0.

Let $\hat{\delta}_{q,X}$ and $\hat{\delta}_{q,Y}$ have their usual meaning. Define $\hat{\delta}^2_{q,XY} = \hat{\delta}^2_{q,X} + \hat{\delta}^2_{q,Y}$. By Corollary S11.4 and the independence of (1.1) and (2.5), it is easy to see that under H_0,

$$T_5 \equiv \sqrt{np^{-1}}\big(\operatorname{Tr}(\hat{G}_{q,X}) - \operatorname{Tr}(\hat{G}_{q,Y})\big)/\hat{\delta}_{q,XY} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

and under H_1, for some $0 < \tilde{\delta} < \infty$

$$\sqrt{np^{-1}}\big(\big(\operatorname{Tr}(\hat{G}_{q,X}) - \operatorname{Tr}(\hat{G}_{q,Y})\big) - \big(\operatorname{Tr}(G_{q,X}) - \operatorname{Tr}(G_{q,Y})\big)\big)/\hat{\delta}_{q,XY} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{\delta}^2).$$

Thus we can use T_5 as a test statistic and reject H_0 if it is large in absolute value.

REMARK 3.2. 1. It may be noted that even though most of the methods in Section 3 are based on the symmetric sum and product of $\{\hat{\Gamma}_u, \hat{\Gamma}^*_u\}$, other polynomials can also be used. We have restricted to the symmetric sum and product only for illustration. However, calculation of the asymptotic variances become increasingly difficult with higher order polynomials.

2. Inference in one-dimensional AR and ARMA models hinge on properties of the partial autocovariance sequences. Similarly, in IVAR and IVARMA models one can define partial autocovariance matrices. Large sample behavior of these matrices could be derived using our approach. Such results can then be applied to inference problems. We have briefly illustrated this idea in Examples 8–11 when coefficient matrices are symmetric and commutative. However, more general situations need further technical machinery and we shall attempt to develop it elsewhere.

3. Simulations show that the eigenvalue distribution of the nonsymmetric $\hat{\Gamma}_u$ are also well behaved; see, for example, Figure 3. However, no theoretical results are known.

Acknowledgment. The referees have provided valuable questions, comments and suggestions. In particular, their ideas have led to significant additions and changes to the manuscript. We remain very grateful to them.

