ASYMPTOTICS OF LARGE VARIANCE-COVARIANCE AND AUTOCOVARIANCE MATRICES

MONIKA BHATTACHARJEE

Indian Statistical Institute, Kolkata
March, 2016

MONIKA BHATTACHARJEE

Thesis submitted to the Indian Statistical Institute

in partial fulfillment of the requirements

for the award of the degree of

Doctor of Philosophy.

March, 2016

Thesis Advisor : Prof. Arup Bose

Indian Statistical Institute

203, B. T. Road, Kolkata, India.

To my grandparents, parents and teachers

Acknowledgement

At the very outset, as I submit this thesis, I convey my gratefulness to my supervisor Prof. Arup Bose for his valuable guidance, inspiration, constant support and as well as for his parental affection throughout the entire period of my Ph.D. Program. His great insight into statistical methodology and techniques have profoundly influenced how I approach statistical research. I am also grateful to him for his valuable time and patience for correction of errors and omissions in my research at every step. I am also forever grateful to him for his silent mental support throughout this period.

I convey my gratitude to other Professors of the Indian Statistical Institute, namely Parthanil Roy, Rajat Subhra Hazra, Krishanu Maulik, Antar Bandyopadhyay, Anil Kumar Ghosh, Probal Chaudhuri, Gopal Krishna Basak, Alok Goswami and Amitesh Dasgupta who helped me with their valuable suggestions and comments on my research. I wish to make special mention of Parthanil Roy and Rajat Subhra Hazra for their suggestions about my future career.

I extend my sincere thanks to my friends without whose help and support this long journey would have been monotonous. My office mates, Munmun di and Soumi were always by my side, whenever I needed their assistance. I also like to thank Maitreyee, Madhura, Sayanti, Tamalika, Raka, Kaiser di, Sedigheh di, Sanghita di and Suchana di for their lovely companionship. I also like to mention Ayan for his constant support and assistance throughout this period.

I cannot forget my professors and teachers in B. Sc., M. Stat. and as well as in my high school, whose contributions are immense towards my present achievement. They not only introduced me to the different aspects of the subject of 'Statistics' but also inspired and encouraged me to proceed further for higher studies.

I owe special gratitude to my parents, whose love, affection, care and inspiration have been all along with me throughout my life. I cannot forget their sacrifices while bringing me up and towards the achievements in my life; these cannot be expressed in words. I have tried to fulfill their expectations and make them happy.

Finally, I also like to express my sincerest thanks to those special people whose names could not be mentioned individually due to space constraint, though their help and support is heartily appreciated.

Contents

Chapter 1

Introduction and summary

New technologies and methods in medical sciences, image processing and the internet, and many other fields of science generate data where the dimension is high and the sample size is small relative to the dimension. For example, microarray data (Dudoit et al. [2002]) contains gene expression for tens of thousands of genes (rows) on a few observations (columns). Another example is fMRI data, which measures the hemodynamic response in hundreds of thousands of voxels (rows) for only a few subjects or replicates (columns). Similarly, the Netflix movie rating data (Bennett and Lanning [2007]) contains the rating information for approximately 480,000 customers (columns) on 18,000 movies (rows). In all these cases, the dimension is large compared to the sample size and may also increase as the next set of measurements become available. Theoretical and practical study of these kind of data has attracted the attention of recent researchers as most of the methods in finite dimensional set up do not work in these cases. This thesis has mainly focused on the following problems in high-dimensional situations. Detailed summary of each chapter is given later.

(a) Suppose the observations $\{X_i : 1 \leq i \leq n\}$ are identically distributed with mean 0 and the $p \times p$ variance-covariance matrix Σ_p. As we are in high-dimensional set up, the dimension $p = p(n) \to \infty$ as the sample size $n \to \infty$. The estimation of the variance-covariance matrix Σ_p is crucial as many statistical analysis such as classification problem, principal component analysis are all based on the variance-covariance matrix. *In Chapter 2, we have discussed estimation of* Σ_p *when* $\{X_i\}$ *are 'weakly' dependent random vectors.*

(b) High-dimensional data are often time series in nature. A very general weak stationary high-dimensional linear time series model is the infinite dimensional moving average process of order ∞ (MA(∞)). A key quantity in the study of these models is the population autocovariance matrices $\{\Gamma_u\}$. *In Chapter 3, we have studied the estimation of* $\{\Gamma_u\}$ *for infinite dimensional MA*(∞) *process.*

(c) Next, we have explored some asymptotic properties of sample autocovariance matrices $\{\hat{\Gamma}_u\}$ for the infinite dimensional MA(q) ($q < \infty$) and MA(∞) processes. Even though these matrices are not consistent, their asymptotic properties, while interesting in their own right, can also be used for statistical applications.

Two most natural ways to study the joint convergence of a collection of random matrices are through

(1) Limiting spectral distribution (LSD) of any symmetric polynomial in these matrices and,

(2) convergence of non-commutative probability space (NCP) generated by these matrices.

Incidentally, (2) with some additional effort implies (1). *In Chapter 4, we have discussed some concepts on random matrix theory and non-commutative probability for studying* (1) *and* (2) *above for* $\{\hat{\Gamma}_u\}$. For more details see Section 1.3.

We have shown that for LSD and NCP convergence purpose, $\{\hat{\Gamma}_u\}$ for MA(q) process can be approximated by $\{\Delta_u\}$ where

$$\Delta_u = \frac{1}{n}\sum_{i,j=0}^{q} \psi_i Z P_{i-j+u} Z^* \psi_j^*, \ \forall u \geq 1, \tag{1.1}$$

Z is a $p \times n$ random matrix with all independent mean 0 and variance 1 entries,

$\{\psi_j\}$ are $p \times p$ matrices (model parameters) and $P_u = ((I(j-i=u)))_{n\times n},\ \forall u.$ *In Chapters 5 and 6, we have studied* (1) *and* (2) *above for* $\{\Delta_u\}$ *and indeed for more general class of random matrices, respectively when* $p/n \to y > 0$ *and* $p/n \to 0$.

In Chapter 7, we have used the results obtained in Chapters 5 and 6 to establish the LSD of any symmetric polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$.

(d) Finally, in Chapter 8, we have discussed *a couple of applications* of our results, obtained so far, in statistical inference:

(i) a model identification problem, namely determination of the unknown order of infinite dimensional moving average (MA) and autoregressive (AR) processes and,

(ii) testing of simple hypotheses by using asymptotic normality of traces of polynomials in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$.

A chapterwise summary of this thesis is given below.

1.1 Summary of Chapter 2

Let $X_i,\ 1 \leq i \leq n$, be p-dimensional identically distributed random vectors with mean 0 and variance covariance matrix Σ_p. As we are in the high-dimensional set up, the dimension $p = p(n) \to \infty$ as the sample size $n \to \infty$. In Chapter 2, we have discussed estimation of Σ_p.

An estimator $\hat{A}_{p,n}$, based on a sample of size n, is called *consistent* (in operator norm) for A_p if

$$||A_p - \hat{A}_{p,n}||_2 \xrightarrow{P} 0, \quad \text{as} \quad n \to \infty, \tag{1.2}$$

where $||\cdot||_2$ is the *operator norm* of a matrix.

In the finite dimensional case i.e. when p is fixed, the sample variance-covariance matrix $\hat{\Sigma}_p$ is a consistent estimator of Σ_p. In high dimensional setting, many researchers have proved that $\hat{\Sigma}_p$ fails to estimate Σ_p consistently, even for i.i.d. $\{X_i\}$. For examples see Johnstone [2001], Johnstone and Lu [2004], Paul [2007] and Johnstone and Lu [2009]. Inconsistency of $\hat{\Sigma}_p$ is also supported by the simulation results given in Example 2.2.1. This inconsistency is due to the increase in the number of unknown parameters along with the sample size. As a remedy we need some restrictions on the parameter space and modifications of the basic estimator $\hat{\Sigma}_p$. This modification is called *covariance regularization*. There are many covariance regularization techniques available for i.i.d. observations in the literature. These have appeared while studying different aspects such as regression, classification or principal component analysis.

One such covariance regularization is due to Bickel and Levina [2008]. It has played a crucial role in Chapter 2 and is discussed in details in Section 2.2.1. They have proved that, uniformly over some fairly natural well conditioned families of covariance matrices, the suitably *banded* and *tapered* $\hat{\Sigma}_p$ are consistent in the *operator norm* for Σ_p when observations are i.i.d. and as long as $n^{-1} \log p \to 0$. They have also obtained explicit rates of these convergence. For more details, see Theorems 2.2.2 and 2.2.6.

However, the independence assumption on $\{X_i\}$ is questionable and many researchers have provided evidence of its lack. For examples see Owen [2005], Klebanov and Yakovlev [2007], Leek and Storey [2008] and Efron [2009]. Efron [2009] proposed the matrix-variate normal distribution as a model for dependent $\{X_i\}$. Allen and Tibshirani [2010] is the only work in this model which has estimated Σ_p. In Section 2.2.2, we have briefly discussed their covariance regularization.

In Section 2.3, we have considered more general models (we call these *weak models*) and have given examples of a huge class of models which are not accommodated by the model assumption of Allen and Tibshirani [2010] (see Examples 2.3.1-2.3.3). We have allowed dependence of appropriate nature among the observations. We have called this dependence the *cross covariance strucuture*. In this chapter we have considered three different restrictions on the cross covariance structure. In the first case, the restriction is on the growth of the powers of the trace of certain matrices derived from the cross covariance structure. In the second case, the dependence among any two columns weakens as the lag between them increases and in the third case we have assumed weak dependence among the high-indexed columns. See Section 2.3.1 for details.

In the first case, we have shown that the convergence rate of the banded estimator is the same as in the i.i.d. case of Theorem 2.2.2 (see Theorem 2.4.4) under a *trace condition*. In Remarks 2.4.5-2.4.8, we have provided some sufficient conditions that imply this trace condition.

The other two weak models do not fall under the purview of Theorem 2.4.4. Under appropriate conditions we have obtained explicit rates of convergence for the banded estimators (see Theorems 2.4.9 and 2.4.11). In particular, for all three cases, the suitably banded variance-covariance matrix continues to remain consistent in operator norm.

Banded estimators are not necessarily positive definite. So we have considered tapered estimators that preserve the positive definiteness of the sample variance-covariance matrix. We have obtained the rate of convergence of the tapered estimator for all the three weak models (see Theorem 2.4.13). In particular the tapered estimator continues to remain consistent in operator norm in these dependent situations.

It is seen that the growth rate of p and the convergence rates for the tapered and the banded estimators are in general slower than the i.i.d. case and there is a trade-off between these rates and the extent of dependency.

1.2 Summary of Chapter 3

Next consider the high dimensional time series model. Some of the more common existing weak stationary high-dimensional time series models in the literature are infinite dimensional IID processes, infinite dimensional finite order moving average processes (MA) and infinite dimensional vector autoregressive processes (IVAR) with *i.i.d.* innovations. Detailed description of the above models is available in Forni and Lippi [2001], Forni et al. [2004] and Chudik and Pesaran [2011]. We have considered the general model, namely the infinite dimensional moving average process of order ∞ (MA(∞)). This is given by

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \forall t, \tag{1.3}$$

where X_t and ε_t are p-dimensional vectors, $\{\varepsilon_t\}$ are i.i.d. with mean 0 and $p \times p$ variance covariance matrix Σ_p, $\{\psi_j\}$ are $p \times p$ *coefficient matrices.* Moreover, as in Chapter 2, the dimension $p = p(n) \to \infty$ as the sample size $n \to \infty$. If $\psi_j = 0$, $\forall j > q$, then it is an infinite dimensional MA(q) process. The IID process is nothing but the MA(0) process. The IVAR(r) process is given by

$$X_t = \sum_{i=1}^{r} A_i X_{t-i} + \varepsilon_t, \quad \forall t, \tag{1.4}$$

where X_t and ε_t are p-dimensional vectors, $\{\varepsilon_t\}$ are i.i.d. with mean 0 and $p \times p$ variance covariance matrix Σ_p, $\{A_i\}$ are $p \times p$ *parameter matrices.* Under appropriate conditions on $\{A_i\}$, (1.4) can also be expressed in the form of (1.3). In Section 3.2, we have briefly described all the above models. The dimension of the above models are not infinite but tends to ∞ with the sample size. However, it has become customary to call them "infinite dimensional".

A key quantity in these models is the sequence of population autocovariance matrices

$$\Gamma_u = E(X_t X_{t+u}^*), \quad \forall u \geq 0. \tag{1.5}$$

In Chapter 3, we have discussed estimation of $\{\Gamma_u\}$ for the process (1.3). Since the size of the population autocovariance matrices $\{\Gamma_u\}$ increases as $p = p(n) \to \infty$, the number of unknown parameters (entries in $\{\Gamma_u\}$) increases. Consequently, just like the sample variance-covariance matrix in Chapter 2, the sample autocovariance matrices

$$\hat{\Gamma}_u = \frac{1}{n} \sum_{t=1}^{n-u} X_t X_{t+u}^*, \quad \forall u \geq 0 \tag{1.6}$$

fail to consistently estimate $\{\Gamma_u\}$.

The existing works on high-dimensional time series have not dealt with the estimation of $\{\Gamma_u\}$. From the experience of Section 1.1, to get consistent estimators of $\{\Gamma_u\}$ we need two things – suitable restrictions on $\{\psi_j\}$ and Σ_p and, appropriate modifications such as banding or tapering on sample autocovariance matrices $\{\hat{\Gamma}_u\}$. In Section 3.3, Theorems 3.3.1 and 3.3.2 provide some restrictions on $\{\psi_j\}$ and Σ_p under which the appropriately banded and tapered version of $\hat{\Gamma}_0$ is consistent for Γ_0. These restrictions are directly borrowed from the developments of Chapter 2 and are seen to be cumbersome and often very difficult to check in general unless there is some additional structure in the model (1.3). Theorems 3.3.1 and 3.3.2 is also silent about other autocovariance matrices.

In Section 3.4, we have identified another appropriate parameter space for $\{\psi_j\}$ and Σ_p to estimate $\{\Gamma_u\}$ for the model (1.3). In Theorem 3.5.1, we have established consistency of the banded and tapered $\{\hat{\Gamma}_u\}$ under the Gaussian assumption, when $n^{-1} \log p \to 0$. We have also derived the convergence rate of these estimators.

As we have discussed earlier, the infinite dimensional MA(q) and IVAR(r) processes are special cases of the model (1.3). Theorems 3.4.5 and 3.4.6 provide appropriate parameter spaces respectively for the matrices $\{\Sigma_p, \psi_j : 0 \leq j \leq q\}$ and $\{\Sigma_p, A_i : 1 \leq i \leq r\}$ so that population autocovariance matrices of the MA(q) and IVAR(r) processes can be consistently estimated. Under these parameter spaces and the Gaussian assumption on the driving process, Theorems 3.5.5 and 3.5.6 state that the banded and tapered $\{\hat{\Gamma}_u\}$ are also consistent for $\{\Gamma_u\}$, when

$n^{-1} \log p \to 0$. Using this consistency, in Section 3.5.1, we have also shown how to obtain consistent estimators for the parameter matrices $\{A_i : 1 \le i \le r\}$ of the IVAR(r) process.

Finally, in Section 3.5.2, we have relaxed the Gaussian assumption on $\{\varepsilon_t\}$ and replaced it by an appropriate condition on the moment generating function. To support our results, some simulations are given in Section 3.6. Our simulations have shown that the convergence rate obtained in Theorem 3.5.8 is quite sharp.

In the subsequent chapters, we have explored further asymptotic properties of $\{\hat{\Gamma}_u\}$. Even though these matrices are not consistent, their asymptotic properties, while interesting in their own right, can also be used for statistical applications.

1.3 Summary of Chapter 4

Note that $\{\hat{\Gamma}_u\}$ are random matrices. A most natural way to look at the large sample behaviour of a collection of random matrices is to study the limiting spectral distribution (LSD) of their polynomials. The *empirical spectral distribution* (ESD) of an $n \times n$ (random) matrix R_n is the (random) probability distribution with mass $1/n$ at each of its eigenvalues. If it converges weakly (almost surely or in probability) to a (non-degenerate) probability distribution, then the latter is called the *limiting spectral distribution* (LSD) of R_n. LSD results for various random matrices have occupied a central position in the literature of random matrix theory (RMT). The so called spectral statistics, useful in statistical application, are functions of this spectral distribution. In case of infinite dimensional moving average processes, many researchers have observed that, under some assumptions, the ESD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ (after appropriate normalization) converges weakly. For example see Liu et al. [2015] and Wang et al. [2015]. Such results are useful in statistical inference. Therefore, the study of the limiting spectral property of $\{\hat{\Gamma}_u\}$ is very important.

In Section 4.2, we have discussed some basic concepts in RMT such as necessary definitions and results with examples, Stieltjes transformation method and moment method to obtain LSD and existing results in the literature, which are relevant to us. We have focussed on the LSD of only symmetric matrices. The LSD of non-symmetric matrices are in general very hard and only a very few results are known in very specific models.

Two specific matrices play a central role in RMT — the Wigner matrix and the independent matrix. A *Wigner matrix* of order p, W_p is a square symmetric random matrix with independent mean 0 variance 1 entries on and above the diagonal. An *independent matrix* of order $p \times n$, Z is a rectangular matrix with all independent mean 0 and variance 1 entries.

Let A_p be a $p \times p$ symmetric non-negative definite matrix, whose LSD exists. In the literature of Wigner matrix, under appropriate moment assumption on the entries of W_p, the LSD of the following matrices are known (a) $p^{-1/2}W_p$ (see for example Anderson et al. [2009]) and (b) $p^{-1/2}A_p^{1/2}W_pA_p^{1/2}$ (Bai and Zhang [2010]).

The classical RMT model for Z assumes $p = p(n) \to \infty$ as $n \to \infty$ and

$$\frac{p}{n} \to y \in [0, \infty). \tag{1.7}$$

For the $y > 0$ case, under appropriate moment assumption on the entries of Z, the LSD of the following matrices are known (c) $n^{-1}ZZ^*$ (Bai and Silverstein [2009]) and (d) $n^{-1}A_p^{1/2}ZZ^*A_p^{1/2}$ (Bai and Silverstein [2009]).

The LSD results for the case $y = 0$ are quite different from the case $y > 0$. Let $\{B_n\}$ be $n \times n$ square symmetric norm bounded non-negative matrices with $\lim_n n^{-1}\mathrm{Tr}(B_n^k) < \infty,\ k = 1, 2$. Under appropriate moment assumption on the entries of Z, the LSD of the following matrices are known (e) $\sqrt{np^{-1}}(n^{-1}ZZ^* - I)$ (Bai and Yin [1988]), (f) $\sqrt{np^{-1}}(n^{-1}A_p^{1/2}ZZ^*A_p^{1/2} - A)$ (Bao [2012]) and (g) $\sqrt{np^{-1}}(n^{-1}A_p^{1/2}ZB_nZ^*Z_P^{1/2} - n^{-1}\mathrm{Tr}(B_n)A_p)$ (Wang and Paul [2014]). Moreover, LSD of (e) and (f) are respectively identical with the LSD of (a) and (b) above.

All the above results have been derived using Stieltjes transform method. As discussed earlier, for a collection of random matrices, we wish to study the LSD of all (symmetric) polynomials in these matrices. The Stieltjes transform method can deal with one polynomial at a time. We have used the moment method which in conjunction with the tools from the non-commutative probability theory, has provided a unified way to study all symmetric polynomials together.

In Section 4.3, we have collected all the necessary concepts and results in non-commutative probability theory. The joint convergence of a class of random matrices can be done in the framework of non-commutative $*$-probability spaces (NCP) generated by them. A *non-commutative $*$-probability space* (NCP) $(\mathcal{A}, \varphi)$ consists of a unital $*$-algebra $\mathcal{A}$ and a linear functional $\varphi : \mathcal{A} \to \mathbb{C}$ (called *state* of $\mathcal{A}$) with $\varphi(1_{\mathcal{A}}) = 1$. Elements of $\mathcal{A}$ are called (non-commutative) variables.

Let $\{a_i, a_i^* : i \geq 1\} \subset \mathcal{A}$. Then

$$\text{Span}\{a_i, a_i^* : i \geq 1\} = \{\Pi(1_{\mathcal{A}}, a_i, a_i^* : i \geq 1) :\ \Pi \text{ is a polynomial}\}. \tag{1.8}$$

Note that $\text{Span}\{a_i, a_i^* : i \geq 1\}$ forms a $*$-sub algebra of $\mathcal{A}$ and is an NCP with the same state φ.

Let $\mathcal{A}_N = \text{Span}\{a_i^{(N)}, a_i^{*(N)} : i \geq 1\}$, $\forall\ N \geq 1$ and $\mathcal{A} = \text{Span}\{a_i, a_i^* : i \geq 1\}$. We say that the sequence of NCP $\{(\mathcal{A}_N, \varphi_N)\}_{N=1}^{\infty}$ *converges* to $(\mathcal{A}, \varphi)$ if for any polynomial Π and $t \geq 1$

$$\lim_{N\to\infty} \varphi_N \left(\Pi(a_i^{(N)}, a_i^{*(N)} : 1 \leq i \leq t)\right) = \varphi\left(\Pi(a_i, a_i^* : 1 \leq i \leq t)\right). \tag{1.9}$$

For a collection of $p \times p$ random matrices $\{A_i : i \geq 1\}$, we are interested in the convergence of the sequence of NCP $(\text{Span}\{A_i, A_i^* : i \geq 1\}, p^{-1}E\text{Tr})$ as $p \to \infty$. Moreover, convergence of NCP is closely related to the convergence of the spectral distribution of a matrix. To establish the LSD of a $p \times p$ symmetric random matrix A_p by the moment method, a crucial step is to show $\lim p^{-1}E\text{Tr}(A_p^k) < \infty$, $\forall k \geq 1$.

Therefore, the convergence of the NCP $(\text{Span}\{A_p\}, p^{-1}E\text{Tr})$ with some additional effort, yields the limiting eigenvalue distribution of A_p. *Similarly, for a collection of $p \times p$ random matrices $\{A_i : i \geq 1\}$, the convergence of the NCP $(Span\{A_i, A_i : i \geq 1\}, p^{-1}ETr)$ with some additional effort, yields the LSD of any symmetric polynomial in $\{A_i, A_i^* : i \geq 1\}$.*

To describe the limiting NCP and LSD, we have often used non-commutative *free variables*. Free variables are analogue of independent variables in commutative probability.

1.4 Summary of Chapter 5

In Chapter 5, we have put to use the machinery developed in Chapter 4 to study the LSD of any symmetric polynomial in the sample autocovariance matrices $\{\hat{\Gamma}_u\}$, along with their joint convergence.

For that we have first expressed these matrices in a suitable form. Recall the independent matrix Z defined in Section 1.3 and the sequence of coefficient matrices $\{\psi_j\}$ in (1.3). Let $\{P_j : j = 0, \pm 1, \pm 2, \ldots\}$ be a sequence of $n \times n$ matrices where P_j has entries equal to one on the j-th upper diagonal and 0 otherwise. Note that $P_0 = I_n$ where I_n is the $n \times n$ identity matrix, and $P_j = P_{-j}^*$, $\forall j$. Define

$$\Delta_u = \frac{1}{n} \sum_{j,j'=0}^{q} \psi_j Z P_{j-j'+u} Z^* \psi_{j'}^*, \quad \forall u = 0, 1, 2, \ldots. \tag{1.10}$$

In Chapter 7, we have proved that $\{\Delta_u\}$ approximates $\{\hat{\Gamma}_u\}$ as far as the LSD and joint convergence are concerned. With this in mind, first we have studied the matrices $\{\Delta_u\}$ in Chapters 5 and 6 respectively for the cases $p/n \to y > 0$ and $p/n \to 0$.

Indeed we have broadened our scope significantly and have dealt with a more general set up where we have

1. *more than one independent matrices,*

2. any $n \times n$ matrices between Z and Z^* instead of typical matrices $\{P_j\}$ and

3. polynomials which contain several (possibly independent) (Z, Z^*) pairs.

Suppose we have matrices $Z_u = ((\varepsilon_{u,t,i}))_{p\times n}, \ 1 \leq u \leq U$, where $\{\varepsilon_{u,t,i} : u, i, j \geq 0\}$ are independent with mean 0 and variance 1. Note that each Z_u is an independent matrix and moreover, they are independent among themselves.

Also suppose $\{B_{2i-1} : 1 \leq i \leq K\}$ and $\{B_{2i} : 1 \leq i \leq L\}$ are constant matrices of order $p \times p$ and $n \times n$ respectively.

Consider all $p \times p$ matrices

$$\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} = \prod_{i=1}^{k_l} \left(n^{-1} A_{l,2i-1} Z_{u_{l,i}} A_{l,2i} Z^*_{u_{l,i}} \right) A_{l,2k_l+1}, \tag{1.11}$$

where $\{A_{l,2i-1}\}$, $\{A_{l,2i}\}$ and $\{Z_{u_{l,i}}\}$ are matrices from the collections $\{B_{2i-1} : 1 \leq i \leq K\}$, $\{B_{2i} : 1 \leq i \leq L\}$ and $\{Z_i : 1 \leq i \leq U\}$ respectively. As the sample variance-covariance matrix $n^{-1}ZZ^*$ (without centering) is a special case of the above matrices, we have called them *generalized dispersion matrices.*

Consider the sequence of NCP $(\mathcal{U}_p, \, p^{-1}E\mathrm{Tr})$, where

$$\mathcal{U}_p = \mathrm{Span}\left(\mathbb{P}_{l,(u_{l,1},\ldots,u_{l,k_l})} : \ l, k_l \geq 1 \right). \tag{1.12}$$

Here we are interested in the convergence of $(\mathcal{U}_p, p^{-1}E\mathrm{Tr})$. For this it is convenient to use embedding.

Recall the Wigner matrix in Section 1.3. We have first embedded Z_u into a Wigner matrix W_u of order $(n + p)$. Thus

$$W_u = \begin{pmatrix} W^{(1u)}_{p\times p} & Z_u \\ Z^*_u & W^{(2u)}_{n\times n} \end{pmatrix}, \tag{1.13}$$

where $\{W^{(iu)} : i = 1, 2, \ u \geq 1\}$ are independent Wigner matrices and are independent of $\{Z_u\}$.

For any matrices B and D of order p and n respectively, let $\bar{B}$ and $\underline{D}$ of order $(n+p)$ be the matrices

$$\bar{B} = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}, \quad \underline{D} = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix}. \tag{1.14}$$

It is easy to see that

$$\bar{\mathbb{P}}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} = \prod_{i=1}^{k_l} \left(n^{-1} \bar{A}_{l,2i-1} W_{u_{l,i}} \underline{A}_{l,2i} W^*_{u_{l,i}} \right) \bar{A}_{l,2k_l+1}. \tag{1.15}$$

Consider the sequence of NCP $(\bar{\mathcal{U}}_p,\ (n+p)^{-1} E\mathrm{Tr})$, where

$$\bar{\mathcal{U}}_p = \mathrm{Span}\left(\bar{\mathbb{P}}_{l,(u_{l,1},\ldots,u_{l,k_l})} :\ l, k_l \geq 1 \right). \tag{1.16}$$

Convergence of $(\bar{\mathcal{U}}_p,\ (n+p)^{-1} E\mathrm{Tr})$ is easy to describe by using tools from non-commutative probability, discussed in Chapter 4. Then we have expressed the limit of $(\mathcal{U}_p, p^{-1} E\mathrm{Tr})$ in terms of the limit of $(\bar{\mathcal{U}}_p,\ (n+p)^{-1} E\mathrm{Tr})$.

In Section 5.2.2, we have provided the idea behind the limit. Then in Theorem 5.3.1 we have stated the result on convergence of $(\bar{\mathcal{U}}_p,\ (n+p)^{-1} E\mathrm{Tr})$ and $(\mathcal{U}_p,\ p^{-1} E\mathrm{Tr})$. The limiting NCP can be expressed in terms of some free variables.

As discussed in Section 1.3, NCP convergence with some additional effort guarantees existence of the LSD. Theorem 5.4.1 states that the LSD of any symmetric polynomial in $\{\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})}\}$ exists and the limit can be expressed in terms of some freely independent variables.

Consider the polynomials $\Delta \in \mathcal{U}_p$ of the form

$$\Delta = \frac{1}{n} \sum_{i=1}^{q} B_{4i-3} Z B_{2i} Z^* B_{4i-1}. \tag{1.17}$$

We have assumed appropriate conditions on $\{B_i\}$ so that Δ is symmetric.

Note that all the existing LSD results in the literature, discussed in Chapter

4 (see (c) and (d) in Section 1.3), are for random matrices which are special cases of Δ. Moreover, the matrices $\{\Delta_u\}$, which are defined in (1.10) and which will approximate $\{\hat{\Gamma}_u\}$, are also special cases of Δ.

However, most of the existing LSD discussed in Chapter 4 (see (c) and (d) in Section 1.3) are in terms of Stieltjes transform. Therefore to show that these results follow from Theorem 5.4.1, we have investigated the Stieltjes transform of the LSD of Δ. In Theorem 5.4.5, we have provided the Stieltjes transform of the LSD of Δ. In Section 5.4.2, we have shown how the existing LSD results ((c) and (d) in Section 1.3)) in the literature follow as special cases of our LSD results.

1.5 Summary of Chapter 6

In Chapter 6, we have studied $\{\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})}\}$ defined in (1.11) as $p, n(p) \to \infty$, $p/n \to 0$. The embedding technique that we have used in Chapter 5, does not work in this case as here the growth of p and n are not comparable. Moreover, very different centering and scaling of $\{\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})}\}$ are needed to get non-degenerate and non-trivial limits. Define the centered and scaled matrices

$$\mathcal{R}_{l,(u_{l,1},\ldots,u_{l,k_l})} = (n/p)^{1/2}(\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} - \mathbb{G}_{l,k_l}), \text{ where} \tag{1.18}$$

$$\mathbb{G}_{l,k_l} = \left(\prod_{i=1}^{k_l} n^{-1}\mathrm{Tr}\,(A_{l,2i})\right)\prod_{i=0}^{k_l} A_{l,2i+1} \tag{1.19}$$

are the centering matrices. We have then considered the convergence of the sequence of NCP $(\mathcal{V}_p, p^{-1}E\mathrm{Tr})$ as $p, n(p) \to \infty$, $p/n \to 0$, where

$$\mathcal{V}_p = \mathrm{Span}\{\mathcal{R}_{l,(u_{l,1},\ldots,u_{l,k_l})} : l, k_l \geq 1\}. \tag{1.20}$$

In Section 6.2.2, we have shown why $\{\mathbb{G}_{l,k_l}\}$ is the correct centering and $\sqrt{np^{-1}}$ is the correct scaling. In Section 6.2.3, we have discussed the idea behind the limit. Then in Theorem 6.3.1, we have stated the result on convergence of this sequence of NCP. The limiting NCP can be expressed in terms of some free variables. In Theorem 6.4.1, we have stated that the LSD of any symmetric polynomial in $\{\mathcal{R}_{l,(u_{l,1},\ldots,u_{l,k_l})}\}$ exists and have expressed it in terms of free variables. In Section 6.4.1, Theorem 6.4.2, we have derived the Stieltjes transform of this LSD. We have also shown how the existing LSD results in the literature (see (e), (f) and (g) in Section 1.3) follow as special cases of our LSD results.

1.6 Summary of Chapter 7

In Chapter 7, we have applied our results of Chapters 5 and 6 to show existence of the LSD of any symmetric polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$ for the infinite dimensional MA processes. In the literature, such results are known only for the particular polynomial $\{\hat{\Gamma}_u + \hat{\Gamma}_u^*\}$ and under very strong assumptions.

In Theorems 7.3.1, 7.3.4, 7.3.15 and 7.3.17, we have shown that under most natural assumptions, the LSD of any symmetric polynomial in $\{\hat{\Gamma}_u\}$ exists for both the cases $p/n \to y > 0$ and $p/n \to 0$. Moreover, apparently for the first time in the literature, we have described the limits in terms of some free variables. Finally we have shown how the existing results follow from our results under *significantly weaker conditions.*

1.7 Summary of Chapter 8

The results obtained in Chapter 7 have plenty of potential for application in high-dimensional time series. In Chapter 8, we have discussed a couple of applications:

1. a model identification problem, namely determination of the *unknown order* of infinite dimensional moving average (MA) and autoregressive (AR) processes and,

2. *testing* of hypotheses using asymptotic normality of traces of polynomials.

In the univariate set up, a plot of the sample autocovariances provides a method to identify the order of an MA process. If the sample autocovariances are almost equal to zero for order $u > \hat{q}$, then $\hat{q}$ is taken to be an estimator of the unknown order. A similar method that uses the sample *partial autocovariances*, is also applicable for AR processes. The theoretical support for this method is the fact that the population autocovariances of order greater than q are all zero for an MA(q) process. Similarly, for an AR(r) process, population partial autocovariances of order greater than r vanish. Moreover, since the sample autocovariances are consistent for the population autocovariances, for large enough sample size, the sample autocovariances are close to population autocovariances.

In the high-dimensional setting, no equivalent method seems to be available in the literature to determine the unknown order of MA and AR processes. The above method cannot be extended naively since, as we have seen in Chapter 3, the sample autocovariance matrices $\{\hat{\Gamma}_u\}$ are not consistent for the population autocovariance matrices $\{\Gamma_u\}$. Nevertheless, we have shown that $\{\hat{\Gamma}_u\}$ can be used graphically for order determination of high-dimensional processes.

For the infinite dimensional MA(q) process, when $p/n \to y > 0$, a result in Chapter 7 has guaranteed that for large sample size, ESD of $\hat{\Gamma}_u\hat{\Gamma}_u^*$ are close for $u > q$ and different for $0 \leq u \leq q$. When $p/n \to 0$, a similar result for $\sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - \Gamma_u - \Gamma_u)$ has also been guaranteed by a result in Chapter 7. This property of sample autocovariance matrices has been used to identify graphically the unknown order of an MA process. For more details see Section 8.2.

To apply a similar idea to an AR(r) process defined in (1.4) with unknown parameter matrices, we first needed consistent estimators of the parameter matrices $\{A_i : 1 \leq i \leq r\}$. Note that we have already obtained such estimators in Chapter 3. Let us denote these estimators by $\{\hat{A}_i^{(r)} : 1 \leq i \leq r\}$. Now, suppose r is unknown. Consider the residual process $\{\hat{\varepsilon}_t^{(s)}\}$ after fitting the AR(s) process using $\{\hat{A}_i^{(s)} : 1 \leq i \leq s\}$. In Theorem 8.3.1 and Remark 8.3.2, we have argued that the residual process $\{\hat{\varepsilon}_t^{(s)}\}$ behaves like an MA(0) process if and only if $s = r$, the true order of the AR process. We have used these results to identify graphically the unknown order of an AR processes.

Linear spectral statistics of a random matrix M are of the form $\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i)$ where $\{\lambda_i\}$ are eigenvalues of M and f is a "suitable" function. Such statistics have been discussed in Diaconis and Evans [2001], Bai and Silverstein [2004] and Bai et al. [2009]. Asymptotic normality of these statistics is extremely useful in statistical inference. While we have not discussed these statistics in general in this thesis, we have dealt with a specific spectral linear statistics, namely traces of polynomials in sample autocovariance matrices. In Section 8.4, we have established the asymptotic normality of these statistics and have suggested how it may be used for testing problems in high-dimensional time series.

Chapter 2

Estimation of large dispersion matrix for dependent observations

2.1 Introduction

In Chapter 1, we have discussed how high dimensional data are becoming more prevalent with the advent of new methods and technologies in bio-sciences, image processing, social network system and internet. Often these data are represented in the form of a $p \times n$ matrix, called a *data matrix*, as follows:

$$X_{p\times n} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{p2} & x_{p3} & \dots & x_{pn} \end{bmatrix} \tag{2.1}$$

where the dimension p is assumed to be increasing with the sample size n i.e., $p = p(n) \to \infty$ as $n \to \infty$. Let $C_{ip} = (x_{1i}, x_{2i}, \dots, x_{pi})^*$, $1 \leq i \leq n$, be the i-th column of the data matrix X. Throughout this chapter, we assume that $\{C_{ip}\}$ are identically distributed with mean 0 and variance-covariance matrix Σ_p. In Chapter 1, we have provided the motivation that in the multivariate set up, estimation of Σ_p is crucial as many statistical procedures such as classification problem, principal component analysis are all based on the variance-covariance matrix. In this chapter, we shall discuss estimation of Σ_p when $\{C_{ip}\}$ are 'weakly' dependent random vectors and when $n, p = p(n) \to \infty$. In particular, we assume $n^{-1} \log p \to 0$ i.e. p is allowed to increase at an exponential rate of n. The meaning of 'weak' dependence will be clear later in Section 2.3.1.

In the finite dimensional case i.e. when p is fixed, the sample variance-covariance matrix is a consistent estimator of Σ_p. In high dimensional setting, many researchers have shown that the sample variance-covariance matrix fails to estimate the population variance-covariance matrix consistently (where consistency is defined in some natural way), even for i.i.d. $\{C_{ip}\}$. This is also supported by the simulation results given in Example 2.2.1. This is due to the increase in the number of unknown parameters along with the sample size. As a remedy we need some restrictions on the parameter space and modifications of the basic estimator, the sample variance-covariance matrix. This modification is called *covariance regularization*. There are many covariance regularization techniques available for i.i.d. observations in the literature. These have appeared while studying different aspects such as regression, classification or principal component analysis.

One such covariance regularization is due to Bickel and Levina [2008]. It shall play a crucial role in this chapter and is discussed in details in Section 2.2.1. They proved that the suitably *banded* and *tapered* sample variance-covariance matrices are consistent in the *operator norm* for the population variance-covariance matrix as long as $n^{-1} \log p \to 0$, uniformly over some fairly natural well conditioned families of covariance matrices. They also obtained the explicit rates of these convergences. For more details, see Theorems 2.2.2 and 2.2.6.

However, the independence assumption on $\{C_{ip}\}$ is questionable and many researchers have provided evidence of its lack. Efron [2009] proposed the matrix-variate normal distribution as a model for dependent $\{C_{ip}\}$. Allen and Tibshirani [2010] is the only work in this model which estimates Σ_p. In Section 2.2.2, we briefly discuss their covariance regularization.

In Section 2.3, we consider more general models (we call these *weak models*) and give examples of a huge class of models which are not accommodated by the model assumption of Allen and Tibshirani [2010] (see Examples 2.3.1-2.3.3). We allow dependence of appropriate nature between the columns. We call this dependence the *cross covariance strucuture*. In this chapter we consider three different restrictions on cross covariance structures. In the first case, the restriction is on the growth of the powers of the trace of certain matrices derived from the cross covariance structure. In the second case, the dependence among any two columns weakens as the lag between them increases and in the third case we assume weak dependence among the high-indexed columns. See Section 2.3.1 for details.

In the first case, we show that the convergence rate of the banded estimator is the same as in the i.i.d. case of Theorem 2.2.2 (see Theorem 2.4.4) under a *trace condition*. In Remarks 2.4.5-2.4.8, we also provide some sufficient conditions that imply this trace condition. The other two weak models do not fall under the purview of Theorem 2.4.4. Under appropriate conditions we obtain explicit rates of convergence for the banded estimators (see Theorems 2.4.9 and 2.4.11). In particular, for all three cases, the suitably banded variance-covariance matrix continues to remain consistent in operator norm.

Banded estimators are not necessarily positive definite. So we consider tapered estimators that preserve the positive definiteness of the sample variance-covariance matrix. In Theorem 2.4.13, we obtain the rates of convergence of the tapered estimator for all the three weak models defined in Section 2.3.1. In particular the tapered estimator continues to remain consistent in operator norm in these dependent situations.

The growth rate of p and the convergence rates for the tapered and the banded estimators are in general slower than the i.i.d. case and there is a trade-off between these rates and the extent of dependency.

The main material of this chapter is taken from Bhattacharjee and Bose [2014a].

2.2 A brief literature review: some necessary results and motivation

For better understating of our models and assumptions for dependent $\{C_{ip}\}$, let us first discuss the standard case when $\{C_{ip}\}$ are i.i.d. with mean 0 and variance-covariance matrix Σ_p. The natural estimator of Σ_p that one can immediately talk about is the sample variance-covariance matrix:

$$\hat{\Sigma}_{p,n} = \frac{1}{n}\sum_{i=1}^{n} C_{ip}C_{ip}^{*}. \tag{2.2}$$

(Throughout this thesis, we deal with only real vectors and matrices. By $*$, here we mean 'transpose' of vectors or matrices.) For each $n \geq 1$, this is a *moment estimator* of Σ_p.

For any square matrix A_p of order p, define

$$\lambda_{\max}(A_p) \quad = \quad \text{the largest eigenvalue of } A_p. \tag{2.3}$$

The L_2 *norm* or the *operator norm* of A_p is defined as

$$||A_p||_2 = \sqrt{\lambda_{\max}(A_p^{*}A_p)}. \tag{2.4}$$

Let $\xrightarrow{P}$ denote convergence in *probability*. An estimator $\hat{A}_{p,n}$, based on a sample of size n, is called *consistent in operator norm* for A_p if

$$||A_p - \hat{A}_{p,n}||_2 \xrightarrow{P} 0, \quad \text{as} \quad n \to \infty. \tag{2.5}$$

Throughout this thesis, by consistent estimator we always mean consistency in operator norm unless otherwise explicitly mentioned.

It is well known that, in finite dimensional set up, $\hat{\Sigma}_{p,n}$ is consistent for Σ_p. But in case of high dimensional set up, as discussed in Chapter 1, many researchers have shown that $\hat{\Sigma}_{p,n}$ may turn out to be a very bad estimator of Σ_p. The following simulation results also support this.

For any matrix M of order $k \times l$, let

$$M(i,j) = m_{ij} = \text{ the } (i,j)\text{-th element of } M,\ \forall 1 \leq i \leq k,\ 1 \leq j \leq l. \tag{2.6}$$

We often write

$$M = ((m_{i,j}))_{k\times l} \text{ or } M = ((M(i,j)))_{k\times l}. \tag{2.7}$$

For a given sequence $\{a_i\}$, let $\text{diag}(a_1, a_2, \ldots, a_k)_{k\times k}$ be the diagonal matrix of order k with diagonal elements $\{a_1, a_2, \ldots, a_k\}$. Let

$$I_k = \text{diag}(1, 1, \ldots, 1)_{k\times k}, \tag{2.8}$$

$$J_k = ((1))_{k\times k}. \tag{2.9}$$

Example 2.2.1. *We choose three different Σ_p, namely $\Sigma_{1,p} = diag(1, 2, \ldots, p)$, $\Sigma_{2,p} = diag(1, 2^{-1}, \ldots, p^{-1})$ and $\Sigma_{3,p} = 0.5I_p + 0.5J_p$. Take $p = [e^{n^{0.2}}]$ and $n = 10, 15, 20, 25$. For each n, consider independent random samples $\{X_{ijk} : i = 1, 2, 3,\ 1 \leq j \leq n,\ 1 \leq k \leq 1000\}$, where*

$$X_{ijk} \sim \mathcal{N}_p(0, \Sigma_{i,p}). \tag{2.10}$$

For each choice of n, we compute

$$R_{in} = \frac{1}{1000}\sum_{k=1}^{1000} ||\frac{1}{n}\sum_{j=1}^{n} X_{ijk}X_{ijk}^{*} - \Sigma_{i,p}||_2. \tag{2.11}$$

The simulation results are reported in the following table.

Table 1: value of $\{R_{in} : i = 1, 2, 3,\ n = 10, 15, 20, 25\}$

n	10	15	20	25
p	8	21	55	149
R_{1n}	7.042	45.687	274.339	1769.052
R_{2n}	6.415	18.288	50.329	81.486
R_{3n}	17.027	38.312	78.915	195.877

The increasing value of $\{R_{in}\}$ *with increase in* n *indicates that the sample variance-covariance matrix fails to estimate the population variance-covariance matrices consistently in the high dimensional setting.*

Note that the total number of "parameters" in Σ_p is same as the total number of entries on or above the diagonal of Σ_p and equals $0.5p(p+1)$. In high dimensional setting, inconsistency of $\hat{\Sigma}_{p,n}$ is due to the increasing number of unknown parameters. Therefore, appropriate restrictions/modifications on Σ_p and $\hat{\Sigma}_{p,n}$ become necessary to obtain a consistent estimator of Σ_p. This is known as *covariance regularization*. In Chapter 1, we have discussed different covariance regularization techniques which have been used for regression or classification problem in the literature of high dimensional data/model for i.i.d. observations. In the following subsection, we review covariance regularization by Bickel and Levina [2008] in details as we shall use their techniques later in this chapter.

2.2.1 Covariance regularization by Bickel and Levina [2008]

They assumed that

$$C_{ip} \overset{\text{i.i.d.}}{\sim} \mathcal{N}_p(0, \Sigma_p),\ \forall p. \tag{2.12}$$

For any square matrix A_p of order p, define

$$\lambda_{\min}(A_p) = \text{ smallest eigenvalue of } A_p. \tag{2.13}$$

Note that in Example 2.2.1, we considered three different kinds of dispersion matrices, where

$$\sup_p \lambda_{\max}(\Sigma_{1,p}) = \infty, \quad \inf_p \lambda_{\min}(\Sigma_{2,p}) = 0 \tag{2.14}$$

and dependence among the i-th and j-th variables in $\Sigma_{3,p}$ does not decrease with increase in $|i-j|$. In all these cases, the above simulation shows very bad results. The parameter space considered by Bickel and Levina [2008] does not allow these kinds of variance-covariance matrices. They considered dispersion matrices which are *well conditioned* and have *polynomially decaying corners.* Such classes of matrices are defined as below.

For all $k \geq 1$, let $\{A_k\}$ be a nested sequence of square matrices where A_k is of order k. Define

$$\begin{aligned} A_\infty \quad = \quad & \text{the } \infty \times \infty \text{ extension of the sequence of matrices } \{A_k\} \\ & \text{in the sense that for all } k \geq 1,\ A_k \text{ is the } k \times k \text{ sub-matrix} \\ & \text{constructed by the first } k \text{ rows and columns of } A_\infty. \end{aligned} \tag{2.15}$$

Note that the relation between A_∞ and $\{A_k\}$ is a bijection.

Well conditioned dispersion matrices. A dispersion matrix Σ_∞ is called *well conditioned* if its eigenvalues are bounded away from both 0 and ∞. For any $\epsilon > 0$, the set of all ϵ-*well conditioned* dispersion matrices is given by:

$$\mathcal{W}(\epsilon) = \{\Sigma_\infty : 0 < \epsilon < \inf_p \lambda_{\max}(\Sigma_p) \leq \sup_p \lambda_{\max}(\Sigma_p) < \epsilon^{-1} < \infty\}. \tag{2.16}$$

Hence, a dispersion matrix Σ_∞ is *well conditioned* if

$$\Sigma_\infty \in \mathcal{W} := \bigcup_{\epsilon > 0} \mathcal{W}(\epsilon). \tag{2.17}$$

Clearly, this class of dispersion matrices avoids matrices like $\Sigma_{1,p}$ and $\Sigma_{2,p}$.

Dispersion matrices with polynomially decaying corner. This class of matrices guarantees weak dependence among i-th and j-th variables when $|i-j|$ becomes large. For this purpose, define the *k-corner measure* of a square matrix $A = ((a_{ij}))$ as

$$T(A, k) = \sup_{j} \sum_{i:|i-j|>k} |a_{ij}|. \tag{2.18}$$

This is nothing but the maximum column sum of the matrix $((a_{ij} I(|i-j| > k))$ derived from A. To quantify the weak dependence among i-th and j-th variable for large $|i-j|$, the corner measure $T(\Sigma_\infty, k)$ should decay as k grows. A dispersion matrix Σ_∞ is said to have a *polynomially decaying corner*, if

$$\Sigma_\infty \in \mathcal{X} := \bigcup_{\alpha, C>0} \mathcal{X}(\alpha, C), \text{ where} \tag{2.19}$$

$$\mathcal{X}(\alpha, C) = \{A : T(A, k) \leq Ck^{-\alpha}, \ \forall k \geq 1\}, \ \alpha, C > 0. \tag{2.20}$$

Clearly, the class $\mathcal{X}$ avoids matrices like $\Sigma_{3,p}$.

Let

$$\mathcal{U} = \bigcup_{\epsilon, \alpha, C>0} \mathcal{U}(\epsilon, \alpha, C), \text{ where} \tag{2.21}$$

$$\mathcal{U}(\epsilon, \alpha, C) = \mathcal{W}(\epsilon) \cap \mathcal{X}(\alpha, C), \ \epsilon, \alpha, C > 0. \tag{2.22}$$

Bickel and Levina [2008] gave consistent estimator of the dispersion matrix Σ_∞ in the class $\mathcal{U}$. We shall quote their results later but first some examples.

Example 2.2.2. *For a symmetric Toeplitz matrix $T = ((t_{ij} = t_{|i-j|}))$, its spectral density is given by (assuming $\sum_{u=-\infty}^{\infty} |t_u| < \infty$)*

$$f_T(x) = \sum_{u=-\infty}^{\infty} t_u e^{iux}, \quad \forall \ 0 \leq x < 2\pi.$$

For a function f on a domain A, we define

$$||f||_\infty = \sup_{x \in A} |f(x)|. \tag{2.23}$$

Let $f^{(m)}$ denote the m-th order derivative of f.

For any $\epsilon, C > 0$ and $m \geq 1$, consider the following class of Toeplitz variance-covariance matrices as

$$\begin{aligned} \mathcal{L}(\epsilon, m, C) &= \{\Sigma_\infty = ((\sigma_{ij})) : \sigma_{ij} = \sigma_{|i-j|} \textit{ with spectral density } f_{\Sigma_\infty}, \\ &\quad 0 < \epsilon < ||f_{\Sigma_\infty}||_\infty < \epsilon^{-1},\ ||f^{(m)}_{\Sigma_\infty}||_\infty \leq C\}. \end{aligned}$$

Bickel and Levina [2008] proved that

$$\mathcal{L}(\epsilon, m, C) \subset \mathcal{U}(\epsilon, m-1, C). \tag{2.24}$$

Example 2.2.3. *For any $\epsilon, m, m_2, C, C_1, C_2 > 0$ and $m_1 \geq 1$, consider the following class of dispersion matrices as*

$$\begin{aligned} \mathcal{K}(m, C) &= \{\Sigma : \sigma_{ii} \leq C i^{-m},\ \forall i\}, \\ \mathcal{T}(\epsilon, m_1, m_2, C_1, C_2) &= \{\Sigma_\infty = A + B :\ A \in \mathcal{L}(\epsilon, m_1, C_1),\ B \in \mathcal{K}(m_2, C_2)\}. \end{aligned}$$

Bickel and Levina [2008] proved that

$$\begin{aligned} \mathcal{T}(\epsilon, m_1, m_2, C_1, C_2) &\subset \mathcal{U}(\epsilon', \alpha, C_3),\ \textit{where} \\ \epsilon' &\leq \epsilon^{-1} + C_2 \\ \alpha &\leq \min\{m_1 - 1, 0.5m_2 - 1\} \\ C_3 &\leq \frac{C_1}{m_1 - 1} + \frac{C_2}{0.5m_2 - 1}. \end{aligned}$$

Note that, the sample variance-covariance matrix $\hat{\Sigma}_{p,n}$, defined in (2.2), cannot be consistent for $\Sigma_\infty \in \mathcal{U}$. This is partially because $\hat{\Sigma}_{p,n}$ has no control over its corner entries. To have control over the corners of $\hat{\Sigma}_{p,n}$, Bickel and Levina [2008] considered *banded* and *taperd* version of the sample variance-covariance matrices defined as below.

Banding. For any square matrix $M = ((m_{ij}))$, the *k-banded version* of M is given by the matrix

$$B_k(M) = ((m_{ij} I(|i-j| \leq k))). \tag{2.25}$$

This provides control over the corners of a matrix by choosing the *banding parameter* k appropriately.

The banded version of a matrix has a connection with its corner measure defined in (2.18). To understand this, consider the $(1,1)$ *norm* of the matrix M as

$$||M||_{(1,1)} = \sup_j \sum_{i \geq 1} |m_{ij}|. \tag{2.26}$$

This is nothing but the maximum column sum of the matrix M. Then the following lemma is immediate.

Lemma 2.2.1. *If $M \in \mathcal{X}(\alpha, C)$, then $||B_k(M) - M||_{(1,1)} = T(M,k) \leq Ck^{-\alpha}$.*

Bickel and Levina [2008] proved that, when $n, p = p(n) \to \infty$ such that $n^{-1} \log p \to 0$, the appropriately banded version of the sample variance-covariance matrix $\hat{\Sigma}_{p,n}$ is consistent in operator norm for the population variance-covariance matrix Σ_p provided $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C)$. Obviously the choice of the banding parameter k will depend on both n and the rate of decay of the corners of Σ_∞ i.e., on α. Let us denote it by $k_{n,\alpha}$. The following theorem due to Bickel and Levina [2008] provides a right choice for $k_{n,\alpha}$. For convenience of our further discussions, we shall discuss the detailed proof of Theorem 2.2.2. Let $\{a_n\}$ and $\{b_n\}$ be two positive sequences. Then by $a_n \asymp b_n$, we mean $-\infty < K_1 < \underline{\lim} \frac{a_n}{b_n} \leq \overline{\lim} \frac{a_n}{b_n} \leq K_2 < \infty$.

Theorem 2.2.2. *(Bickel and Levina [2008]) Suppose* $\{C_{ip}\}$ *satisfies (2.12) and* $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C)$. *Then for* $k_{n,\alpha} \asymp (n^{-1}\log p)^{-\frac{1}{2(1+\alpha)}}$,

$$||B_{k_{n,\alpha}}(\hat{\Sigma}_{p,n}) - \Sigma_p||_2 = O_P(k_{n,\alpha}^{-\alpha}). \tag{2.27}$$

To prove Theorem 2.2.2, we need the following two lemmas. For any square symmetric matrix $M = ((m_{ij}))$, let

$$c_j(M) = \text{ the number of non-zero entries in the } j\text{-th column of } M. \tag{2.28}$$

Define the $||\cdot||_\infty$ norm of M as

$$||M||_\infty = \max_{i,j} |m_{ij}|. \tag{2.29}$$

Lemma 2.2.3. *(Golub and van Loan [1996])* $||M||_2 \le ||M||_{(1,1)} \le (\sup_j c_j(M))||M||_\infty$.

Lemma 2.2.4. *(Saulis and Statulevičius [1991]) Let* χ_n^2 *be a centered chi-square variable with* n *degrees of freedom. Then*

$$P(|\chi_n^2 - n| \ge x) \le e^{-\frac{x^2}{4(2n+x)}}, \ \forall x > 0. \tag{2.30}$$

Now we are prepared to discuss the proof of Theorem 2.2.2 as given in Bickel and Levina [2008].

Proof of Theorem 2.2.2.

Step 1. Since $||\cdot||_2$ is a norm, by triangle inequality, we have

$$||B_{k_{n,\alpha}}(\hat{\Sigma}_{p,n}) - \Sigma_p||_2 \le ||B_{k_{n,\alpha}}(\hat{\Sigma}_{p,n}) - B_{k_{n,\alpha}}(\Sigma_p)||_2 + ||B_{k_{n,\alpha}}(\Sigma_p) - \Sigma_p||_2. \tag{2.31}$$

Step 2. By Lemma 2.2.1,

$$||B_{k_n}(\Sigma_p) - \Sigma_p||_2 = T(\Sigma, k_{n,\alpha}) = O(k_{n,\alpha}^{-\alpha}). \tag{2.32}$$

Step 3. Recall $c_j(M)$ in (2.28). By Lemma 2.2.3, as

$$\sup_j c_j(B_{k_{n,\alpha}}(\hat{\Sigma}_{p,n}) - B_{k_{n,\alpha}}(\Sigma_p)) \leq k_{n,\alpha},$$

we have

$$\begin{aligned} ||B_{k_{n,\alpha}}(\hat{\Sigma}_{p,n}) - B_{k_{n,\alpha}}(\Sigma_p)||_2 &\leq k_{n,\alpha}||B_{k_{n,\alpha}}(\hat{\Sigma}_{p,n}) - B_{k_{n,\alpha}}(\Sigma_p)||_\infty & (2.33)\\ &\leq k_{n,\alpha}||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty. & (2.34)\end{aligned}$$

Step 4. In this step, we prove

$$||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty = O_P(\sqrt{n^{-1}\log p}). \tag{2.35}$$

Proof. To show (2.35), we first prove that for some $C_1, C_2 > 0$,

$$P(||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty \geq t_n) \leq C_1 p^2 e^{-C_2 n t_n^2}, \text{ if } \{t_n\} \text{ is bounded.} \tag{2.36}$$

To prove (2.36), note that

$$\begin{aligned} & P(||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty \geq t_n) & (2.37)\\ \leq\ & P(\max_{j,k}|\frac{1}{n}\sum_{i=1}^n X_{ij}X_{ik} - \sigma_{jk}| \geq t_n), \text{ by (2.29)} \\ =\ & P\left(\bigcup_{j,k}\left[|\frac{1}{n}\sum_{i=1}^n X_{ij}X_{ik} - \sigma_{jk}| \geq t_n\right]\right) \\ \leq\ & \sum_{j,k} P\left(|\frac{1}{n}\sum_{i=1}^n X_{ij}X_{ik} - \sigma_{jk}| \geq t_n\right), \quad \text{by Boole's inequality.}\end{aligned}$$

Let us define, for all $1 \leq i \leq n$ and $1 \leq j, k \leq p$,

$$\begin{aligned}
Z_{ij} &= \frac{X_{ij}}{\sqrt{\Sigma_{jj}}}, \quad \rho_{jk} = \frac{\sigma_{jk}}{\sqrt{\Sigma_{jj}\sigma_{kk}}}, & (2.38)\\
U_{jk}^{i} &= \left(\frac{Z_{ij}+Z_{ik}}{\sqrt{2+2\rho_{jk}}}\right), \quad V_{jk}^{i} = \left(\frac{Z_{ij}-Z_{ik}}{\sqrt{2-2\rho_{jk}}}\right), & (2.39)\\
U_{jk} &= (U_{jk}^{1}, U_{jk}^{2}, \ldots, U_{jk}^{n})^{*}, \; V_{jk} = (V_{jk}^{1}, V_{jk}^{2}, \ldots, V_{jk}^{n})^{*}. & (2.40)
\end{aligned}$$

Now for some $C_1, C_2, C_3 > 0$, we have

$$\begin{aligned}
& P\left(|\frac{1}{n}\sum_{i=1}^{n} X_{ij}X_{ik} - \sigma_{jk}| \geq t_n\right) \\
= \; & P\left(|\sum_{i=1}^{n} 4Z_{ij}Z_{ik} - 4n\rho_{jk}| \geq \frac{4nt_n}{\sqrt{\sigma_{jj}\sigma_{kk}}}\right) \\
\leq \; & P\left(|\sum_{i=1}^{n}(Z_{ij}+Z_{ik})^2 - n(2+\rho_{jk})^2| \geq \frac{2nt_n}{\sqrt{\sigma_{jj}\sigma_{kk}}}\right) \\
& + P\left(|\sum_{i=1}^{n}(Z_{ij}-Z_{ik})^2 - n(2-\rho_{jk})^2| \geq \frac{2nt_n}{\sqrt{\sigma_{jj}\sigma_{kk}}}\right) \\
= \; & P\left(|U_{jk}'U_{jk} - n| \geq \frac{nt_n\sqrt{\sigma_{jj}\sigma_{kk}}}{2(\sqrt{\sigma_{jj}\sigma_{kk}} - \sigma_{jk})^2}\right) \\
& \qquad + P\left(|V_{jk}'V_{jk} - n| \geq \frac{nt_n\sqrt{\sigma_{jj}\sigma_{kk}}}{2(\sqrt{\sigma_{jj}\sigma_{kk}} + \sigma_{jk})^2}\right) & (2.41)\\
= \; & P\left(|U_{jk}'U_{jk} - n| \geq C_1 n t_n\right) + P\left(|V_{jk}'V_{jk} - n| \geq C_1 n t_n\right) & (2.42)\\
= \; & 2P(|\chi_n^2 - n| \geq C_1 n t_n) & (2.43)\\
\leq \; & C_2 e^{-C_3 n t_n^2}, \quad \text{provided } \{t_n\} \text{ is bounded.} & (2.44)
\end{aligned}$$

(2.42) holds true as $\Sigma_\infty = ((\sigma_{ij})) \in \mathcal{W}(\epsilon)$ implies

$$\begin{aligned}
\sigma_{jj} &\leq \sup_p \lambda_{\max}(\Sigma_p) < \epsilon^{-1}, \; \forall j, \quad \text{and} & (2.45)\\
2(\sqrt{\sigma_{jj}\sigma_{kk}} \pm \sigma_{jk}) &\leq (\sigma_{jj} + \sigma_{kk} \pm 2\sigma_{jk}) \leq \sup_p \lambda_{\max}(\Sigma_p) < \epsilon^{-1}, \; \forall j, k. & (2.46)
\end{aligned}$$

(2.43) holds because

$$U_{jk}, V_{jk} \sim \mathcal{N}_n(0, I_n),\ \forall j, k \implies U'_{jk}U_{jk}, V'_{jk}V_{jk} \sim \chi^2_n,\ \forall j, k. \tag{2.47}$$

(2.44) then follows from Lemma 2.2.4.

Hence, by (2.38) and (2.44), (2.36) is proved. □

Now by taking $t_n = C_4\sqrt{n^{-1}\log p}$ for some $C_4 > 2C_3^{-1}$, we have

$$C_2 p^2 e^{-C_3 n t_n^2} = C_2 p^{2-C_3C_4} k_{n,\alpha}^{-C_3C_4} \to 0. \tag{2.48}$$

Hence, the proof of (2.35) is complete.

Step 5. By Step 3, we have

$$||B_{k_{n,\alpha}}(\hat{\Sigma}_{p,n}) - B_{k_{n,\alpha}}(\Sigma_p)||_2 = O_P(k_{n,\alpha}\sqrt{n^{-1}\log p}). \tag{2.49}$$

Step 6. Now to choose $k_{n,\alpha}$ appropriately, by Steps 1, 2 and 5, we have

$$k_{n,\alpha}^{-\alpha} = k_{n,\alpha}\sqrt{n^{-1}\log p} \implies k_{n,\alpha} = \left(n^{-1}\log p\right)^{-\frac{1}{2(1+\alpha)}}. \tag{2.50}$$

This completes the proof of Theorem 2.2.2. □

Remark 2.2.5. *Note that the independence assumption on $\{C_{ip}\}$ is not used from (2.37)-(2.42). It is first used in (2.43) and may not hold in dependent cases. Later in Theorem 2.4.4, we shall see that under some 'weak' dependence among $\{C_{ip}\}$, (2.44) directly follows from (2.42) and hence the same convergence rate as in Theorem 2.2.2 will hold true.*

Tapering. Positive definiteness is a desirable property for estimators of any variance-covariance matrix and the banded version of $\hat{\Sigma}_{p,n}$ may not be a positive definite matrix. As the Schur product or component wise product of two positive definite matrices is always positive definite, one can consider the Schur product of $\hat{\Sigma}_{p,n}$ with an appropriate positive definite matrix to preserve the positive definiteness. This leads to the idea of *tapering*.

Let, $g : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+ \cup \{0\}$ be a continuous, non-increasing function such that $g(0) = 1$ and $\lim_{x\to\infty} g(x) = 0$. Now, we define

$$\begin{aligned} A_n &= \{1, 2, 3, \ldots, p(n)\},\ \forall n \geq 1, && (2.51)\\ \Delta_{g,\tau_{n,\alpha}} &= \sum_{j=0}^{n-1} g\left(\frac{j}{\tau_{n,\alpha}}\right), && (2.52)\\ R_{n,\alpha} &= \left(g\left(\frac{|i-j|}{\tau_{n,\alpha}}\right)\right)_{i,j\in A_n}, \quad \forall n \geq 1 \text{ and for some } \tau_{n,\alpha} > 0, && (2.53)\\ * &= \text{Schur or component wise product of two matrices}, && (2.54)\\ R_{\tau_{n,\alpha}}(A) &= A * R_{n,\alpha},\ \ \forall n \geq 1 \ \text{ and for any } p \times p \text{ matrix } A. && (2.55) \end{aligned}$$

Also, g is such that $R_{n,\alpha}$ is positive definite. One such choice is $g(x) = e^{-|x|}$.

Now we have the following Theorem.

Theorem 2.2.6. *(Bickel and Levina [2008]) Suppose* $\Delta_{g,\tau_{n,\alpha}} \asymp (n^{-1}\log p)^{-\frac{1}{2(1+\alpha)}}$. *Then under the same assumptions as in Theorem 2.2.2, we have*

$$||R_{\tau_{n,\alpha}}(\hat{\Sigma}_{p,n}) - \Sigma_p||_2 = O_P((n^{-1}\log p)^{-\frac{\alpha}{2(1+\alpha)}}). \tag{2.56}$$

Bickel and Levina [2008] did not provide a detailed proof of the above theorem. As the proof is similar to that of Theorem 2.2.2, we also omit the proof.

However, the assumption that $\{C_{ip}\}$ are independently distributed is questionable in applications. As discussed in Chapter 1, many researchers gave specific examples of lack of this independence. Hence, there is need for models which allow dependence among $\{C_{ip}\}$.

Efron [2009] proposed the matrix-variate normal as a model for dependent microarrays. Allen and Tibshirani [2010] is the only work in this model where Σ_p is estimated by considering appropriate regularization. In the following subsection we discuss the covariance regularization considered by Allen and Tibshirani [2010]. In Section 2.3, we shall generalize their model.

2.2.2 Covariance regularization by Allen and Tibshirani [2010]

For the data matrix X, defined in (2.1), let us denote

$$vec(X) = (C_{1p}^*, C_{2p}^*, \ldots, C_{np}^*)^*. \tag{2.57}$$

Allen and Tibshirani [2010] assumed that

$$\begin{aligned} vec(X) &\sim \mathcal{N}_{np}(0, \Omega), \text{ where} & (2.58)\\ \Omega &= \begin{pmatrix} \Sigma_p & \lambda_{21}\Sigma_p & \cdots & \lambda_{n1}\Sigma_p \\ \lambda_{21}\Sigma_p & \Sigma_p & \cdots & \lambda_{n2}\Sigma_p \\ \vdots & \vdots & \cdots & \vdots \\ \lambda_{n1}\Sigma_p & \lambda_{n2}\Sigma_p & \cdots & \Sigma_p \end{pmatrix}, & (2.59) \end{aligned}$$

for some $n \times n$ real symmetric matrix $\Lambda_n = ((\lambda_{ij})),\ \lambda_{ii} = 1,\ \forall i$. Here Λ_n and Σ_p are respectively called *column* and *row covariances*.

