Matrix polynomial generalizations of the sample variance-covariance matrix when $pn^{-1} \to 0$

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

March 2, 2015

Abstract

Let $\{Z_u = ((\varepsilon_{u,i,j}))_{p\times n}\}$ be random matrices where $\{\varepsilon_{u,i,j}\}$ are independently distributed. Suppose $\{A_i\}$, $\{B_i\}$ are non-random matrices of order $p \times p$ and $n \times n$ respectively. Consider all $p \times p$ random matrix polynomials constructed from the above matrices of the form $\mathbb{P} = \left(\prod_{i=1}^{k_l} n^{-1} A_{t_i} Z_{j_i} B_{s_i} Z_{j_i}^*\right) A_{t_{k_l+1}}$ and the corresponding centering polynomials $\mathbb{G} = \left(\prod_{i=1}^{k_l} n^{-1}\mathrm{Tr}\,(B_{s_i})\right) \prod_{i=1}^{k_l+1} A_{t_i}$. We show that under appropriate conditions on the above matrices, the non-commutative $*$- probability space $\mathcal{U}_n = \mathrm{Span}\left\{(n/p)^{1/2}(\mathbb{P} - \mathbb{G})\right\}$ with state $p^{-1}E\mathrm{Tr}$ converges.

As a consequence, the limiting spectral distribution (LSD) of any symmetric polynomial from the above space exists almost surely and they are expressed in terms of semi-circle families and limits of $\{A_i\}$, $\{B_i\}$. These LSD results generalize the results known for the two specific matrices $\sqrt{np^{-1}}(n^{-1}ZZ^* - I)$ and $\sqrt{np^{-1}}(n^{-1}B_1^{1/2}ZB_2Z^*B_1^{1/2} - n^{-1}\mathrm{Tr}(B_2)B_1)$.

As an application, we obtain the almost sure LSD of any symmetrized sample autocovariance matrices for the infinite dimensional MA(q) process.

Key words and phrases. Independent matrix, moment method, Stieltjes transformation, limiting spectral distribution, semi-circle law, non-crossing partition, non-commutative probability space, $*$-algebra, free cumulants, infinite dimensional vector linear process, symmetrized autocovariance matrix.

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

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

1 Introduction

Suppose Z is a $p \times n$ matrix with i.i.d. entries which have finite moment of appropriate order. Suppose further that $p \to \infty$ and $p/n \to 0$. In this regime, limiting spectral distribution (LSD) (see Appendix) results for some specific polynomials in one or two deterministic matrices (A and B) and the matrix Z are known. Here is a brief description of these results that are relevant to this article. The precise conditions that are needed for these results to hold are given in the later sections. All these results were established by using the Stieltjes transformation method, unless otherwise mentioned.

(i) [Bai and Yin, 1988] proved that the almost sure LSD of $\sqrt{np^{-1}}(n^{-1}ZZ^* - I_p)$ is the standard semi-circle law. [Bose et al., 2010] gave a moment method proof of this result. Recall that the semi-circle law is also the almost sure LSD of the $p \times p$ matrix $p^{-1/2}W_p$ where W_p is symmetric with i.i.d. entries whose second moment is finite.

(ii) Let A be a $p \times p$ non-negative definite matrix whose LSD exists. Then [Pan and Gao, 2009] and [Bao, 2012] showed that the LSD of $\sqrt{np^{-1}}(n^{-1}A^{1/2}ZZ^*A^{1/2} - A)$ exists and provided its Stieltjes transform (see Corollary 2.2 (d) (i)). It turns out that this LSD coincides with the LSD of the generalized Wigner matrix $A^{1/2}p^{-1/2}W_pA^{1/2}$ studied by [Bai and Zhang, 2010]. In fact a free probability description of the limit is the following: let a and s be two free variables where a is distributed as the LSD of A and s has the semicircular distribution. Then the LSD is identical to the distribution of the self-adjoint variable $a^{1/2}sa^{1/2}$ (see Corollary 2.2 (a) and (b)).

(iii) Suppose A is as in (ii) and B is an $n \times n$ symmetric matrix such that $b = \lim n^{-1}\mathrm{Tr}(B^2)$ exists. [Wang and Paul, 2014] derived the LSD of $\sqrt{np^{-1}}(n^{-1}A^{1/2}ZBZ^*A^{1/2} - n^{-1}\mathrm{Tr}(B)A)$, which coincides with the distribution of $\sqrt{b}a^{1/2}sa^{1/2}$ (for details see Corollary 2.2).

It may be noted that each of these results deals with one specific matrix polynomial involving one random matrix and one or two deterministic matrices. It is very natural to ask what happens when we allow several random and non-random matrices and also seek the joint limit properties of their polynomials. Moreover, in many high dimensional models, such matrices and polynomials arise quite naturally. Autocovariance matrices of infinite dimensional linear processes is one such application that we discuss later.

A clue for what to expect, not just for the LSD but more generally for the so called *-algebra convergence, is that the matrix B appearing in between Z and Z^* in (iii) above, contributes in the limit only through $n^{-1}Tr(B^2)$ and $n^{-1}Tr(B)$ whereas A, in (i) $-$ (iii) above is, in the limit free of the semicircle variable that arises via ZZ^*. Our goal in this article is to make this idea precise and provide a complete description of the joint convergence.

So, suppose we have finitely many independent *"data" matrices* $Z_u = ((\varepsilon_{u,i,j}))_{p\times n}$, $u \geq 1$ (say) (also called *independent matrices*), where $\{\varepsilon_{u,i,j} : u, i, j \geq 1\}$ are independent.

Also suppose $\{B_{2i-1}\}$, $\{B_{2i}\}$ are constant matrices of order $p \times p$ and $n \times n$ respectively. Consider the $p \times p$ matrix polynomials

$$\mathbb{P}_{l,(u_{l,1},\ldots,u_{l,k_l})} = \Big(\prod_{i=1}^{k_l} n^{-1} A_{l,2i-1} Z_{u_{l,i}} A_{l,2i} Z^*_{u_{l,i}}\Big) A_{l,2k_l+1}, \quad (1.1)$$

$$\mathbb{G}_l = \Big(\prod_{i=1}^{k_l} n^{-1} \mathrm{Tr}\,(A_{l,2i})\Big) \prod_{i=0}^{k_l} A_{l,2i+1}. \quad (1.2)$$

where, $A_{l,2i-1}$, $A_{l,2i}$ and $Z_{u_{l,i}}$, are matrices from the collections $\{B_{2i-1}\}$, $\{B_{2i}\}$ and $\{Z_u\}$ respectively.

Under suitable assumptions discussed in Section 2, we study the joint behaviour of the centered matrices $\Big\{(n/p)^{1/2}(\mathcal{P} - \mathcal{G})\Big\}$. The most natural way to study the joint convergence of these matrices is to consider them as elements of the $*$-sub-algebra of the *non-commutative probability space* (see Definition 4.1) of all $p \times p$ (say) matrices with complex entries (see Example 1) and with the *state* equal to normalized trace or expected normalized trace. We show that the $*$ -probability space

$$\mathcal{U}_n = \mathrm{Span}\Big((p/n)^{-1/2}(\mathbb{P}_{l,(u_{l,1},\ldots,u_{l,k_l})} - \mathbb{G}_l) : \ l \geq 1\Big) \quad (1.3)$$

with the state $p^{-1} E\,\mathrm{Tr}$ converges (see Definition 4.8 for notion of convergence of such spaces). This convergence entails showing convergence of expected normalised traces (moments). The joint moments of the polynomials in (1.3) are obtained using techniques from free probability theory. We describe the limits in terms of a semi-circle family (see (4.15)) and limits of the deterministic matrices (see Theorem 2.1 (a)).

As a consequence, we show that the LSD of any symmetric polynomial in (1.3) exists with probability 1 (see Theorem 2.1 (b)). In Remark 2.1, we provide Stieltjes transformation of these LSDs. As corollaries we derive all the existing results cited above in (i), (ii) and (iii) (see Corollaries 2.1 and 2.2). For comparable results in the case $p/n \to y \neq 0$, see [Marčenko and Pastur, 1967], [Liu et al., 2013], [Jin et al., 2014], [Bhattacharjee and Bose, 2014] and [Bhattacharjee and Bose, 2015].

As a specific application, consider the infinite dimensional vector linear process,

$$X^{(n)}_{t.p} = \sum_{j=0}^{q} \psi_{j,p} \varepsilon_{t-j.p}, \ \forall t = 1, 2, \ldots \text{ (almost surely)}, \quad (1.4)$$

where $X^{(n)}_{t.p}$ and $\varepsilon_{t.p} = (\varepsilon_{t,1}, \varepsilon_{t,2}, \ldots, \varepsilon_{t,p})'$ are p-dimensional vectors. $\{\psi_{j,p}\}$ are $p \times p$ matrices. Often we will omit the suffix p. The *population autocovariance matrices* are defined as

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

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

$$\hat{\Gamma}_{\tau,p} = \frac{1}{n}\sum_{t=\tau+1}^{n} X_{t,p}X^*_{(t-\tau),p}, \quad 0 \le \tau \le n-1. \tag{1.5}$$

[Bhattacharjee and Bose, 2014] studied the joint convergence of autocovariance matrices when $p/n \to y \in (0,\infty)$.

For the case $y = 0$, [Wang et al., 2014] provided the Stieltjes transformation for the LSD of the scaled and centered symmetrized sample autocovariance matrix (1.6) under quite strong conditions on $\{\psi_j\}$ matrices (see Corollary 2.3). However, unlike $(i)-(iii)$, they do not establish any connection of this LSD with the semicircle law.

$$\sqrt{np^{-1}}\left(2^{-1}(\hat{\Gamma}_{\tau,p} + \hat{\Gamma}^*_{\tau,p}) - \sum_{j=0}^{q-\tau}\psi_j\psi^*_{j+\tau}\right) \tag{1.6}$$

Note that the model, (1.4) can be written as

$$X^{(n)}_{t,p} = (\psi^{(n)}_{0,p}\ \ \psi^{(n)}_{1,p}\ \ \psi^{(n)}_{2,p}\ldots\psi^{(n)}_{q,p})(\varepsilon^*_{t,p}\ \ \varepsilon^*_{t-1,p}\ \ \varepsilon^*_{t-2,p}\ldots\varepsilon^*_{t-q,p})^*\ \ \forall t,n \ge 1.$$

Let $\{P_j\}$ be the sequence of $n \times n$ matrices where P_j has upper j-th diagonal to be 1 and 0 otherwise. For example, $P_0 = I_n$, the identity matrix of order n. Then it is easy to see that the LSD of (1.6) and

$$\sqrt{np^{-1}}(\sum_{j,j'=0}^{q}\psi_j Z Pk_{j-j'+i} Z^*\psi^*_{j'} - \sum_{j=0}^{q-\tau}\psi_j\psi_{j+\tau}) \tag{1.7}$$

are identical (see Subsection 3.4). Clearly (1.7) is a symmetric polynomial in $\mathcal{U}_n$ and hence comes under the ambit of our main theorem.

In Corollary 2.3, we provide the LSD of (1.6), for every τ, as a function of a semi-circle variable s and other variables (depending on the limit of the matrices ψ_j) which are free of s. We also derive their Stieltjes transformation as given in [Wang et al., 2014] but now under significantly weaker conditions on the ψ_j. In particular, the LSD of (1.6) are same for $\tau > q$ and different for $0 \le \tau \le q$. We then discuss how this observation can be used in model diagnosis.

Finally, it may be noted that due to the algebra convergence in our main theorem, much more generally than the above corollary, we are also able to claim that the LSD of any properly centered and scaled symmetric (additive, multiplicative or combinations thereof) polynomial of autocovariance matrices exists with probability 1 (see Remark 2.2).

2 Main results

We list below all the assumptions that are required for our main theorem. Let $Z_u = ((\varepsilon_{u,i,j}))_{p\times n}$, $u \ge 1$ be $p \times n$ random matrices which satisfy the following. Assumption

$(A2)$ is replaced by an alternative assumption later for some corollaries and applications.

(A1) $\{\varepsilon_{u,i,j} : u, i, j \geq 1\}$ are independently distributed with $E(\varepsilon_{u,i,j}) = 0$, $E|\varepsilon_{u,i,j}|^2 = 1$ and $\sup_{u,i,j} E|\varepsilon_{u,i,j}|^4 < \infty$.

(A2) For some $\eta > 0,\ 0 < \delta \leq 2$, $P(|\varepsilon_{u,i,j}| < \eta n^{\frac{1}{2+\delta}}) = 1, \forall 1 \leq i \leq 2n,\ 1 \leq j \leq p$.

If $U = 1$, we will write $\varepsilon_{i,j}$ and Z respectively for $\varepsilon_{1,i,j}$ and Z_1.

As discussed in Section 1, depending on where the matrices appear in the polynomial, we need different sets of assumptions. Assumption $(A3)$ ensures that $\{B_{2i-1}\}$ converge as elements of the non-commutative $*$-probability space of matrices with the state as the normalised trace. Assumption $(A4)$ requires the matrices $\{B_{2i}\}$ to converge only in the first and second (joint) moments.

(A3) $\{B_{2i-1} : i \geq 1\}$ are $p \times p$ matrices which are compactly supported and jointly $*$-converge.

(A4) $\{B_{2i} : i \geq 1\}$ are $n \times n$ matrices with bounded spectral norms. For all $i, i' \geq 1$, we have

$$\lim_{n\to\infty} n^{-1}\mathrm{Tr}(B_{2i}) < \infty, \quad \lim_{n\to\infty} n^{-1}\mathrm{Tr}(B_{2i}B_{2i'}) < \infty. \tag{2.1}$$

(A5) $p, n(p) \to \infty$ such that $y_n = pn^{-1} \to 0$.

Consider the free product of the non commutative $*$-probability spaces,

$$\begin{aligned} \mathcal{A} &= \mathrm{Span}\{b_{2i-1}, b_{2i-1}^* : i \geq 1\}, \text{ and} \qquad (2.2)\\ \mathcal{T}_u &= \{s_{l,j,u} : j = 1, 2, \ldots, k_l, l = 1, 2, \ldots, r\},\ u \geq 1 \qquad (2.3)\end{aligned}$$

with state φ such that, by Assumption $(A3)$, for any monomial $m(B_{2i-1}, B_{2i-1}^* : \forall i \geq 1)$,

$$\begin{aligned} \varphi\left(m(b_{2i-1}, b_{2i-1}^* : \forall i \geq 1)\right) &= \lim p^{-1}\mathrm{Tr}\left(m(B_{2i-1}, B_{2i-1}^* : \forall i \geq 1)\right), \qquad (2.4)\\ \varphi(s_{l_1,j_1,t}, s_{l_2,j_2,t}) &= \lim n^{-1}\mathrm{Tr}(A_{l_1,2j_1}, A_{l_2,2j_2}),\ \forall j_1, j_2, l_1, l_2. \qquad (2.5)\end{aligned}$$

Recall $\mathcal{U}_n$ defined in (1.3) and $\{A_{l,2i-1} : 1 \leq i \leq k_l + 1, l \geq 1\} \in \{B_{2i-1}\}_{i\geq 1}$, $\{A_{l,2i} : 1 \leq i \leq k_l, l \geq 1\} \in \{B_{2i}\}_{i\geq 1}$. Let, for all $l \geq 1$ and $1 \leq j \leq k_l$,

$$\begin{aligned} a_{l,-j} &= \Big(\prod_{\substack{i=1\\ i\neq j}}^{k_l} \lim n^{-1}\mathrm{Tr}(A_{l.2i})\Big)\Big(\prod_{i=0}^{j-1} a_{l,2i+1}\Big),\\ c_{l,-j} &= \prod_{i=j}^{k_l} a_{l,2i+1},\\ \mathcal{U} &= \mathrm{Span}\Big(\sum_{j=1}^{k_l} a_{l,-j}s_{l,j,u_{l,j}}c_{l,-j} : \forall u_{l,j} \in \{1, 2, 3\ldots\},\ l \geq 1\Big).\end{aligned}$$

Now we state our main theorem on convergence of the non-commutative $*$-probability space $\mathcal{U}_n$ and on existence of LSD of any symmetric polynomials in $\mathcal{U}_n$.