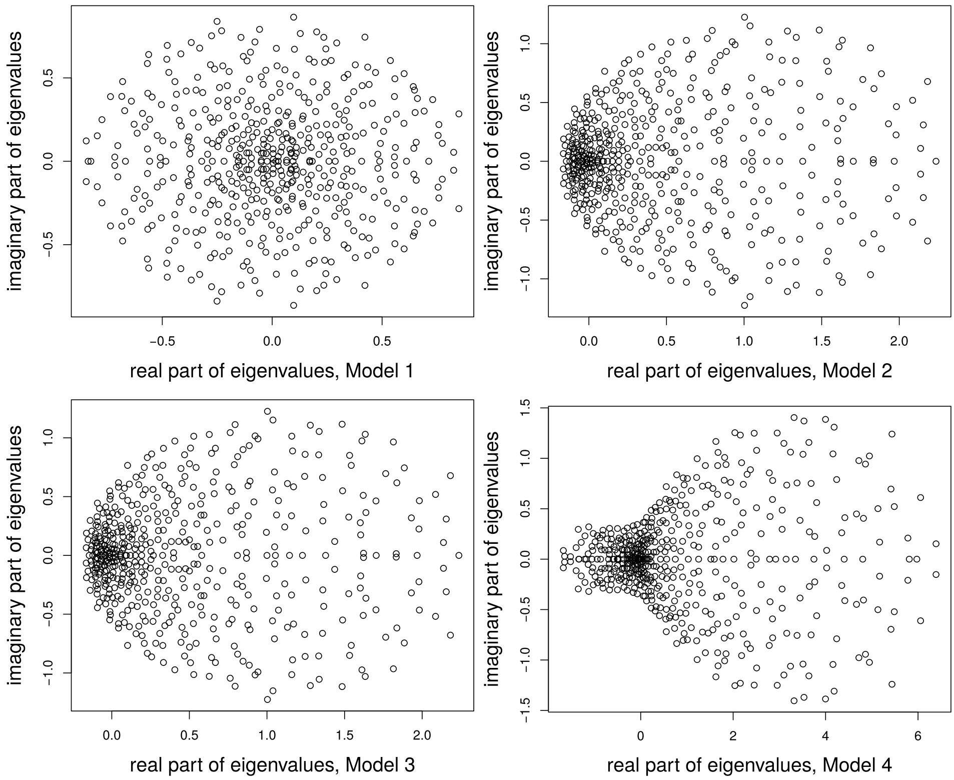

FIG. 3. *ESD of* $\sqrt{np^{-1}}(\hat{\Gamma}_1 - \Gamma_1)$ *for Models* $1-4$ *given in Section* 3.3 *and* $n = 10^3$, $p = n^{0.9}$.

SUPPLEMENTARY MATERIAL

Supplement to "Joint convergence of sample autocovariance matrices when $p/n \to 0$ with application." (DOI: 10.1214/18-AOS1785SUPP; .pdf). The supplementary file provides all technical details, free probability description of LSDs of symmetric polynomials in sample autocovariance matrices and the Stieltjes transform for some of the LSDs.

REFERENCES

BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792

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

BAI, Z., WANG, X. and ZHOU, W. (2009). CLT for linear spectral statistics of Wigner matrices. *Electron. J. Probab.* **14** 2391–2417. MR2556016

BAI, Z. D. and YIN, Y. Q. (1988). Convergence to the semicircle law. *Ann. Probab.* **16** 863–875. MR0929083

BAI, Z., JIANG, D., YAO, J. and ZHENG, S. (2013). Testing linear hypotheses in high-dimensional regressions. *Statistics* **47** 1207–1223. MR3175743

BAO, Z., LIN, L.-C., PAN, G. and ZHOU, W. (2015). Spectral statistics of large dimensional Spearman's rank correlation matrix and its application. *Ann. Statist.* **43** 2588–2623. MR3405605

BASU, S. and MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. MR3357870

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. (2016a). Large sample behaviour of high dimensional autocovariance matrices. *Ann. Statist.* **44** 598–628. MR3476611

BHATTACHARJEE, M. and BOSE, A. (2016b). Polynomial generalizations of the sample variance–covariance matrix when $pn^{-1} \to 0$. *Random Matrices Theory Appl.* **5** 1650014. MR3569842

BHATTACHARJEE, M. and BOSE, A. (2019). Supplement to "Joint convergence of sample autocovariance matrices when $p/n \to 0$ with application." DOI:10.1214/18-AOS1785SUPP.

BHATTACHARJEE, M. (2016). Asymptotics of Large Variance-Covariance and Autocovariance Matrices. PhD Thesis, Indian Statistical Institute.

BROCKWELL, P. J. and DAVIS, R. A. (2006). *Time Series: Theory and Methods. Springer Series in Statistics*. Springer, New York. MR2839251

DIACONIS, P. and EVANS, S. N. (2001). Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.* **353** 2615–2633. MR1828463

HOSKING, J. R. M. (1980). The multivariate portmanteau statistic. *J. Amer. Statist. Assoc.* **75** 602–608. MR0590689

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. *Math. USSR, Sb.* **1** 457–483.

NICA, A. and SPEICHER, R. (2006). *Lectures on the Combinatorics of Free Probability. London Mathematical Society Lecture Note Series* **335**. Cambridge Univ. Press, Cambridge. MR2266879

PFAFFEL, O. and SCHLEMM, E. (2011). Eigenvalue distribution of large sample covariance matrices of linear processes. *Probab. Math. Statist.* **31** 313–329. MR2853681

SHAO, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory* **27** 312–343. MR2782041

WANG, L., AUE, A. and PAUL, D. (2017). Spectral analysis of sample autocovariance matrices of a class of linear time series in moderately high dimensions. *Bernoulli* **23** 2181–2209. MR3648029

XIAO, H. and WU, W. B. (2014). Portmanteau test and simultaneous inference for serial covariances. *Statist. Sinica* **24** 577–599. MR3235390

ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263

ZHENG, S., BAI, Z. and YAO, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591. MR3316190

INFORMATICS INSTITUTE
UNIVERSITY OF FLORIDA
432 NEWELL DRIVE
GAINESVILLE, FLORIDA 32603
USA
E-MAIL: monaiidexp.gamma@gmail.com

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