For any matrix $A = ((a_{ij}))$, let us define

$$||A||_{(1)} = \sum_{i,j} |a_{ij}|, \text{ and } ||A||_{(2)} = \left(\sum_{i,j} |a_{ij}|^2\right)^{1/2}. \tag{2.60}$$

$||\cdot||_{(2)}$ is called the *Hilbert–Schmidt operator norm* or the *Frobenius norm* and note that it is different from $||\cdot||_2$ norm defined in (2.4).

They consider the following penalized log-likelihood

$$\begin{aligned} L(\Sigma_p, \Lambda_n) &= l(\Sigma_p, \Lambda_n) - \rho_c(\alpha||\Lambda_n||_{(1)} + (1-\alpha)||\Lambda_n||_{(2)}) \\ &\quad -\rho_r(\alpha||\Sigma_p||_{(1)} + (1-\alpha)||\Sigma_p||_{(2)}), \end{aligned} \tag{2.61}$$

where $l(\Sigma_p, \Lambda_n)$ is the (Gaussian) log-likelihood of the model (2.58) and, ρ_c and

ρ_r are respectively the column and row penalizing parameters. They obtain an estimator for Σ_p by maximizing (2.61).

The main drawback of the model (2.58) is that the correlation between rows is controlled without considering the effect of the columns; that is,

$$\frac{corr(X_{ki}, X_{lj})}{corr(X_{mi}, X_{mj})} = \frac{\sigma_{kl}}{\sqrt{\sigma_{ll}\sigma_{kk}}}, \ \forall i, j = 1, 2, \ldots, p \text{ and } m = 1, 2, \ldots, n. \tag{2.62}$$

In Examples 2.3.1-2.3.3, in the next section, we shall see that there are many models which are not accommodated by the model considered by Allen and Tibshirani [2010]. We shall then introduce a more general model which can overcome the limitation exhibited in (2.62).

2.3 A more general model and some examples

Recall J_k, A_∞, $*$ and $vec(X)$ defined respectively in (2.9), (2.15), (2.54) and (2.57). We assume that

$$vec(X) \sim \mathcal{N}_{np}(0, \Delta_{np}), \text{ where} \tag{2.63}$$

$$\Delta_{np} = \begin{pmatrix} \Lambda_{11} * \Sigma_p & \Lambda_{21} * \Sigma_p & \cdots & \Lambda_{n1} * \Sigma_p \\ \Lambda_{12} * \Sigma_p & \Lambda_{22} * \Sigma_p & \cdots & \Lambda_{n2} * \Sigma_p \\ \vdots & \vdots & \cdots & \vdots \\ \Lambda_{1n} * \Sigma_p & \Lambda_{2n} * \Sigma_p & \cdots & \Lambda_{nn} * \Sigma_p \end{pmatrix}, \tag{2.64}$$

for some real $p \times p$ matrices $\{\Lambda_{ij}\}$ with $\Lambda_{ij} = \Lambda_{ji}^*$, for all $1 \leq j, i \leq n$ and $\Lambda_{ii} = J_p$ for all $1 \leq i \leq n$.

Recall the set $\mathcal{U}(\epsilon, \alpha, C)$ in (2.22). For the marginal variance-covariance matrix Σ_p, we assume that, for some $\epsilon, \alpha, C > 0$, $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C) \cap \mathcal{V}$, where

$$\mathcal{V} = \{\Sigma_\infty = ((\sigma_{ij})) : \sigma_{ij} \neq 0, \ \forall i, j\}. \tag{2.65}$$

$\Sigma_\infty \in \mathcal{V}$ is an *identifiability condition* since it allows to recover $\{\Lambda_{ij}\}$ uniquely when Δ_{np} and Σ_p are given. Recall the notation in (2.6). By (2.64), note that

$$\Delta((j-1)p+k,(i-1)p+l) \;=\; (\Lambda_{ij} * \Sigma_p)(k,l) = \Lambda_{ij}(k,l)\sigma_{kl}. \quad (2.66)$$

Therefore, the condition $\Sigma_\infty \in \mathcal{V}$, assures that one can recover $\{\Lambda_{ij}\}$ from the matrices Δ_{np} and Σ_p by considering

$$\Lambda_{ij}(k,l) = \frac{\Delta((j-1)p+k,(i-1)p+l)}{\sigma_{kl}},\ \forall 1 \le i,j \le n,\ 1 \le k,l \le p. \quad (2.67)$$

For example, suppose $\{z_t\}$ are one-dimensional and i.i.d. random variables with mean 0 and variance 1. Let

$$w_t = z_t + z_{t-1}, \quad v_t = 0.5v_{t-1} + z_t, \quad \forall t. \quad (2.68)$$

Then $\Sigma_\infty = \mathrm{Var}(w_1, w_2, \ldots) \notin \mathcal{V}$ but $\Sigma_\infty = \mathrm{Var}(v_1, v_2, \ldots) \in \mathcal{V}$. The following examples provide some cases which are accommodated by the model defined in (2.63). We also discuss some cases where the model (2.58) is not applicable but the model (2.63) is.

Example 2.3.1. *Suppose*

$$C_{ip} = A_p C_{(i-1)p} + Z_{ip},\ \forall i = 0, \pm 1, \pm 2, \ldots, \quad (2.69)$$

where each Z_{ip} *is a p-component column vector and i.i.d. with mean zero and* $\mathrm{Var}(Z_{ip}) = \tilde{\Sigma}_p$. *Recall the definition of operator norm in (2.4). Suppose* A_p *is a symmetric square matrix of order p such that* $||A_p||_2 < 1$ *and* $A_p\tilde{\Sigma}_p = \tilde{\Sigma}_p A_p$ *for all p. From the properties of linear operators (see for example Bhatia [2009]), if* $||A_p||_2 < 1$, *then* $(I - A_p)$ *is invertible and* $(I - A_p)^{-1} = (I + A_p + A_p^2 + \ldots)$.

Therefore, it is easy to see that for all $1 \leq i \neq j \leq n$

$$\begin{aligned} \mathrm{Var}(C_{ip}) &= \Sigma_p = (I - A_p^2)^{-1}\tilde{\Sigma}_p, \;\; \textit{and} && (2.70) \\ \mathrm{Cov}(C_{ip}, C_{jp}) &= (I - A_p^2)^{-1} A_p^{|i-j|} = \Sigma_p A_p^{|i-j|}. && (2.71) \end{aligned}$$

Hence,

$$\Delta_{np} = \mathrm{Var}\left(vec(X_{p\times n})\right) = \left(\left(\Sigma_p A_p^{|i-j|} I(i \neq j) + \Sigma_p I(i = j)\right)\right)_{1\leq i,j\leq n}.$$

Recall J_k *defined in (2.9) and the notation in (2.6). As we have mentioned earlier, if* $\Sigma_\infty \in \mathcal{V}$ *then one can express* Δ_{np} *as*

$$\begin{aligned} \Delta_{np} &= ((\Sigma_p * \Lambda_{ij} I(i \neq j) + \Sigma_p * J_p I(i = j)))_{1\leq i,j\leq n}, \;\; \textit{where} && (2.72) \\ \Lambda_{ij}(k,l) &= \frac{(\Sigma_p A_p^{|i-j|})(k,l)}{((I - A_p^2)^{-1}\tilde{\Sigma}_p)(k,l)}, \;\; \forall 1 \leq i \neq j \leq n, \; 1 \leq k,l \leq p. && (2.73) \end{aligned}$$

This satisfies the condition of model (2.58) and (2.59) if for all $1 \leq k,l \leq p, \; 1 \leq i,j \leq n$ *and some* $C_{ij} > 0$,

$$((I - A_p^2)^{-1}\tilde{\Sigma}_p A_p^{|i-j|})(k,l) = C_{ij}((I - A_p^2)^{-1}\tilde{\Sigma}_p)(k,l). \quad (2.74)$$

For example, if $A_p = \alpha I_p$ *for some* $0 < \alpha < 1$*, then the model (2.69) satisfies (2.74). But in general, (2.74) may not hold always. Suppose, for some* $0 < \alpha < 1$,

$$\tilde{\Sigma}_p = I_p \;\; \textit{and} \;\; A_p = ((\alpha I(i + j = p))). \quad (2.75)$$

Then it is easy to see that $\Sigma_p = (1 - \alpha^2)^{-1} I_p$ *and*

$$(I - A_p^2)^{-1}\tilde{\Sigma}_p A_p^{|i-j|} = \begin{cases} (1 - \alpha^2)^{-1}\alpha^{|i-j|} I_p, & \textit{if } |i - j| \textit{ is even} \\ (1 - \alpha^2)^{-1}\alpha^{|i-j|} A_p, & \textit{if } |i - j| \textit{ is odd.} \end{cases} \quad (2.76)$$

Therefore, (2.74) does not hold when $|i-j|$ is odd. This shows that (2.63) is a more general model than (2.58).

It is relevant to address the issue of estimation of $\tilde{\Sigma}_p$ and A_p. This will be addressed in details in Chapter 3.

Example 2.3.2. *Suppose, $\{Z_{ip},\ i=0,\pm1,\pm2,\ldots\}$ is a sequence of p-component column random vectors such that $E(Z_{ip}) = 0\ \forall i$ and $E(Z_{ip}Z^*_{jp}) = D_{|i-j|}\ \forall i,j$. Also, let Y_p be a mean zero p-component column random vector such $\mathrm{Var}(Y_p) = \tilde{\Sigma}_p$ and which is independent of Z_{ip}'s. Recall the Schur product $*$ in (2.54). Define another sequence of p-component mean zero random vector as*

$$C_{ip} = Y_p * Z_{ip},\ i = 0,1,2,\ldots n. \tag{2.77}$$

Clearly, we have $\Delta_{np} = \left(\left(\tilde{\Sigma}_p * D_{|i-j|}\right)\right)_{1\le i,j\le n}$.

Recall J_k defined in (2.9) and the notation in (2.6). Suppose $\Sigma_p = \tilde{\Sigma}_p * D_0$. Then it is easy to see that

$$\Delta_{np} = \left((\Sigma_p * \Lambda_{ij} I(i\neq j) + \Sigma_p * J_p I(i=j))\right)_{1\le i,j\le n},\ \textit{where} \tag{2.78}$$

$$\Lambda_{ij}(k,l) = \frac{D_{|i-j|}(k,l)}{D_0(k,l)},\ \forall 1\le i\neq j\le n,\ 1\le k,l\le p. \tag{2.79}$$

This satisfies the condition of model (2.58) and (2.59) if for all $1\le k,l\le p$, $i\ge 1$ and for some $C_i>0$, we have

$$D_i(k,l) = C_i D_0(k,l). \tag{2.80}$$

But in general, (2.80) may not hold always.

Suppose Z_{ip} satisfies the model (2.69). Then $D_i = (I-A_p)^{-1}\tilde{\Sigma}_p A_p^{|i|}$ for all $i = 0,\pm1,\pm2,\ldots$. Hence, as we have seen in Example 2.3.1, for the choice (2.75) of A_p and $\tilde{\Sigma}_p$, (2.80) is not satisfied and the model (2.58)-(2.59) is not applicable.

Example 2.3.3. *Let*

$$\Delta_{np} = \big(\big(B_p^{i+j} I(i \neq j) + (I - B_p^2)^{-1} I(i = j)\big)\big)_{1 \leq i,j \leq n} \tag{2.81}$$

where B_p is a symmetric $p \times p$ matrix and $||B_p||_2 < 1$ for all p. Then Δ_{np} is always positive semi-definite since

$$\Delta_{np} = \big(B_p\ B_p^2\ \ldots\ B_p^n\big)' \big(B_p\ B_p^2\ \ldots\ B_p^n\big) + Diag\Big(I_p + \sum_{i=1}^{\infty} B_p^{2i} - B_p^{2k},\ k = 1, 2, \ldots, n\Big)$$

where $Diag(A_i,\ i = 1, 2, \ldots, n)$ denotes the block-diagonal matrix with i-th diagonal block as A_i and I_p is the identity matrix of order p.

Recall J_k defined in (2.9) and the notation in (2.6). If $(I - B_p^2)^{-1}(k, l) \neq 0\ \forall k, l$, then we can write

$$\begin{aligned} \Delta_{np} &= \big(\big((I - B_p^2)^{-1} * \Lambda_{ij} I(i \neq j) + (I - B_p^2)^{-1} * J_p I(i = j)\big)\big)_{1 \leq i,j \leq n},\ \textit{where} \\ \Lambda_{ij}(k, l) &= \frac{(B_p^{i+j})(k, l)}{((I - B_p^2)^{-1})(k, l)},\ \forall 1 \leq i \neq j \leq n,\ 1 \leq k, l \leq p. \end{aligned} \tag{2.82}$$

This satisfies the condition of model (2.58) if for all $1 \leq k, l \leq p,\ 1 \leq i, j \leq n$ and some for $C_{ij} > 0$,

$$(B_p^{i+j})(k, l) = C_{ij}((I - B_p^2)^{-1})(k, l). \tag{2.83}$$

For example, if $B_p = \alpha I_p$ for some $0 < \alpha < 1$, then Δ_{np} in (2.81) satisfies (2.83). But in general, (2.83) may not hold always. Suppose, $B_p = A_p$ where A_p is as in (2.75). Then it is easy to see that $(I - B_p^2)^{-1} = (1 - \alpha^2)^{-1} I_p$ and for $i, j \geq 1$,

$$B_p^{i+j} = \begin{cases} \alpha^{i+j} I_p, & \textit{if } i + j \textit{ is even} \\ \alpha^{i+j} B_p, & \textit{if } i + j \textit{ is odd.} \end{cases} \tag{2.84}$$

Therefore, (2.83) does not hold when $i + j$ is odd.

2.3.1 Weak dependence among observations

As promised earlier, we now discuss our assumptions on dependence among observations $\{C_{ip}\}$. As $\{\Lambda_{ij}\}$ indicates dependence among observations, we can separate them out from Δ_{np} and define

$$\nabla_{np} = \begin{bmatrix} J_p & \Lambda_{12} & \Lambda_{13} & \dots & \Lambda_{1n} \\ \Lambda'_{12} & J_p & \Lambda_{23} & \dots & \Lambda_{2n} \\ \Lambda'_{13} & \Lambda'_{23} & J_p & \dots & \Lambda_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Lambda'_{1n} & \Lambda'_{2n} & \Lambda'_{3n} & \dots & J_p \end{bmatrix}.$$

We call it the *covariance structure* of the model (2.63). We now elaborate on the idea of 'weak' dependence mentioned at the beginning of this chapter. We discuss four different assumptions on ∇_{np} showing feeble dependence among columns and provide consistent estimators of Σ_p in each of these cases along with their convergence rates.

(1) A relevant question is under what restrictions on ∇_{np} and Σ_p, can one retain the consistency, preferably with the same convergence rate of the earlier estimators of Σ_p as in case of i.i.d. observations? In Theorem 2.4.4, we shall see that it is sufficient to assume that for some $M > 0$,

$$\sup_{n,j,k} n^{-1} \mathrm{Tr}\left(\left(\Gamma_{\pm}^{jk}\right)^r \right) \le M^r \tag{2.85}$$

where Γ_{+}^{jk} and Γ_{-}^{jk} are two $(n \times n)$ matrices defined by

$$\Gamma_{\pm}^{jk}(p,q) = \begin{cases} \frac{\Lambda_{pq}(jj) \pm \left(\Lambda_{pq}(jk) + \Lambda_{pq}(kj)\right)\rho_{jk} + \Lambda_{pq}(kk)}{2\left(1 \pm \rho_{jk}\right)}, & p \neq q \\ 1, & p = q, \end{cases} \tag{2.86}$$

$1 \le j,k \le p$ and $\rho_{jk} = \sigma_{jk}(\sigma_{jj}\sigma_{kk})^{-\frac{1}{2}}$.

Next, note that, in data where time is one of the latent variables which is responsible for the dependence, one may consider the Toeplitz (stationary) structure $\Lambda_{ij} = \Lambda_{|i-j|}$ for a suitable sequence of matrices $\{\Lambda_i\}$ and if $\Lambda_i = \Lambda^i$, then it yields the autoregressive structure of Example 2.3.1. Example 2.3.2 also has a Toeplitz structure. We have seen an example of Hankel structure $\Lambda_{ij} = \Lambda_{i+j},\ \forall i \neq j$ in Example 2.3.3. Broadly speaking, weak dependence between columns can be modelled by assuming that Λ_{ij} is 'small' when say both indices i and j are large or when $|i-j|$ or $i+j$ is large. While (2.85) demands control over all $\{\Lambda_{ij}\}$, the above assumption (weak dependence for large $|i-j|$ or $i+j$) has control on fewer $\{\Lambda_{ij}\}$. Therefore, results obtained under the assumption (2.85), do not hold true here and hence we need to discuss these cases seperately. Below we provide technical assumptions on these covariance structures so that consistent estimators of Σ_p can be obtained.

Recall $||\cdot||_\infty$ defined in (2.29) and the notation in (2.15). Let, $\{a_n\}_{n=1}^\infty$ be a sequence of non-negative integers such that $n^{-1}a_n < 1,\ \forall n \geq 1$.

(2) Weak dependence among the columns when i and j are large can be modelled as follows:

$$\begin{aligned}
\mathcal{L}_n(a_n) &= \left\{\nabla_{np} : S'(a_n) := \max_{k\geq 1, m\geq 1} ||\Lambda_{a_n+k, a_n+k+m}||_\infty = O\left(n^{-2}a_n\right)\right\}, &(2.87)\\
\mathcal{L}(a_n, n\geq 1) &= \{\nabla_\infty : \nabla_{np} \in \mathcal{L}_n(a_n)\}. &(2.88)
\end{aligned}$$

(3) Weak dependence between i-th and j-th columns when $|i-j|$ is large is modelled as follows:

$$\begin{aligned}
\mathcal{A}_n(a_n) &= \left\{\nabla_{np} = \left(\left(\Lambda_{|i-j|}\right)\right) : S(a_n) := \max_{a_n\leq k\leq n} ||\wedge_k||_\infty = O\left(n^{-2}a_n\right)\right\}, &(2.89)\\
\mathcal{A}(a_n, n\geq 1) &= \left\{\nabla_\infty = \left(\left(\Lambda_{|i-j|}\right)\right) : \nabla_{np} \in \mathcal{A}_n(a_n)\right\}. &(2.90)
\end{aligned}$$

(4) Finally, weak dependence among columns when $(i+j)$ is large, is modelled

by:

$$\mathcal{H}_n(a_n) = \left\{\nabla_{np} = ((\Lambda_{i+j} I\,(i \neq j) + \Lambda_0 I\,(i = j))) : \max_{r \geq a_n} ||\Lambda_r||_\infty = O\left(\frac{a_n}{n^2}\right)\right\}, \tag{2.91}$$

$$\mathcal{H}(a_n, n \geq 1) = \{\nabla_\infty = ((\Lambda_{i+j} I\,(i \neq j) + \Lambda_0 I\,(i = j))) : \nabla_{np} \in \mathcal{H}_n(a_n)\}\,. \tag{2.92}$$

Theorems 2.4.9 and 2.4.11 respectively provide consistent estimator of Σ_p for Cases (2) and (3). Case (4) does not have to be dealt with separately as all bounds for Case (2) will automatically hold for Case (4) due to the following Lemma.

Lemma 2.3.1. $\mathcal{H}\,(a_n, n \geq 1) \subset \mathcal{L}\,([2^{-1}a_n] + 2, n \geq 1)$, *where* $[x]$ *is the largest integer contained in* x.

Proof of the above lemma is immediate by observing that, for $b_n = [2^{-1}a_n] + 2,\ \forall n \geq 1$,

$$\max_{k \geq 1, m \geq 1} ||\Lambda_{b_n+k, b_n+k+m}||_\infty \leq \sup_{r \geq a_n} ||\Lambda_r||_\infty. \tag{2.93}$$

Throughout this chapter, model (2.63) under the assumptions described in (1), (2) and (3), will be referred to as the *weak models*.

2.4 Estimation of Σ_p for the weak model

In this section, we shall provide consistent estimators of Σ_p under the assumptions that ∇_{np} and Σ_p satisfy (2.85) or, $\nabla_\infty \in \mathcal{A}(a_n : n \geq 1)$ or $\mathcal{L}(a_n : n \geq 1)$. Recall the banded and tapered version of a matrix in Section 2.2.1 and the sample variance-covariance matrix $\hat{\Sigma}_{p,n}$ defined in (2.2). As we assume that $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C)$ for some $\epsilon, \alpha, C > 0$, from the experience of Section 2.2.1, we can expect that the banded or tapered version of $\hat{\Sigma}_{p,n}$ can serve our purpose. Let us first concentrate on the banded estimator. The tapered estimator will be discussed later.

2.4.1 Banding

As discussed in the previous section, we are interested not only in just consistency but restrictions on ∇_{np} and Σ_p under which the convergence rate of the banded version of $\hat{\Sigma}_{p,n}$ remains the same as in case of i.i.d. observations dealt with in Theorem 2.2.2. Recall U^i_{jk}, U_{jk}, V^i_{jk} and V_{jk} in (2.39) and (2.40). Note that in Remark 2.2.5, we pointed out that (2.43) may not hold for dependent $\{C_{ip}\}$ as then U^i_{jk} and V^i_{jk} are not independent over i. Recall $\Gamma^{jk}_{\pm}$ defined in (2.86). Note that, under our model assumptions, $U_{jk} \sim \mathcal{N}_n(0, \Gamma^{jk}_{+})$ and $V_{jk} \sim \mathcal{N}_n(0, \Gamma^{jk}_{-})$. Hence the problem boils down to finding conditions on $\Gamma^{jk}_{\pm}$, so that (2.44) follows directly from (2.42). In other words, if $U \sim \mathcal{N}_n(0, \Gamma_{n\times n})$, then under what assumptions on $\Gamma_{n\times n}$ is

$$\begin{aligned} P(|U'U - n| \geq C_1 n t_n) \quad &\leq \quad C_2 e^{-C_3 n t_n^2}, \\ &\text{for any bounded } t_n \text{ and some } C_1, C_2, C_3 > 0\ ? \end{aligned} \tag{2.94}$$

To solve the problem, we need the following lemma on the large deviation rate of a random variable.

Lemma 2.4.1. *(Saulis and Statulevičius [1991]) Suppose $E\xi = 0$ and there exist $\gamma \geq 0$, $H > 0$ and $\overline{\Delta} > 0$ such that*

$$|\mathrm{Cum}_k(\xi)| \leq \left(\frac{k!}{2}\right)^{1+\gamma} \frac{H}{\overline{\Delta}^{k-2}}, \ \ k = 2, 3, 4, \ldots, \tag{2.95}$$

where $|\mathrm{Cum}_k(\xi)| = |\frac{d^k}{dt^k}(\log E(e^{it\xi}))|_{t=0}$, is the k-th order cumulant of ξ. Then for all $x \geq 0$,

$$P\big[\pm\xi \geq x\big] \leq e^{-\frac{x^2}{2}\left(H + x\bar{\Delta}^{\frac{-1}{2\gamma+1}}\right)^{-\frac{2\gamma+1}{\gamma+1}}}.$$

Lemma 2.2.4 easily follows from Lemma 2.4.1 as for $\xi = \big(\chi^2_n - n\big)$, we have $\gamma = 0, H = 4n$ and $\bar{\Delta} = \frac{1}{2}$.

Now to see whether (2.95) is satisfied by $U \sim \mathcal{N}_n(0, \Gamma_{n\times n})$ or not, we need to calculate the characteristic function of $U'U$. For this purpose, we need the following lemma.

Lemma 2.4.2. *Suppose A_k is a $k \times k$ positive definite matrix and recall I_k in (2.8). Then*

$$\int_{\mathbb{R}^k} e^{-\frac{1}{2}y'(A_k - 2itI_k)y} dy = (2\pi)^{\frac{k}{2}} (\det(A_k - 2itI_k))^{-\frac{1}{2}}, \ t \in \mathbb{R}.$$

Proof. Let $\lambda > 0$ be the minimum eigenvalue of A_k. Define f and g as

$$\begin{aligned} g(z) &= (2\pi)^{\frac{k}{2}} \left[\det(A_k - 2zI)\right]^{-\frac{1}{2}}, \ \mathcal{R}e\ z < \lambda, \\ f(z) &= \int_{-\infty}^{\infty} e^{-y'(A_k - 2zI)y} dy, \ \mathcal{R}e\ z < \lambda. \end{aligned}$$

Note that both g and f are well defined. It is easy to check by direct integration that if $z = x \in (-\infty, \lambda)$, then $f(x) = g(x)$. It is also easy to check that both f and g are analytic functions on $\{z : \ \mathcal{R}e\ z < \lambda\}$. Since they agree on $\{z : z = x \in (-\infty, \lambda)\}$, they must be identical functions. Hence $f(it) = g(it), \ t \in \mathbb{R}$ and the proof is complete. □

The following lemma easily follows from the above two lemmas.

Lemma 2.4.3. $U \sim \mathcal{N}_n(0, \Gamma_{n\times n})$ *satisfies (2.94) if for some $M > 0$,*

$$\sup_n \frac{1}{n} \mathrm{Tr}(\Gamma^r_{n\times n}) \le M^r, \ \forall r \ge 1. \tag{2.96}$$

Proof. The characteristic function of $U'U$ is given by

$$\begin{aligned} E(e^{itU'U}) &= (2\pi)^{-n/2} \sqrt{\det(\Gamma^{-1}_{n\times n})} \int_{\mathbb{R}^k} e^{-\frac{1}{2}y'(\Gamma^{-1}_{n\times n} - 2itI_n)y} dy \\ &= (2\pi)^{-n/2} \sqrt{\det(\Gamma^{-1}_{n\times n})} (2\pi)^{\frac{n}{2}} (\det(\Gamma^{-1}_{n\times n} - 2itI_n))^{-\frac{1}{2}}, \text{ by Lemma 2.4.2} \\ &= \left[\det(I_n - 2it\Gamma_{n\times n})\right]^{-\frac{1}{2}}. \end{aligned} \tag{2.97}$$

Hence,

$$\frac{d^r}{dt^r}\log E(e^{itU'U}) = -\frac{1}{2}\sum_{u=1}^{n}\frac{d^r}{dt^r}\log(1-2it\lambda_u)$$

where $\lambda_u,\ 1 \leq u \leq n$, are eigenvalues of $\Gamma_{n\times n}$. So, we have

$$|\mathrm{Cum}_r\,(U'U-n)\,| = |\mathrm{Cum}_r\,(U'U)\,| = |\frac{d^r}{dt^r}\log E(e^{itU'U})|_{t=0} = \frac{1}{2}\sum_{u=1}^{n}(r-1)!2^r\,(\lambda_u)^r.$$

Now, from (2.96), we have

$$|\mathrm{Cum}_r\,(U'U-n)\,| \leq \left(\frac{r!}{2}\right)\frac{4nM^2}{\left(\frac{1}{2M}\right)^{r-2}}.$$

Therefore, (2.95) is satisfied for $\gamma = 0, H = 4nM^2$ and $\bar{\Delta} = (2M)^{-1}$. Hence, using Lemma 2.4.1, for some $C_1, C_2, C_3 > 0$

$$\begin{aligned} P\big[|U'U-n| \geq C_1 nt_n\big] \;&\leq\; e^{-\frac{(C_1nt_n)^2}{2}\left(4nM^2+2C_1Mnt_n\right)^{-1}} \\ &\leq\; C_2e^{-C_3nt_n^2}, \quad \text{provided } t_n \text{ is bounded.} \end{aligned}$$

Therefore, U satisfies (2.94) and hence the proof of Lemma 2.4.3 is complete. $\square$

Lemma 2.4.3 motivates us to state the following theorem providing restrictions on $\{\Gamma_{\pm}^{jk} : j,k \geq 1\}$ so that the banded sample variance-covariance matrix will have the same rate of convergence as in case of i.i.d. observations discussed in Theorem 2.2.2. Recall our model assumptions in (2.63) and the classes of covariance matrices $\mathcal{U}\,(\epsilon,\alpha,C)$ and $\mathcal{V}$ respectively in (2.22) and (2.65). Also recall that Σ_∞ is the $\infty\times\infty$ extension of $\{\Sigma_p\}$. Now we are ready to state the following theorem. This result has appeared in Bhattacharjee and Bose [2014a].

Theorem 2.4.4. *Suppose X satisfies the model assumption (2.63) and $\Sigma_\infty \in \mathcal{U}\,(\epsilon,\alpha,C)\cap\mathcal{V}$ for some $\epsilon,\alpha,C > 0$. Suppose (2.85) holds. Then for $k_{n,\alpha} \asymp \left(n^{-1}\log p\right)^{-\frac{1}{2(\alpha+1)}}$, we have $||B_{k_{n,\alpha}}(\hat{\Sigma}_{p,n}) - \Sigma_p||_2 = O_P(k_{n,\alpha}^{-\alpha})$.*

Proof. Note that by Lemma 2.4.3 and (2.85), (2.42) implies (2.44). Therefore, exactly the same proof as for Theorem 2.2.2 goes through in this case also. □

Often (2.85) is difficult to check. Recall $||\cdot||_\infty$ in (2.29). It is comparatively easy to find $||\Lambda_{ij}||_\infty,\ \forall i, j$. In the following remarks we shall provide some sufficient conditions for (2.85) to hold in terms of $\{||\Lambda_{ij}||_\infty\}$.

Remark 2.4.5. *Suppose, $\Sigma_\infty \in \mathcal{W}(\epsilon)$ for some $\epsilon > 0$ and $\{x_k\}$ is a sequence of non-negative real numbers such that $x_k = x_{-k}$ and $||\Lambda_{ij}||_\infty \le x_{i-j} \quad \forall i \ne j,\ 1 \le i, j \le n$, then (2.85) holds if $\sum |x_k| < \infty$.*

Proof. To prove Remark 2.4.5, we essentially show that

$$\frac{1}{n}\mathrm{Tr}((\Gamma_{\pm}^{jk})^r) \le \left(\sum_{i=1}^{\infty} |x_i|\right)^r,\ \forall 1 \le j, k \le p,\ r \ge 1. \tag{2.98}$$

Fix a $1 \le j, k \le p$ and $r \ge 1$. Note that

$$\frac{1}{n}\mathrm{Tr}((\Gamma_{\pm}^{jk})^r) \quad \le \quad \frac{1}{n}\sum |\Gamma_{\pm}^{jk}(u_1, u_2)\Gamma_{\pm}^{jk}(u_2, u_3)\ldots\Gamma_{\pm}^{jk}(u_r, u_1)|. \tag{2.99}$$

Now, by (2.86)

$$|\Gamma_{\pm}^{jk}(u, v)| \le \left(\frac{(1 + |\rho_{jk}|)^2}{1 - \rho_{jk}^2}\right) ||\Lambda_{uv}||_\infty. \tag{2.100}$$

Moreover, as $\Sigma_\infty \in \mathcal{W}(\epsilon)$

$$1 - \rho_{jk}^2 \ge \sqrt{\sigma_{jj}\sigma_{kk}} - \sigma_{jk}^2 = \det\begin{pmatrix} \sigma_{jj} & \sigma_{jk} \\ \sigma_{jk} & \sigma_{kk} \end{pmatrix} \ge (\inf_p \lambda_{\min}(\Sigma_p))^2 \ge \epsilon^2. \tag{2.101}$$

Now by (2.100), for some $C > 0$

$$|\Gamma_{\pm}^{jk}(u, v)| \le C||\Lambda_{uv}||_\infty \le Cx_{u-v},\ \forall u, v. \tag{2.102}$$

Therefore, by (2.99)

$$\begin{aligned}
\frac{1}{n}\mathrm{Tr}((\Gamma_{\pm}^{jk})^r) &\leq \frac{1}{n}C^r \sum_{u_1\ldots,u_r} x_{u_1-u_2}x_{u_2-u_3}\ldots x_{u_r-u_1} \\
&\leq C^r\left(\sum_{k_1,k_2,\ldots,k_{m-1}=-(n-1)}^{n-1} x_{k_1}x_{k_2}\ldots x_{k_{m-1}}x_{(-\sum_{j=1}^{m-1}k_j)}\right) \qquad (2.103)\\
&\leq C^r\left(\sum_{k=-\infty}^{\infty}|x_k|\right)^r.
\end{aligned}$$

Hence, (2.98) is proved and the proof of Remark 2.4.5 is complete. □

Remark 2.4.6. *Let* $\Lambda_{ij} = 0 \ \ \forall|i-j| > k$. *Then (2.85) holds if*

$$\sum_{l=1}^{k}\left(\sup_{|i-j|=l}||\Lambda_{ij}||_\infty\right) < \infty. \qquad (2.104)$$

Remark 2.4.6 immediately follows from Remark 2.4.5 by observing that

$$||\Lambda_{ij}||_\infty \leq \sup_{|i-j|=l}||\Lambda_{ij}||_\infty,\ \forall|i-j| = l,\ l \geq 1. \qquad (2.105)$$

Remark 2.4.7. *If* $\Lambda_{ij} = \Lambda_{|i-j|} \ \ \forall i,j$ *and* $\Lambda_r = 0 \ \ \forall r > k$, *Then (2.85) will hold if*

$$||\Lambda_r||_\infty < \infty,\ \forall\ 1 \leq r \leq k. \qquad (2.106)$$

Remark 2.4.7 follows from Remark 2.4.6 by observing that $\Lambda_{ij} = \Lambda_l,\ \forall|i-j| = l,\ l \geq 1$ and the sum in (2.104) reduces to $\sum_{l=1}^{k}||\Lambda_l||_\infty$ and it is finite if (2.106) holds.

Now, in the following remark, we provide an example where (2.85) does not hold true. Suppose, $g : [0, 2\pi] \to \mathbb{R}$ is a square integrable function. Then the Fourier coefficients of g are defined as

$$\hat{g}(k) = (2\pi)^{-1}\int_0^{2\pi} g(x)e^{-ikx}dx, \quad k = 0, \pm1, \ldots$$

If g is symmetric (about π), then $\{\hat{g}(k)\}$ are real and $\hat{g}(k) = \hat{g}(-k)\ \forall k$. Let $T_{g,n}$ be the Toeplitz matrix defined by

$$T_{g,n} = ((\hat{g}\,(i-j)))_{1 \le i,j \le n}\,.$$

Remark 2.4.8. *Consider a function $g : [0, 2\pi] \to \mathbb{R}$ which is non-negative, symmetric (about π) and square integrable but is unbounded. Suppose $\Gamma_+^{jk} = T_{g,n},\ \forall j,k$. Then (2.85) does not hold.*

Proof. The proof is an application of Szegö's theorem (see Grenander and Szegö [1958]). Suppose if possible (2.85) holds. Let X_n be a random variable such that

$$P\left(X_n = \lambda_{in}\right) = n^{-1}, \quad i = 1, \ldots, n$$

where $\{\lambda_{1n}, \ldots, \lambda_{nn}\}$ are all the eigenvalues of $T_{g,n}$. By Szegö's theorem $X_n \stackrel{\mathcal{D}}{\to} g(\mathcal{U})$ where $\mathcal{U}$ is a random variable distributed uniformly on $[0, 2\pi]$. Now from inequality (2.85) for all n,

$$EX_n^k = n^{-1} \sum_{i=1}^{n} \lambda_{in}^k = n^{-1} \mathrm{Tr}\left(T_{g,n}^k\right) \le M^k, k = 1, \ldots \tag{2.107}$$

Now, observe that

$$x' T_{g,n} x = \sum_{k,j=1}^{n} x_k x_j \frac{1}{2\pi} \int_0^{2\pi} e^{-i(k-j)x} g(x) dx = \frac{1}{2\pi} \int_0^{2\pi} g(x) |\sum_{k=1}^{n} x_k e^{-ikx}|^2 dx \ge 0.$$

So, $T_{g,n}$ is non-negative definite, that is, X_n is non-negative. Thus (2.107) implies that $\{X_n^k\}$ is uniformly integrable for all $k = 1, 2, \ldots$. As a consequence

$$EX_n^k \to E(g(\mathcal{U}))^k, \quad k = 1, 2 \ldots$$

and using (2.107), $E(g(\mathcal{U}))^k \le M^k, k = 1, \ldots$. From this it is immediate that g is almost everywhere bounded. This contradicts our assumption that g is unbounded. Therefore (2.85) does not hold in this case. □

Recall that Λ_∞ is the $\infty \times \infty$ extension of the matrix Λ_{np} in the sense (2.15) and, the classes of cross covariance structures $\mathcal{A}\,(a_n,\ n \geq 1)$ and $\mathcal{L}\,(a_n,\ n \geq 1)$ are respectively given in (2.90) and (2.88). As discussed in Section 2.3.1, if $\nabla_\infty \in \mathcal{L}\,(a_n,\ n \geq 1)$ or $\mathcal{A}\,(a_n,\ n \geq 1)$, then we cannot say whether (2.85) will hold or not. In these classes, we do not have any control over Λ_{ij} for $\min(i,j) < a_n$ or $|i-j| < a_n$ respectively and moreover $a_n \to \infty$. As the following theorems show, we have a slower rate of convergence for the two classes. This result has appeared in Bhattacharjee and Bose [2014a].

Theorem 2.4.9. *Suppose X satisfies our model assumption (2.63). If $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C) \cap \mathcal{V}$ for some $\epsilon, \alpha, C > 0$ and $\nabla_\infty \in \mathcal{L}(l_n,\ n \geq 1)$ for some non-decreasing sequence $\{l_n\}_{n\geq 1}$ of non-negative integers such that $n^{-1} l_n \log p \to 0$ as $n \to \infty$ and $\liminf n^{-1} l_n^2 \log p > 0$. Then with $k^*_{n,\alpha} \asymp (n^{-1} l_n \log p)^{-\frac{1}{1+\alpha}}$,*

$$||B_{k^*_{n,\alpha}}(\hat{\Sigma}_{p,n}) - \Sigma_p||_2 = O_P\left(k^{*-\alpha}_{n,\alpha}\right).$$

Proof. Steps $1-3$ in the proof of Theorem 2.2.2, continue to hold true when we replace $k_{n,\alpha}$ by $k^*_{n,\alpha}$. However, since the observations are not independent, the rate of convergence of $\hat{\Sigma}_{p,n}$ to Σ_p in $||\cdot||_\infty$ norm as mentioned in Step 4 of the proof of Theorem 2.2.2 does not hold. Instead here we prove

$$||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty = O_P(l_n n^{-1} \log p). \tag{2.108}$$

Note that once (2.108) holds, then steps similar to Steps 5 and 6 in the proof of Theorem 2.2.2, imply

$$||B_{k^*_{n,\alpha}}(\hat{\Sigma}_{p,n}) - B_{k^*_{n,\alpha}}(\Sigma_p)||_2 = O_P(k^*_{n,\alpha} l_n n^{-1} \log p). \tag{2.109}$$

Hence, to choose $k^*_{n,\alpha}$ appropriately, we set

$$k^{*-\alpha}_{n,\alpha} = k^*_{n,\alpha} l_n n^{-1} \log p \implies k^*_{n,\alpha} = \left(l_n n^{-1} \log p\right)^{-\frac{1}{1+\alpha}}. \tag{2.110}$$

Hence proof of Theorem 2.4.9 will be complete if we can show (2.108) holds.

Recall U^i_{jk}, V^i_{jk}, U_{jk} and V_{jk} in (2.41) and (2.40). Now to prove (2.108), again note that the same calculations from (2.37) to (2.42) go through in this case also as the independence assumption is first used in (2.43). Therefore, we have

$$\begin{aligned} & P(||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty \geq t_n) \\ \leq & \sum_{j,k=1}^{p} \left(P\left(|U'_{jk}U_{jk} - n| \geq C_1 n t_n\right) + P\left(|V'_{jk}V_{jk} - n| \geq C_1 n t_n\right)\right). \end{aligned} \tag{2.111}$$

Now, let $U_{jkl_n} = (U^{l_n+1}_{jk}, U^{l_n+2}_{jk}, \ldots, U^n_{jk})'$ and $V_{jkl_n} = (V^{l_n+1}_{jk}, V^{l_n+2}_{jk}, \ldots, V^n_{jk})'$. Suppose $\Gamma^{jkl_n}_{\pm}$ is the variance-covariance matrices of U_{jkl_n} and U_{jkl_n}. Note that $\Gamma^{jkl_n}_{\pm}$ is nothing but a square symmetric positive semi-definite matrix constructed by deleting first l_n rows and columns from $\Gamma^{jk}_{\pm}$. Let $C^{jkl_n}_{\pm} = I_{n-l_n} - \Gamma^{jkl_n\ -1}_{\pm}$. Then we can write

$$\begin{aligned} P\left(|U'_{jk}U_{jk} - n| \geq C_1 n t_n\right) &\leq l_n P(|(U^1_{jk})^2 - 1| \geq C_1 n l_n^{-1} t_n) \\ &\quad + P(|U'_{jkl_n}\Gamma^{jkl_n\ -1}_{+} U_{jkl_n} - (n - l_n)| \geq C_1 n t_n/2) \\ &\quad + P(|U'_{jkl_n} C^{jkl_n}_{+} U_{jkl_n}| \geq C_1 n t_n/2) \\ &= T_1 + T_2 + T_3, \text{ (say)}. \end{aligned} \tag{2.112}$$

Let, $t_n = M l_n n^{-1} \log p \to 0$, for some fix constant $M > 0$. Later M will be chosen appropriately. Now, as $U^i_{jk} \sim \mathcal{N}(0,1)$ for all $i \geq 1$, by Lemma 2.2.4 and for some $C_2, C_3 > 0$, we have

$$T_1 \leq l_n P\left[|\chi^2_1 - 1| \geq C_1 n l_n^{-1} t_n\right] \leq 2 l_n C_2 e^{-C_3 M \log p}. \tag{2.113}$$

Then for some constant $C_4, C_5 > 0$, by Lemma 2.4.1

$$T_2 = P\left[|\chi^2_{(n-l_n)} - (n-l_n)| \geq C_1 n t_n/2\right] \leq C_4 e^{-C_5 \frac{l_n^2}{n}(\log p)^2 M^2}. \tag{2.114}$$

Next, for some $C_6, C_7 > 0$, we have

$$\begin{aligned}
T_3 &= P\left[|\left(U^{jkl_n}\right)'\left(C^{jkl_n}\right)\left(U^{jkl_n}\right)| \geq C_1 n t_n/2\right] \\
&\leq P\left[\max_{U\neq 0} \frac{|\left(U' C^{jkl_n} U\right)|}{U'U}\left(U^{jkl_n}\right)'\left(U^{jkl_n}\right) \geq C_1 n t_n/2\right] \\
&= P\left[\sqrt{\lambda_{\max}\left(C^{jkl_n}\right)'\left(C^{jkl_n}\right)}\left(U^{jkl_n}\right)'\left(U^{jkl_n}\right) \geq C_1 n t_n/2\right] \\
&= P\left[||C^{jkl_n}||_2 \left(U^{jkl_n}\right)'\left(U^{jkl_n}\right) \geq C_1 n t_n/2\right] \\
&\leq nP\left[||C^{jkl_n}||_2\ \chi^2_1 \geq C_1 t_n/2\right] \\
&\leq nC_6 e^{-C_7 \frac{t_n}{||C^{jkl_n}||_2}}, \quad \text{by Lemma 2.2.4.}
\end{aligned}$$

Moreover, it is easy to see that for some $C_8 > 0$, $||C_+^{jkl_n}||_2 \leq n||C_+^{jkl_n}||_\infty \leq nC_8 S'(l_n)$. Hence, putting $t_n = M l_n n^{-1}\log p$, for some constant $C_9,\ C_{10} > 0$,

$$P\left[|\left(U^{jkl_n}\right)'\left(I - \left(\Gamma_+^{jkl_n}\right)^{-1}\right)\left(U^{jkl_n}\right)| \geq C_1 n t_n\right] \leq nC_6 e^{-C_9 \frac{t_n}{nS'(l_n)}} \leq nC_6 e^{-C_{10} M \log p}. \tag{2.115}$$

Similar bound holds for V_{jk} also. By (2.111) to (2.115), for some $C_{11}, C_{12} > 0$ and for all sufficiently large n,

$$\begin{aligned}
& P\left[||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty \geq M n^{-1} l_n \log p\right] \\
\leq\ & 2C_{11} p^2 \left(l_n e^{-C_3 M \log p} + 2e^{-C_5 M^2 \frac{l_n^2}{n}(\log p)^2} + 2ne^{-C_{10} M(\log p)}\right) \\
=\ & C_{12}\left(p^{3-C_3 M} + p^2 e^{-C_5 M^2 \frac{l_n^2}{n}(\log p)^2} + p^{3-C_{10}M}\right).
\end{aligned}$$

If $M > \max\{\frac{3}{C_3}, \frac{3}{C_{10}}\}$, then $p^{3-C_3 M} + p^{3-C_{10}M} \to 0$. The logarithm of the second term is

$$2\log p - C_5 M^2 n^{-1} l_n^2 (\log p)^2 = \log p \left[2 - C_5 M^2 n^{-1} l_n^2 (\log p)\right].$$

Now if $\liminf l_n^2 n^{-1} \log p > 0$ then it is bounded away from zero by S (say). So, if $M > \max\{\frac{3}{c_3}, \frac{3}{C_{10}}, \sqrt{\frac{2}{C_5 S}}\}$, then the second term also tends to zero. This completes the proof of (2.108) and hence the proof of Theorem 2.4.9. $\square$

Remark 2.4.10. *(i) If l_n is of exact order of $(n^{-1}\log p)^{-\frac{1}{2}}$, then the rate of convergence will be same as that for the i.i.d. case. If l_n is of order $(n^{-1}\log p)^{-\beta}$ where β is more than $1/2$, then the rate is slower than the i.i.d. case. Note that $\beta < 1/2$ is not allowed in the theorem as* $\liminf n^{-1} l_n^2 \log p > 0$.

(ii) Theorem 2.4.9 is not applicable in case the sequence $\{l_n\}$ is bounded above. This is because if $n^{-1} l_n \log p \to 0$ then $n^{-1}\log p \to 0$ and hence $n^{-1} l_n^2 \log p \to 0$. Recall $k_{n,\alpha}$ in Theorem 2.2.2. When $\{l_n\}$ is bounded by K, $C_{(K+1)p}$, $C_{(K+2)p}, \ldots$ will be an i.i.d. sample and we can construct the estimator on the basis of this i.i.d. sample i.e. we can consider the $k_{n,\alpha}$ banded version of $\dfrac{1}{n-K} \displaystyle\sum_{i=K+1}^{n} C_{ip} C'_{ip}$ with the same rate as the i.i.d. case.

The next theorem shows consistency of the banded sample variance-covariance matrix when the cross covariance structure $\nabla_\infty \in \mathcal{A}(a_n, n \geq 1)$. This result has appeared in Bhattacharjee and Bose [2014a].

Theorem 2.4.11. *Suppose X satisfies our model assumptions (2.63). If $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C) \cap \mathcal{V}$ for some $\epsilon, \alpha, C > 0$ and $\nabla_\infty \in \mathcal{A}(a_n, n \geq 1)$ for some non-decreasing sequence $\{a_n\}_{n\geq 1}$ of non-negative integers such that $a_n\sqrt{n^{-1}\log p} \to 0$ and $a_n^{-1}\sqrt{n \log p} \to \infty$ as $n \to \infty$. Then with $k_{n,\alpha}^{**} \asymp \left(a_n n^{-\frac{1}{2}}\sqrt{\log p}\right)^{-\frac{1}{1+\alpha}}$,*

$$||B_{k_{n,\alpha}^{**}}(\hat{\Sigma}_{p,n}) - \Sigma_p||_2 = O_P\left(k_{n,\alpha}^{**-\alpha}\right).$$

Proof. As before, note that in the proof of Theorem 2.2.2, Steps $1-3$ hold true when we replace $k_{n,\alpha}$ by $k^{**}_{n,\alpha}$. As observations are not independent, the rate of convergence of $\hat{\Sigma}_{p,n}$ to Σ_p in $||\cdot||_\infty$ norm as mentioned in Step 4 of the proof of Theorem 2.2.2 does not hold and instead here we prove

$$||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty = O_P(a_n\sqrt{n^{-1}\log p}). \tag{2.116}$$

Note that if (2.116) is true, then steps similar to Steps 5 and 6 in the proof of Theorem 2.2.2, will yield

$$||B_{k^{**}_{n,\alpha}}(\hat{\Sigma}_{p,n}) - B_{k^{**}_{n,\alpha}}(\Sigma_p)||_2 = O_P(k^{**}_{n,\alpha}a_n\sqrt{n^{-1}\log p}). \tag{2.117}$$

Then the appropriate choice of $k^{**}_{n,\alpha}$ is obtained by setting

$$k^{**-\alpha}_{n,\alpha} = k^{**}_{n,\alpha}a_n\sqrt{n^{-1}\log p} \implies k^{*}_{n,\alpha} = \left(a_n\sqrt{n^{-1}\log p}\right)^{-\frac{1}{1+\alpha}}. \tag{2.118}$$

Hence proof of Theorem 2.4.11 will be complete if we can show (2.116) holds.

Recall U^i_{jk}, V^i_{jk}, U_{jk} and V_{jk} in (2.41) and (2.40). Now to prove (2.108), again note that the same calculations from (2.37) to (2.42) go through in this case also as the independence assumption is first used in (2.43). Therefore, we have

$$\begin{aligned} P(||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty \geq t_n) \;\leq\; & \sum_{j,k=1}^{p} P\left(|U'_{jk}U_{jk} - n| \geq C_1 n t_n\right) \\ & + \sum_{j,k=1}^{p} P\left(|V'_{jk}V_{jk} - n| \geq C_1 n t_n\right). \end{aligned} \tag{2.119}$$

Let, for all $1 \leq r \leq a_n$,

$$A_{r,a_n} = \{i \in \mathbb{Z}^+ \cup \{0\} : (ia_n + r) \leq n\} \text{ and } C_{r,a_n} = \text{cardinality of } A_{r,a_n}.$$

Let

$$U_{jkra_n} = vec\left(U^i_{jk} : i \in A_{r,a_n}\right) \text{ and } V_{jkra_n} = vec\left(V^i_{jk} : i \in A_{r,a_n}\right).$$

Now, by (2.119), we have,

$$\begin{aligned} P(||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty \geq t_n) &\leq \sum_{j,k}\sum_{r=1}^{a_n} P\left[|(U_{jkra_n})'(U_{jkra_n}) - C_{r,a_n}| \geq \frac{C_1 n t_n}{a_n}\right] \\ &+ \sum_{j,k}\sum_{r=1}^{a_n} P\left[|(V_{jkra_n})'(V_{jkra_n}) - C_{r,a_n}| \geq \frac{C_1 n t_n}{a_n}\right] \end{aligned} \tag{2.120}$$

Recall $\Gamma^{jk}_{\pm}$ in (2.86). Note that for each $1 \leq r \leq a_n$,

$$U_{jkra_n} \sim \mathcal{N}_{C_{r,a_n}}(0, \Gamma^{jkra_n}_{+}) \text{ and } V_{jkra_n} \sim \mathcal{N}_{C_{r,a_n}}(0, \Gamma^{jkra_n}_{-}), \tag{2.121}$$

where $\Gamma^{jkra_n}_{\pm}$ is nothing but the sub-matrix taking A_{r,a_n}-th rows and columns from $\Gamma^{jk}_{\pm}$. Recall I_k in (2.8). Let $C^{jkra_n}_{\pm} = I_{C_{r,a_n}} - \Gamma^{jkar_n}_{\pm}$ $\forall r \geq 1$. Therefore, for some $C_2 > 0$

$$\begin{aligned} P\left[||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty \geq t_n\right] &\leq 2a_n p^2 P\left[|\chi^2_{C_{r,a_n}} - C_{r,a_n}| \geq C_2 n a_n^{-1} t_n\right] \\ &+ \sum_{j,k}\sum_{r=1}^{a_n} P\left[|U_{jkra_n})' C^{jkra_n}_{+} (U_{jkra_n})| \geq C_2 n a_n^{-1} t_n\right] \\ &+ \sum_{j,k}\sum_{r=1}^{a_n} P\left[|(V_{jkra_n})' C^{jkra_n}_{-} (V_{jkra_n})| \geq C_2 n a_n^{-1} t_n\right]. \end{aligned}$$

Again, by Lemma 2.2.4, for $t_n = M a_n (n^{-1}\log p)^{\frac{1}{2}}$ and some $C_3, C_4 > 0$, as $n^{-1}\log p \to 0$ we have

$$P\left[|\chi^2_{C_{r,a_n}} - C_{r,a_n}| \geq C_2 n a_n^{-1} t_n\right] \leq C_3 e^{-C_4 M \log p}. \tag{2.122}$$

Now, as in the proof of Theorem 2.4.9, for some $C_5, C_6 > 0$,

$$P\left[|(U^{jkra_n})'C_+^{jkra_n}(U^{jkra_n})| \geq C_2 n a_n^{-1} t_n\right] \leq nC_5 \exp\big\{ - C_6 t_n a_n^{-1} ||C_+^{jka_n}||_2^{-1}\big\}.$$

Since $\nabla_\infty \in \mathcal{A}(a_n, n \geq 1)$, for some $C_7 > 0$ we have $||C_+^{jkra_n}||_2 \leq C_7 n S(a_n)$. Therefore, putting $t_n = M a_n (n^{-1} \log p)^{\frac{1}{2}}$, we have, for some constant $C_8, C_9 > 0$,

$$P\big[|(U^{jkra_n})'C_+^{jkra_n}(U^{jkra_n})| \geq C_2 {a_n}^{-1} n t_n\big] \leq nC_8 \exp\{-C_9 M {a_n}^{-1}\sqrt{n \log p}\}.$$

Similarly, for some constant $C_{10}, C_{11} > 0$,

$$P\big[|(V_{jkra_n})'C_-^{jkra_n}(V_{jkra_n})| \geq C_2 n a_n^{-1} t_n\big] \leq nC_{10} \exp\{-C_{11} M {a_n}^{-1}\sqrt{n \log p}\}.$$

Hence, for some constant $C_{12}, C_{13}, C_{14} > 0$, we have

$$P\left[||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty \geq t_n\right] \;\; \leq \;\; C_{12}(p^3 e^{-C_{13} M \log p} + p^4 e^{-C_{14} M \frac{\sqrt{n \log p}}{a_n}}). \quad (2.123)$$

Clearly, the first term $\to 0$ as $n \to \infty$ if $M > \frac{3}{C_1}$. Now, since $a_n \sqrt{n^{-1} \log p} \to 0$ and $a_n^{-1}\sqrt{n \log p} \to \infty$, we have, for some constant $C_{14}, C_{16} > 0$,

$$p^4 e^{-C_{14} M \frac{\sqrt{n \log p}}{a_n}} \;\; = \;\; e^{C_{15} \frac{\sqrt{n \log p}}{a_n}\left(a_n \sqrt{\frac{\log p}{n}} - C_{16} M\right)} \to 0.$$

Hence (2.116) is proved and proof of Theorem 2.4.11 is complete. $\square$

Remark 2.4.12. *If $\{a_n\}$ is bounded above, then the rate of convergence reduces to the convergence rate for i.i.d. sample as given in Theorem 2.2.2.*

This completes our discussion on banded estimators of Σ_p.

2.4.2 Tapering

In this section, we consider tapered estimator. Recall the tapered version of $\hat{\Sigma}_{p,n}$ in Section 2.2.1. Let, $g : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+ \cup \{0\}$ be a continuous, non-increasing function such that $g(0) = 1$, $\int_0^\infty g(x) < \infty$ and $1 - g(x) = O(x^\gamma)$ for some $\nu \geq 1$ in some neighborhood of zero. Then we have the following theorem. This result has appeared in Bhattacharjee and Bose [2014a].

Theorem 2.4.13. (a) *Under the conditions of Theorem 2.4.4, if*

$$\tau_{n,\alpha} \asymp \left(n^{-1}\log p\right)^{-\frac{1}{2(1+\gamma)}\left[\frac{\gamma}{1+\alpha}+1\right]},$$

then

$$||R_{\tau_{n,\alpha}}(\hat{\Sigma}_{p,n}) - \Sigma_p||_2 = O_P\left[\left(n^{-1}\log p\right)^{\frac{\gamma\alpha}{2(1+\alpha)(1+\gamma)}}\right].$$

(b) *Under the conditions of Theorem 2.4.9, if*

$$\tau_{n,\alpha} \asymp \left(n^{-1}l_n\log p\right)^{-\frac{1}{2(1+\gamma)}\left[\frac{\gamma}{1+\alpha}+1\right]},$$

then

$$||R_{\tau_{n,\alpha}}(\hat{\Sigma}_{p,n}) - \Sigma_p||_2 = O_P\left[\left(l_n n^{-1}\log p\right)^{\frac{\gamma\alpha}{(1+\alpha)(1+\gamma)}}\right].$$

(c) *Under the conditions of Theorem 2.4.11, if*

$$\tau_{n,\alpha} \asymp \left(a_n n^{-1/2}\sqrt{\log p}\right)^{-\frac{1}{2(1+\gamma)}\left[\frac{\gamma}{1+\alpha}+1\right]},$$

then

$$||R_{\tau_{n,\alpha}}(\hat{\Sigma}_{p,n}) - \Sigma_p||_2 = O_P\left[\left(a_n\sqrt{n^{-1}\log p}\right)^{\frac{\gamma\alpha}{(1+\alpha)(1+\gamma)}}\right].$$

Proof. By Lemma 2.2.3 and triangular inequality,

$$||R_{\tau_{n,\alpha}}\left(\hat{\Sigma}_{p,n}\right) - \Sigma_p||_2 \leq ||R_{\tau_{n,\alpha}}\left(\hat{\Sigma}_{p,n}\right) - R_{\tau_{n,\alpha}}\left(\Sigma_p\right)||_{(1,1)} + ||R_{\tau_{n,\alpha}}\left(\Sigma_p\right) - \Sigma_p||_{(1,1)}.$$

Now, for some constant $C_1 > 0$,

$$||R_{\tau_{n,\alpha}}\left(\hat{\Sigma}_{p,n}\right) - R_{\tau_{n,\alpha}}\left(\Sigma_p\right)||_{(1,1)} \;\; \leq \;\; ||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty\big(2\sum_{l=0}^{p} g(\frac{l}{\tau_{n,\alpha}})\big)$$

$$\leq \tau_{n,\alpha}||\hat{\Sigma}_p - \Sigma_p||_\infty \left[C_1 \int_0^\infty g(x)dx \right]. \tag{2.124}$$

As before we have $||\hat{\Sigma}_{p,n} - \Sigma_p||_\infty = \begin{cases} O_P\left(\sqrt{n^{-1}\log p}\right), & \text{in Theorem 2.4.4} \\ O_P\left(l_n n^{-1}\log p\right), & \text{in Theorem 2.4.9} \\ O_P\left(a_n\sqrt{n^{-1}\log p}\right) & \text{in Theorem 2.4.11.} \end{cases}$

Again, by triangle inequality

$$\begin{aligned} ||R_{\tau_{n,\alpha}}(\Sigma_p) - \Sigma_p||_{(1,1)} \leq\ & ||R_{\tau_{n,\alpha}}(\Sigma_p) - B_{k'_{n,\alpha}}\left[R_{\tau_{n,\alpha}}(\Sigma_p)\right]||_{(1,1)} \\ & + ||B_{k'_{n,\alpha}}\left[R_{\tau_{n,\alpha}}(\Sigma_p)\right] - B_{k'_{n,\alpha}}(\Sigma_p)||_{(1,1)} \\ & + ||B_{k'_{n,\alpha}}(\Sigma_p) - \Sigma_p||_{(1,1)} \end{aligned} \tag{2.125}$$

and by Lemma 2.2.1, we have

$$||B_{k'_{n,\alpha}}(\Sigma_p) - \Sigma_p||_{(1,1)} = O((k'_{n,\alpha})^{-\alpha}) \text{ and}$$

$$||R_{\tau_{n,\alpha}}(\Sigma_p) - B_{k'_{n,\alpha}}\left[R_{\tau_{n,\alpha}}(\Sigma_p)\right]||_{(1,1)} = O((k'_{n,\alpha})^{-\alpha}).$$

Now, for some constant $C_2, C_3 > 0$, as σ_{ij}'s are bounded, for sufficiently large n,

$$\begin{aligned} & ||B_{k'_{n,\alpha}}\left[R_{\tau_{n,\alpha}}(\Sigma_p)\right] - B_{k'_{n,\alpha}}(\Sigma_p)||_{(1,1)} \\ \leq\ & \max_i \sum_{j:|i-j|\leq k'_{n,\alpha}} \left(1 - g\left(\frac{|i-j|}{\tau'_{n,\alpha}}\right)\right)|\sigma_{ij}| \\ \leq\ & C_2 \sum_{l=-k'_{n,\alpha}}^{k'_{n,\alpha}} \left(1 - g\left(\frac{l}{\tau_{n,\alpha}}\right)\right) \leq C_3\left(\frac{k'_{n,\alpha}}{\tau_{n,\alpha}}\right)^\gamma k'_{n,\alpha}. \end{aligned}$$

Now, consider $k'_{n,\alpha} = \begin{cases} (n^{-1}\log p)^{-\frac{\gamma}{2(1+\gamma)(1+\alpha)}}, & \text{for (a)} \\ (l_n n^{-1}\log p)^{-\frac{\gamma}{(1+\gamma)(1+\alpha)}}, & \text{for (b)} \\ \left(a_n n^{-1/2}\sqrt{\log p}\right)^{-\frac{\gamma}{(1+\gamma)(1+\alpha)}}, & \text{for (c)}. \end{cases}$

This completes the proof. □

Remark 2.4.14. *In case of i.i.d. sample, as discussed in Section 2.2.1, Bickel and Levina [2008] claimed, without a detailed proof, that the rate of convergence for the banded and the tapered estimators are same. However, in that article there does not seem to be any condition imposed on the tapering function in the neighborhood of zero. In our proof, there is a term* $||B_{k'_{n,\alpha}}\left[R_{\tau_{n,\alpha}}(\Sigma_p)\right] - B_{k'_{n,\alpha}}(\Sigma_p)||_{(1,1)}$ *(see (2.125)), which is tackled by our condition "* $1 - g(x) = O\left(x^{\gamma}\right)$ *for some* $\gamma \geq 1$ *in some neighborhood of zero", on the tapering function. It appears that under this condition, the convergence rate for the i.i.d. case in Bickel and Levina [2008] would be identical to the rate given in our Theorem 2.4.13* (*a*)*. We did not pursue this.*

To summarize, under weak dependence, the appropriately banded and tapered version of the sample variance-covariance matrix are consistent estimators of the population variance-covariance matrix provided its corners are appropriately decaying and its eigenvalues remain bounded away from 0 and ∞.

We conclude by noting that, as we have discussed in Chapter 1, there are many high dimensional data which are time series in nature. Estimation of population autocovariance matrices of different orders is very important in the analysis of a stationary time series model. The population autocovariance matrix of order 0 is nothing but the population variance-covariance matrix. Therefore, some very specific situations of estimation of autocovariance matrices can be handled by the results of this chapter. The next chapter deals with estimation of autocovariance matrices in details.

Chapter 3

Estimation of large autocovariance matrices for linear process

3.1 Introduction

We have seen in Chapter 1 that there are many high dimensional data which are time series in nature. A huge class of time series models is *weak or covariance stationary* time series. Let $\{X_{t,p} : t = 0, \pm 1, \pm 2, \ldots\}$ be p-dimensional random vectors with $E(X_{t,p}) = 0 \; \forall t$. $\{X_{t,p}\}$ is called a weak/covariance stationary time series if and only if, for all $u \geq 0$, the $p \times p$ matrix

$$\Gamma_{u,p} = E(X_{t,p} X^*_{t+u,p}) \tag{3.1}$$

does not depend on t and is a function of u only. $\Gamma_{u,p}$ is called the *population autocovariance matrix* of order u. Note that $\Gamma_{0,p}$ is the population variance-covariance matrix of $\{X_{t,p}\}$. In this chapter, we are mainly focused on estimating $\{\Gamma_{u,p}\}$ in high dimensional set up.

Some of the more common existing weak stationary high-dimensional time series models in the literature are infinite dimensional IID processes, infinite dimensional finite order moving average processes (MA) and infinite dimensional vector autoregressive processes (IVAR) with *i.i.d.* innovations. A detailed description of the above models is available in Forni and Lippi [2001], Forni et al. [2004] and

Chudik and Pesaran [2011]. We work with the more general high-dimensional time series model, namely the infinite dimensional moving average process of order ∞ (MA(∞)). All the former high-dimensional time series models, under some causality conditions, can be expressed as MA(∞) processes. In Section 3.2, we briefly describe all the above models and their causality conditions.

In this chapter, we are interested in estimating the population autocovariance matrices $\{\Gamma_{u,p}\}$ for the infinite dimensional MA(∞) process. As we have seen in Chapter 2, in high-dimensional setting, the dimension $p = p(n) \to \infty$ as the sample size $n \to \infty$. Therefore, the size of the population autocovariance matrices $\{\Gamma_{u,p}\}$ increases as $p = p(n) \to \infty$ and hence the number of unknown parameters (entries in $\{\Gamma_{u,p}\}$) increases. Consequently, just like the sample variance-covariance matrix in Chapter 2, the sample autocovariance matrices fail to consistently estimate the population autocovariance matrices.

The existing works on high-dimensional time series have not dealt with the estimation of population autocovariance matrices. From the experience of Chapter 2, to get consistent estimators of $\{\Gamma_{u,p}\}$ we need two things – suitable restrictions on the parameter space and appropriate modifications such as banding or tapering on sample autocovariance matrices. In Section 3.3, Theorems 3.3.1 and 3.3.2 provide some restrictions on parameters under which the appropriately banded and tapered version of sample autocovariance of order 0 is consistent for the population autocovariance of order 0. These restrictions are directly borrowed from the development of Chapter 2. But these restrictions on parameters are cumbersome and very difficult to check. Moreover, Theorems 3.3.1 and 3.3.2 are silent about autocovariance matrices of order $u > 0$.

