LARGE SAMPLE BEHAVIOUR OF HIGH DIMENSIONAL AUTOCOVARIANCE MATRICES

MONIKA BHATTACHARJEE
Statistics and Mathematics Unit
Indian Statistical Institute
203 B.T. Road, Kolkata 700108
INDIA
monaiidexp.gamma@gmail.com

ARUP BOSE *
Statistics and Mathematics Unit
Indian Statistical Institute
203 B.T. Road, Kolkata 700108
INDIA
bosearu@gmail.com

This version October 24, 2014

Abstract

We study the large sample behaviour of the sequence of high dimensional sample autocovariance matrices $\{\hat{\Gamma}_i\}_{i\geq 0}$ from an infinite dimensional vector linear process and suggest ways of using our results for model diagnostics.

One way to describe the asymptotic behaviour of any large dimensional random matrix is through its limiting spectral distribution (LSD). The existence of LSD of $\hat{\Gamma}_i + \hat{\Gamma}_i^*$ for every i was proved in Jin *et al.* (2014) and Liu *et al.* (2013) respectively when the observations are from MA(0) and from an infinite dimensional vector linear process with appropriate (strong) assumptions on the coefficient matrices, by the method of Stieltjes transformation.

Under significantly weaker conditions on the coefficient matrices and the driving process, we prove that the expected average trace of any polynomial in these matrices converge. In particular, the LSD of any symmetric polynomials of these matrices, including the matrices $\hat{\Gamma}_i+\hat{\Gamma}_i^*$ and $\hat{\Gamma}_i\hat{\Gamma}_i^*$ 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 of our results.

The behaviour of this limit depends on the order of the process. When the observations are from an MA(0) process, we provide an explicit description for the LSD of $\hat{\Gamma}_i\hat{\Gamma}_i^*$. This behaviour changes when the model is a higher order linear process and consequently it automatically suggests a diagnostic method to determine the order of the infinite dimensional moving average process.

---

[*] Research supported by J.C. Bose National Fellowship, Dept. of Science and Technology, Govt. of India.

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

AMS 2010 Subject Classifications. Primary 37M10, 62M10, 58C40; Secondary 46L54, 62J20.

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 vector autoregressive moving average (IVARMA) processes. These are special cases (see Bhattacharjee & Bose (2014)) of the more general 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

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

For all t, $X_{t,p}^{(n)}$ and $\varepsilon_{t,p}$ are p-dimensional vectors. Precise assumptions of independence and finiteness of moments are discussed later.

We assume that the dimension of the observations increases proportionately with the sample size so that $p = p(n) \to \infty$ and $\frac{p}{n} \to y \in (0, \infty)$. This is one particular type of high dimensional model. The $p \times p$ matrices $\psi_{j,p}^{(n)}, \ \ j \geq 1$ will be called *coefficient matrices* and $\psi_0 = I_p$. The infinite sum in (1.1) exists in the almost sure sense under suitable decay conditions on the $\{\psi_j\}$. If the sum involved has $(q + 1)$ terms, then it will be called an MA(q) process. For convenience of notation, we will usually write p for $p(n)$ and often write ψ_j for $\psi_{j,p}^{(n)}$ etc.

For any time series model, the key quantities are the autocovariance matrices and their estimates. The *population autocovariance matrices* are defined as

$$\Gamma_{k,p} := E(X_{t,p}X^*_{(t-k),p}) = \sum_{j=k+1}^{\infty} \psi_j \psi^*_{j-k}, \ k = 0, 1, \ldots .$$

For each k, the moment estimator of $\Gamma_{k,p}$ is the *sample autocovariance matrix*,

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

We often write Γ_k and $\hat{\Gamma}_k$ respectively for $\Gamma_{k,p}$ and $\hat{\Gamma}_{k,p}$. When p is fixed, under suitable assumptions, $\{\hat{\Gamma}_{k,p}\}$ are weakly/strongly consistent for the population autocovariance matrices $\{\Gamma_{k,p}\}$ (see for example Brockwell & Davis (2009) and Hannan (1970)).

Natural questions which arise are how to estimate these autocovariance matrices, what are the large sample properties of these estimates and how to use these estimates for order determination or diagnostic checks.

It may be mentioned that, this model has been investigated by other researchers. Forni & Lippi (2001) provided sufficient conditions under which model (1.1) can be expressed as a *generalized dynamic-factor model*, Forni *et al.* (2000) proposed consistent estimators of the common components for the *generalized dynamic-factor model* and Forni *et al.* (2004) established their rate of convergence. Assuming sufficient decay conditions on ψ_j and certain regularity conditions on dispersion of $\varepsilon_{t,p}$, Bhattacharjee & Bose (2014) provided consistent estimators for autocovariance matrices in the model (1.1) through banding and smoothing of $\{\hat{\Gamma}_{k,p}\}$.

In this paper, we focus solely on the joint large sample behaviour of $\{\hat{\Gamma}_{k,p}\}$. One way to capture the behaviour of any given sequence of large dimensional random matrices is by studying its limiting spectrum. Suppose R_n is an $n \times n$ real symmetric matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ denote its eigenvalues. The *spectral measure* μ_n of R_n is the measure on $\mathbb{R}$ given by

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}$$

where δ_x is the Dirac delta measure at x. The probability distribution function on $\mathbb{R}$ corresponding to μ_n, is known as the *Empirical Spectral Distribution* (ESD) of R_n. If this ESD converges weakly (either almost surely or in probability) to some unique (non-degenerate) probability distribution, then the limit is called the *limiting spectral distribution* (LSD) of R_n. Incidentally, the study of the limit spectrum of non-symmetric matrices is extremely difficult and only a very few results are known. We do not deal with the spectrum of the non-symmetric autocovariance matrices in this article but with symmetric matrix polynomials in them.

One approach to establish the LSD of a symmetric random matrix is the method of moments, as outlined in Lemma 5.1. For some applications of this method, see Bai (1999) and Basak *et al.* (2014). Another widely used method is that of Stieltjes transformation, which for any measure μ on the real line is the function

$$m_\mu(z) = \int \frac{1}{x - z} \mu(dx), \quad z \in \mathbb{C}^+, \tag{1.3}$$

where $\mathbb{C}^+ := \{x + iy : x \in \mathbb{R}, y > 0\}$. Pointwise convergence of the Stieltjes transform 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 Bai (1993a), Bai (1993b), Bai *et al.* (2007), Bai *et al.* (1999) and Bai & Zhou (2008).

Let $X_t = \varepsilon_t \ \forall t$. Under suitable assumptions on $\{\varepsilon_t\}$, with probability one, the ESD of $\hat{\Gamma}_0$ converges weakly to the Marčenko-Pastur law with parameter y (see for example

Bai & Silverstein (2009)) whose moment sequence is given by

$$\beta_h = \sum_{k=1}^{h} \frac{1}{k}\binom{h-1}{k-1}\binom{h}{k} y^{k-1}, \ \ \forall h = 1, 2, 3, \ldots. \tag{1.4}$$

Jin *et al.* (2014) considered $\frac{1}{2}(\hat{\Gamma}_i + \hat{\Gamma}_i^*)$, when $X_t = \varepsilon_t \ \forall t$ satisfies appropriate assumptions. They showed that the limit Stieltjes transformation satisfies the bi-quadratic equation

$$(1 - y^2 m^2(z))(yzm(z) + y - 1)^2 = 1 \tag{1.5}$$

and therefore the LSD does not depend on the lag i.

Liu *et al.* (2013) considered the model (1.1) and assumed that $\{\varepsilon_{t,i} : t, i = 1, 2, 3, \ldots\}$ are i.i.d. mean 0 variance 1 random variables with finite fourth moment. They also assumed the following conditions on the coefficient matrices $\{\psi_j\}_{j=0}^{\infty}$.

(a) $\psi_j, \ j \in \mathbb{Z}$, are $p \times p$ simultaneously diagonalizable random Hermitian matrices, independent of ε_t, with $||\psi_j|| \leq \bar{\lambda}_{\psi_j}$ for large p, $\sum_{j=0}^{\infty} \bar{\lambda}_{\psi_j} < \infty$ and $\sum_{j=0}^{\infty} j\bar{\lambda}_{\psi_j} < \infty$. Here $|| \cdot ||$ is the operator norm.

(b) There exist: (*i*) continuous functions $f_l(\cdot), \ l \in \mathbb{Z}$, (*ii*) for every p, a distribution F_p^{ψ} with mass $\frac{1}{p}$ on each of the p points denoted as $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p$ subject to multiplicities, (*iii*) for every p, a unitary $p \times p$ matrix U such that for all $j \in \mathbb{Z}$, $U^* \psi_j U = \text{diag}\big(f_j(\lambda_1), f_j(\lambda_2), \ldots, f_j(\lambda_p)\big)$. The functions $f_j(\cdot), \ j \in \mathbb{Z}$, can also depend on p as long as they uniformly converge to continuous functions when $p \to \infty$.

(c) Almost surely, F_p^{ψ} converges weakly to a non-random probability distribution function F^{ψ} as $p \to \infty$.

Under these assumptions, they showed that for every $j = 0, 1, 2, \ldots$ the limiting Stieltjes transformations $m_j(z)$, of the ESD of $\frac{1}{2}(\hat{\Gamma}_j + \hat{\Gamma}_j^*)$ exist and are given by

$$m_j(z) = -\int \frac{dF^{\psi}(\lambda)}{z - \frac{1}{2\pi}\int_0^{2\pi} \frac{\cos(jv)h(\lambda,v)}{1+y\cos(jv)K_j(z,v)} dv} \ \ \forall z \in \mathbb{C}^+, \tag{1.6}$$

where for all $j = 0, 1, 2, \ldots$, $z \in \mathbb{C}^+$ and $v \in (0, \ 2\pi)$,

$$K_j(z, v) = -\int \frac{h(\lambda, v)dF^{\psi}(\lambda)}{z - \frac{1}{2\pi}\int_0^{2\pi} \frac{\cos(jv')h(\lambda,v')}{1+y\cos(jv')K_j(z,v')} dv'} \tag{1.7}$$

$$\text{and} \ \ h(\lambda, v) = |\sum_{l=0}^{\infty} e^{ilv} f_l(\lambda)|^2. \tag{1.8}$$

We also study the joint convergence of (any finitely many) sample autocovariance matrices in a unified way and in the process also relax the above assumptions considered in Liu *et al.* (2013). We also investigate if the above limits can be expressed in some alternative ways so as to provide further insight.

The most natural way to study the joint convergence of matrices is to consider them as elements of the *non-commutative probabilty space* (see Definition 5.1) of all $p \times p$ matrices with complex entries (see Example 1) with the *state* equal to normalized trace or the expected normalized trace. Then we can appeal to the rich theory of convergence of such spaces (see Definition 5.5).

We assume that the coefficient matrices have eigenvalues in a compact set and jointly converge (see Assumptions ($A10$) and ($A11$) in Section 2). These conditions are implied by the conditions (a), (b) and (c) given above. When $\epsilon_{t.i}$ (i-th component of ε_t) are assumed to satisfy ($A3$) or ($A4$), we show that the autocovariances converge jointly (see Theorem 2.1). This argument avoids the use of Stieltjes transform and is based on a more intuitive convergence of moments along with ideas from *free probability* introduced by Voiculescu (1991).

As a consequence of our main theorem, the existence of LSD for any symmetric polynomial including $\{\frac{1}{2}(\hat{\Gamma}_i + \hat{\Gamma}_i^*)\}_{i=0}^{\infty}$ and $\{\hat{\Gamma}_i\hat{\Gamma}_i^*\}_{i=0}^{\infty}$ follows (see Corollary 2.1). The limit can be expressed in terms of certain free random variables. The compact support of $\varepsilon_{t,j}$ and moment conditions are appropriately relaxed by suitable truncation arguments (see Corollary 2.1(b)).

The key idea in the proof is to embed a matrix with independent entries into a *Wigner matrix* and use projection (see (4.5)). As independent Wigner matrices and other deterministic matrices are *asymptotically freely independent* (see Definitions 5.5, 5.4 and Lemma 5.3), we have description of all convergences in terms of freely independent variables. This description and the theory of free probability provide tools for computing limiting moments (see Remark 2.2).

Existing results of Jin *et al.* (2014) and Liu *et al.* (2013) follow as special cases of Theorem 2.1 (see Corollaries 2.2(a), (c) and 2.5).

Corollary 2.2(b) says that, when $X_t = \varepsilon_t \; \forall t$, the LSD of $\left(\frac{n}{p}\right)^2 \hat{\Gamma}_i\hat{\Gamma}_i^*$ is free Bessel$(2, y^{-1})$ characterized by the moment sequence,

$$\beta_h = \sum_{k=1}^{h} \frac{1}{k}\binom{h-1}{k-1}\binom{2h}{k-1} y^{-k}, \tag{1.9}$$

provided ($A1$), ($A2$), ($A7$), ($A8$) and ($A9$) hold. If $y = 0$, then the LSD of $\left(\frac{n}{p}\right)\hat{\Gamma}_i\hat{\Gamma}_i^*$ is the Marčenko-Pastur law with parameter 1.

If the coefficient matrices ψ_j are of the form $\psi_j = \lambda_j I_p \;\; \forall j$, then the LSD of $(\hat{\Gamma}_i + \hat{\Gamma}_i^*)$ has a description in terms of a *compound free Poisson distribution* (see Definition 5.6) and its *free cumulants* have a very neat expression (see (2.15)). In this case, we also derive the Stieltjes transformation, which as a special case verifies (1.6) (see Corollary 2.4).

In Corollary 2.5, we provide a recursion relation for moments of LSD of $(\hat{\Gamma}_i + \hat{\Gamma}_i^*)$ under the assumption that the coefficient matrices are symmetric. This recursion for moments leads to the Stieltjes transformation (1.6), which agrees with the main result of Liu *et al.* (2013).

Under additional assumptions on $\{\psi_j\}$, the above results on $\{\hat{\Gamma}_i\}$ can be extended to the MA(∞) process also (see Corollary 2.6).

Our results may be used to develop diagnostic tests. The problem boils down to testing for independence of the residual process after fitting a model. There is a huge literature on this problem for univariate and multivariate time series models. As the dimension of the autocovariance matrices increases with the sample size, it is not possible to use the finite dimension tests in the infinite dimensional case. Using the changing behaviour of the LSD of the symmetrized autocovaiance matrices depending on whether or not X_t is MA(0) or not, in Section 3, we provide a diagnostic method for identifying the appropriate order of the moving average. We hope to investigate this in details elsewhere.

2 Main Results

Various authors have used different sets of assumptions on the driving preocess $\{\varepsilon_t\}$. We list below all these assumptions and not all of them will be used always.

Let $\varepsilon_{t.i}$ be the i-th component of ε_t.

(A1) $\{\varepsilon_{t.i} : t, i = 1, 2, 3, \ldots\}$ are independent.

(A2) $E(\varepsilon_{t.i}) = 0$ and $E|\varepsilon_{t.i}|^2 = 1$, for all $t, i = 1, 2, 3, \ldots$.

(A3) $\sup_{t,i} E(|\varepsilon_{t.i}|^k) < C_k < \infty, \;\; \forall k = 1, 2, 3, \ldots$.

(A4) For some sequence $\eta_n \downarrow 0, \;\; |\varepsilon_{t.i}| < \eta_n \sqrt{n} \;\; \forall t, i.$

(A5) For any $\eta > 0$, $\eta^{-2}(np)^{-1} \sum_{i=1}^{p} \sum_{t=1}^{n} E(|\varepsilon_{t.i}|^2 I(|\varepsilon_{ti}| > \eta\sqrt{n})) \to 0$.

For some $\delta \in (0, 2]$,

(A6) $\sup_{t,j} E(|\varepsilon_{t.j}|^{2+\delta}) < M < \infty$.

(A7) For any $\eta > 0$, $\frac{1}{\eta^{2+\delta} np} \sum_{j=1}^{p} \sum_{t=1}^{n} E(|\varepsilon_{t.j}|^{2+\delta} I(|\varepsilon_{t.j}| > \eta n^{\frac{1}{2+\delta}})) \to 0$.

(A8) $\sup_{t,j} E|\varepsilon_{t.j}|^4 < M < \infty$..

(A9) The dimension $p = p(n) \to \infty$ and $\frac{p}{n} \to y \in (0, \infty)$.

(A10) For all $j \geq 0$, ψ_j's are compactly supported and jointly converge (see Definition 5.5 for definition of joint convergence).

Theorem 2.1 of this section provides the joint convergence of sample autocovariance matrices. As a corollary, this implies the existence of the LSD of any symmetric polynomials of $\{\hat{\Gamma}_i, \hat{\Gamma}_i^* : i \geq 0\}$, which includes $\hat{\Gamma}_0$, $\hat{\Gamma}_i + \hat{\Gamma}_i^*$, $\hat{\Gamma}_i\hat{\Gamma}_i^*$ for all $i \geq 1$. In particular, this also implies the results of Jin *et al.* (2014) and Liu *et al.* (2013). The theorem also provides an algebraic description of the limits in terms of some natural free variables.

As discussed in Section 1, we use method of moments (Lemma 5.1) to show existence of LSD. The most crucial condition to verify in that lemma is the $(M1)$ condition. Free probability provides the tools for checking this condition.

The Wigner matrix plays a central role in the proof and also in the expression of the limits. For our purposes, this is a symmetric matrix whose all elements on or above diagonal are independent and have mean zero and variance 1. If $W_n = ((w_{ij}))$ is an $n \times n$ Wigner matrix then in the sense of Definition 5.5, $W_n/\sqrt{n}$ with the state as expected normalized trace converges to a semi-circle element s with a corresponding state φ, provided either $(A3)$ or $(A4)$ holds for w_{ij}. Under these assumptions or when the entries are i.i.d. with mean zero and variance one, the almost sure LSD of this matrix is the semi-circular law with probability density function

$$f(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, & -2 \le x \le 2 \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

and the limit state satisfies

$$\varphi(s^k) = \int_{-2}^{2} x^k f(x)dx \;=\; \begin{cases} 0, & \text{if } k \text{ is odd} \\ \frac{k!}{(\frac{k}{2})!(\frac{k}{2}+1)!}, & \text{if } k \text{ is even.} \end{cases} \tag{2.2}$$

Further, it is also well known that if we have k sequences of norm bounded deterministic matrices of order n which converge jointly to say $a_1, \ldots a_k$, then they also converge jointly together with the Wigner matrix (under any of the above assumptions) and, s and $\{a_1, \ldots a_k\}$ are free (see Lemma 5.3).

Our main idea is to write the $\hat{\Gamma}_{k,i}$ as matrix polynomials of simpler matrices which include an ID matrix Z (matrix whose all elements are independent with mean 0 and variance 1), the matrices ψ_j, and matrices whose ith diagonal equals one and rest of the elements are 0. We embed the ID matrix in a symmetric Wigner matrix of a larger dimension and use projections and then appeal to the above results.

To illustrate the idea, consider the MA(0) process $X_t = \varepsilon_t$, for all $t \ge 1$ where Assumption $(A1)$, $(A2)$, $(A9)$ and either $(A3)$ or $(A4)$ hold. In this case the variance-covariance matrix $\hat{\Gamma}_0$ can be written as

$$\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} W \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} W \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} n\hat{\Gamma}_0 & 0 \\ O & 0 \end{pmatrix},$$

for some Wigner matrix W whose entries satisfy either $(A3)$ or $(A4)$. Note that all non-zero eigenvalues of $\left(\begin{smallmatrix} n\hat{\Gamma}_0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $n\hat{\Gamma}_0$ are same. Moreover, W and the deterministic matrices $\left(\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 & 0 \\ 0 & I \end{smallmatrix}\right)$ are asymptotically freely independent (see Lemma 5.3). Using this description, one can easily show that the LSD of $\hat{\Gamma}_0$ in this special case is the Marčenko-Pastur law whose moments are given by (1.4). For details, see Corollary 2.2 (a) and its proof. This approach can be extended to the MA(q) and then to the MA(∞) process.