Theorem 2.1. *Suppose Assumptions* $(A1)-(A5)$ *hold. Then*

(a) $(\mathcal{U}_n, Ep^{-1}Tr) \to (\mathcal{U}, \varphi)$ *and*

(b) *LSD of any self adjoint polynomial* $\mathbb{P}(y_n^{-1/2}(\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} - \mathbb{G}_l) : 1 \leq l \leq r)$ *in* $\mathcal{U}_n$ *exists with probability* 1 *and it is given by* $\mathbb{P}(\sum_{j=1}^{k_l} a_{l,-j}s_{l,j,u_{l,j}}c_{l,-j} : \forall u_{l,j},\ 1 \leq l \leq r)$.

Proof of Theorem 2.1 is given in Subsection 3.1.

Note that, by Assumptions $(A3)$ and $(A4)$, the self-adjoint elements in $\mathcal{U}$ are all bounded random variables and characterized by their moment sequences (see Definition 4.2). The Stieltjes transform (see (4.3)) of the random variables in $\mathcal{U}$ can be described in terms of certain formulae.

Remark 2.1. *Let* $\gamma = \sum_{j=1}^{r} a_j s_j c_j$ *be self-adjoint, where* $\{s_j : \ j = 1,2,\ldots,r\}$ *is a semi-circle family (see (4.15)) with covariances*

$$\varphi(s_{j_1}s_{j_2}) = b_{j_1 j_2},\ \forall j_1, j_2 = 1,2,\ldots,r \tag{2.6}$$

and $\{a_j : j = 1,2,\ldots,r\}$ *is some permutation of* $\{c_j : j = 1,2,\ldots,r\}$. *Further* $\{s_j : \ j = 1,2,\ldots,r\}$ *and* $\{a_j, a_j^*, c_j, c_j^* : j = 1,2,\ldots r\}$ *are freely independent. Then the Stieltjes transformation of* γ *is given by*

$$G_\gamma(z) = -\varphi((z+\beta(z,a))^{-1}),\ \forall z \in \mathbb{C}^+, \text{ where} \tag{2.7}$$

$$\beta(z,a) = -\sum_{j_1,j_2=1}^{r} b_{j_1 j_2} c_{j_1} a_{j_2} \varphi(a_{j_1} c_{j_2}(z+\beta(z,a))^{-1}),\ \forall z \in \mathbb{C}^+. \tag{2.8}$$

We now give some corollaries which follow from Theorems 2.1 and Remark 2.1. In particular, they verify some existing results. In the following corollaries, we relax Assumption $(A2)$ and instead consider that for some $\eta > 0$, $0 < \delta \leq 2$,

$$\frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{2n} E\left(|\varepsilon_{ij}|^{2+\delta} I\left(|\varepsilon_{ij}| > \eta n^{\frac{1}{2+\delta}}\right)\right) = o(1). \tag{2.9}$$

This relaxation of assumption is not for all polynomials in $\mathcal{U}_n$, but only for some specific cases as considered in following corollaries.

Assumption (2.9) holds for i.i.d. random variables $\{\varepsilon_{i,j}\}$ with $0 < \delta < 2/3$ as for some $C, C_1, C_2 > 0$, we have

$$\begin{aligned}
\frac{1}{p^2}\sum_{i=1}^{p}\sum_{j=1}^{2n} E\left(|\varepsilon_{ij}|^{2+\delta} I\left(|\varepsilon_{ij}| > \eta n^{\frac{1}{2+\delta}}\right)\right) &\leq C\frac{n}{p}E\left(|\varepsilon_{ij}|^{2+\delta} I\left(|\varepsilon_{ij}| > \eta n^{\frac{1}{2+\delta}}\right)\right) \\
&\leq C\frac{n}{p}E\left(|\varepsilon_{i,j}|^4\right)^{\frac{2+\delta}{4}}\left(P(|\varepsilon_{i,j}|^4 > n^{\frac{4}{2+\delta}})\right)^{\frac{2-\delta}{4}} \\
&\leq C_1\frac{n}{p}\left(P(|\varepsilon_{i,j}|^4 > n)\right)^{\frac{2-\delta}{4}} \leq C_2\frac{n}{pn^{\frac{3(2-\delta)}{4}}} \\
&= o(1).
\end{aligned} \tag{2.10}$$

Corollary 2.1. (*a*) *Suppose Assumptions* ($A1$)*,* ($A5$) *and (2.9) hold. Then LSD of* $\sqrt{np^{-1}}(n^{-1}ZZ^* - I_p)$ *exists almost surely and is given by the standard semi-circle law.*

(*b*) *Suppose* $\{\varepsilon_{i,j} : i, j = 1, 2, \ldots\}$ *are independently and identically distributed with mean* 0*, variance* 1 *and* $E|\varepsilon_{i,j}|^4 < \infty$*. Then LSD of* $\sqrt{np^{-1}}(n^{-1}ZZ^* - I_p)$ *exists almost surely and is given by the standard semi-circle law. This verifies the Theorem given on page* 864 *of [Bai and Yin, 1988].*

In Corollary 2.3, we have stated a result more general than Corollary 2.1. Therefore, the proof of Corollary 2.1 follows from the proof of Corollary 2.3.

Corollary 2.2. *Consider the following assumptions.*

(A) *Let A be a $p\times p$ compactly supported positive semi-definite Hermitian matrix whose LSD exists. Suppose* $(Span\{A^{1/2}\}, p^{-1}Tr) \to (Span\{a^{1/2}\}, \varphi)$*.*

(B) *Let B be an $n\times n$ Hermitian matrix with bounded spectral norm and* $d_k := \lim n^{-1}Tr(B^k) < \infty,\ \forall k = 1, 2.$

(a) Suppose Assumptions (A)*,* (B)*,* ($A1$)*,* ($A5$) *and (2.9) hold. Then*

$$\left(Span\{\sqrt{np^{-1}}(n^{-1}A^{1/2}ZBZ^*A^{1/2} - n^{-1}Tr(B)A)\}, p^{-1}Tr\right) \to \left(Span\{a^{1/2}sa^{1/2}\}, \varphi\right), \tag{2.11}$$

where s is a semi-circle variable, $Span\{a\}$ *and* $Span\{s\}$ *are freely independent and* $\varphi(s^2) = d_2$*.*

(*b*) *Suppose* $\{\varepsilon_{i,j} : i, j = 1, 2, \ldots\}$ *are independently and identically distributed with mean* 0*, variance* 1 *and* $E|\varepsilon_{i,j}|^4 < \infty$*. We also assume* ($A$) *and* ($B$)*. Then also (2.11) holds.*

(*c*) *Suppose all the assumptions in* (*a*) *or* (*b*) *hold. Then the LSD of* $\sqrt{np^{-1}}(n^{-1}A^{1/2}ZBZ^*A^{1/2} - n^{-1}Tr(B)A)\}$ *exists almost surely. It is distributed as* $a^{1/2}sa^{1/2}$ *with Stieltjes transformation*

$$\begin{aligned} G_{a^{1/2}sa^{1/2}}(z) &= -\varphi((z + ad_2\beta(z))^{-1}),\ \forall z \in \mathbb{C}^+,\ \textit{where} \\ \beta(z) &= -\varphi(a(z + ad_2\beta(z))^{-1}),\ \forall z \in \mathbb{C}^+. \end{aligned}$$

(*d*) *If* $d_2 = 1$*, then* $a^{1/2}sa^{1/2}$ *can be described in the following equivalent ways:*

(*i*) $G_{a^{1/2}sa^{1/2}}(z) = \frac{1}{\sqrt{-1-zG_{a^{1/2}sa^{1/2}}(z)}} G_a\left(\frac{z}{\sqrt{-1-zG_{a^{1/2}sa^{1/2}}(z)}}\right)$

(*ii*) $\mathcal{R}_{a^{1/2}sa^{1/2}}(z) = ((\mathcal{R}_a \boxtimes \mathcal{R}_a \boxtimes Zeta) \circ Sq)(z)$

(*iii*) $S_{a^{1/2}sa^{1/2}a^{1/2}sa^{1/2}}(z) = (S_a(z))^2 S_{s^2}(z) = (1+z)^{-1}(S_a(z))^2.$

See (4.17), (4.8), (4.19), (4.21) and (4.18) respectively for the definitions of the transforms $\mathcal{R}$*,* S*, Zeta,* $\boxtimes$*, Sq and the composition* $\circ$*.*

Theorem 2.1 of [Wang and Paul, 2014] is precisely Corollary 2.2 (*c*) under the assumptions given in Corollary 2.2 (*b*).

2.1 Application in high-dimensional time series

As an application of Theorem 2.1 in high-dimensional statistics, we consider the infinite dimensional vector linear process of order q defined in (1.4). The sequence of autocovariance matrices $\{\Gamma_\tau\}_{\tau\geq 0}$, defined in (1.5), is crucial in time series analysis. In this subsection, we study the joint large sample behaviour of the autocovariance matrices when $p/n \to 0$. Similar results in the case $p/n \to y \in (0,\infty)$ are given in [Bhattacharjee and Bose, 2014].

Consider the following assumptions.

(T) Suppose $(\text{Span}\{\psi_j,\psi_j^* : j=0,1,2,\ldots,q\}, p^{-1}\text{Tr}) \to (\text{Span}\{\gamma_j,\gamma_j^* : j=0,1,2,\ldots,q\},\varphi)$ i.e., for any monomial $m(\psi_j,\psi_j^* : j=0,1,2,\ldots)$, we have

$$\lim p^{-1}\text{Tr}(m(\psi_j,\psi_j^* : j=0,1,2,\ldots,q)) = \varphi(m(\gamma_j,\gamma_j^* : j=0,1,2,\ldots,q)). \quad (2.12)$$

Let $\{s_{j_1,j_2,\tau} : j_1, j_2 = 1,2,\ldots,q\}$ be a semicircle family (see (4.15)) with covariances

$$\begin{aligned}\varphi(s_{j_1,j_2,\tau}s_{k_1,k_2,\tau}) &= \frac{1}{2\pi}\int_0^{2\pi}\cos(j_1-j_2+\tau)\theta\ \cos(k_1-k_2+\tau)\theta\ d\theta \\ &= \begin{cases}0.5, \text{ if } j_1-j_2=k_1-k_2 \text{ or } j_1-j_2+k_1-k_2=-2\tau \\ 0, \text{ otherwise.}\end{cases}\end{aligned} \quad (2.13)$$

Moreover, $\{s_{j_1,j_2,\tau}\}$ and $\{\gamma_j\}$ are free.

Corollary 2.3. *Suppose Assumptions* $(A1)$*,* $(A5)$*,* (T) *and (2.9) hold. Then*

(a) *the LSD* g_τ *of* $\sqrt{np^{-1}}(2^{-1}(\hat{\Gamma}_\tau + \hat{\Gamma}_\tau^*) - \sum_{j=0}^{q-\tau}\psi_j\psi_{j+\tau}^*)$ *exists almost surely, and*

(b) (i) *and* (ii) *are equivalent ways of describing the LSD* g_τ*.*

(i)

$$g_\tau \stackrel{\mathcal{D}}{=} \sum_{j_1,j_2=1}^{q}\gamma_{j_1}s_{j_1,j_2,\tau}\gamma_{j_2}^*. \quad (2.14)$$

(ii) *Stieltjes transformation of* g_τ *is given by,*

$$\begin{aligned}G_{g_\tau}(z) &= -\int\frac{dF(x)}{z+\beta_\tau(z,x)},\ z\in\mathbb{C}^+,\ \textit{where} && (2.15)\\ \beta_\tau(z,x) &= -\int\frac{R_\tau(x,y)dF(y)}{z+\beta_\tau(z,x)},\ z\in\mathbb{C}^+, x\in\mathbb{C}^q, && (2.16)\\ R_\tau(x,y) &= \frac{1}{2\pi}\int_0^{2\pi}\cos^2(\tau\theta)\Psi(x,\theta)\Psi(y,\theta)d\theta,\ x,y\in\mathbb{C}^q,\ \textit{and} && (2.17)\\ \Psi(x,\theta) &= |\sum_{l=0}^{q}x_l e^{il\theta}|^2,\ x=(x_1,\ldots,x_q)\in\mathbb{C}^q, 0<\theta\leq 2\pi, && (2.18)\end{aligned}$$

where F *is the joint distribution function of* $\{\gamma_l,\gamma_l^* : l=0,1,2,\ldots\}$*.*

(c) *Suppose* $\{\varepsilon_{i,j} : i,j=1,2,\ldots\}$ *are independently and identically distributed with mean* 0*, variance* 1 *and* $E|\varepsilon_{i,j}|^4<\infty$*. Then (2.14) and (2.15) continue to hold.*

Proof of Corollary 2.3 is given in Subsection 3.4.

It may be noted that [Wang et al., 2014] established (2.15)-(2.18), under the same assumptions on $\{\varepsilon_{i,j}\}$ as considered in Corollary 2.3 (c) and the following assumption on $\{\psi_j\}$:

(A) There exists U such that $U^*\psi_l U = \text{diag}(f_l(\alpha_j))_{j=1}^p$, where $f_1, f_2, \ldots, f_q$ are bounded continuous functions. As $p \to \infty$, the ESD of $\{\alpha_1, \alpha_2, \ldots, \alpha_p\}$ converges to a compactly supported distribution F.

Note that Condition (A) trivially implies Condition (T).

Remark 2.2. *Suppose* $(A1)$*,* $(A2)$*,* $(A5)$ *and* (T) *hold. Then the limiting spectral distribution of any properly centered and scaled symmetric polynomial of autocovariance matrices exists with probability* 1.

Proof of Remark 2.2 is similar to the proof of Corollary 2.3 and hence we omit the details.

Model diagnostics Corollary 2.3 can be used for diagnostics. While we hope to deal with this elsewhere in details, we make some brief remarks.

There is a huge literature on diagnosis of the appropriate order of a univariate time series model (see for example, Ljung and Box [1978], McLeod [1978], Katayama [2008], Hong [1996], Shao [2011]). The same problem in multivariate set up has also been studied (see Francq and Raïsi [2007], Hosking [1980], Hosking [1981a], Hosking [1981b], Lin and McLeod [1981], Lin and McLeod [2006], Mahdi and McLeod [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. Similar simulation results for the case $p/n \to 0$ are given in Section 3 of [Bhattacharjee and Bose, 2014].

If $X_t \sim$ MA(q) process, then g_τ's are all identical for $\tau > q$. This follows from the fact that no $0 \leq j_1, k_1, j_2, k_2 \leq q$ satisfies $j_1 - j_2 + k_1 - k_2 = -2\tau$ for $\tau > q$ (see (2.13)). Moreover, g_τ's are all different for $0 \leq \tau \leq q$. Therefore, for a given large data set, we can check different even order moments of the ESD of $\sqrt{np^{-1}}(2^{-1}(\hat{\Gamma}_\tau + \hat{\Gamma}_\tau^*) - \sum_{j=0}^{q-\tau} \psi_j \psi_{j+\tau}^*)$ for $\tau = 1, 2, 3, \ldots$. If there is a $q \geq 0$ such that for a fixed $r \geq 1$, the r-th order moments are identical for all $\tau > q$ 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 $q = 0, 1, 2, 3$, where all the coefficient matrices are I_p, the identity matrix of order p, i.e.,

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

We also let $p = \sqrt{n}$. For each fixed $q = 0, 1, 2, 3$ and $r = 2, 4$, we plot the r-th order moments of $\sqrt{np^{-1}}(2^{-1}(\hat{\Gamma}_\tau + \hat{\Gamma}_\tau^*) - \sum_{j=0}^{q-\tau} \psi_j \psi_{j+\tau}^*)$ for $\tau = 1, 2, \ldots, 8$. It is observed from the figures that, for each $r = 2, 4$, the r-th order moments are more or less same for each $\tau = q + 1, q + 2, \ldots, 8$, when X_t is an MA(q) process, $q = 0, 1, 2, 3$.