In Section 3.4, we identify an appropriate parameter space for estimating population autocovariance matrices of any order for the infinite dimensional MA(∞) process. In Section 3.5, Theorem 3.5.1 establishes that the banded and tapered sample autocovariance matrices are consistent for the population autocovariance matrices under the Gaussian assumption on the driving process. We also derive the convergence rate of these estimators.

The infinite dimensional finite order moving average and infinite dimensional autoregressive processes are special cases of the infinite dimensional vector linear process. Theorems 3.4.5 and 3.4.6 provide appropriate parameter spaces for these two processes so that their population autocovariance matrices can be consistently estimated. Under these parameter spaces and the Gaussian assumption on the driving process, Theorems 3.5.5 and 3.5.6 state that the banded and tapered sample autocovariance matrices are also consistent for the population autocovariance matrices. Moreover, the same convergence rates as in Theorem 3.5.1 hold.

Using this consistency, in Section 3.5.1, we show how to obtain consistent estimators for the parameter matrices of an infinite dimensional autogressive process (see Theorem 3.5.8).

In Section 3.5.2, we relax the Gaussian assumption on the driving process. We replace the Gaussian assumption by an appropriate condition on the moment generating function. We show that under this condition, Theorems 3.5.1, 3.5.5, 3.5.6 and 3.5.8 continue to hold.

To support our results, some simulations are given in Section 3.6.

The main material of this chapter is taken from Bhattacharjee and Bose [2014b].

3.2 Models and examples

In this section we shall discuss some high-dimensional time series models. As mentioned in Section 3.1, a very general high dimensional linear time series model is the infinite dimensional moving average process of order ∞ (MA(∞)). This process is given by

$$X_{t,p}^{(n)} = \sum_{j=0}^{\infty} \psi_{j,p}\varepsilon_{t-j}, \quad t, n \geq 1 \ \ \text{(almost surely)}, \tag{3.2}$$

where $\{X_{t,p}^{(n)}\}$ and $\{\varepsilon_t\}$ are both p-dimensional random vectors, $\{\varepsilon_t\}$ are i.i.d. with mean 0 and $p \times p$ variance-covariance matrix Σ_p, $\{\psi_{j,p}\}$ are all $p \times p$ matrices and are called coefficient matrices. We have appropriate conditions on $\{\psi_{j,p}\}$ so that the above sum is meaningful. The dimension $p = p(n) \to \infty$ as the sample size $n \to \infty$. It may be noted that the dimension of $X_{t,p}$ is not infinite. However, since $p \to \infty$, it has become customary to refer to such models as 'infinite dimensional'. This is a weak stationary time series and the population autocovariance matrix of order u is given by

$$\Gamma_{u,p} = \sum_{j=0}^{\infty} \psi_{j,p} \Sigma_p \psi_{j+u,p}^*, \ \forall u \geq 0. \tag{3.3}$$

Clearly in high-dimensional set up, the size of the coefficient matrices $\{\psi_{j,p}\}$ increases as p increases and consequently as we move from the n-th stage to the $(n+1)$-th stage, all the components of $X_{t,p}^{(n)}$ get changed. Hence, in the high dimensional set up, we have the following triangular sequence:

$$\begin{array}{l} X_{1,p(1)}^{(1)} \\ X_{1,p(2)}^{(2)}, X_{2,p(2)}^{(2)} \\ X_{1,p(3)}^{(3)}, X_{2,p(3)}^{(3)}, X_{3,p(3)}^{(3)} \\ \vdots \\ X_{1,p(n)}^{(n)}, X_{2,p(n)}^{(n)}, X_{3,p(n)}^{(n)}, \ldots, X_{n,p(n)}^{(n)} \\ \vdots \end{array} \tag{3.4}$$

and the sample at the n-th stage is the n-th row of this triangular sequence. For convenience, we shall often write $X_{t,p}$ for $\{X_{t,p}^{(n)}\}$.

In all the examples below, we have $p = p(n) \to \infty$ as $n \to \infty$.

Example 3.2.1. *The infinite dimensional IID process is given by*

$$X_{t,p} = \varepsilon_t, \ \forall t \tag{3.5}$$

where $\{\varepsilon_t\}$ *is a set of i.i.d.* p*-dimensional random vectors with mean* 0 *and* $p \times p$ *variance-covariance matrix* Σ_p*.*

(3.5) is a weak stationary time series process with

$$\Gamma_{u,p} = \begin{cases} \Sigma_p, \ \text{if } u = 0 \\ 0, \ \text{otherwise.} \end{cases} \tag{3.6}$$

Note that if $\psi_{0,p} = I_p$ and $\psi_{j,p} = 0, \ \forall j \geq 1$, then the model (3.2) reduces to (3.5).

Example 3.2.2. *The infinite dimensional moving average process of order* r *(*$MA(r)$*) is given by*

$$X_{t,p} = \sum_{i=0}^{r} M_{i,p}\varepsilon_{t-i}, \ t \geq 1 \tag{3.7}$$

where $\{\varepsilon_t\}$ *is as in Example 3.2.1,* $M_{i,p}, i = 0, 1, 2, \ldots, r$ *are square matrices of order* p *and are called parameter matrices,* $M_{0,p} = I_p$ *and* I_p *is as in (2.8) and* r *is a non-negative integer.*

It is easy to see that MA(r) is a weak stationary model and

$$\Gamma_{u,p} = \begin{cases} \sum_{i=0}^{r-u} M_{i,p}\Sigma_p M^*_{i+u,p}, \ \text{for } 0 \leq u \leq r \\ 0, \ \text{otherwise.} \end{cases} \tag{3.8}$$

For $r = 0$, (3.7) is same as the IID process given in Example 3.2.1. If $\psi_{j,p} = M_{j,p}I(0 \leq j \leq r)$, then (3.2) reduces to (3.7).

Example 3.2.3. *The infinite dimensional vector autoregressive process of order*

r (VAR(r)) *is given by*

$$X_{t,p} = \sum_{i=1}^{r} A_{i,p} X_{t-i,p} + \varepsilon_t, \ \ t \geq 1 \tag{3.9}$$

where $\{\varepsilon_t\}$ *is as in Example 3.2.1, the* $p \times p$ *matrices* $A_{i,p}, i = 1, 2, \ldots, r$, *are called the parameter matrices and* r *is a non-negative integer.*

Let I_p be as in (2.8) and

$$\mathbb{C} = \text{ set of all complex numbers.} \tag{3.10}$$

Then we have the following Theorem.

Theorem 3.2.1. *(3.9) is a weak stationary process if for some* $\epsilon > 0$, $\{A_{i,p}\}$ *satisfy*

$$det(I_p - A_{1,p}z - A_{2,p}z^2 - \cdots - A_{r,p}z^r) \neq 0, \ \forall z \in \mathbb{C} \ \textit{such that} \ |z| \leq 1 + \epsilon. \tag{3.11}$$

Under (3.11), (3.9) has the representation,

$$X_{t,p} = \sum_{j=0}^{\infty} \phi_{j,p} \varepsilon_{t-j}, \ \ t \geq 1 \ \ \textit{(almost surely)}, \tag{3.12}$$

where

$$\phi_{0,p} = I_p \ \text{ and } \ \phi_{j,p} = \sum_{i=1}^{j} A_{i,p} \phi_{j-i,p}, j \geq 1. \tag{3.13}$$

Proof. Let $\{z_{i,p} : 1 \leq i \leq r\}$ be r roots of the equation $det(I_{p(n)} - A_{1,p(n)}z - A_{2,p(n)}z^2 - \cdots - A_{r,p(n)}z^r) = 0$, $z \in \mathbb{C}$. Let $\alpha_p = \min\{|z_{i,p}| : 1 \leq i \leq r\}$. By Theorem 11.3.1 in Brockwell and Davis [2009], for each fixed p, (3.9) can be represented as (3.12) with the coefficient matrices (3.13), if

$$\alpha_p > 1. \tag{3.14}$$

Note that (3.11) implies (3.14) for all $p \geq 1$. This completes the proof of Theorem 3.2.1. $\square$

(3.11) is called the *causality* condition for the model (3.9). Under (3.11), the population autocovariance matrix of order u for the model (3.9) is given by

$$\Gamma_{u,p} = \sum_{j=0}^{\infty} \phi_{j,p} \Sigma_p \phi^*_{j+u,p}, \ \forall u \geq 0. \tag{3.15}$$

Thus the infinite dimensional MA(∞) process, defined in (3.2), is a very general time series model.

As we have mentioned in Section 3.1, in this chapter our main goal is to estimate the population autocovariance matrices $\{\Gamma_{u,p}\}$ for the model (3.2). Let $\{X_{t,p} : t = 1, 2, \ldots, n\}$ be a sample of size n from this model. A method of moment estimator of $\Gamma_{u,p}$ is given by the $p \times p$ matrix

$$\hat{\Gamma}_{u,p,n} = \frac{1}{n} \sum_{t=1}^{n-u} X_{t,p} X^*_{t+u,p}. \tag{3.16}$$

$\hat{\Gamma}_{u,p,n}$ is called the *sample autocovariance matrix* of order u based on n observations. Recall the definition of consistency given in (2.5). In finite dimensional set up i.e. when p is fixed, $\hat{\Gamma}_{u,p,n}$ is a consistent estimator of $\Gamma_{u,p}$ as $n \to \infty$.

Now suppose the dimension $p = p(n) \to \infty$ as the sample size $n \to \infty$. Then the size of the population autocovariance matrices $\{\Gamma_{u,p}\}$ increases as p increases and the sample autocovariance matrices $\{\hat{\Gamma}_{u,p,n}\}$ are no more consistent. For evidence see simulation results in Example 2.2.1, which shows $\hat{\Gamma}_{0,p,n}$ is not consistent for $\Gamma_{0,p}$.

3.3 Estimation of $\Gamma_{0,p}$ using results of Chapter 2

This section deals with $\Gamma_{0,p}$ for some very specific cases. These models obey the conditions given in Chapter 2. As in the population variance-covariance matrix estimation in Chapter 2, to get a consistent estimator of $\{\Gamma_{0,p}\}$, we need the following:

(a) suitable restrictions on the coefficient matrices $\{\psi_{j,p}\}$ and on the variance-covariance matrix of $\{\varepsilon_t\}$ i.e. on Σ_p, and

(b) appropriate modification such as banding or tapering of $\{\hat{\Gamma}_{0,p,n}\}$.

Note that $\Gamma_{0,p}$ is the variance-covariance matrix of $\{X_{t,p}\}$. This section provides some restrictions on $\{\psi_{j,p}\}$ and Σ_p so that the appropriately banded and tapered $\hat{\Gamma}_{0,p,n}$ become consistent for $\Gamma_{0,p}$. These restrictions are directly borrowed from the developments of Chapter 2.

Recall the $\infty \times \infty$ extension Σ_∞ of $\{\Sigma_p\}$ in the sense of (2.15). Also recall the class of dispersion matrices $\mathcal{U}(\epsilon, \alpha, C)$ in (2.22). Recall the banded and tapered version of a matrix in Section 2.2.1. Let the tapering function $g : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+ \cup \{0\}$ be continuous, non-increasing such that $g(0) = 1$ and $\lim_{x\to\infty} g(x) = 0$. Let $\Delta_{g,\tau_{n,\alpha}} = \sum_{j=0}^{n-1} g\left(\frac{j}{\tau_{n,\alpha}}\right)$. Then we have the following theorem on infinite dimensional IID or MA(0) process. This is a restatement of Theorems 2.2.2 and 2.2.6.

Theorem 3.3.1. *Consider the model (3.5). Suppose $\varepsilon_t \sim \mathcal{N}_p(0, \Sigma_p)$, $\forall t$ and $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C)$ for some $\epsilon, \alpha, C > 0$. Then*

(a) for $k_{n,\alpha} \asymp (n^{-1} \log p)^{-\frac{1}{2(1+\alpha)}}$, we have

$$||B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p,n}) - \Gamma_{0,p}||_2 = O_P(k_{n,\alpha}^{-\alpha}), \tag{3.17}$$

(b) and for $\Delta_{g,\tau_{n,\alpha}} \asymp (n^{-1} \log p)^{-\frac{1}{2(1+\alpha)}}$, we have

$$||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{0,p,n}) - \Gamma_{0,p}||_2 = O_P(\Delta_{g,\tau_{n,\alpha}}^{-\alpha}). \tag{3.18}$$

Recall the class of dispersion matrices $\mathcal{V}$ in (2.65). Additionally assume $\int_0^\infty g(x) < \infty$ and $1 - g(x) = O\left(x^\gamma\right)$ for some $\nu \geq 1$ in some neighborhood of zero. Then we have the following theorem for the infinite dimensional MA(∞) process.

Theorem 3.3.2. *Consider the model (3.2). Suppose* $\varepsilon_t \sim \mathcal{N}_p(0,\Sigma_p),\ \forall t$, *and for some* $\alpha, \epsilon, C > 0$

$$\sum_{j=0}^{\infty} \psi_{j,p}\Sigma_p\psi_{j,p}^* \in \mathcal{U}(\epsilon,\alpha,C)\cap\mathcal{V} \tag{3.19}$$

and

$$\max_{a_n\le u}\max_{v,w} \frac{|\left(\sum_{j=0}^{\infty}\psi_{j,p}\Sigma_p\psi_{j+u,p}^*\right)(v,w)|}{|\left(\sum_{j=0}^{\infty}\psi_{j,p}\Sigma_p\psi_{j,p}^*\right)(v,w)|} = O(n^{-2}a_n), \tag{3.20}$$

for some a_n *such that* $a_n\sqrt{n^{-1}\log p}\to 0$ *and* $a_n^{-1}\sqrt{n\log p}\to\infty$ *as* $n\to\infty$. *Then*

(a) for $k_{n,\alpha} \asymp (a_n\sqrt{n^{-1}\log p})^{-\frac{1}{1+\alpha}}$, *we have*

$$||B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p,n}) - \Gamma_{0,p}||_2 = O_P(k_{n,\alpha}^{-\alpha}), \tag{3.21}$$

(b) and for $\tau_{n,\alpha} \asymp (a_n\sqrt{n^{-1}\log p})^{-\frac{1}{2(1+\alpha)}[\frac{\gamma}{1+\alpha}+1]}$, *we have*

$$||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{0,p,n}) - \Gamma_{0,p}||_2 = O_P\left[\left(a_n\sqrt{n^{-1}\log p}\right)^{\frac{\gamma\alpha}{(1+\alpha)(1+\gamma)}}\right]. \tag{3.22}$$

Proof. To prove the above theorem, we use Theorems 2.4.11 and 2.4.13 (c). There our approach was to separate out the cross covariance structure ∇_{np}, an $np\times np$ matrix consisting of n^2-many $p\times p$ matrices $\{\Lambda_{ij} : 1\le i,j,\le n\}$. Recall the notation (2.6). Note that, by (2.64) and weak stationarity of (3.2), for all $1\le i,j\le n,\ 1\le v,w\le p$ we have

$$\Lambda_{ij}(v,w) = \Lambda_{|i-j|}(v,w) = \frac{E(X_{i,p}X_{j,p}^*)(v,w)}{E(X_{i,p}X_{i,p}^*)(v,w)} = \frac{\Gamma_{i-j,p}(v,w)}{\Gamma_{0,p}(v,w)}, \tag{3.23}$$

provided $\Gamma_{0,p}(v,w)\neq 0,\ \forall v,w$. Now, by (3.3), for all $1\le v,w\le p$ we have

$$\Lambda_u(v,w) = \frac{\left(\sum_{j=0}^{\infty}\psi_{j,p}\Sigma_p\psi_{j+u,p}^*\right)(v,w)}{\left(\sum_{j=0}^{\infty}\psi_{j,p}\Sigma_p\psi_{j,p}^*\right)(v,w)},\ \forall u\ge 0 \tag{3.24}$$

provided $(\sum_{j=0}^{\infty}\psi_{j,p}\Sigma_p\psi_{j,p}^*)(v,w) \neq 0, \ \forall v,w$. Then Theorem 3.3.2 follows from Theorems 2.4.11 and 2.4.13 (c) provided (3.19) and (3.20) hold. $\square$

It may be observed that conditions (3.19) and (3.20) are cumbersome and difficult to check in general unless there is some additional structure in the model. Example 3.3.1 provides such a case.

Example 3.3.1. *Consider the model (3.2), with $\psi_{j,p} = \theta^j A_p$ for any $0 < \theta < 1$ and $p \times p$ matrix A_p such that all elements of $A_p\Sigma_pA_p^*$ are non-zero. Then for all $1 \leq v,w \leq p$ and $u \geq 1$, we have*

$$\begin{aligned}\Lambda_u(v,w) &= \frac{|\left(\sum_{j=0}^{\infty}\psi_{j,p}\Sigma_p\psi_{j+u,p}^*\right)(v,w)|}{|\left(\sum_{j=0}^{\infty}\psi_{j,p}\Sigma_p\psi_{j,p}^*\right)(v,w)|} \\ &= \frac{\theta^u|\left(A_p\Sigma_pA_p^*\right)(v,w)|\left(\sum_{j=0}^{\infty}\theta^{2j}\right)}{|\left(A_p\Sigma_pA_p^*\right)(v,w)|\left(\sum_{j=0}^{\infty}\theta^{2j}\right)} = \theta^u.\end{aligned}$$

Therefore, $\sup_{u\geq a_n}||\Lambda_u||_\infty = \theta^{a_n}$ *and (3.20) holds.*

Moreover, it is not clear what conditions on the parameter matrices are needed in the IVAR model so that (3.19) and (3.20) can be satisfied. Therefore, we need some directly verifiable conditions on the coefficient matrices.

Finally, the *cross covariance structure* model used in Theorem 3.3.2 is meaningful if and only if all elements of the matrix $\Gamma_{0,p}$ are non-zero (see (3.19)). There are of course many processes where this may not be the case. Here are two simple examples.

Example 3.3.2. $\psi_{j,p} = \theta^j I_p$ *for all j with at least one zero element in Σ_p.*

Example 3.3.3. *Consider $\Sigma_p = I_p$ and $\psi_{j,p}$'s are such that $\psi_{j,p}\psi_{j,p}^*$'s are diagonal. Any asymmetric Toeplitz matrix made of $\{t_i\}_{i=-\infty}^{\infty}$ with $t_i = 0$ for all i except one, are examples of such $\psi_{j,p}$'s.*

Another issue is that, in a time series model, the interest is not only on the marginal variance-covariance matrix $\Gamma_{0,p}$ but also on the entire sequence of autocovariance matrices $\{\Gamma_{u,p}\}$. From Theorems 3.3.1 and 3.3.2, it is not clear how to estimate the *cross covariances* in general. Moreover, the assumptions of Theorem 3.3.2 do not offer any control over the first few cross covariances. Therefore, under those assumptions, it is not possible to estimate them in the high dimensional set up. This leads to the need for identifying appropriate parameter spaces so that all the autocovariance matrices are consistently estimable. In the following section we shall provide such a parameter space.

3.4 Parameter spaces

There are two kinds of parameters in the model (3.2). First, the variance-covariance matrix of the driving process $\{\varepsilon_t\}$ i.e. Σ_p and second, the set of coefficient matrices $\{\psi_{j,p}\}$. Recall the class of matrices having polynomially decaying corner, denoted by $\mathcal{X}(\alpha, C)$ for some $\alpha, C > 0$, in (2.20). Also recall $||\cdot||_\infty$ in (2.29). From the experience of Chapter 2 (see (2.32), (2.45) and (2.46)), we cannot expect the banded or tapered version of the sample autocovariance matrices to be consistent for $\{\Gamma_{u,p}\}$, unless

(i) $\sup_p ||\Gamma_{u,p}||_\infty < \infty,\ \forall u \geq 0$ and

(ii) $\Gamma_{u,p} \in \mathcal{X}(\alpha, C)$ for some $\alpha, C > 0$.

Below we discuss appropriate restrictions on both type of parameters so that $\{\Gamma_{u,p}\}$ satisfy (i) and (ii) above and as a consequence, consistent estimators of $\{\Gamma_{u,p}\}$ can be achieved.

Restrictions on Σ_p. Recall the class of variance-covariance matrices $\mathcal{U}(\epsilon, \alpha, C)$ in (2.22). Recall the notation (2.15). Note that Theorem 3.3.1 provided consistent estimator of $\Gamma_{0,p}$ when the $\infty \times \infty$ extension Σ_∞ of $\{\Sigma_p\}$ is in $\mathcal{U}(\epsilon, \alpha, C)$ for some $\epsilon, \alpha, C > 0$. Since infinite dimensional IID process is a particular case of the model

(3.2), we continue to assume that

$$\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C), \text{ for some } \epsilon, \alpha, C > 0. \tag{3.25}$$

Restrictions on $\{\psi_{j,p}\}$. To define an appropriate parameter space for the sequence of matrices $\{\psi_{j,p}\}_{j=0}^{\infty}$, we assume the following criteria.

For each $j \geq 0$, consider the $\infty \times \infty$ extension of the sequence of matrices $\{\psi_{j,p(n)}\}_{n\geq 1}$ as $\psi_{j,\infty}$ (in the sense (2.15)). Recall the $||\cdot||_{(1,1)}$ norm and the corner measure $T(\cdot,\cdot)$ respectively from (2.26) and (2.18).

(i) **Time lag criterion:** We ensure that the dependence decreases appropriately with the lag. For this purpose, define

$$\max(||\psi_{j,\infty}||_{(1,1)}, ||\psi_{j,\infty}^*||_{(1,1)}) = r_j, \ \forall j \geq 0. \tag{3.26}$$

We define the following class $\Im(\beta, \lambda)$ of sequence of matrices $\{\psi_{j,\infty}\}_{j=0}^{\infty}$ for some $0 < \beta < 1$ and $\lambda \geq 0$.

$$\Im(\beta, \lambda) = \left\{ \{\psi_{j,\infty}\}_{j=0}^{\infty} : \sum_{j=0}^{\infty} r_j^{\beta} < \infty, \sum_{j=0}^{\infty} r_j^{2(1-\beta)} j^{\lambda} < \infty \right\}. \tag{3.27}$$

Note that the summability above implies that the decay rate of r_j cannot be slower than a polynomial rate.

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

$0 < \eta < 1$ as

$$\begin{aligned}\mathcal{G}(C,\alpha,\eta,\nu) &= \Big\{\{\psi_{j,\infty}\} : T\Big(\psi_{j,\infty}, t\sum_{u=0}^{j}\eta^u\Big) < Ct^{-\alpha}r_j j^{\nu}\sum_{u=0}^{j}\eta^{-u\alpha}, \\ &\qquad \sum_{j=0}^{\infty}\frac{r_j r_{j+u} j^{\nu}}{\eta^{\alpha j}} < \infty \;\; \forall u \geq 0\Big\}.\end{aligned}$$

In this chapter we assume $\{\psi_{j,\infty}\} \in \Im(\beta,\lambda) \cap \mathcal{G}(C,\alpha,\eta,\nu)$ for some $\lambda \geq 0$, $C,\alpha,\nu > 0$ and $0 < \beta,\eta < 1$. Recall (i) and (ii) described at the beginning of this section. The next two theorems state that this assumption together with $\Sigma_\infty \in \mathcal{U}(\epsilon,\alpha,C) = \mathcal{X}(\alpha,C) \cap \mathcal{W}(\epsilon)$ for some $\epsilon,\alpha,C > 0$, imply (i) and (ii).

Theorem 3.4.1. *Consider the model (3.2). Suppose $\Sigma_\infty \in \mathcal{W}(\epsilon)$ and $\{\psi_{j,\infty}\} \in \Im(\beta,\lambda)$ for some $\epsilon > 0$, $\lambda \geq 0$ and $0 < \beta < 1$. Then*

$$\sup_p ||\Gamma_{u,p}||_\infty < \infty, \;\; \forall u \geq 0.$$

To prove the above theorem we need the following lemma.

Lemma 3.4.2. *For any two square matrices A and B of same order,*

$$||AB||_\infty \leq \min\{||A||_\infty ||B||_{(1,1)}, ||B||_\infty ||A^*||_{(1,1)}\}.$$

Proof. Recall the notation (2.6). Then

$$\begin{aligned}||AB||_\infty &= \max_{i,j}|AB(i,j)| = \max_{i,j}\sum_k |A(i,k)B(k,j)| \\ &\leq \max_{i,k}|A(i,k)|\max_j\sum_k |B_{k,j}| = ||A||_\infty ||B||_{(1,1)}.\end{aligned}$$

Similarly, one can show that $||AB||_\infty \leq ||B||_\infty ||A^*||_{(1,1)}$. This completes the proof.

□

Proof of Theorem 3.4.1. Note that as $\Sigma_\infty = ((\sigma_{ij})) \in \mathcal{W}(\epsilon)$, we have $|\sigma_{ij}| \leq$

$\sqrt{\sigma_{ii}\sigma_{jj}} \leq \lambda_{\max}(\Sigma_p) < \epsilon^{-1}, \quad \forall i, j$. Therefore,

$$\sup_p ||\Sigma_p||_\infty = ||\Sigma_\infty||_\infty < \epsilon^{-1}. \tag{3.28}$$

Also for the model (3.2), $\Gamma_{u,p} = \sum_{j=0}^{\infty} \psi_{j,p}\Sigma_p\psi^*_{j+u,p}, \ \forall u \geq 0$.

Therefore, by (3.28) and the repeated use of Lemma 3.4.2, we have

$$\begin{aligned}
\sup_p ||\Gamma_{u,p}||_\infty &\leq \sup_p \sum_{j=0}^{\infty} ||\psi_{j,p}\Sigma_p\psi^*_{j+u,p}||_\infty \leq \sup_p \sum_{j=0}^{\infty} ||\psi^*_{j,p}||_{(1,1)}||\Sigma_p\psi^*_{j+u,p}||_\infty \\
&\leq \sup_p \sum_{j=0}^{\infty} ||\psi^*_{j,p}||_{(1,1)}||\Sigma_p||_\infty||\psi^*_{j+u,p}||_{(1,1)} \\
&\leq \epsilon^{-1} \sum_{j=0}^{\infty} r_j r_{j+u} < \infty,
\end{aligned}$$

as $\{\psi_{j,\infty}\} \in \Im(\beta, \lambda)$ for some $\lambda \geq 0$ and $0 < \beta < 1$. □

Note that Theorem 3.4.1 implies (i). The next theorem will imply (ii).

Theorem 3.4.3. *Consider the model (3.2). If $\Sigma_\infty \in \mathcal{X}(\alpha, C)$ and $\{\psi_{j,\infty}\}_{j=0}^{\infty} \in \mathcal{G}(C, \alpha, \eta, \nu)$ for some $C, \alpha, \nu > 0$ and $0 < \eta < 1$, then for all $t > 0$ and some $c' > 0$,*

$$T(\Gamma_{u,p(n)}, t) < c't^{-\alpha}||\Sigma_{p(n)}||_{(1,1)}, \ \forall u \geq 0, \ n \geq 1.$$

Moreover, if $||\Sigma_\infty||_{(1,1)} < \infty$ then $\Gamma_{u,p(n)} \in \mathcal{X}(\alpha, c')$, for all $u \geq 0, \ n \geq 1$.

To prove the above theorem, we need the following lemma on the corner measure $T(\cdot, \cdot)$ and $|| \cdot ||_{(1,1)}$ norm of a square matrix.

Lemma 3.4.4. *Let A and B be two $r \times r$ matrices. Then,*

(i) $T(A, k) \leq T(A, k'), \ \forall \ 0 < k' < k < \infty$.

(ii) $\|AB\|_{(1,1)} \leq \|A\|_{(1,1)}\|B\|_{(1,1)}$,

(iii) $T(AB, (\alpha + \beta)t) \leq \|A\|_{(1,1)} T(B, \alpha t) + \|B\|_{(1,1)} T(A, \beta t)$, *for any* $\alpha, \beta, t > 0$.

Proof. Proofs of (i) and (ii) are trivial. To prove (iii), consider the following steps.

$$\begin{aligned}
& T\left(AB, (\alpha+\beta)\, t\right) \\
\leq\ & \max_k \sum_{j:|j-k|>(\alpha+\beta)t} \left(\sum_{l=1}^{\infty} |a_{jl} b_{lk}| \right) \\
\leq\ & \max_k \sum_{j:|j-k|>(\alpha+\beta)t} \left(\sum_{l:|l-k|\leq \alpha t} |a_{jl} b_{lk}| \right) + \max_k \sum_{j:|j-k|>(\alpha+\beta)t} \left(\sum_{l:|l-k|>\alpha t} |a_{jl} b_{lk}| \right) \\
\leq\ & \left(\max_k \sum_{j:|j-l|>\beta t,\ l:|l-k|\leq \alpha t} |a_{jl} b_{lk}| \right) + \left(\max_k \sum_{j:|j-k|>(\alpha+\beta)t,\ l:|l-k|>\alpha t} |a_{jl} b_{lk}| \right) \\
\leq\ & \left(\max_l \sum_{j:|j-l|>\beta t} |a_{jl}| \right) \left(\max_k \sum_{l=1}^{\infty} |b_{lk}| \right) + \left(\max_k \sum_{l:|l-k|>\alpha t} |b_{lk}| \right) \left(\max_l \sum_{j=1}^{\infty} |a_{jl}| \right) \\
\leq\ & \|B\|_{(1,1)} T\left(A, \beta t\right) + \|A\|_{(1,1)} T(B, \alpha t).
\end{aligned}$$

This completes the proof. $\square$

Now we are ready to prove Theorem 3.4.3.

Proof of Theorem 3.4.3. Let $\delta_p = ||\Sigma_p||_{(1,1)}$. From the properties of $\mathcal{X}(\alpha, C)$ and $\mathcal{G}(C, \alpha, \eta, \nu)$ and, by Lemma 3.4.4 (iii), for some $C_1 > 0$,

$$\begin{aligned}
T(\psi_{j,p}\Sigma_p, (1+\eta+\cdots+\eta^{j+1})t) &\leq \delta_p C t^{-\alpha}(1+\eta^{-\alpha}+\cdots+\eta^{-ja}) r_j j^v + r_j C t^{-\alpha} \eta^{-(j+1)\alpha} \\
&\leq C_1 t^{-\alpha} \delta_p (1+\eta^{-\alpha}+\cdots+\eta^{-(j+1)\alpha}) r_j j^v.
\end{aligned}$$

Again, by Lemma 3.4.4 (ii), (iii) and for some $C_2 > 0$,

$$\begin{aligned}
T(\psi_{j,p}\Sigma_p\psi^*_{j+u,p}, 2(1+\eta+\cdots+\eta^{j+1})t) &\leq r_{j+u} C_1 t^{-\alpha} \delta_p (1+\eta^{-\alpha}+\cdots+\eta^{-(j+1)\alpha}) r_j j^v \\
&\quad + \delta_p r_j C t^{-\alpha} (1+\eta^{-\alpha}+\cdots+\eta^{-(j+1)\alpha}) r_{j+u} j^\nu \\
&\leq C_2 t^{-\alpha} \delta_p (1+\eta^{-\alpha}+\ldots+\eta^{-(j+1)\alpha}) r_j r_{j+u} j^\nu.
\end{aligned}$$

Hence, as $\{\psi_{j,\infty}\} \in \mathcal{G}(C, \alpha, \eta, \nu)$ and by Lemma 3.4.4 (i), for some $C_3, C_4 > 0$, we have

$$T(\Gamma_u, \frac{2}{1-\eta}t) = T(\sum_{j=0}^{\infty} \psi_{j,p}\Sigma_p\psi_{j+u,p}^*, \frac{2}{1-\eta}t) < C_3 t^{-\alpha}\delta_p \sum_{j=0}^{\infty} \frac{r_j r_{j+u}}{\eta^{\alpha j}} j^{\nu} < C_4 t^{-\alpha}\delta_p.$$

Hence the proof of Theorem 3.4.3 is complete. □

Thus, by Theorems 3.4.1 and 3.4.3, it is clear that we need to assume $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C)$ and $\{\psi_{j,\infty}\} \in \Im(\beta, \lambda) \cap \mathcal{G}(C, \alpha, \eta, \nu)$ for some $\lambda \geq 0$, $C, \alpha, \epsilon, \nu > 0$ and $0 < \beta, \eta < 1$, to guarantee $\{\Gamma_{u,p}\}$ have polynomially decaying corners and $\sup_p ||\Gamma_{u,p}||_\infty < \infty$, $\forall u \geq 0$. As mentioned at the beginning of this section, these two conditions will be crucially used when we deal with the banded and tapered sample autocovariance matrices in the next section.

Recall that, infinite dimensional MA(r) processes and IVAR(r) processes defined respectively in Examples 3.2.2 and 3.2.3, are all particular cases of the model (3.2). Therefore, the obvious curiosity is under what condition on $\{M_{i,p}\}$ in Examples 3.2.2 and $\{A_{i,p}\}$ in Examples 3.2.3, would the corresponding coefficient matrices be in $\Im(\beta, \lambda) \cap \mathcal{G}(C, \alpha, \eta, \nu)$ for some $\lambda \geq 0$, $C, \alpha, \nu > 0$ and $0 < \beta, \eta < 1$? The following discussion answers this question.

Parameter space for infinite dimensional MA(r) process. Consider the model (3.7) and its $p \times p$ parameter matrices $\{M_{i,p} : 1 \leq i \leq r\}$ with $M_{0,p} = I_p$. For each $0 \leq i \leq r$, let $M_{i,\infty}$ be the $\infty \times \infty$ extension of the sequence of matrices $\{M_{i,p(n)}\}_{n \geq 1}$ in the sense (2.15). Let us define $M_{i,\infty} = 0$, $\forall i > r$. The following theorem provides a simplified condition on $\{M_{i,\infty} : 0 \leq i \leq r\}$ so that $\{M_{i,\infty} : i \geq 0\} \in \Im(\beta, \lambda) \cap \mathcal{G}(C, \alpha, \eta, \nu)$ for some $\lambda \geq 0$, $C, \alpha, \nu > 0$ and $0 < \beta, \eta < 1$. Recall the class of matrices having polynomially decaying corners, denoted by $\mathcal{X}(\alpha, C)$ for some $\alpha, C > 0$, in (2.20). The following result has appeared in Bhattacharjee and Bose [2014b].

Theorem 3.4.5. *Suppose* $0 < ||M_{i,\infty}||_{(1,1)} < \infty$ *and* $M_{i,\infty} \in \mathcal{X}(\alpha, C)$ *for some*

$\alpha, C > 0$ and for all $1 \leq i \leq r$. Then

$$\{M_{i,\infty} : i \geq 0\} \in \Big(\bigcap_{\substack{0<\beta<1 \\ \lambda \geq 0}} \Im(\beta, \lambda) \Big) \bigcap \Big(\bigcap_{\substack{0<\eta<1 \\ \nu>0}} \mathcal{G}(Cm^{-1}, \alpha, \eta, \nu) \Big),$$

where $m = \min\{||M_{i,\infty}||_{(1,1)} : 1 \leq i \leq r\}$.

Proof. Note that in the model (3.7), r_j as in (3.26) is given by

$$r_j = \begin{cases} ||I_\infty||_{(1,1)} = 1, & \text{if } j = 0 \\ ||\psi_{j,\infty}||_{(1,1)} = ||M_{j,\infty}||_{(1,1)}, & \text{if } 1 \leq j \leq r \\ 0, & \text{if } j > r. \end{cases}$$

Therefore, as there are only finitely many non-zero r_j's, all the summability conditions on $\{r_j\}$ in $\Im(\beta, \lambda)$ and $\mathcal{G}(C, \alpha, \eta, \nu)$ are satisfied for all $\lambda \geq 0$, $\alpha, \nu > 0$ and $0 < \beta, \eta < 1$.

Next, for all $j \geq 0$, $\nu > 0$ and $0 < \eta < 1$, we have

$$\begin{aligned} T\Big(M_{j,\infty}, t\sum_{u=0}^{j} \eta^u\Big) &< Ct^{-\alpha}\Big(\sum_{u=0}^{j} \eta^u\Big)^{-\alpha} < (Cm^{-1})r_j t^{-\alpha} j^{\nu} j^{-\alpha}\Big(j^{-1}\sum_{u=0}^{j} \eta^u\Big)^{-\alpha} \\ &< (Cm^{-1})r_j t^{-\alpha} j^{\nu} j^{-\alpha-1}\sum_{u=0}^{j} \eta^{-u\alpha} < (Cm^{-1})r_j t^{-\alpha} j^{\nu}\sum_{u=0}^{j} \eta^{-u\alpha}. \end{aligned}$$

This completes the proof. □

Parameter space for IVAR(r) process. Consider the model (3.9) in Example 3.2.3. Recall that, under (3.11), the model (3.9) can be represented as a linear process of the form (3.12). For each $i \geq 0$, let $\phi_{i,\infty}$ be the $\infty \times \infty$ extension of the sequence of matrices $\{\phi_{i,p(n)}\}_{n \geq 1}$ in the sense (2.15). Theorem 3.4.6 provides direct conditions on the parameter matrices $\{A_{i,p}\}$ so that the corresponding coefficient matrices $\{\phi_{i,\infty}\} \in \Im(\beta, \lambda) \cap \mathcal{G}(C, \alpha, \eta, \nu)$ for some $\lambda \geq 0$, $C, \alpha, \nu > 0$ and $0 <$

$\beta, \eta < 1$. To state Theorem 3.4.6, we need some preparation.

Let

$$||A_{i,p}||_{(1,1)} = \theta_{i,n} \quad \text{and} \quad ||A^*_{i,p}||_{(1,1)} = \theta'_{i,n}, 1 \leq i \leq r. \tag{3.29}$$

Also let $\{\alpha_{i,n} : i = 1, 2, \ldots, r\}$ and $\{\alpha'_{i,n} : i = 1, 2, \ldots, r\}$ respectively be the roots of the following polynomials.

$$\begin{aligned} 1 - \theta_{1,n}z - \theta_{2,n}z^2 \ldots \theta_{r,n}z^r &= 0, \\ 1 - \theta'_{1,n}z - \theta'_{2,n}z^2 \ldots \theta'_{r,n}z^r &= 0. \end{aligned}$$

For each $1 \leq i \leq r$, let $A_{i,\infty}$ be the $\infty \times \infty$ extension of the sequence of matrices $\{A_{i,p(n)}\}_{n \geq 1}$. Consider the parameter space for $\{A_{i,\infty}\}_{i=1}^r$ as,

$$\mathcal{P}(C, \alpha, \epsilon) = \left\{ \{A_{i,\infty}\}_{i=1}^r : \inf_p \min_{1 \leq i \leq r} (|\alpha_{i,p}|, |\alpha'_{i,p}|) > 1 + \epsilon, \quad \text{and} \quad A_{i,\infty} \in \mathcal{X}(C, \alpha) \ \forall i \right\} \tag{3.30}$$

for some $C, \epsilon, \alpha > 0$. Now, we are prepared to state the following theorem which appeared in Bhattacharjee and Bose [2014b].

Theorem 3.4.6. *If* $\{A_{i,\infty}\}_{i=1}^r \in \mathcal{P}(C, \alpha, \varepsilon)$, *then (3.12) holds. Also,* $\{\phi_{i,\infty}\}_{i=0}^\infty \in \Im(\beta, 0) \bigcap \mathcal{G}(C, \alpha, \eta, 1)$ *for any* $0 < \beta < 1$ *and some* $0 < \eta < 1$.

To prove the above theorem, we need the following two lemmas. Lemma 3.4.7 provides an inequality on matrix norms and Lemma 3.4.8 describes an important property of stationary univariate autoregressive processes.

Lemma 3.4.7. *(Golub and van Loan [1996]) Let* M *be a square matrix. Then* $||M||_2 \leq \sqrt{||M||_{(1,1)}||M^*||_{(1,1)}}$.

Note that Lemma 3.4.7 implies Lemma 2.2.3, when M is a symmetric matrix.

Lemma 3.4.8. *(Brockwell and Davis [2009]) Consider a univariate autoregressive process of order* r*:*

$$x_t = b_1 x_{t-1} + b_2 x_{t-2} + \ldots + b_r x_{t-r} + e_t, \quad \forall t, \tag{3.31}$$

where $\{e_t\}$ *are i.i.d. with mean* 0 *and variance* σ^2*. Now, if* $\{b_i\}$ *satisfies*

$$1 - b_1 z - b_2 z^2 - \ldots - b_r z^r \neq 0, \quad \forall z \in \mathbb{C}, |z| < 1, \tag{3.32}$$

then we have the following representation

$$x_t = \sum_{i=0}^{\infty} d_i e_{t-i}, \quad \text{where} \;\; d_0 = 1, \; d_j = \sum_{i=1}^{j} b_i d_{j-i}, \quad \forall j \geq 1.$$

and moreover, there exists a $0 < \delta < 1$ *and* $c > 0$ *such that* $|d_i| < c\delta^i$*.*

Now, we are ready to prove Theorem 3.4.6.

Proof of Theorem 3.4.6. This proof involves the following three steps.

Step 1: *Proof of (3.12).*

Note that, we need to show that if $\{A_{i,\infty}\}_{i=1}^{r} \in \mathcal{P}(C, \alpha, \varepsilon)$, then it will satisfy condition (3.11). Define the polynomials

$$\begin{aligned} p_1(x) &= \theta_{1,n} x + \theta_{2,n} x^2 + \cdots + \theta_{r,n} x^r \\ p_2(x) &= \theta'_{1,n} x + \theta'_{2,n} x^2 + \cdots + \theta'_{r,n} x^r. \end{aligned}$$

Note that $p_1(0) = p_2(0) = 0$ and both of them are increasing functions of x. Also, as $(1 - p_1(x))$ and $(1 - p_2(x))$ have all their roots strictly greater than $(1 + \varepsilon)$, $p_i(1 + \varepsilon) < 1 \;\; \forall i = 1, 2$. Let us write I and A_i respectively for I_p and $A_{i,p}$ for all $i \geq 1$. Now, for any $|z| \leq 1 + \varepsilon$ and any $x \neq 0$, by Lemma 3.4.7

$$|x' A_1 x z + x' A_2 x z^2 + \cdots + x' A_r x z^r| \leq \sum_{i=1}^{r} \sqrt{\theta_i \theta'_i} |z|^i \leq \frac{1}{2}(p_1(1+\varepsilon) + p_2(1+\varepsilon)) < 1.$$

Hence, there exists no $x \neq 0$ such that $(I - A_1 z - A_2 z^2 \cdots - A_r z^r)x = 0$. Therefore, (3.11) is satisfied.

Step 2: *Proof of* $\{\phi_{i,\infty}\}_{i=0}^{\infty} \in \Im(\beta, 0)$ *for any* $0 < \beta < 1$.

Consider the autoregressive processes

$$\begin{aligned} y_t &= \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_r y_{t-r} + e_t \\ z_t &= \theta'_1 z_{t-1} + \theta'_2 z_{t-2} + \cdots + \theta'_r z_{t-r} + e_t \end{aligned}$$

where $e_t,\ t = 1, 2, ...$ are independently distributed with mean 0 and variance σ^2 and for all $1 \leq i \leq r$, $\theta_i = ||A_{i,\infty}||_{(1,1)}$ and $\theta'_i = ||A^*_{i,\infty}||_{(1,1)}$. If $\{A_{i,\infty}\}_{i=1}^r \in \mathcal{P}(C, \alpha, \varepsilon)$, then by Lemma 3.4.8, we have the representations,

$$y_t = \sum_{i=0}^{\infty} \alpha_i e_{t-i} \ \text{ and } \ z_t = \sum_{i=0}^{\infty} \beta_i e_{t-i}\ \forall t$$

where

$$\alpha_0 = 1,\ \alpha_j = \sum_{i=1}^{j} \theta_i \alpha_{j-i} \ \text{ and } \ \beta_0 = 1,\ \beta_j = \sum_{i=1}^{j} \theta'_i \beta_{j-i}\ \forall j \geq 1$$

and there exist $0 < \delta < 1,\ c > 0$, such that

$$\max(\alpha_i, \beta_i) < c\delta^i\ \forall i.$$

Therefore, using Lemma 3.4.4(i) repeatedly, we have $||\phi_{i,\infty}||_{(1,1)} < \alpha_i$ and hence

$$||\phi_{i,\infty}||_{(1,1)} < c\delta^i \ \ \forall i \ \text{ for some } \ c > 0,\ 0 < \delta < 1. \tag{3.33}$$

Therefore, $\{\phi_{i,\infty}\}_{i=0}^{\infty} \in \Im(\beta, \lambda)$ for any $0 < \beta < 1$ and $\lambda = 0$.

Step 3: *Proof of* $\{\phi_{i,\infty}\} \in \mathcal{G}(C, \alpha, \eta, 1)$ *for some* $0 < \eta < 1$.

By (3.33), the summability condition on $\{||\phi_{i,\infty}||_{(1,1)}\}$ in $\mathcal{G}(C, \alpha, \eta, 1)$ is satisfied. Therefore, it remains to justify the condition on $T(\cdot, \cdot)$ in $\mathcal{G}(C, \alpha, \eta, 1)$.

Now consider any $i \leq r$. Then

$$||\phi_{i,\infty}||_{(1,1)} \leq \sum_{j=1}^{i} ||A_{j,\infty}||_{(1,1)} ||\phi_{i-j,\infty}||_{(1,1)} < c\delta^i.$$

Hence,

$$||A_{i,\infty}||_{(1,1)} < c\delta^i, \ 1 \leq i \leq r.$$

Since $A_{i,\infty} \in \mathcal{X}(C, \alpha)$ $\forall i \leq r$, we have $T(A_{i,\infty}, t) < C_1 \delta^i t^{-\alpha}$, for some $C_1 > 0$. Note that

$$T(\phi_{1,\infty}, (1+\eta)t) < ct^{-\alpha}\delta(1+\eta^{-\alpha}).$$

We now apply induction. Suppose,

$$T(\phi_{j,\infty}, (1+\eta+\cdots+\eta^j)t) < ct^{-\alpha}(1+\eta^{-\alpha}+\cdots+\eta^{-j\alpha})\delta^j j.$$

Then, by Lemma 3.4.4 (iii), for all $j > k$,

$$\begin{aligned}
& T(A_{k,\infty}\phi_{j-k,\infty}, (1+\eta+\cdots+\eta^j)t) \\
\leq \ & \delta^k ct^{-\alpha}(1+\eta^{-\alpha}+\cdots+\eta^{-(j-k)\alpha})\delta^{j-k} j + \delta^{j-k} ct^{-\alpha}\eta^{-(j-k+1)\alpha}\delta^k \\
\leq \ & ct^{-\alpha}\delta^j j(1+\eta^{-\alpha}+\cdots+\eta^{-j\alpha}).
\end{aligned}$$

Since, $\phi_{j+1,\infty} = \sum_{i=0}^{j+1} A_{i,\infty}\phi_{j-i+1,\infty}$, for some $C' > 0$,

$$T(\phi_{j+1,\infty}, (1+\eta+\cdots+\eta^{j+1})t) \leq C' t^{-\alpha}\delta^j j^2 (1+\eta^{-\alpha}+\cdots+\eta^{-(j+1)\alpha}).$$

Hence the proof is complete. □

This complete our discussion on (a) as mentioned at the beginning of Section 3.3. Now we move to (b) i.e. to consistent estimation of autocovariance matrices.

3.5 Estimation of autocovariance matrices

We are now ready to show that appropriate banded and tapered version of the sample autocovariance matrices are consistent for the population autocovariance matrices in the sense of (2.5). Throughout this section, we assume $p = p(n) \to \infty$ as $n \to \infty$ in such a way that $n^{-1} \log p(n) \to 0$.

Recall for any matrix M of order p and $k > 0$, the *k-banded* version of M is as in (2.25). Also recall the tapered version of a matrix as in (2.55). Suppose, $g : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+ \cup \{0\}$ is a continuous, non-increasing function such that $g(0) = 1$, $\int_0^\infty g(x) < \infty$ and $1 - g(x) = O(x^\gamma)$ for some $\nu \geq 1$ in some neighborhood of zero. Then we have the following theorem which has appeared in Bhattacharjee and Bose [2014b].

Theorem 3.5.1. *Consider the model (3.2). Suppose the driving process* $\varepsilon_t \sim \mathcal{N}_p(0, \Sigma_p)$, $\forall t$, $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C)$ *and* $\{\psi_{j,\infty}\} \in \Im(\beta, \lambda) \cap \mathcal{G}(C, \alpha, \eta, \nu)$ *for some* $C, \epsilon, \alpha, \mu > 0$, $\lambda \geq 0$ *and* $0 < \beta, \eta < 1$. *Then for* $k_{n,\alpha} \asymp \left(n^{-1}\log p\right)^{-\frac{1}{2(\alpha+1)}}$, $\tau_{n,\alpha} \asymp (n^{-1} \log p)^{-\frac{1}{2(1+\gamma)}\left[\frac{\gamma}{1+\alpha}+1\right]}$ *and* $u \geq 0$, *we have*

$$\begin{aligned} ||B_{k_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_2 &= O_P(k_{n,\alpha}^{-\alpha}||\Sigma_p||_{(1,1)}), \text{ and} & (3.34)\\ ||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_2 &= O_P\left[\left(n^{-1}\log p\right)^{\frac{\gamma\alpha}{2(1+\alpha)(1+\gamma)}}||\Sigma_p||_{(1,1)})\right]. & (3.35)\end{aligned}$$

To prove the above theorem, we need the following two lemmas. Lemma 3.5.2 provides the convergence rate of the sample autocovariance matrices to their corresponding population autocovariance matrices in $||\cdot||_\infty$ norm for the infinite dimensional IID process. This turns out to be useful since the model (3.2) is driven by an infinite dimensional IID process. Lemma 3.5.3 provides a summability condition which is useful to establish an upper bound to the rate of convergences involved in Theorem 3.5.1.

Lemma 3.5.2. *Suppose, $\{\varepsilon_t\}$ are i.i.d. $\mathcal{N}_p(0, \Sigma_p)$. Then*

$$(i) \quad ||\frac{1}{n}\sum_{t=1}^{n} \varepsilon_t\varepsilon_t^* - \Sigma_p||_\infty = O_P\left(\sqrt{n^{-1}\log p}\right),$$

$$(ii) \quad ||\frac{1}{n}\sum_{t=1}^{n-u} \varepsilon_t\varepsilon_{t+u}^*||_\infty = O_P\left(\sqrt{n^{-1}\log p}\right), \quad \forall u \geq 1.$$

Proof. (i) follows from (2.35). For (ii) Let, $z_{t,i} = \dfrac{\varepsilon_{t,i}}{\sqrt{\Gamma_{0,p.ii}}}$ $\forall i, t$, where $\varepsilon_{t,i}$ is the i-th component of ε_t and $\Gamma_{0,p.ii}$ is the (i,i)th entry of $\Gamma_{0,p}$. Then for some $c_1 > 0$,

$$\begin{aligned} P\left[||\frac{1}{n}\sum_{t=1}^{n-u} \varepsilon_t\varepsilon_{t+u}^*||_\infty > t\right] \; &\leq \; \sum_{l,m} P\left[|\sum_{t\geq 1}^{n-u}\{\frac{(z_{t,l}+z_{t+u,m})^2}{2} - 1\}| > c_1 nt\right] \\ &\quad + \sum_{l,m} P\left[|\sum_{t\geq 1}^{n-u}\{\frac{(z_{t,l}-z_{t+u,m})^2}{2} - 1\}| > c_2 nt\right]. \end{aligned}$$

Since, $\dfrac{(z_{t,l} \pm z_{t+u,m})^2}{2}, t \geq 1$ are all independent χ_1^2 variables, by Lemma 2.30, for some $c_2, c_3 > 0$

$$P[||\frac{1}{n}\sum_{t=1}^{n-u} \varepsilon_t\varepsilon_{t+u}^*||_\infty > t] \leq c_3 p^2 e^{-c_2 nt^2}, \tag{3.36}$$

which tends to 0 as $n \to \infty$ for $t = M\sqrt{n^{-1}\log p}$ and appropriately chosen $M > 0$. Hence, (ii) is proved. □

Lemma 3.5.3. *Let $\{a_j\}_{j=0}^\infty$ be any sequence of positive real members such that $\sum_{j=0}^{\infty} a_j{}^\beta < \infty$ and $\sum_{j=0}^{\infty} a_j^{2(1-\beta)} j^\lambda < \infty$, for some $\lambda > 0$ and $0 < \beta < 1$. Then as $p \to \infty$ for an appropriately chosen $M > 0$,*

$$p^2 \sum_{1\leq i,j<\infty} p^{-\frac{M}{(a_i a_j)^{2(1-\beta)}}} \to 0.$$

Proof. Let $\nu = 2(1-\beta)$. As $\sum_{j=1}^\infty a_j^v j^\lambda < \infty$, we have $\frac{1}{a_j^v} > j^\lambda$ for all $j > N$ and some $N \geq 1$. Now

$$p^2 \sum_{i,j} p^{-\frac{M}{(a_i a_j)^\nu}} \leq \sum_{1 \leq i,j \leq N} p^{-\frac{M}{(a_i a_j)^\nu}+2} + p^2 \sum_{\{1 \leq i,j \leq N\}^c} p^{-M(ij)^\lambda}.$$

As we have finitely many terms, for an appropriately chosen large M, the first term tends to 0 as $p \to \infty$. Now,

$$\begin{aligned} p^2 \sum_{\{1 \leq i,j \leq N\}^c} p^{-M(ij)^\lambda} &\leq p^2 \sum_{k=N}^{\infty} (k - N + 1) p^{-Mk^\lambda} \leq C_1 p^2 \sum_{r=R}^{\infty} r^{\frac{1}{\lambda}} p^{-Mr} \\ &\leq C_1 p^{2-MR} \sum_{r=0}^{\infty} (r+R)^{\frac{1}{\lambda}} p^{-Mr} \leq C_2 p^{2-MR} \end{aligned}$$

for an C_1, C_2 and $R > 0$. This tends to 0 for an appropriately chosen large $M > 0$. Hence the proof is complete. □

Now we are ready to prove Theorem 3.5.1.

Proof of Theorem 3.5.1

Proof of (3.34). By Lemma 3.4.7, we have

$$||B_{k_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_2 \leq \sqrt{||B_{k_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_{(1,1)} ||B_{k,n\alpha}(\hat{\Gamma}^*_{u,p,n}) - \Gamma^*_{u,p}||_{(1,1)}}. \quad (3.37)$$

First, we shall show that

$$||B_{k_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_{(1,1)} = O_P(k_{n,\alpha}^{-\alpha} ||\Sigma_p||_{(1,1)}). \quad (3.38)$$

Using triangle inequality and by Lemma 2.2.3,

$$\begin{aligned} ||B_{k_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_{(1,1)} &\leq ||B_{k_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - B_{k_{n,\alpha}}(\Gamma_{u,p})||_{(1,1)} + T(\Gamma_{u,p}, k_{n,\alpha}) \\ &\leq (2k_{n,\alpha} + 1) ||\hat{\Gamma}_{u,p,n} - \Gamma_{u,p}||_\infty + T(\Gamma_{u,p}, k_{n,\alpha}). \quad (3.39) \end{aligned}$$

By Theorem 3.4.3, we have

$$T(\Gamma_{u,p}, k_{n,\alpha}) = O(k_{n,\alpha}^{-\alpha}||\Sigma_p||_{(1,1)}). \tag{3.40}$$

Using the model (3.2) and Lemma 3.4.4 (ii),

$$||\hat{\Gamma}_{u,p,n} - \Gamma_{u,p}||_\infty \leq \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} r_i r_j ||\frac{1}{n}\sum_{t=1}^{n-u} \varepsilon_{t,j}\varepsilon^*_{(t+u),i} - E_{ij}||_\infty$$

where $E_{ij} = E\varepsilon_{t,j}\varepsilon^*_{(t+u),i}$ $\forall i, j$. Hence for some $C_1 > 0$,

$$\begin{aligned}
& P[||\hat{\Gamma}_{u,p,n} - \Gamma_{u,p}||_\infty > t] \\
\leq \; & P\left[\sum_j\sum_i r_i r_j ||\frac{1}{n}\sum_{t=1}^{n-u} \varepsilon_{t,j}\varepsilon^*_{(t+u),i} - E_{ij}||_\infty > \sum_j\sum_i \frac{C_1 t}{r_i^{-\beta} r_j^{-\beta}}\right] \\
\leq \; & \sum_j\sum_i P\left[||\frac{1}{n}\sum_{t=1}^{n-u} \varepsilon_{t,j}\varepsilon^*_{(t+u),i} - E_{ij}||_\infty > \frac{C_1 t}{r_i^{1-\beta} r_j^{1-\beta}}\right].
\end{aligned}$$

Now, by (2.44) and (3.36), for $t = M\sqrt{n^{-1}\log p}$,

$$P[||\hat{\Gamma}_{u,p,n} - \Gamma_{u,p}||_\infty > t] \leq p^2 \sum_{i,j} p^{-\frac{M}{(r_i r_j)^{2(1-\beta)}}}.$$

By Lemma 3.5.3, this tends to zero as $n \to \infty$. Hence,

$$||\hat{\Gamma}_{u,p,n} - \Gamma_{u,p}||_\infty = O_P(\sqrt{n^{-1}\log p}), \quad \forall u \geq 0. \tag{3.41}$$

and by (3.39) and (3.40), the proof of (3.38) is complete.

Similarly, one can show that

$$||B_{k_{n,\alpha}}(\hat{\Gamma}_{u,p,n})^* - \Gamma^*_{u,p}||_{(1,1)} = O_P(k_{n,\alpha}^{-\alpha}||\Sigma_p||_{(1,1)}). \tag{3.42}$$

Therefore, putting together (3.37), (3.38), (3.41) and (3.42), proof of (3.34) is complete.

Proof of (3.35). By Lemma 3.4.7, we have

$$||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_2 \leq \sqrt{||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_{(1,1)}||R_{\tau_{n,\alpha}}(\hat{\Gamma}^*_{u,p,n}) - \Gamma^*_{u,p}||_{(1,1)}}. \quad (3.43)$$

First, we shall show that

$$||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_{(1,1)} = O_P\Big[\Big(n^{-1}\log p\Big)^{\frac{\gamma\alpha}{2(1+\alpha)(1+\gamma)}}||\Sigma_p||_{(1,1)})\Big]. \quad (3.44)$$

Using triangle inequality,

$$\begin{aligned} ||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_{(1,1)} &\leq ||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - R_{\tau_{n,\alpha}}(\Gamma_{u,p})||_{(1,1)} \\ &\quad + ||R_{\tau_{n,\alpha}}(\Gamma_{u,p}) - \Gamma_{u,p}||_{(1,1)}. \quad (3.45) \end{aligned}$$

Now, for some constant $C_1 > 0$,

$$\begin{aligned} ||R_{\tau_{n,\alpha}}\left(\hat{\Gamma}_{u,p,n}\right) - R_{\tau_{n,\alpha}}\left(\Gamma_{u,p}\right)||_{(1,1)} &\leq ||\hat{\Gamma}_{u,p,n} - \Gamma_{u,p}||_\infty \left(2\sum_{l=0}^{p} g\left(\frac{l}{\tau_{n,\alpha}}\right)\right) \\ &\leq \tau_{n,\alpha}||\hat{\Gamma}_{u,p,n} - \Gamma_{u,p}||_\infty \left[C_1 \int_0^\infty g(x)dx\right]. \end{aligned}$$

Therefore, by (3.41), we have

$$||R_{\tau_{n,\alpha}}\left(\hat{\Gamma}_{u,p,n}\right) - R_{\tau_{n,\alpha}}\left(\Gamma_{u,p}\right)||_{(1,1)} = O_P(\tau_{n,\alpha}\sqrt{n^{-1}\log p}). \quad (3.46)$$

Again, by triangle inequality

$$\begin{aligned} ||R_{\tau_{n,\alpha}}(\Gamma_{u,p}) - \Gamma_{u,p}||_{(1,1)} &\leq ||R_{\tau_{n,\alpha}}(\Gamma_{u,p}) - B_{k'_{n,\alpha}}\big[R_{\tau_{n,\alpha}}(\Gamma_{u,p})\big]||_{(1,1)} \\ &\quad + ||B_{k'_{n,\alpha}}\big[R_{\tau_{n,\alpha}}(\Gamma_{u,p})\big] - B_{k'_{n,\alpha}}(\Gamma_{u,p})||_{(1,1)} \\ &\quad + ||B_{k'_{n,\alpha}}(\Gamma_{u,p}) - \Gamma_{u,p}||_{(1,1)}. \quad (3.47) \end{aligned}$$

By Lemma 2.2.1 and Theorem 3.4.3, we have

$$\begin{aligned} ||B_{k'_{n,\alpha}}(\Gamma_{u,p}) - \Gamma_{u,p}||_{(1,1)} &= O((k'_{n,\alpha})^{-\alpha}||\Sigma_p||_{(1,1)}) \text{ and} && (3.48)\\ ||R_{\tau_{n,\alpha}}(\Gamma_{u,p}) - B_{k'_{n,\alpha}}\left[R_{\tau_{n,\alpha}}(\Gamma_{u,p})\right]||_{(1,1)} &= O((k'_{n,\alpha})^{-\alpha}||\Sigma_p||_{(1,1)}). && (3.49)\end{aligned}$$

Now, by Theorem 3.4.1, for some constant $C_2, C_3 > 0$ and for sufficiently large n,

$$\begin{aligned} & ||B_{k'_{n,\alpha}}\left[R_{\tau_{n,\alpha}}(\Gamma_{u,p})\right] - B_{k'_{n,\alpha}}(\Gamma_{u,p})||_{(1,1)} \\ \leq\ & \max_i \sum_{j:|i-j|\leq k'_{n,\alpha}} \left(1 - g\left(\frac{|i-j|}{\tau'_{n,\alpha}}\right)\right)\left(\sup_p ||\Gamma_{u,p}||_\infty\right) \\ \leq\ & C_2 \sum_{l=-k'_{n,\alpha}}^{k'_{n,\alpha}} \left(1 - g\left(\frac{l}{\tau_{n,\alpha}}\right)\right) \leq C_3 \left(\frac{k'_{n,\alpha}}{\tau_{n,\alpha}}\right)^\gamma k'_{n,\alpha}. \qquad (3.50)\end{aligned}$$

Now, consider $k'_{n,\alpha} = (n^{-1}\log p)^{-\frac{\gamma}{2(1+\gamma)(1+\alpha)}}$. Therefore, by (3.45)-(3.50), the proof of (3.44) is complete.

Similarly, one can show that

$$||R_{\tau_{n,\alpha}}(\hat{\Gamma}^*_{u,p,n}) - \Gamma^*_{u,p}||_{(1,1)} = O_P\Big[\Big(n^{-1}\log p\Big)^{\frac{\gamma\alpha}{2(1+\alpha)(1+\gamma)}}||\Sigma_p||_{(1,1)})\Big]. \qquad (3.51)$$

Hence, by (3.43), (3.44) and (3.51), the proof of (3.35) is complete. Therefore, Theorem 3.5.1 is proved. □

Remark 3.5.4. *Note that the rate of convergence depends not only on the class of parameters of the coefficient and variance-covariance matrix but also on* $\|\Sigma_p\|_{(1,1)}$. *This is to be expected since we are considering linear regression type models. Moreover, if* $\|\Sigma_p\|_{(1,1)}$ *is bounded, then the rate of convergence for the marginal variance-covariance matrix* $\Gamma_{0,p}$ *is same as that for infinite dimensional IID process as given in Theorem 2.2.2.*

We now specialize Theorem 3.5.1 to two cases: the infinite dimensional MA(r) and IVAR(r) processes. The next two theorems follow directly from Theorem 3.5.1

once we invoke Theorems 3.4.5 and 3.4.6. These results appeared in Bhattacharjee and Bose [2014b].

Theorem 3.5.5. *Consider the model (3.7). Suppose the driving process* $\varepsilon_t \sim \mathcal{N}_p(0, \Sigma_p)$, $\forall t$. *Also suppose* $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C)$, $0 < ||M_{i,\infty}||_{(1,1)} < \infty$ *and* $M_{i,\infty} \in \mathcal{X}(\alpha, C)$ *for some* $\epsilon, \alpha, C > 0$ *and for all* $1 \leq i \leq r$. *Then for* $k_{n,\alpha} \asymp \left(n^{-1}\log p\right)^{-\frac{1}{2(\alpha+1)}}$, $\tau_{n,\alpha} \asymp (n^{-1}\log p)^{-\frac{1}{2(1+\gamma)}\left[\frac{\gamma}{1+\alpha}+1\right]}$ *and* $u \geq 0$, *we have*

$$||B_{k_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_2 = O_P(k_{n,\alpha}^{-\alpha}||\Sigma_p||_{(1,1)}), \text{ and} \tag{3.52}$$

$$||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_2 = O_P\Big[\Big(n^{-1}\log p\Big)^{\frac{\gamma\alpha}{2(1+\alpha)(1+\gamma)}}||\Sigma_p||_{(1,1)}\Big]. \tag{3.53}$$

Theorem 3.5.6. *Consider the model (3.9). Suppose the driving process* $\varepsilon_t \sim \mathcal{N}_p(0, \Sigma_p)$, $\forall t$. *Also suppose* $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C)$ *and* $\{A_{i,\infty}\}_{i=1}^r \in \mathcal{P}(C, \alpha, \varepsilon)$ *for some* $\alpha, \epsilon, C > 0$ *and for all* $1 \leq i \leq r$. *Then for* $k_{n,\alpha} \asymp \left(n^{-1}\log p\right)^{-\frac{1}{2(\alpha+1)}}$, $\tau_{n,\alpha} \asymp (n^{-1}\log p)^{-\frac{1}{2(1+\gamma)}\left[\frac{\gamma}{1+\alpha}+1\right]}$ *and* $u \geq 0$, *we have*

$$||B_{k_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_2 = O_P(k_{n,\alpha}^{-\alpha}||\Sigma_p||_{(1,1)}), \text{ and} \tag{3.54}$$

$$||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{u,p,n}) - \Gamma_{u,p}||_2 = O_P\Big[\Big(n^{-1}\log p\Big)^{\frac{\gamma\alpha}{2(1+\alpha)(1+\gamma)}}||\Sigma_p||_{(1,1)}\Big]. \tag{3.55}$$

Our next task is to consistently estimate the parameter matrices $\{A_{i,p} : 1 \leq i \leq r\}$ and the variance-covariance matrix of the driving process $\{\varepsilon_{t,p}\}$ i.e. Σ_p for the IVAR(r) process defined in (3.9).