For any matrix B of order p, we define $\bar{B}$ of order $(n+p)$ as

$$\bar{B} = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.3}$$

Consider the sequence of non-commutative $*$-probability spaces $(\mathcal{A}_n, \varphi_n)$ (see Definition 5.1, Example 1 and (5.12))

$$\mathcal{A}_n = *\text{-algebra generated by } \{n(n+p)^{-1}\bar{\hat{\Gamma}}_i, n(n+p)^{-1}\bar{\hat{\Gamma}}_i^* : i \geq 0\} \text{ and} \tag{2.4}$$

$$\varphi_n = \frac{1}{n+p} E\,\mathrm{Tr}(\cdot). \tag{2.5}$$

To describe the limits, let $(\mathcal{A}, \varphi)$ be a non-commutative $*$-probability space such that

$$\mathcal{A} = *\text{-algebra generated by } s, \{c_i, c_i^* : \ c_0 = c_0^*, i \geq 0\} \text{ and } \{a_i, a_i^* : i \geq 0\}, \tag{2.6}$$

where s is a semi-circular element defined in (2.1) and for any finite monomial $m(c_i, c_i^*, : \ i = 0, 1, 2, \ldots)$ and $m(a_i, a_i^*, : \ i = 0, 1, 2, \ldots)$ from $\mathrm{Span}\{c_i, c_i^* : \ c_0 = c_0^*, i \geq 0\}$ and $\mathrm{Span}\{a_i, a_i^* : i \geq 0\}$ respectively,

$$\varphi(m(c_i, c_i^*, : \ i = 0, 1, 2, \ldots)) = \begin{cases} \frac{1}{1+y}, & \text{if there is same number of } c_i \text{ and } c_i^* \text{ for all } i \geq 1 \\ 0, & \text{otherwise,} \end{cases} \tag{2.7}$$

$$\varphi(m(a_i, a_i^* : \ i = 0, 1, 2, \ldots)) = \frac{y}{1+y} \lim \left\{ \frac{1}{p} \mathrm{Tr}\left(m\left(\psi_j, \psi_j^* : \ j = 0, 1, 2, \ldots \right) \right) \right\} \tag{2.8}$$

and s, $\mathrm{Span}\{c_i, c_i^* : c_0 = c_0^*, i \geq 0\}$ and $\mathrm{Span}\{a_i, a_i^* : i \geq 0\}$ are *freely independent* (see Definition 5.4). In other words, $\mathcal{A}$ is the free product of $\mathrm{Span}\{s\}$, $\mathrm{Span}\{c_i, c_i^* : c_0 = c_0^*, i \geq 0\}$ and $\mathrm{Span}\{a_i, a_i^* : i \geq 0\}$. Also let $c_{-i} = c_i^*$ $\forall i \geq 1$. Let us define for all $i = 0, 1, 2, \ldots$,

$$\gamma_{iq} = \sum_{j=0}^{q} \sum_{j'=0}^{q} a_j s c_{j'-j+i} s a_{j'}^*, \qquad \gamma_{iq}^* = \sum_{j=0}^{q} \sum_{j'=0}^{q} a_{j'} s c_{j'-j+i}^* s a_j^*. \tag{2.9}$$

Next, for all $q \geq 0$, consider the $*$-probability spaces $(\mathcal{B}_q, \varphi)$ (see (5.12)), where

$$\mathcal{B}_q = \mathrm{Span}\{\gamma_{iq}, \gamma_{iq}^* : i \geq 0\} \subset \mathcal{A}. \tag{2.10}$$

Then we have the following theorem. Its proof is given in Section 4.1.

Theorem 2.1. *Suppose $X_t \sim MA(q)$ process defined in (1.1) and* $(A1)$, $(A2)$, $(A9)$, $(A10)$ *and either* $(A3)$ *or* $(A4)$ *hold. Then $(\mathcal{A}_n, \varphi_n)$ converges to $(\mathcal{B}_q, \varphi)$ and $(n(n+p)^{-1}\bar{\hat{\Gamma}}_i$, $n(n+p)^{-1}\bar{\hat{\Gamma}}_i^*)$ are asymptotically distributed as $(\gamma_{iq}, \gamma_{iq}^*)$.*

Remark 2.1. *Theorem 2.1 continues to hold if instead of $\{\varepsilon_{t.i} : t = 1, 2, \ldots, n, \ i = 1, 2, \ldots, p(n)\}$, we have a triangular sequence $\{\varepsilon_{t.i,n} : \ t = 1, 2, \ldots, n, \ i = 1, 2, \ldots, p(n)\}$ with all the necessary obvious changes in the assumptions. This immediately follows when we carefully go through the proof of Theorem 2.1 in Section 4.1.*

Note that any polynomial $\Pi(n(n+p)^{-1}\bar{\hat{\Gamma}}_i, \ n(n+p)^{-1}\bar{\hat{\Gamma}}_i^* : \ i \geq 0)$ of $\{\bar{\hat{\Gamma}}_i, \bar{\hat{\Gamma}}_i^* : \ i \geq 0\}$ is in $(\mathcal{A}_n, \varphi_n)$. Therefore, by Theorem 2.1, $\Pi(\gamma_{iq}, \gamma_{iq}^* : i \geq 0)$ is its limit in $(\mathcal{B}_q, \varphi)$ (see Definition 5.5) and hence

$$\lim \varphi_n(\Pi(n(n+p)^{-1}\bar{\hat{\Gamma}}_i, \ n(n+p)^{-1}\bar{\hat{\Gamma}}_i^* : \ i \geq 0))^h = \varphi(\Pi(\gamma_{iq}, \gamma_{iq}^* : i \geq 0))^h, \ \forall h \geq 1. \tag{2.11}$$

The following corollaries follow from this theorem.

Corollary 2.1. *(a) Suppose $X_t \sim MA(q)$ process defined in (1.1) and $(A1)$, $(A2)$, $(A9)$, $(A10)$ and either $(A3)$ or $(A4)$ hold. Then LSD for any symmetric polynomial, say $\Pi(\hat{\Gamma}_i, \hat{\Gamma}_i^* : i \geq 0)$, of $\{\hat{\Gamma}_i, \hat{\Gamma}_i^* : i \geq 0\}$ exists almost surely. Consequently, for every $i \geq 0$, the LSD of $\hat{\Gamma}_i + \hat{\Gamma}_i^*$ and $\hat{\Gamma}_i\hat{\Gamma}_i^*$ exist almost surely for $MA(q)$ process.*

Moreover, if μ is the probability measure corresponding to any symmetric monomial $m_k(\gamma_{iq}, \gamma_{iq}^* : i \geq 0)$ having k many $\hat{\gamma}_i$ or $\hat{\gamma}_i^*$ and S is the random variable distributed as the LSD of $m_k(\hat{\Gamma}_i, \hat{\Gamma}_i^* : i \geq 0)$, then

$$\mathcal{L}\left(\frac{S}{(1+y)^k}\right) = \frac{1+y}{y}\left(\mu - \frac{1}{1+y}\delta_0\right), \tag{2.12}$$

where δ_0 is the Dirac measure at 0.

(b) Suppose $X_t \sim MA(q)$ process defined in (1.1) and $(A1)$, $(A2)$, $(A9)$ and $(A10)$ hold. Then under the Assumptions $(A6)$ and $(A7)$, LSD of $\{\hat{\Gamma}_i + \hat{\Gamma}_i^*\}_{i\geq 0}$ exist. Moreover, if $(A7)$ and $(A8)$ hold, then LSD of $\{\hat{\Gamma}_i\hat{\Gamma}_i^*\}_{i\geq 0}$ exist. Further, for the $MA(0)$ process, LSD of $\hat{\Gamma}_0$ continues to exist if Assumptions $(A6)$ and $(A7)$ are dropped and instead we assume $(A5)$ holds.

To prove (a), one needs to verify $(M1)$, $(M4)$ (($M2$), if we demand only in probability convergence) and (C) (see Lemma 5.1). $(M1)$ follows from (2.11). Proofs of $(M2)$, $(M4)$ and (C) are given in Section 4.2 (a). Proof of (b) is based on truncation arguments and is given in Section 4.2 (b).

Remark 2.2. *Although there is no closed form expression for arbitrary moments of the above LSD, we can calculate the moments recursively. It is clear that we need to get hold of $\varphi(sb_1sd_1sb_2sd_2\ldots sb_nsd_n)$, $\forall n \geq 1$, where $b_1, b_2, \ldots, b_n \in Span\{a_i, a_i^* : i = 0, 1, 2, \ldots\}$ and $d_1, d_2, \ldots, d_n \in Span\{c_i, c_i^* : i = 0, 1, 2, \ldots\}$. By (5.18) and Lemma 5.2,*

$$\varphi(sb_1sd_1sb_2sd_2\ldots sb_nsd_n) = \sum_{\pi\in NC_2(2n)} \varphi_{K(\pi)}[b_1, d_1, b_2, d_2, \ldots, b_n, d_n],$$

where $NC_2(2n)$ is the the set of all non-crossing pair partitions of $\{1, 2, \ldots, 2n\}$ (see non-crossing partitions in Section 5). Hence, by freeness of $\{b_1, b_2, \ldots, b_n\}$ and $\{d_1, d_2, \ldots, d_n\}$ and Properties 1 to 3 of the Kreweras complement $K(\pi)$ (see Section 5), we have

$$\begin{aligned}\varphi(sb_1sd_1sb_2sd_2\ldots sb_nsd_n) &= \sum_{\pi\in NC(n)} \varphi_\pi[b_1, b_2, \ldots, b_n]\varphi_{K(\pi)}[d_1, d_2, \ldots, d_n] \\ &= \sum_{\pi\in NC(n)} \varphi_\pi[d_1, d_2, \ldots, d_n]\varphi_{K(\pi)}[b_1, b_2, \ldots, b_n],\end{aligned} \tag{2.13}$$

where $NC(n)$ is the the set of all non-crossing partitions of $\{1, 2, \ldots, n\}$ (see non-crossing partitions in Section 5). We use this formula in the proof of next corollaries.

Corollary 2.2. *Consider the $MA(0)$ process i.e. $X_t = \varepsilon_t$ $\forall t$ and suppose Assumptions $(A1), (A2), (A9)$ and either $(A3)$ or $(A4)$ hold. Then the following results (a)-(c) hold.*

(a) Marčenko-Pastur law: *The LSD of* $\hat{\Gamma}_0$ *is the Marčenko-Pastur law. Moreover, the result continues to hold if we assume* $(A5)$ *instead of* $(A3)$ *or* $(A4)$.
(b) Free Bessel law: *The LSD of* $\left(\frac{n}{p}\right)^2 \hat{\Gamma}_i \hat{\Gamma}_i^*$ *is the free Bessel*$(2, y^{-1})$ *law. For* $y = 0$*, the LSD of* $\left(\frac{n}{p}\right) \hat{\Gamma}_i(\varepsilon)\hat{\Gamma}_i^*(\varepsilon)$ *is the Marčenko-Pastur law with parameter* 1*. Moreover, in both cases, the LSD is identical for all* $i \geq 1$*. If we assume* $(A7)$ *and* $(A8)$ *instead of* $(A3)$ *or* $(A4)$*, then also these results holds.*
(c) *The LSD of* $\frac{1}{2}(\hat{\Gamma}_i + \hat{\Gamma}_i^*)$ *are identical for all* $i \geq 1$ *and their common Stieltjes transformation* $m(z)$ *satisfies (1.5). This proves Theorem* 2.1 *of Liu* et al. *(2013)) for the* $MA(0)$ *case and Theorem* 1.1 *of Jin* et al. *(2014). Also this result holds if we assume* $(A6)$ *and* $(A7)$ *instead of* $(A3)$ *or* $(A4)$.

Proof of Corollary 2.2 is given in Section 4.3.

Remark 2.3. *LSD of the sample covariance matrix without independence structures in columns is established in Bai & Zhou (2008).*

Corollary 2.3. *Let* $X_t = \varepsilon_t \ \forall t$ *where* $\{\varepsilon_{t,j}\}$*'s are all i.i.d. random variables with mean* 0 *and variance* 1*. Moreover, suppose* $(A9)$ *holds. Then the Stieltjes transform of the LSD of* $\Sigma^{1/2}\hat{\Gamma}_0\Sigma^{1/2}$ *is given by*

$$m(z) = \int \frac{dF_{\Sigma}(t)}{z - t(1 - y - yzm(z))}, \tag{2.14}$$

where Σ *is a positive definite and symmetric matrix with compactly supported LSD* F_{Σ}*. For a more general result see Bai & Zhou (2008). If* $F_{\Sigma} = \delta_1$*, this reduces to the Stieltjes transform of the Marčenko-Pastur law (see for example Nica & Speicher (2006), Couillet & Debbah (2011) or Bai & Silverstein (2009)).*

Corollary 2.4. *Let* $X_t \sim MA(q)$ *process and suppose Assumptions* $(A1), (A2), (A9)$ *and either* $(A3)$ *or* $(A4)$ *hold. Let* $\psi_j = \lambda_j I_p$*, for all* j*. Then the LSD of* $\frac{1}{2}(\hat{\Gamma}_i + \hat{\Gamma}_i^*)$ *is compound free Poisson (see Definition 5.6) and its* r*-th order free cumulant is given by*

$$K_{ir} = y^{r-1} E_v(\cos(iv)h(\lambda, v))^r, \ \ \forall i \geq 0, \tag{2.15}$$

where h *is given in (1.8) and* $v \sim U(0, 2\pi)$*. Moreover, (2.15) holds if we assume* $(A6)$ *and* $(A7)$ *instead of* $(A3)$ *or* $(A4)$*. These together with Corollary 2.1*(b) *are compatible with the results obtain in Liu* et al. *(2013).*

In the next corollary, we provide a recursion formula for the moments of LSD of $\frac{1}{2}(\hat{\Gamma}_i + \hat{\Gamma}_i^*)$. Let the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ be denoted by $\bar{\gamma}_{uq}$ and v be a $U(0, 2\pi)$ variable independent of $\bar{\gamma}_{uq}$. We define

$$R_{uj}(v) = \varphi\left(| \sum_{l=0}^{q} e^{ilv} a_l|^2 \bar{\gamma}_{uq}^{j-1} \right) \ \ \forall j \geq 1.$$

Now, for any $r \geq 1$, we define

$$\mathcal{P}_r := \{\pi = (i_1, i_2, \ldots, i_t) : t \geq 1,\ i_j \neq 0,\ \forall j,\ \sum_{j=1}^{t} i_j = r\} = \text{ set of all integer partitions } \pi \text{ of } r.$$

For each $\pi \in \mathcal{P}_r$, we also define

$$\begin{aligned} t &:= \text{ total number of digits in } \pi, \\ s &:= \text{ total number of distinct digits in } \pi, \end{aligned}$$

and they are, say, $i_{j_1} < i_{j_2} < \ldots < i_{j_s}$ with corresponding frequencies $f_1, f_2, \ldots, f_s$ respectively in π.

Corollary 2.5. *Let* $X_t \sim MA(q)$ *process and suppose Assumptions* $(A1), (A2), (A9), (A10)$ *and either* $(A3)$ *or* $(A4)$ *hold. Also suppose that* ψ_j *are symmetric. Then, we have*

$$\varphi(\bar{\gamma}^r_{uq}) = E_\nu\Big(\sum_{\pi \in \mathcal{P}_r}\Big(\frac{t!}{\prod_{k=1}^{s} f_k!} y^{t-1}(\cos(i\nu))^t \prod_{k=1}^{t} R_{ui_k}(\nu)\Big)\Big). \tag{2.16}$$

Under the assumptions of Liu *et al.* (2013) (see (a), (b) and (c) in introduction) the Stieltjes transformation obtained there can be easily obtained using (2.16) and the expansion given in (5.5). However, (2.16) remains true more generally if the coefficient matrices are symmetric and $(A1), (A2), (A6), (A7), (A9), (A10)$ hold.

Proof of Corollaries 2.3, 2.4 and 2.5 are given in Sections 4.4, 4.5 and 4.6 respectively.

To treat the MA(∞) case, consider the additional assumption

(A11) $\sum_{j=1}^{\infty} \left\{\varphi(a_j^* a_j a_j^* a_j)\right\}^{\frac{1}{4}} < \infty$ and for every $\varepsilon > 0$, there exists a $p_0 \geq 1$ such that for every $p \geq p_0$

$$\sum_{j=1}^{\infty} \left\{\frac{1}{p}\operatorname{Tr}(\psi_j^* \psi_j \psi_j^* \psi_j)\right\}^{\frac{1}{4}} < \infty.$$

Let us define for all $i \geq 0$,

$$\gamma_{i,\infty} = \sum_{j=0}^{\infty}\sum_{j'=0}^{\infty} a_j s c_{j'-j+i} s a_{j'}^*, \qquad \gamma_{i,\infty}^* = \sum_{j=0}^{\infty}\sum_{j'=0}^{\infty} a_{j'} s c_{j'-j+i}^* s a_j^*. \tag{2.17}$$

Note that the infinite sums exist as $\varphi(\gamma_{i,\infty})$, $\varphi(\gamma_{i,\infty}^*)$ are finite $\forall i$.

Corollary 2.6. (a) *Suppose* $X_t \sim MA(\infty)$ *process and* $(A1)$, $(A2)$, $(A4)$ $(A9)$, $(A10)$ *and* $(A11)$ *hold. Then the LSD of* $\{n(n+p)^{-1}(\bar{\hat{\Gamma}}_i + \bar{\hat{\Gamma}}_i^*)\}_{i\geq 0}$ *and* $\{n^2(n+p)^{-2}\bar{\hat{\Gamma}}_i\bar{\hat{\Gamma}}_i^*\}_{i\geq 0}$ *exist and* *they are given by* $\{\gamma_{i,\infty} + \gamma^*_{i,\infty}\}_{i\geq 0}$ *and* $\{\gamma_{i,\infty}\gamma^*_{i,\infty}\}_{i\geq 0}$ *respectively. Corollaries 2.4 and 2.5 hold for MA*(∞) *process after replacing all q by* ∞.

(b) *The above result holds* (i) *for* $\{n(n+p)^{-1}(\bar{\hat{\Gamma}}_i + \bar{\hat{\Gamma}}^*_i)\}_{i\geq 0}$ *if we assume* $(A6)$ *and* $(A7)$ *instead of* $(A4)$ *and* (ii) *for* $\{n^2(n+p)^{-2}\bar{\hat{\Gamma}}_i\bar{\hat{\Gamma}}^*_i\}_{i\geq 0}$ *if we assume* $(A7)$ *and* $(A8)$ *instead of* $(A4)$.

Proof of Corollary 2.6(a) is given in Section 4.7. Proof of Corollary 2.6(b) can be done similarly using the same arguments as in Section 4.2 (b).

Remark 2.4. *By (2.7), note that the joint distributions of* $\{c_{j-j'+i} : j, j' = 0, 1, 2, \ldots, q\}$ *are identical for all* $i > q$ *and are different for all* $0 \leq i \leq q$. *Therefore, when* X_t *is an MA*(q) *process, the LSD of* $\hat{\Gamma}_i\hat{\Gamma}^*_i$ *are identical for all* $i > q$ *and are different for all* $0 \leq i \leq q$. *Consequently, under MA*(∞) *process, the LSD of* $\hat{\Gamma}_i\hat{\Gamma}^*_i$ *are different for all* $i \geq 0$. *The same result holds for the matrices* $\hat{\Gamma}_i + \hat{\Gamma}^*_i$ *, for* $i \geq 1$. *In Section 3, we use this remark to suggest an appropriate order determination method.*

3 Model diagnostics

There is a huge literature on diagnosis of the appropriate order of a time series model (see for example, Ljung & Box (1978), Mc.Leod (1978), Katayama (2008), Hong (1996), Shao (2011)). The same problem in finite dimensional set up has also been studied (see Francq & Raïsi (2007), Hosking (1980), Hosking (1981b), Hosking (1981a), Lin & Mc.Leod (1981), Lin & Mc.Leod (2006), Mahdi & Mc.Leod (2012)). As the dimension of the autocovariance matrices increases with the sample size, it is not possible to use the finite dimensional tests in the infinite dimensional case. Here we provide a diagnostic method for identifying the appropriate order of the moving average.