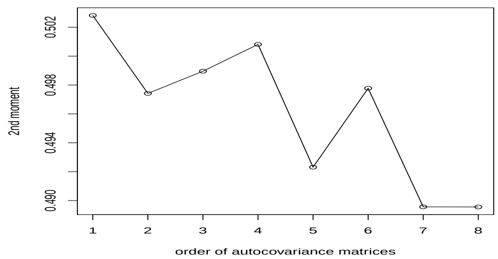

Figure 4.1*:* *MA*(0) *process:* 2*nd moment*

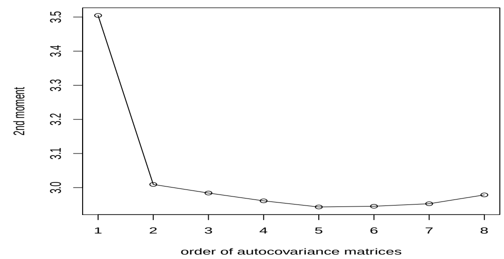

Figure 4.2 *MA*(1) *process:* 2*nd moment*

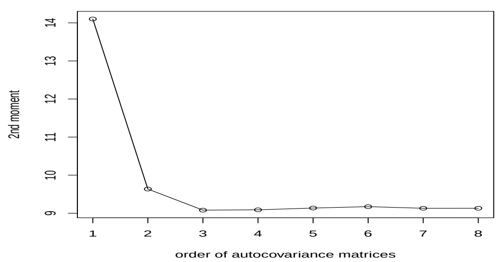

Figure 4.3*:* *MA*(2) *process:* 2*nd moment*

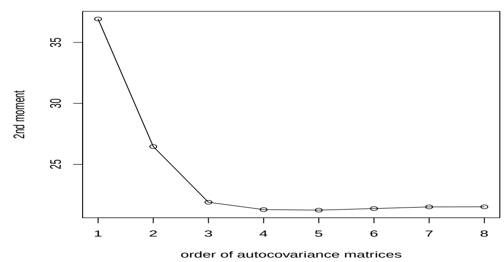

Figure 4.4 *MA*(3) *process:* 2*nd moment*

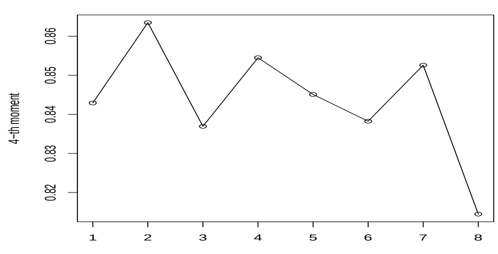

Figure 4.5*:* *MA*(0) *process:* 4*th moment*

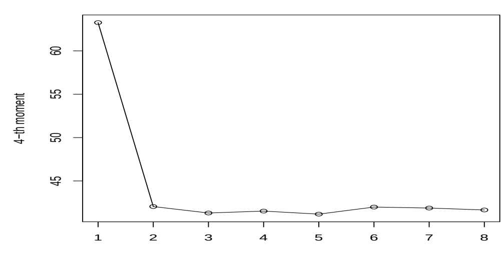

Figure 4.6 *MA*(1) *process:* 4*th moment*

3 Proofs

In this section, we prove Theorem 2.1 and the remarks and corollaries stated in Section 2.

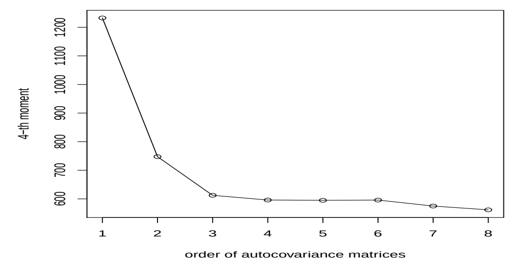

Figure 4.7: *MA*(2) *process:* 4*th moment*

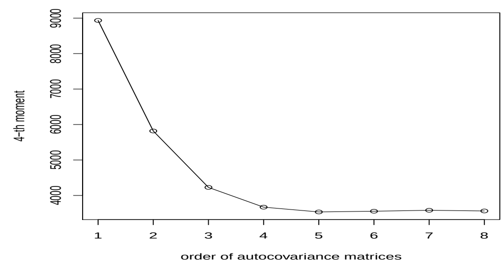

Figure 4.8 *MA*(3) *process:* 4*th moment*

3.1 Proof of Theorem 2.1

Proof of Theorem 2.1 (*a*)*:* First assume $U = 1$. Then it is enough to show that, for $1 \leq T \leq r$ and $1 \leq l_1 < l_2 < \ldots < l_T \leq r$,

$$\lim y_n^{-T/2} E\mathrm{Tr}\left(\prod_{t=1}^{T}(\mathbb{P}_{l_t} - \mathbb{G}_{l_t})\right) \to \varphi\left(\prod_{t=1}^{T}\left(\sum_{j=1}^{k_{l_t}} a_{l_t,-j} s_{l_t,j} c_{l_t,-j}\right)\right). \tag{3.1}$$

For convenience of notation, we verify (3.1) only for the case $T = r$. For other values of T, the arguments are similar.

Fix $k \geq 1$. Let $B_1, B_3, \ldots, B_{2k+1}$ be $(k+1)$ matrices of order $p \times p$, which are compactly supported and jointly converge. Also let $B_2, B_4, \ldots, B_{2k}$ be k norm bounded matrices of order $n \times n$. Let $B_l(i, j)$ be the (i, j)-th element of matrix B_l. Then, we have

$$\begin{aligned} & \lim \frac{E}{n^k p} \mathrm{Tr}(B_1 Z B_2 Z^* B_3 Z \ldots Z^* B_{2k+1}) \qquad (3.2)\\ = \ & \lim \frac{E}{n^k p} \sum_{\substack{i_1, i_2, \ldots, i_{2k+1} \\ j_1, j_2, \ldots, j_{2k}}} B_1(i_1, i_2)\varepsilon_{i_2,j_1} B_2(j_1, j_2)\varepsilon_{i_3,j_2} B_3(i_3, i_4) \ldots \varepsilon_{i_{2k+1},j_{2k}} B_{2k+1}(i_{2k+1}, i_1). \end{aligned}$$

Note that the left side of (3.1) involves limits similar to (3.2). Lemma 3.1 provides an upper bound for these limits.

To have a non-zero contribution from any summand, by independence, each index in the set $\{(i_2, j_1), (i_3, j_2), \ldots, (i_{2k+1}, j_{2k})\}$ must be matched. We thus need to study the partitions of a set of size $2k$, say, $\{1, 2, \ldots, 2k\}$. For a particular partition, the indices are matched within blocks and are not matched between blocks. For example, $\{\{1, 2\}, \{3, 4\}, \ldots, \{2k-1, 2k\}\}$ indicates $(i_2, j_1) = (i_3, j_2), (i_4, j_3) = (i_5, j_4), \ldots, (i_{2k}, j_{2k-1}) = (i_{2k+1}, j_{2k})$. Let,

$$\begin{aligned} \mathcal{P}_{2k} &= \text{set of all partitions of } \{1, 2, \ldots, 2k\} \text{ having no singleton block}, \\ \mathrm{NC}(2k) &= \text{set of all non-crossing partitions of } \{1, 2, \ldots, 2k\}, \\ \mathrm{NC}_2(2k) &= \text{set of all non-crossing pair partitions of } \{1, 2, \ldots, 2k\}. \end{aligned}$$

Note that, as $E(\varepsilon_{i,j}) = 0 \ \forall i, j$, here only $\mathcal{P}_{2k}$ contributes. The following lemma provides us an upper bound for the contribution of any partition in $\mathrm{NC}(2k) \cap \mathcal{P}_{2k}$. Recall that $y_n = pn^{-1}$.

Lemma 3.1. *Suppose Assumption* (*A*2) *holds. Consider any non-crossing partition* σ *in* $NC(2k) \cap \mathcal{P}_{2k}$ *having* K_i *many blocks of size* $i \geq 2$ *of which let* $K_{i,1}$ *and* $K_{i,2}$ *start with odd and even indices respectively. Then for some* $C > 0$, $C_n \downarrow 0$, *we have*

$$\begin{aligned} & |\text{Contribution of } \sigma \text{ in } \frac{E}{n^k p} Tr(A_1 Z A_2 Z^* A_3 Z \ldots Z^* A_{2k+1})| \\ \leq & \begin{cases} C y_n^{K_{2,2}}, & \text{if } K_i = 0,\ \forall i \geq 3 \\ C_n y_n^{K_{2,2}+0.5\sum_{i\geq 2} K_{2i-1,1} + \sum_{i\geq 2}(K_{2i,1}+K_{2i-1,2}) + 2\sum_{i\geq 2} K_{2i,2}}, & \text{otherwise.} \end{cases} \end{aligned} \tag{3.3}$$

Proof. Since σ is a non-crossing partition, it always has a block with adjacent indices. Further, if we drop that block and rename the indices to the right of the dropped block by increment 1, then again we have a block with adjacent idices.

Case 1: Suppose σ has at least one even sized block with adjacent elements. Note that this summand is of either of the following forms:

(I) it starts with an odd index and is of size $2(s-j)$ i.e. of the form $n^{-(s-j)} D_1 Z A_{2j} Z^* A_{2j+1} Z A_{2j+2} Z^* \ldots A_{2s} Z^* D_2$, and

(II) it starts with an even index and is of size $2(s-j)$ i.e. of the form $n^{-(s-j)} E_1 Z^* A_{2j+1} Z A_{2j+2} Z^* \ldots A_{2s+1} Z E_2$.

The matrices above satisfy the following condition:

(G) D_1, D_2 and E_1, E_2 may involve Z, Z^*'s. All Z, Z^*'s in $Z A_{2j} Z^* A_{2j+1} Z A_{2j+2} Z^* \ldots A_{2s} Z^*$ and $Z^* A_{2j+1} Z A_{2j+2} Z^* \ldots A_{2s+1} Z$ are matched with each other and are not matched with any Z or Z^* in D_1, D_2, E_1, E_2.

Now in (I), by Assumptions (A2), (A3) and (A4),

$$\begin{aligned} & n^{-(s-j)} |\frac{E}{p} \mathrm{Tr}(D_1 Z A_{2j} Z^* A_{2j+1} Z A_{2j+2} Z^* \ldots A_{2s} Z^* D_2)| \\ \leq\ & C |\frac{E}{p} \mathrm{Tr}(D_1 D_2)| \left(I(s-j=1) + \frac{n^{s-j-2}}{n^{s-j-1}} I(s-j>1) \right) \\ \leq\ & \begin{cases} C\, |\frac{E}{p} \mathrm{Tr}(D_1 D_2)|, & \text{if } s-j=1 \\ C_n\, |\frac{E}{p} \mathrm{Tr}(D_1 D_2)|\, y_n, & \text{if } s-j>1. \end{cases} \end{aligned} \tag{3.4}$$

Also, in (II), similarly

$$\begin{aligned} & n^{-(s-j)} |\frac{E}{p} \mathrm{Tr}\left(E_1 Z^* A_{2j+1} Z A_{2j+2} Z^* \ldots A_{2s+1} Z E_2 \right)| \\ \leq\ & C |\frac{E}{p} \mathrm{Tr}(E_1 E_2)| y_n \left(I(s-j=1) + \frac{n^{s-j-2}}{n^{s-j-1}} I(s-j>1) \right) \\ \leq\ & \begin{cases} C\, |\frac{E}{p} \mathrm{Tr}(E_1 E_2)|\, y_n, & \text{if } s-j=1 \\ C_n\, |\frac{E}{p} \mathrm{Tr}(E_1 E_2)|\, y_n^2, & \text{if } s-j>1. \end{cases} \end{aligned} \tag{3.5}$$

Case 2**.** Suppose σ does not have any even sized block with adjacent elements. Since it is non-crossing, it should have at least one odd sized block with adjacent elements. These summands are of one of the following forms:

$$\begin{array}{ll} (I) & n^{-(s-j)-0.5} D_1 Z A_{2j} Z^* A_{2j+1} Z \dots A_{2s+1} Z D_2 \\ (II) & n^{-(s-j)-0.5} E_1 Z^* A_{2j+1} Z \dots B_{2s+1} Z^* E_2. \end{array}$$

Here also D_1, D_2, E_1, E_2, $ZA_{2j}Z^*A_{2j+1}Z \dots A_{2s+1}Z$ and $Z^*A_{2j+1}Z \dots B_{2s+1}Z^*$ satisfy (G) as in Case 1. Now note that

$$\begin{aligned} & n^{-(s-j)-0.5} \, |\frac{E}{p}\mathrm{Tr}(D_1 Z A_{2j} Z^* A_{2j+1} Z \dots A_{2s+1} Z D_2)| \\ \leq & \sqrt{\frac{E}{p}\mathrm{Tr}(D_2 D_1 D_1' D_2') n^{-2s+2j-1} p^{-1} \mathrm{Tr}(Z A_{2j} Z^* \dots A_{2s+1} Z Z^* A_{2s+1}^* \dots A_{2j}^* Z^*)}, \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} & n^{-(s-j)-0.5} \, |\frac{E}{p}\mathrm{Tr}(E_1 Z^* A_{2j+1} Z \dots A_{2s+1} Z^* E_2)| \\ \leq & \sqrt{\frac{E}{p}\mathrm{Tr}(E_2 E_1 E_1' E_2') n^{-2s+2j-1} p^{-1} \mathrm{Tr}(Z^* A_{2j+1} Z \dots A_{2s+1} Z^* Z A_{2s+1}^* \dots A_{2j+1}^* Z)}. \end{aligned} \tag{3.7}$$

and both $D_2D_1D_1'D_2'$, $E_2E_1E_1'E_2'$ have at least one even sized block with adjacent elements. So we are back to Case 1. Now, the proof of the lemma is easily completed by a repeated application of (3.4)- (3.7). □

To prove (3.1), One of the terms that we shall need to compute is $\lim y_n^{-r/2}\frac{E}{p}\mathrm{Tr}(\prod_{l=1}^r \mathbb{P}_l)$. Therefore, we need to deal with set of all partitions of $\{1, 2, \dots, 2\sum_{l=1}^r k_l\}$. For this purpose, we introduce some notation and terminology in the following subsection.

3.1.1 Useful discussion on partitions

For any poset S, let

$$\begin{aligned} \mathrm{NC}(S) &= \text{set of all non-crossing partitions of } S, \\ \mathrm{NC}_2(S) &= \text{set of all non-crossing pair partitions of } S, \\ \mathcal{P}(S) &= \text{set of all partitions of } S \text{ having no singleton block.} \end{aligned}$$

Parts: We define

$$B_l = \{2k_{l-1}+1, 2k_{l-1}+2, \dots, 2k_l\}, \; k_0 = 0, \; \forall l = 1, 2, \dots r. \tag{3.8}$$

We call B_l to be the l-th part and there are r such parts.

***Type* 1, *Type* 2 partitions and disjoint decomposition of $\mathcal{P}(B_l)$:** We call a partition in $NC_2(B_l)$ to be of *Type* 1 if all its blocks are of the form (odd, even). Thus a typical *Type* 1 partition is of the form $\{(2k_{l-1}+1, 2k_{l-1}+2), (2k_{l-1}+3, 2k_{l-1}+4), \ldots, (2k_l - 1, 2k_l)\}$.