3.5.1 Estimation of parameter matrices for IVAR(r)

By right multiplying both sides of (3.9) with $X_{t-k,p}^*$, $k = 1, 2, \ldots r$ successively and then taking expectation, we have

$$\Gamma_{1,p}^* = A_{1,p}\Gamma_{0,p} + A_{2,p}\Gamma_{1,p} + \ldots + A_{r,p}\Gamma_{r-1,p} \tag{3.56}$$

$$\Gamma_{2,p}^* = A_{1,p}\Gamma_{1,p}^* + A_{2,p}\Gamma_{0,p} + \ldots + A_{r,p}\Gamma_{r-2,p}$$

$$\vdots$$

$$\Gamma^*_{r,p} = A_{1,p}\Gamma^*_{r-1,p} + A_{2,p}\Gamma^*_{r-2,p} + \ldots + A_{r,p}\Gamma_{0,p}.$$

Let

$$\mathcal{Y}_{r,n} = (\Gamma_{1,p}, \Gamma_{2,p}, \ldots, \Gamma_{r,p})^*, \quad \mathcal{A}_{r,n} = (A^*_{1,p}, A^*_{2,p}, \ldots, A^*_{r,p})^*$$

and let $G_{r,n}$ be a block matrix with r^2 many $p \times p$ blocks

$$G_{r,n}(i,j) = \Gamma_{|i-j|,p} I(i < j) + \Gamma^*_{|i-j|,p} I(i \geq j),\ 1 \leq i, j \leq r. \tag{3.57}$$

Then from (3.56) we have,

$$\mathcal{Y}_{r,n} = G_{r,n}\mathcal{A}_{r,n}. \tag{3.58}$$

This is analogous to the Yule-Walker equations for finite dimensional AR process. The following lemma implies the invertibility of the matrix $G_{r,n}$ for all $n \geq 1$. Recall the definition of $\lambda_{\min}$ in (2.13).

Lemma 3.5.7. *Fix any $n \geq 1$. If $\lambda_{\min}(\Gamma_{0,p(n)}) > 0$ and $||\Gamma_{h,p(n)}||_2 \to 0$ as $h \to \infty$, then $G_{r,n}$ is non-singular.*

Proof. Suppose that $G_{q,p}$ is non-singular but $G_{q+1,p}$ is singular. Then there exist $a, a_1, a_2, \ldots, a_q$ such that

$$a^* X_{q+1,p} = \sum_{j=1}^{q} a_j^* X_{j,p} \text{ a.s..}$$

By stationarity,

$$a^* X_{q+h+1,p} = \sum_{j=1}^{q} a_j^* X_{h+j,p}\ \forall h \geq 1 \text{ a.s..}$$

So, $\forall K \geq q+1, \exists\, a_1^{(K)}, a_2^{(K)}, \ldots, a_q^{(K)}$ such that $A^{(K)} = (a_1^{(K)*}, a_2^{(K)*}, \ldots, a_q^{(K)*}), Y_{q,p} =$

$(X_{1,p}^*, X_{2,p}^* \ldots, X_{q,p}^*)^*$ and $a^* X_{K,p} = A^{(K)} Y_{q,p}$. Hence,

$$a^*\Gamma_{0,p}a = A^{(K)}G_{q,p}A^{(K)^*} \;\; \geq \;\; \lambda_1 A^{(K)}A^{(K)^*} = \lambda_1 \sum_{i=1}^{q} ||a_i^{(K)}||_2$$

where λ_1 is the smallest eigenvalue of $G_{q,p}$. Therefore, $||a_i^{(K)}||_2$ are bounded function of K for all i. Again,

$$\begin{aligned} a^* X_{K,p} &= A^{(K)} Y_{q,p} \Rightarrow a^* X_{K,p} X_{K,p}^* a = A^{(K)} Y_{q,p} X_{K,p}^* a \\ \Rightarrow a^* \Gamma_{0,p} a &= \sum_{j=1}^{q} a_j^{(K)*} \Gamma_{K-j,p} a. \end{aligned}$$

Hence,

$$|a^*\Gamma_{0,p}a| \leq \sum_{j=1}^{q} ||a_j^{(K)}||_2 ||\Gamma_{K-j,p}||_2 ||a||_2 \leq C \sum_{j=1}^{q} ||\Gamma_{K-j,p}||_2$$

for some $C > 0$ and tends to zero as $K \to \infty$. So, $a^*\Gamma_{0,p}a = 0$ for some $a \neq 0$. This contradicts the assumption $\lambda_{\min}(\Gamma_{0,p}) > 0$. Hence, the result holds as $G_{1,p} = \Gamma_{0,p}$ is non-singular. $\square$

Recall the class of dispersion matrices $\mathcal{W}(\epsilon)$ in (2.16). It is easy to see that, for the model (3.9), if $\{A_{i,\infty}\} \in \mathcal{P}(C, \alpha, \epsilon)$, $\Sigma_\infty \in \mathcal{W}(\epsilon)$ for some $C, \alpha, \epsilon > 0$, then $||\Gamma_{h,p(n)}||_2 \to 0$ as $h \to \infty$ and for all $n \geq 1$. The above statement follows because, by Theorem 3.4.6, (3.9) can be represented in the form (3.12), (3.33) holds and

$$\Gamma_{u,p} = \sum_{j=0}^{\infty} \phi_{j,p} \Sigma_p \psi_{j+u,p}^*, \quad \forall u \geq 0. \tag{3.59}$$

Therefore, by (3.33), Lemma 3.4.7 and as $\Sigma_\infty \in \mathcal{W}(\epsilon)$, for some $C_1 > 0$, we have

$$\begin{aligned} ||\Gamma_{u,p}||_2 &\leq \sum_{j=0}^{\infty} ||\phi_{j,p}||_2 ||\Sigma_p||_2 ||\psi_{j+u,p}^*||_2 \\ &\leq \epsilon^{-1} \sum_{j=0}^{\infty} \sqrt{||\phi_{j,p}||_{(1,1)} ||\phi_{j,p}^*||_{(1,1)}} \sqrt{||\phi_{j,p}||_{(1,1)} ||\phi_{j+u,p}^*||_{(1,1)}} \end{aligned}$$

$$\begin{aligned} &\leq \quad C_1 \epsilon^{-1} \delta^u \left(\sum_{j=0}^{\infty} \delta^{2j} \right), \quad 0 < \delta < 1. \\ &\to \quad 0, \quad \text{as } u \to \infty \text{ and for all } n \geq 1. \end{aligned} \tag{3.60}$$

Therefore, for the model (3.9), if $\Gamma_{0,p(n)}$ is non-singular for each $n \geq 1$, then

$$\mathcal{A}_{r,n} = G_{r,n}^{-1} \mathcal{Y}_{r,n} \tag{3.61}$$

i.e., each $A_{i,p}$ is the finite sum of the finite products of $\{\Gamma_{u,p}, \Gamma_{u,p}^{-1},\ 1 \leq u \leq r\}$. Hence, (3.61) provides consistent estimates of A_i, once we replace the population autocovariance matrices by their consistent estimates. We illustrate this by the IVAR(1) model. Similar technique is also applicable to estimate the parameter matrices for other finite order IVAR processes. The following result appeared in Bhattacharjee and Bose [2014b].

Theorem 3.5.8. *Consider the model (3.9) for $r = 1$. Suppose $\varepsilon_t \sim \mathcal{N}_p(0, \Sigma_p)$, $\Sigma_\infty \in \mathcal{U}(\epsilon, \alpha, C)$ and $A_{1,\infty} \in \mathcal{A}(\delta, C, \alpha)$ for some $\epsilon, \alpha, C > 0$ and $0 < \delta < 1$, where*

$$\mathcal{A}(\delta, C, \alpha) = \left\{ A_\infty : \max(||A_\infty||_{(1,1)}, ||A_\infty^*||_{(1,1)}) < (1 - \delta),\ A_\infty,\ A_\infty^* \in \mathcal{X}(C, \alpha) \right\}.$$

Suppose all the inverses below exist. Then for $k_{n,\alpha} \asymp \left(n^{-1} \log p\right)^{-\frac{1}{2(\alpha+1)}}$ and $\tau_{n,\alpha} \asymp (n^{-1} \log p)^{-\frac{1}{2(1+\gamma)}\left[\frac{\gamma}{1+\alpha}+1\right]}$,

(i) $||B_{k_{n,\alpha}}(\hat{\Gamma}_{1,p,n})(B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p,n}))^{-1} - A_{1,p}||_2 = O_P(k_{n,\alpha}^{-\alpha} ||\Sigma_p||_{(1,1)})$,

(ii) $||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{1,p,n})(R_{\tau_{n,\alpha}}(\hat{\Gamma}_{0,p,n}))^{-1} - A_{1,p}||_2 = O_P\Big[\Big(n^{-1} \log p\Big)^{\frac{\gamma\alpha}{2(1+\alpha)(1+\gamma)}} ||\Sigma_p||_{(1,1)})\Big]$,

(iii) $||\hat{\Sigma}_{p,n,\alpha} - \Sigma_p||_2 = O_P(||\Sigma_p||_{(1,1)} k_{n,\alpha}^{-\alpha})$ *and*

(iv) $||\hat{\hat{\Sigma}}_{p,n,\alpha} - \Sigma_p||_2 = O_P\Big[\Big(n^{-1} \log p\Big)^{\frac{\gamma\alpha}{2(1+\alpha)(1+\gamma)}} ||\Sigma_p||_{(1,1)})\Big]$,

where

$$\hat{\Sigma}_{p,n,\alpha} \quad = \quad B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p}) - B_{k_{n,\alpha}}(\hat{\Gamma}_{1,p})(B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p}))^{-1} B_{k_{n,\alpha}}(\hat{\Gamma}_{1,p}^*),$$

$$\hat{\hat{\Sigma}}_{p,n,\alpha} = R_{\tau_{n,\alpha}}(\hat{\Gamma}_{0,p}) - R_{\tau_{n,\alpha}}(\hat{\Gamma}_{1,p})(R_{\tau_{n,\alpha}}(\hat{\Gamma}_{0,p}))^{-1}R_{\tau_{n,\alpha}}(\hat{\Gamma}^*_{1,p}).$$

To prove the above theorem, we need the following lemma.

Lemma 3.5.9. *(Bhatia [2009]) If A and B are invertible and $||A-B||_2 \leq ||A^{-1}||_2^{-1}$, then*

$$||B^{-1} - A^{-1}||_2 \leq \frac{||A^{-1}||_2^2||A-B||_2}{1-||A^{-1}||_2||A-B||_2}. \tag{3.62}$$

Proof of Theorem 3.5.8. It is easy to see that, $A_{1,\infty} \in \mathcal{A}(\delta, C, \alpha)$ implies $A_{1,\infty} \in \mathcal{P}(C, \alpha, \delta(1-\delta)^{-1})$. Therefore, the conclusions of Theorem 3.5.6 hold.

Using Lemma 3.5.9, for large n,

$$\begin{aligned}
||(B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p,n}))^{-1} - \Gamma_{0,p}^{-1}||_2 &\leq \frac{||\Gamma_{0,p}^{-1}||_2^2||B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p,n}) - \Gamma_{0,p}||_2}{1-||\Gamma_{0,p}^{-1}||_2||B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p,n}) - \Gamma_{0,p}||_2}, \\
||(R_{\tau_{n,\alpha}}(\hat{\Gamma}_{0,p,n}))^{-1} - \Gamma_{0,p}^{-1}||_2 &\leq \frac{||\Gamma_{0,p}^{-1}||_2^2||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{0,p,n}) - \Gamma_{0,p}||_2}{1-||\Gamma_{0,p}^{-1}||_2||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{0,p,n}) - \Gamma_{0,p}||_2}.
\end{aligned}$$

If $n^{-1}\log p \to 0$, then for some $C > 0$ and for sufficiently large n,

$$\begin{aligned}
||(B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p,n}))^{-1} - \Gamma_{0,p}^{-1}||_2 &\leq C||B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p,n}) - \Gamma_{0,p}||_2, \\
||(R_{\tau_{n,\alpha}}(\hat{\Gamma}_{0,p,n}))^{-1} - \Gamma_{0,p}^{-1}||_2 &\leq C||R_{\tau_{n,\alpha}}(\hat{\Gamma}_{0,p,n}) - \Gamma_{0,p}||_2.
\end{aligned} \tag{3.63}$$

Therefore, by Theorem 3.5.6

$$\begin{aligned}
||(B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p,n}))^{-1} - \Gamma_{0,p}^{-1}||_2 &= O_P(||\Sigma_p||_{(1,1)}k_{n,\alpha}^{-\alpha}), \\
||(R_{\tau_{n,\alpha}}(\hat{\Gamma}_{0,p,n}))^{-1} - \Gamma_{0,p}^{-1}||_2 &= O_P\Big[\Big(n^{-1}\log p\Big)^{\frac{\gamma\alpha}{2(1+\alpha)(1+\gamma)}}||\Sigma_p||_{(1,1)})\Big].
\end{aligned} \tag{3.64}$$

Again, by the fact

$$||AB - CD||_2 \leq ||A-C||_2||B-D||_2 + ||A-C||_2||D||_2 + ||C||_2||B-D||_2$$

and using $A_{1,p} = \Gamma_{1,p}\Gamma_{0,p}^{-1}$, (i) and (ii) follow.

Results (iii) and (iv) are immediate from the relation

$$\Gamma_{0,p} - A_{1,p}\Gamma_{1,p}^{*} - \Gamma_{1,p}A_{1,p}^{*} + A_{1,p}\Gamma_{0,p}A_{1,p}^{*} = \Sigma_p.$$

This completes the proof of Theorem 3.5.8. $\square$

Next we shall relax the Gaussian assumption on the driving process $\{\varepsilon_{t,p}\}$ in Theorems 3.5.1, 3.5.5, 3.5.6 and 3.5.8 and, Lemma 3.5.2.

3.5.2 Relaxing the Gaussian assumption

The Gaussianity assumption made so far (see Theorems 3.5.5, 3.5.6 and 3.5.8) may seem to be a very strong restriction. However, note that in the proofs of these theorems, Gaussian assumption is used only while invoking Theorem 3.5.1. Moreover, the proof of Theorem 3.5.1 uses the Gaussian assumption only via application of Lemma 3.5.2. Our goal is to now replace the Gaussian assumption by a suitable weaker assumption in Lemma 3.5.2.

The cue to the answer lies in Theorem 2.2.2. Bickel and Levina [2008] first proved the consistency of the variance-covariance matrix $\hat{\Sigma}_p$ for the IID process under the assumption $\varepsilon_t \sim \mathcal{N}_p(0, \Sigma_p)$. Later they relaxed this assumption and proved (2.27) under the weaker assumption that,

$$\sup_{j\geq 1} E(e^{\lambda\varepsilon_{t,j}}) < \infty \quad \text{for all } |\lambda| < \lambda_0 \text{ and some } \lambda_0 > 0, \tag{3.65}$$

where $\varepsilon_{t,j}$ is the j-th element of ε_t.

As a prelude we need the following lemma. For $n \geq 1$, let $U_1, U_2, \ldots, U_n$ be independent random variables with

$$EU_j = 0 \quad \text{and} \quad \sigma_j^2 = Var(U_j) > 0, j = 1, 2, \ldots.$$

Set

$$S_n = \sum_{j=1}^{n} U_j \quad \text{and} \quad B_n^2 = \sum_{j=1}^{n} \sigma_j^2, \quad Z_n = \frac{S_n}{B_n}.$$

We say that $\{U_j\}$ *satisfies condition* (P), if there exist positive constants $A, C, C_1, C_2, \ldots$ such that for all $|z| < A$ and $j = 1, 2, \ldots,$

$$|\frac{\ln E(e^{zU_j})}{Z^2}| \le C_j^2 \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{B_n^2} \sum_{j=1}^{n} C_j^2 \le C. \tag{3.66}$$

Lemma 3.5.10. *Suppose a sequence of random variables $\{U_j\}$ with $EU_j = 0$ and $\sigma_j^2 = Var(U_j) > 0$ satisfies condition (P). Then there exist some $A,\ C > 0$ such that*

$$|Cum_k(Z_n)| \le \frac{k!C}{(AB_n)^{k-2}}\ \forall k \ge 3.$$

That is, for the random variable $\xi = Z_n$, the conclusion of Lemma 2.4.1 holds with $\nu = 0, H = 2C, \bar{\Delta} = AB_n$. In particular, if U_i are i.i.d. then, (3.66) holds if

$$|\frac{\ln E(e^{zU_1})}{z^2}| \le C, \quad \textit{for all } |z| < A, \quad \textit{for some} \ \ A,\ C > 0. \tag{3.67}$$

Also for a random variable U with $EU = 0$, if there exists $A',\ C' > 0$ such that $E(e^{\lambda U}) \le C$ for all $|\lambda| < A$, then (3.67) holds.

Proof. The first part of the lemma is easy to show and the proof is given in Saulis and Statulevičius [1991]. Therefore, it remains to prove the last statement. The cumulants $\{K_n\}$ of a random variable U are defined by the cumulant generating function

$$g(z) = \log(E(e^{zU})) = \sum_{n=1}^{\infty} K_n \frac{z^n}{n!}.$$

The cumulants are related to the moments $\{\mu_n' = E(U^n)\}$ by the following recursion formula

$$K_n = \mu_n' - \sum_{m=1}^{n-1} \binom{n-1}{m-1} K_m \mu_{n-m}'.$$

As all the moments of U exist, $K_n, n = 1, 2, \ldots$ also exist. Moreover, $K_1 = \mu_1' = 0$

and $K_2 = \mu_2' - \mu_1' = \mu_2'$. Hence,

$$|\frac{g(z)}{z^2}| \leq \frac{\mu_2'}{2!} + \sum_{n=3}^{\infty} K_n \frac{|A|^{n-2}}{n!}.$$

This completes the proof. □

Lemma 3.5.11. *Let $\{\varepsilon_t\}$ be i.i.d. with mean* 0 *and variance-covariance matrix Σ_p. Suppose (3.65) holds. Then (i) and (ii) of Lemma 3.5.2 hold.*

Proof. (i) follows from Bickel and Levina [2008]. For (ii), using Lemma 3.5.10, we need the existence of Moment generating function of $\frac{(Z_{t,j} \pm Z_{(t+u),l})^2}{2} - 1 \ \forall j, l$ in some neighborhood of zero. This existence follows from the fact that $(x+y)^2 < 2(x^2+y^2)$. □

Thus the conclusions of Theorems 3.5.1, 3.5.5, 3.5.6, 3.5.8 and Lemma 3.5.2 hold true if we assume (3.65) instead of $\varepsilon_t \sim \mathcal{N}_p(0, \Sigma_p)$.

3.6 Simulations

Consider the IVAR model (3.9) for $r = 1$. This section reports simulations for this model with two different choices of the parameter matrix $A_{1,p}$ which have the Toeplitz structure. As we move away from the main diagonal, in one case the entries decrease exponentially and in the other case, they decrease polynomially. Our simulations show that the convergence rate obtained in Theorem 3.5.8 is quite sharp. Establishing the exact rate appears to be a very difficult problem and we did not pursue this issue.

Example 3.6.1. ***Exponentially decaying corners:*** *Consider the IVAR(1) model with the symmetric parameter matrix $A_{1,\infty} = ((\ (-0.5)^{|i-j|}\ ))$. Note that*

$$||A_{1,\infty}||_{(1,1)} \ = \ ||A^*_{1,\infty}||_{(1,1)} \leq 1 + 2\sum_{u=1}^{\infty}(-0.5)^u = 1 - 2/3,$$

$$T(A_{1,\infty}, k) \leq 2 \sum_{u=k+1}^{\infty} (-0.5)^u \leq (2/3)(0.5)^k < (2/3)k^{-1}.$$

Therefore, $A_{1,\infty} \in \mathcal{A}(2/3, 2/3, 1)$ *and the conclusion of Theorem 3.5.8 hold.*

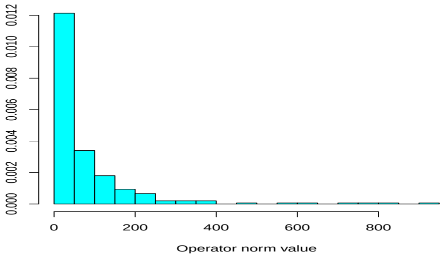
Example 3.6.2. ***Polynomially decaying corners:*** *Consider the* $IVAR(1)$ *model with the symmetric parameter matrix* $A_{1,\infty} = ((\ (-1)^{|i-j|}(|i-j|+1)^{-\beta}\ ))$, *for some* $\beta > 1$. *Note that, in this case*

$$\begin{aligned} ||A_{1,\infty}||_{(1,1)} &= ||A^*_{1,\infty}||_{(1,1)} \leq 1 + 2\sum_{u=1}^{\infty}(-1)^u(u+1)^{-\beta} \leq 1 - 2(2^{-\beta} - 3^{-\beta}), \\ T(A_{1,\infty}, k) &\leq 2 \sum_{u=k+1}^{\infty} (-1)^u(u+1)^{-\beta} \leq 2\int_k^{\infty} x^{-\beta}dx = 2(\beta-1)^{-1}k^{-(\beta-1)}. \end{aligned}$$

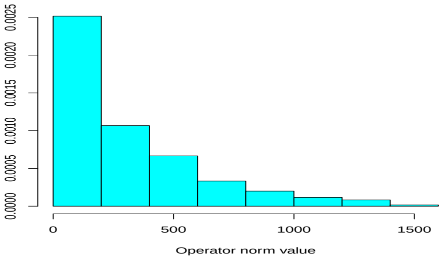
Therefore, $A_{1,\infty} \in \mathcal{A}(2(2^{-\beta}-3^{-\beta}), 2(\beta-1)^{-1}, \beta-1)$ *and the conclusion of Theorem 3.5.8 hold. For the following simulations, we choose* $\beta = 1.01, 1.1, 1.2$ *and* 1.5.

Recall I_k in (2.8). We let $\varepsilon_t \sim \mathcal{N}_p(0, I_p)$, $\forall t$. In each case, we draw the histogram for the values of $||B_{k_{n,\alpha}}(\hat{\Gamma}_{1,p,n})(B_{k_{n,\alpha}}(\hat{\Gamma}_{0,p,n}))^{-1} - A_{1,p}||_2$ using $R = 300$ replications. We consider two combinations of n an p, namely $n = 20,\ \ p = e^{\sqrt{n}} \sim 87$ and $n = 40,\ \ p = e^{\sqrt{n}} \sim 558$.

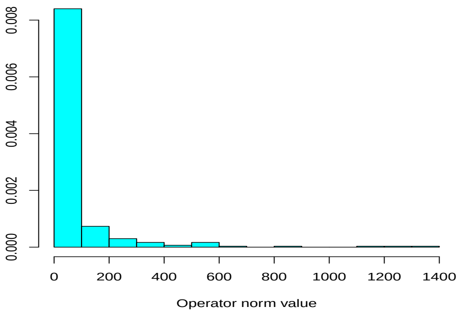
Note that most of the mass is concentrated near zero. Expectedly, the accuracy is sharper in the first example than in the second example. Moreover, as β increases, the histogram has more mass near zero and there is some mass in the high values of the tail. Some stray values beyond the range given in the figures were observed over the different sets of simulations but overall most of the mass was concentrated in the range $(0, 600)$. This indicates that the rates of convergence are probably quite sharp. The theoretical results on the exact rate of convergence appear hard to derive. We have not investigate this issue in this thesis.



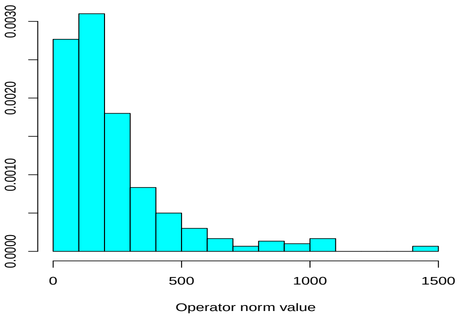
Example 3.6.1 n = *20*



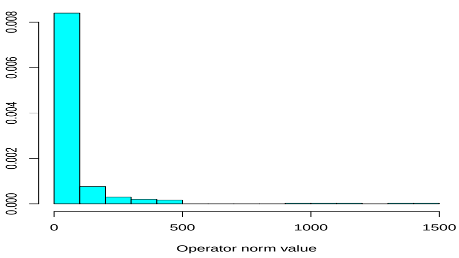
Example 3.6.1 n = *40*



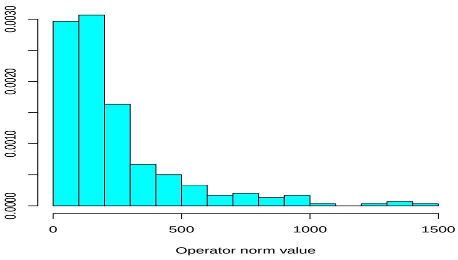
Example 3.6.2 $\beta = 1.01, \quad n = 20$



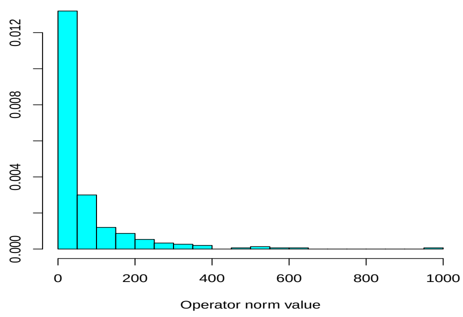
Example 3.6.2 $\beta = 1.01, \quad n = 40$



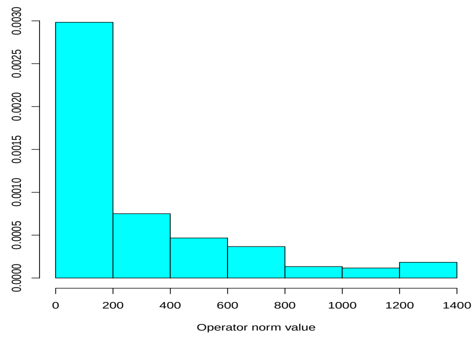
Example 3.6.2 $\beta = 1.1, \quad n = 20$



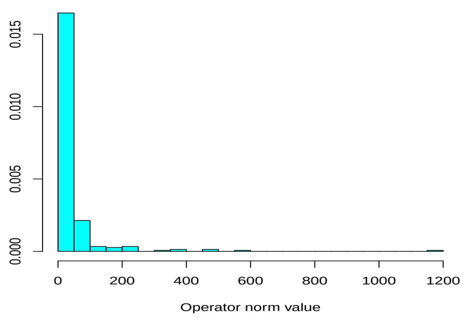
Example 3.6.2 $\beta = 1.1, \quad n = 40$



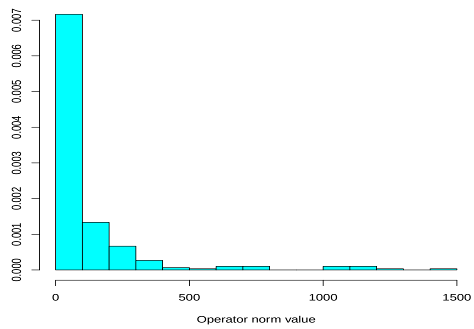
Example 3.6.2 $\beta = 1.2, \quad n = 20$



Example 3.6.2 $\beta = 1.2, \quad n = 40$



Example 3.6.2 $\beta = 1.5, \quad n = 20$



Example 3.6.2 $\beta = 1.5, \quad n = 40$

Chapter 4

Limiting spectral distribution and free probability

4.1 Introduction

In Chapter 3, we have seen that a very general high dimensional linear time series model is the infinite dimensional moving average process of order ∞ (MA(∞)) defined in (3.2) and a key quantity to analyze this model is the sequence of sample autocovariance matrices $\{\hat{\Gamma}_u\}$ defined in (3.16). There we used some regularization, namely banding and tapering on $\{\hat{\Gamma}_u\}$ so that they turn out to be consistent estimator for their population counterpart. In the next few chapters, we explore further asymptotic properties of $\{\hat{\Gamma}_u\}$. Even though these matrices are not consistent, their asymptotic properties, while interesting in their own right, can also be used for statistical applications.

A most natural way to look at the large sample behaviour of the sample autocovariance matrices is to study their limiting spectral distribution. Such results for various random matrices occupy a central position in the literature of random matrix theory (RMT). The so called spectral statistics, useful in statistical application, are functions of this spectral distribution. In case of infinite dimensional moving average processes, many researchers observed, under some assumptions, the eigenvalue distribution of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ (after appropriate normalization) weakly converges to a non-degenerate distribution. For example see Liu et al. [2015] and Wang et al. [2015]. This may be useful to determine the nature of the temporal dependence in high-dimensional settings. Therefore, the study of the limiting spectral property of sample autocovariance matrices is very important.

Another way to study the joint convergence of the sample autocovariance matrices is to consider the convergence of the sequence of non-commutative $*$-probability spaces (NCP) generated by them. Moreover, convergence of NCP is closely related to the convergence of the spectral distribution of a matrix. For a $p \times p$ symmetric matrix A_p, the convergence of the NCP generated by A_p with some additional effort, yields the limiting eigenvalue distribution of A_p.

This chapter collects all the basic concepts and results in RMT and non-commutative probability literature that we shall need. These will be crucially used in Chapters 5, 6, 7 and 8 and we shall refer to them frequently.

4.2 LSD and related discussions

One of the primary objects of interest of a matrix is its spectral distribution. The following definition of spectral distributions are for both random and non-random matrices.

Definition 4.2.1. *(ESD, EESD and LSD) The empirical spectral distribution (ESD) of a $p \times p$ (random) matrix R_p is the (random) probability distribution with mass p^{-1} at each of its eigenvalues. If it converges weakly (almost surely) to a (non-degenerate) non-random probability distribution, then the latter will be called the limiting spectral distribution (LSD) of R_p.*

The expectation of ESD is called EESD. This is a non-random probability distribution function.

For a non-random matrix, ESD and EESD are identical.
Consider the following examples on LSD of non-random matrices. We shall use these examples again in Chapter 7.

Example 4.2.1. *Let* $A_p = 0.5I_p$, *where* I_p *is as in (2.8). As all its eigenvalues are* 0.5, *its ESD is degenerate at* 0.5 *and the LSD is also so.*

Example 4.2.2. *Let* $B_p = 0.5(I_p + J_p)$, *where* I_p *and* J_p *are respectively as in (2.8) and (2.9). Note that* B_p *has* $(p-1)$*-many eigenvalues equal to* 0.5 *and one eigenvalue equals* $1 + 0.5(p-1)$. *Therefore, ESD of* B_p, *say* F^{B_p}, *can be written as*

$$F^{B_p}(x) = \begin{cases} 0, & \text{if } -\infty < x < 0.5 \\ 1 - p^{-1}, & \text{if } 0.5 \le x < 1 + 0.5(p-1) \\ 1, & \text{if } 1 + 0.5(1-p) \le x < \infty. \end{cases} \tag{4.1}$$

Hence, LSD of B_p *is degenerate at* 0.5.

Example 4.2.3. *Let* $C_p = ((I(i = j, 1 \le i \le [p/2]) - I(i = j, [p/2]+1 \le i \le p)))$, *where* $[x]$ *denotes the largest integer contained in* x. *Its* $[p/2]$*-many eigenvalues are equal to* 1 *and* $(p - [p/2])$*-many eigenvalues are equal to* -1. *Therefore, the LSD of* C_p *is the distribution* $2Ber(0.5) - 1$, *where* $Ber(0.5)$ *is the Bernoulli variable with success probability* 0.5

Example 4.2.4. *Let* $D_p = ((I(i + j = p + 1)))$. *In this case also, LSD of* D_p *is the distribution* $2Ber(0.5) - 1$.

Incidentally, the study of the limit spectrum of non-hermitian matrices is extremely difficult and very few results are known for general random non-hermitian sequences. The sample autocovariances $\{\hat{\Gamma}_u : u \ge 1\}$ are not hermitian. We shall only consider certain symmetrized version of these matrices.

Two widely used approaches to establish the LSD of symmetric square random matrices are the (i) *moment method* and the (ii) *Stieltjes transformation method.* Below is a brief description of these methods. For more details see Bai and Silverstein [2009]. In this thesis, we shall primarily use the moment method but we shall also use Stieltjes transforms occasionally.

Moment method. The h-th order moment of the ESD of any $p\times p$ real symmetric random matrix R_p equals $\beta_h(R_p) := \frac{1}{p}\text{tr}(R_p^h)$. Consider the following conditions.

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

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

(C) The sequence $\{\beta_h\}$ satisfies Carleman's condition, $\displaystyle\sum_{h=1}^{\infty} \beta_{2h}^{-\frac{1}{2h}} = \infty$.

Then we have the following lemma. We omit its proof. For more details, see for example Bai and Silverstein [2009] and Bose et al. [2010].

Lemma 4.2.1. *If (M1), (M4) and (C) hold, then the ESD of R_p converges almost surely to the distribution F determined uniquely by the moments $\{\beta_h\}$.*

(M1) is the most crucial condition in this method as it identifies the moments of the LSD. Later in Section 4.3, we shall see that the (M1) condition for R_p is ensured by the convergence of an appropriate sequence of NCP generated by R_p. This will be more clear in Section 4.3 after Definition 4.3.4.

The following lemma is relevant to us.

Lemma 4.2.2. *(a) Let $\{\mu_p\}$ be a sequence of probability measures on $\mathbb{R}$. Suppose for all $k \geq 1$ and for some $C > 0$,*

$$\lim_{p\to\infty} \int_{\mathbb{R}} x^k d\mu_p = m_k \text{ and } |m_k| \leq C^k. \tag{4.2}$$

Then $\{m_k\}$ is a moment sequence and there is a unique probability measure μ on $\mathbb{R}$ such that

$$m_k = \int_{\mathbb{R}} x^k d\mu,\ \forall k \geq 1. \tag{4.3}$$

and as $p \to \infty$, $\{\mu_p\}$ converges weakly to μ.

(b) Let A_p be a real symmetric matrix of order p. Suppose for all $k \geq 1$ and for

some $C > 0$,

$$\lim_{p\to\infty} \frac{1}{p} E\,Tr(A_p^k) = m_k \text{ and } |m_k| \leq C^k. \tag{4.4}$$

Then there is a unique probability measure μ *on* $\mathbb{R}$ *such that (4.3) holds.*

Proof. (a) follows as by (4.2), $\{m_k\}$ is a moment sequence and $\sum_{k=1}^{\infty} m_{2k}^{-1/2k} = \infty$. (b) follows from (a) by observing that

$$\frac{1}{p} E\mathrm{Tr}(A_p^k) = \int_{\mathbb{R}} x^k d\mu_p, \ \forall k \geq 1, \tag{4.5}$$

where μ_p is EESD of A_p. □

Stieltjes transformation method. Another widely used method to establish the LSD is the Stieltjes transformation method. The Stieltjes transformation for any random variable X or its probability measure μ on $\mathbb{R}$ equals

$$m_X(z) = m_\mu(z) = \int \frac{1}{x-z}\mu(dx), \qquad z \in \mathbb{C}^+ := \{x+iy : x \in \mathbb{R}, y > 0\}. \tag{4.6}$$

Note that the integral above is always finite for $z \in \mathbb{C}^+$. Pointwise convergence of Stieltjes transforms to a Stieltjes transform implies the convergence of the corresponding distributions. This convergence is usually proved for ESD by linking the Stieltjes transform of the ESD to the resolvent and showing convergence by martingale convergence methods.

Let μ be a compactly supported probability measure on $\mathbb{R}$ with $\mu(-K, K) = 1$ for some $K > 0$. Then we have the following formal power series expansion of the Stieltjes transformation $m(z)$ for $|z| > K$ and $z \in \mathbb{C}^+$,

$$m_\mu(z) = -\frac{1}{z} E_\mu\left(\frac{1}{1-\frac{X}{z}}\right) = -\frac{1}{z} - \frac{E_\mu(X)}{z^2} - \frac{E_\mu(X^2)}{z^3} - \cdots . \tag{4.7}$$

This relation is crucial in linking the moment approach and the Stieltjes transform approach. Since $m_\mu(z)$ is analytic for $z \in \mathbb{C}^+$, in principle it suffices to identify it only for large enough $z \in \mathbb{C}^+$.

The following observation is useful.

Lemma 4.2.3. *Suppose a random variable X has Stieltjes transformation $m_X(z)$. Then for any $\sigma > 0$, the Stieltjes transformation $m_{\sigma X}(z)$ of σX is given by*

$$m_{\sigma X}(z) = \sigma^{-1} m_X(z\sigma^{-1}), \quad \forall z \in \mathbb{C}^+. \tag{4.8}$$

Proof. Let the distribution function of X be F. Then by the definition of Stieltjes transformation given in (4.6), we have

$$m_{\sigma X}(z) = \int \frac{dF(x)}{\sigma x - z} = \int \frac{\sigma^{-1} dF(x)}{x - z\sigma^{-1}} = \sigma^{-1} m_X(z\sigma^{-1}). \tag{4.9}$$

Hence the proof is complete. □

Two specific random matrices play a central role in RMT.

Definition 4.2.2. *(Wigner matrix) For our purposes a Wigner matrix is a square symmetric random matrix with independent mean 0 variance 1 entries on and above the diagonal. We denote a Wigner matrix of order p by W_p.*

Definition 4.2.3. *(Independent matrix) An independent matrix is a rectangular matrix with all independent mean 0 and variance 1 entries. We denote an independent matrix of order $p \times n$ by $Z_{p\times n}$.*

We shall often write W and Z respectively for W_p and $Z_{p\times n}$, if there is no confusion about the dimension of the matrices. As we proceed, further restrictions will be imposed on the entries of these matrices as required.

4.2.1 Existing results on W

Wigner [1958] derived the weak limit of the EESD of $p^{-1/2}W_p$ when the entries are i.i.d. Gaussian. The limiting distribution has p.d.f.

$$f(x) = \begin{cases} \frac{1}{4\pi}\sqrt{4-x^2}, & \text{if } -2 < x < 2 \\ 0, & \text{otherwise} \end{cases} \tag{4.10}$$

with the moment sequence

$$\beta_h = \begin{cases} \frac{h!}{(h/2)!(1+h/2)!}, & \text{if } h \text{ is even} \\ 0, & \text{if } h \text{ is odd.} \end{cases} \tag{4.11}$$

This is known as the standard semi-circle law. Its Stieltjes transformation $m(z)$ satisfies the quadratic equation (only one solution yields a valid Stieltjes transform)

$$m^2(z) + zm(z) + 1 = 0, \quad \forall z \in \mathbb{C}^+. \tag{4.12}$$

It can be shown easily that if $\{\beta_h\}$ satisfies (4.11), then

$$\sum_{h=1}^{\infty} \beta_{2h}^{-1/2h} = \infty. \tag{4.13}$$

We shall need the above facts later in Chapters 5 and 6.

Arnold [1967] and Arnold [1971] showed that the ESD of $p^{-1/2}W_p$ almost surely converges in distribution to the standard semi-circle law, when the entries of W_p are i.i.d. with finite 4-th moment. There has been much subsequent development on the necessary and sufficient conditions for the convergence of the ESD of $p^{-1/2}W_p$. The following theorem (see for example Anderson et al. [2009]) is relevant for us.

Consider the following classes of independent random variables.

$$\mathcal{L}_r \;=\; \text{set of all collections of independent random variables} \tag{4.14}$$

$$\begin{aligned}
& \{\epsilon_{i,j} : i, j \geq 1\} \text{ such that } \sup_{i,j} E|\varepsilon_{i,j}|^r < \infty, \\
\mathcal{L} &= \bigcap_{r=1}^{\infty} \mathcal{L}_r, \qquad (4.15) \\
C(\delta, p) &= \text{set of all collections of random variables } \{\varepsilon_{i,j} : i, j \geq 1\} \text{ such that} \\
& P(|\varepsilon_{i,j}| \leq \eta_p p^{\frac{1}{2+\delta}}) = 1, \quad \forall i, j \text{ and for some } \eta_p \downarrow 0 \text{ as } p \to \infty. \qquad (4.16)
\end{aligned}$$

Theorem 4.2.4. *Let $W_p = ((\omega_{i,j}))$ be a Wigner matrix of order p. Suppose $\{\omega_{i,j} : 1 \leq i, j \leq p\} \in \mathcal{L} \cup C(0, p)$ $\forall p \geq 1$. Then, as $p \to \infty$, the LSD of $p^{-1/2} W_p$ is the standard semi-circle law.*

Next we consider a very specific polynomial in a Wigner and a deterministic matrix considered by Bai and Zhang [2010]. To state their theorem, we need the following class of independent random variables. Let

$$\begin{aligned}
U(\delta) &= \text{set of all collections of independent random variables } \{\varepsilon_{i,j} : i, j \geq 1\} \\
& \text{such that } \lim \frac{\eta^{-(2+\delta)}}{np} \sum_{i=1}^{p} \sum_{j=1}^{n} E(|\varepsilon_{i,j}|^{2+\delta} I(|\varepsilon_{i,j}| > \eta p^{\frac{1}{2+\delta}}) = 0 \\
& \text{for all } \eta > 0. \qquad (4.17)
\end{aligned}$$

By a sequence of nested matrices $\{B_r\}$, we mean that for each $r \geq 1$, the sub-matrix constructed by the first r rows and columns of B_{r+1} is B_r. Consider the following class of matrices:

$$\begin{aligned}
\mathcal{NND} &= \text{set of all sequences of non-negative definite symmetric} \qquad (4.18) \\
& \text{nested matrices } \{B_r\} \text{ whose LSD exists.}
\end{aligned}$$

Theorem 4.2.5. *(Bai and Zhang [2010]) Let $W_p = ((\omega_{i,j}))$ be a Wigner matrix of order p and A_p be a non-random square matrix of order p. Suppose $\{\omega_{i,j} : 1 \leq i, j \leq p\} \in \mathcal{L} \cup C(0, p)$ $\forall p$ or $\{\omega_{i,j} : i, j \geq 1\} \in U(0)$ and $\{A_p\} \in \mathcal{NND}$. Let F^A denote the LSD of A_p. Then, as $p \to \infty$, the ESD of $p^{-1/2} A_p^{1/2} W_p A_p^{1/2}$ converges weakly (almost surely) to a non-random probability distribution whose*

Stieltjes transform $m(z)$ uniquely solves the following system of equations

$$
\begin{aligned}
m(z) &= -z^{-1} - z^{-1}g^2(z), && (4.19)\\
g(z) &= \int \frac{t}{-z - tg(z)} dF^A(t), \quad \forall z \in \mathbb{C}^+. && (4.20)
\end{aligned}
$$

Note that Theorem 4.2.5 assumes weaker conditions on W than Theorem 4.2.4. (4.12) can be derived from (4.19) and (4.20) by putting $A_p = I_p$, where I_p is as in (2.8). Therefore, by Example 4.2.1, F^A is degenerate at 1. Hence, by (4.20)

$$g^2(z) = -(zg(z) + 1). \tag{4.21}$$

Therefore, substituting (4.21) in (4.19), we have

$$zm(z) = -1 + (zg(z) + 1) \implies m(z) = g(z). \tag{4.22}$$

Hence, by (4.21), $m(z)$ satisfies (4.12).

Now we shall consider the existing results on the independent matrix $Z_{p\times n}$. The classical RMT model for $Z_{p\times n}$ assumes $p = p(n) \to \infty$ as $n \to \infty$ and

$$\frac{p}{n} \to y \in [0, \infty). \tag{4.23}$$

The LSD results for the cases $y > 0$ and $y = 0$ are significantly different. Hence we discuss them separately. Note that for the $y > 0$ case, as p and n are comparable, it does not really matter whether we are assuming '$p = p(n) \to \infty$ as $n \to \infty$' or '$n = n(p) \to \infty$ as $p \to \infty$'. But for the case $y = 0$, to be technically correct we shall assume the latter.

4.2.2 Existing results on Z when $p/n \to y > 0$

The well known *Wishart matrix* (without centering) of order p can be written as $S_p = n^{-1}ZZ^*$. Marčenko and Pastur [1967] derived the LSD of S_p when the entries of Z are i.i.d. and have finite fourth moment. For $y \in (0,1]$, the limiting distribution has p.d.f.

$$f_y(x) = \begin{cases} \frac{\sqrt{(b_+(y)-x)(x-b_-(y))}}{2\pi yx}, & \text{if } b_-(y) < x < b_+(y) \\ 0, & \text{otherwise.} \end{cases} \tag{4.24}$$

where $b_\pm(y) = (1 \pm \sqrt{y})^2$. If $y \in (1,\infty)$, the limiting distribution is a mixture of a point mass at 0 and the p.d.f. $f_{1/y}$ with weights $1 - y^{-1}$ and y^{-1}, respectively. It is known as the *Marčenko-Pastur law* with parameter y, say, MP_y, $0 < y < \infty$. This law has the moment sequence

$$\beta_h = \sum_{k=1}^{h} \frac{1}{k} \binom{h-1}{k-1} \binom{h}{k-1} y^{k-1}, \quad \forall h \geq 1. \tag{4.25}$$

Its Stieltjes transformation $m(z)$ satisfies the quadratic equation (only one solution yields a valid Stieltjes transform)

$$yz(m(z))^2 + (y + z - 1)m(z) + 1 = 0, \quad \forall z \in \mathbb{C}^+. \tag{4.26}$$

Over the years several researchers reduced the moment assumptions in Marčenko and Pastur [1967]. For examples, one can consult Wachter [1978] and Yin [1986]. The version of the Marčenko-Pastur law with minimal moment conditions appears to be the following.

Theorem 4.2.6. *(Bai and Silverstein [2009]) Let $Z_{p\times n} = ((z_{i,j}))$ be an independent matrix of order $p \times n$ and let $S_p = n^{-1}ZZ^*$ be the Wishart matrix of order p. Suppose $\{z_{i,j} : i,j \geq 1\} \in U(0)$. Then, as $n, p(n) \to \infty$ and $p/n \to y > 0$, the ESD of S_p almost surely converges in distribution to* MP_y.

The following theorem discusses the convergence of the ESD of $A_p^{1/2}S_pA^{1/2}$, where A_p is as in Theorem 4.2.5. Recall the class of matrices $\mathcal{NND}$ in (4.18).

Theorem 4.2.7. *(Bai and Silverstein [2009]) Let $Z_{p\times n} = ((z_{i,j}))$ be an independent matrix of order $p \times n$ and $S_p = n^{-1}ZZ^*$ be the Wishart matrix of order p. Suppose $\{z_{i,j} : i, j \geq 1\} \in U(0)$. Let $\{A_p\} \in \mathcal{NND}$ with LSD F^A. Then, as $n, p(n) \to \infty$ and $p/n \to y > 0$, the ESD of $A_p^{1/2}S_pA^{1/2}$ almost surely converges in distribution to the probability distribution with Stieltjes transformation $m(z)$ which uniquely solves*

$$m(z) = \int \frac{dF^A(t)}{t(1 - y - yzm(z)) - z}. \tag{4.27}$$

Note that Theorem 4.2.6 is a particular case of Theorem 4.2.7. (4.26) can be derived from (4.27) once we put $A_p = I_p$, where I_p is as in (2.8). Then, by Example 4.2.1, F^A is degenerate at 1. Now, by (4.27) we have

$$\begin{aligned} m(z) &= \frac{1}{1 - y - yzm(z) - z} \Rightarrow m(z)(1 - y - yzm(z) - z) - 1 = 0 \\ \Rightarrow \quad & yzm^2(z) + (y - 1 + z)m(z) + 1 = 0. \end{aligned} \tag{4.28}$$

Hence, (4.26) is established from (4.27).

4.2.3 Existing results on Z when $p/n \to 0$

The LSD results for the case $y = 0$ are quite different from the case $y > 0$. This is because if we put $y = 0$ in the results obtained for $y > 0$, we obtain a degenerate distribution. For example, if we put $y = 0$ in (4.25), β_h will be 1 for all $h \geq 1$. Therefore, LSD of S_p turns out to be degenerate at 1. Hence we need appropriate centering and scaling on S_p. Some of the existing results in this regime are given below. Moreover, as mentioned just before Section 4.2.2, to be technically correct, all the results below assume $p \to \infty$, $n = n(p) \to \infty$ as $p \to \infty$.

Theorem 4.2.8. *(Bai and Yin [1988]) Let $Z_{p\times n}$ be an independent matrix whose* *entries are i.i.d. and have finite fourth order moment. Then, as $p/n \to 0$, the almost sure LSD of $\sqrt{np^{-1}}(n^{-1}ZZ^* - I_p)$ exists and it is distributed as the standard semi-circle variable with pdf (4.10).*

Theorem 4.2.9. *(Bao [2012]) Let $Z_{p\times n}$ be an independent matrix whose entries are i.i.d. and have finite fourth moment. Suppose $\{A_p\} \in \mathcal{NND}$ with LSD F^A. Then as $p/n \to 0$, the almost sure LSD of $\sqrt{np^{-1}}(n^{-1}A^{1/2}ZZ^*A^{1/2} - A)$ exists and its Stieltjes transform $m(z)$ uniquely solves the system of equations (4.19) and (4.20).*

Remark 4.2.10. *Consider all the assumptions in Theorems 4.2.5 and 4.2.9. Then the almost sure LSD of $p^{-1/2}A^{1/2}WA^{1/2}$ and $\sqrt{np^{-1}}(n^{-1}A^{1/2}ZZ^*A^{1/2} - A)$ are identical.*

Recall the notation (2.4). Next, we consider the following class of $r \times r$ deterministic matrices:

$$\begin{aligned} \mathcal{N} \;=\; & \text{set of all sequences of symmetric non-negative definite nested} \quad (4.29)\\ & \text{matrices } \{B_r\} \text{ such that } \sup_r ||B_r||_2 < \infty \text{ and } \lim r^{-1}\mathrm{tr}(B_r^i) \text{ exists}\\ & \text{for } i = 1, 2. \end{aligned}$$

Theorem 4.2.11. *(Wang and Paul [2014]) Let $Z_{p\times n}$ be an independent matrix whose entries are i.i.d. and have finite fourth moment and, A_p and B_n be two deterministic matrices. Suppose $\{A_p\} \in \mathcal{NND}$ with LSD F^A. Suppose $\{B_n\} \in \mathcal{N}$ and $\lim n^{-1}\mathrm{tr}(B_n^2) = d_2$. Then as $p/n \to 0$, the almost sure LSD of $\sqrt{np^{-1}}(n^{-1}A_p^{1/2}ZB_nZ^*A_p^{1/2} - A_pn^{-1}\mathrm{Tr}(B_n))$ exists and its Stieltjes transform $m(z)$ uniquely solves the system of equations*

$$m(z) = -\int \frac{dF^A(t)}{z + d_2tg(z)} \tag{4.30}$$

$$g(z) = -\int \frac{tdF^A(t)}{z + d_2tg(z)}, \ \forall z \in \mathbb{C}^+. \tag{4.31}$$

Note that,

$$zm(z) = -\int \frac{zdF^A(t)}{z + d_2tg(z)} = -1 + d_2g(z)\int \frac{tdF^A(t)}{z + d_2tg(z)} = -1 - d_2g^2(z).$$

Therefore, an equivalent way to write (4.30) is,

$$m(z) = -z^{-1} - z^{-1}d_2g^2(z), \quad \forall z \in \mathbb{C}^+. \tag{4.32}$$

Corollary 4.2.12. *Theorem 4.2.9 follows immediately from Theorem 4.2.11 by putting $B_n = I_n$ and $d_2 = 1$.*

Corollary 4.2.13. *Consider all the assumptions in Theorems 4.2.5 and 4.2.11. Then the almost sure LSD of $\sqrt{d_2}p^{-1/2}A^{1/2}WA^{1/2}$ and $\sqrt{np^{-1}}(n^{-1}A^{1/2}ZBZ^*A^{1/2} - An^{-1}\mathrm{Tr}(B))$ are identical.*

Proof. Suppose LSD of $p^{-1/2}A^{1/2}WA^{1/2}$ is distributed as the random variable X. Therefore, LSD of $p^{-1/2}\sqrt{d_2}A^{1/2}WA^{1/2}$ is distributed as $\sqrt{d_2}X$. By Theorem 4.2.5, the Stieltjes transformation of X is given by

$$\begin{aligned} m_X(z) &= -z^{-1} - z^{-1}g_X^2(z), &(4.33)\\ g_X(z) &= \int \frac{t}{-z - tg_X(z)}dF^A(t), \quad \forall z \in \mathbb{C}^+. &(4.34)\end{aligned}$$

Therefore, by Lemma 4.2.3, the Stieltjes transformation of $\sqrt{d_2}X$ is

$$\begin{aligned} m_{\sqrt{d_2}X}(z) &= d_2^{-1/2}m_X(d_2^{-1/2}z) = -z^{-1} - z^{-1}g_X^2(d_2^{-1/2}z), \text{ where}\\ g_X(d_2^{-1/2}z) &= \int \frac{\sqrt{d_2}t}{-z - t\sqrt{d_2}g_X(d_2^{-1/2}z)}dF^A(t). &(4.35)\end{aligned}$$

Let us define

$$g_{\sqrt{d_2}X}(z) = d_2^{-1/2}g_X(d_2^{-1/2}z).$$

Therefore,

$$g_{\sqrt{d_2}X}(x) = \int \frac{tdF^A(t)}{-z - t\sqrt{d_2}g_X(d_2^{-1/2}z)} = \int \frac{tdF^A(t)}{-z - td_2 g_{\sqrt{d_2}X}(z)}.$$

Hence,

$$\begin{aligned} m_{\sqrt{d_2}X}(z) &= -z^{-1} - z^{-1}d_2 g^2_{\sqrt{d_2}X}(z), \text{ where} \\ g_{\sqrt{d_2}X}(z) &= \int \frac{t}{-z - td_2 g_{\sqrt{d_2}X}(z)} dF^A(t). \end{aligned}$$

This agrees with (4.19) and (4.20). Hence it completes the proof. □

4.3 Free probability and related notions

This section will serve as a brief introduction to *free probability* and related notions that will be used in later chapters. The material of this section is mostly taken from Nica and Speicher [2006], unless otherwise mentioned. As mentioned in Section 4.1, a most natural way to study the joint convergence of several matrices is through the convergence of the non-commutative $*$-probability space (NCP) generated by the matrices. As matrices are non-commutative objects, appearance of non-commutative spaces is not surprising. As we know, commutative (classical) random variables are attached to a probability space $(\mathcal{S}, E)$, which consists of a σ-field $\mathcal{S}$ and an expectation operator E. Similarly, non-commutative variables are attached to an NCP. In the following subsection, we shall define NCP, its convergence and some related notions.

4.3.1 NCP and its convergence

Definition 4.3.1. *A non-commutative $*$-probability space (NCP), $(\mathcal{A}, \varphi)$, consists of a unital $*$-algebra $\mathcal{A}$ over $\mathbb{C}$ and a unital linear functional*

$$\varphi : \mathcal{A} \to \mathbb{C} \quad \textit{such that} \quad \varphi(1_{\mathcal{A}}) = 1.$$

Thus, φ is the analogue of the (classical) expectation operator and is called a *state* of the algebra $\mathcal{A}$. The elements $a \in \mathcal{A}$ are called non-commutative *random variables* in $(\mathcal{A}, \varphi)$. If $a = a^*$, then a is called *self-adjoint.* The state φ is *tracial* if

$$\varphi(ab) = \varphi(ba), \quad \forall a, b \in \mathcal{A}. \tag{4.36}$$

The state φ is *positive* if

$$\varphi(a^* a) \geq 0, \quad \forall a \in \mathcal{A}. \tag{4.37}$$

In this thesis, φ will always be tracial and positive.

The following are some examples of non-commutative $*$-probability spaces.

Example 4.3.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space in the classical sense, i.e. Ω is a non-empty set, $\mathcal{F}$ is a σ-algebra of subsets of Ω and $P : \mathcal{F} \to [0, 1]$ is a probability measure. Let $\mathcal{A} = L^\infty(\Omega, P) =$ set of all measurable functions $a : \Omega \to \mathbb{C}$ such that $P(w : \sup_{\omega \in \Omega} |a(\omega)| < \infty) = 1$ and let φ be defined by*

$$\varphi(a) = \int_\Omega a(\omega) dP(\omega), \quad a \in \mathcal{A}.$$

Then $(\mathcal{A}, \varphi)$ is an NCP.

Example 4.3.2. *Let d be a positive integer. Let $\mathcal{M}_d(\mathbb{C})$ be the algebra of $d \times d$ matrices with complex entries and usual matrix multiplication, and let*

$\mathrm{tr} : \mathcal{M}_d(\mathbb{C}) \to \mathbb{C}$ *be the normalized trace,*

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

Then $(\mathcal{M}_d(\mathbb{C}), \mathrm{tr})$ is an NCP, where the $*$-operation is to take both the transpose of the matrix and the complex conjugate of the entries.

Also let $\mathcal{M}_{(d)}(\mathbb{C})$ be the $*$-algebra of all $d \times d$ random matrices with usual matrix multiplication, then $(\mathcal{M}_{(d)}(\mathbb{C}), E\mathrm{tr})$ forms an NCP.

The following lemma provides two inequalities. Proof of this lemma is trivial and hence we omit it (see Nica and Speicher [2006] for part (a)).

Lemma 4.3.1. *Suppose $(\mathcal{A}, \varphi)$ is an NCP. Let $a, b, a_1, a_2, \ldots, a_k \in \mathcal{A}$ and φ be positive. Then the following results hold.*

(a) $|\varphi(ab)| \leq \sqrt{\varphi(a^*a)\varphi(b^*b)}$.

(b) Moreover, if φ is tracial, then there exists $h_1, h_2, \ldots h_k \geq 1$ such that

$$|\varphi(a_1a_2 \ldots a_k)| \leq \prod_{i=1}^{k} (\varphi(a_i^*a_i)^{h_i})^{1/2h_i}.$$

Since we shall always work with φ which is tracial and positive, Lemma 4.3.1 (a) and (b) shall hold.

Often we deal with $*$-sub-algebras of $\mathcal{M}_d(\mathbb{C})$ and $\mathcal{M}_{(d)}(\mathbb{C})$.

Definition 4.3.2. *($*$-sub-algebra and span) Let $\mathcal{B}$ be a unital $*$-sub-algebra of $\mathcal{A}$. Then $(\mathcal{B}, \varphi)$ also forms an NCP. Let $1_{\mathcal{A}}$ be the identity element of $\mathcal{A}$. Consider $t \geq 1$. Let $\Pi(1_{\mathcal{A}}, a_i, a_i^* : 1 \leq i \leq t) \in \mathcal{A}$ be any polynomial of $\{1_{\mathcal{A}}, a_i, a_i^* : 1 \leq i \leq t\} \subset \mathcal{A}$. Then*

$$\mathit{Span}\{a_i, a_i^* : 1 \leq i \leq t\} = \{\Pi(1_{\mathcal{A}}, a_i, a_i^* : 1 \leq i \leq t) : \ \Pi \textit{ is a polynomial}\}. \quad (4.38)$$

Span$\{a_i, a_i^* : 1 \leq i \leq t\}$ *is called the* $*$*-algebra generated by* $\{a_i, a_i^* : 1 \leq i \leq t\}$. *Equipped with* φ, *it is a unital* $*$*-algebra and is called the NCP generated by* $\{a_i, a_i^* : 1 \leq i \leq t\}$.

By span of a collection of infinitely many non-commutative variables $\{a_i, a_i^* : i \geq 1\}$, *we mean*

$$\text{Span}\{a_i, a_i^* : i \geq 1\} = \{\Pi(1_{\mathcal{A}}, a_{i_k}, a_{i_k}^* : 1 \leq k \leq t) : \ i_k, t \geq 1, \ \Pi \text{ is a polynomial}\}$$

and $(\text{Span}\{a_i, a_i^* : i \geq 1\}, \varphi)$ *again forms an NCP.*

For example, consider a class of $d \times d$ (random) matrices $\{M_i : 1 \leq i \leq r\}$. Then $(\text{Span}\{M_i, M_i^* : 1 \leq i \leq r\}, E\text{tr})$ is an NCP.

The distribution and moments of non-commutative variables are defined as follows.

Definition 4.3.3. *(Distribution and moments) Let* $(\mathcal{A}, \varphi)$ *be an NCP. Let* $\Pi(a, a^*) \in \mathcal{A}$ *be any polynomial in* $a, a^* \in \mathcal{A}$. *Then* $\{\varphi(\Pi(a, a^*)) : \Pi \text{ is a polynomial}\}$ *is called the* $*$*-distribution of* a *or* a^*. *In particular, if* $a \in \mathcal{A}$ *is self-adjont, then* $\{\varphi(a^k)\}_{k=1}^{\infty}$ *is called the distribution of* a.

Consider $t \geq 1$. *Let* $\Pi(a_i, a_i^* : 1 \leq i \leq t) \in \mathcal{A}$ *be any polynomial in* $\{a_i : 1 \leq i \leq t\} \subset \mathcal{A}$. *Then* $\{\varphi\left(\Pi(a_i, a_i^* : 1 \leq i \leq t)\right) : \Pi \text{ is a polynomial}\}$ *is called the joint distribution of* $\{a_i : 1 \leq i \leq t\}$.

For a collection of infinitely many non-commutative variables $\{a_i : i \geq 1\}$, $\{\varphi\left(\Pi(a_{i_k}, a_{i_k}^* : 1 \leq k \leq t)\right) : \quad i_k, t \geq 1, \ \Pi \text{ is a polynomial}\}$ *is its joint distribution. Likewise we can also define the distribution in all the above cases. For simplicity we shall write "distribution" for "*$*$*-distribution".*

Now we shall define convergence of NCP.

Definition 4.3.4. *(Convergence of NCP) Let* $\mathcal{A}_N = \text{Span}\{a_i^{(N)}, a_i^{*(N)} : i \geq 1\}, \ \forall \ N \geq 1$ *and* $\mathcal{A} = \text{Span}\{a_i, a_i^* : i \geq 1\}$. *We say that the sequence of NCP*

$\{(\mathcal{A}_N, \varphi_N)\}_{N=1}^{\infty}$ *converges to* $(\mathcal{A}, \varphi)$ *if for any polynomial* Π *and* $t \geq 1$

$$\lim_{N\to\infty} \varphi_N \left(\Pi(a_i^{(N)}, a_i^{*(N)} : 1 \leq i \leq t)\right) = \varphi\left(\Pi(a_i, a_i^* : 1 \leq i \leq t)\right). \tag{4.39}$$

This is also described as the joint convergence of $\{a_i^{(N)}, a_i^{*(N)} : i \geq 1\}$ *to* $\{a_i, a_i^* : i \geq 1\}$. *If* $\{\varphi_N\}$ *are tracial and positive, then* φ *is also so (see (4.36) and (4.37)). For a fixed* $i \geq 1$, *we say that* a_i^N *converges in distribution to* a_i *if for any polynomial* Π,

$$\lim \varphi_N(\Pi(a_i^{(N)}, a_i^{(N)*})) = \varphi(\Pi(a_i, a_i^*)). \tag{4.40}$$

Remark 4.3.2. *Suppose we are given a unital* $*$*-sub-algebra* $\mathcal{A}_N$ *as above and the left side of (4.39) exists for all polynomial* Π. *Then we can construct a polynomial algebra* $\mathcal{A}$ *of indeterminates* $\{a_i, a_i^*\}$ *which also includes an identity* $1_{\mathcal{A}}$. *We can define* φ *on* $\mathcal{A}$ *by the equation (4.39). Then* $(\mathcal{A}, \varphi)$ *is an NCP and* $(\mathcal{A}_N, \varphi_N) \to (\mathcal{A}, \varphi)$ *in the above sense.*

Let A_p be a square real symmetric random matrix of order p. Note that the convergence of $(\text{Span}\{A_p\}, p^{-1}E\text{Tr})$ guarantees the (M1) condition in the moment method (see Lemma 4.2.1). The following lemma connects LSD and NCP convergence. Proof of this lemma follows immediately from Lemma 4.2.1.

Lemma 4.3.3. *Let* A_p *be a square real symmetric random matrix of order* p. *Suppose* $(\text{Span}\{A_p\}, p^{-1}E\text{Tr}) \to (\text{Span}\{a\}, \varphi)$. *Moreover suppose (M4) holds and* $\{\varphi(a^k)\}$ *satisfies (C). Then as* $p \to \infty$, *the almost sure LSD of* A_p *exists and it is uniquely determined by the moment sequence* $\{\varphi(a^k)\}$.

The following theorem (see for example Anderson et al. [2009]) is relevant in this context and shall be used later in Chapter 5.

Theorem 4.3.4. *Let* W_p *be a Wigner matrix of order* p. *Then under the same* *assumptions as in Theorem 4.2.4,*

$$(Span\{p^{-1/2}W_p\}, E\mathrm{tr}) \to (Span\{s\}, \varphi), \tag{4.41}$$

where s is the standard semi-circle variable with $\varphi(s^h) = \beta_h$, $\forall h$ and $\{\beta_h\}$ is as in (4.11).

Often the limit of random matrices (in the sense of Definition 4.3.4) can be expressed in terms of some *freely independent variables.* Free variables in the non-commutative world, is the analogue of independent random variables in the commutative world. In the next few sections, we develop the basics of free probability.

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

In the non-commutative case, we also have the notion of joint cumulants, called *free cumulants.* These can be uniquely obtained from the moments and vice-versa via a different Möbius transformation and its inverse on the POSET of all *non-crossing partitions.* Non-commutative variables are said to be *free* (freely independent) if and only if all their mixed free cumulants vanish.

All the technicalities and definitions are given in Section 4.3.3. Before that, we need some notions from the theory of partitions, specially non-crossing partitions.

4.3.2 Möbius function, non-crossing partitions and Kreweras complement

We first define the Möbius function which will be useful in the next section to define free cumulants.