By Remark 2.4, when X_t is an MA(q) process, the LSD of $(\frac{n}{p})^2\hat{\Gamma}_i\hat{\Gamma}^*_i$ are identical for $i > q$ and different for $0 \leq i \leq q$. Moreover, when X_t is an MA(∞) process, the LSD of $(\frac{n}{p})^2\hat{\Gamma}_i\hat{\Gamma}^*_i$ are all different. Therefore, for a given large data set, we can check different order moments of the ESD of $(\frac{n}{p})^2\hat{\Gamma}_i\hat{\Gamma}^*_i$ for $i = 1, 2, 3, \ldots$. If there is a $q \geq 0$ such that for a fixed $r \geq 1$, the r-th order moments are identical for all $i > q$ and this holds for all $r \geq 1$ (in application for sufficiently many r's), then q will be the appropriate order. Moreover, if there is no such q such that the above phenomenon holds, then the data set is either from an MA(∞) process or it is not from any linear model.

To convince ourselves of the above diagnostic method, we consider a particular MA(q) process, for $q = 0, 1, 2, 3, 4, 5$, where all the coefficient matrices are I_p, identity matrix of order p, i.e.,

$$X_t = \sum_{k=0}^{q} \varepsilon_{t-k}, \quad q = 0, 1, 2, 3.$$

We also consider $y = 1$. For each fixed $q = 0, 1, 2, 3$ and $r = 1, 2, 3$, we plot the r-th order moments of $(\frac{n}{p})^2\hat{\Gamma}_i\hat{\Gamma}^*_i$ for $i = 1, 2, \ldots, 8$. It is observed from the figures that, for each $r = 1, 2, 3$, the r-th order moments are more or less same for each $i = q+1, q+2, \ldots, 8$, when X_t is an MA(q) process, $q = 0, 1, 2, 3$.

Also, note that the 1st, 2nd and 3rd order moments under MA(0) process are near about 1, 3 and 12 respectively, which are respectively the 1st , 2nd and 3rd order moments of Bessel(2,1) law.

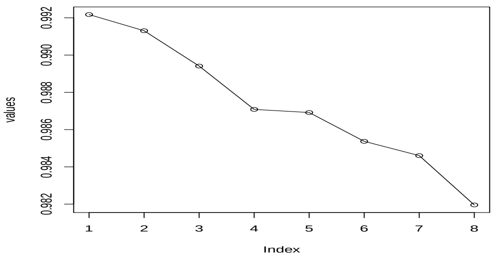

Figure 4.1*:* *MA*(0) *process:* 1*st moment*

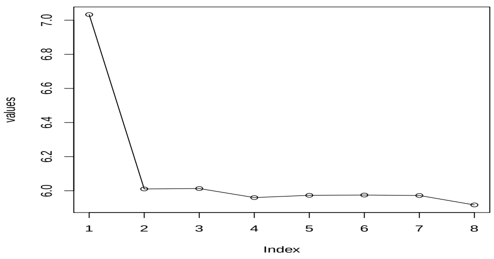

Figure 4.2 *MA*(1) *process:* 1*st moment*

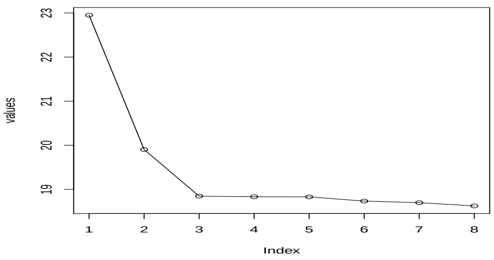

Figure 4.3*:* *MA*(2) *process:* 1*st moment*

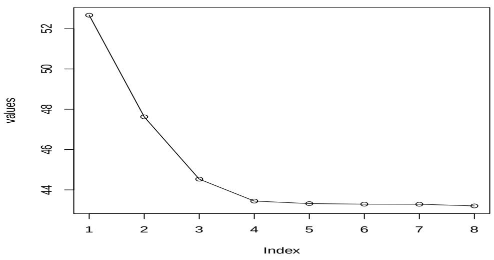

Figure 4.4 *MA*(3) *process:* 1*st moment*

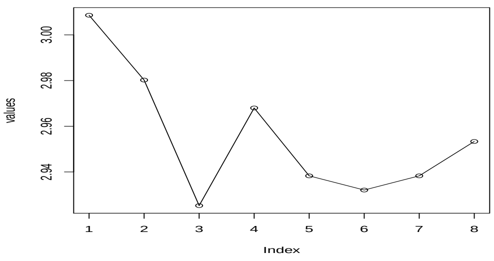

Figure 4.5*:* *MA*(0) *process:* 2*nd moment*

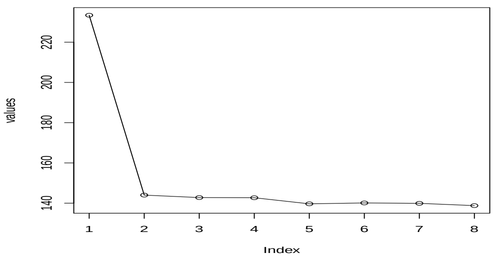

Figure 4.6 *MA*(1) *process:* 2*nd moment*

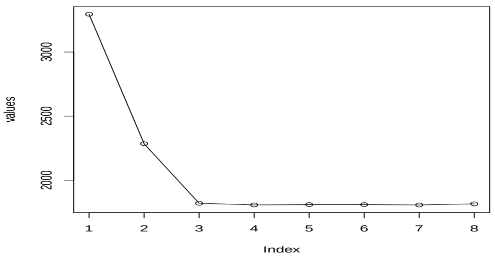

Figure 4.7*:* *MA*(2) *process:* 2*nd moment*

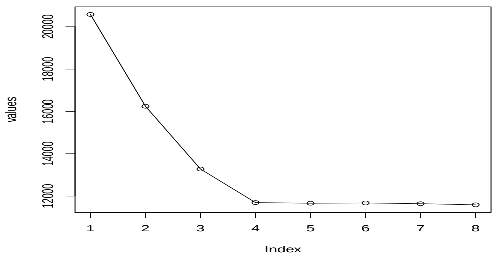

Figure 4.8 *MA*(3) *process:* 2*nd moment*

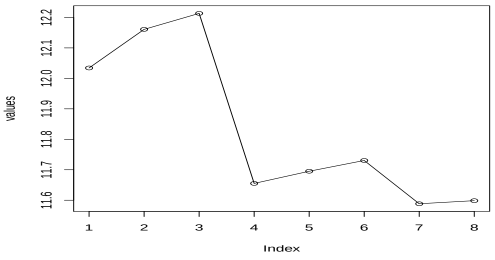

Figure 4.9*:* *MA*(0) *process:* 3*rd moment*

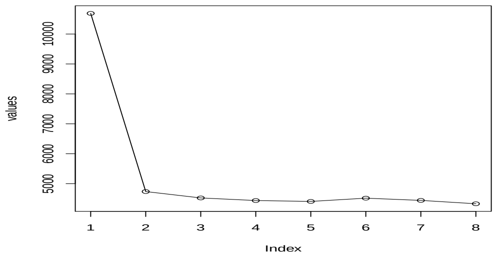

Figure 4.10 *MA*(1) *process:* 3*rd moment*

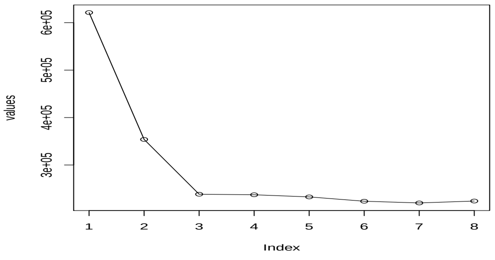

Figure 4.11*:* *MA*(2) *process:* 3*rd moment*

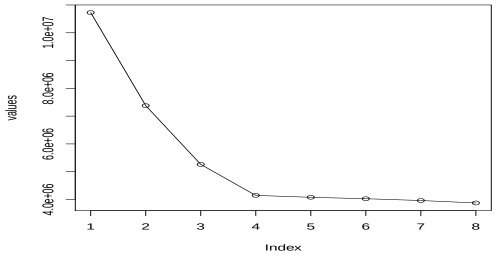

Figure 4.12 *MA*(3) *process:* 3*rd moment*

4 Proofs

4.1 Proof of Theorem 2.1

Let $\hat{\Gamma}_k(\varepsilon)$ be the k-th order sample autocovariance matrix of the process $\{\varepsilon_t\}$. We define

$$\Delta_k = \sum_{j=0}^{q} \sum_{j'=0}^{q} \psi_j \hat{\Gamma}_{j'-j+k}(\varepsilon) \psi_{j'}^*, \quad \forall k \geq 0. \tag{4.1}$$

Later we will show that $\hat{\Gamma}_k$ and Δ_k have same limits in average trace. Before, that we will show that Theorem 2.1 is still true if we replace $\mathcal{A}_n$ by

$$\mathcal{A}_n^{\Delta} = *\text{-algebra generated by } \{n(n+p)^{-1}\bar{\Delta}_k, n(n+p)^{-1}\bar{\Delta}_k^* : k \geq 0\} \tag{4.2}$$

with the same state φ_n. We define a $p \times n$ matrix

$$Z = (\varepsilon_1 \ \varepsilon_2 \ \ldots \varepsilon_n). \tag{4.3}$$

It is called the ID matrix as all its elements are independently distributed with mean 0 and variance 1. We embed Z into a Wigner matrix W of order $(n+p)$. Thus

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

where $W^{(1)}$ and $W^{(2)}$ are two independent Wigner matrices of order p and n respectively and also independent of Z and satisfy either $(A3)$ or $(A4)$. For any matrix D of order n, we define $\underline{D}$ of order $(n+p)$ as

$$\underline{D} = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix}. \tag{4.4}$$

Therefore, we can write

$$n(n+p)^{-1}\bar{\Delta}_i = (n+p)^{-1} \sum_{j=0}^{q} \sum_{j'=0}^{q} \bar{\psi}_j W \underline{P}_{j'-j+i} W \bar{\psi}_{j'}^*, \quad \forall i \geq 0. \tag{4.5}$$

To get a handle on polynomials of $\bar{\Delta}_i$'s, consider the following three sequences of non-commutative $*$-algebras with the state φ_n defined in (2.5):

$$\begin{aligned}
\mathcal{M}_n &= \text{Span}\{(n+p)^{-1/2}W, \bar{\psi}_j, \bar{\psi}_j^*, \underline{P}_j, \underline{P}_j^* : \ j = 0, 1, 2, \ldots\}, \\
\mathcal{M}_n^{(1)} &= \text{Span}\{\underline{P}_j, \underline{P}_j^* : \ j = 0, 1, 2, \ldots\}, \\
\mathcal{M}_n^{(2)} &= \text{Span}\{\bar{\psi}_j, \bar{\psi}_j^* : \ j = 0, 1, 2, \ldots\}.
\end{aligned}$$

Note that $\mathcal{A}_n^{\Delta}$, $\mathcal{M}_n^{(1)}$, $\mathcal{M}_n^{(2)}$ and $\text{Span}\{(n+p)^{-1/2}W\}$ are all sub-algebras of $\mathcal{M}_n$. First we concentrate on the finite marginal monomials coming from any one of $\text{Span}\{(n+p)^{-1/2}W\}$ or $\mathcal{M}_n^{(1)}$ or $\mathcal{M}_n^{(2)}$.

Note that the LSD of $(n+p)^{-1/2}W$ is the standard semi-circular law (see Arnold (1967), Arnold (1971), Wigner (1958)). Therefore, for $h \geq 1$, $\lim \varphi_n((n+p)^{-1/2}W)^h = \varphi(s^h)$ (see (2.2)).

Moreover, it is not hard to verify that for any finite monomials $m(\underline{\mathrm{P}}_j, \underline{\mathrm{P}}_j^* : \ j = 0,1,2,\ldots)$ and $m(\bar{\psi}_j, \bar{\psi}_j^* : \ j = 0,1,2,\ldots)$, we have

$$\lim \varphi_n\left(m\left(\underline{\mathrm{P}}_j, \underline{\mathrm{P}}_j^* : \ j = 0,1,2,\ldots\right)\right) = \begin{cases} \frac{1}{1+y}, & \text{if there is same number of } \underline{\mathrm{P}}_j \text{ and } \underline{\mathrm{P}}_j^* \text{ for all } j \geq 1 \\ 0, & \text{otherwise.} \end{cases}$$

$$\lim \varphi_n\left(m\left(\bar{\psi}_j, \bar{\psi}_j^* : \ j = 0,1,2,\ldots\right)\right) = \frac{y}{1+y} \lim \frac{1}{p} \operatorname{Tr}\left(m\left(\psi_j, \psi_j^* : \ j = 0,1,2,\ldots\right)\right).$$

Let us define two non-commutative $*$-probability spaces $(\mathcal{A}_1, \varphi_1)$ and $(\mathcal{A}_2, \varphi_2)$ via the above limits.

$$\begin{aligned} \mathcal{A}^{(1)} &= \operatorname{Span}\{c_i, \ c_i^* : c_0 = c_0^*, \ i = 0,1,2,\ldots\}, \\ \mathcal{A}^{(2)} &= \operatorname{Span}\{a_i, \ a_i^* : i = 0,1,2,\ldots\}, \end{aligned}$$

$$\varphi_1(m(c_i, c_i^*, : \ i = 0,1,2,\ldots)) = \begin{cases} \frac{1}{1+y}, & \text{if there is same number of } c_i \text{ and } c_i^* \text{ for all } i \geq 1 \\ 0, & \text{otherwise,} \end{cases}$$

$$\varphi_2(m(a_i, a_i^* : \ i = 0,1,2,\ldots)) = \frac{y}{1+y} \lim \left\{\frac{1}{p} \operatorname{Tr}\left(m\left(\psi_j, \psi_j^* : \ j = 0,1,2,\ldots\right)\right)\right\}.$$

Then by Assumption $(A10)$, $\left(\mathcal{M}_n^{(2)}, \varphi_n\right)$ converges to $\left(\mathcal{A}^{(2)}, \varphi_2\right)$ (see Definition 5.5). Thus $\{c_i, c_i^* : i \geq 0\}$ and $\{a_i, a_i^* : i \geq 0\}$ are respectively the limits of $\{\underline{\mathrm{P}}_i, \underline{\mathrm{P}}_i^* : i \geq 0\}$ and $\{\bar{\psi}_j, \bar{\psi}_j^* : j \geq 0\}$ respectively. For any $k \geq 1$, s^k is the limit of $((n+p)^{-1/2}W)^k$ in $\operatorname{Span}\{(n+p)^{-1/2}W\}$.

Now, to handle joint monomials, by Lemma 5.3, we know that, $\operatorname{Span}\{(n+p)^{-1/2}W\}$ and $\operatorname{Span}\left(\mathcal{M}_n^{(1)} \cup \mathcal{M}_n^{(2)}\right)$ are asymptotically freely independent. Let us define

$$\mathcal{M}_n^{(3)} = \operatorname{Span}\left\{(n+p)^{-1}W\underline{\mathrm{P}}_j W, \ \ (n+p)^{-1}W\underline{\mathrm{P}}_j^* W : j = 0,1,2,\ldots\right\}. \tag{4.6}$$

Then by Lemma 5.4, $\mathcal{M}_n^{(2)}$ and $\mathcal{M}_n^{(3)}$ are asymptotically freely independent. This yields a subtle observation: *even though $\mathcal{M}_n^{(1)}$ and $\mathcal{M}_n^{(2)}$ are not asymptotically freely independent, to compute the limits of φ_n for different elements from $\mathcal{A}_n^{\Delta}$, defined in (4.2), we can assume that they are asymptotically freely independent.* As a consequence, the limit of $(\mathcal{M}_n, \varphi_n)$ is the free product of $\operatorname{Span}\{s\}$, $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$, which is nothing but $(\mathcal{A}, \varphi)$ defined in (2.6). Therefore $(\mathcal{A}_n^{\Delta}, \varphi_n)$ converges to $(\mathcal{B}_q, \varphi)$. Hence, the proof will be complete if we can show that for any monomial m,

$$\lim\left(\varphi_n(m(\hat{\Gamma}_i, \hat{\Gamma}_i^* : i \geq 0)) - \varphi_n(m(\Delta_i, \Delta_i^* : i \geq 0))\right) = 0 \tag{4.7}$$

Note that, when $X_t \sim MA(q)$ process, we can write Model (1.1) as

$$X_{t,p}^{(n)} = (\psi_{0,p}^{(n)} \ \ \psi_{1,p}^{(n)} \ \ \psi_{2,p}^{(n)} \ldots \psi_{q,p}^{(n)})(\varepsilon_{t,p}^* \ \ \varepsilon_{t-1,p}^* \ \ \varepsilon_{t-2,p}^* \ldots \varepsilon_{t-q,p}^*)^* \ \ \forall t, n \geq 1.$$

Therefore, by (1.2), the sample autocovariance matrix of order k is given by

$$\begin{aligned}
n\hat{\Gamma}_{k,p} &= \sum_{t=k+1}^{n} (\psi_{0,p}^{(n)}\ \psi_{1,p}^{(n)}\ \ldots \psi_{q,p}^{(n)})(\varepsilon_{t,p}^{*}\ \ldots \varepsilon_{t-q,p}^{*})^{*}(\varepsilon_{t-k,p}^{*}\ \ldots \varepsilon_{t-k-q,p}^{*})(\psi_{0,p}^{(n)}\ \psi_{1,p}^{(n)}\ \ldots \psi_{q,p}^{(n)})^{*} \\
&= \sum_{j=0}^{q} \sum_{j'=0}^{q} \psi_{j,p}^{(n)}\, \hat{\Gamma}_{j'-j+k}(\varepsilon)\, \psi_{j',p}^{(n)*} - \sum_{j=0}^{q} \sum_{\substack{j'=0 \\ j-j'\neq k}}^{q} \psi_{j,p}^{(n)} \left(\sum_{t=n-j+1}^{n} \varepsilon_{t,p}\varepsilon_{t-(j'+k-j)}^{*} \right) \psi_{j',p}^{(n)*} \\
&\quad + \sum_{j=0}^{q} \sum_{\substack{j'=0 \\ j-j'\neq k}}^{q} \psi_{j,p}^{(n)} \left(\sum_{t=k-j+1}^{j'+k-j} \varepsilon_{t,p}\varepsilon_{t-(j'+k-j)}^{*} \right) \psi_{j',p}^{(n)^{*}} + \sum_{j=0}^{q} \psi_{j,p}^{(n)} \left(\sum_{t=n-j+1}^{n} \varepsilon_{t,p}\varepsilon_{t,p}^{*} \right) \psi_{j-k,p}^{(n)*} \\
&\quad + \sum_{j=0}^{q} \psi_{j,p}^{(n)} \left(\sum_{t=k-j+1}^{0} \varepsilon_{t,p}\varepsilon_{t,p}^{*} \right) \psi_{j-k,p}^{(n)*} \\
&= \Delta_k + R_{1n} + R_{2n} + R_{3n} + R_{4n},\ \text{(say)}. \qquad (4.8)
\end{aligned}$$

Let $\|\cdot\|$ be the operator norm. Then we have

$$\lim E\|n^{-1}R_{1n}\| \leq \lim \sum_{j=0}^{q} \sum_{\substack{j'=0 \\ j-j'\neq k}}^{q} \|\psi_{j,p}^{(n)}\| \left(n^{-1} \sum_{t=n-j+1}^{n} E\|\varepsilon_{t,p}\varepsilon_{t-(j'+k-j)}^{*}\| \right) \|\psi_{j',p}^{(n)^{*}}\| = 0.$$

Similarly, $\lim E\|n^{-1}R_{2n}\| = 0$. Also, by Lemma 5.3, we have for all $r \geq 1$,

$$\begin{aligned}
\lim E\frac{1}{p}\operatorname{Tr}(n^{-1}R_{3n})^{r} &= n^{-r} \sum_{\substack{j_u=0 \\ u=1(1)r}}^{q} \sum_{\substack{t_u=n-j_u+1 \\ u=1(1)r}}^{n} E\frac{1}{p}\operatorname{Tr}(\Pi_{u=1}^{r}\psi_{j_u,p}^{(n)}\varepsilon_{t_u,p}\varepsilon_{t_u,p}^{*}\psi_{j_u-k,p}^{(n)*}) \\
&= \lim n^{-r}O(p^{r-1}) = 0.
\end{aligned}$$

Similarly, $\lim E\frac{1}{p}\operatorname{Tr}(n^{-1}R_{4n})^{r} = 0\ \forall r$. Note that $(m(\hat{\Gamma}_i, \hat{\Gamma}_i^{*} : i \geq 0) - m(\Delta_i, \Delta_i^{*} : i \geq 0))$ involves monomials with at least one of R_{1n} R_{2n}, R_{3n} or R_{4n}. Hence, (4.7) is proved and the proof of Theorem 2.1 is complete.