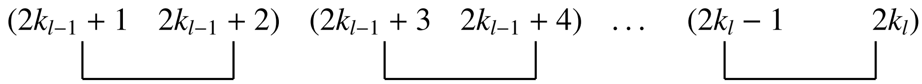

Figure 1: *Type* 1 partition in B_l

It is denoted by $\mathcal{T}_1(B_l)$. Any partition in $\mathcal{T}_2(B_l) := \mathcal{P}(B_l) - \mathcal{T}_1(B_l)$ is called a Type 2 partition in B_l. Clearly, $\mathcal{P}(B_l) = \mathcal{T}_1(B_l) \cup \mathcal{T}_2(B_l)$ is a disjoint decomposition of $\mathcal{P}(B_l)$.

Next we consider the disjoint decomposition of $\mathcal{P}(\cup_{l=1}^r B_l)$. For this purpose, we consider the following partitions.

$\mathcal{P}^*$ partitions: For all $1 \le T \le r$ and $1 \le l_1 < l_2 < \ldots < l_T \le r$, we define

$$\begin{aligned}\mathcal{P}^*(B_{l_1}, B_{l_2}, \ldots, B_{l_T}) \quad = \quad & \text{set of all partitions } \sigma \in \mathcal{P}(\cup_{t=1}^T B_{l_t}) \text{ such that} \\ & \sigma \vee \{B_{l_1}, B_{l_2}, \ldots, B_{l_T}\} = 1_{\cup_{t=1}^T B_{l_t}} \end{aligned} \tag{3.9}$$

That means any partition in $\mathcal{P}^*(B_{l_1}, B_{l_2}, \ldots, B_{l_T})$ joins all the parts $B_{l_1}, B_{l_2}, \ldots, B_{l_T}$.

For example, let $l_1 = 1, l_2 = 2$ and $k_1 = k_2 = 4$ and hence $B_1 = \{1,2,3,4\}$, $B_2 = \{5,6,7,8\}$. Then in Figure 2, $\sigma_1 \in \mathcal{P}^*(B_1, B_2)$ whereas $\sigma_2 \notin \mathcal{P}^*(B_1, B_2)$.

Figure 2: $\mathcal{P}^*$ partitions

Restricted partitions on parts: Let σ be any partition in $\mathcal{P}(\cup_{l=1}^r B_l)$. Let $\sigma(B_l)$ denote the set of all such blocks in σ which together form B_l. This is called restricted partition of σ on the part B_l. Therefore, by definition, if there is at least one block in σ which consists of some elements from B_l itself and some elements from B_l^c, then $\sigma(B_l) = \phi$. For example, in Figure 2, $\sigma_1(B_1) = \phi, \sigma_1(B_2) = \phi$ and $\sigma_2(B_1) = \{\{1,4\},\{2,3\}\}, \sigma_2(B_2) = \{\{5,8\},\{6,7\}\}$.

$\mathcal{G}$ and $\mathcal{G}^*$ partitions: For all $1 \le T \le r$ and $1 \le l_1 < l_2 < \ldots < l_T \le r$, we define

$$\begin{aligned}\mathcal{G}_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^r B_l) \quad = \quad & \text{set of all partitions } \sigma \in \mathcal{P}(\cup_{l=1}^r B_l) \text{ such that} \\ & \sigma \vee \{B_1, B_2, \ldots, B_r\} = \{B_{l_1}, B_{l_2}, \ldots, B_{l_T}, 1_{\cup_{l \notin \{l_1,l_2,\ldots,l_T\}} B_l}\} \end{aligned} \tag{3.10}$$

that means $\sigma \in \mathcal{G}_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^r B_l) \iff \sigma(B_{l_t}) \in \mathcal{P}(B_{l_t}), \forall t = 1,2,\ldots,T$ and $\sigma(B_l) = \phi, \forall l \ne l_t, t = 1,2,\ldots,T$.

Here is an example. Let $r = 4$ and $k_1 = k_2 = k_3 = k_4 = 4$. Then $B_i = \{4i-3, 4i-2, 4i-1, 4i\}$, $\forall i = 1, 2, 3, 4$. Consider the following partitions.

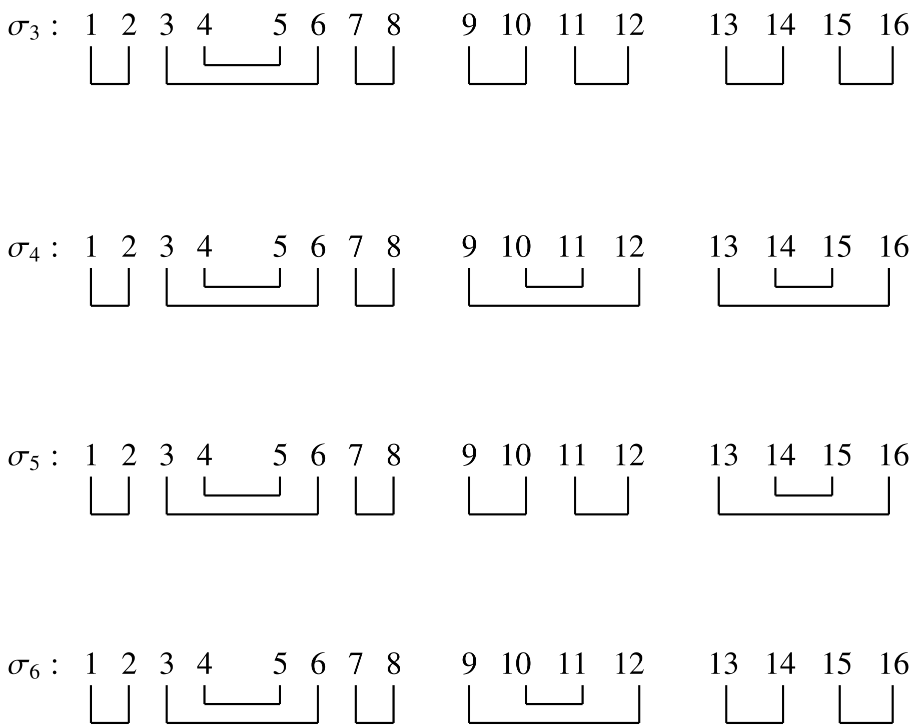

Figure 3: Some partitions in $\mathcal{G}_{3,4,r=4}(\cup_{l=1}^{4} B_l)$

Here $\sigma_i \in \mathcal{G}_{3,4,r=4}(\cup_{l=1}^{4} B_l), \forall i = 3, 4, 5, 6$.

Next, we define $\mathcal{G}^*_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^{r} B_l) \subset \mathcal{G}_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^{r} B_l)$ as follows.

$$\begin{aligned} \sigma \in \mathcal{G}^*_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^{r} B_l) \quad \Longleftrightarrow \quad & \sigma(B_{l_t}) \in \mathcal{T}_2(B_{l_t}),\ \forall t = 1, 2, \ldots, T \text{ and} \\ & \sigma(B_l) = \phi,\ \forall l \neq l_t,\ t = 1, 2, \ldots, T. \end{aligned} \tag{3.11}$$

Also, the complement of $\mathcal{G}^*$ is defined as

$$\mathcal{G}^{*c}_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^{r} B_l) = \mathcal{G}_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^{r} B_l) - \mathcal{G}^*_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^{r} B_l). \tag{3.12}$$

In Figure 3, $\sigma_4 \in \mathcal{G}^*_{3,4,r=4}(\cup_{l=1}^{4} B_l)$ and $\sigma_3, \sigma_5, \sigma_6 \in \mathcal{G}^{*c}_{3,4,r=4}(\cup_{l=1}^{4} B_l)$.

Disjoint decomposition of $\mathcal{P}(\cup_{l=1}^{r} B_l)$: Here, we have the following decomposition,

$$\begin{aligned} \mathcal{P}(\cup_{l=1}^{r} B_l) &= \Big(\bigcup_{\substack{1 \le l_1 < l_2 < \ldots < l_T \le r \\ 1 \le T \le r}} \mathcal{G}_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^{r} B_l) \Big) \bigcup \mathcal{P}^*(B_l : l = 1, 2, \ldots, r) \\ &= \Big(\bigcup_{\substack{1 \le l_1 < l_2 < \ldots < l_T \le r \\ 1 \le T \le r}} \mathcal{G}^*_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^{r} B_l) \Big) \bigcup \Big(\bigcup_{\substack{1 \le l_1 < l_2 < \ldots < l_T \le r \\ 1 \le T \le r}} \mathcal{G}^{*c}_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^{r} B_l) \Big) \bigcup \mathcal{P}^*(B_l : l = 1, 2, \ldots, r). \\ &= F_1 \cup F_2 \cup F_3, \text{ (say).} \end{aligned} \tag{3.13}$$

This decomposition will be used in the proof of (3.1). We will show that there are respectively negligible and no contributions of F_1 and F_2 in $y_n^{-r/2} E\mathrm{Tr}\,(\prod_{l=1}^{r}(\mathbb{P}_l - \mathbb{G}_l))$. Moreover, the only partitions from F_3 which contribute in $y_n^{-r/2} E\mathrm{Tr}\,(\prod_{l=1}^{r}(\mathbb{P}_l - \mathbb{G}_l))$, are characterized by the following B^*-partitions.

B^*-partition: For every $1 \le l_1 < l_2 \le r$ and $1 \le i \le k_{l_1}$, $1 \le j \le k_{l_2}$, we define

$$\begin{aligned}
\mathbb{B}_{i,j}(l_1,l_2) \;=\; & \Big\{\{2\sum_{l=1}^{l_1-1} k_l + 1, 2\sum_{l=1}^{l_1-1} k_l + 2\}, \ldots, \{2\sum_{l=1}^{l_1-1} k_l + 2i-3, 2\sum_{l=1}^{l_1-1} k_l + 2i-2\},\\
& \{2\sum_{l=1}^{l_1-1} k_l + 2i-1, 2\sum_{l=1}^{l_2-1} k_l + 2j\}, \{2\sum_{l=1}^{l_1-1} k_l + 2i, 2\sum_{l=1}^{l_2-1} k_l + 2j-1\},\\
& \{2\sum_{l=1}^{l_1-1} k_l + 2i+1, 2\sum_{l=1}^{l_1-1} k_l + 2i+2\}, \ldots, \{2\sum_{l=1}^{l_1} k_l - 1, 2\sum_{l=1}^{l_1} k_l\},\\
& \{2\sum_{l=1}^{l_2-1} k_l + 1, 2\sum_{l=1}^{l_2-1} k_l + 2\}, \ldots, \{2\sum_{l=1}^{l_2-1} k_l + 2j-3, 2\sum_{l=1}^{l_2-1} k_l + 2j-2\},\\
& \{2\sum_{l=1}^{l_2-1} k_l + 2j+1, 2\sum_{l=1}^{l_2-1} k_l + 2j+2\}, \ldots, \{2\sum_{l=1}^{l_2} k_l - 1, 2\sum_{l=1}^{l_2} k_l\}\Big\}
\end{aligned}$$

Clearly, the first, third, fourth and fifth lines are empty respectively for $i = 1, k_{l_1}, j = 1, k_{l_2}$. For all $1 \le l_1 < l_2 \le r$, we define

$$\mathbb{B}(l_1,l_2) \;=\; \{\mathbb{B}_{i,j}(l_1,l_2) : 1 \le i \le k_{l_1}, 1 \le j \le k_{l_2}\}.$$

For every $\sigma = \{V_1, V_2, \ldots, V_{r/2}\} \in \mathrm{NC}_2(r)$, we define

$$\mathcal{B}_{\sigma,r} \;=\; \mathbb{B}(V_1) \times \mathbb{B}(V_2) \times \ldots \times \mathbb{B}(V_{r/2}),$$

where " $\times$ " is the Cartesian product between sets and

$$B_r^* \;=\; \cup_{\sigma \in \mathrm{NC}_2(2T)} \mathcal{B}_\sigma \subset \mathcal{P}^*(B_l : l = 1, 2, \ldots, r). \tag{3.14}$$

Note that if r is odd, then $\mathrm{NC}_2(r) = \phi$ and hence $\mathcal{B}_r^* = \phi$. We shall see that if r is odd, $\lim y_n^{-r/2} E\mathrm{Tr}\,(\prod_{l=1}^{r}(\mathbb{P}_l - \mathbb{G}_l)) = 0$ and if r is even, partitions from B_r^* are only contributing partitions in $\lim y_n^{-r/2} E\mathrm{Tr}\,(\prod_{l=1}^{r}(\mathbb{P}_l - \mathbb{G}_l))$.

The following observations are immediate from Lemma 3.1.

(a) The partition, whose contribution is $\Omega(1)$ [1], has $K_{2,1} = k$ and $K_{2,2} = 0, K_i = 0,\ \forall i \ge 3$. Note that the only such partition (and it is of *Type* 1) is the above one. Therefore, for all $l = 1, 2, \ldots, r$, $\mathcal{T}_1(B_l)$ contributes $\Omega(1)$.

[1] $a_n = \Omega(b_n) \iff \{\frac{a_n}{b_n}\}$ is bounded but does not tend to 0.

(b) Also there is no partition in $\mathcal{P}_{2k}$ which can contribute $O(\sqrt{y_n})$. This is because, to get $1/2$ in power of y_n, we need partitions with $K_{2i'+1,1} = 1, K_{2i'+1,2} = 0$ for some i' and $K_{2i-1} = 0, \forall i \neq i'$, $K_{2i} = 0, \forall i$, but this is impossible. Therefore, contribution of $\mathcal{T}_2(B_l) \cap NC(B_l)$ in $E\text{Tr}\left(\prod_{l=1}^{r}(\mathbb{P}_l - \mathbb{G}_l)\right)$ is $O(y_n)$ from l-th part B_l. Now note that the partitions in $\mathcal{T}_2(B_l) - NC(B_l)$ i.e. all crossing partitions in $\mathcal{T}_2(B_l)$ can be obtained from the partitions in $\mathcal{T}_2(B_l) \cap NC(B_l)$ by swapping indices. Swapping of indices raises power of y_n in the corresponding term. Hence, contribution of $\mathcal{T}_2(B_l)$ in $E\text{Tr}\left(\prod_{l=1}^{r}(\mathbb{P}_l - \mathbb{G}_l)\right)$ is $O(y_n)$ from the l-th part B_l.

(c) Suppose $\sigma \in \mathcal{P}(\cup_{t=1}^{T} B_{l_t})$ be such that $\sigma \vee \{B_{l_1}, B_{l_2}, \ldots, B_{l_T}\} = \cup_{t=1}^{T} B_{l_t}$. To join any two parts, at least one block must start with an even index. Hence the contribution of σ in $E\text{Tr}\left(\prod_{l=1}^{r}(\mathbb{P}_l - \mathbb{G}_l)\right)$ is $O(y_n^{T-1})$.

(d) Contribution of $\mathcal{P}^*(B_l, l \neq l_1, l_2, \ldots, l_T)$ in $E\text{Tr}\left(\prod_{l=1}^{r}(\mathbb{P}_l - \mathbb{G}_l)\right)$ is $O(y_n^{(r-T)/2})$ when $r - T$ is even and $O(y_n^{(r-T+1)/2})$ when $r - T$ is odd.

(e) Hence, by (b), (d), contribution of $\mathcal{G}^*_{l_1,l_2,\ldots,l_T,r}(\cup_{l=1}^{r} B_l)$ in $E\text{Tr}\left(\prod_{l=1}^{r}(\mathbb{P}_l - \mathbb{G}_l)\right)$ is $O(y_n^{(r+T)/2})$. As a consequence, the contribution of F_1 is $o(y_n^{r/2})$.