Let P be a finite POSET with the partial order $\leq$. We also assume that P is a *lattice*. Let us denote

$$P^{(2)} := \{(\pi, \sigma) | \pi, \sigma \in P, \pi \leq \sigma\}. \tag{4.42}$$

For $F, G : P^{(2)} \to \mathbb{C}$, their *convolution* $F * G$ is the function from $P^{(2)}$ to $\mathbb{C}$ defined by:

$$(F * G)(\pi, \sigma) \quad = \quad \sum_{\substack{\rho \in P \\ \pi \leq \rho \leq \sigma}} F(\pi, \rho) G(\rho, \sigma). \tag{4.43}$$

G is called the inverse of F, if

$$(F * G)(\pi, \sigma) = (G * F)(\pi, \sigma) = I(\pi = \sigma), \ \forall \pi \leq \sigma \in P. \tag{4.44}$$

It is easy to show that any $F : P^{(2)} \to \mathbb{C}$ is *invertible* with respect to the above convolution if and only if $F(\pi, \pi) \neq 0$ for every $\pi \in P$ (see Proposition 10.4 in Nica and Speicher [2006])

The *Zeta function* of P, $\xi : P^{(2)} \to \mathbb{C}$ is defined by

$$\xi(\pi, \sigma) = 1, \ \ \forall \ (\pi, \sigma) \in P^{(2)}. \tag{4.45}$$

The inverse of ξ under the above convolution is called the *Möbius function* of P and is denoted by $\mu(\cdot, \cdot)$. Therefore,

$$(\xi * \mu)(\pi, \sigma) = (\mu * \xi)(\pi, \sigma) = I(\pi = \sigma), \ \forall \pi \leq \sigma \in P. \tag{4.46}$$

Now we shall define non-crossing partitions and the related notions.

Let S be a finite totally ordered set. We call $\pi = \{V_1, V_2, \ldots, V_r\}$ a *partition* of the set S if and only if V_i $(1 \leq i \leq r)$ are pairwise disjoint, non-void subsets of S such that $V_1 \cup V_2 \cup \ldots \cup V_r = S$. We call $V_1, V_2, \ldots, V_r$ the *blocks* of π. Given two elements $p, q \in S$, we write $p \sim_\pi q$ if p and q belong to the same block of π.

A partition π is called a *pair partition* if each block of π contains exactly two elements.

Let $\pi = \{V_1, V_2, \ldots, V_r\}$ and $\sigma = \{U_1, U_2, \ldots, U_k\}$ be two partitions of S. Then we call $\pi \leq \sigma$ if for any fixed $1 \leq i \leq r$ there exists a $1 \leq j \leq k$ such that $V_i \subset U_j$. Therefore, the set of all partitions of S forms a POSET.

A partition π of the set S is called *crossing* if there exists $p_1 < q_1 < p_2 < q_2$ in S such that $p_1 \sim_\pi p_2$ and $q_1 \sim_\pi q_2$ but (p_1, p_2) and (q_1, q_2) are not in the same block. If π is not crossing, then it is called *non-crossing*.

Consider the following sets:

$$\begin{aligned}
NC(n) &= \text{set of all non-crossing partitions of } \{1, 2, 3, \ldots, n\}, && (4.47)\\
NC_2(2n) &= \text{set of all non-crossing pair partitions of } \{1, 2, 3, \ldots, 2n\}, && (4.48)\\
NCE(2n) &= \text{set of } \pi \in NC(2n) \text{ such that every block of } \pi \text{ has even cardinality.} && (4.49)
\end{aligned}$$

Note that the above sets of partitions are all POSET and have smallest and largest elements. For example, in $NC(n)$, the smallest element is $0_n = \{\{1\}, \{2\}, \ldots, \{n\}\}$ and the largest element is $1_n = \{1, 2, \ldots, n\}$. Moreover, $NC_2(2n) \subset NCE(2n) \subset NC(2n)$.

It can be shown that $\{\beta_h\}$ given in (4.11) satisfies

$$\beta_{2h} = \#NC_2(2h), \ \forall h \geq 1. \tag{4.50}$$

Thus the semi-circle law is intimately connected to non-crossing partitions.

Later we shall need the Möbius function on the POSET $NC(n)$. It can be shown that

$$\mu(0_n, 1_n) = (-1)^{n-1} C_{n-1}, \text{ where } C_n = \frac{1}{n+1}\binom{2n}{n}, \ \forall n \geq 1. \tag{4.51}$$

$\mu(\pi, \sigma)$ for other pair $\pi \leq \sigma \in NC(n)$ can be obtained as follows. it is known that the interval $[\pi, \sigma]$ has the *canonical factorization*

$$[\pi, \sigma] \equiv NC(1)^{k_1} \times NC(2)^{k_2} \times NC(n)^{k_n}. \tag{4.52}$$

Let $s_n = \mu(0_n, 1_n), \forall n$. Then it can be proved that

$$\mu(\pi, \sigma) = s_1^{k_1} s_2^{k_2} \ldots s_n^{k_n}. \tag{4.53}$$

For details, see Chapter 10 in Nica and Speicher [2006].

Next we define a useful notion of partition theory, called *Kreweras complement.*

Kreweras complement. The 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 $K(\pi) \in NC(\bar{1}, \bar{2}, \ldots, \bar{n}) \equiv NC(n)$ is defined to be the biggest element among those $\sigma \in NC(\bar{1}, \bar{2}, \ldots, \bar{n})$ which have the property that

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

Let π be a partition of the set S and $A \subset S$. Then by $K_\pi | A$ we mean the restriction of K_π on A. The following properties of Kreweras complement are useful. See Chapter 9 of Nica and Speicher [2006] for details.

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

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

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

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

4.3.3 Free cumulants and free independence

We first present the notion of *free cumulants* and then use it to define *free independence*. Let $(\mathcal{A}, \varphi)$ be an NCP. Define a sequence of *multilinear functionals* $(\varphi_n)_{n\in\mathbb{N}}$ on $\mathcal{A}^n$ via

$$\varphi_n(a_1, a_2, \ldots, a_n) := \varphi(a_1 a_2 \ldots a_n). \tag{4.54}$$

Define recursively a family of 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

$$\varphi_\pi[a_1, a_2, \ldots, a_n] := \varphi(V_1)[a_1, a_2, \ldots, a_n] \cdots \varphi(V_r)[a_1, a_2, \ldots, a_n], \tag{4.55}$$

where $\varphi(V)[a_1, a_2, \ldots, a_n] := \varphi_s(a_{i_1}, a_{i_2}, \ldots, a_{i_s}) = \varphi(a_{i_1} a_{i_2} \ldots a_{i_s})$ for $V = (i_1 < i_2 < \ldots < i_s)$. Observe the use of ( ) and [ ] in (4.54) and (4.55). Define the *joint free cumulant* of order n of $(a_1, a_2, \ldots, a_n)$ as

$$\kappa_n(a_1, a_2, \ldots, a_n) = \sum_{\sigma \in NC(n)} \varphi_\sigma[a_1, a_2, ..., a_n] \mu(\sigma, 1_n), \tag{4.56}$$

where μ is the Möbius function on $NC(n)$. $\kappa_n(a_1, a_2, \ldots, a_n)$ is called a *mixed cumulant* if at least one pair a_i, a_j are different and $a_i \neq a_j^*$. For any $\epsilon_i =$

$1, *$, $\forall 1 \leq i \leq n$, $\kappa_n(a^{\epsilon_1}, a^{\epsilon_2}, \ldots, a^{\epsilon_n})$ is called a *marginal free cumulant* of order n of $\{a, a^*\}$. For a self-adjoint element a, $\kappa_n(a, a, \ldots, a)$ is called the n-th free cumulant of a. We denote it by $\kappa_n(a)$. Note that mixed/marginal free cumulants are all special cases of joint free cumulant.

Just as in (4.55), $\{\kappa_n(a_1, a_2, \ldots, a_n) : n \geq 1\}$ has a multiplicative extension $\{\kappa_\pi : \pi \in NC(n)\}$. It is known that (see Proposition 11.4 in Nica and Speicher [2006])

$$\kappa_\pi[a_1, a_2, ..., a_n] := \sum_{\substack{\sigma \in NC(n) \\ \sigma \leq \pi}} \varphi_\sigma[a_1, a_2, \ldots, a_n]\mu(\sigma, \pi), \tag{4.57}$$

where μ is the Möbius function on $NC(n)$. Note that

$$\begin{aligned} \varphi_{1_n}[a_1, a_2, \ldots, a_n] &= \varphi_n(a_1, a_2, \ldots, a_n) = \varphi(a_1 a_2 \ldots a_n), \\ \kappa_{1_n}[a_1, a_2, \ldots, a_n] &= \kappa_n(a_1, a_2, \ldots, a_n). \end{aligned}$$

Moreover, it can be shown that

$$\varphi(a_1 a_2 \ldots a_n) = \sum_{\pi \leq 1_n} \kappa_\pi[a_1, a_2, \cdots, a_n]. \tag{4.58}$$

Therefore, there is a one-to-one correspondence between free cumulants and moments.

Also, it is known that a self-adjoint variable s is the standard semi-circle if and only if

$$\kappa_n(s) = \begin{cases} 1, & \text{if } n = 2 \\ 0, & \text{if } n \neq 2. \end{cases} \tag{4.59}$$

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

$$C(z) = 1 + \sum_{n=1}^{\infty} \kappa_n(a) z^n \quad \forall z \in \mathbb{C} \text{ and when the series is defined.} \tag{4.60}$$

$C(z)$ satisfies the relation

$$C(-m(z)) = -zm(z), \tag{4.61}$$

where $m(z)$ is the Stieltjes transformation defined in (4.6). If there exist a $C > 0$ such that $|\kappa_n(a)| \leq C^n \ \forall n$, Then (4.60) exists for all $|z| < C^{-1}$. In that case (4.61) makes sense $\forall z \in \mathbb{C}^+$, $|z|$ large, since $m(z) \to 0$ as $|z| \to \infty$.

The free cumulant generating function $C(z)$ and relation (4.61) are useful to derive Stieltjes transformation of a random variable from its free cumulant generating function. We shall use it later.

The following classes of non-commutative variables will be useful in Chapter 6.

Definition 4.3.5. *$\{s_1, s_2, \ldots, s_k\}$ is called a semi-circle family if for each i, s_i is self-adjoint and all their joint free cumulants vanish except for order 2.*

Definition 4.3.6. *$\{c_1, c_2, \ldots, c_k\}$ is called a circular family if all their joint free cumulants vanish except of order 2 and $\kappa_2(c_i, c_j) = \kappa_2(c_i^*, c_j^*) = 0$ for all i, j.*

Now we are ready to define *free independence* of random variables and *free product* of NCP.

Definition 4.3.7. *(Free independence) Let $(\mathcal{A}, \varphi)$ be an NCP. Consider unital $*$-sub-algebras $(\mathcal{A}_i)_{i \in I}$ of $\mathcal{A}$. Then $(\mathcal{A}_i)_{i \in I}$ are freely independent (strictly speaking $*$-free) 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 $\kappa_n(a_1, a_2, \ldots, a_n) = 0$ whenever there exist $1 \leq l, k \leq n$ with $i(l) \neq i(k)$. In short, all mixed cumulants of $(a_1, a_2, \ldots, a_n)$ are zero whenever at least two of them are from different $\mathcal{A}_i$.*

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

Definition 4.3.8. *(Free product) Let $(\mathcal{A}_i, \varphi_i)_{i\in I}$ be a family of NCP. Then there exists an NCP $(\mathcal{A}, \varphi)$, called free product of $(\mathcal{A}_i, \varphi_i)_{i\in I}$, such that $\mathcal{A}_i \subset \mathcal{A}, i \in I$ are freely independent in $(\mathcal{A}, \varphi)$ and $\varphi|\mathcal{A}_i = \varphi_i$.*

A consequence of freeness is that all joint moments of free variables are computable in terms of the moments of the individual variables. Of course, the algorithm for computing moments under freeness is different from (and more complicated than) the product rule under usual independence. In the following Section we shall discuss such an algorithm.

4.3.4 Algorithm to compute moments of free variables

Let s be the standard semi-circle variable. Recall that κ_r denotes the r-th order free cumulant defined in (4.56). Let $\{w_i : 1 \leq i \leq k\}$ be a family of non-commutative variables *which is closed under $*$ operation* and satisfies

$$\kappa_r(w_{l_1}, w_{l_2}, \ldots, w_{l_r}) = 0, \ \forall r \neq 2, \ l_1, l_2, \ldots, l_r \geq 1. \tag{4.62}$$

Moreover, s, $\{w_1, w_2, \ldots, w_k\}$, $\{b_i, b_i^*\}$ and $\{d_i, d_i^*\}$ are free. Later in Chapters 5 and 6, we shall encounter moments of the form

$$\varphi(d_0 s b_1 s d_1 s b_2 s d_2 \ldots s b_n s d_r) \text{ and } \varphi(w_{l_1} b_{l_1} w_{l_2} b_{l_2} w_{l_3} b_{l_3} w_{l_4} b_{l_4} \ldots w_{l_r} b_{l_r}). \tag{4.63}$$

For all $l_1, l_2, \ldots, l_r, r \geq 1$. In this section, we shall discuss an algorithm for computing the expressions in (4.63) in terms of the moments of s, $\{w_1, w_2, \ldots, w_k\}$, $\{b_i, b_i^*\}$ and $\{d_i, d_i^*\}$. The following lemma is useful for this purpose.

Lemma 4.3.5. *Let $(\mathcal{A}, \varphi)$ be an NCP and consider random variables $a_1, a_2, \ldots a_n$, $b_1, b_2, \ldots, b_n \in \mathcal{A}$ such that $\mathrm{Span}\{a_i, a_i^* : 1 \leq i \leq n\}$ and $\mathrm{Span}\{b_i, b_i^* : 1 \leq i \leq n\}$*

are freely independent. Then we have

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

where $K(\pi)$ is the Kreweras Complement of π defined in Section 4.3.2.

The next lemma is useful to compute the first term of (4.63).

Lemma 4.3.6. *Suppose φ is tracial. Then the following holds.*

(a)

$$\begin{aligned}
&\varphi(d_0sb_1sd_1sb_2\dots sd_n)\\
&= \sum_{\pi\in NC_2(2n)} \varphi_{K(\pi)}[b_1,d_1,b_2,d_2,\dots,b_n,d_nd_0] \quad (4.64)\\
&= \sum_{\pi\in NC(n)} \varphi_\pi[b_1,b_2,\dots,b_n]\varphi_{K(\pi)}[d_1,d_2,\dots,d_nd_0] \quad (4.65)\\
&= \sum_{\pi\in NC(n)} \varphi_\pi[d_1,d_2,\dots,d_nd_0]\varphi_{K(\pi)}[b_1,b_2,\dots,b_n]. \quad (4.66)
\end{aligned}$$

(b) Fix $1 = k_0 < k_1 < \dots < k_t \le n$ and let $\mathcal{S} \subset NC_2(2n)$ be given by

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

Then

$$\begin{aligned}
&\sum_{\pi\in\mathcal{S}} \varphi_{K(\pi)}[b_1,d_1,b_2,d_2,\dots,b_n,d_nd_0] \quad (4.67)\\
&= \varphi\Big(\prod_{s=0}^{t} b_{k_s}\Big)\prod_{s=1}^{t+1}\varphi(d_{k_{s-1}}sb_{k_{s-1}+1}sd_{k_{s-1}+1}\dots sd_{k_s-1}),
\end{aligned}$$

where $k_0 = 1, d_{k_{t+1}-1} = d_nd_0$.

Proof. Relation (4.64) follows from Lemma 4.3.5. By freeness of $\{b_i\}$ and $\{d_i\}$, and by Properties $1-3$ of $K(\pi)$ in Section 4.3.2, (4.65) and (4.66) follow from

(4.64). We now prove (4.67).

Consider the following subsets of $\{1, 2, \ldots, 2n\}$ as

$$\begin{aligned} \mathcal{S}_0 &= \{2k_i - 1, 2k_i : 1 \leq i \leq t\}, \\ \mathcal{S}_i &= \{2k_{i-1} + 1, 2k_{i-1} + 2, \ldots, 2k_i - 2\},\ 1 \leq i \leq t+1,\ k_{t+1} = n+1. \end{aligned}$$

By $NC(\mathcal{S}_i)$ and $NC_2(\mathcal{S}_i)$, respectively, we mean the sets of all non-crossing and non-crossing pair partitions of indices in $\mathcal{S}_i$. Let

$$\sigma_0 = \{\{2k_i, 2k_{i+1} - 1\} : 0 \leq i \leq t,\ k_{t+1} = 1\} \in NC_2(\mathcal{S}_0). \tag{4.68}$$

Note that, as $\mathcal{S}$ contains only non-crossing partitions, we have

$$\mathcal{S} = \{\sigma_0 \cup \sigma_1 \cup \ldots \cup \sigma_{t+1} : \sigma_i \in NC_2(\mathcal{S}_i),\ \forall 1 \leq i \leq (t+1)\}. \tag{4.69}$$

Now to understand the nature of the Kreweras complement $K(\pi)$ for $\pi \in \mathcal{S}$, consider the following subsets of $\{1, 2, \ldots, 2n\}$. For all $1 \leq i \leq t+1$,

$$\begin{aligned} W_i &= \{2k_{i-1}, 2k_{i-1} + 1, \ldots, 2k_i - 2\}, \\ W_i^- &= \{2k_{i-1} + 1, 2k_{i-1} + 2, \ldots, 2k_i - 3\}, \\ 1_{W_i} &= \{2k_{i-1}, 2k_{i-1} + 1, \ldots, 2k_i - 2\}, \\ 0_{W_i^-} &= \{2k_{i-1} + 1\}, \{2k_{i-1} + 2\}, \ldots, \{2k_i - 3\}. \end{aligned}$$

Since the Kreweras complement $K(\pi)$ is non-crossing, it will be of the following form,

$$K(\pi) = \tau_0 \cup \tau_1(\sigma_1) \cup \tau_2(\sigma_2) \cup \ldots \cup \tau_{t+1}(\sigma_{t+1}), \tag{4.70}$$

where the blocks $\{\tau_0, \tau_i(\sigma_i)\}$ satisfy

$$\begin{aligned}
\tau_0 &= \{2k_i - 1 : 0 \leq i \leq t\} \\
\{\{2k_{i-1}, 2k_i - 2\}, 0_{W_i^-}\} &\leq \tau_i(\sigma_i) \leq 1_{W_i},\ \forall 1 \leq i \leq t+1.
\end{aligned}$$

Note that $\tau_i(\sigma_i)$ depends only on $\sigma_i \in NC_2(\mathcal{S}_i)$ but no other σ_k, $k \neq i$. Hence, by multiplicative property (4.55) of φ,

$$\begin{aligned}
&\sum_{\pi \in \mathcal{S}} \varphi_{K(\pi)}[b_1, d_1, b_2, d_2, \ldots, b_n, d_n d_0] \qquad (4.71)\\
&= \varphi(\prod_{i=0}^{t} b_{k_s})\Bigg(\prod_{i=1}^{t+1} \sum_{\sigma_i \in NC_2(\mathcal{S}_i)} \varphi_{\tau_i(\sigma_i)}(b_{k_{i-1}+1}, d_{k_{i-1}+1}, \ldots, b_{k_i - 1}, d_{k_i - 1} d_{k_{i-1}})\Bigg).
\end{aligned}$$

Now, note that the set of blocks $\mathcal{G}_i = \{\tau_i(\sigma_i) : \sigma_i \in NC_2(\mathcal{S}_i)\}$ has one-to-one correspondence with the set of Kreweras complements

$$\{K(\pi) :\ \pi \in NC_2(2k_i - 2k_{i-1} - 2)\}.$$

This one-to-one correspondence is obvious when one sets $2k_{i-1} + j = j$, $\forall 1 \leq j \leq 2k_i - 2k_{i-1} - 3$, and $\{2k_{i-1}, 2k_i - 2\} = 2k_i - 2k_{i-1} - 2$.

Let $2k_i - 2k_{i-1} - 2 = \omega_i$ (say). Hence, by (4.71), we have

$$\begin{aligned}
&\sum_{\pi \in \mathcal{S}} \varphi_{K(\pi)}[b_1, d_1, b_2, d_2, \ldots, b_n, d_n d_0] \qquad (4.72)\\
&= \varphi(\prod_{i=0}^{t} b_{k_s})\Bigg(\prod_{i=1}^{t+1} \sum_{\sigma_i \in NC_2(\omega_i)} \varphi(b_{k_{i-1}+1}, d_{k_{i-1}+1}, \ldots, b_{k_i - 1}, d_{k_i - 1} d_{k_{i-1}})\Bigg) \\
&= \varphi(\prod_{i=0}^{t} b_{k_s})\Bigg(\prod_{i=1}^{t+1} \varphi(d_{k_{i-1}} s b_{k_{i-1}+1} s d_{k_{i-1}+1} \ldots s d_{k_i - 1}\Bigg).
\end{aligned}$$

Hence (4.67) is justified. □

Next we shall see how to compute the second term of (4.63). By Lemma 4.3.5, we have for all $m \geq 1$,

$$\varphi(w_{l_1}b_{l_1}w_{l_2}b_{l_2}\dots w_{l_r}b_{l_r}) = \begin{cases} 0, \text{ if } r = 2m-1 \\ \displaystyle\sum_{\pi \in NC_2(2m)} \varphi_{K(\pi)}[b_{l_1}, \dots, b_{l_{2m}}]\kappa_\pi[w_{l_1}, \dots, w_{l_{2m}}], \\ \qquad\qquad \text{if } r = 2m. \end{cases} \tag{4.73}$$

Note that $\pi \in NC_2(2m)$ can be expressed as $\{(t_1, t_2), (t_3, t_4), \dots, (t_{2m-1}, t_{2m})\}$, where $t_1 < t_2, t_3 < t_4, \dots, t_{2m-1} < t_{2m}$. Hence, by (4.73), we have

$$\varphi(w_{l_1}b_{l_1}w_{l_2}b_{l_2}\dots w_{l_r}b_{l_r}) = \begin{cases} 0, \text{ if } r = 2m-1, \\ \displaystyle\sum_{\pi=\{(t_1,t_2),\dots,(t_{2m-1},t_{2m})\}} \Big(\prod_{i=1}^m \kappa_2(w_{l_{t_{2i-1}}}, w_{t_{l_{2i}}})\Big) \\ \qquad \varphi_{K(\pi)}[b_{l_1}, b_{l_2}, \dots, b_{l_k}], \qquad \text{if } r = 2m. \end{cases} \tag{4.74}$$

The last equality holds due to multilinear property of κ_π.

4.3.5 Some existing results on the joint convergence of random matrices

The following theorem states the joint convergence of several Wigner and deterministic matrices. See for example Anderson et al. [2009].

Theorem 4.3.7. *Let $W_p^{(1)}, W_p^{(2)}, \dots, W_p^{(r)}$ be r independent Wigner matrices of order p such that each matrix individually satisfies the assumptions of Theorem 4.2.4. Let $D_p^{(1)}, D_p^{(2)}, \dots, D_p^{(2q)}$ be $2q$ constant matrices of order p with bounded norm such that, for $\epsilon = 0, 1$, ($Span\{D_p^{(2i-\epsilon)}, D_p^{(2i-\epsilon)*} : 1 \leq i \leq q\}, p^{-1}\mathrm{Tr}$) converges. Then the following statements hold. As $p \to \infty$,*

(a) $p^{-1/2}W_p^{(1)}, p^{-1/2}W_p^{(2)}, \dots, p^{-1/2}W_p^{(r)}$ are asymptotically free.

(b) For $\epsilon = 0$ *or* 1*, the collections* $\{p^{-1/2}W_p^{(i)}\}$ *and* $\{D_p^{(2i-\epsilon)}, D_p^{(2i-\epsilon)*}\}$ *are asymptotically free.*

(c) The collections $\{p^{-1}W_p^{(i)}D_p^{(2j)}W_p^{(i)}, p^{-1}W_p^{(i)}D_p^{(2j)*}W_p^{(i)} : i, j \geq 1\}$ *and* $\{D_p^{(2i-1)}, D_p^{(2i-1)*}\}$ *are asymptotically free.*

(d) Let $\epsilon_i = 1, *, \forall 1 \leq i \leq 2k$*. To compute* $\lim p^{-1}E\mathrm{Tr}(p^{-k}\prod_{i=1}^{K} D_p^{(2i-1)\epsilon_{2i-1}}W_p^{(i)}D_p^{(2i)\epsilon_{2i}}W_p^{(i)})$ *one can assume that the collections* $\{W_p^{(i)}\}$*,* $\{D_p^{(2i-1)}, D_p^{(2i-1)*}\}$ *and* $\{D_p^{(2i)}, D_p^{(2i)*}\}$ *are asymptotically free.*

Proof. For (a) and (b) see Anderson et al. [2009]. (c) follows from (a), (b) and Theorem 11.12, Page 180 of Nica and Speicher [2006]. (d) is immediate from (a), (b) and (c). □

Next we define the *compound free Poisson distribution.* This will be useful in Chapters 5 and 6 to describe the LSD of $\{\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)\}_{u \geq 0}$ when the coefficient matrices $\psi_j = \lambda_j I_p$, for all $j \geq 0$ and I_p is as in (2.8). For any measure μ, let $m_n(\mu) = \int x^n d\mu(x),\ n \geq 1$ be its moments (assumed to be finite). Let $\{k_n(\mu)\}$ denote the corresponding free cumulants.

Definition 4.3.9. *(Compound free Poisson distribution) A probability measure* μ *on* $\mathbb{R}$ *with free cumulants*

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

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

Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. Let $s, a \in \mathcal{A}$ be such that s is a semi-circle variable with moment sequence (4.11), and moreover s and a are free. Then the free cumulants of sas are given by

$$k_n(sas, sas, \ldots, sas) = \varphi(a^n)\ \forall n \geq 1. \tag{4.75}$$

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

Recall the classes of independent random variables $\mathcal{L}$ and C respectively from (4.14) and (4.16). It can be shown that the limiting free cumulants of $n^{-1}ZAZ^*$, for any $p \times n$ independent matrix $Z = ((z_{i,j}))_{p\times n}$ such that $\{z_{i,j} : 1 \leq i \leq p,\ 1 \leq j \leq n\} \in \mathcal{L} \cup C(\delta, p),\ \forall p \geq 1$ and for some $\delta > 0$ and for any self-adjoint matrix A of order n with compactly supported LSD a, is given by

$$\lim_n k_r(n^{-1}ZAZ^*, n^{-1}ZAZ^*, \ldots, n^{-1}ZAZ^*) = y^{r-1}\varphi(a^r),\ \forall r \geq 1. \tag{4.76}$$

Here $p = p(n)$ is such that $pn^{-1} \to y \in (0, \infty)$. Therefore, LSD of $n^{-1}ZAZ^*$ is the compound free Poisson distribution with rate y^{-1} and jump distribution ya.

Chapter 5

Joint convergence of generalized dispersion matrices when $p/n \to y > 0$

5.1 Introduction

We can now put to use the machinery developed in Chapter 4 to study the LSD of any symmetric polynomial in the sample autocovariance matrices $\{\hat{\Gamma}_u\}$, along with their joint convergence.

For that we first need to express these matrices in a suitable form. Recall the independent matrix Z in Definition 4.2.3 and the sequence of coefficient matrices $\{\psi_j\}$ in (3.2). Let $\{P_j : j = 0, \pm 1, \pm 2, \ldots\}$ be a sequence of $n \times n$ matrices where P_j has entries equal to one on the j-th upper diagonal and 0 otherwise. Note that $P_0 = I_n$ where I_n is the $n \times n$ identity matrix, and $P_j = P^*_{-j}$, $\forall j$. Define

$$\Delta_u = \frac{1}{n} \sum_{j,j'=0}^{q} \psi_j Z P_{j-j'+u} Z^* \psi^*_{j'}, \quad \forall u = 0, 1, 2, \ldots. \tag{5.1}$$

In Chapter 7, we shall prove that $\{\Delta_u\}$ approximates $\{\hat{\Gamma}_u\}$ as far as the LSD and joint convergence are concerned. With this in mind, first we study the matrices $\{\Delta_u\}$ in this chapter and Chapter 6 respectively for the cases $p/n \to y > 0$ and $p/n \to 0$.

Indeed we broaden our scope significantly and deal with a more general set up *where we have*

1. *more than one independent matrices,*

2. *any $n \times n$ matrices between Z and Z^* instead of typical matrices $\{P_j\}$ and*

3. *polynomials which contain several (possibly independent) (Z, Z^*) pairs.*

Suppose we have matrices $Z_u = ((\varepsilon_{u,t,i}))_{p\times n},\ 1 \leq u \leq U$, where $\{\varepsilon_{u,t,i} : u, i, j \geq 0\}$ are independent with mean 0 and variance 1. Note that each Z_u is an independent matrix and moreover, they are independent among themselves.

Also suppose $\{B_{2i-1} : 1 \leq i \leq K\}$ and $\{B_{2i} : 1 \leq i \leq L\}$ are constant matrices of order $p \times p$ and $n \times n$ respectively.

Consider all $p \times p$ matrices

$$\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} = \prod_{i=1}^{k_l} \left(n^{-1} A_{l,2i-1} Z_{u_{l,i}} A_{l,2i} Z^*_{u_{l,i}} \right) A_{l,2k_l+1}, \tag{5.2}$$

where $\{A_{l,2i-1}\}$, $\{A_{l,2i}\}$ and $\{Z_{u_{l,i}}\}$ are matrices from the collections $\{B_{2i-1}, B^*_{2i-1} : 1 \leq i \leq K\}$, $\{B_{2i}, B^*_{2i} : 1 \leq i \leq L\}$ and $\{Z_i : 1 \leq i \leq U\}$ respectively. As the sample variance-covariance matrix (without centering) is a special case of the above matrices, we call them *generalized dispersion matrices.*

Consider the sequence of NCP ($\mathcal{U}_p$, $p^{-1}E\mathrm{Tr}$), where

$$\mathcal{U}_p = \mathrm{Span}\left(\mathbb{P}_{l,(u_{l,1},\ldots,u_{l,k_l})} :\ l, k_l \geq 1 \right). \tag{5.3}$$

Note that $\mathcal{U}_p$ forms a $*$-algebra. Here we are interested in the convergence of $(\mathcal{U}_p, p^{-1}E\mathrm{Tr})$.

As $\{Z_u\}$, $\{B_{2i-1}\}$ and $\{B_{2i}\}$ are all of different orders, it is not possible directly to define an algebra of these matrices. Therefore, it becomes difficult to describe the limit of $(\mathcal{U}_p, p^{-1}E\mathrm{Tr})$ directly in terms of the limits of $\{Z_u\}$, $\{B_{2i-1}\}$ and $\{B_{2i}\}$. The solution is to embed all these matrices into matrices of order $(n + p)$.

Recall the Wigner matrix in Definition 4.2.2. We first embed Z_u into a Wigner matrix W_u of order $(n+p)$. Thus

$$W_u = \begin{pmatrix} W^{(1u)}_{p\times p} & Z_u \\ Z_u^* & W^{(2u)}_{n\times n} \end{pmatrix}, \tag{5.4}$$

where $\{W^{(iu)} : i = 1, 2, \ u \geq 1\}$ are independent Wigner matrices and are independent of $\{Z_u\}$.

For any matrices B and D of order p and n respectively, let $\bar{B}$ and $\underline{D}$ of order $(n+p)$ be the matrices

$$\bar{B} = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}, \quad \underline{D} = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix}. \tag{5.5}$$

It is easy to see that

$$\bar{\mathbb{P}}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} = \prod_{i=1}^{k_l} \left(n^{-1} \bar{A}_{l,2i-1} W_{u_{l,i}} \underline{A}_{l,2i} W^*_{u_{l,i}} \right) \bar{A}_{l,2k_l+1}. \tag{5.6}$$

Note that the right side of (5.6) is a polynomial in Wigner and deterministic matrices.

Consider the sequence of NCP $(\bar{\mathcal{U}}_p,\ (n+p)^{-1}E\text{Tr})$, where

$$\bar{\mathcal{U}}_p = \text{Span}\left(\bar{\mathbb{P}}_{l,(u_{l,1},\ldots,u_{l,k_l})} :\ l, k_l \geq 1 \right). \tag{5.7}$$

Note that $\bar{\mathcal{U}}_p$ also forms a $*$-algebra. Convergence of $(\bar{\mathcal{U}}_p,\ (n+p)^{-1}E\text{Tr})$ is easy to describe by using Theorem 4.3.7. Then we express the limit of $(\mathcal{U}_p, p^{-1}E\text{Tr})$ in terms of the limit of $(\bar{\mathcal{U}}_p,\ (n+p)^{-1}E\text{Tr})$.

In Section 5.2.2, we provide the idea behind the limit. Then in Theorem 5.3.1 we state the result on convergence of $(\bar{\mathcal{U}}_p,\ (n+p)^{-1}E\text{Tr})$ and $(\mathcal{U}_p,\ p^{-1}E\text{Tr})$. The limiting NCP can be expressed in terms of some free variables.

As discussed in Lemma 4.3.3 of Chapter 4, NCP convergence with some additional effort guarantees existence of the LSD. Theorem 5.4.1 states that the LSD of any symmetric polynomial in $\{\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})}\}$ exists and the limit can be expressed in terms of some freely independent variables.

As we have seen in Chapter 4, most of the existing LSD results are obtained using the Stieltjes transformation method. It is not clear how this method could be used for any arbitrary symmetric polynomial. We fall back on the moment method to derive LSD of such polynomials. To link these two methods, in Theorem 5.4.5, we derive the Stieltjes transformation of the LSD of some specific matrices. In Section 5.4.2, we show how the existing LSD results in the literature follow as special cases of our LSD results. *The main material of this chapter is taken from Bhattacharjee and Bose [2015a].*

5.2 Preliminaries

5.2.1 Assumptions

We first list all the assumptions that are required.

First, let us describe our assumption on $\{Z_u\}$. Recall the independent matrix in Definition 4.2.3. Let $Z_u = ((\varepsilon_{u,i,j}))_{p\times n},\ 1 \leq u \leq U$ be $p \times n$ independent matrices. Therefore, $\{\varepsilon_{u,i,j} : u, i, j \geq 1\}$ are independently distributed with $E(\varepsilon_{u,i,j}) = 0$, $E|\varepsilon_{u,i,j}|^2 = 1$. Recall the classes $\mathcal{L}$ and C respectively in (4.14) and (4.16). We assume that

(A1) $\{\varepsilon_{u,i,j} : 1 \leq i \leq p,\ 1 \leq j \leq n\} \in \mathcal{L} \cup C(\delta, p)\ \forall p \geq 1$ for some $\delta \in (0, 2]$ and for all $1 \leq u \leq U$.

If there is only one u i.e., if $U = 1$, we will write $\varepsilon_{i,j}$ and Z respectively for $\varepsilon_{1,i,j}$ and Z_1. Assumption $(A1)$ will be weakened later for some corollaries and applications by means of truncation techniques.

Now we move to the assumptions on the deterministic matrices $\{B_i\}$.

(A2) $\{B_{2i-1} : 1 \leq i \leq K\}$ are norm bounded $p \times p$ matrices and $(\text{Span}(B_{2i-1}, B_{2i-1}^* : 1 \leq i \leq K), p^{-1}\text{Tr})$ converges.

(A3) $\{B_{2i} : 1 \leq i \leq L\}$ are norm bounded $n \times n$ matrices and $(\text{Span}(B_{2i}, B_{2i}^* : 1 \leq i \leq L), n^{-1}\text{Tr})$ converges.

Note that we do not assume the joint convergence of $\{B_i : i \geq 1\}$.

5.2.2 Idea behind the limit of $(\mathcal{U}_p, p^{-1}E\mathbf{Tr})$

To see how freeness comes into the picture and hence how it motivates the limiting NCP of $(\mathcal{U}_p, p^{-1}E\text{Tr})$, let us focus on a particular element of $\mathcal{U}_p$:

$$P = n^{-1}(B_1Z_1B_2Z_1^*B_3 + B_5Z_1B_4Z_1^*B_7 + B_9Z_1B_6Z_1^*B_{11} + B_{13}Z_1B_8Z_1^*B_{15}).$$

In this section, we consider appropriate conditions on $\{B_{4i-j} : 1 \leq i \leq 4, j = 1, 3\}$ and $\{B_{2i} : 1 \leq i \leq 4\}$ so that P is self-adjoint. For illustration we consider only a self-adjoint polynomial. Similar idea works for non-self-adjoint polynomials also. Our primary goal is to show that for all $r \geq 1$, $\lim p^{-1}E\text{Tr}(P^r)$ exists.

We embed Z_1 into the Wigner matrix W_1 as given in (5.4). For any matrices B and D of order p and n respectively, recall the matrices $\bar{B}$ and $\underline{D}$ in (5.5).

Note that for any integer r, if the right/left side limits below exist, then

$$\lim(n+p)^{-1}\text{Tr}(\bar{B}^r) = y(1+y)^{-1}\lim p^{-1}\text{Tr}(B^r), \tag{5.8}$$

$$\lim(n+p)^{-1}\text{Tr}(\underline{D}^r) = (1+y)^{-1}\lim n^{-1}\text{Tr}(D^r), \text{ and} \tag{5.9}$$

$$\lim p^{-1}\text{Tr}(P^r) = y^{-1}(1+y)\lim(n+p)^{-1}\text{Tr}(\bar{P}^r). \tag{5.10}$$

Thus to deal with P, we consider the corresponding matrix $\bar{P}$. Note that

$$n\bar{P} = \bar{B}_1W_1\underline{B}_2W_1\bar{B}_3 + \bar{B}_5W_1\underline{B}_4W_1\bar{B}_7 + \bar{B}_9W_1\underline{B}_6W_1\bar{B}_{11} + \bar{B}_{13}W_1\underline{B}_8W_1\bar{B}_{15}.$$

Thus $\bar{P}^r$ involves polynomials in these enlarged matrices. So it is a question of computing the limiting trace of such polynomials. Note that by (A2) and (A3), for any monomial m,

(1) $\lim(n+p)^{-1}\mathrm{Tr}(m(\{\bar{B}_{2i-1}, \bar{B}^*_{2i-1} : 1 \leq i \leq 8\}))$ and $\lim(n+p)^{-1}\mathrm{Tr}(m(\{\underline{B}_{2i}, \underline{B}^*_{2i} : 1 \leq i \leq 4\}))$ exist.

Therefore, as $p(n), n \to \infty$, $(\mathrm{Span}\{\bar{B}_{2i-1}, \bar{B}^*_{2i-1} : 1 \leq i \leq 8\}, (n+p)^{-1}E\mathrm{Tr})$ and $(\mathrm{Span}\{\underline{B}_{2i}, \underline{B}^*_{2i} : 1 \leq i \leq 4\}, (n+p)^{-1}E\mathrm{Tr})$ converge respectively to $(\mathrm{Span}\{\bar{b}_{2i-1}, \bar{b}^*_{2i-1} : 1 \leq i \leq 8\}, \varphi_1)$ and $(\mathrm{Span}\{\underline{b}_{2i}, \underline{b}^*_{2i} : 1 \leq i \leq 4\}, \varphi_2)$, say.

(2) Recall the class of independent variables $\mathcal{L}$ and $\mathcal{C}$ respectively in (4.14) and (4.16). By Theorem 4.3.4, if $\{\varepsilon_{1,i,j}\} \in \mathcal{L} \cap C(0,p)$, then $\lim E(n+p)^{-1}\mathrm{Tr}((n+p)^{-1/2}W_1)^r = E(s^r)$ i.e. $(\mathrm{Span}\{(n+p)^{-1/2}W_1\}, (n+p)^{-1}E\mathrm{Tr})$ converges to $(\mathrm{Span}\{s\}, \varphi_3)$, say, where s is a standard semi-circle variable with the moment sequence (4.11).

(3) Finally, by Theorem 4.3.7 (d) and for the polynomial $\bar{P}$, *in the limit*, the matrices $(n+p)^{-1/2}W_1$, $\{\bar{B}_{2i-1}, \bar{B}^*_{2i-1} : 1 \leq i \leq 8\}$ and $\{\underline{B}_{2i}, \underline{B}^*_{2i} : 1 \leq i \leq 4\}$ are free variables.

Recall the free product of NCP in Definition 4.3.8. By (3), $s, \{\bar{b}_{2i-1}, \bar{b}^*_{2i-1} : 1 \leq i \leq 8\}$ and $\{\underline{b}_{2i}, \underline{b}^*_{2i} : 1 \leq i \leq 4\}$ are free in some NCP $(\mathcal{A}, \varphi)$.

Thus using the above observations (1)-(3) in conjunction with equations (5.8)-(5.10), we can conclude that $\lim(n+p)^{-1}\mathrm{Tr}(\bar{P}^r)$ and $\lim p^{-1}\mathrm{Tr}(P^r)$ exist and

$$\begin{aligned}
\lim \frac{1}{n+p}\mathrm{Tr}(\bar{P}^r) &= \lim \left(\frac{n+p}{n}\right)^r \left(\frac{1}{n+p}\mathrm{Tr}(\sum_{i=1}^4 \bar{B}_{4i-3}\frac{W_1}{\sqrt{n+p}}\underline{B}_{2i}\frac{W_1}{\sqrt{n+p}}\bar{B}_{4i-1})^r\right) \\
&= \varphi\left((1+y)\sum_{i=1}^4 \bar{b}_{4i-3}s\underline{b}_{2i}s\bar{b}_{4i-1}\right)^r \quad \text{and} \qquad (5.11) \\
\lim p^{-1}\mathrm{Tr}(P^r) &= y^{-1}(1+y)\varphi\left((1+y)\sum_{i=1}^4 \bar{b}_{4i-3}s\underline{b}_{2i}s\bar{b}_{4i-1}\right)^r. \qquad (5.12)
\end{aligned}$$

The right side of (5.12), involving free variables, are then the limit moments of P and can be computed using Lemma 4.3.6.

This is the idea we implement for the general matrices (5.2). To embed $\{Z_u\}$, we need independent Wigner matrices $\{W_u\}$ and by Theorem 4.3.7 (a), these yield U many free semi-circle variables. The limits can then be expressed in terms of polynomials in the free semi-circle variables, and the limits of $(\text{Span}\{\bar{B}_{2i-1}, \bar{B}^*_{2i-1} : 1 \leq i \leq 8\}, (n+p)^{-1}E\text{Tr})$ and $(\text{Span}\{\underline{B}_{2i}, \underline{B}^*_{2i}\}, (n+p)^{-1}E\text{Tr})$, where the two limit collections are free.

5.3 NCP convergence result

Now we rigorously state our result for the joint convergence of the general matrices (5.2). To describe the limits, consider $(\mathcal{S} = \text{Span}\{s_u : u \geq 1\}, \varphi_s)$ to be an NCP of free standard semi-circular variables $\{s_u\}$. Suppose that

$$(\text{Span}\{B_{2i-1}, B^*_{2i-1} : 1 \leq i \leq K\}, \frac{1}{p}\text{Tr}) \;\to\; (\mathcal{A}_{\text{odd}} = \text{Span}\{b_{2i-1}, b^*_{2i-1} : 1 \leq i \leq K\}, \varphi_{\text{odd}}), \tag{5.13}$$

$$(\text{Span}\{\bar{B}_{2i-1}, \bar{B}^*_{2i-1} : 1 \leq i \leq K\}, \frac{1}{n+p}\text{Tr}) \;\to\; (\bar{\mathcal{A}}_{\text{odd}} = \text{Span}\{\bar{b}_{2i-1}, \bar{b}^*_{2i-1} : 1 \leq i \leq K\}, \bar{\varphi}_{\text{odd}}), \tag{5.14}$$

$$(\text{Span}\{B_{2i}, B^*_{2i} : 1 \leq i \leq L\}, \frac{1}{n}\text{Tr}) \;\to\; (\mathcal{A}_{\text{even}} = \text{Span}\{b_{2i}, b^*_{2i} : 1 \leq i \leq L\}, \varphi_{\text{even}}). \tag{5.15}$$

$$(\text{Span}\{\underline{B}_{2i}, \underline{B}^*_{2i} : 1 \leq i \leq L\}, (n+p)^{-1}\text{Tr}) \;\to\; (\bar{\mathcal{A}}_{\text{even}} = \text{Span}\{\underline{b}_{2i}, \underline{b}^*_{2i} : 1 \leq i \leq L\}, \bar{\varphi}_{\text{even}}). \tag{5.16}$$

Therefore, for any polynomial Π,

$$\varphi_{\text{odd}}(\Pi(b_{2i-1}, b^*_{2i-1} : 1 \leq i \leq K)) = y^{-1}(1+y)\bar{\varphi}_{\text{odd}}(\Pi(\bar{b}_{2i-1}, \bar{b}^*_{2i-1} : 1 \leq i \leq K)), \tag{5.17}$$

$$\varphi_{\text{even}}(\Pi(b_{2i}, b_{2i}^* : 1 \leq i \leq L)) \;=\; (1+y)\bar{\varphi}_{\text{even}}(\Pi(\underline{b}_{2i}, \underline{b}_{2i}^* : 1 \leq i \leq L)). \quad (5.18)$$

Recall the free product of NCP in Definition 4.3.8. Let

$$(\mathcal{A}, \bar{\varphi}) \;=\; \text{free product of } (\mathcal{S}, \varphi_s),\ (\bar{\mathcal{A}}_{\text{odd}}, \bar{\varphi}_{\text{odd}}) \text{ and } (\bar{\mathcal{A}}_{\text{even}}, \bar{\varphi}_{\text{even}}). \quad (5.19)$$

Consider the subalgebra $\bar{\mathcal{U}}$ of $\mathcal{A}$ as

$$\bar{\mathcal{U}} = \text{Span}\Bigg(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} := (1+y)^{k_l} \prod_{i=1}^{k_l} \big(\bar{a}_{l,2i-1} s_{u_{l,i}} \underline{a}_{l,2i} s_{u_{l,i}}\big) \bar{a}_{l,2k_l+1} : l \geq 1\Bigg) \quad (5.20)$$

where $\bar{a}_{l,2i-1} \in \{\bar{b}_{2i-1}, \bar{b}_{2i-1}^*\}, \underline{a}_{l,2i} \in \{\underline{b}_{2i}, \underline{b}_{2i}^*\}$ and $s_{u_{l,i}} \in \{s_u\}$. Note that $\bar{\mathcal{U}}$ forms a $*$-algebra.

Then we have the following Theorem. This result may not be available in the literature in exactly this form. However, the ideas are already available in the free probability literature. For example see Benaych-Georges [2009], Benaych-Georges [2010] and Benaych-Georges and Nadakuditi [2012].

Theorem 5.3.1. *Suppose Assumptions $(A1)-(A3)$ hold and $n, p(n) \to \infty$, $p/n \to y > 0$. Then*

(a) $(\bar{\mathcal{U}}_p, (n+p)^{-1} E\text{Tr}) \to (\bar{\mathcal{U}}, \bar{\varphi})$, *and*

(b) for any polynomial Π,

$$\lim \frac{1}{p} E\text{Tr}(\Pi(\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1)) = \frac{1+y}{y} \bar{\varphi}(\Pi(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1)). \quad (5.21)$$

Hence, $(\mathcal{U}_p, p^{-1}E\text{Tr})$ *converges. The limit NCP may be denoted as* $(\mathcal{U} := \text{Span}(p_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1), \varphi)$, *say, where*

$$\varphi(\Pi(p_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1)) = \frac{1+y}{y} \bar{\varphi}(\Pi(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1)).$$

Proof. (a) By Definition 4.3.4, it is enough to show that for any polynomial Π,

$$\lim \frac{1}{n+p} E\mathrm{Tr}(\Pi(\{\bar{\mathbb{P}}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})}\})) \;=\; \bar{\varphi}(\Pi(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})})). \tag{5.22}$$

For this, we first embed $\{Z_u\}$ into the Wigner matrices $\{W_u\}$ of order $(n+p)$ as given in (5.4). Recall $\bar{B}$ and $\underline{D}$ respectively for the matrices B and D of orders p and n in (5.5). Therefore,

$$\Pi(\{\bar{\mathbb{P}}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})}\}) = \Pi\Bigg(\{\prod_{i=1}^{k_l}\Big(n^{-1}\bar{A}_{l,2i-1}W_{u_{l,i}}\underline{A}_{l,2i}W^*_{u_{l,i}}\Big)\bar{A}_{l,2k_l+1}\}\Bigg). \tag{5.23}$$

By using (5.14), (5.16) and Theorem 4.3.7 (a), (d), the NCP (Span($\{\bar{B}_{2i-1}, \bar{B}^*_{2i-1} : i \geq 1\}, (n+p)^{-1}\mathrm{Tr}$), (Span$\{\underline{B}_{2i}, \underline{B}^*_{2i} : i \geq 1\}, (n+p)^{-1}\mathrm{Tr}$) and (Span$\{(n+p)^{-1/2}W_u : 1 \leq u \leq U\}, (n+p)^{-1}E\mathrm{Tr}$) respectively converge to $(\bar{\mathcal{A}}_{\text{odd}}, \bar{\varphi}_{\text{odd}})$, $(\bar{\mathcal{A}}_{\text{even}}, \bar{\varphi}_{\text{even}})$ and $(\mathcal{S}, \varphi_s)$ and they are asymptotically free. Note that $\{\bar{B}_{2i-1}, \bar{B}^*_{2i-1}\}$ and $\{\underline{B}_{2i}, \underline{B}^*_{2i}\}$ are not in general asymptotically free. They are asymptotically free in polynomials where $\{B_{2i-1}, B^*_{2i-1}\}$ and $\{B_{2i}, B^*_{2i}\}$ are respectively enclosed within (Z^*, Z) and (Z, Z^*). Therefore, observing (5.19), (5.22) holds.

(b) Note that for any polynomial Π,

$$\begin{aligned}
&\lim p^{-1}E\mathrm{Tr}(\Pi(\{\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1\})) \\
= \;& \lim \frac{n+p}{p}(n+p)^{-1}E\mathrm{Tr}(\Pi(\{\bar{\mathbb{P}}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1\})) \\
= \;& y^{-1}(1+y)\lim(n+p)^{-1}E\mathrm{Tr}(\Pi(\{\bar{\mathbb{P}}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1\})) \\
= \;& y^{-1}(1+y)\bar{\varphi}(\Pi(\{\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1\})) \\
= \;& \varphi(\Pi(\{p_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1\})), \text{ (say)}.
\end{aligned}$$

This completes the proof of Theorem 5.3.1 (b). □

Remark 5.3.2. *It is easy to see that $\bar{\varphi}$ and φ both are tracial and positive (see (4.36), (4.37) and (4.40)).*

5.4 LSD of symmetric polynomials in $\{\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})}\}$

The following Theorem guarantees the existence of the LSD of any symmetric polynomial in $\{\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l}}\}$.

Theorem 5.4.1. *Suppose (A1)-(A3) hold and* $n, p(n) \to \infty$, $p/n \to y > 0$. *Then the LSD of any symmetric polynomial* $\Pi(\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1)$ *exists almost surely and it is uniquely determined by the (usual) moment sequence*

$$\begin{aligned}
& \lim \frac{1}{p} E\mathrm{Tr}(\Pi(\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1))^k \qquad (5.24)\\
= \;& \varphi(\Pi(p_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1))^k \\
= \;& y^{-1}(1+y)\bar{\varphi}(\Pi(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1))^k, \ \forall k \geq 1.
\end{aligned}$$

Proof. By Lemma 4.2.1, we need to establish (M1), (M4) and (C) as described in the moment method in Section 4.2. The (M1) condition is nothing but (5.21) in Theorem 5.3.1 (b). Now we shall establish (M4) and (C).

Proof of (M4). To establish (M4), we need the following lemma, whose proof is very technical and is deferred to Section 5.5.

Lemma 5.4.2. *Suppose (A1)-(A3) hold and* $n, p(n) \to \infty$, $p/n \to y > 0$. *Let* $\mathbb{P}_u \in \mathrm{Span}\{\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})}\}$, $u \geq 0$. *Let for* $1 \leq i \leq T$, $m_i(\mathbb{P}_u, \mathbb{P}_u^* : u \geq 0)$ *be polynomials. Let*

$$\mathcal{P}_i = \mathrm{Tr}(m_i(\mathbb{P}_u, \mathbb{P}_u^* : u \geq 0)) \text{ and } \mathcal{P}_i^0 = E\mathcal{P}_i.$$

For $d \geq 1$, *define*

$$\mathcal{S}_d = \text{set of all pair partitions } \{(i_1,i_2),(i_3,i_4),\ldots,(i_{2d-1},i_{2d})\} \text{ of } \{1,2,\ldots,2d\}.$$

Then, for all $d \geq 1$,

$$\begin{aligned}
&\lim E\left[\Pi_{i=1}^{T}\left(\mathcal{P}_i - \mathcal{P}_i^0\right)\right] \qquad (5.25)\\
= &\begin{cases} 0 \;\; \textit{if } T = 2d-1, \\ \sum_{\mathcal{S}_d} \prod_{k=1}^{d} \lim E\left[\left(\mathcal{P}_{i_{2k-1}} - \mathcal{P}^0_{i_{2k-1}}\right)\left(\mathcal{P}_{i_{2k}} - \mathcal{P}^0_{i_{2k}}\right)\right], \;\; \textit{if } T = 2d. \end{cases}
\end{aligned}$$

For any polynomial $\Pi(\mathbb{P}_u, \mathbb{P}_u^* : \; u \geq 0)$, taking $T = 4$ and $\mathcal{P}_i = \text{Tr}\left(\Pi(\mathbb{P}_u, \mathbb{P}_u^* : \; u \geq 0)\right)^h$ in Lemma 5.4.2, we have,

$$E\left[\frac{1}{p}\text{Tr}\left(\Pi(\mathbb{P}_u, \mathbb{P}_u^* : \; u \geq 0)\right)^h - E\left(\frac{1}{p}\text{Tr}\left(\Pi(\mathbb{P}_u, \mathbb{P}_u^* : \; u \geq 0)\right)^h\right)\right]^4 = O(p^{-4}) = O(n^{-4})$$

and hence (M4) is established.

Proof of (C). We have to show, for any symmetric polynomial Π,

$$\sum_{k=1}^{\infty}\left(y^{-1}(1+y)\bar{\varphi}(\Pi(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1))^{2k}\right)^{-1/2k} = \infty. \qquad (5.26)$$

Now note that

$$\begin{aligned}
&\sum_{k=1}^{\infty}\left(y^{-1}(1+y)\bar{\varphi}(\Pi(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1))^{2k}\right)^{-1/2k} \\
\geq \;\; &\frac{y}{1+y}\sum_{k=1}^{\infty}\left(\bar{\varphi}(\Pi(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1))^{2k}\right)^{-1/2k}.
\end{aligned}$$

Therefore, to prove (5.26), it is enough to show that

$$\sum_{k=1}^{\infty}\left(\bar{\varphi}(\Pi(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1))^{2k}\right)^{-1/2k} = \infty. \qquad (5.27)$$

Moreover, to prove (5.27), it is enough to show that for some $C > 0$,

$$\bar{\varphi}(\Pi(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1))^{2k} \leq C^{2k}, \;\; \forall k \geq 1. \qquad (5.28)$$

The following lemma is useful in this proof.

Lemma 5.4.3. *Let s be a standard semicircle variable. For all $\{\bar{a}_{2i-1}\} \in \{\bar{b}_{2i-1}, \bar{b}^*_{2i-1}\}$, $\{\bar{a}_{2i}\} \in \{\bar{b}_{2i}, \bar{b}^*_{2i}\}$, $h \geq 1$ and for some $C_1 > 0$, we have*

$$|\bar{\varphi}(\bar{a}_1 s \underline{a}_2 s \bar{a}_3 \ldots \underline{a}_{2h} s)| \leq C_1^{2h}.$$

Proof. Recall $||\cdot||_2$ defined in (2.4). By Assumptions (A2) and (A3), there exists $C > 0$ such that

$$\begin{aligned}
\sup_{1\leq i\leq K} \sup_p ||B_{2i-1}||_2 &= \sup_{1\leq i\leq K} \sup_p ||\bar{B}_{2i-1}||_2 \leq C, \quad \text{and} && (5.29)\\
\sup_{1\leq i\leq L} \sup_n ||B_{2i}||_2 &= \sup_{1\leq i\leq L} \sup_n ||\bar{B}_{2i}||_2 \leq C. && (5.30)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\bar{\varphi}(\bar{b}^*_{2i-1}\bar{b}_{2i-1})^h &= \lim \frac{1}{n+p}\mathrm{Tr}(\bar{B}^*_{2i-1}\bar{B}_{2i-1})^h \\
&\leq \sup_p ||\bar{B}^*_{2i-1}\bar{B}_{2i-1}||_2^h \leq C^{2h}, \ \forall h \geq 1, \ 1 \leq i \leq K. \quad (5.31)
\end{aligned}$$

Similarly,

$$\bar{\varphi}(\underline{b}^*_{2i}\underline{b}_{2i})^h \leq C^{2h}, \quad \forall h \geq 1, \ 1 \leq i \leq L. \quad (5.32)$$

Also note that, for all $\bar{a}_{2i-1} \in \{\bar{b}_{2i-1}, \bar{b}^*_{2i-1} : 1 \leq i \leq K\}$, $\underline{a}_{2i} \in \{\underline{b}_{2i}, \underline{b}^*_{2i} : 1 \leq i \leq L\}$ and $h \geq 1$, by Lemma 4.3.1 (b), there exists $\{h_i : 1 \leq i \leq 2h\}$ such that

$$|\bar{\varphi}(\bar{a}_1\underline{a}_2\bar{a}_3 \ldots \bar{a}_{2h-1}\underline{a}_{2h})| \leq \prod_{i=1}^{h} \left(\bar{\varphi}(\bar{a}^*_{2i-1}\bar{a}_{2i-1})^{h_{2i-1}}\right)^{1/h_{2i-1}} \prod_{i=1}^{h} \left(\bar{\varphi}(\underline{a}^*_{2i}\underline{a}_{2i})^{h_{2i}}\right)^{1/h_{2i}}.$$

Hence, by (5.31) and (5.32)

$$|\bar{\varphi}(\bar{a}_1\underline{a}_2\bar{a}_3 \ldots \bar{a}_{2h-1}\underline{a}_{2h})| \leq C^{2h}, \ \forall h \geq 1. \quad (5.33)$$

Therefore, applying (4.50) and (4.64) and using the fact that $\#NC_2(2h) \leq 2^{2h}, \ \forall h \geq 1$,

$$|\bar{\varphi}(\bar{a}_1 s \underline{a}_2 s \bar{a}_3 \ldots \underline{a}_{2h} s)| \leq C^{2h}(\#NC_2(2h)) \leq (2C)^{2h}.$$

Hence the proof of Lemma 5.4.3 is complete. □

Now by (5.20), note that

$$\begin{aligned} &\Pi(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1) = \sum_{i=1}^{T} g_i, \quad \text{where} \\ &g_i = \bar{a}_{1,i} s \underline{a}_{2,i} s \cdots \underline{a}_{2l_i,i} s, \quad \forall i \geq 1, \end{aligned} \tag{5.34}$$

$\bar{a}_{2j-1,i} \in \{\bar{b}_{2i-1}, \bar{b}^*_{2i-1} : i \geq 1\}$ and $\underline{a}_{2j,i} \in \{\underline{b}_{2i}, \underline{b}^*_{2i} : i \geq 1\}$. Now, by Lemmas 4.3.1 (b) and 5.4.3, there is a $C_1, C_2 > 0$ such that

$$\begin{aligned} &\bar{\varphi}(\Pi(\bar{p}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : l \geq 1))^{2k} = \bar{\varphi}\left(\sum_{i=1}^{T} g_i\right)^{2k} \\ = &\sum_{1 \leq i_1, i_2, \ldots, i_{2k} \leq T} \bar{\varphi}(g_{i_1} g_{i_2} \ldots g_{i_{2k}}) \\ \leq &\sum_{1 \leq i_1, i_2, \ldots, i_{2k} \leq T} |\bar{\varphi}(g_{i_1} g_{i_2} \ldots g_{i_{2k}})| \\ \leq &\ C_1^{2\sum_{j=1}^{2k} l_{i_j}} T^{2k} \leq C_2^{2k}. \end{aligned} \tag{5.35}$$

Hence, (5.28) is proved and (C) follows. This completes the proof of Theorem 5.4.1. □

5.4.1 Stieltjes transform

We enlarge the collections $\{B_{2i-1}\}$ and $\{B_{2i}\}$ so that they are closed under $*$ operation. All results proved so far continue to remain valid. Consider the polynomials

$\Delta \in \mathcal{U}_p$ of the form

$$\Delta = \frac{1}{n}\sum_{i=1}^{q} B_{4i-3} Z B_{2i} Z^* B_{4i-1}. \tag{5.36}$$

We assume appropriate conditions on $\{B_i\}$ so that Δ is symmetric.

Note that all the existing LSD results in the literature, discussed in Chapter 4, are for random matrices which are special cases of Δ. Moreover, the matrices $\{\Delta_u\}$, which are defined in (5.1) and which will approximate the sample autocovariance matrices, are also special cases of Δ. By Theorem 5.4.1, under (A1)-(A3), the almost sure LSD of Δ exists and it is characterized by the moment sequence

$$\lim \frac{1}{p} E\mathrm{Tr}(\Delta)^k = \frac{1+y}{y}\bar{\varphi}(\bar{\delta})^k, \ \forall k \geq 1, \tag{5.37}$$

where

$$\bar{\delta} = (1+y)\sum_{i=1}^{q} \bar{b}_{4i-3} s \underline{b}_{2i} s \bar{b}_{4i-1}. \tag{5.38}$$

Recall that $\{\bar{b}_{2i-1}\}$ and $\{\underline{b}_{2i}\}$ are respectively limits of $\{\bar{B}_{2i-1}\}$ and $\{\underline{B}_{2i}\}$. Moreover, s, $\{\bar{b}_{2i-1}\}$ and $\{\underline{b}_{2i}\}$ are free (by Theorem 5.4.1, as far as computing limits of polynomials of the form (5.36) is concerned).

However, most of the existing LSD discussed in Chapter 4 are in terms of the Stieltjes transform. Therefore to show how these results follow from Theorem 5.4.1, we need to study the Stieltjes transform of the LSD of Δ. The following theorem provides this Stieltjes transform.

Note that $\bar{\delta}$ is self-adjoint and $\bar{\varphi}$ is positive. By (A2) and (A3), there is $C > 0$ such that $|\bar{\phi}(\bar{\delta}^k)| \leq C^k$, $\forall k$. Hence, by Lemma 4.2.2 (b), there is a unique probability measure on $\mathbb{R}$, say $\bar{\mu}$, characterized by the moment sequence $\{\bar{\phi}(\bar{\delta}^k)\}$. Let μ be the probability measure on $\mathbb{R}$ corresponding to the LSD of Δ. Note that by (5.37),

$$\int_{\mathbb{R}} x^k d\mu = \frac{1+y}{y} \int_{\mathbb{R}} x^k d\bar{\mu}, \ \forall k \geq 1. \tag{5.39}$$

Let $m_{\bar{\mu}}(z)$ and $m_{\mu}(z)$ be respectively the Stieltjes transforms of $\bar{\mu}$ and μ. We first describe $m_{\bar{\mu}}(z)$. Then it is easy to express $m_{\mu}(z)$ in terms of $m_{\bar{\mu}}(z)$.

To describe $m_{\bar{\mu}}(z)$, we write infinite sums of the form $\sum_{1 \leq i_1, i_2, \ldots, i_k < \infty} a_{i_1} a_{i_2} \ldots a_{i_k}$ in the sense that

$$\varphi\big(\sum_{1 \leq i_1, i_2, \ldots, i_k < \infty} a_{i_1} a_{i_2} \ldots a_{i_k}\big) = \sum_{1 \leq i_1, i_2, \ldots, i_k < \infty} \varphi(a_{i_1} a_{i_2} \ldots a_{i_k}),$$

whenever

$$\sum_{1 \leq i_1, i_2, \ldots, i_k < \infty} |\varphi(a_{i_1} a_{i_2} \ldots a_{i_k})| < \infty. \tag{5.40}$$

Moreover, we write $(1-a)^{-1} := \sum_{i=0}^{\infty} a^i$.

Let,

$$\begin{aligned} d &= \{\bar{b}_{4i-3} : 1 \leq i \leq q\}, \quad e = \{\bar{b}_{4i-1} : 1 \leq i \leq q\}, \quad \text{and} && (5.41) \\ f &= \{\underline{b}_{2i} : 1 \leq i \leq q\}, \quad h(d,e,f) = (1+y) \sum_{i=1}^{q} \bar{b}_{4i-3} \underline{b}_{2i} \bar{b}_{4i-1}. && (5.42) \end{aligned}$$

Define

$$R_j(f) = \bar{\varphi}\left(h(d,e,f)\bar{\delta}^{j-1} | f\right) := (1+y) \sum_{i=0}^{q} \bar{\varphi}(\bar{b}_{4i-3} \bar{b}_{4i-1} \bar{\delta}^{j-1}) \underline{b}_{2i}. \tag{5.43}$$

Note that $R_j(f) \in \mathcal{A}, \ \forall j \geq 1$. Let,

$$\begin{aligned} K(z,f) &= \sum_{i=1}^{\infty} z^{-i} R_i(f) && (5.44) \\ B(d,e,z) &= \bar{\varphi}\left(h(d,e,f)(1+K(z,f))^{-1} | d, e\right), && (5.45) \end{aligned}$$

$$
\begin{aligned}
&:= (1+y)\sum_{i=1}^{q}\bar{\varphi}\left(\underline{b}_{2i}(1+yK(z,f))^{-1}\right)\bar{b}_{4i-3}\bar{b}_{4i-1} &(5.46)\\
&= (1+y)\sum_{j=0}^{\infty}\sum_{i=1}^{q}\bar{\varphi}(\underline{b}_{2i}(-K(z,f))^{i})\bar{b}_{4i-3}\bar{b}_{4i-1}. &(5.47)\\
G(d,e,z) &= (B(d,e,z)-z)^{-1} = z^{-1}\sum_{i=0}^{\infty}z^{-i}(B(d,e,z))^{-i}. &(5.48)
\end{aligned}
$$

The following lemma guarantees the existence of $K(z,f), B(d,e,z)$ and $G(d,e,z)$.

Lemma 5.4.4. *$K(z,f), B(d,e,z)$ and $G(d,e,z)$ exist for all sufficiently large $|z|$, in the sense of (5.40).*

Proof. Note that there is a $C > 0$ such that for any $\{\bar{a}_{2i-1}\} \in \{\bar{b}_{2i-1}, \bar{b}^*_{2i-1}\}$, $\{\underline{a}_{2i}\} \in \{\underline{b}_{2i}, \underline{b}^*_{2i}\}$ and $h \geq 1$, we have

$$|\bar{\varphi}(\bar{a}_1\bar{a}_3\ldots\bar{a}_{2h-1})| \leq C^h, \quad |\bar{\varphi}(\underline{a}_2\underline{a}_4\ldots\underline{a}_{2h})| \leq C^h \text{ and } |\bar{\varphi}(\bar{\delta})^h| \leq C^h. \tag{5.49}$$

Proof of (5.49) is along the same lines as the proof of Condition (C) in Theorem 5.4.1. Hence we omit it.