4.2 Proof of Corollary 2.1

(a) We use Lemma 5.1 to prove the *existence* of LSD of any symmetric polynomials $\Pi(\hat{\Gamma}_i, \hat{\Gamma}_i^{*} : i \geq 0)$ of $\{\hat{\Gamma}_i, \hat{\Gamma}_i^{*} : i \geq 0\}$. As discussed in the paragraph after Corollary 2.1, (M1) of Lemma 5.1 automatically follows from Theorem 2.1. We only need to establish (M4) and (C).

Proof of Carleman's condition (C)**:** Consider any symmetric polynomial $\Pi(\hat{\Gamma}_i, \hat{\Gamma}_i^{*} : i \geq 0) = \sum_i t_i m_{k_i}$ (say), where t_i's are constants. Now, by Remark 2.2 and Assumption (A10), it is easy to show that, there exists a $K > 1$ such that

$$\varphi(\Pi(\hat{\Gamma}_i, \hat{\Gamma}_i^{*} : i \geq 0)^{2h} \leq K^{2h} C_{2h\max\{k_i : i \geq 1\}},$$

where $C_r = \dfrac{(2r)!}{(r)!(r+1)!}\ \forall r = 1, 2, \ldots.$

For example, consider

$$\begin{aligned}
\lim E\left(p^{-1}\operatorname{Tr}(\hat{\Gamma}_0)^{2h}\right) &= \frac{(1+y)^{2h+1}}{y}\varphi\left(\sum_{j=0}^{q}\sum_{j'=0}^{q} a_j s c_{j-j'} s a_{j'}^{*}\right)^{2h} \\
&= \frac{(1+y)^{2h+1}}{y}\sum_{\substack{j_r : r=1,2,\ldots,2h \\ j'_r : r=1,2,\ldots,2h}} \varphi\left(a_{j_1} s c_{j_1-j'_1} s a_{j_1}^{*} a_{j_2} s c_{j_2-j'_2} s a_{j_2}^{*} \ldots a_{j'_{2h}}^{*}\right) \\
&= \frac{(1+y)^{2h+1}}{y}\sum_{\substack{j_r : r=1,2,\ldots,2h \\ j'_r : r=1,2,\ldots,2h}} \sum_{\pi \in NC_2(4h)} \varphi_{k(\pi)}\left[c_{j_1-j'_1}, a_{j_1}^{*} a_{j_2}, c_{j_2-j'_2}, \ldots, a_{j'_{2h}}^{*} a_{j_1}\right] \\
&= \frac{(1+y)^{2h+1}}{y}\sum_{\substack{j_r : r=1,2,\ldots,2h \\ j'_r : r=1,2,\ldots,2h \\ \pi \in NC_2(2h)}} \varphi_{\pi}(c_{j_1-j'_1}, \ldots, c_{j_{2h}-j'_{2h}})\varphi_{k(\pi)}(a_{j_1}^{*} a_{j_2}, \ldots, a_{j'_{2h}}^{*} a_{j_1}) \\
&\leq K^{2h} C_{2h}, \qquad \text{for some } K > 1, \text{ by Assumption } (A10)\ .
\end{aligned}$$

Now note that $K^{2h} C_{2h \max\{k_i : i \geq 1\}}$ satisfies Carleman's condition and hence Condition (C) is established.

Now it remains to show $(M4)$ to conclude almost sure convergence. Incidentally, the $(M2)$ condition, which implies probability convergence, is easier to verify. We omit the detailed argument of $(M4)$ and only sketch the main steps. To show $(M2)$, We have to show, for all $h \geq 1$,

$$E\left(\frac{1}{p}\operatorname{Tr}\left(\Pi(\hat{\Gamma}_i, \hat{\Gamma}_i^{*} : \ i \geq 0)\right)^{h}\right)^{2} - E^{2}\left(\frac{1}{p}\operatorname{Tr}\left(\Pi(\hat{\Gamma}_i, \hat{\Gamma}_i^{*} : \ i \geq 0)\right)^{h}\right) \to 0. \tag{4.9}$$

Let m_k be any monomial of order k i.e. it involves k many $\hat{\Gamma}_i$ or $\hat{\Gamma}_i^{*}$, $i \geq 0$. It is enough to show that

$$E(p^{-1}\operatorname{Tr}(m_{k_1}) p^{-1}\operatorname{Tr}(m_{k_2})) - E(p^{-1}\operatorname{Tr}(m_{k_1}))E(p^{-1}\operatorname{Tr}(m_{k_2})) \to 0. \tag{4.10}$$

Note that $E(p^{-1}\operatorname{Tr}(m_{k_1}) p^{-1}\operatorname{Tr}(m_{k_2}))$ involves $(2k_1 + 2k_2)$ many Z and Z^{*} and in the denominator, we have $p^2 n^{k_1+k_2}$. Also, by Remark 2.2, only non-crossing partitions of $\{Z, Z^{*}\}$ will contribute. By Property 4 of Kreweras complement, the non-crossing partitions for $\{Z, Z^{*}\}$, which join $\{1, 2, \ldots, 2k_1\}$ and $\{2k_1 + 1, 2k_1 + 2, \ldots, 2k_1 + 2k_2\}$ contribute $O(n^{-1})$. For example, consider the following string of equalities which are valid upto $o(1)$ terms:

$$E(p^{-1}\operatorname{Tr}(\hat{\Gamma}_0))^2 = \frac{1}{n^2 p^2} E\left[\sum_{j=0}^{q}\sum_{j'=0}^{q}\operatorname{Tr}\left(\psi_j Z P_{j-j'} Z^{*} \psi_{j'}^{*}\right)\right]^2$$

$$
\begin{aligned}
&= \frac{1}{n^2p^2}E\left[\sum_{j,j'=0}^{q}\sum_{k,k'=0}^{q}\operatorname{Tr}\left(\psi_j Z P_{j-j'} Z^*\psi_{j'}^*\right)\operatorname{Tr}\left(\psi_k Z P_{k-k'} Z^*\psi_{k'}^*\right)\right]\\
&= \frac{1}{n^2p^2}E\left[\sum_{\substack{1\le j,j',\\ k,k'\le q}}\ \sum_{\substack{1\le i_1,\ldots,i_6\le p\\ 1\le t_1,\ldots,t_4\le n}}\left\{(\psi_j)_{i_1i_2}Z_{i_2t_1}(P_{j-j'})_{t_1t_2}Z_{i_3t_2}(\psi_{j'})_{i_1i_3}(\psi_k)_{i_4i_5}Z_{i_5t_3}(P_{k-k'})_{t_3t_4}Z_{i_6t_4}(\psi_{k'})_{i_4i_6}\right\}\right].
\end{aligned}
$$

Consider $i_2 = i_6, t_1 = t_4, i_3 = i_5$ and $t_2 = t_3$. Note that this contributes

$$\frac{1}{np}\sum_{j=0}^{q}\sum_{k=0}^{q}\sum_{j'=0}^{q}\sum_{k'=0}^{q}\left\{\frac{1}{p}\operatorname{Tr}(\psi_j\psi_k^*, \psi_k\psi_{j'}^*)\right\}\left\{\frac{1}{n}\operatorname{Tr}(P_{j-j'}P_{k-k'})\right\} = O(n^{-1}).$$

Moreover, by Theorem 2.1, the remaining non-crossing partitions of $\{Z, Z^*\}$ contribute same as that of $E(p^{-1}\operatorname{Tr}(m_{k_1}))E(p^{-1}\operatorname{Tr}(m_{k_2}))$. Hence, (4.10) is proved. As a consequence (4.9) and hence $(M2)$ is established.

Now we sketch the argument for $(M4)$. Using the same arguments as in the proof of $(M2)$, it can be shown that

$$E\left(\frac{1}{p}\operatorname{Tr}\left(\Pi(\hat\Gamma_i, \hat\Gamma_i^* \,:\, i\ge 0)\right)^h\right)^r = E^r\left(\frac{1}{p}\operatorname{Tr}\left(\Pi(\hat\Gamma_i, \hat\Gamma_i^* \,:\, i\ge 0)\right)^h\right) + O(n^{-r+1})\ \ \forall r \ge 1.$$

Therefore, $E\left[\frac{1}{p}\operatorname{Tr}\left(\Pi(\hat\Gamma_i, \hat\Gamma_i^* \,:\, i\ge 0)\right)^h - E\left(\frac{1}{p}\operatorname{Tr}\left(\Pi(\hat\Gamma_i, \hat\Gamma_i^* \,:\, i\ge 0)\right)^h\right)\right]^4 = O(n^{-3})$ and hence $(M4)$ is established.

The proof of Corollary 2.1(a) is complete.

(b) Suppose $X_t \sim$ MA(0) process and Assumptions $(A1)$, $(A2)$, $(A9)$ $(A10)$ hold. Then the existence of the LSD of $\hat\Gamma_0$, under Assumption $(A5)$, is well known (see for example Bai & Silverstein (2009)). Jin *et al.* (2014) established the existence of the LSD of $\{\hat\Gamma_i + \hat\Gamma_i^*\}_{i\ge 1}$ under Assumption $(A6)$ and $(A7)$.

Now let $X_t \sim$ MA(q), $q \ge 1$ process and suppose Assumptions $(A1)$, $(A2)$, $(A6)$, $(A7)$ and $(A9)$ hold. Here we will establish the existence of the LSD of $\{\hat\Gamma_i + \hat\Gamma_i^*\}_{i\ge 0}$. Let

$$
\begin{aligned}
\tilde\varepsilon_{ti} &= \varepsilon_{ti}I(|\varepsilon_{ti}| < \eta_n n^{\frac{1}{2+\delta}}),\ \hat\varepsilon_{ti} = \tilde\varepsilon_{ti} - E(\tilde\varepsilon_{ti}),\ \forall t, i \text{ and some } \eta_n \downarrow 0,\\
\sigma_{ti}^2 &= E|\hat\varepsilon_{ti}|^2,\ \Delta = n^{-\frac{\delta}{4+2\delta}}, X_{ti} = 2Ber(0.5) - 1,\ \text{i.i.d. for all } t, i,\\
\bar\varepsilon_{ti} &= \begin{cases} X_{ti}, \text{ if } \sigma_{ti}^2 < 1 - \Delta,\\ \frac{\hat\varepsilon_{ti}}{\sigma_{ti}}, \text{ otherwise,}\end{cases}\\
\hat\Gamma_i(\varepsilon), \tilde\Gamma_i(\varepsilon), \hat{\hat\Gamma}_i(\varepsilon), \bar\Gamma_i(\varepsilon) &= i\text{-th order sample autocovariance matrix of } \{\varepsilon_{ti}\}, \{\tilde\varepsilon_{ti}\}, \{\hat\varepsilon_{ti}\}, \{\bar\varepsilon_{ti}\}\\
&\quad \text{(respectively)}\\
\hat T_i &= \sum_{j,j'=0}^{q}\psi_j\hat\Gamma_{j-j'+i}(\varepsilon)\psi_{j'}^*,\ \tilde T_i = \sum_{j,j'=0}^{q}\psi_j\tilde\Gamma_{j-j'+i}(\varepsilon)\psi_{j'}^*,
\end{aligned}
$$

$$\hat{\hat{T}}_i \;=\; \sum_{j,j'=0}^{q} \psi_j \hat{\hat{\Gamma}}_{j-j'+i}(\varepsilon)\psi^*_{j'},\;\; \bar{T}_i = \sum_{j,j'=0}^{q} \psi_j \bar{\Gamma}_{j-j'+i}(\varepsilon)\psi^*_{j'}.$$

Let F^A denote the ESD of the matrix A and L denote the Lévy metric on the space of probability distribution functions. Existence of the LSD of $\{\bar{T}_i + \bar{T}_i^*\}_{i\geq 0}$ follows by Remark 2.1 and Corollary 2.1. We will actually show that the LSD of $\{\hat{\Gamma}_i + \hat{\Gamma}_i^*\}_{i\geq 0}$ is same as that of $\{\bar{T}_i + \bar{T}_i^*\}_{i\geq 0}$ by showing that $L(F^{\hat{\Gamma}_i+\hat{\Gamma}_i^*}, F^{\bar{T}_i+\bar{T}_i^*}) \to 0\ \forall i \geq 0$ a.s.. Note that

$$\begin{aligned} L(F^{\hat{\Gamma}_i+\hat{\Gamma}_i^*}, F^{\bar{T}_i+\bar{T}_i^*}) \;&\leq\; L(F^{\hat{\Gamma}_i+\hat{\Gamma}_i^*}, F^{\hat{T}_i+\hat{T}_i^*}) + L(F^{\hat{T}_i+\hat{T}_i^*}, F^{\tilde{T}_i+\tilde{T}_i^*}) \\ &\quad + L(F^{\tilde{T}_i+\tilde{T}_i^*}, F^{\hat{\hat{T}}_i+\hat{\hat{T}}_i^*}) + L(F^{\hat{\hat{T}}_i+\hat{\hat{T}}_i^*}, F^{\bar{T}_i+\bar{T}_i^*}). \end{aligned} \tag{4.11}$$

We will show that each of the terms in the right side of (4.11) tends to 0 almost surely.

Proof of $L(F^{\hat{\Gamma}_i+\hat{\Gamma}_i^*}, F^{\hat{T}_i+\hat{T}_i^*}) \to 0$**:** By Theorem A.43 and Lemma B.18 in Bai & Silverstein (2009), we have

$$\begin{aligned} L(F^{\hat{\Gamma}_i+\hat{\Gamma}_i^*}, F^{\hat{T}_i+\hat{T}_i^*}) \;&\leq\; 2p^{-1}\left(\mathrm{rank}(R_{1n}) + \mathrm{rank}(R_{2n}) + \mathrm{rank}(R_{3n}) + \mathrm{rank}(R_{4n})\right) \\ &\leq\; \frac{8q}{p} \to 0 \text{ a.s..} \end{aligned} \tag{4.12}$$

where R_{1n}, R_{2n}, R_{3n} and R_{4n} are as in (4.8).

Proof of $L(F^{\hat{T}_i+\hat{T}_i^*}, F^{\tilde{T}_i+\tilde{T}_i^*}) \to 0$**:** By Theorem A.43 and Lemma B.18 in Bai & Silverstein (2009), we have for some $C > 0$

$$\begin{aligned} L(F^{\hat{T}_i+\hat{T}_i^*}, F^{\tilde{T}_i+\tilde{T}_i^*}) \;&\leq\; \frac{1}{p}\mathrm{rank}(\hat{T}_i + \hat{T}_i^* - \tilde{T}_i - \tilde{T}_i^*) \leq \frac{2}{p}\mathrm{rank}(\hat{T}_i - \tilde{T}_i) \\ &\leq\; \frac{1}{p}\mathrm{rank}\left(\sum_{j,j'=0}^{q} \psi_j \left(\hat{\Gamma}_{j-j'+i}(\varepsilon) - \tilde{\Gamma}_{j-j'+i}(\varepsilon)\right)\psi^*_{j'}\right) \\ &\leq\; \frac{C}{p}\mathrm{rank}(\hat{\Gamma}_i(\varepsilon) - \tilde{\Gamma}_i(\varepsilon)) \to 0 \text{ a.s.} \end{aligned} \tag{4.13}$$

(see page 1210 in Jin *et al.* (2014)).

Proof of $L(F^{\tilde{T}_i+\tilde{T}_i^*}, F^{\hat{\hat{T}}_i+\hat{\hat{T}}_i^*}) \to 0$**:** By Theorem A.45 in Bai & Silverstein (2009), we have for some $C > 0$,

$$L(F^{\tilde{T}_i+\tilde{T}_i^*}, F^{\hat{\hat{T}}_i+\hat{\hat{T}}_i^*}) \;\leq\; \|\tilde{T}_i + \tilde{T}_i^* - \hat{\hat{T}}_i - \hat{\hat{T}}_i^*\| \leq C\|\tilde{\Gamma}_i(\varepsilon) - \hat{\hat{\Gamma}}_i(\varepsilon)\| \to 0 \text{ a.s.} \tag{4.14}$$

(see page 1211 in Jin *et al.* (2014)).