(f) Let $\sigma \in \mathcal{P}^*(B_l, l = 1, 2, \ldots, r) = F_3$ be such that $\sigma \vee \{B_1, B_2, \ldots, B_r\} = \{V_1, V_2, \ldots, V_k\}$, where $V_i \neq B_j$, $\forall i, j$. Then by (c), contribution of σ is $O(y_n^{\sum_{j=1}^{k}(|V_j|-1)}) = O(y_n^{r-k})$. Here, the maximum possible value of k is $[r/2]$.

Therefore, if r is odd, the contribution of F_3 in $E\text{Tr}\left(\prod_{l=1}^{r}(\mathbb{P}_l - \mathbb{G}_l)\right)$ is $o(y_n^{r/2})$. If r is even, by Lemma 3.1, the set of partitions in F_3, which contributes $\Omega(y_n^{r/2})$ in $E\text{Tr}\left(\prod_{l=1}^{r}(\mathbb{P}_l - \mathbb{G}_l)\right)$ is B_r^*. Clearly, contribution of $F_3 - B_r^*$ is $o(y_n^{r/2})$.

To prove (3.1), we can now argue as follows. First, by (e), there is negligible contribution of F_1. Second, by (f), there is negligible contribution of $F_3 - B_r^*$. Finally, there is no contribution of F_2. To see this, consider any $\sigma \in F_2$, with exactly $1 \leq s \leq r$ many parts having *Type* 1 partitions. Then the number of times the partition σ appears in left side of (3.1) is $\sum_{j=0}^{s}(-1)^j\binom{s}{j} = 0$. Therefore, there is no contribution of F_2.
Hence, by (3.13), the proof of Theorem 2.1 (a) is complete for $U = 1$.

Now assume $U > 1$. Then not all partitions in B_r^* contribute. By independence of different Z_u's, a partition in B_r^* contributes if the matched Z, Z^* matrices have same u indices. This immediately implies Theorem 2.1 (a) for $U > 1$.

Proof of Theorem 2.1 (b)*:* First we prove a necessary lemma. Let

$$\Pi_i = p^{-1}\text{Tr}\Big(\prod_{j=1}^{t_i} y_n^{-1/2}(\mathbb{P}_{i_j} - G_{i_j})\Big),\ \Pi_i^0 = E(\Pi_i),\ \forall i = 1, 2, \ldots, T, \tag{3.15}$$

and $\{i_j : j = 1, 2, \ldots, t_i,\ i = 1, 2, \ldots, T\} \in \{1, 2, \ldots, r\}$.

Lemma 3.2. *Suppose* $(A1)-(A5)$ *hold. Then*

$$\sum_{\epsilon_i=0,1}(-1)^{T-(\sum_{i=1}^{T}\epsilon_i)}E\Big(\prod_{i=1}^{T}\Pi_i^{\epsilon_i}(\Pi_i^0)^{1-\epsilon_i}\Big)=\sum_{a=0}^{T}(-1)^{T-a}\sum_{\substack{\epsilon_i=0,1,\ \forall i\\ \sum_{i=1}^{T}\epsilon_i=a}}E\Big(\prod_{i=1}^{T}\Pi_i^{\epsilon_i}(\Pi_i^0)^{1-\epsilon_i}\Big)=O(p^{-T/2}).$$

Proof. Recall the definition of parts in (3.8). Let us call $G_i=\cup_{j=1}^{t_i}B_{i_j}$ to be the i-th group, $i=1,2,\ldots,T$. Using the same arguments as above to show negligibility of F_2, any partition σ of $\cup_{i=1}^{T}G_i$ with $\sigma(G_i)=G_i$ for at least one i has no contribution to the above sum. Moreover, similar arguments as (f) above show that the contribution of the partition σ of $\cup_{i=1}^{T}G_i$ with $\sigma(G_i)=\phi,\ \forall i$ is $O(p^{-T/2})$. Hence the proof of the lemma is complete. □

To prove Theorem 2.1 (b), we use the moment method (see Lemma 4.1). Condition $(M1)$ follows from Theorem 2.1 (a). Conditions $(M2)$ and $(M4)$ follow by applying Lemma 3.2 for $T=2$ and 4 respectively. Condition (C) follows from the fact that, for any bounded random variables $a_1,a_2,\ldots$ and semi-circle family $\{s_1,s_2,\ldots,s_{2h}\}$ and for some $C,C'>0$

$$|E(a_1s_1a_2s_2a_3s_3,\ldots,a_{2h}s_{2h})|\le C^{2h}E(s_1s_2\ldots s_{2h})\le C'^{2h}E(s_1^{2h}).$$

Therefore, Theorem 2.1 (b) is proved and hence the proof of Theorem 2.1 is complete.

3.2 Proof of Remark 2.1

Let $\{b_{j_1j_2}\}_{j_1,j_2=1}^{r}$ be as (2.6). Let us define

$$\begin{aligned}R_{i,j_1,j_2} &= \varphi(a_{j_1}c_{j_2}\gamma^{i-1}),\ \forall j_1,j_2=1,2,\ldots,r,\ i\ge 1,\\ R_i &= \sum_{j_1,j_2=1}^{r} b_{j_1j_2}c_{j_1}a_{j_2}R_{i,j_1,j_2},\ \forall i\ge 1,\\ \beta(z,a) &= -\sum_{i=1}^{\infty} z^{-i}R_i, \forall z\in\mathbb{C}^+.\end{aligned}$$

Note that $\varphi(\gamma^{2h-1})=0$ and $\varphi(R_{2h})=0,\ \forall h\ge 1$. By Lemma 4.2, we have

$$\begin{aligned}\varphi(\gamma^{2h}) &= \sum_{j_1,j_2,\ldots,j_{2h}}\varphi(\prod_{k=1}^{2h}a_{j_k}s_{j_k}c_{j_k})\\ &= \sum_{j_1,j_2,\ldots,j_{2h}}\sum_{\pi\in \mathrm{NC}_2(2h)}\varphi_{K(\pi)}(c_{j_1}a_{j_2},c_{j_2}a_{j_3},\ldots,c_{j_{2h}}a_{j_1})k_\pi(s_{j_1},\ldots,s_{j_{2h}}).\end{aligned}\quad(3.16)$$

To simplify (3.16), consider the following decomposition of $\mathrm{NC}_2(2h)$.

$$\mathrm{NC}_2(2h) = \cup_{i=1}^{h}C_{i,h},\ \text{where}$$

$$C_{i,h} \quad = \quad \text{set of all } \sigma \in \text{NC}_2(2h) \text{ such that } \{1, 2i\} \in \sigma.$$

Note that the contribution of $\{\{1, 2h\}, \{2, 3\}, \{4, 5\}, \ldots, \{2h-2, 2h-1\}\} \in C_{h,h}$ to right side of (3.16), is $\varphi(R_1^h)$. Now,

$$\varphi(\gamma^2) = \text{contribution of } C_{1,1} = \varphi(R_1).$$

Again,

$$\varphi(\gamma^4) = \text{contribution of } C_{1,2} + \text{contribution of } C_{2,2} = \varphi(R_3 + R_1^2).$$

Next,

$$\varphi(\gamma^6) = \text{contribution of } C_{1,3} + \text{contribution of } C_{2,3} + \text{contribution of } C_{3,3} = \varphi(R_5 + R_1 R_3 + (R_3 R_1 + R_1^3)).$$

Now, let us define the set of all ordered partitions of K in t blocks as follows,

$$S_{K,t} = \{(i_1, i_2, \ldots, i_t) : i_1, i_2, \ldots, i_t \in \mathbb{N}, \sum_{j=1}^{t} i_j = r\}, \ \forall K \geq 1, 1 \leq t \leq K.$$

Then, one can show easily by induction on h that

$$\varphi(\gamma^{2h}) = \varphi\left(\sum_{t=1}^{h} \sum_{i_1, i_2, \ldots, i_t \in S_{2h-t,t}} \prod_{j=1}^{t} R_{i_j} \right), \forall h \geq 1.$$

We omit the tedious details. Hence, using the power series expansion (4.4),

$$\begin{aligned}
G_\gamma(z) &= \varphi((\gamma - z)^{-1}) = -z^{-1} \sum_{h=0}^{\infty} z^{-2h} \varphi(\gamma^{2h}) = -\varphi\left(z^{-1} \sum_{h=0}^{\infty} z^{-2h} \sum_{t=1}^{h} \sum_{i_1, i_2, \ldots, i_t \in S_{2h-t,t}} \prod_{j=1}^{t} R_{i_j} \right) \\
&= -\varphi\left(z^{-1} \sum_{t=0}^{\infty} z^{-t} \sum_{h=t}^{\infty} \sum_{i_1, i_2, \ldots, i_t \in S_{2h-t,t}} \prod_{j=1}^{t} z^{-i_j} R_{i_j} \right) = -\varphi\left(z^{-1} \sum_{t=0}^{\infty} z^{-t} \left(\sum_{i=1}^{\infty} z^{-i} R_i \right)^t \right) \\
&= -\varphi\left(z^{-1} \sum_{t=0}^{\infty} z^{-t} (-\beta(z, a))^t \right) = -\varphi((z + \beta(z, a))^{-1}).
\end{aligned}$$

Similarly, one can show easily by induction on h and the assumption $\{a_j : j = 1, 2, \ldots, r\} = \{c_j : j = 1, 2, \ldots, r\}$, that

$$\begin{aligned}
R_{2h+1} &= \sum_{j_1, j_2 = 1}^{r} b_{j_1 j_2} a_{j_1} c_{j_2} \varphi\left(c_{j_1} a_{j_2} \sum_{t=1}^{h} \sum_{i_1, i_2, \ldots, i_t \in S_{2h-t,t}} \prod_{j=1}^{t} R_{i_j} \right), \\
\beta(z, a) &= -\sum_{j_1, j_2 = 1}^{r} b_{j_1 j_2} c_{j_1} a_{j_2} \varphi(a_{j_1} c_{j_2} (z + \beta(z, a))^{-1}).
\end{aligned}$$

Hence the proof of Remark 2.1 is complete.

3.3 Proof of Corollary 2.2

Corollary 2.2 (a) follows trivially from Theorem 2.1 and the truncation arguments given in Subsection 3.4. Corollary 2.2 (b) follows easily from Corollary 2.2 (a) and (2.10). Corollary 2.2 (c) follows from Theorem 2.1 (b) and Remark 2.1. Proof of Corollary 2.2 (d) is as follows.

Proof of (d) (i)*:* Note that by Corollary 2.2 (c),

$$G_{a^{1/2}sa^{1/2}}(z) = -E\left((z + a\beta(z))^{-1}\right) = -(\beta(z))^{-1}G_a(-z(\beta(z))^{-1}). \tag{3.17}$$

Also,

$$\begin{aligned}\beta(z) &= -E\left(a(z + a\beta(z))^{-1}\right) = -(\beta(z))^{-1}E\left(a(a + z(\beta(z))^{-1})^{-1}\right)\\ &= -(\beta(z))^{-1}\left(1 - z(\beta(z))^{-1}G_a(-z(\beta(z))^{-1})\right)\\ &= -(\beta(z))^{-1}\left(1 + zG_{a^{1/2}sa^{1/2}}(z)\right).\end{aligned}$$

Therefore, $\beta(z) = -\sqrt{-1 - zG_{a^{1/2}sa^{1/2}}(z)}$. We take the negative sign of the square root because $\lim_{|z|\to\infty} z\beta(z) = -E(a)$ and $\lim_{|z|\to\infty} \sqrt{-1 - zG_{a^{1/2}sa^{1/2}}(z)} = E(a)$. Hence

$$G_{a^{1/2}sa^{1/2}}(z) = \frac{1}{\sqrt{-1 - zG_{a^{1/2}sa^{1/2}}(z)}}G_a\left(\frac{z}{\sqrt{-1 - zG_{a^{1/2}sa^{1/2}}(z)}}\right).$$

Proof of equivalence of (i) *and* (iii) *in* (d)*:*

$$\begin{aligned}
S_{a^{1/2}sa^{1/2}a^{1/2}sa^{1/2}}(z) &= S_{(sa)^2}(z) = S_{s^2}(z)S_a^2(z) = (z+1)^{-1}S_a^2(z)\\
&\iff \chi_{(sa)^2}(z) = z(z+1)^{-2}S_a^2(z) = z^{-1}\chi_a^2(z) \ \ (by(4.5) - (4.8))\\
&\iff z = (\Psi_{(sa)^2}(z))^{-1}\chi_a^2(\Psi_{(sa)^2}(z)) \iff z\,\Psi_{(sa)^2}(z) = \chi_a^2(\Psi_{(sa)^2}(z))\\
&\iff \chi_a(\Psi_{(sa)^2}(z)) = \sqrt{z\Psi_{(sa)^2}(z)} \iff \Psi_{(sa)^2}(z) = \Psi_a\left(\sqrt{z\Psi_{(sa)^2}(z)}\right)\\
&\iff -z^{-1}G_{(sa)^2}(z^{-1}) - 1 = \Psi_a\left(\sqrt{-G_{(sa)^2}(z^{-1}) - z}\right)\\
&\iff -zG_{(sa)^2}(z) - 1 = \Psi_a\left(\sqrt{-G_{(sa)^2}(z) - z^{-1}}\right)\\
&\iff G_{(sa)^2}(z) = -z^{-1}\left(\Psi_a\left(\sqrt{-G_{(sa)^2}(z) - z^{-1}}\right) + 1\right)\\
&\iff G_{(sa)^2}(z) = -z^{-1}\left(\Psi_a\left(\frac{\sqrt{-zG_{(sa)^2}(z) - 1}}{\sqrt{z}}\right) + 1\right)\\
&\iff G_{(sa)^2}(z) = -z^{-1}\left(-\frac{\sqrt{z}}{\sqrt{-zG_{(sa)^2}(z) - 1}}G_a\left(\frac{\sqrt{z}}{\sqrt{-zG_{(sa)^2}(z) - 1}}\right)\right)\\
&\iff G_{(sa)^2}(z^2) = \frac{1}{z}\frac{1}{\sqrt{-z^2G_{(sa)^2}(z^2) - 1}}G_a\left(\frac{z}{\sqrt{-z^2G_{(sa)^2}(z^2) - 1}}\right)
\end{aligned}$$

$$\iff \quad z^{-1}G_{(sa)}(z) = \frac{1}{z}\frac{1}{\sqrt{-zG_{(sa)}(z)-1}}G_a\left(\frac{z}{\sqrt{-zG_{(sa)}(z)-1}}\right), \text{ (by (4.10))}$$
$$\iff \quad G_{a^{1/2}sa^{1/2}}(z) = \frac{1}{\sqrt{-1-zG_{a^{1/2}sa^{1/2}}(z)}}G_a\left(\frac{z}{\sqrt{-1-zG_{a^{1/2}sa^{1/2}}(z)}}\right).$$

Proof of equivalence of (*ii*) and (*iii*) in (*d*) follows easily by Lemma 4.3, (4.18) and (4.20). We omit the details.

Hence, the proof of Corollary 2.2 (*d*) is complete.