We first show that

$$(K(z,f))^r = \sum_{1\leq j_1,j_2,\ldots,j_r<\infty} z^{-\sum_{k=1}^{r} j_k}\bar{\varphi}(\prod_{k=0}^{r} R_{j_k}(f)) \tag{5.50}$$

exists. For this, we have to show

$$\sum_{1\leq j_1,j_2,\ldots,j_r<\infty} |z|^{-\sum_{k=1}^{r} j_k}|\bar{\varphi}(\prod_{k=0}^{r} R_{j_k}(f))| < \infty, \ \forall r \geq 1. \tag{5.51}$$

Now, note that, by (5.43)

$$R_j(f) = \bar{\varphi}(h(d,e,f)\bar{\delta}^{j-1}|f) = \sum_{i=0}^{q}\bar{\varphi}(\bar{b}_{4i-3}\bar{b}_{4i-1}\bar{\delta}^{j-1})\underline{b}_{2i}.$$

Therefore,

$$
\begin{aligned}
|\bar\varphi(\prod_{k=0}^{r} R_{j_k}(f))| &\leq \sum_{i_1,i_2,\ldots,i_r=0}^{q} \left(\prod_{k=1}^{r} |\bar\varphi(\bar b_{4i_k-3}\bar b_{4i_k-1}\bar\delta^{j_k-1})|\right) |\bar\varphi(\prod_{k=1}^{r} \underline{b}_{2i_k})| \\
&\leq \sum_{i_1,i_2,\ldots,i_r=0}^{q} \left(\prod_{k=1}^{r} \left(\bar\varphi(\bar b_{4i_k-3}\bar b_{4i_k-1}\bar b^*_{4i_k-1}\bar b^*_{4i_k-3})\bar\varphi(\bar\delta)^{2j_k-2}\right)^{1/2}\right) |\bar\varphi(\prod_{k=1}^{r} \underline{b}_{2i_k})| \\
&\leq (q+1)^r C^{2r+\sum_{k=1}^{r} j_k}, \quad \text{by (5.49)} \\
&\leq C_1^{\sum_{k=1}^{r} j_k}, \quad \text{for some } C_1 > C > 0. \qquad (5.52)
\end{aligned}
$$

Therefore, for $|z| > \eta C_1$, $\eta > 1$, $(K(z,f))^r$ in (5.50) exists and moreover, for all $r \geq 1$, we have

$$
\bar\varphi(K(z,f))^r = \bar\varphi\left(\sum_{j=1}^{\infty} \frac{R_j(f)}{z^j}\right)^r = \left(\sum_{j_1,j_2,\ldots,j_r=1}^{\infty} \frac{\bar\varphi(R_{j_1}(f)R_{j_2}(f)\ldots R_{j_r}(f))}{z^{j_1}z^{j_2}\ldots z^{j_r}}\right).
$$

Therefore,

$$
\begin{aligned}
|\bar\varphi(K(z,f))^r| &\leq \left(\sum_{j_1,j_2,\ldots,j_r=1}^{\infty} \frac{|\bar\varphi(R_{j_1}(f)R_{j_2}(f)\ldots R_{j_r}(f))|}{|z|^{j_1}|z|^{j_2}\ldots|z|^{j_r}}\right) \\
&\leq \sum_{j_1,j_2,\ldots,j_r=1}^{\infty} \frac{C_1^{\sum_{k=1}^{r} j_k}}{|z|^{j_1}|z|^{j_2}\ldots|z|^{j_r}}, \quad \text{by (5.52)} \\
&\leq \left(\sum_{j=1}^{\infty} \frac{C_1^j}{(\eta C_1)^j}\right)^r \leq \left(\frac{1}{\eta-1}\right)^r, \quad \text{as } \eta > 1. \qquad (5.53)
\end{aligned}
$$

Note that by (5.47), (we shall show below that this infinite sum in (5.54) exists)

$$
B(d,e,z) = \sum_{i=0}^{\infty}\sum_{j=0}^{q} \bar\varphi(\underline{b}_{2j}(-K(z,f))^i)\bar b_{4j-3}\bar b_{4j-1} = \sum_{i=0}^{\infty} A_i, \text{ say.} \qquad (5.54)
$$

Now, for all $|z| > \eta C_1$, $\eta > 1$,

$$
|\bar\varphi(A_{i_1}A_{i_2}\ldots A_{i_r})| = \sum_{j_1,j_2,\ldots,j_r=0}^{q} \left(\prod_{k=1}^{r} |\bar\varphi(\underline{b}_{2j_k}(-K(z,f))^{i_k}|\right) |\bar\varphi(\prod_{k=1}^{r} \bar b_{4j_k-3}\bar b_{4j_k-1})|
$$

$$
\begin{aligned}
&\leq \sum_{j_1,j_2,\ldots,j_r=0}^{q} \left(\prod_{k=1}^{r} \left(\bar{\varphi}(\underline{b}_{2j_k}\underline{b}_{2j_k}^*)\bar{\varphi}(K(z,f))^{2i_k}\right)^{1/2}\right) |\bar{\varphi}(\prod_{k=1}^{r} \bar{b}_{4j_k-3}\bar{b}_{4j_k-1})| \\
&\leq \sum_{j_1,j_2,\ldots,j_r=0}^{q} \left(\prod_{k=1}^{r} C^2 \left(\frac{1}{\eta-1}\right)^{2i_k}\right)^{1/2} C^{2r} \\
&\qquad \text{by (5.49) and (5.53)} \\
&\leq (q+1)^r C^{3r} \left(\prod_{k=1}^{r} \left(\frac{1}{\eta-1}\right)^{i_k}\right) \qquad (5.55)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sum_{1\leq i_1,i_2,\ldots,i_r<\infty} |\bar{\varphi}(A_{i_1}A_{i_2}\ldots A_{i_r})| &\leq (q+1)^r C^{3r} \left(\sum_{i=0}^{\infty} \left(\frac{1}{\eta-1}\right)^{i}\right)^r, \quad \text{(by (5.55))} \\
&\leq \left(\frac{(q+1)C(\eta-1)}{(\eta-2)}\right)^r, \text{ as } \eta > 2 \\
&\leq C_2^r, \quad \text{for some } C_2 > C_1 > C > 0. \qquad (5.56)
\end{aligned}
$$

Therefore, observing (5.54), $(B(d,e,z))^r$ exists for all $|z| > \eta C_1$, $\eta > 2$. Now, using the same arguments as in the existence of $K(z,f)$ above, it is easy to see that $G(d,e,z)$ defined in (5.48) exists for all $|z| > C_2$. This completes the proof of Lemma 5.4.4. □

The following Theorem provides $m_{\bar{\mu}}(z)$ and $m_{\mu}(z)$.

Theorem 5.4.5. *Suppose (A1)-(A3) hold and $n, p(n) \to \infty$, $p/n \to y > 0$.*

(a) Then for $z \in \mathbb{C}^+$, $|z|$ large, $m_{\bar{\mu}}(z)$ is given by

$$m_{\bar{\mu}}(z) = \bar{\varphi}(G(d,e,z)), \qquad (5.57)$$

where $G(d,e,z)$ satisfies (5.44), (5.45) and (5.48). Moreover, $K(z,f)$ in (5.44) also satisfies

$$K(z,f) = \bar{\varphi}(h(d,e,f)(B(d,e,z)-z)^{-1}|f) \qquad (5.58)$$

$$:= \quad (1+y)\sum_{i=1}^{q} \bar{\varphi}(\bar{b}_{4i-3}\bar{b}_{4i-1}(B(d,e,z)-z)^{-1})\underline{b}_{2i}. \tag{5.59}$$

(b) For $z \in \mathbb{C}^+$, $m_\mu(z)$ is given by

$$m_{\bar{\mu}}(z) = \frac{y}{1+y}m_\mu(z) - \frac{1}{1+y}\frac{1}{z}. \tag{5.60}$$

To prove Theorem 5.4.5, we shall prove a lemma that provides a recursion formula for the moments of $\bar{\mu}$. For convenience of writing, let us denote $D_i = \bar{B}_{4i-3}$, $E_i = \bar{B}_{4i-1}$, $d_i = \bar{b}_{4i-3}$ and $e_i = \bar{b}_{4i-1}$, $f_i = \underline{b}_{2i}$ for all $1 \leq i \leq q$. For any polynomial $\Pi = \Pi(D_j, D_j^*, E_j, E_j^* : j \geq 0)$, let $\Pi^0 = \Pi(d_j, d_j^*, e_j, e_j^* : j \geq 0)$. Recall $\{R_j(f)\}$ defined in (5.43). For all $j \geq 0$, let

$$S_j(f,\Pi) \quad = \quad \bar{\varphi}\left(\bar{\Pi}^0 h(d,e,f)\bar{\delta}^{j-1}|f\right) := \sum_{i=0}^{q} \bar{\varphi}(\bar{\Pi}^0\bar{b}_{4i-3}\bar{b}_{4i-1}\bar{\delta}^{j-1})\underline{b}_{2i}. \tag{5.61}$$

Lemma 5.4.6. *Suppose (A1)-(A3) hold and $n, p(n) \to \infty$, $p/n \to y > 0$. Then for any polynomial $\Pi = \Pi(D_j, D_j^*, E_j, E_j^* : j \geq 0)$, we have*

$$\lim (n+p)^{-1} E\mathrm{Tr}(\Pi\bar{\Delta}^r) \quad = \quad \sum_{t=1}^{r}\bar{\varphi}\Big[\sum_{\substack{1\leq i_1,i_2,\ldots,i_t\leq r \\ \sum_{j=1}^t i_j = r}} S_{i_1}(f,\Pi)\Big(\prod_{k=2}^{t} R_{i_k}(f)\Big)\Big].$$

Proof. By Theorem 5.4.1, it is immediate that

$$\lim (n+p)^{-1} E\mathrm{Tr}(\Pi\Delta^r) = \bar{\varphi}(\Pi^0\bar{\delta}^r) = (1+y)^r \sum_{\substack{j_k=1 \\ 1\leq k\leq r}}^{q} \bar{\varphi}\left(\Pi^0\prod_{k=1}^{r} d_{j_k} s f_{j_k} s e_{j_k}\right)$$

$$= \sum_{\sigma\in NC_2(2r)} \tau_\sigma, \quad \text{by (4.64)} \tag{5.62}$$

where

$$\tau_\sigma = (1+y)^r \sum_{\substack{j_k=1 \\ 1\leq k\leq r}}^{q} \bar{\varphi}_{K(\sigma)}[f_{j_1}, e_{j_1}d_{j_2}, f_{j_2}, e_{j_2}d_{j_3}, \ldots, e_{j_r}d_{j_1}\Pi^0],$$

and $K(\sigma)$ is the Kreweras complement of σ defined in Section 4.3.2. Now to compute (5.62), we consider the decomposition of $NC_2(2r) = \cup_{t=1}^{r}\mathcal{P}_t^{2r}$, where $\mathcal{P}_1^{2r} = \{\sigma \in NC_2(2r) : \{1,2\} \in \sigma\}$ and for all $2 \leq t \leq r$,

$$\begin{aligned}\mathcal{P}_t^{2r} &= \{\sigma \in NC_2(2r) :\ \{2k_0-1, 2k_t\}, \{2k_0, 2k_1-1\}, \{2k_1, 2k_2-1\}, \ldots, \\ &\qquad \{2k_{t-2}, 2k_{t-1}-1\} \in \sigma,\ 1 = k_0 < k_1 < k_2 < \ldots < k_{t-1} \leq r\}.\end{aligned}$$

Hence, (5.62) is equivalent to

$$\lim (n+p)^{-1} E\mathrm{Tr}(\Pi\bar{\Delta}^r) = \sum_{t=1}^{r} \mathcal{T}_t, \tag{5.63}$$

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

$$\mathcal{T}_t = \sum_{\sigma\in\mathcal{P}_t^{2r}} \tau_\sigma = \sum_{1=k_0<k_1<k_2<\ldots<k_{t-1}\leq r} g(t, k_1, k_2, \ldots, k_{t-1}), \tag{5.64}$$

and

$$\begin{aligned}&(1+y)^{-t-1} g(t+1, k_1, k_2, \ldots, k_t) \qquad (5.65)\\ =& \sum_{1\leq j_{k_s}\leq q} \bar{\varphi}\Big(\prod_{s=0}^{t} f_{j_{k_s}}\Big) \prod_{s=0}^{t} \bar{\varphi}\big(e_{j_{k_s}} \bar{\delta}^{k_{s+1}-k_s-1} d_{j_{k_{(s+1)}}}\big),\\ &\qquad \text{(by Lemma 4.3.6 (b) and where } k_{t+1} = r+1 \text{ and } d_{j_{k_{t+1}}} = d_{j_{k_0}}\Pi^0) .\end{aligned}$$

Therefore, (5.65) is equal to

$$\begin{aligned}&= \sum_{1\leq j_{k_s}\leq q} \bar{\varphi}\Big(\prod_{s=0}^{t} f_{j_{k_s}}\Big) \prod_{s=0}^{t} \bar{\varphi}\big(e_{j_{k_s}} \bar{\delta}^{k_{s+1}-k_s-1} d_{j_{k_{(s+1)}}}\big)\\ &= \bar{\varphi}\Big(\sum_{j_{k_s}, j_{k_{s+1}}} \prod_{s=0}^{t} f_{j_{k_s}} \bar{\varphi}\big(e_{j_{k_s}} \bar{\delta}^{k_{s+1}-k_s-1} d_{j_{k_{(s+1)}}}\big)\Big),\\ &\qquad\qquad \text{where } j_{k_{t+1}} = j_{k_0}.\end{aligned}$$

Note that

$$\sum_{j'_{k_s}, j_{k_{s+1}}} f_{j_{k_s}} \bar{\varphi}\big(e_{j_{k_s}} \bar{\delta}^{k_{s+1}-k_s-1} d_{j_{k_{(s+1)}}}\big) = \begin{cases} (1+y)^{-1} R_{(k_{s+1}-k_s)}(f), \ 1 \le s \le t-1 \\ (1+y)^{-1} S_{(r+1-k_t)}(f, \Pi), \ s = t. \end{cases} .$$

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

$$g(t+1, k_1, k_2, \ldots, k_t) \;=\; \bar{\varphi}(S_{(r+1-k_t)}(f, \Pi) \prod_{s=0}^{t-1} R_{(k_{s+1}-k_s)}(f)).$$

Therefore, by (5.64), for all $1 \le t \le r$,

$$\begin{aligned} \mathcal{T}_t &= \sum_{\substack{1=k_0<k_1<\ldots<k_t \le r \\ k_{t+1}=r+1}} \bar{\varphi}\big[S_{(r+1-k_t)}(f, \Pi) \prod_{s=0}^{t-1} R_{(k_{s+1}-k_s)}(f)\big] \\ &= \sum_{\substack{1 \le i_1, i_2, \ldots, i_t \le r \\ i_1+i_2+\ldots+i_t=r}} \bar{\varphi}(S_{i_1}(f, \Pi) \prod_{s=2}^{t} R_{i_s}(f)). \end{aligned}$$

Hence, by using (5.63) and (5.64), Lemma 5.4.6 follows. □

Now we are ready to prove Theorem 5.4.5.

Proof of Theorem 5.4.5. (a) Define

$$D = \sum_{i=1}^{\infty} z^{-i} \bar{\delta}^i. \tag{5.66}$$

Note that by the last part of (5.49), D exists for sufficiently large $|z|$. Moreover, by (4.7), $\varphi(D) = m_{\bar{\mu}}(z)$.

To establish (5.58), note that

$$\begin{aligned} K(z, f) &= \sum_{i=1}^{\infty} z^{-i} \bar{\varphi}(h(d, e, f) \bar{\delta}^{i-1} | f), \text{ by (5.43) and (5.44)} \qquad (5.67) \\ &= -z^{-1} \bar{\varphi}(h(d, e, f) | f) - z^{-1} \bar{\varphi}(h(d, e, f) D | f) \end{aligned}$$

where for f' with same property as in f, we have

$$\begin{aligned}
&\bar{\varphi}(h(d,e,f)D|f)\\
=\;& \sum_{r=1}^{\infty} z^{-r}\sum_{t=1}^{r}\bar{\varphi}\Big[\Big(\sum_{\substack{1\leq i_1,i_2,\cdots,i_t\leq r\\ i_1+i_2+\cdots+i_t=r}} S_{i_1}(f',h(d,e,f))\prod_{s=2}^{t}R_{i_s}(f')\Big)|f\Big]\\
&\hfill\text{(by Lemma 5.4.6)}\\
=\;& \sum_{t=1}^{\infty}\bar{\varphi}\Big[\sum_{r=t}^{\infty} z^{-r}\Big(\sum_{\substack{1\leq i_1,i_2,\cdots,i_t\leq r\\ i_1+i_2+\cdots+i_t=r}} S_{i_1}(f',h(d,e,f))\prod_{s=2}^{t}R_{i_s}(f')\Big)|f\Big]\\
=\;& \sum_{t=1}^{\infty}\bar{\varphi}\left[(K(z,f'))^{t-1}\left(\sum_{r=1}^{\infty} z^{-r}S_r(f',h(d,e,f))\right)|f\right]\\
=\;& \bar{\varphi}\left[(1+K(z,f'))^{-1}\left(\sum_{r=1}^{\infty} z^{-r}S_r(f',h(d,e,f))\right)|f\right]\\
=\;& \bar{\varphi}\left(\sum_{r=1}^{\infty} z^{-r}\bar{\delta}^{r-1}h(d,e,f)\bar{\varphi}\left[h(d,e,f)(1+yK(z,f'))^{-1}\right]|f\right)\\
=\;& z^{-1}\bar{\varphi}(h(d,e,f)B(d,e,z)|f)+z^{-1}\varphi(Dh(d,e,f)B(d,e,z)|f).
\end{aligned}$$

In a similar fashion, using $h(d,e,f)B(d,e,z)$ instead of $h(d,e,f)$ in the above steps,

$$\bar{\varphi}(Dh(d,e,f)B(d,e,z)|f)=z^{-1}\bar{\varphi}(h(d,e,f)B^2(d,e,z)|f)+z^{-1}\bar{\varphi}(Dh(d,e,f)B^2(d,e,z)|f).$$

Finally iterating we have

$$\bar{\varphi}(h(d,e,f)D|f)=\sum_{r=1}^{\infty} z^{-r}\bar{\varphi}(h(d,e,f)B^r(d,e,z)|f). \tag{5.68}$$

Hence,

$$\begin{aligned}
&K(z,f)\\
=\;& -z^{-1}\bar{\varphi}\big(\sum_{r=0}^{\infty} h(d,e,f)z^{-r}B^r(d,e,z)|f\big)=\bar{\varphi}(h(d,e,f)(B(d,e,z)-z)^{-1}|f), \quad (5.69)
\end{aligned}$$

which is (5.58) in Theorem 5.4.5.

Note that the above steps from (5.67) leading to (5.69) remain valid if we replace $h(d, e, f)$ by 1 in (5.67). This yields (instead of (5.69)),

$$\begin{aligned} m_{\bar{\mu}}(z) &= -z^{-1} \sum_{i=0}^{\infty} \left(\frac{B(d, e, z)}{z} \right)^i & (5.70) \\ &= \bar{\varphi}((B(d, e, z) - z)^{-1}), & (5.71) \end{aligned}$$

which is (5.57) in Theorem 5.4.5. Hence the proof of Theorem 5.4.5 (a) is complete.

(b) Let δ_0 be the degenerate probability measure at 0. Then (5.60) follows immediately by noting that

$$\bar{\mu} = \frac{y}{1+y}\mu + \frac{1}{1+y}\delta_0. \tag{5.72}$$

Hence the proof of Theorem 5.4.5 is complete. □

5.4.2 Corollaries

This section collects corollaries of three kinds. Corollaries 5.4.7, 5.4.8 and 5.4.9 will be useful later when we deal with LSD of sample autocovariance matrices. Corollaries 5.4.10 and 5.4.11 show how the existing LSD results in the literature can be quickly derived using Theorem 5.4.5. Finally, Corollary 5.4.12 provides completely new results.

Recall $\{\Delta_u\}$ defined in (5.1) and the coefficient matrices $\{\psi_j\}$ in (3.2). Suppose $\{\psi_j\} \subset \{B_{2i-1}, B_{2i-1}^*\}$ i.e. we assume:

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

Suppose

$$(\text{Span}\{\psi_j, \psi_j^* : j \geq 0\}, p^{-1}\text{Tr}) \;\to\; (\text{Span}\{\eta_j, \eta_j^* : j \geq 0\}, \varphi_{\text{odd}}), \tag{5.73}$$

$$(\mathrm{Span}\{\bar{\psi}_j, \bar{\psi}_j^* : j \geq 0\}, (n+p)^{-1}\mathrm{Tr}) \quad \to \quad (\mathrm{Span}\{\bar{\eta}_j, \bar{\eta}_j^* : j \geq 0\}, \bar{\varphi}_{\mathrm{odd}}). \tag{5.74}$$

Recall the NCP $(\mathcal{A}_{\mathrm{odd}}, \varphi_{\mathrm{odd}})$ and $(\bar{\mathcal{A}}_{\mathrm{odd}}, \bar{\varphi}_{\mathrm{odd}})$ respectively defined in (5.13) and (5.14). Clearly the NCP in the right side of (5.73) and (5.74) are $*$-sub-algebras of $(\mathcal{A}_{\mathrm{odd}}, \varphi_{\mathrm{odd}})$ and $(\bar{\mathcal{A}}_{\mathrm{odd}}, \bar{\varphi}_{\mathrm{odd}})$, respectively. Recall that $\bar{\varphi}$ is the state corresponding to the free product given in (5.19). Therefore, by Definition 4.3.8, the restriction of $\bar{\varphi}$ on $\bar{\mathcal{A}}_{\mathrm{odd}}$ is $\bar{\varphi}_{\mathrm{odd}}$. Also note that for any polynomial Π

$$\begin{aligned}
\bar{\varphi}(\Pi(\bar{\eta}_j, \bar{\eta}_j^* : j \geq 0)) &= \bar{\varphi}_{\mathrm{odd}}(\Pi(\bar{\eta}_j, \bar{\eta}_j^* : j \geq 0)) \\
&= \frac{y}{1+y}\varphi_{\mathrm{odd}}(\Pi(\eta_j, \eta_j^* : j \geq 0)). \qquad (5.75)
\end{aligned}$$

Now we have the following corollary. This corollary will be useful later in Chapter 7, when we deal with the Stieltjes transform of the LSD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$.

Corollary 5.4.7. *Suppose (A1), (B) hold and $n, p(n) \to \infty$, $p/n \to y > 0$. Then the almost sure LSD of $\frac{1}{2}(\Delta_u + \Delta_u^*)$ exists and its Stieltjes transform $m_u(z)$, for $z \in \mathbb{C}^+$ and $|z|$ large, is given by*

$$\begin{aligned}
m_u(z) &= \varphi_{\mathrm{odd}}((B_u(\lambda, z) - z)^{-1}), \quad \textit{where} && (5.76)\\
K_u(z,\theta) &= \varphi_{\mathrm{odd}}(h(\lambda,\theta)(B_u(\lambda, z) - z)^{-1}), && (5.77)\\
&:= \sum_{j,k=0}^{q} \varphi_{\mathrm{odd}}(\eta_j\eta_k^*(B_u(\lambda, z) - z)^{-1})e^{i(j-k)\theta} \\
h(\lambda,\theta) &= (\sum_{j=0}^{q} e^{ij\theta}\eta_j)(\sum_{j=0}^{q} e^{-ij\theta}\eta_j^*), \ \lambda = \{\eta_j, \eta_j^* : j \geq 0\}, && (5.78)\\
B_u(\lambda, z) &= E_\theta\left(\cos(u\theta)h(\lambda,\theta)(1 + y\cos(u\theta)K(z,\theta))^{-1}\right), && (5.79)\\
&:= \sum_{j,k=0}^{q} \eta_j\eta_k^* E_\theta\left(\cos(u\theta)e^{i(j-k)\theta}(1 + y\cos(u\theta)K(z,\theta))^{-1}\right).
\end{aligned}$$

and θ is a $U(0, 2\pi)$ random variable.

Proof. First note that $\{\Delta_u\}$ satisfy the form (5.2). Moreover, under (B) and (5.73), $\{\psi_j\}$ satisfy (A2) and (5.13). Also note that the matrices $\{P_u : u =$

$0, \pm 1, \pm 2, \ldots\}$ satisfy (A3). Suppose,

$$(\mathrm{Span}\{\underline{P}_u : u = 0, \pm 1, \ldots\}, (n+p)^{-1}\mathrm{Tr}) \to (\mathrm{Span}\{\underline{c}_u : u = 0, \pm 1, \ldots\}, \bar{\varphi}). \tag{5.80}$$

Then for any polynomial Π

$$\begin{aligned} \bar{\varphi}(\Pi(\{\underline{c}_u : u = 0, \pm 1, \ldots\})) &= \lim \frac{1}{n+p}\mathrm{Tr}(\Pi(\{\underline{P}_u : u = 0, \pm 1, \ldots\})) \\ &= \frac{1}{1+y} E_\theta(\Pi(\{e^{iu\theta} : u = 0, \pm 1, \ldots\})), \end{aligned} \tag{5.81}$$

where $\theta \sim U(0, 2\pi)$.

Now applying Theorem 5.4.5, (5.38), (5.41) and (5.42) reduce to

$$\begin{aligned} \bar{\delta} &= \frac{1+y}{2} \sum_{j,k=0}^{q} \bar{\eta}_j s(\underline{c}_{j-k+u} + \underline{c}_{j-k-u})\bar{\eta}_k^*, \\ d &= (\bar{\eta}_j : 0 \le j \le q), \ \ e = (\bar{\eta}_j^* : 0 \le j \le q), \\ f &= (\underline{c}_{j-k+u} : 0 \le j, k \le q), \\ h(d, e, f) &= \frac{1+y}{2} \sum_{j,k=0}^{q} \bar{\eta}_j(\underline{c}_{j-k+u} + \underline{c}_{j-k-u})\bar{\eta}_k^*. \end{aligned} \tag{5.82}$$

Also define

$$g = (e^{i(j-k+u)\theta} : 0 \le j, k \le q).$$

Now, (5.45) reduces to

$$\begin{aligned} B(d, e, z) &= \bar{\varphi}(h(d, e, f)(1 + K(z, f))^{-1} | d, e) \\ &= \frac{1+y}{2} \sum_{j,k=0}^{q} \bar{\eta}_j \bar{\eta}_k^* \bar{\varphi}((\underline{c}_{j-k+u} + \underline{c}_{j-k-u})(1 + K(z, f))^{-1}), \\ &= \frac{1+y}{2} \sum_{j,k=0}^{q} \bar{\eta}_j \bar{\eta}_k^* \frac{1}{1+y} E_\theta\left(e^{i(j-k)\theta}(2\cos(u\theta))(1 + K(z, g))^{-1}\right), \quad \text{by (5.81)} \\ &= E_\theta\left(h(d, \theta)(\cos(u\theta))(1 + K(z, g))^{-1}\right). \end{aligned} \tag{5.83}$$

Let $\lambda = (\eta_j : 0 \leq j \leq q)$. Then (5.58) reduces to

$$\begin{aligned} K(z,f) &= \bar{\varphi}(h(d,e,f)(B(d,e,z)-z)^{-1}|f) \\ &= \frac{1+y}{2}\sum_{j,k=0}^{q}(\underline{c}_{j-k+u}+\underline{c}_{j-k-u})\bar{\varphi}(\bar{\eta}_j\bar{\eta}_k^*(B(d,e,z)-z)^{-1}), \\ &= \frac{1+y}{2}\sum_{j,k=0}^{q}(\underline{c}_{j-k+u}+\underline{c}_{j-k-u})\frac{y}{1+y}\varphi_{\text{odd}}(\eta_j\eta_k^*(B(\lambda,\lambda^*,z)-z)^{-1}), \\ &\qquad \text{by (5.75)} \\ &= \frac{y}{2}\sum_{j,k=0}^{q}(\underline{c}_{j-k+u}+\underline{c}_{j-k-u})\varphi_{\text{odd}}(\eta_j\eta_k^*(B(\lambda,\lambda^*,z)-z)^{-1}). \end{aligned} \tag{5.84}$$

Let $B(\lambda,\lambda^*,z) = B(\lambda,z)$, say. Therefore,

$$\begin{aligned} K(z,g) &= y\cos(u\theta)\varphi_{\text{odd}}(h(\lambda,\theta)(B(\lambda,z)-z)^{-1}) = y\cos(u\theta)K(z,\theta), \quad \text{where} \\ K(z,\theta) &= \varphi_{\text{odd}}(h(\lambda,\theta)(B(\lambda,z)-z)^{-1}). \end{aligned} \tag{5.85}$$

This establishes (5.77). Now, by (5.83), we have

$$B(\lambda,z) = E_\theta\left(h(d,\theta)(\cos(u\theta))(1+y\cos(u\theta)K(z,\theta))^{-1}\right). \tag{5.86}$$

This establishes (5.79). Now, (5.57) reduces to

$$m_{\bar{\mu}}(z) = \bar{\varphi}((B(d,e,z)-z)^{-1}) = \frac{y}{1+y}\varphi_{\text{odd}}(B(\lambda,z)-z)^{-1}) - \frac{1}{1+y}\frac{1}{z}.$$

Therefore, by (5.60), we have

$$m_u(z) = \varphi_{\text{odd}}(B(\lambda,z)-z)^{-1}), \tag{5.87}$$

where $B(\lambda,z)$ satisfies (5.83) and (5.85). This establishes (5.76) and completes the proof of Corollary 5.4.7. $\square$

Apart from the cumbersome expressions (5.76)-(5.79) of the Stieltjes transformation, in general, there is no further simplified form of the LSD of $\Delta_u + \Delta_u^*$. We have a better description of the LSD of $\Delta_u + \Delta_u^*$ in the special case $\psi_0 = I_p$, $\psi_j = \lambda_j I_p$, $\lambda_j \in \mathbb{R}$, for all $j \geq 1$. The following corollary will be useful later in Chapter 7 and there we shall also obtain the limit Stieltjes transform. Recall the compound free Poisson distribution in Definition 4.3.9.

Corollary 5.4.8. *Suppose (A1) holds and $n, p(n) \to \infty$, $p/n \to y > 0$. Let $\psi_0 = I_p$, $\psi_j = \lambda_j I_p$, $1 \leq j \leq q$. Then the LSD of $\frac{1}{2}(\Delta_u + \Delta_u^*)$ is the compound free Poisson whose r-th order free cumulant equals*

$$\kappa_{ur} = y^{r-1} E_\theta(\cos(u\theta)\tilde{h}(\lambda, \theta))^r, \quad \forall i \geq 0, \tag{5.88}$$

where

$$\tilde{h}(\lambda, \theta) = |\sum_{j=0}^{q} e^{ij\theta}\lambda_j|^2, \ \lambda_0 = 1, \ \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_q) \text{ and } \theta \sim U(0, 2\pi). \tag{5.89}$$

Proof. Note that we can write

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

and hence

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

Note that by (5.81) and for all $r \geq 1$,

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

Therefore, invoking the discussion around (4.76), the LSD of $\frac{1}{2}(\Delta_u + \Delta_u^*)$ is a compound free Poisson with the r-th order free cumulant

$$y^{r-1} \lim \frac{1}{n} \mathrm{Tr}\big(\frac{1}{2} \sum_{j,j'=0}^{q} \lambda_j \lambda_{j'} (P_{j-j'+u} + P^*_{j-j'+u})\big)^r = y^{r-1} E_\theta(\cos(u\theta)\tilde{h}(\lambda,\theta))^r,$$

where $\tilde{h}$ is as given in (5.89) and $\theta \sim U(0, 2\pi)$. Hence, the proof of Corollary 5.4.8 is complete. □

The following corollary will be invoked later in Chapter 7, when we deal with the Stieltjes transform of the LSD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ for the MA(0) process.

Corollary 5.4.9. *Suppose (A1) holds and* $n, p(n) \to \infty$*,* $p/n \to y > 0$*. Then for each* $u \geq 1$*, LSD of* $(2n)^{-1}Z(P_u + P_u^*)Z^*$ *exists almost surely and its Stieltjes transform* $m(z)$ *is given by the solution (only one solution is a valid solution) of*

$$(1 - y^2 m^2(z))(yzm(z) + y - 1)^2 = 1, \quad z \in \mathbb{C}^+. \tag{5.90}$$

Proof. By Theorem 5.4.1, LSD of $(2n)^{-1}Z(P_u + P_u^*)Z^*$ exists almost surely. To obtain the Stieltjes transform, we now use Corollary 5.4.7. So assume $z \in \mathbb{C}^+$ and $|z|$ large. Note that $(2n)^{-1}Z(P_u + P_u^*)Z^* = \Delta_u,\ \forall u \geq 1$ with

$$\psi_0 = I_p, \quad \psi_j = 0,\ \forall j \geq 1. \tag{5.91}$$

By (5.91) and (5.73), $\eta_0 = 1_{\mathcal{A}_{\text{odd}}}$ (the identity element of $\mathcal{A}_{\text{odd}}$) and $\eta_j = 0,\ \forall j \geq 1$. Therefore, (5.78) reduces to

$$\lambda \;=\; 1_{\mathcal{A}_{\text{odd}}}, \quad h(\lambda,\theta) = 1_{\mathcal{A}_{\text{odd}}}. \tag{5.92}$$

By (5.76) and (5.77), we have ($z \in \mathbb{C}^+$, $|z|$ large)

$$K_u(z,\theta) \;=\; \varphi_{\text{odd}}(1_{\mathcal{A}_{\text{odd}}}(B_u(1_{\mathcal{A}_{\text{odd}}}, z) - z)^{-1}) = \varphi_{\text{odd}}((B_u(1_{\mathcal{A}_{\text{odd}}}, z) - z)^{-1}), \tag{5.93}$$

$$m_u(z) \;=\; \varphi_{\text{odd}}((B_u(1_{\mathcal{A}_{\text{odd}}}, z) - z)^{-1}). \tag{5.94}$$

Therefore, by (5.93) and (5.94), we have ($z \in \mathbb{C}^+$, $|z|$ large)

$$m_u(z) = K_u(z, \theta). \tag{5.95}$$

By (5.79) and (5.95), we have ($z \in \mathbb{C}^+$, $|z|$ large)

$$\begin{aligned} B_u(1_{\mathcal{A}_{\text{odd}}}, z) &= E_\theta(\cos(u\theta)1_{\mathcal{A}_{\text{odd}}}(1 + y\cos(u\theta)m_u(z))^{-1}) \\ &= E_\theta(\cos(u\theta)(1 + y\cos(u\theta)m_u(z))^{-1})1_{\mathcal{A}_{\text{odd}}} \\ &= \left(\frac{1}{2\pi}\int_0^{2\pi} \frac{\cos(u\theta)}{1 + y\cos(u\theta)m_u(z)}\right)1_{\mathcal{A}_{\text{odd}}}. \end{aligned} \tag{5.96}$$

Hence, by (5.94) and (5.96) and, for $z \in \mathbb{C}^+$ and $|z|$ large, the Stieltjes transform of the LSD of $\frac{1}{2n}Z(P_u + P_u^*)Z^*$ satisfies,

$$\begin{aligned} m_u(z) &= \varphi_{\text{odd}}(B(1_{\mathcal{A}_{\text{odd}}}, z) - z)^{-1} \\ &= -\varphi_{\text{odd}}\left(z - \frac{1}{2\pi}\int_0^{2\pi} \frac{\cos(u\theta)\ d\theta}{1 + ym_u(z)\cos(u\theta)}1_{\mathcal{A}_{\text{odd}}}\right)^{-1} \\ &= -\left(z - \frac{1}{2\pi}\int_0^{2\pi} \frac{\cos(u\theta)\ d\theta}{1 + ym_u(z)\cos(u\theta)}\right)^{-1}. \end{aligned} \tag{5.97}$$

Now by contour integration, it can be shown that

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{\cos(u\theta)\ d\theta}{1 + ym_u(z)\cos(u\theta)} \;=\; \frac{1}{ym_u(z)} - \frac{2}{y^2m_u^2(z)}\frac{1}{\omega_1 - \omega_2}$$

where ω_1 and ω_2 are two roots of $\omega^2 + 2(ym_u(z))^{-1}\omega + 1 = 0$ with $|\omega_1| > 1$, $|\omega_2| < 1$ and $(\omega_1 - \omega_2)^{-2} = \frac{y^2m_u^2(z)}{4(1 - y^2m_u^2(z))}$. Therefore, by (5.97), for $z \in \mathbb{C}^+$ and $|z|$ large, we have

$$\frac{-1}{m_u(z)} = z - \frac{1}{ym_u(z)} + \frac{2(\omega_1 - \omega_2)^{-1}}{y^2m^2(z)} \implies ((1-y) - zym_u(z))^2(1 - y^2m_u^2(z)) = 1.$$

Hence, (5.90) is established for $z \in \mathbb{C}^+$ and $|z|$ large. Using analyticity of $m_u(z)$, (5.90) continues to hold for all $z \in \mathbb{C}^+$. □

Recall that in Theorems 4.2.6 and 4.2.7, we stated the LSD of $n^{-1}ZZ^*$ and $n^{-1}A^{1/2}ZZ^*A^{1/2}$ where A is a $p \times p$ symmetric, non-negative definite matrix. Recall the class $U(\delta)$ in (4.17). Among other things, there we assumed that $\{\varepsilon_{i,j} : i, j \geq 1\} \in U(0)$.

Now if we are willing to work under the stronger assumption (A1) and norm bounded A, then those conclusions follow from Theorems 5.4.1 and 5.4.5. If one carefully follows the existing proofs of Theorems 4.2.6 and 4.2.7 given in Bai and Silverstein [2009], he/she can see that these are first proved under (A1) and when A is norm bounded. Then to relax these assumptions, appropriate truncations on the entries of Z and on the ESD of A are used. The same truncation arguments can also be used for the following two corollaries. Recall the class of matrices $\mathcal{NND}$ in (4.18).

Corollary 5.4.10. *Suppose (A1) holds and $n, p(n) \to \infty$, $p/n \to y > 0$. Suppose $\{A_p\} \in \mathcal{NND}$, norm bounded and has LSD F^A. Then the almost sure LSD of $n^{-1}A_p^{1/2}ZZ^*A_p^{1/2}$ is given by (4.27). The same LSD result continues to hold if we relax the norm bounded assumption on A and, instead of (A1), assume $\{\varepsilon_{i,j} : i, j \geq 1\} \in U(0)$.*

Proof. We shall prove only the first part. Since the proof of the second part involves standard truncation as discussed, we shall omit it.

To prove the first part, we again use Corollary 5.4.7. So assume $z \in \mathbb{C}^+$ and $|z|$ large. Note that $n^{-1}A_p^{1/2}ZZ^*A_p^{1/2} = \Delta_0$ with

$$\psi_0 = A_p^{1/2}, \quad \psi_j = 0, \ \forall j \geq 1. \tag{5.98}$$

Let $a \sim F^A$. As A_p is symmetric and non-negative definite, $a^{1/2}$ is meaningful.

By (5.98) and (5.73), $\eta_0 = a^{1/2}$ and $\eta_j = 0,\ \forall j \geq 1$. Therefore, (5.78) reduces to

$$\lambda \;=\; a^{1/2}, \quad h(\lambda, \theta) = a^{1/2}a^{1/2} = a. \tag{5.99}$$

By (5.76) and (5.77), we have for $z \in \mathbb{C}^+$, $|z|$ large (since $u = 0$),

$$\begin{aligned}
K_0(z, \theta) &= \varphi_{\text{odd}}(a(B_0(a^{1/2}, z) - z)^{-1}) \\
&= \int \frac{adF^A(a)}{B_0(a^{1/2}, z) - z} = K(z), \text{ say,} && (5.100) \\
m_0(z) &= \varphi_{\text{odd}}((B_0(a^{1/2}, z) - z)^{-1}) = \int \frac{dF^A}{B_0(a^{1/2}, z) - z}. && (5.101)
\end{aligned}$$

Now by (5.79) and (5.100), we have for $z \in \mathbb{C}^+$, $|z|$ large (since $u = 0$),

$$\begin{aligned}
B_0(a^{1/2}, z) &= E_\theta(\cos(0\theta)a(1 + y\cos(0\theta)K(z)^{-1}) \\
&= E_\theta(a(1 + yK(z))^{-1}) = \frac{a}{1 + yK(z)}. && (5.102)
\end{aligned}$$

Hence, by (5.100), (5.101) and (5.102), we have for $z \in \mathbb{C}^+$, $|z|$ large,

$$\begin{aligned}
zm_0(z) &= \int \frac{zdF^A}{B_0(a^{1/2}, z) - z} = \int \frac{B_0(a^{1/2}, z)dF^A}{B_0(a^{1/2}, z) - z} - 1 \\
&= \frac{1}{1 + yK(z)} \int \frac{adF^A}{B_0(a^{1/2}, z) - z} - 1 = \frac{K(z)}{1 + yK(z)} - 1. \\
&= \frac{1}{y} \frac{yK(z)}{1 + yK(z)} - 1 = -\frac{1}{y} \frac{1}{1 + yK(z)} + \frac{1}{y} - 1. && (5.103)
\end{aligned}$$

Therefore, by (5.103) and (5.102), we have for $z \in \mathbb{C}^+$, $|z|$ large,

$$a(zym_0(z) + y - 1) = -\frac{a}{1 + yK(z)} = -B_0(a^{1/2}, z). \tag{5.104}$$

Now substituting the value of $B_0(a^{1/2}, z)$ obtained in (5.104) into (5.101), we have

$$m_0(z) = -\int \frac{dF^A}{z - a(zym_0(z) + y - 1)} \quad z \in \mathbb{C}^+,\ |z| \text{ large.}$$

Therefore, for $z \in \mathbb{C}^+$ and $|z|$ large, (4.27) is proved. Using analyticity, (4.27) continues to hold for all $z \in \mathbb{C}^+$. $\square$

We now give an alternative free probability proof of Theorem 4.2.6. Recall the well known Marčenko-Pastur law MP_y with parameter y in Section 4.2.2.

Corollary 5.4.11. *Suppose (A1) holds and $n, p(n) \to \infty$, $p/n \to y > 0$. Then the almost sure LSD of $n^{-1}ZZ^*$ exists and it is distributed as MP_y where the moment sequence and the Stieltjes transform respectively satisfy (4.25) and (4.26). The result continues to hold if instead of (A1), we assume $\{\varepsilon_{i,j} : i, j \geq 1\} \in U(0)$.*

Proof. Again we shall prove only the first part. We have already established (4.27) for general A_p in corollary 5.4.10. Put $A_p^{1/2} = I_p$, where I_p is as in (2.8). Then (4.26) follows immediately.

Next we show (4.25) using Theorem 5.4.1. Let $B_1 = I_p$ and $B_2 = I_n$. Then note that $n^{-1}ZZ^* = n^{-1}B_1 Z B_2 Z^* B_1$. Moreover, $\bar{B}_1 \to a_0, \underline{B}_2 \to c_0$, where a_0 and c_0 are both Bernoulli random variables with success probabilities $y(1+y)^{-1}$ and $(1+y)^{-1}$ respectively. Let s be a semi-circle variable and suppose s, a_0 and c_0 are free. Observe that, by (5.24),

$$\lim p^{-1} E\mathrm{Tr}(n^{-1}ZZ^*)^h = y^{-1}(1+y)\bar{\varphi}\left((1+y)a_0 s c_0 s a_0\right)^h, \forall h \geq 1.$$

Hence, by (4.66), the h-th moment of the LSD of $n^{-1}ZZ^*$ is given by

$$\frac{(1+y)^{h+1}}{y} \sum_{\pi \in NC(h)} \bar{\varphi}_\pi[a_0^2, a_0^2, \ldots, a_0^2] \bar{\varphi}_{K(\pi)}[c_0, c_0, \ldots, c_0]. \tag{5.105}$$

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

$$\begin{aligned} \bar{\varphi}_\pi[a_0^2, a_0^2, \ldots, a_0^2] &= \bar{\varphi}_\pi[a_0, a_0, \ldots, a_0] = y^k(1+y)^{-k}, \\ \bar{\varphi}_\pi[c_0, c_0, \ldots, c_0] &= (1+y)^{-k}. \end{aligned}$$

By Property 4 of Kreweras complement in Section 4.3.2, if $\pi \in NC(h)$ has k

blocks then $K(\pi)$ has $(h-k+1)$ many blocks. Therefore, (5.105) equals

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

which is the h-th moment of the Marčenko-Pastur law (see (4.25)). For the last equality see page 144 of Nica and Speicher [2006]. This completes the proof of Corollary 5.4.11. □

The following corollary does not seem to be known in the literature.

Corollary 5.4.12. *Suppose (A1) and $n, p(n) \to \infty$, $p/n \to y > 0$ hold. Then for each $u \geq 1$, LSD of $p^{-2}ZP_uZ^*ZP_{-u}Z^*$ exists almost surely and is the Bessel$(2, y^{-1})$ (see Banica et al. [2011]) law whose hth moment satisfies*

$$\beta_h = \sum_{k=1}^{h} \frac{1}{k}\binom{h-1}{k-1}\binom{2h}{k-1} y^{-k},\ h \geq 1. \tag{5.106}$$

Proof. To establish (5.106), we again use Theorem 5.4.1. Let $B_1 = I_p$ and $B_2 = P_u$. Then note that $p^{-2}ZP_uZ^*ZP_{-u}Z^* = (n/p)^2 n^{-2} B_1ZB_2Z^*B_1ZB_2^*Z^*B_1$. Note that $\bar{B}_1 \to a_0$ where a_0 is the Bernoulli random variables with success probability $y(1+y)^{-1}$. Also $(\underline{B}_2, \underline{B}_2^*) \to (c, c^*)$, where c and c^* commute and

$$\bar{\varphi}(c^k c^{*l}) = \frac{1}{1+y} I(k = l). \tag{5.107}$$

Let s be the semi-circle variable and suppose s, a_0 and $\{c, c^*\}$ are free. Observe that, by (5.24),

$$\begin{aligned}
\lim p^{-1} E\mathrm{Tr}\left(p^{-2}ZP_uZ^*ZP_{-u}Z^*\right)^h &= \lim \left(\frac{n}{p}\right)^{2h} p^{-1}E\mathrm{Tr}\left(n^{-2}ZP_uZ^*ZP_{-u}Z^*\right)^h \\
&= y^{-(2h+1)}(1+y)\bar{\varphi}((1+y)^2 a_0 scs a_0^2 scs a_0), \forall h \geq 1.
\end{aligned}$$

Recall $NCE(2n)$ in (4.49). By (4.66), the h-th moment of the LSD of

$p^{-2}ZP_uZ^*ZP_{-u}Z^*$ is given by

$$\frac{(1+y)^{2h+1}}{y^{2h+1}} \sum_{\pi \in NC(2h)} \bar{\varphi}_{K(\pi)}[a_0, a_0, \ldots, a_0]\bar{\varphi}_{\pi}[c, c^*, c, c^* \ldots, c, c^*]. \quad (5.108)$$

Note that $\bar{\varphi}_{\pi}[c, c^*, c, c^* \ldots, c, c^*] = 0$ if $\pi \in NC(2h) - NCE(2h)$. If $\pi \in NCE(2h)$ has k many blocks, then $\bar{\varphi}_{\pi}[c, c^*, c, c^* \ldots, c, c^*] = (1+y)^k$. Note that by Property 4 of Kreweras complement in Section 4.3.2, $K(\pi)$ has $2h+1-k$ blocks and hence $\bar{\varphi}_{K(\pi)}[a_0, a_0, \ldots, a_0] = y^{2h+1-k}(1+y)^{2h+1-k}$. Therefore (5.108) equals

$$\begin{aligned} & y^{-2h} \sum_{k=1}^{h} \#\{\pi \in NCE(2h) : \ \pi \text{ has } k \text{ blocks}\}\ y^{2h+1-k-1}, \\ = \ & \sum_{k=1}^{h} \frac{1}{k} \binom{h-1}{k-1} \binom{2h}{k-1} y^{-k}, \end{aligned}$$

where the last equality follows from Lemma 4.1 of Edelman [1980]. The final expression is indeed the h-th moment of the free Bessel$(2, y^{-1})$ law. This proves Corollary 5.4.12. $\square$

5.5 Proof of Lemma 5.4.2

We first prove the result for monomials $\{m_i\}$ in $\{\mathbb{P}_u, \mathbb{P}_u^*\}$ involving *one independent matrix* Z. Once the result is proved for monomials, it is easy to see that it continues to hold for polynomials (see Step 1 below). Moreover it will also be clear that the arguments continue to hold when more than one independent matrices $\{Z_u\}$ are involved. The proof is completed in two steps.

Step 1. First we show that it is enough to prove the lemma for monomials. Consider arbitrary $p \times p$ matrices $\{A_{ik}, C_{ik} : 1 \le k \le r_i, 1 \le i \le T\}$ $\subset$ Span$\{B_{2i-1}, B_{2i-1}^*\}$ and $n \times n$ matrices $\{B_{ik} : 1 \le k \le r_i, 1 \le i \le T\} \subset$

$\text{Span}\{B_{2i}, B_{2i}^*\}$. Define

$$\pi_i = n^{-r_i}\text{Tr}(\prod_{k=1}^{r_i} A_{ik}ZB_{ik}Z^*C_{ik}) \text{ and } \pi_i^0 = E\pi_i,\ 1 \leq i \leq T. \tag{5.109}$$

For all $d \geq 1$, consider the equations

$$\begin{aligned} &\lim E\left[\Pi_{i=1}^T\left(\pi_i - \pi_i^0\right)\right] \\ = &\begin{cases} 0 \text{ if } T = 2d-1, \\ \sum_{\mathcal{S}_d}\prod_{k=1}^d \lim E\big[\big(\pi_{i_{2k-1}} - \pi^0_{i_{2k-1}}\big)\big(\pi_{i_{2k}} - \pi^0_{i_{2k}}\big)\big],\ \text{if } T = 2d. \end{cases} \end{aligned} \tag{5.110}$$

We now prove that (5.110) $\implies$ (5.25) i.e. it is enough to prove Lemma 5.4.2 for monomials only.

In the next step, we prove (5.110).

Note that for some matrices $\{A_{iks}, C_{iks}\} \in \text{Span}\{B_{2i-1}, B_{2i-1}^*\}$ and $\{B_{iks}\} \in \text{Span}\{B_{2i}, B_{2i}^*\}$, we can write

$$\begin{aligned} \mathcal{P}_i &= \sum_{k=1}^{t_i}\text{Tr}\left(n^{-r_k}\prod_{s=1}^{r_k} A_{iks}ZB_{iks}Z^*C_{iks}\right) = \sum_{k=1}^{t_i} S_{ik},\ \text{(say)}. \\ \mathcal{P}_i^0 &= \sum_{i=1}^{t_i} E(S_{ik}) = \sum_{i=1}^{t_i} S_{ik}^0,\quad \text{(say)}. \end{aligned}$$

Therefore,

$$\begin{aligned} \lim E(\prod_{i=1}^T(\mathcal{P}_i - \mathcal{P}_i^0)) &= \lim E(\prod_{i=1}^T\sum_{k=1}^{t_i}(S_{ik} - S_{ik}^0)) = \lim E(\prod_{i=1}^T\sum_{k=1}^{t_i} T_{ik}),\quad \text{(say)} \\ &= \lim E(\sum_{1\leq k_i\leq t_i}\prod_{i=1}^T T_{ik_i}) = \sum_{1\leq k_i\leq t_i}\lim E(\prod_{i=1}^T T_{ik_i}) \\ &= \begin{cases} 0,\quad \text{if } T = 2d-1 \\ \sum_{1\leq k_i\leq t_i}\sum_{\mathcal{S}_d}\prod_{s=1}^d \lim E(T_{i_{2s-1}k_{i_{2s-1}}}T_{i_{2s}k_{i_{2s}}}),\ \text{if } T = 2d. \end{cases} \end{aligned}$$

The last equality holds by (5.110). Therefore, (5.25) follows from (5.110) when T is odd.

When T is even, we have

$$\begin{aligned} &\sum_{\mathcal{S}_d}\prod_{s=1}^{d}\sum_{k_{2s-1},k_{2s}} \lim E(T_{i_{2s-1}k_{i_{2s-1}}}T_{i_{2s}k_{i_{2s}}}) \\ =& \sum_{\mathcal{S}_d}\prod_{k=1}^{d}\lim E\big[\big(\mathcal{P}_{i_{2k-1}}-\mathcal{P}^0_{i_{2k-1}}\big)\big(\mathcal{P}_{i_{2k}}-\mathcal{P}^0_{i_{2k}}\big)\big]. \end{aligned}$$

Therefore, (5.25) follows from (5.110) for all $T \geq 1$.

Step 2. **Proof of (5.110).** Let $A(i,j)$ be the (i,j)-th element of the matrix A. Note that, for all $1 \leq i \leq T$,

$$\begin{aligned} n^{r_i}\pi_i &= \mathrm{Tr}(\prod_{k=1}^{r_i} A_{ik}ZB_{ik}Z^*C_{ik}) \qquad (5.111)\\ =& \sum_{\substack{1\leq u_t^{(i)}\leq p,\ 1\leq v_s^{(i)}\leq p \\ 1\leq t\leq 3r_i,\ 1\leq s\leq 2r_i}} \prod_{\substack{k=1 \\ u_{3r_i+1}^{(i)}=u_1^{(i)}}}^{r_i} A_{ik}(u_{3k-2}^{(i)},u_{3k-1}^{(i)})\varepsilon_{u_{3k-1}^{(i)},v_{2k-1}^{(i)}}B_{ik}(v_{2k-1}^{(i)},v_{2k}^{(i)}) \\ & \varepsilon_{u_{3k}^{(i)},v_{2k}^{(i)}}C_{ik}(u_{3k}^{(i)},u_{3k+1}^{(i)}). \end{aligned}$$

For fix $1 \leq i \leq T$, we define

$$\mathcal{U}_i = \{(u_{3k+\delta}^{(i)}, v_{2k+\delta}^{(i)}) : 1 \leq k \leq r_i, \delta = -1, 0, u_{3r_i+1}^{(i)} = u_1^{(i)}\}. \qquad (5.112)$$

Note that $\mathcal{U}_i$ is the set of all indices attached with ε's in the expansion of π_i given in (5.111). An index $(u_{3k+\delta}^{(i)}, v_{2k+\delta}^{(i)})$ is said to be matched if there is at least one $(k',\delta',i') \neq (k,\delta,i)$ with $(u_{3k+\delta}^{(i)}, v_{2k+\delta}^{(i)}) = (u_{3k'+\delta'}^{(i')}, v_{2k'+\delta'}^{(i')})$. Now note that $E\left[\Pi_{i=1}^{T}\left(\pi_i - \pi_i^0\right)\right]$ involves all indices in $\cup_{i=1}^{T}\mathcal{U}_i$ (if we expand $\{\pi_i\}$ using the last equality of (5.111)). As $\{\varepsilon_{i,j}\}$ are independent and have mean 0, all indices in $\cup_{i=1}^{T}\mathcal{U}_i$ need to be matched to guarantee a non-zero contribution. For each

$1 \leq i \leq T$, consider the following sets of matched indices.

$$\begin{aligned} B_i &= \text{set of all matchings where for each } (k,\delta), \text{ there is at least one} \\ &\quad (k',\delta') \neq (k,\delta) \text{ with } (u^{(i)}_{3k+\delta}, v^{(i)}_{2k+\delta}) = (u^{(i)}_{3k'+\delta'}, v^{(i)}_{2k'+\delta'}) \text{ and for } i \neq i', \\ &\quad \text{there is no } (k',\delta',i') \text{ such that } (u^{(i)}_{3k+\delta}, v^{(i)}_{2k+\delta}) = (u^{(i')}_{3k'+\delta'}, v^{(i')}_{2k'+\delta'}) \ . \end{aligned}$$

Consider the disjoint decomposition $\cup_{i=1}^{T+1} C_i$ of all possible matchings of indices in $\cup_{i=1}^{T} \mathcal{U}_i$, where

$$C_1 = B_1, \ C_i = (\cap_{j=1}^{i-1} B_j^c) \cap B_i \ \forall 2 \leq i \leq T, \ C_{T+1} = \cap_{i=1}^{T} B_i^c. \tag{5.113}$$

Let for any set A, E_A be the usual expectation restricting on the set A. We shall first show that

$$E\left[\Pi_{i=1}^{T}\left(\pi_i - \pi_i^0\right)\right] = E_{C_{T+1}}(\Pi_{i=1}^{T}\pi_i). \tag{5.114}$$

For this purpose, we need more analysis for the set C_i. Define

$$\begin{aligned} \mathcal{S}_i &= \text{set of all matchings of indices in } \mathcal{U}_i, \text{and} \\ \mathcal{S}_{-i} &= \text{set of all matchings of indices in } \cup_{j\neq i}\mathcal{U}_j \text{ such that for each } 1 \leq j < i \\ &\quad \text{there is at least one index in } \mathcal{U}_j \text{ which matches with some index} \\ &\quad \text{in } \mathcal{U}_k, \ k \neq j, i. \end{aligned}$$

Note that

$$C_i = (\cap_{j=1}^{i-1} B_j^c) \cap B_i = \{(\sigma_1 \cup \sigma_2) : \sigma_1 \in \mathcal{S}_i, \ \sigma_2 \in \mathcal{S}_{-i}\}. \tag{5.115}$$

Also note that

$$E(\Pi_{i=1}^{T}(\pi_i - \pi_i^0)) = E_{C_i}((\Pi_{j=1}^{i}\pi_j)\Pi_{j=i+1}^{T}(\pi_j - \pi_j^0)) + \text{ other terms.}$$

Then for all $2 \leq i \leq T$, we have

$$E_{C_i}(\Pi_{j=1}^{i}\pi_j\Pi_{j=i+1}^{T}(\pi_j - \pi_j^0)) = \sum_{\sigma \in C_i} E_\sigma(\Pi_{j=1}^{i}\pi_j\Pi_{j=i+1}^{T}(\pi_j - \pi_j^0)) \quad (5.116)$$

$$= \sum_{\sigma_1 \in \mathcal{S}_i,\ \sigma_2 \in \mathcal{S}_{-i}} E_{\sigma_1}(\pi_i)E_{\sigma_2}(\Pi_{j=1}^{i-1}\pi_j\Pi_{j=i+1}^{T}(\pi_j - \pi_j^0))$$

[as $C_i \subset B_i$ and under B_i, $\{\varepsilon_{u,v} : (u,v) \in \mathcal{U}_i\}$ are independent of $\{\varepsilon_{u,v} : (u,v) \in \cup_{j \neq i}\mathcal{U}_j\}$]

$$= \left(\sum_{\sigma_1 \in \mathcal{S}_i} E_{\sigma_1}(\pi_i)\right)\left(\sum_{\sigma_2 \in \mathcal{S}_{-i}} E_{\sigma_2}(\Pi_{j=1}^{i-1}\pi_j\Pi_{j=i+1}^{T}(\pi_j - \pi_j^0))\right)$$

$$= \pi_i^0\left(E_{\cap_{j=1}^{i-1}B_j^c}(\Pi_{j=1}^{i-1}\pi_j\Pi_{j=i+1}^{T}(\pi_j - \pi_j^0)))\right).$$

Similarly,

$$E_{B_1}(\pi_1\Pi_{i=2}^{T}(\pi_i - \pi_i^0)) = \pi_1^0 E(\Pi_{i=2}^{T}(\pi_i - \pi_i^0)). \quad (5.117)$$

Now, the left side of (5.114) equals,

$$E\left[\Pi_{i=1}^{T}\left(\pi_i - \pi_i^0\right)\right]$$
$$= E_{B_1}(\pi_1\Pi_{i=2}^{T}(\pi_i - \pi_i^0)) + E_{B_1^c}(\pi_1\Pi_{i=2}^{T}(\pi_i - \pi_i^0)) - \pi_1^0 E(\Pi_{i=2}^{T}(\pi_i - \pi_i^0))$$
$$= E_{B_1^c}(\pi_1\Pi_{i=2}^{T}(\pi_i - \pi_i^0)),$$
$$= E_{B_1^c \cap B_2}(\pi_1\pi_2\Pi_{i=3}^{T}(\pi_i - \pi_i^0)) + E_{B_1^c \cap B_2^c}(\pi_1\pi_2\Pi_{i=3}^{T}(\pi_i - \pi_i^0))$$
$$-\pi_2^0 E_{B_1^c}(\pi_1\Pi_{i=3}^{T}(\pi_i - \pi_i^0))$$
$$= E_{B_1^c \cap B_2^c}(\pi_1\pi_2\Pi_{i=3}^{T}(\pi_i - \pi_i^0)), \quad \text{(by (5.116), for } T = 2)$$
$$\vdots$$
$$= E_{B_1^c \cap B_2^c \cap \ldots \cap B_T^c}(\Pi_{i=1}^{T}\pi_i), \text{ by repeated application of (5.116) for } 3 \leq i \leq T.$$
$$= E_{C_{T+1}}(\Pi_{i=1}^{T}\pi_i).$$

Therefore, (5.114) is established.

Next we shall analyze the set C_{T+1} and identify the set of matchings which contribute in the limit.

Two index sets $\mathcal{U}_i$ and $\mathcal{U}_{i'}$ are said to be connected if there is (k, δ) and (k', δ') with $(u^{(i)}_{3k+\delta}, v^{(i)}_{2k+\delta}) = (u^{(i')}_{3k+\delta}, v^{(i')}_{2k'+\delta'})$. Also a collection of index sets $\{\mathcal{U}_{i_1}, \mathcal{U}_{i_2}, \ldots, \mathcal{U}_{i_s}\}$, $s \geq 2$, is said to form a connected group if for each $1 \leq k \leq s-1$, $\mathcal{U}_{i_k}$ and $\mathcal{U}_{i_{k+1}}$ is connected. Note that, in a typical matching in C_{T+1}, for each i, $\mathcal{U}_i$ is connected with some other $\mathcal{U}_{i'}$, $i' \neq i$. Therefore, each matching in C_{T+1} corresponds to some disjoint connected groups each of length at least 2. Consider the following disjoint decomposition of C_{T+1}.

$$\begin{aligned} C_{T+1} &= \bigcup_{\substack{2 \leq g_1, g_2, \ldots, g_R \leq T \\ \sum_{j=1}^R g_j = T,\ R \geq 1}} G(g_1, g_2, \ldots, g_R), \text{ where} \\ G(g_1, g_2, \ldots, g_R) &= \text{set of all such matchings in } C_{T+1} \text{ which form exactly} \\ &\quad\ R \text{ connected groups of length } g_1, g_2, \ldots, g_R. \end{aligned}$$

Hence, by (5.114), we have

$$\begin{aligned} E\left[\Pi_{i=1}^T \left(\pi_i - \pi_i^0\right)\right] &= E_{C_{T+1}}(\Pi_{i=1}^T \pi_i) \qquad (5.118) \\ &= \sum_{\substack{2 \leq g_1, g_2, \ldots, g_R \leq T \\ \sum_{j=1}^R g_j = T,\ R \geq 1}} E_{G(g_1, g_2, \ldots, g_R)}(\Pi_{i=1}^T \pi_i). \end{aligned}$$

Consider a typical matching σ in $G(g_1, g_2, \ldots, g_R)$ with connected groups $G_{\sigma 1}, G_{\sigma 2}, \ldots,$ $G_{\sigma R}$ respectively of lengths $g_1, g_2, \ldots, g_R$. Note that, for a fixed σ, $\{G_{\sigma k} : 1 \leq k \leq R\}$ forms a partition of $\{\mathcal{U}_k : 1 \leq k \leq T\}$. Also note that, if $i \neq j$, then no index in $G_{\sigma i}$ matches with any index in $G_{\sigma j}$. Hence, by independence of $\{\varepsilon_{i,j}\}$,

$$\begin{aligned} E_{G(g_1, g_2, \ldots, g_R)}(\Pi_{i=1}^T \pi_i) &= \sum_{\sigma} \prod_{k=1}^R E_{G_{\sigma k}}(\pi_1, \pi_2, \ldots, \pi_T), \text{ where} \qquad (5.119) \\ E_{G_{\sigma k}}(\pi_1, \pi_2, \ldots, \pi_T) &= E_{G_{\sigma k}}\Bigg(\prod_{\substack{i_j:\ \mathcal{U}_{i_j} \in G_{\sigma k} \\ 1 \leq j \leq g_j}} \pi_{i_j}\Bigg), \ \forall 1 \leq k \leq R, \end{aligned}$$

and $E_{G_{\sigma k}}$ is the usual expectation restricting on the matchings in $G_{\sigma k}$. For the time being assume that the following claim is true. We shall prove the claim later.

Claim. $E_{G_{\sigma k}}(\pi_1, \pi_2, \ldots, \pi_T) = O(n^{-g_k+2}), \ \forall \sigma, k.$

Therefore, for all $\sigma \in G(g_1, g_2, \ldots, g_R)$,

$$\prod_{k=1}^{R} E_{G_{\sigma k}}(\pi_1, \pi_2, \ldots, \pi_T) = O(n^{-\sum(g_j - 2)}) = O(n^{-T+2R}) \tag{5.120}$$

As $G(g_1, g_2, \ldots, g_R)$ is a finite set, by (5.119), we have

$$E_{G(g_1,g_2,\ldots,g_R)}(\Pi_{i=1}^{T} \pi_i) = O(n^{-T+2R}). \tag{5.121}$$

Note that as $g_1, g_2, \ldots, g_R \geq 2$, the maximum possible value of R is $[T/2]$, the greatest integer $\leq T/2$.

First suppose T is odd. Then we always have $T - 2R > 0$ and hence, using (5.121), $\lim E_{G(g_1,g_2,\ldots,g_R)}(\Pi_{i=1}^{T} \pi_i) = 0$. As a consequence, using (5.118), we have

$$\lim E\left[\Pi_{i=1}^{T}\left(\pi_i - \pi_i^0\right)\right] = 0, \text{ if } T \text{ is odd}, \tag{5.122}$$

proving (5.110) when T is odd.