Proof of $L(F^{\hat{\hat{T}}_i+\hat{\hat{T}}_i^*}, F^{\bar{T}_i+\bar{T}_i^*}) \to 0$**:** By Corollary A.41 in Bai & Silverstein (2009), we have

$$L^3(F^{\hat{\hat{T}}_i+\hat{\hat{T}}_i^*}, F^{\bar{T}_i+\bar{T}_i^*}) \;\leq\; \frac{1}{p}\,\mathrm{Tr}\left((\hat{\hat{T}}_i + \hat{\hat{T}}_i^* - \bar{T}_i - \bar{T}_i^*)(\hat{\hat{T}}_i + \hat{\hat{T}}_i^* - \bar{T}_i - \bar{T}_i^*)^*\right)$$

$$
\begin{aligned}
&\leq \frac{4}{p}\operatorname{Tr}\Big((\hat{\hat{T}}_i - \bar{T}_i)(\hat{\hat{T}}_i - \bar{T}_i)^*\Big)\\
&= \frac{4}{p}\sum_{j,j',k,k'=0}^{q}\operatorname{Tr}\Big(\psi_j\Big(\hat{\hat{\Gamma}}_{j-j'+i}(\varepsilon) - \bar{\Gamma}_{j-j'+i}(\varepsilon)\Big)\psi^*_{j'}\\
&\qquad\qquad \psi_{k'}\Big(\hat{\hat{\Gamma}}_{k-k'+i}(\varepsilon) - \bar{\Gamma}_{k-k'+i}(\varepsilon)\Big)^*\psi^*_k\Big).
\end{aligned}
$$

Therefore, it is enough to show that

$$
p^{-1}\operatorname{Tr}(A(\hat{\hat{\Gamma}}_i(\varepsilon) - \bar{\Gamma}_i(\varepsilon))BB^*(\hat{\hat{\Gamma}}_i(\varepsilon) - \bar{\Gamma}_i(\varepsilon))^*A^*) \to 0, \tag{4.15}
$$

for any $A, B \in \operatorname{Span}\{\psi_j, \psi^*_j : j \geq 0\}$. The proof of (4.15) given below goes along the same lines as the proof of $p^{-1}\operatorname{Tr}((\hat{\hat{\Gamma}}_i(\varepsilon) - \bar{\Gamma}_i(\varepsilon))(\hat{\hat{\Gamma}}_i(\varepsilon) - \bar{\Gamma}_i(\varepsilon))^*) \to 0$ given in Jin *et al.* (2014). In our case we have the extra factors of A, B etc. Let

$$
\begin{aligned}
\hat{\alpha}_k &= (2n)^{-1/2}(\hat{\varepsilon}_{k1}, \hat{\varepsilon}_{k2}, \ldots, \hat{\varepsilon}_{kp})^T, & \bar{\alpha}_k &= (2n)^{-1/2}(\bar{\varepsilon}_{k1}, \bar{\varepsilon}_{k2}, \ldots, \bar{\varepsilon}_{kp})^T,\\
\hat{U} &= (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{n-i}), & \bar{U} &= (\bar{\alpha}_1, \bar{\alpha}_2, \ldots, \bar{\alpha}_{n-i}),\\
\hat{V} &= (\hat{\alpha}_{1+i}, \hat{\alpha}_{2+i}, \ldots, \hat{\alpha}_n), & \bar{V} &= (\bar{\alpha}_{1+i}, \bar{\alpha}_{2+i}, \ldots, \bar{\alpha}_n).
\end{aligned}
$$

Then,

$$
\begin{aligned}
&p^{-1}\operatorname{Tr}(A(\hat{\hat{\Gamma}}_i - \bar{\Gamma}_i)BB^*(\hat{\hat{\Gamma}}_i - \bar{\Gamma}_i)^*A^*)\\
=\;& p^{-1}\operatorname{Tr}(A(\hat{U}\hat{V}^* - \bar{U}\bar{V}^*)BB^*(\hat{U}\hat{V}^* - \bar{U}\bar{V}^*)^*A^*)\\
=\;& p^{-1}\operatorname{Tr}(A((\hat{U} - \bar{U})\hat{V}^* + \bar{U}(\hat{V} - \bar{V})^*)BB^*((\hat{U} - \bar{U})\hat{V}^* + \bar{U}(\hat{V} - \bar{V})^*)^*A^*)\\
\leq\;& 2p^{-1}\operatorname{Tr}(A(\hat{U} - \bar{U})\hat{V}^*BB^*\hat{V}(\hat{U} - \bar{U})^*A^*)\\
&+2p^{-1}\operatorname{Tr}(A\bar{U}(\hat{V} - \bar{V})^*BB^*(\hat{V} - \bar{V})\bar{U}^*A^*).
\end{aligned}
$$

Now, we have for some $C > 0$, with $A = ((a_{ij}))$ and $B = ((b_{ij}))$,

$$
\begin{aligned}
&p^{-1}\operatorname{Tr}(A(\hat{U} - \bar{U})\hat{V}^*BB^*\hat{V}(\hat{U} - \bar{U})^*A^*)\\
\leq\;& \frac{C}{n^3}\sum_{u,v}|\sum_{l,k,j} a_{ul}(\hat{\varepsilon}_{kl} - \bar{\varepsilon}_{kl})\hat{\varepsilon}^*_{(k+i)j}b_{jv}|^2\\
=\;& \frac{C}{n^3}\sum_{u,v}\sum_{l_1,k_1,j_1}\sum_{l_2,k_2,j_2}\Big(a_{ul_1}(\hat{\varepsilon}_{k_1l_1} - \bar{\varepsilon}_{k_1l_1})\hat{\varepsilon}^*_{(k_1+i)j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k_2+i)j_2}(\hat{\varepsilon}_{k_2l_2} - \bar{\varepsilon}_{k_2l_2})^*a^*_{ul_2}\Big)\\
=\;& J_1 + J_2 + J_3 + J_4 + J_5,
\end{aligned}
$$

where,

$$
\begin{aligned}
J_1 &= \frac{C}{n^3}\sum_{u,v}\sum_{\substack{l_1,l_2,j_1,j_2\\ k_1>k_2,\ k_1\neq k_2+i}}\Big(a_{ul_1}(\hat{\varepsilon}_{k_1l_1} - \bar{\varepsilon}_{k_1l_1})\hat{\varepsilon}^*_{(k_1+i)j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k_2+i)j_2}(\hat{\varepsilon}_{k_2l_2} - \bar{\varepsilon}_{k_2l_2})^*a^*_{ul_2}\Big),\\
J_2 &= \frac{C}{n^3}\sum_{u,v}\sum_{\substack{l_1,j_1,l_2,\\ j_2,k_2}}\Big(a_{ul_1}(\hat{\varepsilon}_{(k_2+i)l_1} - \bar{\varepsilon}_{(k_2+i)l_1})\hat{\varepsilon}^*_{(k_2+2i)j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k_2+i)j_2}(\hat{\varepsilon}_{k_2l_2} - \bar{\varepsilon}_{k_2l_2})^*a^*_{ul_2}\Big),
\end{aligned}
$$

$$\begin{aligned}
J_3 &= \frac{C}{n^3}\sum_{u,v}\sum_{\substack{l_1,l_2,j_1,j_2\\ k_2>k_1,\ k_2\neq k_1+i}} \left(a_{ul_1}(\hat{\varepsilon}_{k_1l_1}-\bar{\varepsilon}_{k_1l_1})\hat{\varepsilon}^*_{(k_1+i)j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k_2+i)j_2}(\hat{\varepsilon}_{k_2l_2}-\bar{\varepsilon}_{k_2l_2})^*a^*_{ul_2}\right),\\
J_4 &= \frac{C}{n^3}\sum_{u,v}\sum_{\substack{l_1,j_1,l_2,\\ j_2,k_1}} \left(a_{ul_1}(\hat{\varepsilon}_{k_1l_1}-\bar{\varepsilon}_{k_1l_1})\hat{\varepsilon}^*_{(k_1+i)j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k_1+2i)j_2}(\hat{\varepsilon}_{(k_1+i)l_2}-\bar{\varepsilon}_{(k_1+i)l_2})^*a^*_{ul_2}\right),\\
J_5 &= \frac{C}{n^3}\sum_{\substack{u,v,l_1,l_2\\ j_1,j_2,k}} \left[a_{ul_1}(\hat{\varepsilon}_{kl_1}-\bar{\varepsilon}_{kl_1})\hat{\varepsilon}^*_{(k+i)j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k+i)j_2}(\hat{\varepsilon}_{kl_2}-\bar{\varepsilon}_{kl_2})^*a^*_{ul_2}\right].
\end{aligned}$$

Note that $E(J_1)=E(J_2)=E(J_3)=E(J_4)=0$. Moreover for some $C_1,C_2,C_3>0$,

$$\begin{aligned}
\operatorname{Var}(J_1) &= E(J_2)^2 \le \frac{C_1}{n^6}\sum_{u_1,v_1}\sum_{\substack{l_1,l_2,j_1,j_2\\ k_1>k_2,\ k_1\neq k_2+i}}\sum_{u_2,v_2}\sum_{\substack{l_3,l_4,j_3,j_4\\ k_3>k_4,\ k_3\neq k_4+i}} E\Big[a_{u_1l_1}(\hat{\varepsilon}_{k_1l_1}-\bar{\varepsilon}_{k_1l_1})\hat{\varepsilon}^*_{(k_1+i)j_1}b_{j_1v_1}b^*_{j_2v_1}\\
&\qquad \hat{\varepsilon}_{(k_2+i)j_2}(\hat{\varepsilon}_{k_2l_2}-\bar{\varepsilon}_{k_2l_2})^*a^*_{u_1l_2}a_{u_2l_3}(\hat{\varepsilon}_{k_3l_3}-\bar{\varepsilon}_{k_3l_3})\hat{\varepsilon}^*_{(k_3+i)j_3}b_{j_3v_2}b^*_{j_4v_2}\\
&\qquad \hat{\varepsilon}_{(k_4+i)j_4}(\hat{\varepsilon}_{k_4l_4}-\bar{\varepsilon}_{k_4l_4})^*a^*_{u_2l_4}\Big]\\
&\le \frac{C_2}{n^4}\sum_{\substack{u_1,u_2\\ v_1,v_2}}\sum_{\substack{l_1,l_2\\ j_1,j_2}}\left(a_{u_1l_1}b_{j_1v_1}b^*_{j_2v_1}a^*_{u_1l_2}a_{u_2l_1}b_{j_1v_2}b^*_{j_2v_2}a^*_{u_2l_2}\right)\\
&\le \frac{C_3}{n^2}(n^{-1}\operatorname{Tr}(A^2A^{*2}))(n^{-1}\operatorname{Tr}(B^2B^{*2})) = O(n^{-2}).
\end{aligned}$$

Also for some $C_1,C_2>0$,

$$\begin{aligned}
\operatorname{Var}(J_2) &= E(J_2)^2 \le \frac{C_1}{n^6}\sum_{\substack{u_1,v_1\\ u_2,v_2}}\sum_{\substack{l_1,j_1,l_2,\\ j_2,k_2}}\sum_{\substack{l_3,j_3,l_4,\\ j_4,k_4}} E\Big[a_{u_1l_1}(\hat{\varepsilon}_{(k_2+i)l_1}-\bar{\varepsilon}_{(k_2+i)l_1})\hat{\varepsilon}^*_{(k_2+2i)j_1}b_{j_1v_1}b^*_{j_2v_1}\\
&\qquad \hat{\varepsilon}_{(k_2+i)j_2}(\hat{\varepsilon}_{k_2l_2}-\bar{\varepsilon}_{k_2l_2})^*a^*_{u_1l_2}a_{u_2l_3}(\hat{\varepsilon}_{(k_4+i)l_3}-\bar{\varepsilon}_{(k_4+i)l_3})\hat{\varepsilon}^*_{(k_4+2i)j_3}\\
&\qquad b_{j_3v_2}b^*_{j_4v_2}\hat{\varepsilon}_{(k_4+i)j_4}(\hat{\varepsilon}_{k_4l_4}-\bar{\varepsilon}_{k_4l_4})^*a^*_{u_2l_4}\Big]\\
&\le \frac{C_2}{n^4}\sum_{\substack{u_1,v_1\\ u_2,v_2}}\sum_{\substack{l_1,j_1,l_2,\\ j_2}}\sum_{\substack{l_3,j_3,l_4,\\ j_4}}\left(a_{u_1l_1}b_{j_1v_1}b^*_{j_2v_1}a^*_{u_1l_2}a_{u_2l_3}b_{j_1v_2}b^*_{j_4v_2}a^*_{u_2l_2}\right) = O(n^{-2}).
\end{aligned}$$

Similarly one can show that $\operatorname{Var}(J_3)=O(n^{-2}),\ \operatorname{Var}(J_4)=O(n^{-2})$.

Let $\tilde{\tilde{\varepsilon}}_{ti} = \varepsilon_{ti}I(|\varepsilon_{ti}| > \eta_n n^{\frac{1}{2+\delta}}),\ \forall t,i$. Therefore, as $E(\varepsilon_{ti})=0$, note that $E(\tilde{\varepsilon}_{ti}) = -E(\tilde{\tilde{\varepsilon}}_{ti}),\ \forall t,i$. Also note that

$$1 = \operatorname{Var}(\varepsilon_{ti}) = \operatorname{Var}\left(\tilde{\varepsilon}_{ti}-E(\tilde{\varepsilon}_{ti})+\tilde{\tilde{\varepsilon}}_{ti}-E(\tilde{\tilde{\varepsilon}}_{ti})\right) = \sigma^2_{ti}+\operatorname{Var}(\tilde{\tilde{\varepsilon}}_{ti})+2(E(\tilde{\tilde{\varepsilon}}_{ti}))^2.$$

Therefore, Using $(A6)$, for some $C>0$

$$(1-\sigma^2_{ti}) \le 2E(\tilde{\tilde{\varepsilon}}^2_{ti}) \le 2C(P(|\varepsilon_{ti}| > \eta_n n^{\frac{1}{2+\delta}}))^{\frac{\delta}{2+\delta}} \le 2C\eta_n^{-\delta}n^{-\frac{\delta}{2+\delta}} \tag{4.16}$$

Let $E = \{(t,i) : \ \sigma_{ti}^2 < 1 - \Delta\}$. Then if $(t,i) \notin E$, we have for some $C > 0$ (see last line of page 1214 in Jin *et al.* (2014)),

$$(1 - \sigma_{ti}^{-1})^2 \leq C\eta_n^{-2\delta} n^{-\frac{2\delta}{2+\delta}}. \tag{4.17}$$

Moreover note that if $(t,i) \in E$, then $\frac{1-\sigma_{ti}^2}{\Delta} > 1$. Then by (4.16) and (4.17), we have for some $C_1, C_2 > 0$,

$$\begin{aligned} E(J_5) &= \frac{C}{n^3} \sum_{u,v,l_1,j_1,k} a_{ul_1} E|\hat{\varepsilon}_{kl_1} - \bar{\varepsilon}_{kl_1}|^2 E|\hat{\varepsilon}_{(k+i)j_1}|^2 b_{j_1v} b_{j_1v}^* a_{ul_1}^* \\ &\leq \frac{C_1}{n^3} \sum_{\substack{u,v,j_1 \\ (k,l_1)\in E}} a_{ul_1} b_{j_1v} b_{j_1v}^* a_{ul_1}^* \left(\frac{1 - \sigma_{kl_1}^2}{\Delta} \right) + \frac{C_2}{n^3} \sum_{\substack{u,v,j_1 \\ (k,l_1)\notin E}} a_{ul_1} b_{j_1v} b_{j_1v}^* a_{ul_1}^* \left(1 - \sigma_{kl}^{-1}\right)^2 \\ &= O(n^{\frac{-\delta}{4+2\delta}}) + O(n^{\frac{-2\delta}{2+\delta}}). \end{aligned}$$

Therefore,

$$E(p^{-1} \operatorname{Tr}(A(\hat{U} - \bar{U})\hat{V}^* BB^* \hat{V}(\hat{U} - \bar{U})^* A^*)) \to 0.$$

Similarly one can show that for some $\epsilon > 0$, $\operatorname{Var}(J_5) = O(n^{-1-\epsilon})$ and as a consequence we have

$$\operatorname{Var}(p^{-1} \operatorname{Tr}(A(\hat{U} - \bar{U})\hat{V}^* BB^* \hat{V}(\hat{U} - \bar{U})^* A^*)) = O(n^{-1-\epsilon}).$$

Hence,

$$p^{-1} \operatorname{Tr}(A(\hat{U} - \bar{U})\hat{V}^* BB^* \hat{V}(\hat{U} - \bar{U})^* A^*) \to 0, a.s.. \tag{4.18}$$

Similarly,

$$p^{-1} \operatorname{Tr}(A\bar{U}(\hat{V} - \bar{V})^* BB^* (\hat{V} - \bar{V})\bar{U}^* A^*) \to 0, a.s.. \tag{4.19}$$

Hence by (4.18) and (4.19), (4.15) is proved. Also by (4.12), (4.13), (4.14) and (4.15),

$$L(F^{\hat{\Gamma}_i + \hat{\Gamma}_i^*}, F^{\bar{T}_i + \bar{T}_i^*}) \to 0, a.s..$$

Additionally if we assume $\sup_{t,i} E|\varepsilon_{ti}|^4 < M < \infty$, then the LSD of $\{\hat{\Gamma}_i \hat{\Gamma}_i^*\}_{i \geq 0}$ exists. This follows from the following inequalities.

$$\begin{aligned} L(F^{\hat{\Gamma}_i \hat{\Gamma}_i^*}, F^{\hat{T}_i \hat{T}_i^*}) &\leq p^{-1}\operatorname{rank}(\hat{\Gamma}_i \hat{\Gamma}_i^* - \hat{T}_i \hat{T}_i^*) \\ &\leq p^{-1}\operatorname{rank}((\hat{\Gamma}_i - \hat{T}_i)\hat{\Gamma}_i^* + \hat{T}_i(\hat{\Gamma}_i - \hat{T}_i)^*) \\ &\leq 2p^{-1}\operatorname{rank}(\hat{\Gamma}_i - \hat{T}_i), \\ L(F^{\hat{T}_i \hat{T}_i^*}, F^{\tilde{T}_i \tilde{T}_i^*}) &\leq 2p^{-1}\operatorname{rank}(\hat{T}_i - \tilde{T}_i), \end{aligned}$$

and by Corollary A.41 of Bai & Silverstein (2009)

$$\begin{aligned} L^4(F^{\tilde{T}_i \tilde{T}_i^*}, F^{\hat{\hat{T}}_i \hat{\hat{T}}_i^*}) &\leq 2p^{-2} \operatorname{Tr}(\tilde{T}_i \tilde{T}_i^* + \hat{\hat{T}}_i \hat{\hat{T}}_i^*) \operatorname{Tr}((\tilde{T}_i - \hat{\hat{T}}_i)(\tilde{T}_i - \hat{\hat{T}}_i)^*) \\ &\leq 2p^{-1} \operatorname{Tr}(\tilde{T}_i \tilde{T}_i^* + \hat{\hat{T}}_i \hat{\hat{T}}_i^*)||\tilde{T}_i - \hat{\hat{T}}_i||^2, \\ L^4(F^{\bar{T}_i \bar{T}_i^*}, F^{\hat{\hat{T}}_i \hat{\hat{T}}_i^*}) &\leq 2p^{-2} \operatorname{Tr}(\bar{T}_i \bar{T}_i^* + \hat{\hat{T}}_i \hat{\hat{T}}_i^*) \operatorname{Tr}((\bar{T}_i - \hat{\hat{T}}_i)(\bar{T}_i - \hat{\hat{T}}_i)^*). \end{aligned}$$

We omit the details. This completes the proof of Corollary 2.1(b).

4.3 Proof of Corollary 2.2

Note that by Corollary 2.1(b), it is enough to work under Assumption ($A3$) or ($A4$).

(a) By Theorem 2.1, the limit of $n(n+p)^{-1}\bar{\hat{\Gamma}}_0$ is γ_{00}, where $\gamma_{00} = \gamma^*_{00} = a_0 s c_0 s a_0$. By (2.7) and (2.8),

$$\varphi(a_0^k) = \frac{y}{1+y} \text{ and } \varphi(c_0^k) = \frac{1}{1+y} \text{ for all } k \geq 1.$$

Note that (see page 144 of Nica & Speicher (2006)),

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

Hence, by (2.12) and (2.13), the h-th order moment of the LSD of $\hat{\Gamma}_0$ is given by

$$\begin{aligned}
& \frac{(1+y)^{h+1}}{y} \sum_{\pi \in NC(h)} \varphi_\pi[a_0^2, a_0^2, \ldots, a_0^2]\varphi_{K(\pi)}[c_0, c_0, \ldots, c_0] \\
= & \sum_{k=1}^{h} \#\{\pi \in NC(h) : \pi \text{ has } k \text{ blocks}\}\, y^{k-1} \\
= & \sum_{k=1}^{h} \frac{1}{k}\binom{h-1}{k-1}\binom{h}{k-1} y^{k-1},
\end{aligned}$$

which is the h-th order moment of the Marčenko-Pastur law (see (1.4)). The first equality holds by Property 4 of the Kreweras complement (see Section 5). This proves (a) under either ($A3$) or ($A4$).

(b) Observe that, by Theorem 2.1, the limit of $n^2(n+p)^{-2}\bar{\hat{\Gamma}}_i\bar{\hat{\Gamma}}_i^*$ is $\gamma_{i0}\gamma^*_{i0}$. By (2.9), $\gamma_{i,0} = a_0 s c_i s a_0$ and $\gamma^*_{i,0} = a_0 s c_i^* s a_0$. Since the marginal distribution of all the c_i's are same for $i \geq 1$, using free independence of Span$\{s\}$, Span$\{a_i, a_i^* : i \geq 0\}$ and Span$\{c_i, c_i^* : c_0 = c_0^*, i \geq 0\}$, the LSD of $\left(\frac{n}{p}\right)^2 \hat{\Gamma}_i(\varepsilon)\hat{\Gamma}_i^*(\varepsilon)$ are identical for all $i \geq 1$.