3.4 Proof of Corollary 2.3

(*a*) When $X_t \sim MA(q)$ process, we can write Model (1.4) as

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

Let $\hat{\Gamma}_k(\varepsilon)$ be the k-th order sample autocovariance matrix of ε. Therefore, by (1.5), the sample autocovariance matrix of order k is given by

$$\begin{aligned}
n\hat{\Gamma}_{k,p} &= \sum_{t=k+1}^{n}(\psi_{0,p}^{(n)} \ \psi_{1,p}^{(n)} \ \dots \psi_{q,p}^{(n)})(\varepsilon_{t,p}^{*} \ \dots \varepsilon_{t-q,p}^{*})^{*}(\varepsilon_{t-k,p}^{*} \ \dots \varepsilon_{t-k-q,p}^{*})(\psi_{0,p}^{(n)} \ \psi_{1,p}^{(n)} \ \dots \psi_{q,p}^{(n)})^{*} \\
&= n\sum_{j=0}^{q}\sum_{j'=0}^{q}\psi_{j,p}^{(n)}\,\hat{\Gamma}_{j'-j+k}(\varepsilon)\,\psi_{j',p}^{(n)*} - \sum_{j=0}^{q}\sum_{\substack{j'=0\\ j-j'\neq k}}^{q}\psi_{j,p}^{(n)}\left(\sum_{t=n-j+1}^{n}\varepsilon_{t,p}\varepsilon_{t-(j'+k-j)}^{*}\right)\psi_{j',p}^{(n)*} \\
&\quad + \sum_{j=0}^{q}\sum_{\substack{j'=0\\ j-j'\neq k}}^{q}\psi_{j,p}^{(n)}\left(\sum_{t=k-j+1}^{j'+k-j}\varepsilon_{t,p}\varepsilon_{t-(j'+k-j)}^{*}\right)\psi_{j',p}^{(n)^{*}} + \sum_{j=0}^{q}\psi_{j,p}^{(n)}\left(\sum_{t=n-j+1}^{n}\varepsilon_{t,p}\varepsilon_{t,p}^{*}\right)\psi_{j-k,p}^{(n)*} \\
&\quad + \sum_{j=0}^{q}\psi_{j,p}^{(n)}\left(\sum_{t=k-j+1}^{0}\varepsilon_{t,p}\varepsilon_{t,p}^{*}\right)\psi_{j-k,p}^{(n)*} \\
&= n\Delta_k + R_{1n} + R_{2n} + R_{3n} + R_{4n}, \text{ (say)}. \qquad (3.18)
\end{aligned}$$

Let F^A denote the ESD of the matrix A and L denote the Lévy metric between two distribution functions. Let $C_n = \sum_{j=0}^{q-\tau}\psi_j\psi_{j+\tau}^{*}$. By Theorem A.43 and Lemma B.18 in [Bai and Silverstein, 2010], we have for some $C > 0$,

$$\begin{aligned}
L(F^{\sqrt{np^{-1}}(\hat{\Gamma}_i+\hat{\Gamma}_i^{*}-C_n)}, F^{\sqrt{np^{-1}}(\Delta_i+\Delta_i^{*}-C_n)}) &\leq p^{-1}[\text{rank}(R_{1n}) + \text{rank}(R_{2n}) + \text{rank}(R_{3n}) + \text{rank}(R_{4n})] \\
&\leq \frac{4Cq}{p} \to 0 \text{ a.s..}
\end{aligned}$$

Therefore, LSD of $\sqrt{np^{-1}}(\hat{\Gamma}_i + \hat{\Gamma}_i^{*} - C_n)$ and $\sqrt{np^{-1}}(\Delta_i + \Delta_i^{*} - C_n)$ are identical. Let P_i be the $n \times n$ matrix with upper i-th diagonal 1 and 0 otherwise. Note that $P_0 = I_n$ and

$P_{-i} = P_i^*$. Then,

$$\Delta_i = \sum_{j=0}^{q} \sum_{j'=0}^{q} \psi_{j,p}^{(n)} \hat{\Gamma}_{j'-j+k}(\varepsilon) \psi_{j',p}^{(n)*} = n^{-1} \sum_{j=0}^{q} \sum_{j'=0}^{q} \psi_{j,p}^{(n)} Z P_{j'-j+k} Z^* \psi_{j',p}^{(n)*}. \tag{3.19}$$

Therefore, under Assumptions ($A1$), ($A2$) and ($A5$) and by Theorem 2.1 (b), the LSD of $\sqrt{np^{-1}}(\Delta_i + \Delta_i^* - C_n)$ exists. Clearly, this LSD is distributed as g_i defined in (2.14). Hence, Corollary 2.3 (a) and (b) (i) are proved under Assumption ($A2$).

Equivalence of (i) *and* (ii) *in* (b)*:* Now we derive the Stieltjes transformation of g_τ and establish (ii). In this case, applying Remark 2.1, we have

$$\begin{aligned}
-\beta(z,a) &= \sum_{j_1,j_2,k_1,k_2=1}^{q} \frac{1}{2\pi} \int_0^{2\pi} \cos(j_1 - j_2 + \tau)\theta \, \cos(k_1 - k_2 + \tau)\theta \, \gamma_{j_2}\gamma_{k_1}\varphi\left(\gamma_{j_1}\gamma_{k_2}(z + \beta(z,a))^{-1}\right) d\theta \\
&= \frac{1}{8\pi} \int_0^{2\pi} e^{2i\tau\theta} \sum_{j_1,j_2,k_1,k_2=1}^{q} e^{i(j_1-j_2+k_1-k_2)\theta} \gamma_{j_2}\gamma_{k_1}\varphi\left(\gamma_{j_1}\gamma_{k_2}(z + \beta(z,a))^{-1}\right) d\theta \\
&\quad + \frac{1}{8\pi} \int_0^{2\pi} e^{-2i\tau\theta} \sum_{j_1,j_2,k_1,k_2=1}^{q} e^{i(-j_1+j_2-k_1+k_2)\theta} \gamma_{j_2}\gamma_{k_1}\varphi\left(\gamma_{j_1}\gamma_{k_2}(z + \beta(z,a))^{-1}\right) d\theta \\
&\quad + \frac{1}{8\pi} \int_0^{2\pi} \sum_{j_1,j_2,k_1,k_2=1}^{q} e^{i(j_1-j_2-k_1+k_2)\theta} \gamma_{j_2}\gamma_{k_1}\varphi\left(\gamma_{j_1}\gamma_{k_2}(z + \beta(z,a))^{-1}\right) d\theta \\
&\quad + \frac{1}{8\pi} \int_0^{2\pi} e^{2i\tau\theta} \sum_{j_1,j_2,k_1,k_2=1}^{q} e^{i(-j_1+j_2+k_1-k_2)\theta} \gamma_{j_2}\gamma_{k_1}\varphi\left(\gamma_{j_1}\gamma_{k_2}(z + \beta(z,a))^{-1}\right) d\theta \\
&= \frac{1}{8\pi} \int_0^{2\pi} (e^{2i\tau\theta} + e^{-2i\tau\theta} + 2)\Psi(a,\theta)\varphi\left(\Psi(b,\theta)(z + \beta(z,a))^{-1}\right) d\theta \\
&= \varphi\left(\frac{1}{2\pi} \int_0^{2\pi} \cos^2(\tau\theta)\Psi(a,\theta)\Psi(b,\theta)(z + \beta(z,a))^{-1} d\theta \,|a\right) \\
&= \varphi\left(R_\tau(a,b)(z + \beta(z,a))^{-1} \,|a\right).
\end{aligned}$$

Hence, Corollary 2.3 (b) (ii) is proved.

Now we show that Corollary 2.3 (a) and (b) remain true even if we drop ($A2$) and use the more relaxed condition (2.9). This is achieved by truncation arguments similar to those given in page $1210 - 1217$ of [Jin et al., 2014]. Let X_t be as (1.4) and suppose Assumptions ($A1$), ($A5$) and (2.9) hold. Let

$$\begin{aligned}
\tilde{\varepsilon}_{ti} &= \varepsilon_{ti} I(|\varepsilon_{ti}| < \eta n^{\frac{1}{2+\delta}}), \ \hat{\varepsilon}_{ti} = \tilde{\varepsilon}_{ti} - E(\tilde{\varepsilon}_{ti}), \ \forall t, i \text{ and some } \eta > 0, \\
\sigma_{ti}^2 &= E|\hat{\varepsilon}_{ti}|^2, \ \Delta = n^{-\frac{\delta}{4+2\delta}}, X_{ti} = 2Ber(0.5) - 1, \text{ i.i.d. for all } t, i, \\
\bar{\varepsilon}_{ti} &= \begin{cases} X_{ti}, \text{ if } \sigma_{ti}^2 < 1 - \Delta, \\ \frac{\hat{\varepsilon}_{ti}}{\sigma_{ti}}, \text{ otherwise}, \end{cases} \\
\hat{\Gamma}_i(\varepsilon), \tilde{\Gamma}_i(\varepsilon), \hat{\hat{\Gamma}}_i(\varepsilon), \bar{\Gamma}_i(\varepsilon) &= i\text{-th order sample autocovariance matrix of } \{\varepsilon_{ti}\}, \{\tilde{\varepsilon}_{ti}\}, \{\hat{\varepsilon}_{ti}\}, \{\bar{\varepsilon}_{ti}\}
\end{aligned}$$

(respectively)

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

Since $\{\bar{\varepsilon}_{t,i}\}$ satisfy the stronger assumption $(A2)$, the existence of the LSD of $\sqrt{np^{-1}}(\bar{T}_i + \bar{T}_i^* - C_n)$ is guaranteed by the previous arguments and it is distributed as g_i defined in (2.14).

We will actually show that the LSD of $\sqrt{np^{-1}}(\hat{\Gamma}_i + \hat{\Gamma}_i^* - C_n)$ is same as that of $\sqrt{np^{-1}}(\bar{T}_i + \bar{T}_i^* - C_n)$. Note that

$$\begin{aligned}
& L(F^{\sqrt{np^{-1}}(\hat{\Gamma}_i+\hat{\Gamma}_i^*-C_n)}, F^{\sqrt{np^{-1}}(\bar{T}_i+\bar{T}_i^*-C_n)}) \\
\leq\ & L(F^{\sqrt{np^{-1}}(\hat{\Gamma}_i+\hat{\Gamma}_i^*-C_n)}, F^{\sqrt{np^{-1}}(\hat{T}_i+\hat{T}_i^*-C_n)}) + L(F^{\sqrt{np^{-1}}(\hat{T}_i+\hat{T}_i^*-C_n)}, F^{\sqrt{np^{-1}}(\tilde{T}_i+\tilde{T}_i^*-C_n)}) \\
& + L(F^{\sqrt{np^{-1}}(\tilde{T}_i+\tilde{T}_i^*-C_n)}, F^{\sqrt{np^{-1}}(\hat{\hat{T}}_i+\hat{\hat{T}}_i^*-C_n)}) + L(F^{\sqrt{np^{-1}}(\hat{\hat{T}}_i+\hat{\hat{T}}_i^*-C_n)}, F^{\sqrt{np^{-1}}(\bar{T}_i+\bar{T}_i^*-C_n)}). \\
=\ & T_1 + T_2 + T_3 + T_4, \text{ (say)}. \qquad (3.20)
\end{aligned}$$

It is enough to show that $T_i \to 0$ almost surely for all $i = 1, 2, 3, 4$.

By Theorem A.43 and Lemma B.18 in [Bai and Silverstein, 2010], we have for some $C > 0$, with R_{1n}, R_{2n}, R_{3n} and R_{4n} as in (3.18),

$$\begin{aligned}
T_1 &\leq p^{-1}(\text{rank}(R_{1n}) + \text{rank}(R_{2n}) + \text{rank}(R_{3n}) + \text{rank}(R_{4n})) \\
&\leq \frac{4Cq}{p} \to 0 \text{ a.s.}. \qquad (3.21)
\end{aligned}$$

By Theorem A.43 and Lemma B.18 in [Bai and Silverstein, 2010], we have for some $C, C_1 > 0$

$$\begin{aligned}
T_2 &\leq \frac{1}{p}\text{rank}(\hat{T}_i + \hat{T}_i^* - \tilde{T}_i - \tilde{T}_i^*) \leq \frac{2}{p}\text{rank}(\hat{T}_i - \tilde{T}_i) \\
&\leq \frac{1}{p}\text{rank}\left(\sum_{j,j'=0}^{q} \psi_j \left(\hat{\Gamma}_{j-j'+i}(\varepsilon) - \tilde{\Gamma}_{j-j'+i}(\varepsilon)\right)\psi_{j'}^*\right) \\
&\leq \frac{C}{p}\text{rank}(\hat{\Gamma}_i(\varepsilon) - \tilde{\Gamma}_i(\varepsilon)) \\
&\leq \frac{C_1}{p}\sum_{j=1}^{p}\sum_{t=1}^{n+i} I(|\varepsilon_{t,j}| \geq \eta p^{1/(2+\delta)}). \qquad (3.22)
\end{aligned}$$

Also, we have

$$E\left(\frac{1}{p}\sum_{j=1}^{p}\sum_{t=1}^{n+i} I(|\varepsilon_{tj}| \geq \eta p^{1/(2+\delta)})\right)$$

$$\leq \frac{1}{\eta^{2+\delta}p^2}\sum_{j=1}^{p}\sum_{t=1}^{n+i}E\left(|\varepsilon_{t,j}|^{(2+\delta)}I(|\varepsilon_{tj}|\geq \eta p^{1/(2+\delta)})\right)=o(1) \tag{3.23}$$

and

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{p}\sum_{j=1}^{p}\sum_{t=1}^{n+i}I(|\varepsilon_{tj}|\geq \eta p^{1/(2+\delta)})\right)\\
\leq\ & \frac{1}{\eta^{2+\delta}p^3}\sum_{j=1}^{p}\sum_{t=1}^{n+i}E\left(|\varepsilon_{t,j}|^{(2+\delta)}I(|\varepsilon_{tj}|\geq \eta p^{1/(2+\delta)})\right)=o(p^{-1}).
\end{aligned} \tag{3.24}$$

Applying Bernstein's inequality and (3.23), (3.24), for all $\epsilon > 0$ and large p, we have for some $C, C_1 > 0$,

$$P\left(\frac{1}{p}\sum_{j=1}^{p}\sum_{t=1}^{n+i}I(|\varepsilon_{t,j}|>\eta p^{1/(2+\delta)})\geq\epsilon\right)\leq Ce^{-C_1 p}.$$

Therefore, by Borel-Cantelli lemma, we have

$$T_2 \to 0 \text{ a.s..} \tag{3.25}$$

Let $\hat{\gamma}_k = n^{-1/2}(\hat{\varepsilon}_{k,1}, \hat{\varepsilon}_{k,2}, \ldots, \hat{\varepsilon}_{k,p})'$ and $\tilde{\gamma}_k = n^{-1/2}(\tilde{\varepsilon}_{k,1}, \tilde{\varepsilon}_{k,2}, \ldots, \tilde{\varepsilon}_{k,p})'$. By Theorem $A.45$ in [Bai and Silverstein, 2010], we have for some $C, C_1 > 0$,

$$\begin{aligned}
T_3 &\leq \sqrt{np^{-1}}\|\tilde{T}_i+\tilde{T}_i^*-\hat{\hat{T}}_i-\hat{\hat{T}}_i^*\|\leq C\sqrt{np^{-1}}\|\tilde{\Gamma}_i(\varepsilon)-\hat{\hat{\Gamma}}_i(\varepsilon)\|\\
&\leq C_1\sqrt{np^{-1}}\|\sum_{k=1}^{n}(\hat{\gamma}_kE\tilde{\gamma}_{k+i}^*+\hat{\gamma}_{k+i}E\tilde{\gamma}_k^*)\|+C_1\sqrt{np^{-1}}\|\sum_{k=1}^{n}(E\hat{\gamma}_kE\tilde{\gamma}_{k+i}^*+E\hat{\gamma}_{k+i}E\tilde{\gamma}_k^*)\|.
\end{aligned}$$

For the second part, we have for some $C > 0$,

$$\begin{aligned}
&\sqrt{np^{-1}}\|\sum_{k=1}^{n}(E\hat{\gamma}_kE\tilde{\gamma}_{k+i}^*+E\hat{\gamma}_{k+i}E\tilde{\gamma}_k^*)\|\\
\leq\ &\sqrt{(np)^{-1}}\sum_{k=1}^{n}\sum_{j=1}^{p}|E(\varepsilon_{k,j}I(|\varepsilon_{k,j}|>\eta p^{1/(2+\delta)}))E(\varepsilon_{k+i,j}I(|\varepsilon_{k+i,j}|>\eta p^{1/(2+\delta)}))|\\
\leq\ &C\frac{p}{\sqrt{np}}p^{-2}\sum_{k=1}^{n+i}\sum_{j=1}^{p}E(|\varepsilon_{k,j}|^{2+\delta}I(|\varepsilon_{k,j}|>\eta p^{1/(2+\delta)}))=o(1).
\end{aligned}$$