Now suppose T is even, say $T = 2d$. Then note that

$$T - 2R \begin{cases} = 0, \text{ for } G(2, 2, \ldots, 2), \ R = d \\ > 0, \text{ otherwise.} \end{cases} \tag{5.123}$$

Therefore, by (5.121),

$$\lim E_{G(g_1,g_2,\ldots,g_R)}(\Pi_{i=1}^{T} \pi_i) = 0, \text{ if } G(g_1, g_2, \ldots, g_R) \neq G(2, 2, \ldots, 2), \tag{5.124}$$

and hence by (5.118), we have

$$\lim E\left[\Pi_{i=1}^{T}\left(\pi_i - \pi_i^0\right)\right] = \lim E_{G(2,2,\ldots,2)}(\Pi_{i=1}^{T}\pi_i). \tag{5.125}$$

It remains to identify the right side of (5.125) as the right side of (5.110). Note that a typical matching in $G(2,2,\ldots,2)$ involves d groups each with length 2. Hence, there is a one-to-one correspondence of $G(2,2,\ldots,2)$ and $\mathcal{S}_d$, set of all pair partitions of $\{1,2,\ldots,2d\}$. The one-to-one correspondence is as follows. Consider $\sigma = \{(i_1,i_2),(i_3,i_4),\ldots,(i_{2d-1},i_{2d})\} \in \mathcal{S}_d$, then for every $1 \le k \le d$, $\{\mathcal{U}_{i_{2k-1}}, \mathcal{U}_{i_{2k}}\}$ forms a connected group. Therefore, by (5.119), we have

$$E_{G(2,2,\ldots,2)}(\Pi_{i=1}^{T}\pi_i) = \sum_{\sigma\in\mathcal{S}_d}\prod_{k=1}^{d} E_{\{\mathcal{U}_{i_{2k-1}},\, \mathcal{U}_{i_{2k}}\}}(\pi_1,\pi_2,\ldots,\pi_T). \tag{5.126}$$

Let D be the set of all such matchings of indices in $\mathcal{U}_{i_{2k-1}} \cup \mathcal{U}_{i_{2k}}$ such that $\{\mathcal{U}_{i_{2k-1}}, \mathcal{U}_{i_{2k}}\}$ are connected. Note that

$$\begin{aligned}
& E_{\{\mathcal{U}_{i_{2k-1}},\, \mathcal{U}_{i_{2k}}\}}(\pi_1,\pi_2,\ldots,\pi_T) \qquad (5.127)\\
= & \sum_{\sigma\in D} E_\sigma(\pi_{i_{2k-1}}\pi_{i_{2k}})\\
= & E\left(\left(\pi_{i_{2k-1}} - \pi^0_{i_{2k-1}}\right)\left(\pi_{i_{2k}} - \pi^0_{i_{2k}}\right)\right), \text{ by (5.114) for } T = 2.
\end{aligned}$$

Therefore, by (5.126) and (5.127), we have

$$E_{G(2,2,\ldots,2)}(\Pi_{i=1}^{T}\pi_i) = \sum_{\sigma\in\mathcal{S}_d}\prod_{k=1}^{d} E\left(\left(\pi_{i_{2k-1}} - \pi^0_{i_{2k-1}}\right)\left(\pi_{i_{2k}} - \pi^0_{i_{2k}}\right)\right). \tag{5.128}$$

Now substituting (5.128) in (5.125), we have established (5.110) for $T = 2d$. Therefore, by Steps 1 and 2, proof of (5.25) and hence of Lemma 5.4.2 is complete when one independent matrix is involved, provided the claim is true.

Proof of claim. As $\{\pi_i\}$ are commutative, it is enough to show that

$$E_C(\pi_1\pi_2\ldots\pi_g) = O(n^{-g+2}), \tag{5.129}$$

where C is the set of all matchings of indices in $\cup_{i=1}^{g}\mathcal{U}_i$ such that $\{\mathcal{U}_i : 1 \le i \le g\}$ forms a connected group. Recall $\{\pi_i\}$ from (5.111). Consider the following decomposition of C.

$$\begin{aligned} C = & \bigcup_{1\le t_j\le k_j\le r_j} \mathcal{C}(k_j, t_j : 1 \le j \le g), \text{ where} \\ \mathcal{C}(k_j, t_j : 1 \le j \le g) = & \text{ set of all matchings (may be pair, non-pair,} \\ & \text{crossing, non-crossing) in } C \text{ such that} \\ & (u^{(i)}_{3k_i}, v^{(i)}_{2k_i}) = (u^{(i+1)}_{3t_{i+1}-1}, v^{(i+1)}_{2t_{i+1}-1}), \ \forall 1 \le i \le g-1. \end{aligned} \tag{5.130}$$

Therefore,

$$E_C(\pi_1\pi_2\ldots\pi_g) = \sum_{1\le t_j\le k_j\le r_j} E_{\mathcal{C}(k_j,t_j:1\le j\le g)}(\pi_1\pi_2\ldots\pi_g). \tag{5.131}$$

Now for convenience of writing, let us denote, for all $1 \le i \le g$,

$$\begin{aligned} D_i &= (\Pi_{k=1}^{t_i-1} A_{ik}(Z/\sqrt{n})B_{ik}(Z^*/\sqrt{n})C_{ik})A_{it_i}, \\ E_i &= B_{it_i}(Z^*/\sqrt{n})C_{it_i}(\Pi_{k=t_i+1}^{k_i-1} A_{ik}(Z/\sqrt{n})B_{ik}(Z^*/\sqrt{n})C_{ik})A_{ik_i}ZB_{ik_i}, \\ F_i &= C_{ik_i}(\Pi_{k=k_i+1}^{r_i} A_{ik}(Z/\sqrt{n})B_{ik}(Z^*/\sqrt{n})C_{ik}), \text{ and} \end{aligned}$$

hence,

$$n\pi_i = \operatorname{Tr}(D_iZE_iZ^*F_i), \ \forall 1 \le i \le g. \tag{5.132}$$

Therefore,

$$\begin{aligned}
& n^g E_{\mathcal{C}(k_j,t_j:1\le j\le g)}(\pi_1\pi_2\ldots\pi_g) \\
= \; & E_{\mathcal{C}(k_j,t_j:1\le j\le g)}\big(\Pi_{i=1}^g \mathrm{Tr}(D_i Z E_i Z^* F_i)\big), \text{ (by (5.132))} \\
= \; & \sum_{\substack{\{u_{ij},v_{ik},\ 1\le i\le g\} \\ j=1,2,3,\ k=1,2}} E_{\mathcal{C}(k_j,t_j:1\le j\le g)}\big(\Pi_{i=1}^g D_i(u_{i1},u_{i2})\varepsilon_{u_{i2},v_{i1}} E_i(v_{i1},v_{i2})\varepsilon_{u_{i3},v_{i2}} F_i(u_{i3},u_{i1})\big) \\
= \; & \sum_{\sigma\in\mathcal{C}(k_j,t_j:1\le j\le g)} \sum_{\substack{\{u_{ij},v_{ik},\ 1\le i\le g\} \\ j=1,2,3,\ k=1,2}} E_\sigma\big(\Pi_{i=1}^g D_i(u_{i1},u_{i2})\varepsilon_{u_{i2},v_{i1}} E_i(v_{i1},v_{i2})\varepsilon_{u_{i3},v_{i2}} F_i(u_{i3},u_{i1})\big).
\end{aligned}$$

Now by (5.130), we have $(u_{i3}, v_{i2}) = (u_{(i+1)2}, v_{(i+1)1})$, $\forall 1 \le i \le g-1$ and therefore,

$$\begin{aligned}
& \sum_{\substack{\{u_{ij},v_{ik},\ 1\le i\le g\} \\ j=1,2,3,\ k=1,2}} E_\sigma\big(\Pi_{i=1}^g D_i(u_{i1},u_{i2})\varepsilon_{u_{i2},v_{i1}} E_i(v_{i1},v_{i2})\varepsilon_{u_{i3},v_{i2}} F_i(u_{i3},u_{i1})\big) \\
= \; & E_\sigma\big(\mathrm{Tr}(Z(\Pi_{i=1}^g E_i Z^* Z) Z^* (\Pi_{i=1}^g F_{g+1-i} D_{g+1-i}))\big).
\end{aligned}$$

Hence,

$$\begin{aligned}
& n^g E_{\mathcal{C}(k_j,t_j:1\le j\le g)}(\pi_1\pi_2\ldots\pi_g) \\
= & \sum_{\sigma\in\mathcal{C}(k_j,t_j:1\le j\le g)} E_\sigma\big(\mathrm{Tr}(Z(\Pi_{i=1}^g E_i Z^* Z) Z^* (\Pi_{i=1}^g F_{g+1-i} D_{g+1-i}))\big).
\end{aligned} \tag{5.133}$$

Using the same idea as in the proof of (M1) condition for Theorem 3.1, one can show that

$$\begin{aligned}
& \lim n^{-2} E_\sigma \mathrm{Tr}(Z(\Pi_{i=1}^g E_i Z^* Z) Z^* (\Pi_{i=1}^g F_{g+1-i} D_{g+1-i})) \\
= & \begin{cases} O(1), & \text{if } \sigma \text{ is non-crossing pair matching} \\ o(1), & \text{if } \sigma \text{ is not non-crossing or pair matching.} \end{cases}
\end{aligned}$$

Hence by (5.133), $E_{\mathcal{C}(k_j,t_j:1\le j\le g)}(\pi_1\pi_2\ldots\pi_g) = O(n^{-g+2})$. Therefore, by (5.131), (5.129) follows and the claim is established.

This completes the proof of Lemma 5.4.2 for one independent matrix Z. Note that if we have more than one independent matrix $\{Z_i\}$, then also the above proof will remain unchanged except $\varepsilon_{u,v}$ (the (u, v)-th element of Z) will be replaced by $\varepsilon_{i,u,v}$ (the (u, v)-th element of Z_i). $\square$

Chapter 6

Joint convergence of generalized dispersion matrices when $p/n \to 0$

6.1 Introduction

In this chapter, we are interested in the joint convergence of the class of matrices $\{\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})}\}$ defined in (5.2) as $p, n(p) \to \infty$, $p/n \to 0$. In Chapter 5, we dealt with the same problem when $n, p(n) \to \infty$, $p/n \to y > 0$. There we used asymptotic freeness of Wigner and deterministic matrices after embedding matrices of different orders into larger square matrices of same order. For example see (5.4) and (5.5) and the proof of Theorem 5.3.1. The embedding technique that we used in Chapter 5, does not work in this case as here the growth of p and n are not comparable. Moreover, recall the statements of Theorems 4.2.8-4.2.11. There we saw that very different centering and scaling on $\{\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})}\}$ are needed, to get non-degenerate limits. Define the centered and scaled matrices

$$\mathcal{R}_{l,(u_{l,1},\ldots,u_{l,k_l})} = (n/p)^{1/2}(\mathbb{P}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} - \mathbb{G}_{l,k_l}), \text{ where} \tag{6.1}$$

$$\mathbb{G}_{l,k_l} = \left(\prod_{i=1}^{k_l} n^{-1}\mathrm{Tr}\,(A_{l,2i})\right)\prod_{i=0}^{k_l} A_{l,2i+1} \tag{6.2}$$

are the centering matrices. We then consider the convergence of the sequence of NCP $(\mathcal{V}_p, p^{-1}E\text{Tr})$ as $p, n(p) \to \infty$, $p/n \to 0$, where

$$\mathcal{V}_p = \text{Span}\{\mathcal{R}_{l,(u_{l,1},\ldots,u_{l,k_l})} : l, k_l \geq 1\}. \tag{6.3}$$

Note that $\mathcal{V}_p$ forms a $*$-algebra. In Section 6.2.2, we show why $\{\mathbb{G}_{l,k_l}\}$ is the correct centering and $\sqrt{np^{-1}}$ is the correct scaling. In Section 6.2.3, we discuss the idea behind the limit. Then in Theorem 6.3.1, we state the result on convergence of this sequence of NCP. The limiting NCP can be expressed in terms of some free variables. Theorem 6.4.1 states that the LSD of any symmetric polynomial in $\{\mathcal{R}_{l,(u_{l,1},\ldots,u_{l,k_l})}\}$ exists and can be expressed in terms of free variables. In Section 6.4.1, Theorem 6.4.2, we derive the Stieltjes transform of this LSD. We also show how the existing LSD results in the literature follow as special cases of our LSD results. In Chapters 7 and 8, we will use these results for statistical inference in high-dimensional time series. *The main material of this chapter is taken from Bhattacharjee and Bose [2015b].*

6.2 Preliminaries

6.2.1 Assumptions

We first list all the assumptions that are required for the convergence of $(\mathcal{V}_p, p^{-1}E\text{Tr})$ as $p, n(p) \to \infty$, $p/n \to 0$. Some of these have already appeared in Chapter 5. For convenience of the reader, we state them again.

Recall the independent matrix in Definition 4.2.3. Let $Z_u = ((\varepsilon_{u,i,j}))_{p\times n}$, $1 \leq u \leq U$ be $p \times n$ independent matrices. Therefore, $\{\varepsilon_{u,i,j} : u, i, j \geq 1\}$ are independently distributed with $E(\varepsilon_{u,i,j}) = 0$, $E|\varepsilon_{u,i,j}|^2 = 1$. Recall the classes $\mathcal{L}$ and C respectively in (4.14) and (4.16). We assume that

(A1) $((\varepsilon_{u,i,j})) \in \mathcal{L} \cup C(\delta, p)$ for some $\delta \in (0, 2]$ and for all $1 \leq u \leq U$.

Recall that (A1) was also assumed in Chapter 5. It will be weakened later for some corollaries and applications. If there is only one u i.e., if $U = 1$, we will write $\varepsilon_{i,j}$ and Z respectively for $\varepsilon_{1,i,j}$ and Z_1.

Now we move to the assumptions on the deterministic matrices $\{B_i\}$. The following assumption on $\{B_{2i-1}\}$ was also made in Chapter 5.

(A2) $\{B_{2i-1} : 1 \leq i \leq K\}$ are norm bounded $p \times p$ matrices and $(\text{Span}(B_{2i-1}, B^*_{2i-1} : 1 \leq i \leq K), p^{-1}\text{Tr})$ converges.

Recall the following convergence in (5.13),

$$(\text{Span}\{B_{2i-1}, B^*_{2i-1} : 1 \leq i \leq K\}, p^{-1}\text{Tr}) \;\to\; (\mathcal{A}_{\text{odd}} = \text{Span}\{b_{2i-1}, b^*_{2i-1} : 1 \leq i \leq K\}, \varphi_{\text{odd}}). \quad (6.4)$$

Recall that in Chapter 5, we assumed (A3) for $\{B_{2i}\}$. That assumption is now replaced by the following:

(A3a) $\{B_{2i} : 1 \leq i \leq L\}$ are $n \times n$ matrices with bounded spectral norms. For all $1 \leq i, i' \leq L$, $\epsilon_1, \epsilon_2 = 1$ or $*$, we assume

$$(a)\; -\infty < \lim_{n\to\infty} n^{-1}\text{Tr}(B^{\epsilon_1}_{2i}) < \infty, \quad (b)\; -\infty < \lim_{n\to\infty} n^{-1}\text{Tr}(B^{\epsilon_1}_{2i} B^{\epsilon_2}_{2i'}) < \infty. \quad (6.5)$$

Thus $\{B_{2i}\}$ may not converge jointly. Only moments of polynomials of degree 1 *and* 2 *are assumed to converge.*

6.2.2 Centering and Scaling

To see the necessity of the appropriate centering and scaling on matrix polynomials, let us consider the following example.

Example 1. Let $H = n^{-1}A_1Z_1A_2Z_1^*A_1^*$, where $A_1, A_1^* \in \{B_{2i-1}, B^*_{2i-1}\}$, $A_2 \in \{B_{2i}, B^*_{2i}\}$ and $A_2 = A_2^*$. Recall the convergence in (6.4). Let $\{a_1, a_1^*\} \in$

$\{b_{2i-1}, b_{2i-1}^*\}$ denote the limits of $\{A_1, A_1^*\}$. Let

$$d_0 = \lim n^{-1}\mathrm{Tr}(A_2). \tag{6.6}$$

By (A3a), the right side of (6.6) exists and is finite. Using simple algebra, under (A1), it is easy to see that

$$\begin{aligned}\lim \frac{1}{p} E\mathrm{Tr}(H) &= \left(\lim \frac{1}{p}\mathrm{Tr}(A_1A_1^*)\right)\left(\lim \frac{1}{n}\mathrm{Tr}(A_2)\right) \\ &= \varphi_{\mathrm{odd}}(d_0 a_1 a_1^*), \text{ by (A2), (A3a) and (6.6),} \end{aligned} \tag{6.7}$$

and

$$\begin{aligned}\lim \frac{1}{p} E\mathrm{Tr}(H^2) &= \left(\lim \frac{1}{p}\mathrm{Tr}(A_1A_1^*)^2\right)\left(\lim \frac{1}{n}\mathrm{Tr}(A_2)\right)^2 \\ &\quad + \left(\lim \frac{p}{n}\right)\left(\lim \frac{1}{p}\mathrm{Tr}(A_1A_1^*)\right)^2\left(\lim \frac{1}{n}\mathrm{Tr}(A_2^2)\right) \\ &= \varphi_{\mathrm{odd}}(d_0 a_1 a_1^*)^2, \\ &\qquad \text{by (A2), (A3a) and (6.6) and, as } p/n \to 0. \end{aligned} \tag{6.8}$$

Similarly, under (A1), (A2), (A3a) and if $p/n \to 0$, we have

$$\lim \frac{1}{p} E\mathrm{Tr}(H^h) = \varphi_{\mathrm{odd}}(d_0 a_1 a_1^*)^h, \ \forall h > 2. \tag{6.9}$$

Therefore, H converges to $d_0 a_1 a_1^*$. Consider the matrix $G = n^{-1}\mathrm{Tr}(A_2)A_1A_1^*$. Note that by (A2) and (A3a), G also converges to $d_0 a_1 a_1^*$. Therefore, there is no contribution of the random matrix Z_1 in the limit of H. This is not desirable because such results cannot be used in any statistical application.

To get a non-trivial limit of H, we need appropriate centering and scaling. By (6.7) and as $G \to d_0 a_1 a_1^*$, the appropriate centering for H is G. Now to find the suitable scaling, consider the following computation.

$$\begin{aligned} & \lim p^{-1} E\mathrm{Tr}((H-G)^2) \\ = \; & \lim p^{-1} E\mathrm{Tr}(H)^2 + \lim p^{-1}\mathrm{Tr}(G)^2 - 2\lim p^{-1} E\mathrm{Tr}(HG). \end{aligned} \tag{6.10}$$

Now, under (A1), it is easy to see that

$$\begin{aligned} \frac{1}{p} E\mathrm{Tr}(H^2) &= \left(\frac{1}{p}\mathrm{Tr}(A_1A_1^*)^2\right)\left(\frac{1}{n}\mathrm{Tr}(A_2)\right)^2 \\ &\quad + \left(\frac{p}{n}\right)\left(\frac{1}{p}\mathrm{Tr}(A_1A_1^*)\right)^2\left(\frac{1}{n}\mathrm{Tr}(A_2^2)\right) + O(1/n), \\ p^{-1}\mathrm{Tr}(G^2) &= \left(\frac{1}{n}\mathrm{Tr}(A_2)\right)^2\left(\frac{1}{p}\mathrm{Tr}(A_1A_1^*)^2\right), \\ p^{-1}\mathrm{Tr}(HG) &= \left(\frac{1}{n}\mathrm{Tr}(A_2)\right)^2\left(\frac{1}{p}\mathrm{Tr}(A_1A_1^*)^2\right). \end{aligned} \tag{6.11}$$

Therefore, by (6.10), we have

$$p^{-1} E\mathrm{Tr}((H-G)^2) = \left(\frac{p}{n}\right)\left(\frac{1}{p}\mathrm{Tr}(A_1A_1^*)\right)^2\left(\frac{1}{n}\mathrm{Tr}(A_2^2)\right) + O(1/n).$$

Hence, an appropriate scaling for $(H-G)$ is $\sqrt{np^{-1}}$ and under (A1), (A2), (A3a) and $p/n \to 0$, we have

$$\lim p^{-1} E\mathrm{Tr}(\sqrt{np^{-1}}(H-G))^2 = d_1\varphi_{\text{odd}}(a_1a_1^*)^2, \text{ where } d_1 = \lim \frac{1}{n}\mathrm{Tr}(A_2^2).$$

Moreover, under (A1), (A2), (A3a) and $p/n \to 0$, one can easily see that

$$\lim p^{-1} E\mathrm{Tr}(\sqrt{np^{-1}}(H-G))^4 = 2d_1^2\varphi_{\text{odd}}(a_1a_1^*)^2(\varphi_{\text{odd}}(a_1a_1^*))^2. \tag{6.12}$$

Therefore, the limit of $\sqrt{np^{-1}}(H-G)$ is not trivial. For more precise description of the limit see the next two examples where we shall see the contribution of Z_1

in the limit via freeness. In general $(\mathcal{V}_p, p^{-1}E\text{Tr})$ is the appropriate NCP to work with.

6.2.3 Idea behind the limiting NCP of $(\mathcal{V}_p, p^{-1}E\mathbf{Tr})$

To see how freeness comes into the picture and hence how it motivates the limiting NCP of $(\mathcal{V}_p, p^{-1}E\text{Tr})$, let us focus on the following two examples.

Example 2. Consider the following three polynomials

$$\pi_1 = \sqrt{np^{-1}}\left(n^{-1}A_1Z_1A_2Z_1^*A_3 - n^{-1}\text{Tr}(A_2)A_1A_3\right), \tag{6.13}$$
$$\pi_2 = \sqrt{np^{-1}}\left(n^{-1}A_1Z_2A_2Z_2^*A_3 - n^{-1}\text{Tr}(A_2)A_1A_3\right), \tag{6.14}$$
$$\pi_3 = \sqrt{np^{-1}}\left(n^{-1}A_5Z_1A_6Z_1^*A_7 - n^{-1}\text{Tr}(A_6)A_5A_7\right), \tag{6.15}$$

where $A_1, A_3, A_5, A_7 \in \{B_{2i-1}, B_{2i-1}^* : i \geq 1\}$ and $A_2, A_6 \in \{B_{2i}, B_{2i}^* : i \geq 1\}$. Suppose $\{A_i\}$ are norm bounded matrices. Let us first focus on the marginal convergence of any π_i, say π_1.

As discussed in Definition 4.3.4, convergence of π_1 is equivalent to the convergence of $p^{-1}E\text{Tr}(\Pi(\pi_1, \pi_1^*))$ for all polynomial Π. Using simple matrix algebra, under (A1), one can easily see that

$$\lim \frac{1}{p}E\text{Tr}(\pi_1) = 0,\ \lim \frac{1}{p}E\text{Tr}(\pi_1^2) = \big(\lim \frac{1}{n}\text{Tr}(A_2^2)\big)\big(\lim \frac{1}{p}\text{Tr}(A_1A_3)\big)^2, \tag{6.16}$$
$$\lim \frac{1}{p}E\text{Tr}(\pi_1\pi_1^*) = \big(\lim \frac{1}{n}\text{Tr}(A_2A_2^*)\big)\big(\lim \frac{1}{p}\text{Tr}(A_1A_1^*)\big)\big(\lim \frac{1}{p}\text{Tr}(A_3A_3^*)\big). \tag{6.17}$$

By (A2), $\lim \frac{1}{p}\text{Tr}(A_1A_1^*), \lim \frac{1}{p}\text{Tr}(A_3A_3^*), \lim \frac{1}{p}\text{Tr}(A_1A_3) < \infty$ and by (A3a), $\lim \frac{1}{n}\text{Tr}(A_2^2), \lim \frac{1}{n}\text{Tr}(A_2A_2^*) < \infty$.

Recall (6.4). Let $\{a_1, a_3, a_1^*, a_3^*\} \in \text{Span}\ \{b_{2i-1}, b_{2i-1}^* : i \geq 1\}$ denote the limit of $\{A_1, A_3, A_1^*, A_3^*\}$. Also by (A3a),

$$\lim n^{-1}\text{Tr}(A_2^{\epsilon_1}A_2^{\epsilon_2}) < \infty,\ \forall \epsilon_1, \epsilon_2 = 1, *. \tag{6.18}$$

Recall the free cumulant κ_l of order l in (4.56). By enlarging the NCP of $\{b_{2i-1}, b_{2i-1}^*\}$ if necessary, let w_1 be a variable which is free of $\{a_1, a_3, a_1, a_3^*\}$ and whose all marginal free cumulants of order greater than two are 0 and the first two free cumulants satisfy

$$\kappa_1(w_1^{\epsilon_1}) = 0,\ k_2(w_1^{\epsilon_1}, w_1^{\epsilon_2}) = \lim n^{-1}\mathrm{Tr}(A_2^{\epsilon_1}A_2^{\epsilon_2}),\ \forall \epsilon_1, \epsilon_2 = 1, *. \tag{6.19}$$

Denote the state of the above enlarged NCP by φ_0. Therefore, the restriction of φ_0 on $\{b_{2i-1}, b_{2i-1}^*\}$ is φ_{odd}. Using the algorithm for computing moments of free variables given in Section 4.3.4, one can easily see that

$$\begin{aligned}
\varphi_0(a_1w_1a_3) &\overset{\text{by (4.64)}}{=} 0 \text{ and} && (6.20)\\
\varphi_0(a_1w_1a_3)^2 &\overset{\text{by (4.64)}}{=} (\varphi_{\text{odd}}(a_1a_3))^2 k_2(w_1, w_1)\\
&\overset{\text{by (6.19)}}{=} (\varphi_{\text{odd}}(a_1a_3))^2\big(\lim \frac{1}{n}\mathrm{Tr}(A_2^2)\big)\\
&= \big(\lim \frac{1}{n}\mathrm{Tr}(A_2^2)\big)\big(\lim \frac{1}{p}\mathrm{Tr}(A_1A_3)\big)^2,\\
\varphi(a_1w_1a_3a_3^*w_1^*a_1^*) &\overset{\text{by (4.64)}}{=} (\varphi_{\text{odd}}(a_1a_1^*))(\varphi_{\text{odd}}(a_3a_3^*))k_2(w_1, w_1^*)\\
&\overset{\text{by (6.19)}}{=} (\varphi_{\text{odd}}(a_1a_1^*)\varphi_{\text{odd}}(a_3a_3^*))\big(\lim \frac{1}{n}\mathrm{Tr}(A_2A_2^*)\big)\\
&= \big(\lim \frac{1}{n}\mathrm{Tr}(A_2A_2^*)\big)\big(\lim \frac{1}{p}\mathrm{Tr}(A_1A_1^*)\big)\big(\lim \frac{1}{p}\mathrm{Tr}(A_3A_3^*)\big). && (6.21)
\end{aligned}$$

Therefore by (6.16), (6.17) and (6.20)-(6.21),

$$\begin{aligned}
\lim \frac{1}{p}E\mathrm{Tr}(\pi_1) &= \varphi_0(a_1w_1a_3), && (6.22)\\
\lim p^{-1}E\mathrm{Tr}(\pi_1)^2 &= \varphi_0(a_1w_1a_3)^2 \text{ and} && (6.23)\\
\lim p^{-1}E\mathrm{Tr}(\pi_1\pi_1^*) &= \varphi_0((a_1w_1a_3)(a_1w_1a_3)^*). && (6.24)
\end{aligned}$$

Similarly, one can also show that, for all $T \geq 1$ and $\epsilon_1, \epsilon_2, \ldots, \epsilon_T = 1, *$, we have

$$\lim p^{-1}E\mathrm{Tr}(\pi_1^{\epsilon_1}\pi_2^{\epsilon_2}\ldots\pi_T^{\epsilon_T}) = \varphi_0((a_1w_1a_3)^{\epsilon_1}(a_1w_1a_3)^{\epsilon_2}\ldots(a_1w_1a_3)^{\epsilon_T}). \tag{6.25}$$

Therefore,

$$(\text{Span}\{\pi_1, \pi_1^*\}, p^{-1}E\text{Tr}) \to (\text{Span}\{\alpha_1, \alpha_1^*\}, \varphi_0), \quad \text{where } \alpha_1 := a_1 w_1 a_3. \tag{6.26}$$

Note that the right side of the above equations (6.22), (6.23) and (6.25) can be in principle computed by using the distribution of $\{a_1, a_3, a_1^*, a_3^*\}$, the distribution of w_1 and, the freeness of these two collections.

Moreover by (A3a),

$$\lim n^{-1}\text{Tr}(A_6^{\epsilon_1} A_6^{\epsilon_2}) < \infty, \ \forall \epsilon_1, \epsilon_2 = 1, *. \tag{6.27}$$

Then under (A1), (A2) and (A3a), one can similarly see that,

$$\begin{aligned}
(\text{Span}\{\pi_2, \pi_2^*\}, p^{-1}E\text{Tr}) &\to (\text{Span}\{\alpha_2, \alpha_2^*\}, \varphi_0), \quad \alpha_2 := a_1 w_2 a_3, & (6.28)\\
(\text{Span}\{\pi_2, \pi_2^*\}, p^{-1}E\text{Tr}) &\to (\text{Span}\{\alpha_3, \alpha_3^*\}, \varphi_0), \quad \alpha_3 := a_5 w_3 a_7 & (6.29)
\end{aligned}$$

where $\{a_5, a_7, a_5^*, a_7^*\}$ is the limit of $\{A_5, A_7, A_5^*, A_7^*\}$ and w_2, w_3 have exactly the same free cumulants given in (6.19) as w_1 except A_2 is replaced by A_6 for w_3.

Now we shall discuss the convergence of $(\text{Span}\{\pi_1, \pi_2, \pi_3, \pi_1^*, \pi_2^*, \pi_3^*\}, p^{-1}E\text{Tr})$. By (A3a),

$$\lim n^{-1}\text{Tr}(A_2^{\epsilon_1} A_6^{\epsilon_2}) < \infty, \quad \forall \epsilon_1, \epsilon_2 = 1, *. \tag{6.30}$$

Suppose the marginal cumulants of $\{w_1, w_2, w_3\}$ are as before and the joint cumulants are as follows.

$$\begin{aligned}
&\kappa_r(w_{i_1}^{\epsilon_1}, w_{i_2}^{\epsilon_2}, \ldots, w_{i_r}^{\epsilon_r}) = 0, \ \forall r > 2, \ i_1, i_2 \ldots i_r = 1, 2, 3, \ \epsilon_1, \epsilon_2, \ldots, \epsilon_r = 1, *.\\
&\kappa_2(w_1^{\epsilon_1}, w_2^{\epsilon_2}) = \kappa_2(w_2^{\epsilon_1}, w_3^{\epsilon_2}) = 0 \text{ and } \kappa_2(w_1^{\epsilon_1}, w_3^{\epsilon_2}) = \lim n^{-1}\text{Tr}(A_2^{\epsilon_1} A_6^{\epsilon_2}). \quad (6.31)
\end{aligned}$$

Using arguments similar to those in the marginal cases, under (A1), (A2) and

(A3a), one can show that

$$\begin{aligned}
\lim p^{-1} E\mathrm{Tr}(\pi_1\pi_2) &= \lim p^{-1} E\mathrm{Tr}(\pi_2\pi_3) = 0 = \varphi_0(\alpha_1\alpha_2) = \varphi_0(\alpha_2\alpha_3) \text{ and} \\
\lim p^{-1} E\mathrm{Tr}(\pi_1\pi_3) &= \big(\lim n^{-1}\mathrm{Tr}(A_2A_6)\big)\big(\lim p^{-1}\mathrm{Tr}(A_1A_7)\big)\big(\lim p^{-1}\mathrm{Tr}(A_3A_5)\big) \\
&= \varphi_0(\alpha_1\alpha_3).
\end{aligned}$$

Moreover, one can indeed show that for $T \geq 1$, $\epsilon_1, \epsilon_2, \ldots, \epsilon_T = 1, *$ also

$$\lim p^{-1} E\mathrm{Tr}(\pi_{i_1}^{\epsilon_1}\pi_{i_2}^{\epsilon_2}\ldots\pi_{i_T}^{\epsilon_t}) = \varphi_0(\alpha_{i_1}^{\epsilon_1}\alpha_{i_2}^{\epsilon_2}\ldots\alpha_{i_T}^{\epsilon_T}),\ i_1, i_2, \ldots \in \{1, 2, 3\}.$$

Hence,

$$(\mathrm{Span}\{\pi_1, \pi_2, \pi_3, \pi_1^*, \pi_2^*, \pi_3^*\}, p^{-1}E\mathrm{Tr}) \ \to (\mathrm{Span}\{\alpha_1, \alpha_2, \alpha_3, \alpha_1^*, \alpha_2^*, \alpha_3^*\}, \varphi_0). \tag{6.32}$$

Example 3. This example is on a larger sized polynomial. Let

$$\begin{aligned}
S_1 &= A_1Z_1A_2Z_1^*A_3, \qquad S_2 = A_5Z_1A_6Z_1^*A_7, \\
G_1 &= \big(n^{-1}\mathrm{Tr}(A_2)\big)A_1A_3, \qquad G_2 = \big(n^{-1}\mathrm{Tr}(A_6)\big)A_5A_7, \\
g_1 &= \lim n^{-1}\mathrm{Tr}(A_2)a_1a_3, \quad g_2 = \lim n^{-1}\mathrm{Tr}(A_6)a_5a_7.
\end{aligned}$$

Consider the polynomial

$$\pi_4 = \sqrt{np^{-1}}(S_1S_2 - G_1G_2) \tag{6.33}$$

(which corresponds to the case $k_l = 2$ in $\mathcal{V}_p$). Now

$$(\mathrm{Span}\{G_1, G_2, G_1^*, G_2^*\}, p^{-1}E\mathrm{Tr}) \to (\mathrm{Span}\{g_1, g_2, g_1^*, g_2^*\}, \varphi_{\mathrm{odd}}). \tag{6.34}$$

Recall π_1 and π_3 respectively in (6.13) and (6.14). Note that

$$\pi_1 = \sqrt{np^{-1}}(S_1 - G_1) \quad \text{and} \quad \pi_3 = \sqrt{np^{-1}}(S_2 - G_2).$$

To understand the convergence of π_4, observe that

$$\pi_4 = \pi_1 G_2 + G_1 \pi_2 + \sqrt{pn^{-1}}\pi_1\pi_2. \tag{6.35}$$

Hence, by the previous example,

$$\begin{aligned}
\pi_4 &\rightarrow \alpha_1 g_2 + g_1 \alpha_2 + 0.\alpha_1\alpha_2 \text{ (since } p/n \rightarrow 0) \\
&= a_1 w_1 a_3 g_2 + g_1 a_5 w_3 a_7 \\
&= a_1 w_1 a_3 \big(\lim n^{-1}\mathrm{Tr}(A_6)\big) a_5 a_7 + \big(\lim n^{-1}\mathrm{Tr}(A_2)\big) a_1 a_3 a_5 w_3 a_7 \qquad (6.36)
\end{aligned}$$

where (w_1, w_3) are as in the previous example.

As a prelude to using the argument in the above example more generally, note that each summand has only one w variable. The expression in (6.36) can be visualized as follows.

$$\pi_4 \approx S_1 S_2 \approx A_1(Z_1 A_2 Z_1^*)A_3 A_5(Z_1 A_6 Z_1^*)A_7. \tag{6.37}$$

Each pair (Z_i, Z_i^*) gives rise to a w variable. For example, the first pair gives a w_1 and it contributes $a_1 w_1 a_3 \big(\lim n^{-1}\mathrm{Tr}(A_6)\big) a_5 a_7$. Similarly, the second pair gives a w_3 and it contributes $\big(\lim n^{-1}\mathrm{Tr}(A_2)\big) a_1 a_3 a_5 w_3 a_7$. Then the limit is the *sum* of these two (see (6.36)). Later we shall refer to variables on the left and right of any w as a and c variables respectively. For example (6.36) can be written as

$$\pi_4 \rightarrow a_{-1} w_1 c_{-1} + a_{-2} w_3 c_{-2} \text{ (say)},$$

where $a_{-1} = a_1$, $c_{-1} = a_3\big(\lim n^{-1}\text{Tr}(A_6)\big)a_5a_7$, $a_{-2} = \big(\lim n^{-1}\text{Tr}(A_2)\big)a_1a_3a_5$, $c_{-2} = a_7$. Moreover, from the above examples it is intuitively clear why the centered polynomials $\{\sqrt{np^{-1}}(\mathbb{P} - \mathbb{G}) = \mathcal{R}\}$ should also converge jointly and what would be their limits. We make these ideas precise in the next section.

6.3 NCP convergence result

To describe the limit, define a family of variables $\mathcal{T} = \{w_{u,l,i} : u, l, i \geq 1\}$ (note that $w_{u,l,i}$ is attached to the matrix $A_{l,2i}$ and Z_u-index u), where for all $l_j, u_j, i_j \geq 1$, $\epsilon_j = 1, *, \forall j \geq 1$,

$$\kappa_r(w^{\epsilon_j}_{u_j,l_j,i_j} : 1 \leq j \leq r) \;=\; \begin{cases} \lim n^{-1}\text{Tr}(A^{\epsilon_1}_{l_1,2i_1}A^{\epsilon_2}_{l_2,2i_2}), \text{ if } r = 2 \text{ and } u_1 = u_2 \\ 0, \text{ if } r \neq 2 \text{ or } u_1 \neq u_2, \end{cases}$$

and they are free of $\{b_{2i-1}, b^*_{2i-1} : 1 \leq i \leq K\}$. In the language of free probability, $\mathcal{A}_u = \{w_{u,l,i} : l, i \geq 1\}$, $1 \leq u \leq U$ are free. The above sequence of free cumulants naturally defines a state on $\text{Span}\{w_{u,l,i} : u, l, i \geq 1\}$, say φ_w.

Two special cases are worth mentioning. Recall Definitions 4.3.5 and 4.3.6 of the semi-circle family and the circular family of non-commutative variables. If B_{2i}, $i \geq 1$ are self-adjoint, then each $w_{u,l,j}$ can be taken to be self-adjoint and $\mathcal{T}$ is a semi-circle family. On the other hand, if $\lim n^{-1}\text{Tr}(B^2_{2i}) = \lim n^{-1}\text{Tr}(B^{*2}_{2i}) = 0$, $\forall i \geq 1$, then $\mathcal{T}$ is a circular family.

Now we formalize the definition of the left and the right variables a and c. Motivated by the ideas given towards the end of Section 6.2.3, in general, for $\mathcal{V}_p$, let for all $l \geq 1$ and $1 \leq j \leq k_l$,

$$a_{l,-j} \;=\; \Big(\prod_{i=1}^{j-1}\lim \frac{1}{n}\text{Tr}(A_{l,2i})\Big)\Big(\prod_{i=0}^{j-1} a_{l,2i+1}\Big), \tag{6.38}$$

$$c_{l,-j} \;=\; \Big(\prod_{i=j+1}^{k_l}\lim \frac{1}{n}\text{Tr}(A_{l,2i})\Big)\prod_{i=j}^{k_l} a_{l,2i+1}, \tag{6.39}$$

$$\alpha_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} = \sum_{j=1}^{k_l} a_{l,-j} w_{u_{l,j},l,j} c_{l,-j}. \tag{6.40}$$

Let,

$$(\mathcal{B}, \varphi_0) = \text{free product of } (\mathcal{A}_{\text{odd}}, \varphi_{\text{odd}}) \text{ and } (\text{Span}\{w_{u,l,i} : u, l, i \geq 1\}, \varphi_w). \tag{6.41}$$

Consider the following $*$-sub-algebra of $\mathcal{B}$,

$$\mathcal{V} = \text{Span}\big(\alpha_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : \ \forall u_{l,j} \in \{1, 2, 3\ldots\}, \ l \geq 1\big).$$

Now we have all the ingredients to state the following joint convergence theorem (see Bhattacharjee and Bose [2015b]). Proof of this theorem is very technical. We provide the proof later in Section 6.6. This result is the corner stone to obtain the LSD of symmetric polynomials in $\{\mathcal{R}_{l,(u_{l,1},\ldots,u_{l,k_l})}\}$ in the next section.

Theorem 6.3.1. *Suppose Assumptions* (A1), (A2) *and* (A3a) *hold and* $p, n(p) \to \infty$, $p/n \to 0$. *Then*

(*a*) $(\mathcal{V}_p, Ep^{-1}\text{Tr}) \to (\mathcal{V}, \varphi_0)$.

(*b*) $\big(Span\{\sqrt{np^{-1}}(n^{-1}Z_{u_j}B_{2j}Z_{u_j}^* - n^{-1}\text{Tr}(B_{2j})) : u_j \in \{1, 2, \ldots\}, j \geq 1\}, p^{-1}E\text{Tr}\big)$ *and* $(Span(B_{2i-1} : 1 \leq i \leq K), p^{-1}\text{Tr})$ *are asymptotically free.*

Remark 6.3.2. *It is easy to see that φ_0 is tracial and positive (see (4.36), (4.37) and (4.40)).*

6.4 LSD of symmetric polynomials in $\{\mathcal{R}_{l,(u_{l,1},\ldots,u_{l,k_l})}\}$

The following theorem (see Bhattacharjee and Bose [2015b]) guarantees the existence of the LSD of any symmetric polynomial in $\{\mathcal{R}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})}\}$.

Theorem 6.4.1. *Suppose Assumptions* (A1), (A2) *and* (A3a) *hold and* $p, n(p) \to \infty$, $p/n \to 0$. *Then the LSD of any self-adjoint polynomial* $\mathbb{P}(\mathcal{R}_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} :$

$1 \le l \le r)$ *in* $\mathcal{V}_p$ *exists with probability* 1 *and it is given by* $\mathbb{P}(\alpha_{l,(u_{l,1},u_{l,2},\ldots,u_{l,k_l})} : 1 \le l \le r)$.

Proof. To prove the theorem, by Lemma 4.2.1, we need to establish the conditions (M1), (M4) and (C) as described in the moment method in Section 4.2. The (M1) condition is immediate from Theorem 6.3.1. Proof of (M4) and (C) go through the same lines as the proof of (M4) and (C) in the proof Theorem 5.4.1. We omit the similar and tedious technical details. Hence the proof of Theorem 6.4.1 is complete. □

6.4.1 Stieltjes transform of the LSD

By utilizing Assumptions $(A2)$ and $(A3a)$, we can verify that the self-adjoint elements in $\mathcal{V}$ have moments with nice bounds. Hence they uniquely define proper probability distributions of usual bounded random variables. In principle we know how to calculate the moments of these variables.

On the other hand, the existing LSD results in the literature are mostly in terms of the Stieltjes transform. To show how these existing results follow from Theorem 6.4.1, we need to express our LSD results in terms of Stieltjes transforms. So we first establish a general Stieltjes transform result.

Let

$$\gamma = \sum_{j=1}^{r} (a_j w_j c_j + c_j^* w_j^* a_j^*), \tag{6.42}$$

where $\{w_j, w_j^*\} \subset \{w_{u,l,i} : u, l, i \ge 1\}$ is a family of non-commutative variables which satisfy

$$\kappa_r(w_{j_i}^{\epsilon_i} : 1 \le i \le r) = \begin{cases} b_{j_1,j_2,\epsilon_1,\epsilon_2}, \text{ if } r = 2 \\ 0, \text{ if } r \ne 2, \end{cases} \tag{6.43}$$

for all $j_i \ge 1$, $\epsilon_i = 1, *$, $i \ge 1$, $\{a_j, c_j\} \subset \{b_{2i-1}\}$ and $\{a_j\}$ is some permutation of $\{c_j\}$. Further $\{w_j, w_j^*\}$ and $\{a_j, a_j^*\}$ are freely independent. Recall the state φ_0 in

(6.41), whose restriction on $\mathcal{A}_{\text{odd}}$ is φ_{odd}.

The general Stieltjes transform formula given below looks messy. However, when we specialize to specific choices of elements in $\{\mathcal{R}_{l,(u_{l,1},\ldots,u_{l,k_l})}\}$, their limit γ have specific form and the formulae will be significantly simplified. We shall deal with some special cases in the next section.

Theorem 6.4.2. *For $z \in \mathbb{C}^+$, $|z|$ large, the Stieltjes transform of γ is given by*

$$
\begin{aligned}
m_\gamma(z) &= -\varphi_0((z+\beta(z,a))^{-1}), && (6.44)\\
&= -\varphi_{\text{odd}}((z+\beta(z,a))^{-1}), \ \textit{where} && (6.45)\\
\beta(z,a) &= -\sum_{j_1,j_2=1}^{r}\Big(b_{j_1,j_2,1,1}c_{j_1}a_{j_2}\varphi_0(\frac{a_{j_1}c_{j_2}}{z+\beta(z,a)}) + b_{j_1,j_2,1,*}c_{j_1}c^*_{j_2}\varphi_0(\frac{a_{j_1}a^*_{j_2}}{z+\beta(z,a)})\Big)\\
&\quad -\sum_{j_1,j_2=1}^{r}\Big(b_{j_1,j_2,*,1}a^*_{j_1}a_{j_2}\varphi_0(\frac{c^*_{j_1}c_{j_2}}{z+\beta(z,a)}) + b_{j_1,j_2,*,*}a^*_{j_1}c^*_{j_2}\varphi_0(\frac{c^*_{j_1}a^*_{j_2}}{z+\beta(z,a)})\Big) && (6.46)\\
&= \textit{same expression with } \varphi_0 \textit{ replaced by } \varphi_{\text{odd}}.
\end{aligned}
$$

Here $(z+\beta(z,a))^{-1} = z^{-1}\sum_{i=0}^{\infty} z^{-i}(-\beta(z,a))^i$.

Using the same arguments as in Lemma 5.4.4, it is easy to see that the power series above are all meaningful for large $|z|$.

Proof of Theorem 6.4.2. Throughout, $|z|$ is assumed to be sufficiently large for any relevant expression to be meaningful.

For all $i \geq 1$, define

$$
\begin{aligned}
R_i &= \sum_{j_1,j_2=1}^{r} b_{j_1,j_2,1,1}c_{j_1}a_{j_2}\varphi_0(a_{j_1}c_{j_2}\gamma^{i-1}) + \sum_{j_1,j_2=1}^{r} b_{j_1,j_2,1,*}c_{j_1}c^*_{j_2}\varphi_0(a_{j_1}a^*_{j_2}\gamma^{i-1})\\
&\quad + \sum_{j_1,j_2=1}^{r} b_{j_1,j_2,*,1}a^*_{j_1}a_{j_2}\varphi_0(c^*_{j_1}c_{j_2}\gamma^{i-1}) + \sum_{j_1,j_2=1}^{r} b_{j_1,j_2,*,*}a^*_{j_1}c^*_{j_2}\varphi_0(c^*_{j_1}a^*_{j_2}\gamma^{i-1}), && (6.47)\\
\beta(z,a) &= -\sum_{i=1}^{\infty} z^{-i}R_i. && (6.48)
\end{aligned}
$$

Note that $\varphi_0(\gamma^{2h-1}) = 0$ and $\varphi_0(R_{2h}) = 0,\ \forall h \geq 1$. By (4.73), we have

$$\begin{aligned}
\varphi_0(\gamma^{2h}) &= \sum_{\epsilon_1,\epsilon_2,\ldots,\epsilon_{2h}} \sum_{j_1,j_2,\ldots,j_{2h}} \varphi_0(\prod_{k=1}^{2h} a_{j_k}^{\epsilon_k} w_{j_k}^{\epsilon_k} c_{j_k}^{\epsilon_k}) \\
&= \sum_{\epsilon_1,\epsilon_2,\ldots,\epsilon_{2h}} \sum_{j_1,j_2,\ldots,j_{2h}} \sum_{\pi \in NC_2(2h)} \varphi_{0K(\pi)}(c_{j_1}a_{j_2}, c_{j_2}a_{j_3}, \ldots, c_{j_{2h}}a_{j_1}) k_\pi(w_{j_1}, \ldots, w_{j_{2h}}).
\end{aligned} \tag{6.49}$$

For any subset A of $NC(n)$, by contribution of A in $\varphi_0(\gamma^{2h})$, we mean

$$\sum_{\epsilon_1,\epsilon_2,\ldots,\epsilon_{2h}} \sum_{j_1,j_2,\ldots,j_{2h}} \sum_{\pi \in A} \varphi_{0K(\pi)}(c_{j_1}a_{j_2}, c_{j_2}a_{j_3}, \ldots, c_{j_{2h}}a_{j_1}) k_\pi(w_{j_1}, \ldots, w_{j_{2h}}).$$

To simplify (6.49), consider the following decomposition of $NC_2(2h)$.

$$\begin{aligned}
NC_2(2h) &= \cup_{i=1}^h \mathcal{C}_{i,h}, \text{ where} \\
\mathcal{C}_{i,h} &= \text{set of all } \sigma \in NC_2(2h) \text{ such that } \{1, 2i\} \in \sigma.
\end{aligned}$$

Note that the contribution of $\{\{1,2h\},\{2,3\},\{4,5\},\ldots,\{2h-2,2h-1\}\} \in \mathcal{C}_{h,h}$ to right side of (6.49), is $\varphi_0(R_1^h)$. Now,

$$\begin{aligned}
\varphi_0(\gamma^2) &= \text{contribution of } \mathcal{C}_{1,1} \text{ in } \varphi_0(\gamma^2) \\
&= \varphi_0(R_1).
\end{aligned}$$

Again,

$$\begin{aligned}
\varphi_0(\gamma^4) &= \text{contribution of } \mathcal{C}_{1,2} \text{ in } \varphi_0(\gamma^4) + \text{contribution of } \mathcal{C}_{2,2} \text{ in } \varphi_0(\gamma^4) \\
&= \varphi_0(R_3 + R_1^2).
\end{aligned}$$

Next,

$$\varphi_0(\gamma^6) = \text{contribution of } \mathcal{C}_{1,3} \text{ in } \varphi_0(\gamma^6) + \text{contribution of } \mathcal{C}_{2,3} \text{ in } \varphi_0(\gamma^6)$$

$$\begin{aligned} &+\text{contribution of } \mathcal{C}_{3,3} \text{ in } \varphi_0(\gamma^6) \\ =\ & \varphi_0(R_5 + R_1R_3 + (R_3R_1 + R_1^3)). \end{aligned}$$

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

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

Then, one can show easily by induction on h that

$$\varphi_0(\gamma^{2h}) = \varphi_0\left(\sum_{t=1}^{h} \sum_{i_1,i_2,\ldots,i_t \in S_{2h-t,t}} \prod_{j=1}^{t} R_{i_j}\right), \forall h \geq 1.$$

We omit the tedious details. Hence, using the power series expansion (4.7) for Stieltjes transformation, we have

$$\begin{aligned} m_\gamma(z) &= \varphi_0((\gamma - z)^{-1}) = -z^{-1}\sum_{h=0}^{\infty} z^{-2h}\varphi_0(\gamma^{2h}) && (6.50) \\ &= -\varphi_0\Big(z^{-1}\sum_{h=0}^{\infty} z^{-2h} \sum_{t=1}^{h} \sum_{i_1,i_2,\ldots,i_t \in S_{2h-t,t}} \prod_{j=1}^{t} R_{i_j}\Big) \\ &= -\varphi_0\Big(z^{-1}\sum_{t=0}^{\infty} z^{-t} \sum_{h=t}^{\infty} \sum_{i_1,i_2,\ldots,i_t \in S_{2h-t,t}} \prod_{j=1}^{t} z^{-i_j} R_{i_j}\Big) \\ &= -\varphi_0\Big(z^{-1}\sum_{t=0}^{\infty} z^{-t} \Big(\sum_{i=1}^{\infty} z^{-i} R_i\Big)^t\Big) \\ &= -\varphi_0\left(z^{-1}\sum_{t=0}^{\infty} z^{-t}(-\beta(z,a))^t\right) = -\varphi_0((z + \beta(z,a))^{-1}). && (6.51) \end{aligned}$$

Similarly, one can easily show by induction on h and the assumption $\{a_j : j =$

$1, 2, \ldots, r\} = \{c_j : j = 1, 2, \ldots, r\}$, that

$$\begin{aligned}
R_{2h+1} \quad = \quad & \sum_{j_1,j_2=1}^{r} b_{j_1,j_2,1,1} a_{j_1} c_{j_2} \varphi_0\Big(c_{j_1} a_{j_2} \sum_{t=1}^{h} \sum_{i_1,i_2,\ldots,i_t \in S_{2h-t,t}} \prod_{j=1}^{t} R_{i_j}\Big) \\
& + \sum_{j_1,j_2=1}^{r} b_{j_1,j_2,1,*} a_{j_1} a^*_{j_2} \varphi_0\Big(c_{j_1} c^*_{j_2} \sum_{t=1}^{h} \sum_{i_1,i_2,\ldots,i_t \in S_{2h-t,t}} \prod_{j=1}^{t} R_{i_j}\Big) \\
& + \sum_{j_1,j_2=1}^{r} b_{j_1,j_2,*,1} c^*_{j_1} c_{j_2} \varphi_0\Big(a^*_{j_1} a_{j_2} \sum_{t=1}^{h} \sum_{i_1,i_2,\ldots,i_t \in S_{2h-t,t}} \prod_{j=1}^{t} R_{i_j}\Big) \\
& + \sum_{j_1,j_2=1}^{r} b_{j_1,j_2,*,*} c^*_{j_1} a^*_{j_2} \varphi_0\Big(a^*_{j_1} c^*_{j_2} \sum_{t=1}^{h} \sum_{i_1,i_2,\ldots,i_t \in S_{2h-t,t}} \prod_{j=1}^{t} R_{i_j}\Big).
\end{aligned}$$

Now (6.46) is immediate from the power series expansion of $\beta(z, a)$ in (6.48) and using calculations similar to (6.50)-(6.51). Hence the proof of Theorem 6.4.2 is complete. $\square$

6.4.2 Corollaries

The next corollaries and remarks discuss some special cases, which follow from Theorems 6.3.1, 6.4.1 and 6.4.2. Some of them imply the existing results stated in Theorems 4.2.8-4.2.11 while others are new results. Recall the classes $\mathcal{NND}$ and $\mathcal{N}$ respectively in (4.18) and (4.29).

Corollary 6.4.3. *Let $Z_{p\times n}$ be an independent matrix whose entries satisfy (A1). Let A_p and B_n be norm bounded deterministic matrices. Suppose $\{A_p\} \in \mathcal{NND}$ with LSD F^A. Suppose $\{B_n\} \in \mathcal{N}$ and $\lim n^{-1} Tr(B^2) = d_2$. Let (a, s) be free in some NCP $(\mathcal{B}, \varphi_0)$ where $a \sim F^A$ and s is a standard semi-circle variable. Suppose $p, n(p) \to \infty$, $p/n \to 0$. Then*

(a) $\Big(Span\{ \sqrt{\frac{n}{p}} (\frac{1}{n} A^{1/2} Z B Z^* A^{1/2} - \frac{1}{n}\mathrm{Tr}(B) A)\}, \frac{1}{p} E\mathrm{Tr} \Big) \to \big(Span\{a^{1/2}\sqrt{d_2} s a^{1/2}\}, \varphi_0\big)$

(b) The LSD of $\sqrt{np^{-1}}(n^{-1}A^{1/2}\ ZBZ^*A^{1/2} - n^{-1}\mathrm{Tr}(B)A)$ exists almost surely and it is distributed as $a^{1/2}\sqrt{d_2} s a^{1/2}$ whose Stieltjes transform satisfies the pair of equations (4.30) and (4.31). Here (a, s) is as in (a) above.

Proof. As the assumptions on A_p and B_n in the above corollary respectively satisfy (A2) and (A3a), (a) follows immediately from Theorem 6.3.1 (a). Also the first part of (b) follows from Theorem 6.4.1.

To derive the Stieltjes transform, note that by (6.42), we have $a^{1/2}\sqrt{d_2}sa^{1/2} = \gamma$, where

$$a_1 = a_1^* = c_1 = c_1^* = \frac{1}{\sqrt{2}}a^{1/2},\ w_1 = w_1^* = s,\ a_j = c_j = w_j = 0\ \forall j > 1. \quad (6.52)$$

Also by (6.43)

$$b_{j_1,j_2,\epsilon_1,\epsilon_2} = \kappa_2(w_{j_1}^{\epsilon_1}, w_{j_2}^{\epsilon_2}) = \begin{cases} \kappa_2(\sqrt{d_2}s, \sqrt{d_2}s) = d_2, & \forall \epsilon_1, \epsilon_2 = 1 \text{ or } *,\ j_1 = j_2 = 1 \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, for $z \in \mathbb{C}^+$ and large $|z|$, (6.46) reduces to

$$\begin{aligned} \beta(z,a) &= -4\left[d_2\frac{a}{2}\varphi_0\left(\frac{a}{2}(z+\beta(z,a))^{-1}\right)\right] \\ &= -d_2a\varphi_0(a(z+\beta(z,a))^{-1}) = d_2ag(z), \text{ say,} \end{aligned} \quad (6.53)$$

where

$$g(z) = -\varphi_0(a(z+\beta(z,a))^{-1}) = -\varphi_0(a(z+d_2ag(z))^{-1}) = -\int_{\mathbb{R}} \frac{tdF^A(t)}{z+d_2tg(z)}.$$

Hence, (4.31) is established. Now, for $z \in \mathbb{C}^+$ and large $|z|$, (6.44) reduces to

$$m_\gamma(z) = -\varphi_0((z+\beta(z,a))^{-1}) = -\varphi_0((z+d_2ag(z))^{-1}) = -\int_{\mathbb{R}} \frac{dF^A(t)}{z+d_2tg(z)}.$$

This established (4.30) for large $|z|$. Now note that both sides of (4.30) are analytic. Hence, using analyticity, (4.30) continues to hold for all $z \in \mathbb{C}^+$. $\square$

Recall the classes $\mathcal{L}_4$ and $U(\delta)$ respectively in (4.14) and (4.17). Consider the following weak assumptions on the entries of the independent matrix Z.

(A4) $\{\varepsilon_{i,j} : i, j \geq 1\} \in \mathcal{L}_4 \cap U(\delta)$ for some $\delta \in (0, 2]$.

(A5) $\{\varepsilon_{i,j} : i, j \geq 1\}$ are *i.i.d.* with mean 0, variance 1 and $E|\varepsilon_{i,j}|^4 < \infty$.

Remark 6.4.4. *The LSD result in Corollary 6.4.3 continues to hold if we replace (A1) by (A4) or (A5). This relaxation is possible by first observing that Theorem 6.4.1 is applicable to appropriately truncated variables and then using a suitable metric to estimate the distance between the ESD of the original and the truncated version. We omit the tedious details of this argument, specially because the proof of Corollary 7.3.18* (c) *given later is also along the same lines. In addition, we can drop the norm boundedness assumption on A_p by truncating the ESD of the matrix A. For details of this argument see Section* 3.1 *of Wang and Paul [2014].*

Thus we have proved Theorem 4.2.11 (Wang et al. [2015]) as a special case of our results. If B is taken to be the identity matrix, this implies Theorem 4.2.9 (Bao [2012]). Of course, if both A and B are taken to be the identity matrices, this implies Theorem 4.2.8 (Bai and Yin [1988]).

Recall the compound free Poisson distribution in Definition 4.3.9.

Corollary 6.4.5. *Suppose all the assumptions in Corollary 6.4.3 hold. Then the LSD of $\sqrt{np^{-1}}(n^{-1}ZB\ Z^* - \frac{1}{n}\mathrm{Tr}(B))A\ \sqrt{np^{-1}}(n^{-1}ZBZ^* - \frac{1}{n}\mathrm{Tr}(B))$ is the compound free Poisson distribution with rate* 1, *and jump distribution same as the distribution of $d_2 a$.*

Proof. From the discussions around Definition 4.3.9, it is clear that if a semi-circle variable s and another variable a are freely independent, then for any constant $c > 0$, $\sqrt{c}sa\sqrt{c}s$ has the compound free Poisson distribution with rate 1 and jump distribution ca. Therefore, Corollary 6.4.5 is immediate since by Theorem 6.3.1 (a), $\sqrt{np^{-1}}(n^{-1}ZBZ^* - \frac{1}{n}\mathrm{Tr}(B))$ converges to $\sqrt{d_2}s$, where s is the standard semi-circle variable and by Theorem 6.3.1 (b), s and a are freely independent. □

The following corollary will be used later in Chapter 7, when we shall deal with the LSD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$. Recall $\{\Delta_u\}$ in (5.1). We consider the same assumptions on

$\{\psi_j\}$ as in Corollary 5.4.7. For convenience of the reader, we state it again.

Suppose $\{\psi_j\} \subset \{B_{2i-1}, B_{2i-1}^*\}$ i.e. we assume:

(B) $\{\psi_j\}$ are norm bounded and

$$(\text{Span}\{\psi_j, \psi_j^* : j \geq 0\}, p^{-1}\text{Tr}) \to (\text{Span}\{\eta_j, \eta_j^* : j \geq 0\}, \varphi_{\text{odd}}) \text{ (say)}. \tag{6.54}$$

Recall the NCP $(\mathcal{A}_{\text{odd}}, \varphi_{\text{odd}})$ in (6.4). Clearly the NCP in the right side of (6.54) is a $*$-sub-algebra of $(\mathcal{A}_{\text{odd}}, \varphi_{\text{odd}})$. Recall that φ_0 is the state corresponding to the free product given in (6.41). Therefore, by Definition 4.3.8, the restriction of φ_0 on $\mathcal{A}_{\text{odd}}$ is φ_{odd}.

To describe the Stieltjes transform below, for $x = (x_1, \ldots, x_q)$, $x_l \in \mathcal{A}_{\text{odd}}\ \forall l$, we define

$$\Psi(x, \theta) = \left(\sum_{l=0}^{q} x_l e^{il\theta}\right)\left(\sum_{l=0}^{q} x_l^* e^{-il\theta}\right) = \sum_{l_1, k_1 = 0}^{q} x_{l_1} x_{k_1}^* e^{i(l_1 - k_1)\theta}. \tag{6.55}$$

For $x = (x_1, \ldots, x_q), y = (y_1, y_2, \ldots, y_q)$, $x_l, y_l \in \mathcal{A}_{\text{odd}}\ \forall l$, we define

$$\begin{aligned}
R_u(x, y) &= \frac{1}{2\pi}\int_0^{2\pi} \cos^2(u\theta)\Psi(x,\theta)\Psi(y,\theta)d\theta \qquad (6.56)\\
&= \sum_{l_1,l_2,k_1,k_2=0}^{q} x_{l_1}x_{k_1}^* y_{l_2} y_{k_2}^* \frac{1}{2\pi}\int_0^{2\pi} \cos^2(u\theta) e^{i(l_1-k_1+l_2-k_2)\theta} d\theta.\\
&= 0.5 \sum_{l_1,l_2,k_1,k_2=0}^{q} x_{l_1}x_{k_1}^* y_{l_2} y_{k_2}^* I(l_1-k_1+l_2-k_2=0)\\
&\quad +0.25 \sum_{l_1,l_2,k_1,k_2=0}^{q} x_{l_1}x_{k_1}^* y_{l_2} y_{k_2}^* I(l_1-k_1+l_2-k_2=2u)\\
&\quad +0.25 \sum_{l_1,l_2,k_1,k_2=0}^{q} x_{l_1}x_{k_1}^* y_{l_2} y_{k_2}^* I(l_1-k_1+l_2-k_2=-2u).
\end{aligned}$$

For $x = (x_1, \ldots, x_q)$, $x_l \in \mathcal{A}_{\text{odd}}\ \forall l$ and $\eta = (\eta_1, \eta_2, \ldots, \eta_q)$ ($\{\eta_j\}$ are as in (6.54)), we define

$$\begin{aligned}
\beta_u(z,x) &= -\varphi_0(R_u(x,\eta)(z+\beta_u(z,\eta))^{-1}|x) \qquad (6.57)\\
&:= -0.5 \sum_{l_1,l_2,k_1,k_2=0}^{q} x_{l_1}x_{k_1}^* \varphi_0(\eta_{l_2}\eta_{k_2}^*(z+\beta_u(z,\eta))^{-1}) I(l_1-k_1+l_2-k_2=0)\\
&\quad -0.25 \sum_{l_1,l_2,k_1,k_2=0}^{q} x_{l_1}x_{k_1}^* \varphi_0(\eta_{l_2}\eta_{k_2}^*(z+\beta_u(z,\eta))^{-1}) I(l_1-k_1+l_2-k_2=2u)\\
&\quad -0.25 \sum_{l_1,l_2,k_1,k_2=0}^{q} x_{l_1}x_{k_1}^* \varphi_0(\eta_{l_2}\eta_{k_2}^*(z+\beta_u(z,\eta))^{-1}) I(l_1-k_1+l_2-k_2=-2u).
\end{aligned}$$

Now we are ready to state the following corollary. Recall $\{\Gamma_u\}$ in (3.3).

Corollary 6.4.6. *Suppose (A1), (B) and (6.54) hold and $p, n(p) \to \infty$, $p/n \to 0$. Then the almost sure LSD of $\frac{1}{2}\sqrt{np^{-1}}(\Delta_u + \Delta_u^* - \Gamma_u - \Gamma_u^*)$ exists and its Stieltjes transform is given by*

$$m_u(z) = -\varphi_0((z+\beta_u(z,\eta))^{-1}), \quad z \in \mathbb{C}^+, |z| \text{ large}, \qquad (6.58)$$

where $\beta_u(z,\eta)$ satisfies (6.57).