Also, by (2.8), for all $k \geq 1$, $\varphi(a_0^k) = \frac{y}{1+y}$. For any monomial $m(\{c_i, c_i^*\})$, by (2.7), we know $\varphi(m(\{c_i, c_i^*\}))$. Hence using (2.12) and (2.13), the h-th order moment of the LSD of $\left(\frac{n}{p}\right)^2 \hat{\Gamma}_i(\varepsilon)\hat{\Gamma}_i^*(\varepsilon)$ is given by

$$\begin{aligned}
& y^{-2h}\frac{(1+y)^{2h+1}}{y} \sum_{\pi \in NC(2h)} \varphi_{K(\pi)}[a_0, a_0, \ldots, a_0]\varphi_\pi[c_i, c_i^*, c_i, c_i^* \ldots, c_i, c_i^*] \\
= & \; y^{-2h} \sum_{k=1}^{h} \#\{\pi \in NCE(2h) : \pi \text{ has } k \text{ blocks}\}\, y^{2h+1-k-1} \\
= & \sum_{k=1}^{h} \frac{1}{k}\binom{h-1}{k-1}\binom{2h}{k-1} y^{-k}
\end{aligned}$$

where as in Corollary 2.2(a), the first equality holds by Property 4 of the Kreweras complement and the last equality follows from Lemma 4.1 of Edelman (1980). The final expression is indeed the h-th order moment of the Bessel$(2, y^{-1})$ law.

If $p \to \infty$ such that $y = 0$, then the h-th order moment of the LSD of $\left(\frac{n}{p}\right)\hat{\Gamma}_i(\varepsilon)\hat{\Gamma}_i^*(\varepsilon)$ is $\frac{1}{h}\binom{2h}{h-1}$. Hence, the LSD is the Marčenko-Pastur law with parameter 1. This proves (b) under either ($A3$) or ($A4$).

(c) Note that, by Theorem 2.1, the LSD of $\frac{1}{2}(\bar{\hat{\Gamma}}_i + \bar{\hat{\Gamma}}_i^*)$ is same as the distribution of

$$a_0(\sqrt{1+y}\ s)\left(\frac{c_i + c_i^*}{2}\right)(\sqrt{1+y}\ s)\ a_0 = u \text{ (say)},$$

where a_0, Span$\{c_i, c_i^*\}$ and s are freely independent. The marginal distributions of Span$\{c_i, c_i^*\}$ and a_0 are given by (2.8) and (2.7) respectively. Therefore, by (5.11), the S-transform of u is given by,

$$S_u(z) = S_{a_0}(z) S_{(1+y)s^2}(z) S_{\frac{c_i + c_i^*}{2}}(z). \tag{4.20}$$

Moreover, by (5.6), we have

$$m_u(z) = m(z)\frac{y}{1+y} - \frac{1}{z}\frac{1}{1+y}. \tag{4.21}$$

Now, it is known that (see for example Couillet & Debbah (2011)),

$$S_{(1+y)s^2}(z) \quad = \quad \frac{1}{(1+y)(1+z)}. \tag{4.22}$$

Also, by (5.7), (5.9) and (5.10), we have

$$\begin{aligned} \Psi_{a_0}(z) &= \frac{z}{1-z}\frac{y}{1+y} \\ \implies \quad \chi_{a_0}(z) &= \frac{(1+y)z}{y + (1+y)z} \\ \implies \quad S_{a_0}(z) &= \frac{z+1}{z}\chi_{a_0}(z) = \frac{(1+y)(z+1)}{y+(1+y)z}. \end{aligned} \tag{4.23}$$

The distribution of $\frac{c_i + c_i^*}{2}$ is $\frac{1}{1+y}a + \delta_0 \frac{y}{1+y}$, where a has the arcsine law supported on $(-1, 1)$ and δ_0 is a random variable degenerate at 0 (see for example Nica & Speicher (2006)). Also, note that the Stieltjes transformation of a is given by (see for example Nica & Speicher (2006))

$$m_a(z) = \frac{-1}{\sqrt{z^2 - 1}}. \tag{4.24}$$

Therefore, by (5.6), (4.24), (5.8) and (5.9) we have

$$m_{\frac{c_i + c_i^*}{2}}(z) = \frac{-1}{\sqrt{z^2-1}}\frac{1}{1+y} - \frac{y}{1+y}\frac{1}{z}$$

$$
\begin{aligned}
\implies\quad & -\frac{1}{z}m_{\frac{c_i+c_i^*}{2}}(z^{-1}) = \frac{1}{\sqrt{1-z^2}}\frac{1}{1+y}+\frac{y}{1+y}\\
\implies\quad & \frac{1}{\sqrt{1-\chi^2_{\frac{c_i+c_i^*}{2}}(z)}}\frac{1}{1+y} = z+1-\frac{y}{1+y}\\
\implies\quad & \chi^2_{\frac{c_i+c_i^*}{2}}(z) = 1-((z+1)(y+1)-y)^{-2} = 1-(z(y+1)+1)^{-2}.
\end{aligned}
$$

Hence, by (4.20), (4.22), (4.23), (4.25) and (5.10), we have

$$
\begin{aligned}
& \chi_u(z) = \left(1-((z(y+1)+1)^{-2}\right)^{1/2}((1+y)z+y)^{-1}\\
\implies\quad & z = \left(1-((\Psi_u(z)(y+1)+1)^{-2}\right)^{1/2}((1+y)\Psi_u(z)+y)^{-1}\\
\implies\quad & z^2\,((1+y)\Psi_u(z)+y)^2 = \left(1-((\Psi_u(z)(y+1)+1)^{-2}\right)\\
\implies\quad & \left(-(1+y)m_u(z^{-1})-z\right)^2 = 1-\left(-(1+y)m_u(z^{-1})z^{-1}-y\right)^{-2}\\
\implies\quad & \left(-(1+y)m_u(z)-z^{-1}\right)^2 = 1-(-(1+y)m_u(z)z-y)^{-2}\\
\implies\quad & (-yzm(z)+(1-y))^2(1-y^2m^2(z)) = 1, \quad \text{[by (4.21)]}.
\end{aligned}
$$

This agrees with the result of Jin *et al.* (2014).

By Theorem 2.1 of Liu *et al.* (2013), the Stieltjes transformation of LSD of $\frac{1}{2}(\hat{\Gamma}_i+\hat{\Gamma}_i^*)$ satisfies,

$$
m(z) = -\left(z-\frac{1}{2\pi}\int_0^{2\pi}\frac{\cos\theta\,d\theta}{1+ym(z)\cos\theta}\right)^{-1}. \tag{4.25}
$$

Now,

$$
\begin{aligned}
\frac{1}{2\pi}\int_0^{2\pi}\frac{\cos\theta\,d\theta}{1+ym(z)\cos\theta} &= \frac{1}{2\pi ym(z)}\int_0^{2\pi}\frac{(1+ym(z)\cos\theta)-1}{1+ym(z)\cos\theta}d\theta\\
&= \frac{1}{ym(z)}-\frac{1}{2\pi ym(z)}\int_0^{2\pi}\frac{1}{1+ym(z)\cos\theta}d\theta\\
&= \frac{1}{ym(z)}-\frac{1}{2\pi ym(z)}\int_0^{2\pi}\frac{1}{1+(1/2)ym(z)(e^{i\theta}+e^{-i\theta})}d\theta\\
&= \frac{1}{ym(z)}-\frac{1}{2\pi iym(z)}\int_\gamma\frac{1}{1+(1/2)ym(z)(\omega+\omega^{-1})}\frac{d\omega}{\omega}\\
&= \frac{1}{ym(z)}-\frac{2}{2\pi iy^2m^2(z)}\int_\gamma\frac{d\omega}{\omega^2+2(ym(z))^{-1}\omega+1}\\
&= \frac{1}{ym(z)}-\frac{2}{y^2m^2(z)}\frac{1}{\omega_1-\omega_2}
\end{aligned}
$$

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 $(\omega_1-\omega_2)^{-2}=\frac{y^2m^2(z)}{4(1-y^2m^2(z))}$. Also $\gamma=\{e^{i\theta}:0\le\theta<2\pi\}$ i.e. contour of unit circle.

Therefore, by (4.25), we have

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

Therefore, Theorem 2.1 of Liu *et al.* (2013) also agrees with (1.5).

4.4 Proof of Corollary 2.3

Let σ be the LSD of $\bar{\Sigma}_p$. Also let $m(z)$ and $m_a(z)$ be respectively the Stieltjes transformations of LSD of $\Sigma^{1/2}\hat{\Gamma}_0(\varepsilon)\Sigma^{1/2}$ and $a = (1+y)\sigma^{1/2}sc_0s\sigma^{1/2} \in \mathcal{B}_0$. Then by Theorem 2.1 and (5.6),

$$
m_a(z) = m(z)\frac{y}{1+y} + \frac{1}{z(1+y)}. \tag{4.26}
$$

By using (5.11) for free variables repeatedly,

$$
\begin{aligned}
S_a(z) &= S_{sc_0s(1+y)\sigma}(z) \text{ since } a \text{ and } sc_0s(1+y)\sigma \text{ are identically distributed} \\
&= S_{sc_0s}(z)S_{(1+y)\sigma}(z) \text{ since } sc_0s \text{ and } \sigma \text{ are free} \\
&= S_{s^2}(z)S_{c_0}(z)S_{(1+y)\sigma}(z) \text{ since } \varphi \text{ is tracial and } s \text{ and } c_0 \text{ are free.}
\end{aligned} \tag{4.27}
$$

Now, $S_{s^2}(z) = (1+z)^{-1}$ (see for example Burda (2013)) . By (5.7), (5.9) and (5.10), we have

$$
\Psi_{c_0}(z) = \frac{z}{1-z}\frac{1}{1+y}, \quad \chi_{c_0}(z) = \frac{z(1+y)}{1+z(1+y)}, \quad S_{c_0}(z) = \frac{(1+y)(z+1)}{1+z(1+y)}.
$$

Hence, by (4.27), we have

$$
S_a(z) = \frac{1+y}{1+z(1+y)}S_{(1+y)\sigma}(z) \implies \chi_a(z) = \frac{1+y}{1+z(1+y)}\chi_{(1+y)\sigma}(z).
$$

Therefore, using the definition of Ψ and χ transformations (see (5.7), (5.8) and (5.9)), we have

$$
\begin{aligned}
& \frac{1+y}{1+(1+y)\Psi_a(z)}\Big(\chi_{(1+y)\sigma}\big(\Psi_a(z)\big)\Big) = z \\
\Longrightarrow \quad & \Psi_{(1+y)\sigma}\left(\frac{z\big(1+(1+y)\Psi_a(z)\big)}{1+y}\right) = \Psi_a(z)
\end{aligned}
$$

$$\begin{aligned}
&\implies && \Psi_{(1+y)\sigma}\left(\frac{1+(1+y)\Psi_a(z^{-1})}{z(1+y)}\right) = \Psi_a(z^{-1}) \\
&\implies && \Psi_{(1+y)\sigma}\left(-m_a(z) - \frac{y}{z(1+y)}\right) = -zm_a(z) - 1 \\
&\implies && \frac{y}{1+y}\int \frac{-m_a(z)t(1+y) - \frac{ty}{z}}{1+m_a(z)t(1+y)+\frac{ty}{z}} dF_\Sigma(t) = -zm_a(z) - 1 \\
&\implies && \int\left(\frac{m_a(z)ty + y + \frac{ty^2}{z(1+y)}}{yz + zm_a(z)ty(1+y) + \frac{ty}{z}}\right) dF_\Sigma(t) = -m_a(z).
\end{aligned}$$

Substituting (4.26) in the above expression, one can easily see that $m(z)$ satisfies (2.14).

An alternative approach for y=1: For $y = 1$, by (5.20), (2.14) reduces to

$$c(-m(z)) = -zm(z) = \int \frac{dF_\Sigma(t)}{1+tm(z)} = 1 + \sum_{k=1}^{\infty} \varphi(t^k)(-m(z))^k,$$

where $\varphi(t^k) = \int t^k dF_\Sigma(t)$. Thus the k-th order free cumulant of LSD of $\Sigma^{1/2}\hat{\Gamma}_0\Sigma^{1/2}$ is $\varphi(t^k)$ for all $k \geq 1$. Note that the LSD of $\Sigma^{1/2}\hat{\Gamma}_0\Sigma^{1/2}$ and $n^{-1}Z^*\Sigma Z$ are same for $y = 1$, where Z is given by (4.3). By (5.23), the k-th order free cumulant of $n^{-1}Z^*\Sigma Z$ is also given by $\varphi(t^k)$ for all $k \geq 1$.

4.5 Proof of Corollary 2.4

By Corollary 2.1(b), it is enough to work under Assumption ($A3$) or ($A4$). Note that, in the present case, LSD of ψ_j, for all j, is δ_{λ_j}. For any i, let P_i be the $n \times n$ matrix with upper ith diagonal 1. Note that $P_0 = I_n$. We can write

$$n\hat{\Gamma}_i = Z\left(\sum_{j=0}^{q}\sum_{j'=0}^{q} \lambda_j\lambda_{j'}P_{j-j'+i}\right)Z^*, \qquad n\hat{\Gamma}_i^* = Z\left(\sum_{j=0}^{q}\sum_{j'=0}^{q} \lambda_j\lambda_{j'}P^*_{j-j'+i}\right)Z^*.$$

Therefore,

$$\frac{n}{2}(\hat{\Gamma}_i + \hat{\Gamma}_i^*) = Z\left(\frac{1}{2}\sum_{j=0}^{q}\sum_{j'=0}^{q} \lambda_j\lambda_{j'}(P_{j-j'+i} + P^*_{j-j'+i})\right)Z^*,$$

whose LSD is a compound free Poisson (see Definition 5.6) and by (5.22), its r-th order free cumulant is given by

$$K_{ir} = y^{r-1}\lim \frac{1}{n}\operatorname{Tr}\left(\frac{1}{2}\sum_{j=0}^{q}\sum_{j'=0}^{q} \lambda_j\lambda_{j'}(P_{j-j'+i} + P^*_{j-j'+i})\right)^r = y^{r-1}E_\nu(\cos(i\nu)h(\lambda,\nu))^r, \tag{4.28}$$

where h is given as in (1.8) and $\nu \sim U(0, 2\pi)$.

Now, we will show that (2.15) matches with the results in Liu *et al.* (2013). By (1.6), (1.7), (1.8) and (5.19), we have

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{\cos(i\nu)K_i(z,\nu)d\nu}{1+y\cos(i\nu)K_i(z,\nu)}-1=zm_i(z)=-C(-m_i(z)). \tag{4.29}$$

Moreover, since $\psi_j=\lambda_j I_p$ for all j, then for all $i\geq 0$, $K_i(z,\nu)=m_i(z)h(\lambda,\nu)$. Hence, by (4.29) and (5.20), we have

$$\begin{aligned} C(-m_i(z)) &= -\frac{1}{2\pi}\int_0^{2\pi}\frac{\cos(i\nu)m_i(z)h(\lambda,\nu)d\nu}{1+y\cos(i\nu)m_i(z)h(\lambda,\nu)}+1 \\ &= 1+\sum_{r=1}^{\infty}y^{r-1}E_\nu(\cos(i\nu)h(\lambda,\nu))^r(-m_i(z))^r, \end{aligned} \tag{4.30}$$

where $C(\cdot)$ is the free cumulant generating function defined in (5.19). Therefore, for all $r\geq 1$, coefficient of $(-m_i(z))^r$ in (4.30) agrees with K_{ir} in (2.15).

4.6 Proof of Corollary 2.5

We prove the corollary only for $\hat{\Gamma}_0$ when X_t is an MA(1) process. The proofs are similar for other matrices and for $MA(q)$ process. For convenience, we write $R_i(\nu)$ instead of $R_{0i}(\nu)$. Also by Corollary 2.1(b), it is enough to work under Assumption ($A3$) or ($A4$). To prove (2.16), we consider the following decomposition of $NC_2(2r)$. For $1\leq j\leq r$, let

Π_j^{2r} := set of all non-crossing pair partition of $\{1,2,\ldots,2r\}$ such that one block is $(1,2j)$.

Note that $NC_2(2r)=\cup_{j=1}^r\Pi_j^{2r}$. We will show that the contribution of Π_j^{2r} to $\varphi(\bar{\gamma}_{01}^r)$ is

$$E_\nu\Big(R_{r-j+1}(\nu)\sum_{\pi\in\mathcal{P}_{j-1}}\Big(\frac{t!}{\prod_{k=1}^s f_k!}y^t\prod_{k=1}^t R_{i_k}(\nu)\Big)\Big),\ \ \forall 1\leq j\leq r. \tag{4.31}$$

Then it follows that

$$\varphi(\bar{\gamma}_{01}^r)=\sum_{j=1}^r E_\nu\Big(R_{r-j+1}(\nu)\sum_{\pi\in\mathcal{P}_{j-1}}\Big(\frac{t!}{\prod_{k=1}^s f_k!}y^t\prod_{k=1}^t R_{i_k}(\nu)\Big)\Big), \tag{4.32}$$

which involves all partitions in $\mathcal{P}_r$. Note that for a typical partition $\pi\in\mathcal{P}_r$, coefficient of $\prod_{k=1}^t R_{i_k}(\nu)$ in (4.32) is $y^{t-1}\sum_{j=1}^s\frac{(t-1)!}{(f_j-1)!\prod_{k\neq j}f_k!}$, which is exactly equal to $y^{t-1}\frac{t!}{\prod_{k=1}^s f_k!}$. Hence, (2.16) will be proved provided (4.31) is proved. We prove this by induction on r.