For the first part, note that

$$np^{-1}\|\sum_{k=1}^{n}(\hat{\gamma}_kE\tilde{\gamma}_{k+i}^*+\hat{\gamma}_{k+i}E\tilde{\gamma}_k^*)\|^2\leq 2np^{-1}\left(\|\sum_{k=1}^{n}\hat{\gamma}_kE\tilde{\gamma}_{k+i}^*\|^2+\|\sum_{k=1}^{n}\hat{\gamma}_{k+i}E\tilde{\gamma}_k^*\|^2\right).$$

Now, for some $C > 0$, we have

$$
\begin{aligned}
& np^{-1}\|\sum_{k=1}^{n}\hat{\gamma}_k E\tilde{\gamma}_{k+i}^{*}\|^2 \le C(np)^{-1}\sum_{j=1}^{p}\sum_{l=1}^{p}\left(\sum_{k=1}^{n}\hat{\varepsilon}_{k,j}E\tilde{\varepsilon}_{k+i,l}\right)^2 \\
= \;& C(np)^{-1}\sum_{j=1}^{p}\sum_{l=1}^{p}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}(\hat{\varepsilon}_{k_1,j}E\tilde{\varepsilon}_{k_1+i,l}\hat{\varepsilon}_{k_2,j}E\tilde{\varepsilon}_{k_2+i,l}) \\
= \;& C(np)^{-1}\sum_{j=1}^{p}\sum_{l=1}^{p}\left(\sum_{k_1=1}^{n}\hat{\varepsilon}_{k_1,j}^2(E\tilde{\varepsilon}_{k_1+i,l})^2 + \sum_{k_1\neq k_2}\hat{\varepsilon}_{k_1,j}E\tilde{\varepsilon}_{k_1+i,l}\hat{\varepsilon}_{k_2,j}E\tilde{\varepsilon}_{k_2+i,l}\right) \\
= \;& J_{11} + J_{12}, \text{ (say)}.
\end{aligned} \tag{3.26}
$$

As $E(\hat{\varepsilon}_{t,i}^4), E(\tilde{\varepsilon}_{t,i}^4) < \infty$, there exists constant C_1, C_2 and C_3 such that

$$
\begin{aligned}
EJ_{11} &= C(np)^{-1}\sum_{j=1}^{p}\sum_{l=1}^{p}\sum_{k_1=1}^{n}\hat{\varepsilon}_{k_1,j}^2(E\tilde{\varepsilon}_{k_1+i,l})^2 \le C(np)^{-1}\sum_{j=1}^{p}\sum_{l=1}^{p}\sum_{k_1=1}^{n}(E(|\varepsilon_{k_1,l}|I(|\varepsilon_{k_1,l}| > \eta p^{1/(2+\delta)})))^2 \\
&\le C(np)^{-1}\eta^{-2(1+\delta)}p^{-2(1+\delta)/(2+\delta)}\sum_{j=1}^{p}\sum_{l=1}^{p}\sum_{k_1=1}^{n}E(|\varepsilon_{k_1,l}|^{2+\delta}I(|\varepsilon_{k_1,l}| > \eta p^{1/(2+\delta)})))^2 = O(p^{-\delta/(2+\delta)}).
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}J_{11} &= C^2(np)^{-2}\sum_{j=1}^{p}\sum_{k_1=1}^{n}E(\hat{\varepsilon}_{k_1,j}^2 - E\hat{\varepsilon}_{k_1,j}^2)^2\left(\sum_{l=1}^{p}(E\tilde{\varepsilon}_{k_1+i,l})^2\right)^2 \\
&\le C_2(np)^{-2}\sum_{j=1}^{p}\sum_{k_1=1}^{n}E(\tilde{\varepsilon}_{k_1,j}^4)\left(p\eta^{-2(1+\delta)}p^{-2(1+\delta)/(2+\delta)}\right)^2 = O(p^{-1-4\delta/(2+\delta)}).
\end{aligned}
$$

Therefore, $J_{11} \to 0$ a.s.. Further, we have

$$
\begin{aligned}
\mathrm{Var}J_{12} &= C(np)^{-2}\sum_{j=1}^{p}\sum_{k_1\neq k_2}E\hat{\varepsilon}_{k_1,j}^2E\hat{\varepsilon}_{k_2,j}^2\left(\sum_{l=1}^{p}E\tilde{\varepsilon}_{k_1+i,l}E\tilde{\varepsilon}_{k_2+i,l}\right)^2 \\
&\le C_3(np)^{-2}\sum_{j=1}^{p}\sum_{k_1\neq k_2}\left(p\eta^{-2(1+\delta)}p^{-2(1+\delta)/(2+\delta)}\right)^2 = O(p^{-1-2\delta/(2+\delta)}),
\end{aligned}
$$

which implies $J_{12} \to 0$, a.s.. Hence, we have $\|\sum_{k=1}^{n}\hat{\gamma}_k E\tilde{\gamma}_{k+i}^{*}\|^2 \to 0$, a.s.. Similarly, $\|\sum_{k=1}^{n}\hat{\gamma}_{k+i}E\tilde{\gamma}_{k}^{*}\|^2 \to 0$, a.s.. Thus,

$$T_3 \to 0, \text{ a.s..}$$

We now finally prove $T_4 \to 0$ almost surely. By Corollary $A.41$ in Bai and Silverstein [2010], we have

$$T_4^3 \;\le\; \frac{n}{p^2}\,\mathrm{Tr}\left((\hat{\hat{T}}_i + \hat{\hat{T}}_i^{*} - \bar{T}_i - \bar{T}_i^{*})(\hat{\hat{T}}_i + \hat{\hat{T}}_i^{*} - \bar{T}_i - \bar{T}_i^{*})^{*}\right)$$

$$\begin{aligned}
&\le \frac{4n}{p^2}\operatorname{Tr}\left((\hat{\hat{T}}_i-\bar{T}_i)(\hat{\hat{T}}_i-\bar{T}_i)^*\right)\\
&= \frac{4n}{p^2}\sum_{j,j',k,k'=0}^{q}\operatorname{Tr}\Big(\psi_j\left(\hat{\hat{\Gamma}}_{j-j'+i}(\varepsilon)-\bar{\Gamma}_{j-j'+i}(\varepsilon)\right)\psi_{j'}^*\\
&\qquad\qquad \psi_{k'}\left(\hat{\hat{\Gamma}}_{k-k'+i}(\varepsilon)-\bar{\Gamma}_{k-k'+i}(\varepsilon)\right)^*\psi_k^*\Big).
\end{aligned}$$

Therefore, it is enough to show that

$$np^{-2}\operatorname{Tr}(A(\hat{\hat{\Gamma}}_i(\varepsilon)-\bar{\Gamma}_i(\varepsilon))BB^*(\hat{\hat{\Gamma}}_i(\varepsilon)-\bar{\Gamma}_i(\varepsilon))^*A^*)\to 0,\ \text{a.s.} \tag{3.27}$$

for any $A,B\in\operatorname{Span}\{\psi_j,\psi_j^*\,:\,j\ge 0\}$. The proof of (3.27) 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 page 1210 − 1217 of [Jin et al., 2014]. In our case we have the extra factors of A, B etc. Let

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

Then,

$$\begin{aligned}
&np^{-2}\operatorname{Tr}(A(\hat{\hat{\Gamma}}_i(\varepsilon)-\bar{\Gamma}_i(\varepsilon))BB^*(\hat{\hat{\Gamma}}_i(\varepsilon)-\bar{\Gamma}_i(\varepsilon))^*A^*)\\
=\ &np^{-2}\operatorname{Tr}(A(\hat{U}\hat{V}^*-\bar{U}\bar{V}^*)BB^*(\hat{U}\hat{V}^*-\bar{U}\bar{V}^*)^*A^*)\\
=\ &np^{-2}\operatorname{Tr}(A((\hat{U}-\bar{U})\hat{V}^*+\bar{U}(\hat{V}-\bar{V})^*)BB^*((\hat{U}-\bar{U})\hat{V}^*+\bar{U}(\hat{V}-\bar{V})^*)^*A^*)\\
\le\ &2np^{-2}\operatorname{Tr}(A(\hat{U}-\bar{U})\hat{V}^*BB^*\hat{V}(\hat{U}-\bar{U})^*A^*)\\
&+2np^{-2}\operatorname{Tr}(A\bar{U}(\hat{V}-\bar{V})^*BB^*(\hat{V}-\bar{V})\bar{U}^*A^*).
\end{aligned}$$

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

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

where,

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

$$J_3 = \frac{Cn}{p^2n^2}\sum_{u,v}\sum_{\substack{l_1,l_2,j_1,j_2\\ k_2>k_1,\ k_2\neq k_1+i}}\left(a_{ul_1}(\hat{\varepsilon}_{k_1l_1}-\bar{\varepsilon}_{k_1l_1})\hat{\varepsilon}^*_{(k_1+i)j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k_2+i)j_2}(\hat{\varepsilon}_{k_2l_2}-\bar{\varepsilon}_{k_2l_2})^*a^*_{ul_2}\right),$$

$$J_4 = \frac{Cn}{p^2n^2}\sum_{u,v}\sum_{\substack{l_1,j_1,l_2,\\ j_2,k_1}}\left(a_{ul_1}(\hat{\varepsilon}_{k_1l_1}-\bar{\varepsilon}_{k_1l_1})\hat{\varepsilon}^*_{(k_1+i)j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k_1+2i)j_2}(\hat{\varepsilon}_{(k_1+i)l_2}-\bar{\varepsilon}_{(k_1+i)l_2})^*a^*_{ul_2}\right),$$

$$J_5 = \frac{Cn}{p^2n^2}\sum_{\substack{u,v,l_1,l_2\\ j_1,j_2,k}}\left[a_{ul_1}(\hat{\varepsilon}_{kl_1}-\bar{\varepsilon}_{kl_1})\hat{\varepsilon}^*_{(k+i)j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k+i)j_2}(\hat{\varepsilon}_{kl_2}-\bar{\varepsilon}_{kl_2})^*a^*_{ul_2}\right].$$

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

$$\begin{aligned}
\operatorname{Var}(J_1) &= E(J_2)^2 \leq \frac{C_1n^2}{p^4n^4}\sum_{u_1,v_1}\sum_{\substack{l_1,l_2,j_1,j_2\\ k_1>k_2,\ k_1\neq k_2+i}}\sum_{u_2,v_2}\sum_{\substack{l_3,l_4,j_3,j_4\\ k_3>k_4,\ k_3\neq k_4+i}} E\Big[a_{u_1l_1}(\hat{\varepsilon}_{k_1l_1}-\bar{\varepsilon}_{k_1l_1})\hat{\varepsilon}^*_{(k_1+i)j_1}b_{j_1v_1}b^*_{j_2v_1} \\
&\quad \hat{\varepsilon}_{(k_2+i)j_2}(\hat{\varepsilon}_{k_2l_2}-\bar{\varepsilon}_{k_2l_2})^*a^*_{u_1l_2}a_{u_2l_3}(\hat{\varepsilon}_{k_3l_3}-\bar{\varepsilon}_{k_3l_3})\hat{\varepsilon}^*_{(k_3+i)j_3}b_{j_3v_2}b^*_{j_4v_2} \\
&\quad \hat{\varepsilon}_{(k_4+i)j_4}(\hat{\varepsilon}_{k_4l_4}-\bar{\varepsilon}_{k_4l_4})^*a^*_{u_2l_4}\Big] \\
&\leq \frac{C_2n^2}{p^4n^2}\sum_{\substack{u_1,u_2\\ v_1,v_2}}\sum_{\substack{l_1,l_2\\ j_1,j_2}}\left(a_{u_1l_1}b_{j_1v_1}b^*_{j_2v_1}a^*_{u_1l_2}a_{u_2l_1}b_{j_1v_2}b^*_{j_2v_2}a^*_{u_2l_2}\right) \\
&\leq \frac{C_3}{p^2}(p^{-1}\operatorname{Tr}(A^2A^{*2}))(p^{-1}\operatorname{Tr}(B^2B^{*2})) = O(p^{-2}).
\end{aligned}$$

Also for some $C_1, C_2 > 0$,

$$\begin{aligned}
\operatorname{Var}(J_2) &= E(J_2)^2 \leq \frac{C_1n^2}{p^4n^4}\sum_{\substack{u_1,v_1\\ u_2,v_2}}\sum_{\substack{l_1,j_1,l_2,\\ j_2,k_2}}\sum_{\substack{l_3,j_3,l_4,\\ j_4,k_4}} E\Big[a_{u_1l_1}(\hat{\varepsilon}_{(k_2+i)l_1}-\bar{\varepsilon}_{(k_2+i)l_1})\hat{\varepsilon}^*_{(k_2+2i)j_1}b_{j_1v_1}b^*_{j_2v_1} \\
&\quad \hat{\varepsilon}_{(k_2+i)j_2}(\hat{\varepsilon}_{k_2l_2}-\bar{\varepsilon}_{k_2l_2})^*a^*_{u_1l_2}a_{u_2l_3}(\hat{\varepsilon}_{(k_4+i)l_3}-\bar{\varepsilon}_{(k_4+i)l_3})\hat{\varepsilon}^*_{(k_4+2i)j_3} \\
&\quad b_{j_3v_2}b^*_{j_4v_2}\hat{\varepsilon}_{(k_4+i)j_4}(\hat{\varepsilon}_{k_4l_4}-\bar{\varepsilon}_{k_4l_4})^*a^*_{u_2l_4}\Big] \\
&\leq \frac{C_2}{p^4}\sum_{\substack{u_1,v_1\\ u_2,v_2}}\sum_{\substack{l_1,j_1,l_2,\\ j_2}}\sum_{\substack{l_3,j_3,l_4,\\ j_4}}\left(a_{u_1l_1}b_{j_1v_1}b^*_{j_2v_1}a^*_{u_1l_2}a_{u_2l_3}b_{j_1v_2}b^*_{j_4v_2}a^*_{u_2l_2}\right) = O(p^{-2}).
\end{aligned}$$

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

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 $C > 0$

$$(1-\sigma^2_{ti}) \leq 2E(\tilde{\tilde{\varepsilon}}^2_{ti}) \leq 2C(P(|\varepsilon_{ti}| > \eta_n n^{\frac{1}{2+\delta}}))^{\frac{\delta}{2+\delta}} \leq 2C\eta_n^{-\delta}p^{-\frac{\delta}{2+\delta}} \tag{3.28}$$

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]),

$$(1-\sigma_{ti}^{-1})^2 \leq C\eta_n^{-2\delta} p^{-\frac{2\delta}{2+\delta}}. \tag{3.29}$$

Moreover note that if $(t,i) \in E$, then $\frac{1-\sigma_{ti}^2}{\Delta} > 1$. Then by (3.28) and (3.29), we have for some $C_1, C_2 > 0$,

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

Therefore,

$$E(np^{-2}\operatorname{Tr}(A(\hat{U}-\bar{U})\hat{V}^* BB^* \hat{V}(\hat{U}-\bar{U})^* A^*)) \to 0.$$

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

$$\operatorname{Var}(np^{-2}\operatorname{Tr}(A(\hat{U}-\bar{U})\hat{V}^* BB^* \hat{V}(\hat{U}-\bar{U})^* A^*)) = O(p^{-1-\epsilon}).$$

Hence,

$$np^{-2}\operatorname{Tr}(A(\hat{U}-\bar{U})\hat{V}^* BB^* \hat{V}(\hat{U}-\bar{U})^* A^*) \to 0, a.s.. \tag{3.30}$$

Similarly,

$$np^{-2}\operatorname{Tr}(A\bar{U}(\hat{V}-\bar{V})^* BB^* (\hat{V}-\bar{V})\bar{U}^* A^*) \to 0, a.s.. \tag{3.31}$$

Hence by (3.30) and (3.31), (3.27) is proved. Also by (3.21), (3.22), (3.26) and (3.27),

$$T_4 \to 0, a.s..$$

Hence, the proof of Corollary 2.3 (a) and (b) is complete. Corollary 2.3 (c) follows trivially since (2.10) holds.