Consider $r=1$. Then we have only Π_1^2. Let $\phi_n=Ep^{-1}\operatorname{Tr}$. Therefore, the contribution of Π_1^2 in $\varphi(\bar{\gamma}_{01})$ is given by

$$\lim\phi_n(ZZ^*+AZZ^*A+AZCZ^*+ZC^*Z^*A)=E_\nu(\varphi(1+a^2+2a\cos\theta))=E_\nu(R_1(\nu)). \tag{4.33}$$

Also note that the contribution of Π_1^4 in $\varphi(\bar{\gamma}_{01}^2)$ is given by

$$\lim \phi_n((ZZ^*+AZZ^*A+AZCZ^*+ZC^*Z^*A)\hat{\Gamma}_{01}) = E_v(\varphi(1+a^2+2a\cos v)\bar{\gamma}_{01}) = E_v(R_2(v)). \tag{4.34}$$

Therefore, by (4.33) and (4.34), it is easy to see that the contribution of Π_1^{2r} in $\varphi(\bar{\gamma}_{01}^r)$ is $E_v(R_r(v))$. This is obtained just by introducing $(r-2)$ more $\bar{\gamma}_{01}$ in (4.34). Moreover, note that the contribution of Π_2^4 i.e. $\{\{1,4\},\{2,3\}\}$ in $\varphi(\bar{\gamma}_{01}^2)$ is $E_v(yR_1(v)R_1(v))$. Therefore,

$$\begin{aligned}\varphi(\bar{\gamma}_{01}^2) &= \text{contribution of } \Pi_1^4 = \{\{1,2\},\{3,4\}\} + \text{contribution of } \Pi_2^4 = \{\{1,4\},\{2,3\}\}\\ &= E_v(R_2(v)) + E_v(yR_1^2(v)). \end{aligned} \tag{4.35}$$

Hence, we get Π_2^6 from Π_2^4 by appending $\bar{\gamma}_{01}$ at the end i.e. by $\{\{1,4\},\{2,3\},\mathbf{\{5,6\}}\}$ and the contribution is $E_v(yR_1(v)R_2(v))$. The partition $\{\{1,6\},\{2,5\},\mathbf{\{3,4\}}\}$ is also obtained from Π_2^4 just by introducing one more $\bar{\gamma}_{01}$ inside the loop of $\{2,3\}$ and then renaming $3,4,5,6$ as $5,6,3,4$ respectively. Thus the contribution of $\{\{1,6\},\{2,5\},\mathbf{\{3,4\}}\}$ is $E_v(yR_2(v)R_1(v))$. Therefore,

$$\begin{aligned}\varphi(\bar{\gamma}_{01}^3) &= \text{contribution of } \Pi_1^6 + \text{contribution of } \Pi_2^6 = \{\{1,4\},\{2,3\},\{5,6\}\}\\ &\quad +\text{contribution of } \Pi_3^6 = \{\{1,6\},\{2,5\},\{3,4\}\} \cup \{\{1,6\},\{2,3\},\{4,5\}\}\\ &= E_v(R_3(v)) + E_v(yR_1(v)R_2(v)) + E_v(yR_2(v)R_1(v))\\ &\quad +\text{contribution of } \{\{1,6\},\{2,3\},\{4,5\}\}. \end{aligned} \tag{4.36}$$

Suppose (4.31) is true for $r = u$. Then, we get any partition $\{i_1, i_2, \ldots, i_t\} \in \mathcal{P}_{u+1}$ with not all $i_j = 1$, just by introducing one more $\bar{\gamma}_{01}$ in the partitions of $\mathcal{P}_k$. Hence, the contribution of all such partitions of $\mathcal{P}_{u+1}$ equals

$$\begin{aligned}&\sum_{j=1}^{u} E_v\Big(R_{u-j+2}(v) \sum_{\pi\in\mathcal{P}_{j-1}} \Big(\frac{t!}{\prod_{k=1}^{s} f_k!} y^t \prod_{k=1}^{t} R_{i_k}(v)\Big)\Big)\\ &+ \; E_v\Big(R_1(v) \sum_{\substack{\pi\in\mathcal{P}_u\\ \text{at least one } i_k > 1}} \Big(\frac{t!}{\prod_{k=1}^{s} f_k!} y^t \prod_{k=1}^{t} R_{i_k}(v)\Big)\Big). \end{aligned} \tag{4.37}$$

The first term of (4.37) covers the contribution of $\cup_{j=1}^{u} \Pi_j^{2(u+1)}$. Moreover, the second term of (4.37) covers the contribution of all partitions in $\Pi_{u+1}^{2(u+1)}$ which contain at least one {odd, even} type blocks except $\{1, 2(u+1)\}$. This is because of the extra $\bar{\gamma}_{01}$ term for which one more {odd, even} type block gets included in the partition. Therefore, (4.37) covers the contribution of all partitions in $NC(2(u+1))$ except

$$\{\{1, 2(u+1)\}, \{2,3\}, \{4,5\}, \ldots, \{2u, 2u+1\}\}, \tag{4.38}$$

which is under $\Pi_{u+1}^{2(u+1)}$ and contains no {odd, even} type block except $\{1, 2(u+1)\}$ (see for example (4.36)). It is very easy to see that the contribution of $\{\{1,4\},\{2,3\}\}$ in $\varphi(\bar{\gamma}_{01}^2)$ is $E_v(yR_1^2(v))$. Suppose the contribution of

$$\{\{1, 2u\}, \{2,3\}, \{4,5\}, \ldots, \{2(u-2), 2u-1\}\} \tag{4.39}$$

in $\varphi(\bar{\gamma}_{01}^{u})$ is $E_{\nu}(y^{u-1}R_{1}^{u}(\nu))$. Now to find the contribution of the partition (4.38) from (4.39), we may just introduce one of either $Z^{*}Z$, $Z^{*}A^{2}Z$, $Z^{*}AZC$ or $Z^{*}AZC^{*}$ in the loop formed by the block $\{1, 2(u+1)\}$. In this case, Z^{*} and Z will form a $\{\text{even}, \text{odd}\}$ type block and by Remark (2.2), this contributes an additional factor of $yR_{1}(\nu)$. Hence, the contribution of the partition (4.38) is $E_{\nu}(y^{u}R_{1}^{u+1}(\nu))$. Thus,

$$\text{the contribution of the partition (4.39) in } \varphi(\bar{\gamma}_{01}^{u}) \text{ is } E_{\nu}(y^{u-1}R_{1}^{u}(\nu)),\ \forall u \geq 1. \tag{4.40}$$

Therefore, combining (4.35),(4.36), (4.37) and (4.40), proof of (4.31) follows by induction.

4.7 Proof of Corollary 2.6

Note that by Corollary 2.1(b), it is enough to work under Assumption $(A3)$ or $(A4)$. Here we will only prove the existence of LSD for $\hat{\Gamma}_{0}$. Similar proof will work for $\{\hat{\Gamma}_{i} + \hat{\Gamma}_{i}^{*}\}_{i\geq 1}$ and $\{\hat{\Gamma}_{i}\hat{\Gamma}_{i}^{*}\}_{i\geq 1}$. Let us denote

$F_{p,q}^{(1)}$, $F_{q}^{(1)}$: ESD and LSD of $\hat{\Gamma}_{0}$ for MA(q) process respectively,

$F_{p,\infty}^{(1)}$, $F_{\infty}^{(1)}$: ESD and LSD of $\hat{\Gamma}_{0}$ for MA(∞) process respectively.

Then it is enough to prove that with d given by (5.3), for every $\varepsilon > 0$,

$(S1)$ there exists a $q_{0} \geq 1$ such that $d(F_{q}^{(1)}, F_{\infty}^{(1)}) < \varepsilon$, $\forall q \geq q_{0}$,

$(S2)$ there exists a $p_{0} \geq 1$, such that for every $p \geq p_{0}$, there is $q_{1} \geq 1$ such that $q \geq q_{1}$ implies $d(F_{p,\infty}^{(1)}, F_{p,q}^{(1)}) < \varepsilon$,

$(S3)$ there exists a $p_{1} \geq 1$ such that $d(F_{p,q}^{(1)}, F_{q}^{(1)}) < \varepsilon$ for all $p \geq p_{1}$.

Note that p_{0} and p_{1} depend only on ε. Also q_{0} depends only on ε and p_{0}. Now,

$$\begin{aligned}
&\varphi\left[\sum_{j=0}^{\infty}\sum_{j'=0}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*} - \sum_{j=0}^{q}\sum_{j'=0}^{q} a_{j}sc_{j'-j}sa_{j'}^{*}\right]^{2} \\
= \ & \varphi\left[\sum_{j=q+1}^{\infty}\sum_{j'=q+1}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*} + \sum_{j=q+1}^{\infty}\sum_{j'=0}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*} + \sum_{j=0}^{\infty}\sum_{j'=q+1}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*}\right]^{2} \\
= \ & \varphi\Bigg[\sum_{j,j'=q+1}^{\infty}\sum_{k,k'=q+1}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*}a_{k}sc_{k'-k}sa_{k'}^{*} + \sum_{j,k=q+1}^{\infty}\sum_{j',k'=0}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*}a_{k}sc_{k'-k}sa_{k'}^{*} \\
& + \sum_{j,k=0}^{\infty}\sum_{j',k'=q+1}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*}a_{k}sc_{k'-k}sa_{k'}^{*} + \sum_{j,k'=q+1}^{\infty}\sum_{j',k=0}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*}a_{k}sc_{k'-k}sa_{k'}^{*} \\
& + \sum_{j,k'=0}^{\infty}\sum_{j',k=q+1}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*}a_{k}sc_{k'-k}sa_{k'}^{*} + \sum_{j,j',k=q+1}^{\infty}\sum_{k'=0}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*}a_{k}sc_{k'-k}sa_{k'}^{*} \\
& + \sum_{j,k,k'=q+1}^{\infty}\sum_{j'=0}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*}a_{k}sc_{k'-k}sa_{k'}^{*} + \sum_{j,j',k'=q+1}^{\infty}\sum_{k=0}^{\infty} a_{j}sc_{j'-j}sa_{j'}^{*}a_{k}sc_{k'-k}sa_{k'}^{*}
\end{aligned}$$

$$+\sum_{j=0}^{\infty}\sum_{j',k,k'=q+1}^{\infty} a_j s c_{j'-j} s a_{j'}^* a_k s c_{k'-k} s a_{k'}^* \Bigg]. \tag{4.41}$$

Now using the linearity of φ, (2.7) and Remark 2.2, we have

$$\begin{aligned}
\varphi(a_j s c_{j'-j} s a_{j'}^* a_k s c_{k'-k} s a_{k'}^*) &\leq |\varphi(a_{j'}^* a_k)\varphi(a_{k'}^* a_j)| + |\varphi(a_j a_{j'}^* a_k a_{k'}^*)| \\
&\leq (\varphi(a_j^* a_j)\varphi(a_{j'}^* a_{j'})\varphi(a_k^* a_k)\varphi(a_{k'}^* a_{k'}))^{1/2} \\
&\quad +(\varphi(a_j^* a_j a_j^* a_j)\varphi(a_{j'}^* a_{j'} a_{j'}^* a_{j'})\varphi(a_k^* a_k a_k^* a_k)\varphi(a_{k'}^* a_{k'} a_{k'}^* a_{k'}))^{1/4}.
\end{aligned}$$

Hence, (4.41) is bounded above by

$$\begin{aligned}
&\left(\sum_{j=q+1}^{\infty}\sqrt{\varphi(a_j^* a_j)}\right)^4 + 4\left(\sum_{q+1}^{\infty}\sqrt{\varphi(a_j^* a_j)}\right)^2\left(\sum_{j=0}^{\infty}\sqrt{\varphi(a_j^* a_j)}\right)^2 \\
&+4\left(\sum_{j=q+1}^{\infty}\sqrt{\varphi(a_j^* a_j)}\right)^3\left(\sum_{j=0}^{\infty}\sqrt{\varphi(a_j^* a_j)}\right) + \left(\sum_{j=q+1}^{\infty}\sqrt[4]{\varphi(a_j^* a_j a_j^* a_j)}\right)^4 \\
&+4\left(\sum_{j=q+1}^{\infty}\sqrt[4]{\varphi(a_j^* a_j a_j^* a_j)}\right)^2\left(\sum_{j=0}^{\infty}\sqrt[4]{\varphi(a_j^* a_j a_j^* a_j)}\right)^2 \\
&+4\left(\sum_{j=q+1}^{\infty}\sqrt[4]{\varphi(a_j^* a_j a_j^* a_j)}\right)^3\left(\sum_{j=0}^{\infty}\sqrt[4]{\varphi(a_j^* a_j a_j^* a_j)}\right) \to 0, \text{ as } q \to \infty,
\end{aligned}$$

by Assumption $(A11)$. Therefore, $(S1)$ holds. If we replace a_j, a_j^*, c_j, c_j^* and s by the matrices $\bar{\psi}_j, \bar{\psi}_j{}^*, \underline{\mathrm{P}}_j, \underline{\mathrm{P}}_j^*$ and W respectively, then by Assumption $(A10)$, the same calculations as above proves $(S2)$. $(S3)$ follows from Corollary 2.1. Hence, Corollary 2.6 is proved for $\hat{\Gamma}_0$.

5 Appendix

In this section, we discuss the concepts and results from the literature of non-commutative probability which have been used in the proof of Theorem 2.1 and its corollaries in Section 2.

Moment method. We begin with a result (see Bai & Silverstein (2009)), which outlines the moment method for showing the existence of LSD. This was used in the proof of existence of the LSDs for symmetric polynomials of $\{\hat{\Gamma}_i, \hat{\Gamma}_i^* : i \geq 0\}$ (see Corollary 2.1 and Section 4.2).

The h-th order raw moment of the ESD of an $n\times n$ real symmetric matrix R_n is given by

$$\beta_h(R_n) := \frac{1}{n}\sum_{i=1}^{n}\lambda_i^h = \frac{1}{n}\mathrm{Tr}(R_n^h). \tag{5.1}$$

Consider the following conditions.

(M1) For every $h \geq 1, E(\beta_h(R_n)) \to \beta_h$.

(M2) $\mathrm{Var}(\beta_h(R_n)) \to 0, \ \ \forall h \geq 1$.

(M4) $\sum_{n=1}^{\infty} E(\beta_h(R_n) - E(\beta_h(R_n)))^4 < \infty, \ \ \forall h \geq 1$.

(C) The sequence $\{\beta_h\}$ satisfies Carleman's condition,

$$\sum_{h=1}^{\infty} \beta_{2h}^{-\frac{1}{2h}} = \infty. \tag{5.2}$$

Lemma 5.1. *If (M1), (M2) and (C) hold, then ESD of R_n converges in probability to a unique probability distribution F determined by the moment sequence $\{\beta_h\}$. Further the convergence is almost sure if $(M4)$ holds.*

Metric for truncation arguments. For the truncation argument in the proof of Corollary 2.6, we use a Mallow's metric (or the Wasserstein metric) on the space of all probability distributions with finite second moment.

Let F and G be two distribution functions with finite second moment. Then the Mallow's distance between F and G is defined as

$$d(F,G) = \Big[\inf_{X \sim F, Y \sim G} \mathrm{E}\,|X-Y|^2\Big]^{1/2},$$

where the infimum is taken over all possible joint distributions of (X, Y) such that their marginal distributions are respectively F and G. It is known that $d(F_n, F) \to 0$ if and only if F_n converges to F weakly and $\displaystyle\int x^2 dF_n(x) \to \int x^2 dF(x)$. This metric is useful in random matrices due to the following two upper bounds.

Let A, B be two $n \times n$ real symmetric matrices with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_n$, respectively. Let F^A and F^B be the ESD of A and B respectively. Then by Lemma 2.3 of Bai (1999), we have

$$d^2(F^A, F^B) \leq n^{-1}\,\mathrm{Tr}(A-B)^2. \tag{5.3}$$

Also suppose A and B are $p \times n$ real matrices. Let $X = AA^*$ and $Y = BB^*$ and F^X, F^Y be ESD of X and Y respectively. Then

$$d^2(F^X, F^Y) \leq \frac{2}{p^2}\,\mathrm{Tr}(X+Y)\,\mathrm{Tr}[(A-B)(A-B)^*]. \tag{5.4}$$

Stieltjes and related transforms. We needed the Stieltjes and related transforms to check the agreement of our results with the exisiting results (for example Corollaries 2.2 (c) and 2.3). We provide a brief discussion on various transforms following Couillet & Debbah (2011).

(a) If all moments of a probability measure μ are finite then we have the following formal power series expansion of the Stieltjes transformation $m_\mu(z)$ defined in (1.3),

$$m_\mu(z) = -\frac{1}{z}E_\mu\left(\frac{1}{1-\frac{x}{z}}\right) = -\frac{1}{z} - \frac{E_\mu(x)}{z^2} - \frac{E_\mu(x^2)}{z^3} - \dots. \tag{5.5}$$

(b) Recall the definition of $\bar{M}$ given in (2.3). By (1.3), if $m_M(z)$ is the Stieltjes transform of the ESD of a $p \times p$ matrix M, then the Stieltjes transform of the $(n+p) \times (n+p)$ matrix $\bar{M}$ is given by

$$m_{\bar{M}}(z) = \frac{y}{1+y}m_M(z) + \frac{1}{z}\frac{1}{1+y}. \tag{5.6}$$

(c) Ψ transformation. Given a probability measure μ on $\mathbb{R}$ with compact support, the Ψ-transformation $\Psi_\mu(z)$ is defined by

$$\Psi_\mu(z) = \int \frac{zt}{1-zt}d\mu(t). \tag{5.7}$$

One can easily show that

$$\Psi_\mu\left(\frac{1}{z}\right) = -zm_\mu(z) - 1, \tag{5.8}$$

where m_μ is the Steiltjes transformation of the measure μ.

(d) χ-transformation. χ_μ is the unique function analytic in a neighbourhood of zero, that satisfies

$$\chi_\mu(\Psi_\mu(z)) = z \text{ for } |z| \text{ small enough.} \tag{5.9}$$

(e) S transformation. The S-transformation S_μ of the measure μ is given by

$$S_\mu(z) = \frac{1+z}{z}\chi_\mu(z). \tag{5.10}$$

The S-transformation $S_{\mu\boxtimes\nu}$ of the measure $\mu \boxtimes \nu$, free product of two measures μ and ν (see Definition 5.4), satisfies

$$S_{\mu\boxtimes\nu} = S_\mu S_\nu. \tag{5.11}$$

Non-commutative probability space.

Definition 5.1. A *non-commutative $*$-probability space* $(\mathcal{A}, \varphi)$ consists of a unital $*$-algebra $\mathcal{A}$ over $\mathbb{C}$ and a unital linear functional (state)

$$\varphi : \mathcal{A} \to \mathbb{C}, \;\; \varphi(1_{\mathcal{A}}) = 1.$$

The elements $a \in \mathcal{A}$ are called non-commutative *random variables* in $(\mathcal{A}, \varphi)$. Moreover, if $a = a^*$, then a is called *self-adjoint*. The state φ is *tracial* if

$$\varphi(ab) = \varphi(ba), \;\; \forall a, b \in \mathcal{A}.$$

The following example of non-commutative $*$-probability space is relevant for us.

Example 1. Let d be a positive integer. Let $\mathcal{M}_d(\mathbb{C})$ be the $*$-algebra of $d \times d$ matrices with complex entries and with usual matrix multiplication, and let $tr : \mathcal{M}_d(\mathbb{C}) \to \mathbb{C}$ be the normalized trace,

$$tr(a) = \frac{1}{d}\sum_{i=1}^{d} \alpha_{ii} \quad \forall a = ((\alpha_{ij}))_{i,j=1}^{d} \in \mathcal{M}_d(\mathbb{C}).$$

Then $(\mathcal{M}_d(\mathbb{C}), tr)$ is a $*$-probability space.

If the entries of $\mathcal{M}_{(d)}(\mathbb{C})$ are random, then it is a non-commutative $*$-probability spaced with $\varphi = E\ tr$.

Let M be any square self-adjoint matrix with LSD a supported on $\mathbb{R}$. Then for all $k \geq 1$, $\lim \varphi_n(M^k)$ equals $E(a^k)$.

Let $\mathcal{B}$ be a unital $*$-sub-algebra of $\mathcal{A}$. Then $(\mathcal{B}, \varphi)$ also forms a non-commutative $*$-probability space. Consider $t \geq 1$. Let $\Pi(a_i, a_i^* : 1 \leq i \leq t) \in \mathcal{A}$ be any polynomial of $\{a_i, a_i^* : 1 \leq i \leq t\} \subset \mathcal{A}$. Then

$$\text{Span}\{a_i, a_i^* : 1 \leq i \leq t\} = \{\Pi(a_i, a_i^* : 1 \leq i \leq t) : \ \Pi \text{ is a polynomial}\} \tag{5.12}$$

is the $*$-algebra generated by $\{a_i, a_i^* : 1 \leq i \leq t\}$. Clearly it is a non-commutative $*$-probability space.

For example, $\mathcal{A}_n$ in (2.4) is a unital $*$-sub-algebra of $\mathcal{M}_{n+p}(\mathbb{C})$ (see Example 1) and $(\mathcal{A}_n, \varphi_n)$ forms a non-commutative $*$-probability space. Similarly, $(\mathcal{A}, \varphi)$ defined by (2.6), (2.2), (2.8) and (2.7), $(\mathcal{B}_q, \varphi)$ defined by (2.10), are all non-commutative $*$-probability spaces.

Definition 5.2. Let $\Pi(a, a^*) \in \mathcal{A}$ be any polynomial in $a, a^* \in \mathcal{A}$. Then $\{\varphi(\Pi(a, a^*)) : \Pi(a, a^*) \in \mathcal{A}\}$ is called the $*$-distribution of a or a^*. In particular, if a is self-adjont, then the sequence $\{\varphi(a^k)\}_{k=1}^{\infty}$, is the distribution of the random variable $a \in \mathcal{A}$. It often defines a unique probability measure on $\mathbb{R}$ with this as its moment sequence. For example, the distribution of a semi-circular variable is given by (2.2) which are the moments of the (unique) semi-circle law.

Consider $t \geq 1$. Let $\Pi(a_i, a_i^* : 1 \leq i \leq t) \in \mathcal{A}$ be any polynomial in $\{a_i : 1 \leq i \leq t\} \subset \mathcal{A}$. Then $\{\varphi\left(\Pi(a_i, a_i^* : 1 \leq i \leq t)\right) : \Pi \in \mathcal{A}\}$ determines the joint distribution of $\{a_i : 1 \leq i \leq t\}$. For example, (2.7) and (2.8) provide the joint distribution of $\{c_i, c_i^* : i \geq 0\}$ and $\{a_i, a_i^* : i \geq 0\}$ respectively.

Non-crossing partitions and Kreweras complement. Non-crossing partitions form the core of the concept of free independence. Kreweras complement is a useful tool to compute various moments of polynomials of free variables (see for example Remark 2.2).

A partition of a set say, $\{1, 2, \ldots n\}$ is said to be non-crossing if for any two blocks V_1 and V_2 of the partition, there does not exist $a_{i,1}, a_{i,2} \in V_i$, $i = 1, 2$ such that $a_{1,1} <$

$a_{2,1} < a_{1,2} < a_{2,2}$. This is a POSET with the natural ordering which stipulates that the one block partition (say 1_n) is the largest element and the n bock partition (say 0_n) is the smallest. It is closed under the natural max and min operations. Let

$$NC(n) = \{\pi : \pi \text{ is a non-crossing partition of } \{1, 2, 3, \ldots, n\}\}. \tag{5.13}$$

$$NC_2(2n) = \{\pi : \pi \text{ is a non-crossing pair partition of } \{1, 2, 3, \ldots, 2n\}\}. \tag{5.14}$$

$$NCE(2n) = \{\pi \in NC(2n) : \text{ every block of } \pi \text{ has even cardinality}\}. \tag{5.15}$$

The Kreweras complementation map $K : NC(n) \to NC(n)$ is defined as follows. We consider additional numbers $\bar{1}, \bar{2}, \ldots, \bar{n}$ and interlace them with $1, 2, \ldots, n$ in the following alternating way:

$$1, \bar{1}, 2, \bar{2}, \ldots, n, \bar{n}.$$

Let π be a non-crossing partition of $\{1, 2, \ldots, n\}$. Then its Kreweras complement $K(\pi) \in NC(\bar{1}, \bar{2}, \ldots, \bar{n}) \sim NC(1, 2, \ldots, n)$ is defined to be the biggest element among those $\sigma \in NC(\bar{1}, \bar{2}, \ldots, \bar{n})$ which have the property that

$$\max(\pi, \sigma) \in NC(1, \bar{1}, 2, \bar{2}, \ldots, n, \bar{n}).$$

The following properties are known. Properties $1 - 3$ have been used in Remark 2.2 and Property 4 was used in Corollaries 2.2 (*a*) and (*b*).

Property 1. $K : NC(n) \to NC(n)$ is a bijection.

Property 2: $K(NCE(2n))$ is in bijection with the set of all such π in $NC(2n)$ such that every block of π is contained either in $\{1, 3, \ldots, 2n-1\}$ or in $\{2, 4, \ldots, 2n\}$.

Property 3: $NC_2(2n) \ni \pi \to (K(\pi)|\{1, 3, \ldots, 2n-1\})$ is a bijection between $NC_2(2n)$ and $NC(1, 3, \ldots, 2n-1)$.

Property 4: Let $|\pi|$ be the total number of blocks in any partition π. Then for any $\pi \in NC(n)$, we have $|\pi| + |K(\pi)| = n + 1$.

Free cumulants and free independence.

Definition 5.3. Let $(\mathcal{A}, \varphi)$ be a non-commutative $*$-probability space. Define multilinear functionals $(\varphi_n)_{n\in\mathbb{N}}$ on $\mathcal{A}$ via

$$\varphi_n(a_1, a_2, \ldots, a_n) := \varphi(a_1 a_2 \ldots a_n).$$

Define recursively a family of multiplicative, multilinear functionals $\varphi_\pi (n \geq 1, \pi \in NC(n))$ by the following formula. If $\pi = \{V_1, V_2, \ldots, V_r\} \in NC(n)$, then

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

where $\varphi(V)[a_1, a_2, \ldots, a_n] := \varphi_s(a_{i_1}, a_{i_2}, \ldots, a_{i_s})$ for $V = (i_1 < i_2 < \ldots < i_s)$. Then the *free cumulants* $(k_\pi)_{\pi \in NC(n), n \geq 1}$ are the multiplicative, multilinear functionals defined by

$$k_\pi[a_1, a_2, ..., a_n] := \sum_{\substack{\sigma \in NC(n) \\ \sigma \leq \pi}} \varphi_\sigma[a_1, a_2, \ldots, a_n]\mu(\sigma, \pi), \tag{5.16}$$

where μ is the Möbius function on the POSET $NC(n)$. Note that the Möbius function depends only on partitions and not on the variables whose cumulants are being calculated. For each $n \geq 1$, we put $k_n := k_{1_n}$. (5.16) is equivalent to the statement that, for all $n \in \mathbb{N}$ and all $a_1, a_2, \ldots, a_n \in \mathcal{A}$, we have

$$k_n(a_1, a_2, \ldots, a_n) = \sum_{\sigma \in NC(n)} \varphi_\sigma[a_1, a_2, ..., a_n]\mu(\sigma, 1_n). \tag{5.17}$$

For any variable a, the numbers $k_n(a) = k_n(a, a, \ldots, a)\ \forall n \geq 1$ are called the *free cumulants of* a.

If s is a semi-circular random variable, then it is well known that

$$k_n(s) = \begin{cases} 1, & \text{if } n = 2 \\ 0, & \text{if } n > 2. \end{cases} \tag{5.18}$$

The free cumulant generating function $C(z)$ of a self-adjoint random variable a is defined as the formal power series

$$C(z) = 1 + \sum_{n=1}^{\infty} k_n(a) z^n \quad \forall z \in \mathbb{C}. \tag{5.19}$$

$C(z)$ satisfies the relation

$$C(-m(z)) = -zm(z), \tag{5.20}$$

where $m(z)$ is the Stieltjes transformation defined in (1.3).

The free cumulant generating function $C(z)$ and relation (5.20) can often be used to derive the Stieltjes transformation of a random variable (see Corollary 2.4 and alternative proof of Corollary 2.2) (c)).

Free independence of random variables and sub-algebras can be defined through free cumulants.

Definition 5.4. Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. Consider unital sub-algebras $(\mathcal{A}_i)_{i \in I}$ of $\mathcal{A}$. Then $(\mathcal{A}_i)_{i \in I}$ are *freely independent* if for all $n \geq 2$ and for all $a_i \in \mathcal{A}_{i(j)}$ $(j = 1, 2, \ldots, n)$ with $i(1), i(2), \ldots, i(n) \in I$, we have $k_n(a_1, a_2, \ldots, a_n) = 0$ whenever there exist $1 \leq l, k \leq n$ with $i(l) \neq i(k)$.

Also, let $(\mathcal{A}_i, \varphi_i)_{i \in I}$ be a family of non-commutative probability spaces, then there exists a non-commutative probability space $(\mathcal{A}, \varphi)$, called *free product* of $(\mathcal{A}_i, \varphi_i)_{i \in I}$, such that $\mathcal{A}_i \subset \mathcal{A}, i \in I$ are freely independent in $(\mathcal{A}, \varphi)$ and $\varphi|\mathcal{A}_i = \varphi_i$.

Let μ and ν be compactly supported probability measures on $\mathbb{R}^+$ and $\mathbb{R}$ respectively. Then their free product $\mu \boxtimes \nu$ is defined as the distribution of $\sqrt{x} y \sqrt{x}$, where x and y have μ and ν respectively as their distribution and they are freely independent. This is useful in (5.11).

(5.17) and Definition 5.4 provide tools for showing free independence of two subalgebras.

Next we state how to compute φ functions under free independence.

Lemma 5.2. *Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and consider random variables $a_1, a_2, \ldots a_n, b_1, b_2, \ldots, b_n \in \mathcal{A}$ such that $\{a_1, a_2, \ldots a_n\}$ and $\{b_1, b_2, \ldots, b_n\}$ are freely independent. Then we have*

$$\varphi(a_1 b_1 a_2 b_2 \ldots a_n b_n) = \sum_{\pi \in NC(n)} k_\pi[a_1, a_2, \ldots, a_n] \varphi_{K(\pi)}[b_1, b_2, \ldots, b_n],$$

where $K(\pi)$ is the Kreweras complement.

Convergence of NCP and freeness

Definition 5.5. Let $\{(\mathcal{A}_N, \varphi_N)\}_{N=1}^{\infty}$ be a sequence of non-commutative $*$-probability spaces. We say that this sequence converges to a non-commutative $*$-probability space $(\mathcal{A}, \varphi)$ if for each $a^{(N)} \in \mathcal{A}_N$, there is a *corresponding element* (limit) $a \in \mathcal{A}$ such that

$$\lim_{N \to \infty} \varphi_N \left(\Pi(a_i^{(N)}, a_i^{*(N)} : 1 \leq i \leq t) \right) = \varphi \left(\Pi(a_i, a_i^* : 1 \leq i \leq t) \right), \text{ for any polynomial } \Pi \text{ and } t \geq 1.$$

In particular, for a fix $i \geq 1$, we say that a_i^N converges in distribution to a_i if $\lim \varphi_N(a_i^{(N)k}) = \varphi(a_i^k)$ $\forall k \geq 1$. Moreover, for fixed $t \geq 1$, by joint convergence of $\{a_i^{(N)}, a_i^{*(N)} : 1 \leq i \leq t\}$ to $\{a_i, a_i^* : 1 \leq i \leq t\}$, we mean $(\text{Span}\{a_i^{(N)}, a_i^{*(N)} : 1 \leq i \leq t\}, \varphi_N)$ converges to $(\text{Span}\{a_i, a_i^* : 1 \leq i \leq t\}, \varphi)$.

The well known Lemma 5.3 provides the joint convergence of p- many independent Wigner and q-many deterministic matrices and their asymptotic freeness. In particular, if W_n is a Wigner matrix of order $n \times n$, then $n^{-1/2} W_n$ converges in distribution to a semi-circle law determined by the moment sequence (2.2). Proof of Lemma 5.3 under the Assumption $(A3)$, is given in Zeitouni *et al.* (2010). Similar proof works under the Assumption $(A4)$.

Lemma 5.3. *Let, for each $N \in \mathbb{N}$, $A_N^{(1)}, A_N^{(2)}, \ldots, A_N^{(p)}$ be p independent Wigner matrices and let $D_N^{(1)}, D_N^{(2)}, \ldots, D_N^{(q)}$ be q constant matrices with bounded norm, which converge in distribution for $N \to \infty$, i.e.*

$$D_N^{(1)}, D_N^{(2)}, \ldots, D_N^{(q)} \xrightarrow{d} d_1, d_2, \ldots, d_q,$$

for some $d_1, d_2, \ldots d_q \in (\mathcal{A}, \varphi)$. Then, under Assumption either $(A3)$ or $(A4)$ on the entries of the Wigner matrices,

$$A_N^{(1)}, A_N^{(2)}, \ldots, A_N^{(p)}, D_N^{(1)}, D_N^{(2)}, \ldots, D_N^{(q)} \xrightarrow{\mathcal{D}} s_1, s_2, \ldots, s_p, d_1, d_2, \ldots, d_q,$$

where each s_i is a standard semi-circular element and $s_1, s_2, \ldots, s_p$, $\{d_1, d_2, \ldots, d_q\}$ are free.

The following lemma was used in the proof of Theorem 2.1.

Lemma 5.4. *Span*$\{\bar{\psi}_i, \bar{\psi}_i^* \,:\, i \geq 0\}$ *and Span* $\left\{(n+p)^{-1}W\underline{P}_iW, (n+p)^{-1}W\underline{P}_i^*W \,:\, i \geq 0\right\}$ *are asymptotically freely independent.*

This implies that though Span$\{\bar{\psi}_i, \bar{\psi}_i^* \,:\, i \geq 0\}$ and Span$\{\underline{P}_i, \underline{P}_i^* \,:\, i \geq 0\}$ are not freely independent, to compute limits of φ_n for any polynomials of $\{\hat{\Gamma}_i, \hat{\Gamma}_i^* \,:\, i \geq 0\}$, they asymptotically behave like freely independent non-commutative $*$-probability spaces.

Compound free Poisson distribution

Definition 5.6. A probability measure μ on $\mathbb{R}$ with free cumulants

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

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.

Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. Let $s, a \in \mathcal{A}$ be such that s is a semi-circular element, defined by the moment sequence (2.2), and moreover s and a are free. Then the free cumulants of sas are given by

$$k_n(sas, sas, \ldots, sas) = \varphi(a^n) \ \forall n \geq 1. \tag{5.21}$$

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

Let Z be any $p \times n$ ID matrix satisfying $(A1)$, $(A2)$, $(A9)$ and either $(A3)$ or $(A4)$. Let A be a self-adjoint matrix of order p with compactly supported LSD a. Here $p = p(n)$ is such that $pn^{-1} \to y \in (0, \infty)$. Then it can be shown that the limiting free cumulants of ZAZ^* are given by

$$\lim_n k_r(ZAZ^*, ZAZ^*, \ldots, ZAZ^*) = y^{r-1}\varphi(a^r), \ \forall r \geq 1. \tag{5.22}$$

Therefore, asymptotically ZAZ^* is a compound free Poisson variable with rate y^{-1} and jump measure ya. Similarly, for any matrix B of order n having compactly supported LSD b, the limiting free cumulants of Z^*BZ are given by

$$\lim_n k_r(Z^*BZ, Z^*BZ, \ldots, Z^*BZ) = y\varphi(b^r), \ \forall r \geq 1, \tag{5.23}$$

i.e. Z^*BZ is a compound free Poisson random variable with rate y and jump measure b.

This was useful to describe the LSD of $\{\frac{1}{2}(\hat{\Gamma}_i + \hat{\Gamma}_i^*)\}_{i \geq 0}$ when the coefficient matrices are $\psi_j = \lambda_j I_p$, for all $j \geq 0$ (see Corollary 2.4). It was also used to obtain the Stieltjes transformation in (2.14) from our results (see alternative proof of Corollary 2.2 (c)).

Acknowledgment: We thank Octavio Arizmendi and Carlos Vargas Obieta for interacting and helpful discussions on free probability.

References

Arnold, L. 1967. On the asymptotic distribution of eigenvalues of the random matrices. *Aust. J. Math. Anal. Appl.*, **20**, 262–268.

Arnold, L. 1971. On Wigner's semicircle law for the eigenvalues of random matrices. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **19**(3), 191–198.

Bai, Z. D. 1993a. Convergence rate of expected spectral distributions of large random matrices. Part I. Wigner matrices. *Ann. Probab.*, **21**(2), 625–648.

Bai, Z. D. 1993b. Convergence rate of expected spectral distributions of large random matrices. Part II. Sample covariance matrices. *Ann. Probab.*, **21**(2), 649–672.

Bai, Z. D. 1999. Methodologies in spectral analysis of large dimensional random matrices, a review. *Stat Sinica*, **9**, 611–677.

Bai, Z. D., & Silverstein, J. W. 2009. *Spectral analysis of large dimensional random matrices.* Springer.

Bai, Z. D., Miao, B. Q., & Tsay, J. 1999. Remarks on the convergence rate of the spectral distributions of Wigner matrices. *J. Theoret. Probab.*, **12**(2), 301–311.

Bai, Z. D., Miao, B. Q., & Tsay, J. 2007. A note on the convergence rate of the spectral distributions of large dimensional random matrices. *Statist. Probab. Lett.*, **34**, 95–101.

Bai, Zhidong, & Zhou, Wang. 2008. Large sample covariance matrices without independence structures in columns. *Statist. Sinica*, **18**(2), 425–442.

Basak, A., Bose, A., & Sen, S. 2014. Limiting spectral distribution of sample autocovariance matrices. *Bernoulli Journal*, **20**(3), 1234–1259.

Bhattacharjee, M., & Bose, A. 2014. Estimation of Autocovariance matrices for Infinite Dimensional Vector Linear Process. *J. Time Series Anal.*, **35**(3), 262–281.

Brockwell, Peter J, & Davis, Richard A. 2009. *Time series: theory and methods*. Springer.

Burda, Z. 2013. Free products of large random matrices – a short review of recent developments. *J. Phys.: Conf. Ser.*, **473**(012002), conference 1.

Couillet, R., & Debbah, M. 2011. *Random matrix methods for wireless communications.* Cambridge, UK: Cambridge University Press.

Edelman, P. H. 1980. Chain enumeration and non-crossing partitions. *Discrete Mathematics*, **31**, 171–180.

Forni, M., & Lippi, M. 2001. The Generalized Factor Model: representation Theory. *Econometric Theory*, **17**, 1113–1141.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. 2000. The Generalized Dynamic Factor Model: identification and Estimation. *Review of Economics and Statistics*, **82**, 540–554.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. 2004. The Generalized Dynamic Factor Model: consistency and Rates. *J. Econometrics*, **119**, 231–235.

Francq, C., & Raïsi, H. 2007. Multivariate Portmanteau test for autoregressive models with uncorrelated but nonindependent errors. *J. Time Series Anal.*, **28**(3), 454–470.

Hannan, E. 1970. *Multiple Time Series*. New York: John wiley and Sons, Inc.

Hong, Y. 1996. Consistent testing for serial correlation of unknown form. *Econometrica*, **64**, 837–864.

Hosking, J. R. M. 1980. The multivariate portmanteau statistic. *J. Amer. Statist. Assoc.*, **75**(371), 602–607.

Hosking, J. R. M. 1981a. Equivalent forms of multivariate portmanteau statistic. *J. Roy. Statist. Soc. B*, **43**(2), 261–262.

Hosking, J. R. M. 1981b. Lagrange multiplier tests of mutivariate time sereis models. *J. Roy. Statist. Soc. B (Methodological)*, **43**(2), 219–230.

Jin, Baisuo, Wang, Chen, Bai, Z. D., Nair, K. Krishnan, & Harding, Matthew. 2014. Limiting spectral distribution of a symmetrized auto-cross covariance matrix. *Ann. Appl. Probab.*, **24**(3), 1199–1225.

Katayama, N. 2008. An improvement of the Portmanteau statistic. *J. Time Series Anal.*, **29**(2), 359–378.

Lin, J. W., & Mc.Leod, A.I. 1981. Distribution of the residual autocorrelation in multivariate ARMA time series models. *J. Roy. Statist. Soc. B*, **43**(2), 231–239.

Lin, J. W., & Mc.Leod, A.I. 2006. Improved Pena-Rodriguez portmanteau test. *Computational statistics and data analysis*, **51**(3), 1731–1738.

Liu, H., Aue, A., & Paul, D. 2013. On the Marcenko-Pastur law for linear time series. *arXiv:1310.7270*.

Ljung, G. M., & Box, G. E. P. 1978. On a measure of lack of fit in time series models. *Biometrika*, **65**, 297–303.

Mahdi, Esam, & Mc.Leod, Lan. 2012. Improved multivariate portmanteau test. *J. Time Series Anal.*, **33**(2), 211–222.

Mc.Leod, A.I. 1978. On the distribution of residual autocorrelations in Box-Jenkins models. *J. Roy. Statist. Soc. B*, **40**, 296–302.

Nica, A., & Speicher, R. 2006. *Lectures on the combinatorics of free probability*. Cambridge, UK: Cambridge University Press.

Shao, X. 2011. Testing for White Noise under unknown dependence and its application to diagnostic checking for time series models. *Econometric Theory, 1-32*.

Voiculescu, Dan. 1991. Limit laws for random matrices and free products. *Invent. Math.*, **104**(1), 201–220.

Wigner, E. P. 1958. On the distribution of the roots of certain symmetric matrices. *Ann. of Math.*, **67**(2), 325–327.

Zeitouni, O., Anderson, G. W., & Guionnet, A. 2010. *An Introduction to Random Matrices*. New York: Cambridge University Press.