4 Appendix

In this section, we discuss the concepts and results from the literature of non-commutative probability, Stieltjes and related transformations etc., which have been used in the proof of Theorems 2.1 and the remarks and corollaries in Section 2.

Moment method. We begin with a result (see Bai and Silverstein [2010]), which outlines the moment method for showing the existence of LSD. This was used in the proof of Theorem 2.1 (b).

The empirical spectral distribution (ESD) of an $n \times n$ real symmetric matrix R_n is the probability distribution which puts mass $1/n$ at each of its eigenvalues. The h-th order raw moment of the ESD is given by

$$\beta_h(R_n) := \frac{1}{n}\sum_{i=1}^{n} \lambda_i^h = \frac{1}{n}\operatorname{Tr}(R_n^h). \tag{4.1}$$

Consider the following conditions.

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

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

(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,

$$\sum_{h=1}^{\infty} \beta_{2h}^{-\frac{1}{2h}} = \infty. \tag{4.2}$$

Lemma 4.1. *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.*

Stieltjes and related transforms. We needed the Stieltjes and related transforms to prove Corollary 2.2 (d). We provide a brief discussion on various transforms following Couillet and Debbah [2011]. Let μ be a probability distribution on the real line.

(a) The Stieltjes transformation of μ is the function

$$G_\mu(z) = \int \frac{1}{x-z}\mu(dx), \quad z \in \mathbb{C}^+, \tag{4.3}$$

where $\mathbb{C}^+ := \{x + iy : x \in \mathbb{R}, y > 0\}$. Pointwise convergence of the Stieltjes transform to a Stieltjes transform implies the weak convergence of the corresponding distributions. If μ has compact support then we have the following formal power series expansion of the Stieltjes transformation $G_\mu(z)$

$$G_\mu(z) = -\frac{1}{z}E_\mu\left(\frac{1}{1-\frac{x}{z}}\right) = -\frac{1}{z} - \frac{E_\mu(x)}{z^2} - \frac{E_\mu(x^2)}{z^3} - \ldots. \tag{4.4}$$

(b) Ψ transformation. Ψ-transformation is defined by

$$\Psi_\mu(z) = \int \frac{zt}{1-zt}d\mu(t). \tag{4.5}$$

It is easy to see that

$$\Psi_\mu\left(\frac{1}{z}\right) = -zG_\mu(z) - 1. \tag{4.6}$$

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

$$\chi_\mu(\Psi_\mu(z)) = z \text{ for } |z| \text{ small enough.} \tag{4.7}$$

(d) S transformation. The S-transformation S_μ is given by

$$S_\mu(z) = \frac{1+z}{z}\chi_\mu(z). \tag{4.8}$$

If $\mu \boxtimes \nu$ is the free product of two measures μ and ν (see page 33 after Definition 4.4), then it is known that

$$S_{\mu\boxtimes\nu} = S_\mu S_\nu. \tag{4.9}$$

(e) Let X be a random variable symmetric about 0. Then $G_X(z) = G_{-X}(z)$. Therefore, we have

$$\begin{aligned} G_{X^2}(z^2) &= E((z^2 - X^2)^{-1}) = \frac{1}{2z}\left(E((z-x)^{-1}) + E((z+x)^{-1})\right) \\ &= \frac{1}{2z}(G_X(z) + G_{-X}(z)) = \frac{G_X(z)}{z}. \end{aligned} \tag{4.10}$$

Now, we discuss the concepts and results of non-commutative probability space following [Nica and Speicher, 2006]. We needed these in Section 2.

Non-commutative probability space.

Definition 4.1. 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 said to be *self-adjoint*. The state φ is *tracial* if

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

The following example of a 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,

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

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

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

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

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 $\{a_i, a_i^* : i \in \mathbb{Z}\} \subset \mathcal{A}$. Then

$$\mathrm{Span}\{a_i, a_i^* : i \in \mathbb{Z}\} = \{\Pi(a_i, a_i^* : i \in \mathbb{Z}) : \ \Pi \text{ is a polynomial}\} \tag{4.11}$$

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

For example, $(\mathcal{A}, \varphi)$ and $(\mathcal{D} := \mathrm{Span}(\{a_{l,2i+1}, a_{l.2i+1}^* : 1 \leq l \leq r, 0 \leq i \leq k_l\} \cup (\cup_{u=1}^{U} \mathcal{T}_u)), \varphi)$ (see (2.2) and (2.3)) are non-commutative $*$-probability spaces. $\mathcal{U}$ in Theorem 2.1 is a non-commutative $*$-sub algebra of $\mathcal{D}$. $\{\mathcal{U}_n\}_{n\geq 1}$ in Theorem 2.1 is a sequence of non-commut-ative $*$-probability spaces.

Definition 4.2. 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 called 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.

Consider $t \geq 1$. Let $\Pi(\{a_{l,2i+1}, a_{l.2i+1}^* : 1 \leq l \leq r, 0 \leq i \leq k_l\}) \in \mathcal{A}$ be any polynomial in $\{a_{l,2i+1}, a_{l.2i+1}^* : 1 \leq l \leq r, 0 \leq i \leq k_l\} \subset \mathcal{A}$. Then $\{\varphi(\Pi(a_{l,2i+1}, a_{l.2i+1}^* : 1 \leq l \leq r, 0 \leq i \leq k_l)) : \Pi \text{ polynomial}\}$ determines the joint distribution of $\{a_{l,2i+1}, a_{l.2i+1}^* : 1 \leq l \leq r, 0 \leq i \leq k_l\}$. For example, (2.4) provides the joint distribution of $\{a_{l,2i+1}, a_{l.2i+1}^* : 1 \leq l \leq r, 0 \leq i \leq k_l\}$.

Non-crossing partitions and Kreweras complement. Non-crossing partitions form the core of the concept of free independence. Among other things, Kreweras complement serves as a useful tool to compute moments of polynomials of free variables (see for example, Lemma 4.2). We provide a brief description of the main concepts and results on these here.

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

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

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

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

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

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

Free cumulants and free independence.

Definition 4.3. 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

$$k_\pi[a_1, a_2, ..., a_n] := \sum_{\substack{\sigma\in NC(n)\\ \sigma\leq\pi}} \varphi_\sigma[a_1, a_2, \ldots, a_n]\mu(\sigma, \pi), \tag{4.12}$$

where μ is the Möbius function on the POSET $NC(n)$. Note that the Möbius function depends only on partitions and ordering and not on the variables whose cumulants are being calculated. For each $n \geq 1$, we put $k_n := k_{1_n}$. (4.12) 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

$$k_n(a_1, a_2, \ldots, a_n) = \sum_{\sigma\in NC(n)} \varphi_\sigma[a_1, a_2, ..., a_n]\mu(\sigma, 1_n). \tag{4.13}$$

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

$$k_n(s) = \begin{cases} 1, & \text{if } n = 2 \\ 0, & \text{if } n > 2. \end{cases} \tag{4.14}$$

$\{s_i : i \in I\}$ is said to form a semi-circle family with the variance-covariance matrix $((b_{ij}))_{I\times I}$ if

$$\begin{aligned} k_2(s_i, s_j) &= b_{ij},\ \forall i, j \in I \text{ and} \\ k_n(s_{i_1}, s_{i_2}, \ldots, s_{i_n}) &= 0,\ \forall n \geq 2,\ i_1, i_2, \ldots, i_n \in I. \end{aligned} \tag{4.15}$$

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

Definition 4.4. 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 $i(l) \neq i(k)$.

(4.13) in conjunction with Definition 4.4 is helpful for showing free independence of two sub-algebras.

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 the *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 relevant in (4.9).

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

Lemma 4.2. *Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and consider random variables $\{a_1, a_2, \ldots a_n\}$ and $\{b_1, b_2, \ldots, b_n\}$ which are freely independent. Then we have*

$$\varphi(a_1b_1a_2b_2\ldots a_nb_n) = \sum_{\pi\in NC(n)} k_\pi[a_1, a_2, \ldots, a_n]\varphi_{K(\pi)}[b_1, b_2, \ldots, b_n],$$

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

Next we discuss some properties of power series and generating functions that we needed in Corollary 2.2 (d) and its proof.

Power series and generating functions: Let s be a positive integer. We denote by Θ_s the set of all formal power series of the form

$$f(z_1, \ldots, z_s) = \sum_{n=1}^{\infty} \sum_{i_1,\ldots,i_n=1}^{s} \alpha_{i_1,\ldots,i_n} z_{i_1} \ldots z_{i_n}, \tag{4.16}$$

with $\alpha_{i_1,\ldots,i_n} \in \mathbb{C}$ $(\forall n \geq 1,\ \forall 1 \leq i_1, \ldots, i_n \leq s)$, and where $z_1, \ldots, z_s$ are non-commutative indeterminates.

Let $f \in \Theta_s$ be as in (4.16). For every $n \geq 1$ and $1 \leq i_1, \ldots, i_n \leq s$ we denote

$$\alpha_{i_1,\ldots,i_n} =: \mathrm{Cf}_{(i_1,\ldots,i_n)}(f).$$

Now let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and let $a_1, \ldots, a_s$ be an s-tuple of elements of $\mathcal{A}$. Consider the family of all free cumulants of $a_1, \ldots, a_s$,

$$\{k_n(a_{i_1} \ldots a_{i_n}) :\ n \geq 1,\ 1 \leq i_1, \ldots, i_n \leq s\}.$$

With these numbers we form the series $\mathcal{R}_{a_1,\ldots,a_s}$, called the R-transform of $a_1, \ldots, a_s$:

$$\mathcal{R}_{a_1,\ldots,a_s}(z_1, \ldots, z_s) = \sum_{n=1}^{\infty} \sum_{i_1,\ldots,i_n=1}^{s} k_n(a_{i_1} \ldots a_{i_n}) z_{i_1} \ldots z_{i_n}. \tag{4.17}$$

For $s = 1$, let $f(z) = \sum_{n=1}^{\infty} \alpha_n z^n \in \Theta_1$. Then we define

$$(f \circ \mathrm{Sq})(z) := \sum_{n=1}^{\infty} \alpha_n z^{2n}. \tag{4.18}$$

It is easy to see that if a is an even element i.e. if all odd order cumulants are zero, then $\mathcal{R}_a(z) = \mathcal{R}_{a^2}(z) \circ \mathrm{Sq}$

Definition 4.5. Zeta$_s$: Zeta$_s \in \Theta_s$ is defined as

$$\mathrm{Zeta}_s(z_1, \ldots, z_s) := \sum_{n=1}^{\infty} \sum_{i_1,\ldots,i_n=1}^{s} z_{i_1} \ldots z_{i_n}. \tag{4.19}$$

If $s = 1$, we simply write Zeta. Note that

$$\mathrm{Zeta} = \mathcal{R}_{s^2}, \tag{4.20}$$

where s is a standard semi-circle variable.

Definition 4.6. $\boxtimes$ convolution: On Θ_s we define a binary operation $\boxtimes$, by the following rule: for every $f, g \in \Theta_s$ and for every $n \geq 1$, $1 \leq i_1, i_2, \ldots, i_n \leq s$, the coefficient of order $(i_1, \ldots, i_n)$ of $f \boxtimes g$ is:

$$\mathrm{Cf}_{(i_1,\ldots,i_n)}(f \boxtimes g) := \sum_{\pi \in \mathrm{NC}(n)} \mathrm{Cf}_{(i_1,\ldots,i_n);\pi}(f)\mathrm{Cf}_{(i_1,\ldots,i_n);K(\pi)}(g). \tag{4.21}$$

We then have the following Lemma.

Lemma 4.3. *Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and let $\{a_1, \ldots, a_s\}$ and $\{b_1, \ldots, b_s\}$ be free. Then the R-transform of the s-tuple $(a_1b_1, \ldots, a_sb_s)$ is*

$$\mathcal{R}_{a_1b_1,\ldots,a_sb_s} = \mathcal{R}_{a_1,\ldots,a_s} \boxtimes \mathcal{R}_{b_1,\ldots,b_s}.$$

Definition 4.7. Convergence of NCP and freeness

Definition 4.8. 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 $(\mathrm{Span}\{a_i^{(N)}, a_i^{*(N)} : 1 \leq i \leq t\}, \varphi_N)$ converges to $(\mathrm{Span}\{a_i, a_i^* : 1 \leq i \leq t\}, \varphi)$.
For example, Theorem 2.1 says $(\mathcal{U}_n, Ep^{-1}\mathrm{Tr})$ converges to $(\mathcal{U}, \varphi)$.

References

Z. D. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices.* Springer, 2010.

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

Z. D. Bai and L. X. Zhang. The limiting spectral distribution of the product of the Wigner matrix and a nonnegative definite matrix. *J. Multivariate Anal.*, 101(9): 1927–1949, 2010. doi: 10.1016/j.jmva.2010.05.002.

Z. Bao. Strong convergence of ESD for the generalized sample covariance matrices when $p/n \to 0$. *Statist. Probab. Lett.*, 82(5):894–901, 2012. doi: 10.1016/j.spl.2012.01.012.

M. Bhattacharjee and A. Bose. Large sample behaviour of high dimensional autocovariance matrices. *Tech. Rep., R2/2014*, $http : //www.isical.ac.in/ \sim statmath/html /publication/large\_sample\_behaviour.pdf. Submitted\ for\ publication$, 2014.

M. Bhattacharjee and A. Bose. Matrix polynomial generalizations of sample variance-covariance matrices when $p/n \to y \in (0, \infty)$. *Manuscript in progress*, 2015.

A. Bose, S. Gangopadhyay, and A. Sen. Limiting spectral distribution of XX' matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(3):677–707, 2010. doi: 10.1214/09-AIHP329.

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

C. Francq and H. Raïsi. Multivariate portmanteau test for autoregressive models with uncorrelated but non-independent errors. *J. Time Series Anal.*, 28(3):454–470, 2007.

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

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

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

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

B. Jin, C. Wang, Z. D. Bai, K. K. Nair, and M. Harding. Limiting spectral distribution of a symmetrized auto-cross covariance matrix. *Ann. Appl. Probab.*, 24(3):1199–1225, 2014.

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

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

J. W. Lin and A. McLeod. Improved Pẽna-Rodriguez portmanteau test. *Computational statistics and data analysis*, 51(3):1731–1738, 2006.

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

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

E. Mahdi and L. McLeod. Improved multivariate portmanteau test. *J. Time Series Anal.*, 33(2):211–222, 2012.

V. Marčenko and L. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1:457–483, 1967.

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

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

G. M. Pan and J. T. Gao. Asymptotic theory for sample covariance matrix under cross-sectional dependence. *Manuscript*, 2009.

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

L. Wang and D. Paul. Limiting spectral distribution of renormalized separable sample covariance matrices when $p/n \to 0$. *J. Multivariate Anal.*, 126:25–52, 2014. doi: 10.1016/j.jmva.2013.12.015.

L. Wang, A. Aue, and D. Paul. Limiting spectral distribution of normalised sample autocovariance matrices of linear time series when $p, n \to \infty$ such that $pn^{-1} \to 0$. *Manuscript in progress*, 2014.