Proof. First note that $\{\frac{1}{2}\sqrt{np^{-1}}(\Delta_u + \Delta_u^* - \Gamma_u - \Gamma_u^*)\}$ satisfy the form (6.1) with $\{B_{2i}\} = \{P_i\}$. Moreover, under (B) and (6.54), $\{\psi_j\}$ satisfy (A2) and (6.4). Also note that the matrices $\{P_u : u = 0, \pm 1, \pm 2, \ldots\}$ satisfy (A3a) and (5.81). Hence, by Theorem 6.4.1, the LSD of $\{\frac{1}{2}\sqrt{np^{-1}}(\Delta_u + \Delta_u^* - \Gamma_u - \Gamma_u^*)\}$ is given by

$$\gamma = \sum_{j,k=0}^{q} \eta_j w_{u,j,k} \eta_k^*, \text{ where } w_{u,j,k}^* = w_{u,k,j}, \quad \text{and}$$

$$\begin{aligned}
&\kappa_r(w_{u,j_l,k_l} : 1 \le l \le r)\\
= &\begin{cases} \lim \frac{1}{n}\operatorname{Tr}\left(\frac{(P_{j_1-k_1+u}+P_{j_1-k_1-u})}{2}\frac{(P_{j_2-k_2+u}+P_{j_2-k_2-u})}{2}\right), & \text{if } r=2\\ 0, \quad r \neq 2 \end{cases}
\end{aligned}$$

$$\overset{\text{by (5.81)}}{=} \begin{cases} \frac{1}{8\pi}\int_0^{2\pi} e^{i(j_1-k_1+j_2-k_2+2u)\theta} + \frac{1}{8\pi}\int_0^{2\pi} e^{i(j_1-k_1+j_2-k_2-2u)\theta} \\ +\frac{1}{4\pi}\int_0^{2\pi} e^{i(j_1-k_1+j_2-k_2)\theta}, \qquad \text{if } r=2 \\ 0, \quad r \neq 2, \end{cases}$$

and, $\{\eta_j\}$ ($\{\eta_j\}$ are as in (6.54)) and $\{w_{u,j,k}\}$ are free.

Now Theorem 6.4.2 can be applied to get the Stieltjes transform of γ. By (6.46), $-\beta_u(z,x)$ equals

$$\begin{aligned} & \frac{1}{8\pi}\int_0^{2\pi} e^{2iu\theta} \sum_{j_1,j_2,k_1,k_2=1}^{q} e^{i(j_1-k_1+j_2-k_2)\theta} x_{j_1}^* x_{k_1} \varphi_0\left(\eta_{j_2}\eta_{k_2}^*(z+\beta(z,\eta))^{-1}\right) d\theta \\ & +\frac{1}{8\pi}\int_0^{2\pi} e^{-2iu\theta} \sum_{j_1,j_2,k_1,k_2=1}^{q} e^{i(j_1-k_1+j_2-k_2)\theta} x_{j_1}^* x_{k_1} \varphi_0\left(\eta_{j_2}\eta_{k_2}^*(z+\beta(z,\eta))^{-1}\right) d\theta \\ & +\frac{1}{4\pi}\int_0^{2\pi} \sum_{j_1,j_2,k_1,k_2=1}^{q} e^{i(j_1-k_1+j_2-k_2)\theta} x_{j_1}^* x_{k_1} \varphi_0\left(\eta_{j_2}\eta_{k_2}^*(z+\beta(z,\eta))^{-1}\right) d\theta \\ = \; & \frac{1}{8\pi}\int_0^{2\pi} (e^{2iu\theta} + e^{-2iu\theta} + 2)\Psi(x,\theta)\varphi_0\left(\Psi(\eta,\theta)(z+\beta(z,\eta))^{-1}\right) d\theta \\ = \; & \varphi_0\left(\frac{1}{2\pi}\int_0^{2\pi} \cos^2(u\theta)\Psi(x,\theta)\Psi(\eta,\theta)(z+\beta(z,\eta))^{-1} d\theta | x\right) \\ = \; & \varphi_0\left(R_u(x,\eta)(z+\beta(z,\eta))^{-1}|x\right). \end{aligned}$$

Hence, (6.57) is proved. Now by (6.44), (6.58) holds for large $|z|$. □

6.5 A necessary lemma

To prove Theorem 6.3.1, we will frequently need to calculate expressions of the form

$$\frac{1}{p} E\mathrm{Tr}\left(A_1 \frac{Z}{\sqrt{n}} A_2 \frac{Z^*}{\sqrt{n}} A_3 \frac{Z}{\sqrt{n}} \cdots \frac{Z^*}{\sqrt{n}} A_{2k+1}\right). \tag{6.59}$$

Recall the classes $\mathcal{L}$ and C respectively in (4.14) and (4.16). Note that in $(A1)$ we assume **one** of the following.

$(A1a)$ $((\varepsilon_{i,j})) \in C(\delta, n)$ for some $\delta \in (0, 2]$.

$(A1b)$ $((\varepsilon_{i,j})) \in \mathcal{L}$.

Under $(A1a)$ for some $C > 0$,

$$E|\varepsilon_{a,b}|^r \leq (\sqrt{n})^{r-4} E|\varepsilon_{a,b}|^4 \leq C n^{r/2-2}, \ \forall r \geq 4. \tag{6.60}$$

Under $(A1b)$ too, for some $C > 0$,

$$E|\varepsilon_{a,b}|^r \leq C \leq C n^{r/2-2}, \ \forall r \geq 4. \tag{6.61}$$

Recall $||\cdot||_2$ in (2.4). Let $A(i,j)$ be the (i,j)-th element of the matrix A. Note that, as $\{A_i\}$ are norm bounded matrices (by $(A2)$ and $(A3a)$), for some $C > 0$ and for all a, b,

$$\begin{aligned} |A_i(a,b)| &\leq \sqrt{(A_i A_i^*)(a,a)}, \quad \text{by Cauchy-Schwartz inequality} \\ &\leq ||A_i||_2 < C, \ \forall 1 \leq i \leq 2k+1. \end{aligned} \tag{6.62}$$

We shall use (6.60), (6.61) and (6.62) to prove Lemma 6.5.1. Now,

$$\begin{aligned} & \frac{1}{p} E\mathrm{Tr}\left(A_1 \frac{Z}{\sqrt{n}} A_2 \frac{Z^*}{\sqrt{n}} A_3 \frac{Z}{\sqrt{n}} \cdots \frac{Z^*}{\sqrt{n}} A_{2k+1}\right) \qquad (6.63) \\ = \; & \frac{1}{n^k p} E \sum_{u,v} \left(\prod_{i=1}^{k} A_{2i-1}(u_{2i-1}, u_{2i}) \varepsilon_{u_{2i}, v_{2i-1}} A_{2i}(v_{2i-1}, v_{2i}) \varepsilon_{u_{2i+1}, v_{2i}}\right) A_{2k+1}(u_{2k+1}, u_1). \end{aligned}$$

Let

$$T = \{(u_2, v_1), (u_3, v_2), (u_4, v_3), (u_5, v_4), \ldots, (u_{2k}, v_{2k-1}), (u_{2k+1}, v_{2k})\}. \tag{6.64}$$

Note that $\mathcal{T}$ is the set of all indices attached with ε's. In (6.59), (Z, Z^*) appear alternately and there are k such (Z, Z^*). When $\delta = 0$, $(u_{2i+\delta}, v_{2i+\delta-1})$ is attached with the i-th Z. Similarly, when $\delta = -1$, $(u_{2i+\delta}, v_{2i+\delta-1})$ is attached with the i-th

Z^*. Pairs of indices $(u_{2i+\delta}, v_{2i+\delta-1})$ and $(u_{2i'+\delta'}, v_{2i'+\delta'-1})$ are said to be matched if $(i', \delta') \neq (i, \delta)$ and $(u_{2i+\delta}, v_{2i+\delta-1}) = (u_{2i'+\delta'}, v_{2i'+\delta'-1})$. As $\{\varepsilon_{i,j}\}$ are independent with mean 0, only matched indices in $\mathcal{T}$ need to be considered.

Consider the natural bijection between $\mathcal{T}$ in (6.64) and $\{1, 2, \ldots, 2k\}$. A matching in $\mathcal{T}$ forms a partition of $\mathcal{T}$, where the matched indices form blocks. Since we have ruled out singleton blocks, the relevant set of matchings of $\mathcal{T}$ is in bijection (induced by the above bijection) with

$$\mathcal{P}_{2k} \;=\; \text{set of all partitions of } \{1, 2, \ldots, 2k\} \text{ having no singleton block.} \tag{6.65}$$

Let for any set A, E_A be the usual expectation restricting on the set A. Then from the above discussions,

$$\begin{aligned}
& E\mathrm{Tr}\left(A_1 \frac{Z}{\sqrt{n}} A_2 \frac{Z^*}{\sqrt{n}} A_3 \ldots \frac{Z^*}{\sqrt{n}} A_{2k+1}\right) \\
= \; & E_{\mathcal{P}_{2k}} \mathrm{Tr}\left(A_1 \frac{Z}{\sqrt{n}} A_2 \frac{Z^*}{\sqrt{n}} A_3 \ldots \frac{Z^*}{\sqrt{n}} A_{2k+1}\right) \\
= \; & \sum_{\sigma \in \mathcal{P}_{2k}} E_\sigma \mathrm{Tr}\left(A_1 \frac{Z}{\sqrt{n}} A_2 \frac{Z^*}{\sqrt{n}} A_3 \ldots \frac{Z^*}{\sqrt{n}} A_{2k+1}\right).
\end{aligned} \tag{6.66}$$

Now to compute (6.66), let us first concentrate on $\sigma \in \mathcal{P}_{2k} \cap NC(2k)$, where $NC(2k)$ is as in (4.47). Lemma 6.5.1 provides an upper bound for the terms in (6.66). This will be useful in the proof of Theorem 6.3.1.

Lemma 6.5.1. *Suppose* $(A1)$ *holds and the matrices* $\{A_i\}$ *are all norm bounded. Suppose* $\sigma \in NC(2k) \cap \mathcal{P}_{2k}$ *has* $K_{i,\sigma}$ *blocks of size* $i \geq 2$ *of which* $K_{i,1,\sigma}$ *and* $K_{i,2,\sigma}$ *start with odd and even indices respectively. Then for some* $C > 0$,

$$(a) \qquad \left|\frac{1}{p} E_\sigma \mathrm{Tr}\left(A_1 \frac{Z}{\sqrt{n}} A_2 \frac{Z^*}{\sqrt{n}} \ldots \frac{Z^*}{\sqrt{n}} A_{2k+1}\right)\right| \leq C y_n^{K_{2,2,\sigma}} (y_n p^{-1})^{\sum_{i \geq 2}(0.5 K_{2i-1,\sigma} + K_{2i,\sigma})}. \tag{6.67}$$

(*b*) *The upper bound in (6.67) also holds for*

$$|E_\sigma(A_1\frac{Z}{\sqrt{n}}A_2\frac{Z^*}{\sqrt{n}}A_3\frac{Z}{\sqrt{n}}\ldots\frac{Z^*}{\sqrt{n}}A_{2k+1})(u,u')|,\ \forall 1 \le u, u' \le p. \quad (6.68)$$

Proof. We shall prove (a) and (b) simultaneously and use induction on k. Let $k = 1$. Note that $NC(2) \cap \mathcal{P}_2 = \{\{1,2\}\}$. By (6.63), for some $C > 0$, we have

$$\begin{aligned}
|\frac{1}{p}E\mathrm{Tr}\left(A_1\frac{Z}{\sqrt{n}}A_2\frac{Z^*}{\sqrt{n}}A_3\right)| &= |\frac{1}{np}E_{\{1,2\}}\sum_{u_1,a,b,c,d}A_1(u_1,a)\varepsilon_{a,b}A_2(b,d)\varepsilon_{c,d}A_3(c,u_1)| \\
&= |\frac{1}{np}E_{\{a=c,b=d\}}\sum_{u_1,a,b}A_1(u_1,a)\varepsilon_{a,b}A_2(b,d)\varepsilon_{c,d}A_3(c,u_1)| \\
&= |\frac{1}{np}\sum_{u_1,a,b}A_1(u_1,a)A_2(b,b)A_3(a,u_1)E(\varepsilon_{a,b}^2)| \\
&= |\frac{1}{p}\big(\mathrm{Tr}(A_1A_3)\big)||\frac{1}{n}\mathrm{Tr}(A_2)|, \quad \text{as } E(\varepsilon_{a,b}^2) = 1 \\
&\le \sqrt{\frac{1}{p}\big(\mathrm{Tr}(A_1^*A_1)\big)\frac{1}{p}\big(\mathrm{Tr}(A_3^*A_3)\big)}\ |\frac{1}{n}\mathrm{Tr}(A_2)| \\
&\le ||A_1||_2\ ||A_2||_2\ ||A_3||_2 \\
&\le C, \ \text{as } \{A_i\} \text{ are norm bounded.}
\end{aligned}$$

Therefore, as $K_{2,1,\{1,2\}} = 1$, $K_{2,2,\{1,2\}} = 0$ and $K_{i,\{1,2\}} = 0\ \forall i \ge 3$, ($a$) is proved for $k = 1$. Next, again for some $C > 0$,

$$\begin{aligned}
&|E\left(A_1\frac{Z}{\sqrt{n}}A_2\frac{Z^*}{\sqrt{n}}A_3\right)(u_1,u_2)| \\
= \ &|\frac{1}{n}E_{\{1,2\}}\sum_{a,b,c,d}A_1(u_1,a)\varepsilon_{a,b}A_2(b,d)\varepsilon_{c,d}A_3(c,u_2)| \\
= \ &|\frac{1}{n}E_{\{a=c,b=d\}}\sum_{a,b}A_1(u_1,a)\varepsilon_{a,b}A_2(b,d)\varepsilon_{c,d}A_3(c,u_2)| \\
= \ &|\frac{1}{n}\sum_{a,b}A_1(u_1,a)A_2(b,b)A_3(a,u_2)E(\varepsilon_{a,b}^2)| \\
= \ &|\big((A_1A_3)(u_1,u_2)\big)|\ |\frac{1}{n}\mathrm{Tr}(A_2)|, \quad \text{as } E(\varepsilon_{a,b}^2) = 1 \\
\le \ &C, \ \text{by applying (6.62) on } A_1A_3 \text{ and } A_2.
\end{aligned}$$

Hence, (b) is proved for $k = 1$.

Suppose (a) and (b) hold for all $k \leq m-1$. Now we shall show that they are true for $k = m$ also.

Since σ is non-crossing, it always has at least one block B with adjacent indices. If we drop any one of those blocks, then again we have a non-crossing partition, say σ^*, of the remaining indices. Note that $\sigma^* \in NC(2k) \cap \mathcal{P}_{2k}$ for some $k \leq m-1$. Therefore, (6.67) holds for σ^*. Then four situations can arise depending on the length of B (even/odd) and the index of the starting element of B (even/odd). Here we shall show the details for the case where B is of even length and starts with an odd index. Similar argument works for other cases.

Now, σ has a block $B = \{2j-1, 2j, \ldots, 2s\}$. Then there exists (a, b) such that

$$(u_{2i+\delta}, v_{2i+\delta-1}) \begin{cases} = (a,b), \forall j \leq i \leq s, \delta = 0, 1 \\ \neq (a,b), \forall i < j \text{ or } i > s,\ \delta = 0, 1. \end{cases}$$

Moreover, note that $\sigma^* \in NC(2(m-s+j-1)) \cap \mathcal{P}_{2(m-s+j-1)}$ and

$$\begin{aligned} K_{2i-1,1,\sigma^*} &= K_{2i-1,1,\sigma},\ K_{2i-1,2,\sigma^*} = K_{2i-1,2,\sigma},\ K_{2i,2,\sigma^*} = K_{2i,2,\sigma},\ \forall i \geq 1, \\ K_{2i,1,\sigma^*} &= K_{2i,1,\sigma},\ \forall i \neq s-j+1,\ K_{2(s-j+1),1,\sigma^*} = K_{2(s-j+1),1,\sigma} - 1. \end{aligned} \tag{6.69}$$

Let $D_1 = A_1 \frac{Z}{\sqrt{n}} A_2 \frac{Z^*}{\sqrt{n}} A_3 \ldots \frac{Z^*}{\sqrt{n}} A_{2j-1}$, $D_2 = A_{2s+1} \frac{Z}{\sqrt{n}} A_{2s+2} \frac{Z^*}{\sqrt{n}} A_{2s+3} \ldots A_{2k} \frac{Z^*}{\sqrt{n}} A_{2k+1}$.

Case I. Let $s - j = 0$. Then by (6.63), for some $C_1, C_2 > 0$, we have

$$\begin{aligned} & |\frac{1}{p} E_\sigma \mathrm{Tr}\left(A_1 \frac{Z}{\sqrt{n}} A_2 \frac{Z^*}{\sqrt{n}} A_3 \frac{Z}{\sqrt{n}} \ldots \frac{Z^*}{\sqrt{n}} A_{2k+1}\right)| \\ = & |\frac{1}{np} E_\sigma \sum_{u_1,a,b} D_1(u_1,a)\varepsilon_{a,b} A_{2s}(b,b)\varepsilon_{a,b} D_2(a,u_1)| \\ = & |\frac{1}{np} \sum_{u_1,a,b} E_{\sigma^*}(D_1(u_1,a) A_{2s}(b,b) D_2(a,u_1)) E(\varepsilon_{a,b}^2)| \\ = & |\frac{1}{p} E_{\sigma^*}\big(\mathrm{Tr}(D_1 D_2)\big)||\frac{1}{n}\mathrm{Tr}(A_{2s})|, \text{ as } E(\varepsilon_{a,b}^2) = 1 \end{aligned}$$

$$\begin{aligned}
&\leq C_1 \Big|\frac{1}{p} E_{\sigma^*}\big(\mathrm{Tr}(D_1 D_2)\big)\Big|, \text{ by applying (6.62) on } A_{2s} \\
&\leq C_2 y_n^{K_{2,2,\sigma^*}} (y_n p^{-1})^{\sum_{i \geq 2}(0.5 K_{2i-1,\sigma^*} + K_{2i,\sigma^*})} \\
&\qquad\qquad\qquad \text{by applying } (a) \text{ on } k = m-1 \\
&= C_2 y_n^{K_{2,2,\sigma}} (y_n p^{-1})^{\sum_{i \geq 2}(0.5 K_{2i-1,\sigma} + K_{2i,\sigma})}, \text{ by (6.69).}
\end{aligned}$$

Hence, (a) is proved for $k = m$ and $s - j = 0$.

Case II. Let $s - j > 0$. Then by (6.63), for some $C_1, C_2, C_3, C_4 > 0$, we have

$$\begin{aligned}
& \Big|\frac{1}{p} E_\sigma \mathrm{Tr}\left(A_1 \frac{Z}{\sqrt{n}} A_2 \frac{Z^*}{\sqrt{n}} A_3 \frac{Z}{\sqrt{n}} \ldots \frac{Z^*}{\sqrt{n}} A_{2k+1}\right)\Big| \\
=\ & \Big|\frac{1}{n^{s-j+1}p} E_\sigma \sum_{u_1,a,b} D_1(u_1,a)\Big(\prod_{i=j}^{s-1} \varepsilon_{a,b} A_{2i}(b,b) \varepsilon_{a,b} A_{2i+1}(a,a)\Big) \\
&\qquad \varepsilon_{a,b} A_{2s}(b,b) \varepsilon_{a,b} D_2(a,u_1)\Big| \\
=\ & \Big|\frac{1}{n^{s-j+1}p}\Big(\sum_{a,b}\Big(\prod_{i=j}^{s-1} A_{2i+1}(a,a) \prod_{i=j}^{s} A_{2i}(b,b)\Big) \\
&\qquad E_{\sigma^*}\Big(\sum_{u_1} D_1^*(a,u_1) D_2^*(u_1,a)\Big) E(\varepsilon_{a,b}^{2s-2j+2})\Big)\Big| \\
\leq\ & \left[\frac{1}{np}\sum_{a,b}\Big(\prod_{i=j}^{s-1} A_{2i+1}^2(a,a) \prod_{i=j}^{s} A_{2i}^2(b,b)\Big)\Big(E_{\sigma^*}\Big(\sum_{u_1} D_1^*(a,u_1) D_2^*(u_1,a)\Big)\Big)^2\right]^{1/2} \\
& \frac{1}{n^{s-j}}\left[\frac{1}{np}\sum_{a,b}(E(\varepsilon_{a,b}^{2s-2j+2}))^2\right]^{1/2}, \text{ (by Cauchy-Schwartz inequality on } \Sigma_{a,b}) \\
\leq\ & \left[\frac{1}{np}\sum_{a,b}\Big(\prod_{i=j}^{s-1} A_{2i+1}^2(a,a) \prod_{i=j}^{s} A_{2i}^2(b,b)\Big)\Big(E_{\sigma^*}\Big(\sum_{u_1} D_1^*(a,u_1) D_2^*(u_1,a)\Big)\Big)^2\right]^{1/2} \\
&\qquad \frac{1}{n^{s-j}}\left[\frac{1}{np}\sum_{a,b}(C_1 n^{s-j-1} \sup_{a,b} E(\varepsilon_{a,b}^4))^2\right]^{1/2}, \text{ (by (6.60) and (6.61))} \\
\leq\ & \left[\frac{1}{np}\sum_{a,b}\Big(\prod_{i=j}^{s-1} A_{2i+1}^2(a,a) \prod_{i=j}^{s} A_{2i}^2(b,b)\Big)\right. \\
& \left.\Big(E_{\sigma^*}\Big(\sum_{u_1} D_1^*(a,u_1) D_2^*(u_1,a)\Big)\Big)^2\right]^{1/2} C_2 \frac{n^{s-j-1}}{n^{s-j}}, \quad \text{(as } \sup_{a,b} E(\varepsilon_{a,b}^4) < \infty)
\end{aligned}$$

$$\begin{aligned}
&\le \frac{C_3}{n}\left[\frac{1}{np}\sum_{a,b}\left(E_{\sigma^*}\left(\sum_{u_1} D_1^*(a,u_1)D_2^*(u_1,a)\right)\right)^2\right]^{1/2}, \\
&\qquad\qquad \text{by applying (6.62) on } \left(\prod_{i=j}^{s-1} A_{2i+1}^2(a,a)\prod_{i=j}^{s} A_{2i}^2(b,b)\right) \\
&\le \frac{C_3}{n}\left[\frac{1}{p}\sum_{a}\left(E_{\sigma^*}\left(D_1^*D_2^*(a,a)\right)\right)^2\right]^{1/2} = \frac{C_3}{p}y_n\left[\frac{1}{p}\sum_{a}\left(E_{\sigma^*}\left(D_1^*D_2^*(a,a)\right)\right)^2\right]^{1/2} \\
&\le C_4\frac{y_n}{p}y_n^{K_{2,2,\sigma^*}}(y_np^{-1})^{\sum_{i\ge 2}(0.5K_{2i-1,\sigma^*}+K_{2i,\sigma^*})} \\
&\qquad\qquad \text{by applying } (b) \text{ on } \sigma^* \text{ for } k=m-s+j-1 \\
&= C_4\frac{y_n}{p}y_n^{K_{2,2,\sigma}}(y_np^{-1})^{\sum_{i\ge 2}0.5K_{2i-1,\sigma}}(y_np^{-1})^{\sum_{s-j+1\neq i\ge 2}K_{2i,\sigma}+(K_{2(s-j+1),1,\sigma}-1)+K_{2(s-j+1),2,\sigma}}, \\
&\qquad\qquad \text{(by (6.69))} \\
&= C_4 y_n^{K_{2,2,\sigma}}(y_np^{-1})^{\sum_{i\ge 2}(0.5K_{2i-1,\sigma}+K_{2i,\sigma})}.
\end{aligned}$$

Therefore, (a) is proved for $k=m$ and $s-j>0$ and hence proof of (a) is complete. One can similarly prove (b).

This completes the proof of Lemma 6.5.1. □

6.6 Proof of Theorem 6.3.1

(a) Here we prove the theorem only for $U=1$. Similar argument works for $U>1$. Note that for $U=1$, we have only $\{\mathbb{P}_{l,(1,1,\ldots,1)}\}$, $\{\mathcal{R}_{l,(1,1,\ldots,1)}\}$, $\{w_{1,l,j}\}$ and $\{\alpha_{l,(1,1,\ldots,1)}\}$. Let us denote them respectively by $\{\mathbb{P}_l\}$, $\{\mathcal{R}_l\}$, $\{w_{l,j}\}$ and $\{\alpha_l\}$. Let us write $\{\mathbb{G}_l\}$ for $\{\mathbb{G}_{l,k_l}\}$.

Let π be any polynomial. Then by Definition 4.3.4, it is enough to prove

$$\lim p^{-1}E\mathrm{Tr}(\pi(\mathcal{R}_l : l\ge 1)) = \varphi(\pi(\alpha_l : l\ge 1)). \tag{6.70}$$

Note that, for some $\{\mathcal{R}_{l_{ti}}\}$ from $\{\mathcal{R}_l\}$ and constants $\{c_i\}$, we can write $\pi(\mathcal{R}_l : l\ge$

$1) = \sum_{i=1}^{T} c_i \prod_{t=1}^{T_i} \mathcal{R}_{l_{ti}}$. Therefore, it is enough to establish, for each $1 \le i \le T$

$$\lim p^{-1} E\mathrm{Tr}(\prod_{i=1}^{T_i} \mathcal{R}_{l_{ti}}) = \begin{cases} 0, \text{ if } T_i \text{ is odd} \\ \varphi(\prod_{i=1}^{T_i} \alpha_{l_{ti}}), \text{ if } T_i \text{ is even.} \end{cases} \tag{6.71}$$

For simplicity of notation, we prove only

$$\lim p^{-1} E\mathrm{Tr}(\prod_{l=1}^{T} \mathcal{R}_{l}) = \begin{cases} 0, \text{ if } T \text{ is odd} \\ \varphi(\prod_{t=1}^{T} \alpha_{l}), \text{ if } T \text{ is even.} \end{cases} \tag{6.72}$$

Similar argument works to prove the more general (6.71).

Let $A(i,j)$ be the (i,j)-th element of the matrix A. For convenience we write $\mathrm{Tr}(\prod_{l=1}^{T} \mathcal{R}_l)$ in the form

$$\mathrm{Tr}(\prod_{l=1}^{T} \mathcal{R}_l) = \sum_{\substack{u_{l,3k_l+1}=u_{l+1,1} \\ u_{T+1,1}=u_{11}}} \prod_{l=1}^{T} \mathcal{R}_l(u_{l,1}, u_{l,3k_l+1}), \text{ where} \tag{6.73}$$

$$\mathcal{R}_l(u_{l,1}, u_{l,3k_l+1}) = \sqrt{\frac{n}{p}} \Big(\mathbb{P}_l(u_{l,1}, u_{l,3K_l+1}) - \mathbb{G}_l(u_{l,1}, u_{l,3K_l+1}) \Big), \text{ and} \tag{6.74}$$

$$\begin{aligned} \mathbb{P}_l(u_{l,1}, u_{l,3k_l+1}) \quad &= \quad n^{-k_l} \sum_{\substack{u_{l,j}, v_{l,j} \\ u_{l,j} \ne u_{l,1}, u_{l,3K_l+1}}} \prod_{\substack{i: u_{l,3i}=u_{l,3i+1} \\ 1 \le i \le k_l - 1}} \Big(A_{l,2i-1}(u_{l,3i-2}, u_{l,3i-1}) \\ & \varepsilon_{u_{l,3i-1}, v_{l,2i-1}} A_{l,2i}(v_{l,2i-1}, v_{l,2i}) \varepsilon_{u_{l,3i}, v_{l,2i}} \Big) A_{l,2k_l+1}(u_{l,3k_l}, u_{l,3k_l+1}). \end{aligned} \tag{6.75}$$

For each $1 \le l \le T$, we define

$$\mathcal{I}_l = \{(u_{l,3i+\delta}, v_{l,2i+\delta}) : \delta = -1, 0, \ 1 \le i \le k_l\}. \tag{6.76}$$

Note that $\mathcal{I}_l$ is the set of all indices attached with ε's in the expansion of $\mathcal{R}_l$ given in (6.73)-(6.75). An index $(u_{l,3k+\delta}, v_{l,2k+\delta})$ is said to be matched if there is at least one $(k', \delta', l') \neq (k, \delta, l)$ with $(u_{l,3k+\delta}, v_{l,2k+\delta}) = (u_{l',3k'+\delta'}, v_{l',2k'+\delta'})$. Now note that $E\left(\mathrm{Tr}(\prod_{l=1}^{T} \mathcal{R}_l)\right)$ involves all indices in $\cup_{l=1}^{T} \mathcal{I}_l$. As $\{\varepsilon_{i,j}\}$ are independent and have mean 0, all indices in $\cup_{l=1}^{T} \mathcal{I}_l$ need to be matched to guarantee a non-zero contribution. For each $1 \leq l \leq T$, consider the following sets of matched indices.

$$\begin{aligned} B_l \;\; = \;\; & \text{set of all matchings where for each } (k,\delta), \text{ there is at least one} \\ & (k', \delta') \neq (k, \delta) \text{ with } (u_{l,3k+\delta}, v_{l,2k+\delta}) = (u_{l,3k'+\delta'}, v_{l,2k'+\delta'}) \text{ and for } l \neq l', \\ & \text{there is no } (k', \delta', l') \text{ such that } (u_{l,3k+\delta}, v_{l,2k+\delta}) = (u_{l',3k'+\delta'}, v_{l',2k'+\delta'}). \end{aligned} \tag{6.77}$$

Consider the disjoint decomposition $\cup_{l=1}^{T+1} C_l$ of all possible matchings of indices in $\cup_{l=1}^{T} \mathcal{I}_l$, where

$$C_1 = B_1,\; C_l = (\cap_{j=1}^{l-1} B_j^c) \cap B_l \;\forall 2 \leq l \leq T,\; C_{T+1} = \cap_{l=1}^{T} B_l^c. \tag{6.78}$$

Let for any set A, E_A be the usual expectation restricting on the set A. Then we have the following lemma.

Lemma 6.6.1. *Suppose Assumptions* (A1)*,* (A2) *and* (A3a) *hold. Then*
(*i*) $E(y_n^{-1/2}\mathbb{P}_l(u_2,u_3)) = y_n^{-1/2}\mathbb{G}_l(u_2,u_3) + O(y_n^{1/2})$.
(*ii*) $\lim \frac{1}{p} E_{C_l} \mathrm{Tr}\left(y_n^{-l/2} \prod_{i=1}^{l} \mathbb{P}_i \prod_{i=l+1}^{T} \mathcal{R}_i\right) = \lim \frac{1}{p} E_{\cap_{i=1}^{l-1} B_i^c} \mathrm{Tr}\left(y_n^{-l/2} \prod_{i=1}^{l-1} \mathbb{P}_i \mathbb{G}_l \prod_{i=l+1}^{T} \mathcal{R}_i\right)$.
(*iii*) $\lim \frac{1}{p} E\mathrm{Tr}(\prod_{l=1}^{T} \mathcal{R}_l) = \lim \frac{1}{p} E_{C_{T+1}} \mathrm{Tr}(y_n^{-T/2} \prod_{l=1}^{T} \mathcal{P}_l)$.

Proof. (*i*) Recall that $\mathbb{P}_l = A_{l,1}\frac{Z}{\sqrt{n}}A_{l,2}\frac{Z^*}{\sqrt{n}}A_{l,3}\frac{Z}{\sqrt{n}}\cdots\frac{Z^*}{\sqrt{n}}A_{l,2k_l+1}$. Consider the partition $\sigma^* = \{\{1,2\},\{3,4\},\ldots,\{2k_l-1,2k_l\}\}$. Note that

$$E_{\sigma^*}(y_n^{-1/2}\mathbb{P}_l(u_2,u_3)) = y_n^{-1/2}\mathbb{G}_l(u_2,u_3). \tag{6.79}$$

Recall $\mathcal{P}_{2k}$ in (6.65). Let $\mathcal{P}_{2k}^c$ = set of all partitions of $\{1,2\ldots,2k\} - \mathcal{P}_{2k}$. Note that,

$$
\begin{aligned}
E(y_n^{-1/2}\mathbb{P}_l(u_2,u_3)) &= E_{\sigma^*}(y_n^{-1/2}\mathbb{P}_l(u_2,u_3)) + \sum_{\sigma\in\mathcal{P}_{2k_l}^c} E_\sigma(y_n^{-1/2}\mathbb{P}_l(u_2,u_3)) \\
&\quad + \sum_{\sigma\in NC(2k_l)\cap\mathcal{P}_{2k_l}-\{\sigma^*\}} E_\sigma(y_n^{-1/2}\mathbb{P}_l(u_2,u_3)) \\
&\quad + \sum_{\sigma\in\mathcal{P}_{2k_l}-NC(2k_l)} E_\sigma(y_n^{-1/2}\mathbb{P}_l(u_2,u_3)) \\
&= T_1+T_2+T_3+T_4, \ (say). \qquad (6.80)
\end{aligned}
$$

As each partition in $\mathcal{P}_{2k_l}^c$ has at least one singleton block, $T_2 = 0$. Also a partition in $NC(2k_l)\cap\mathcal{P}_{2k_l} - \{\sigma^*\}$ contains either a block of length 2 and starts with an even index or a block of length longer than 2. Hence, by Lemma 6.5.1 (b), $T_3 = O(y_n^{1/2})$. Moreover, crossing partitions in $\mathcal{P}_{2k_l} - NC(2k_l)$ have more restrictions on indices than that of partitions in $\mathcal{P}_{2k_l}\cap NC(2k_l) - \{\sigma^*\}$. Therefore contribution of $\mathcal{P}_{2k_l} - NC(2k_l)$ in $E(y_n^{-1/2}\mathbb{P}_l(u_2,u_3))$ is smaller than the contribution of the latter. Therefore, $T_4 = O(y_n^{1/2})$. Hence, by (6.79) and (6.80), the proof of Lemma 6.6.2 (i) is complete.

(ii) To prove (ii), we need more analysis for the set C_l. Define

$$
\begin{aligned}
\mathcal{S}_l &= \text{set of all matchings of indices in } \mathcal{I}_l, \text{and} \\
\mathcal{S}_{-l} &= \text{set of all matchings of indices in } \cup_{j\neq l}\mathcal{I}_j \text{ such that for each } 1\le j<l, \\
&\quad\ \text{there is at least one index in } \mathcal{I}_j \text{ which matches with some index} \\
&\quad\ \text{in } \mathcal{I}_k,\ k\neq j,l.
\end{aligned}
$$

Note that

$$
C_l = (\cap_{j=1}^{l-1}B_j^c)\cap B_l = \{(\sigma_1\cup\sigma_2) : \sigma_1\in\mathcal{S}_l,\ \sigma_2\in\mathcal{S}_{-l}\}. \qquad (6.81)
$$

Let us denote

$$W_{P,l} \;:=\; (np^{-1})^{l/2}\prod_{i=1}^{l}\mathbb{P}_i, \quad W_{R,l} := \prod_{i=l+1}^{T}\mathcal{R}_i.$$

Then for all $2 \le l \le T$, we have

$$\begin{aligned}
p^{-1}E_{C_l}\mathrm{Tr}(W_{P,l}W_{R,l}) \;&= \sum_{\sigma\in C_l} p^{-1}E_\sigma \mathrm{Tr}(W_{P,l-1}\sqrt{np^{-1}}\mathbb{P}_l W_{R,l}) \qquad (6.82)\\
= \;& \sum_{\sigma\in C_l} p^{-1}\sum_u E_\sigma\left(W_{P,l-1}(u_1,u_2)\sqrt{np^{-1}}\mathbb{P}_l(u_2,u_3)W_{R,l}(u_3,u_1)\right)\\
= \;& \sum_{\sigma_1\in\mathcal{S}_l,\ \sigma_2\in\mathcal{S}_{-l}} p^{-1}\sum_u E_{\sigma_1}(\sqrt{np^{-1}}\mathbb{P}_l(u_2,u_3))E_{\sigma_2}(W_{P,l-1}(u_1,u_2)W_{R,l}(u_3,u_1))\\
& \qquad [\text{as } C_l\subset B_l \text{ and even under } B_l,\ \{\varepsilon_{u,v}:(u,v)\in\mathcal{I}_l\} \text{ are}\\
& \qquad\qquad \text{independent of } \{\varepsilon_{u,v}:(u,v)\in\cup_{j\neq l}\mathcal{I}_j\}]\\
= \;& p^{-1}\sum_u\left(\sum_{\sigma_1\in\mathcal{S}_l}E_{\sigma_1}(\sqrt{np^{-1}}\mathbb{P}_l(u_2,u_3))\right)\left(\sum_{\sigma_2\in\mathcal{S}_{-l}}E_{\sigma_2}(W_{P,l-1}(u_1,u_2)W_{R,l}(u_3,u_1))\right)\\
= \;& p^{-1}\sum_u E(\sqrt{np^{-1}}\mathbb{P}_l(u_2,u_3))\left(\sum_{\sigma_2\in\mathcal{S}_{-l}}E_{\sigma_2}(W_{P,l-1}(u_1,u_2)W_{R,l}(u_3,u_1))\right)\\
& \qquad \left(E_{\cap_{i=1}^{l-1}B_i^c}(W_{P,l-1}(u_1,u_2)W_{R,l}(u_3,u_1))\right), \quad [\text{by } (a)]\\
= \;& p^{-1}E_{\cap_{i=1}^{l-1}B_i^c}\mathrm{Tr}\left(W_{P,l-1}\sqrt{np^{-1}}\mathbb{G}_lW_{R,l}\right) + O((p/n)^{1/2}).
\end{aligned}$$

Hence, the proof of (ii) is complete.

$$\begin{aligned}
(iii) \quad & \lim p^{-1}E\mathrm{Tr}\left[\Pi_{i=1}^{T}\mathcal{R}_i\right]\\
& = \lim p^{-1}E_{B_1}\mathrm{Tr}(\sqrt{np^{-1}}\mathbb{P}_1W_{R,1}) + \lim p^{-1}E_{B_1^c}\mathrm{Tr}(\sqrt{np^{-1}}\mathbb{P}_1W_{R,1})\\
& \qquad\qquad - \lim p^{-1}E\mathrm{Tr}\left(\sqrt{np^{-1}}\mathbb{G}_1W_{R,1}\right)\\
& = \lim p^{-1}E_{B_1^c}\mathrm{Tr}(\sqrt{np^{-1}}\mathbb{P}_1W_{R,1}), \quad (\text{by } (ii) \text{ for } l=1)\\
& = \lim p^{-1}E_{B_1^c\cap B_2}\mathrm{Tr}(W_{P,2}W_{R,2}) + \lim p^{-1}E_{B_1^c\cap B_2^c}\mathrm{Tr}(W_{P,2}W_{R,2})\\
& \qquad\qquad - \lim p^{-1}E_{B_1^c}\mathrm{Tr}(W_{P,1}\sqrt{np^{-1}}\mathbb{G}_2W_{R,2})\\
& = \lim p^{-1}E_{B_1^c\cap B_2^c}\mathrm{Tr}(W_{P,2}W_{R,2}), \quad (\text{by } (ii), \text{ for } l=2)
\end{aligned}$$

$$\vdots$$

$$= \lim p^{-1} E_{B_1^c \cap B_2^c \cap \ldots \cap B_T^c} \mathrm{Tr}(\Pi_{i=1}^T \mathbb{P}_i), \text{ by repeated application of } (ii) \text{ for } l \geq 3.$$

$$= \lim (np^{-1})^{T/2} p^{-1} E_{C_{T+1}} \mathrm{Tr}(\Pi_{i=1}^T \mathbb{P}_i).$$

Therefore, (iii) is established.

Thus proof of Lemma 6.6.1 is complete. □

Now we get back to the proof of the Theorem. Therefore, by Lemma 6.6.1 (iii), we have

$$\lim p^{-1} E\left(\mathrm{Tr}(\prod_{l=1}^{T} \mathcal{R}_l) \right) = \lim \left(\frac{n}{p}\right)^{T/2} p^{-1} E_{C_{T+1}} \mathrm{Tr}(\Pi_{l=1}^T \mathbb{P}_l). \tag{6.83}$$

Next we shall analyze the set C_{T+1} and identify the set of matchings which contribute in the limit.

Two index sets $\mathcal{I}_i$ and $\mathcal{I}_{i'}$ are said to be connected if there is (k, δ) and (k', δ') with $(u_{i,3k+\delta}, v_{i,2k+\delta}) = (u_{i',3k'+\delta'}, v_{i',2k'+\delta'})$, where $(u_{i,3k+\delta}, v_{i,2k+\delta}) \in \mathcal{I}_i$ and $(u_{i',3k'+\delta'}, v_{i',2k'+\delta'}) \in \mathcal{I}_{i'}$. Also a collection of index sets $\{\mathcal{I}_{i_1}, \mathcal{I}_{i_2}, \ldots, \mathcal{I}_{i_s}\}$, $s \geq 2$, is said to form a connected group if for each $1 \leq k \leq s-1$, $\mathcal{I}_{i_k}$ and $\mathcal{I}_{i_{k+1}}$ is connected. Note that, in a typical matching in C_{T+1}, for each i, $\mathcal{I}_i$ is connected with some other $\mathcal{I}_{i'}$, $i' \neq i$. Therefore, each matching in C_{T+1} corresponds to some disjoint connected groups each of length at least 2. Consider the following disjoint decomposition of C_{T+1}.

$$C_{T+1} = \bigcup_{\substack{2 \leq g_1, g_2, \ldots, g_R \leq T \\ \sum_{j=1}^R g_j = T, \ R \geq 1}} G(g_1, g_2, \ldots, g_R), \text{ where} \tag{6.84}$$

$$G(g_1, g_2, \ldots, g_R) = \text{set of all such matchings in } C_{T+1} \text{ which form exactly } R \text{ connected groups of length } g_1, g_2, \ldots, g_R. \tag{6.85}$$

Note that $R \leq T/2$ and equality holds if T is even and $g_i = 2, \forall i$. Now we have the following lemma.

Lemma 6.6.2. *Suppose Assumptions (A1), (A2) and (A3a) hold. Then*

$$\lim \frac{1}{p} E_{G(g_1,g_2,\ldots,g_R)} \mathrm{Tr}(n^{T/2} p^{-T/2} \prod_{i=1}^{T} \mathbb{P}_i) = O(y_n^{T/2-R}).$$

Proof. Let $D(g_1, g_2, \ldots, g_R)$ be the set of all non-crossing pair matchings in C_{T+1} which form exactly R connected groups of lengths $g_1, g_2, \ldots, g_R$. Note that $D(g_1, g_2, \ldots, g_R) \subset G(g_1, g_2, \ldots, g_R)$. We shall first show that

$$\lim \frac{1}{p} E_{D(g_1,g_2,\ldots,g_R)} \mathrm{Tr}(n^{T/2} p^{-T/2} \prod_{i=1}^{T} \mathbb{P}_i) = O(y_n^{T/2-R}). \tag{6.86}$$

Under $D(g_1, g_2, \ldots, g_R)$, to connect two index sets $\mathcal{I}_l$ and $\mathcal{I}_{l'}$, there must be a matching of the type $(u_{l,3i}, v_{l,2i}) = (u_{l',3i'-1}, v_{l',2i'-1})$ for some i and i'. Note that they respectively correspond to the i-th Z^* in $\mathbb{P}_l$ and i'-th Z in $\mathbb{P}_{l'}$. Therefore, under $D(g_1, g_2, \ldots, g_R)$, to connect $\mathcal{I}_l$ and $\mathcal{I}_{l'}$, there must be a block which starts with an even index. Now to form a connected group of length g (say), we need to connect g many index sets $\mathcal{I}_{l_1}, \mathcal{I}_{l_2}, \ldots, \mathcal{I}_{l_g}$ (say) and hence there must be $(g-1)$ matchings of the form $(u_{l_k,3i'_k}, v_{l_k,2i'_k}) = (u_{l_{k+1},3i_{k+1}-1}, v_{l_{k+1},2i_{k+1}-1})$ for some $i_k \leq i'_k$ and for all $1 \leq k \leq g-1$. Therefore, under $D(g_1, g_2, \ldots, g_R)$, to form a connected group of length g, there must be $(g-1)$ blocks which start with an even index. Hence by Lemma 6.5.1, as we have R connected groups of lengths $g_1, g_2, \ldots, g_R$,

$$\begin{aligned} & \lim \frac{1}{p} E_{D(g_1,g_2,\ldots,g_R)} \mathrm{Tr}(n^{T/2} p^{-T/2} \prod_{i=1}^{T} \mathbb{P}_i) \\ = \; & O(y_n^{-T/2+\sum(g_i-1)}) = O(y_n^{-T/2+T-R}) = O(y_n^{T/2-R}). \end{aligned} \tag{6.87}$$

Let

$$F(g_1, g_2, \ldots, g_R) = G(g_1, g_2, \ldots, g_R) - D(g_1, g_2, \ldots, g_R).$$

Then, by Lemma 6.5.1 and (6.87),

$$\lim \frac{1}{p} E_{F(g_1,g_2,\ldots,g_R)} \mathrm{Tr}(n^{T/2}p^{-T/2}\prod_{i=1}^{T}\mathbb{P}_i) = o(y_n^{T/2-R}). \tag{6.88}$$

Hence, by (6.87) and (6.88), proof of Lemma 6.6.2 is complete. □

Getting back to the proof of the theorem, by (6.83) and Lemma 6.6.2, we have

$$\begin{aligned}
&\lim p^{-1} E\mathrm{Tr}\left[\Pi_{i=1}^{T}\mathcal{R}_i\right] = \lim p^{-1} E_{C_{T+1}} \mathrm{Tr}(n^T p^{-T}\Pi_{i=1}^{T}\mathbb{P}_i) \\
= & \sum_{\substack{2\le g_1,g_2,\ldots,g_R\le T \\ \sum_{j=1}^{R} g_j=T,\ R\ge 1}} \lim p^{-1} E_{G(g_1,g_2,\ldots,g_R)} \mathrm{Tr}(n^T p^{-T}\Pi_{i=1}^{T}\mathbb{P}_i) \\
= & \begin{cases} 0, \text{ if } T \text{ is odd} \\ \lim p^{-1} E_{G(2,2,\ldots,2)} \mathrm{Tr}(n^{T/2}p^{-T/2}\Pi_{i=1}^{T}\mathbb{P}_i), \text{ if } T \text{ is even.} \end{cases}
\end{aligned} \tag{6.89}$$

Therefore, (6.72) is proved for odd T.

It remains to show that (6.72) and (6.89) are equivalent when T is even. Let $T = 2m$ and $D(2,2,\ldots,2)$ be the set of all non-crossing pair matchings in $G(2,2,\ldots,2)$. Then from the proof of Lemma 6.6.2, it is obvious that

$$\lim p^{-1} E_{G(2,2,\ldots,2)} \mathrm{Tr}(n^m p^{-m}\Pi_{i=1}^{2m}\mathbb{P}_i) = \lim p^{-1} E_{D(2,2,\ldots,2)} \mathrm{Tr}(n^m p^{-m}\Pi_{i=1}^{2m}\mathbb{P}_i). \tag{6.90}$$

Note that $D(2,2,\ldots,2)$ is the set of all non-crossing pair matchings each of which has $T/2$ many connected groups of length 2. Moreover, observe that we need at least one block start with even index to get a connected group of length 2. Hence, each matching in $D(2,2,\ldots,2)$ has at least $T/2$ blocks start with even index. Now consider $\mathcal{C} \subset D(2,2,\ldots,2)$ of matchings which have exactly $T/2$ many blocks starts with even index.

$$\mathcal{C} = \{\sigma_{\tau,(i_1,i_2,\ldots,i_{2m})} : \tau \in NC_2(2m)\}, \tag{6.91}$$

where for each $\tau = \{(l_1, l_2), (l_3, l_4), \ldots, (l_{2m-1}, l_{2m})\} \in NC_2(2m)$, $l_{2k-1} < l_{2k}$ for all k, we have

$$\begin{aligned}\sigma_{\tau,(i_1,i_2,\ldots,i_{2m})} &= \{(u_{l_{2k-1},3i_{2k-1}}, v_{l_{2k-1},2i_{2k-1}}) = (u_{l_{2k},3i_{2k}-1}, v_{l_{2k},2i_{2k}-1}), \forall 1 \leq \\ &\qquad k \leq m, (u_{l,3i-1}, v_{l,2i-1}) = (u_{l,3i}, v_{l,2i}),\ \forall i \neq i_l, 1 \leq l \leq 2m\}. \end{aligned} \tag{6.92}$$

Note that $D(2, 2, \ldots, 2) - \mathcal{C}$ has more than $T/2$ blocks that start with an even index. Therefore, by Lemma 6.5.1,

$$\lim p^{-1} E_{D(2,2,\ldots,2)-\mathcal{C}} \mathrm{Tr}(n^m p^{-m} \Pi_{i=1}^{2m} \mathbb{P}_i) = 0. \tag{6.93}$$

Hence by (6.89), (6.90) and (6.93), we have

$$\lim p^{-1} E \mathrm{Tr}\left[\Pi_{i=1}^{2m} \mathcal{R}_i\right] = \lim p^{-1} E_{\mathcal{C}} \mathrm{Tr}(n^m p^{-m} \Pi_{i=1}^{2m} \mathbb{P}_i). \tag{6.94}$$

Hence, it remains to show that the right sides of (6.72) and (7.119) match for $T = 2m$. Now it is easy to show that

$$\begin{aligned} &\sum_{\tau \in NC_2(2m)} \lim p^{-1} E_{\sigma_{\tau,(i_1,i_2,\ldots,i_{2m})}} \mathrm{Tr}(n^m p^{-m} \Pi_{i=1}^{2m} \mathbb{P}_i) \qquad (6.95) \\ = &\sum_{\tau \in NC_2(2m)} \Big[\varphi_{0K(\tau)}(c_{k,-i_k} a_{(k+1) \bmod 2m, -i_{(k+1) \bmod 2m}} : 1 \leq k \leq 2m) \\ &\qquad\qquad \kappa_\tau(w_{k,i_k} : 1 \leq k \leq 2m) \Big] \\ = &\ \varphi_0\Big(\prod_{k=1}^{2m} a_{k,-i_k} w_{k,i_k} c_{k,-i_k}\Big). \end{aligned}$$

$$\begin{aligned} \text{Now} \quad & \lim p^{-1} E_{\mathcal{C}} \mathrm{Tr}(n^m p^{-m} \Pi_{i=1}^{2m} \mathbb{P}_i) \\ = &\sum_{i_1,i_2,\ldots,i_{2m}} \sum_{\tau \in NC_2(2m)} \lim p^{-1} E_{\sigma_{\tau,(i_1,i_2,\ldots,i_{2m})}} \mathrm{Tr}(n^m p^{-m} \Pi_{i=1}^{2m} \mathbb{P}_i) \end{aligned}$$

$$\begin{aligned} &= \sum_{i_1, i_2, \ldots, i_{2m}} \varphi_0\Big(\prod_{k=1}^{2m} a_{k,-i_k} w_{k,i_k} c_{k,-i_k}\Big) = \varphi_0\Big(\prod_{k=1}^{2m} \sum_{i_k} a_{k,-i_k} w_{k,i_k} c_{k,-i_k}\Big) \\ &= \varphi_0\Big(\prod_{l=1}^{2m} \alpha_l\Big). \end{aligned} \tag{6.96}$$

Hence, by (7.119) and (6.96), $\lim p^{-1} E \mathrm{Tr}\,[\Pi_{i=1}^{2m} \mathcal{R}_i] = \varphi(\prod_{l=1}^{2m} \alpha_l)$. Therefore, proof of Theorem 6.3.1 (a) is complete.

(b) Proof of (b) is immediate from Theorem 6.3.1 (a) by observing the fact that proof of (a) will go through if instead of $\{A_{l,2i-1}\} \subset \{B_{2i-1}, B^*_{2i-1}\}$, we assume $\{A_{l,2i-1}\} \subset \mathrm{Span}\{B_{2i-1}, B^*_{2i-1}\}$.

Hence the proof of Theorem 6.3.1 is complete. $\square$

Chapter 7

Limiting spectral distribution of sample autocovariance matrices

7.1 Introduction

This chapter focuses on the LSD of symmetric polynomials in sample autocovariance matrices $\{\hat{\Gamma}_u\}$ for the infinite dimensional moving average processes. In the literature, such results are known only for the particular polynomial $\{\hat{\Gamma}_u+\hat{\Gamma}_u^*\}$ and under quite restrictive assumptions on the coefficient matrices $\{\psi_j\}$. In Section 7.2, we collect all the existing results in the literature.

We make use of the general results developed in Chapters 5 and 6 to deal with all symmetric polynomials in $\{\hat{\Gamma}_u\}$. In Theorems 7.3.1, 7.3.4, 7.3.15 and 7.3.17, we show that under significantly weaker conditions on $\{\psi_j\}$, the LSD of any symmetric polynomial in $\{\hat{\Gamma}_u\}$ exists for both the cases $p/n \to y > 0$ and $p/n \to 0$. Moreover, apparently for the first time in the literature, we describe the limits in terms of some free variables. Finally we show how the existing LSD results follow from our result.

In the next chapter we will see some statistical applications of these results.

The main material of this chapter is taken from Bhattacharjee and Bose [2015a] and Bhattacharjee and Bose [2015b].

7.2 Existing results and motivation

First consider the simplest case, the MA(0) process, defined in (3.5):

$$X_{t,p} = \varepsilon_t, \ \forall t. \tag{7.1}$$

For convenience, let us write X_t for $X_{t,p}$. Let $\varepsilon_{t,i}$ be the i-th element of ε_t. Consider the following assumption

(B1) $\{\varepsilon_{i,j}\}$ are independently distributed with mean 0 and variance 1.

Recall the class of independent random variables defined in (4.17):

$$\begin{aligned} U(\delta) \ &= \ \text{set of all collections of independent random variables } \{\varepsilon_{i,j} : i, j \geq 1\} \\ &\quad \text{such that } \lim \frac{\eta^{-(2+\delta)}}{np} \sum_{i=1}^{p} \sum_{j=1}^{n} E(|\varepsilon_{i,j}|^{2+\delta} I(|\varepsilon_{i,j}| > \eta p^{\frac{1}{2+\delta}}) = 0 \\ &\quad \text{for all } \eta > 0. \end{aligned} \tag{7.2}$$

Also recall the Marčenko-Pastur law MP_y with parameter $y > 0$ satisfying the moment sequence (4.25). The following results are known in the literature

Theorem 7.2.1. *(Bai and Silverstein [2009]) Consider the model (3.5) (or (7.1)). Suppose (B1) holds and $\{\varepsilon_{i,j} : i, j \geq 1\} \in U(0)$. Let $n, p(n) \to \infty$, $p/n \to y > 0$. Then the almost sure LSD of $\hat{\Gamma}_0$ is the MP_y law.*

Theorem 7.2.2. *(Jin et al. [2014]) Consider the model (3.5) (or (7.1)). Suppose (B1) holds and $\{\varepsilon_{i,j} : i, j \geq 1\} \in U(\delta)$ for some $\delta \in (0, 2]$. Let $n, p(n) \to \infty$, $p/n \to y > 0$. Then the almost sure LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ are identical for all $u \geq 1$ and the common limiting Stieltjes transformation satisfies (5.90).*

Next consider the infinite dimensional MA(∞) process defined in (3.2):

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \ \forall t. \tag{7.3}$$

Consider the following assumption.

(B2) $\{\varepsilon_{i,j}\}$ are i.i.d. random variables with mean 0, variance 1 and $E|\varepsilon_{i,j}|^4 < \infty$.

Recall I_k in (2.8). Then the following results are known.

Theorem 7.2.3. *(Pfaffel and Schlemm [2011]) Consider the model (3.2) (or (7.3)) with $\psi_{j,p} = \lambda_j I_p,\ \lambda_j \in \mathbb{R},\ \lambda_0 = 1$ and $\sum_{j=1}^{\infty} |\lambda_j| < \infty$. Suppose (B2) holds and $n, p(n) \to \infty,\ p/n \to y > 0$. Then for each $u \geq 1$, the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ exists almost surely and the limiting Stieltjes transformation $m_u(z)$ satisfies the following equations (only one solution yields a valid Stieltjes transform)*

$$\begin{aligned} z &= -\frac{1}{m_u(z)} + \frac{1}{2\pi}\int_0^{2\pi} \frac{d\theta}{y m_u(z) + f^{-1}(\theta)},\ z \in \mathbb{C}^+ \text{ where} && (7.4)\\ f(\theta) &= \cos(u\theta)|\sum_{k=0}^{\infty} \lambda_k e^{ik\theta}|^2. && (7.5)\end{aligned}$$

Now let us consider general $\{\psi_j\}$ which satisfy the following assumption.

(WAP) $\{\psi_j\}$ are Hermitian, simultaneously diagonalizable and norm bounded. There are continuous functions $f_j : \mathbb{R}^m \to \mathbb{R}$ and a $p \times p$ unitary matrix U such that $U\psi_j U^* = \text{diag}(f_j(\alpha_1), f_j(\alpha_2), \ldots, f_j(\alpha_p))$, $\alpha_j \in \mathbb{R}^m$ for all j and some positive integer m. The distribution which puts mass $1/p$ at each α_i, converges weakly to a compactly supported probability distribution F on $\mathbb{R}^m$.

Then the following results are known.

Theorem 7.2.4. *(Liu et al. [2015]) Consider the model (3.2) (or (7.3)). Suppose (B2) and (WAP) hold, $\sum_{j=0}^{\infty} |f_j(\alpha)| < \infty,\ \forall \alpha \in \mathbb{R}^m$ and $n, p(n) \to \infty,\ p/n \to y > 0$. Then for each $u \geq 1$, the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ exists almost surely and the limiting Stieltjes transformation $m_u(z)$ satisfies*

$$m_u(z) = \int_{\mathbb{R}^m} \left(\frac{1}{2\pi}\int_0^{2\pi} \frac{\cos(u\theta')h_1(\alpha,\theta')d\theta'}{1 + y\cos(u\theta')K_u(z,\theta')} - z \right)^{-1} dF(\alpha),\ z \in \mathbb{C}^+ \text{ where} \quad (7.6)$$

$$
\begin{aligned}
K_u(z,\theta) &= \int_{\mathbb{R}^m}\left(\frac{1}{2\pi}\int_0^{2\pi}\frac{\cos(u\theta')h_1(\alpha,\theta')d\theta'}{1+y\cos(u\theta')K_u(z,\theta')}-z\right)^{-1}h_1(\alpha,\theta)dF(\alpha),\ z\in\mathbb{C}^+ \quad (7.7)\\
h_1(\alpha,\theta) &= |\sum_{j=0}^{\infty}e^{ij\theta}f_j(\alpha)|^2,\quad \alpha\in\mathbb{R}^m. \quad (7.8)
\end{aligned}
$$

Theorem 7.2.5. *(Wang et al. [2015]) Consider the model (3.2) (or (7.3)). Suppose (B2) and (WAP) hold, $\sum_{j=0}^{\infty}|f_j(\alpha)| < \infty\ \forall\alpha\in\mathbb{R}^m$ and $p, n(p)\to\infty$, $p/n\to 0$. Then for each $u\geq 1$, the LSD of $\frac{1}{2}(\hat{\Gamma}_u+\hat{\Gamma}_u^*)$ exists almost surely and the limiting Stieltjes transformation $m_u(z)$ satisfies*

$$
\begin{aligned}
m_u(z) &= -\int_{\mathbb{R}^m}\frac{dF(\alpha)}{z+\beta_u(z,\alpha)},\ z\in\mathbb{C}^+,\ \textit{where} \quad (7.9)\\
\beta_u(z,x) &= -\int_{\mathbb{R}^m}\frac{R_u(x,\alpha)dF(\alpha)}{z+\beta_u(z,\alpha)},\ z\in\mathbb{C}^+, x\in\mathbb{R}^m, \quad (7.10)\\
R_u(x,y) &= \frac{1}{2\pi}\int_0^{2\pi}\cos^2(u\theta)h_1(x,\theta)h_1(y,\theta)d\theta,\ x,y\in\mathbb{R}^m, \quad (7.11)
\end{aligned}
$$

and $h_1(\cdot,\cdot)$ is as in (7.8).

The assumption (WAP) on $\{\psi_j\}$ made by Liu et al. [2015] and Wang et al. [2015] is however, quite restrictive. Later Liu et al. [2015] replaced (WAP) by the assumption that $\{\psi_j\}$ are Toeplitz matrices with suitable decay conditions on their entries. Even so, this excludes many interesting linear processes. For example, consider the following MA(2) process

$$X_t = \varepsilon_t + C\varepsilon_{t-1} + D\varepsilon_{t-2},\ \forall t, \quad (7.12)$$

and C and D are respectively as in Examples 4.2.3 and 4.2.4:

$$C = ((I(1\leq i=j\leq [p/2]) - I([p/2]+1\leq i=j\leq p))),\ D = ((I(i+j=p+1))),$$

and $[x]$ is the largest integer in x. Note that for the above model (7.12), the coefficient matrices C and D are neither simultaneously diagonalizable (as $CD\neq DC$) nor are they Toeplitz matrices. Hence Theorems 7.2.4 and 7.2.5 are not applicable for the model (7.12).

To indicate another limitation of (WAP), suppose further that $\varepsilon_t \sim \mathcal{N}(0, I_p)$. Let U be a unitary matrix such that $U\psi_j U^* =: \Lambda_j$ (say) are diagonal matrices. Since $U\varepsilon_t$ and ε_t are identically distributed and $UU^* = I_p$, as far as the LSD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ is concerned, (3.2) (or (7.3)) is equivalent to the model

$$X_{t,i} \ (i\text{-th component of } X_t) \ = \sum_{j=0}^{\infty} \psi_{j,(i,i)} \varepsilon_{t-j,i}, \forall i \geq 1 \tag{7.13}$$

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

First note that all the existing works concentrate on $\hat{\Gamma}_u + \hat{\Gamma}_u^*$. We may be interested in other functions of Γ_u. For example, if we wish to study the *singular values* of $\hat{\Gamma}_u$, we need to consider $\hat{\Gamma}_u \hat{\Gamma}_u^*$. This gives rise to a completely different LSD problem. Indeed, one may consider more general symmetrizations that involve several $\hat{\Gamma}_u$. As we may recall, in the one dimensional case, all tests for white noise are based on quadratic functions of autocovariances. See for example Shao [2011] and Xiao and Wu [2014]. The analogous objects in our model are quadratic polynomials in autocovariances. Thus we are naturally led to the consideration of matrix polynomials of autocovariances.

Second, as we have seen above, the (WAP) condition is fairly strong. We shall replace this condition by a more natural and much weaker joint convergence assumption (Assumption (B) in Section 7.3).

Finally, all the above results are derived using Stieltjes transformation method. While it is conceivable that this method can be potentially used to tackle these cases, it seems to be rather cumbersome and needlessly lengthy to do so and shall at best be a case by case study.

In the next section, we provide a unified method to study the LSD of symmetric polynomials of the autocovariance matrices using the tools and results from Chapters 4-6. We do not use Stieltjes transforms at all except to cross-check our results with the existing results, all of which follow as special cases. In the next chapter, we shall use these results in statistical applications.

7.3 LSD of any symmetric polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$

We first list the assumptions. Recall the following classes of independent random variables in (4.14), (4.15) and (4.16):

$$\begin{aligned}
\mathcal{L}_r &= \text{set of all collections of independent random variables} \qquad (7.14)\\
&\quad \{\epsilon_{i,j} : i, j \geq 1\} \text{ such that } \sup_{i,j} E|\varepsilon_{i,j}|^r < \infty,\\
\mathcal{L} &= \bigcap_{r=1}^{\infty} \mathcal{L}_r, \qquad (7.15)\\
C(\delta, p) &= \text{set of all collections of random variables } \{\varepsilon_{i,j} : i, j \geq 1\} \text{ such that}\\
&\quad P(|\varepsilon_{i,j}| \leq \eta_p p^{\frac{1}{2+\delta}}) = 1, \quad \forall i, j \text{ and for some } \eta_p \downarrow 0 \text{ as } p \to \infty. \qquad (7.16)
\end{aligned}$$

Consider the following assumption on $\{\varepsilon_{i,j}\}$.

(B3) $\{\varepsilon_{i,j} : 1 \leq i \leq p,\ 1 \leq j \leq n\} \in \mathcal{L} \cup C(\delta, p)$ for all $p \geq 1$ and for some $\delta > 0$.

Later we relax $(B3)$ for specific polynomials.

We consider the same assumption on $\{\psi_j\}$ as in Corollaries 5.4.7 and 6.4.6. For convenience of the reader, here we state it again. Recall the collection of some $p \times p$ matrices $\{B_{2i-1}\}$ satisfying Assumption (A2) in Chapters 5 and 6.

Suppose $\{\psi_j\} \subset \{B_{2i-1}, B_{2i-1}^*\}$ i.e. we assume:

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

Suppose

$$(\text{Span}\{\psi_j, \psi_j^* : j \geq 0\}, p^{-1}\text{Tr}) \to (\text{Span}\{\eta_j, \eta_j^* : j \geq 0\}, \varphi_{\text{odd}}), \qquad (7.17)$$

$$(\text{Span}\{\bar{\psi}_j, \bar{\psi}_j^* : j \geq 0\}, (n+p)^{-1}\text{Tr}) \to (\text{Span}\{\bar{\eta}_j, \bar{\eta}_j^* : j \geq 0\}, \bar{\varphi}_{\text{odd}}). \qquad (7.18)$$

Recall the NCP $(\mathcal{A}_{\text{odd}}, \varphi_{\text{odd}})$ and $(\bar{\mathcal{A}}_{\text{odd}}, \bar{\varphi}_{\text{odd}})$ respectively in (5.13) and (5.14). Clearly the NCP at right side of (7.17) and (7.18) are sub-algebras of $(\mathcal{A}_{\text{odd}}, \varphi_{\text{odd}})$ and $(\bar{\mathcal{A}}_{\text{odd}}, \bar{\varphi}_{\text{odd}})$. Moreover, for any polynomial Π,

$$\begin{aligned}
\varphi_{\text{odd}}(\Pi(\eta_j, \eta_j^* : j \geq 0)) &= \frac{1+y}{y}\bar{\varphi}_{\text{odd}}(\Pi(\bar{\eta}_j, \bar{\eta}_j^* : j \geq 0)) \\
&= \frac{1+y}{y}\bar{\varphi}(\Pi(\bar{\eta}_j, \bar{\eta}_j^* : j \geq 0)).
\end{aligned}$$

7.3.1 LSD for $p/n \to y > 0$

To describe the LSD of any symmetric polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$, we need the matrices $\{P_i\}$ defined at the beginning of Section 5.1. This is because, later we shall see that for LSD purposes, $\{\hat{\Gamma}_u\}$ can be approximated by $\{\Delta_u\}$ (see (5.1)) where $\{\Delta_u\}$ is of the form (5.2) with $\{B_{2i}\} = \{P_i : i = 0, \pm 1, \pm 2, \ldots\}$. Hence by (5.16), (5.81) and (5.80), we now have

$$\begin{gathered}
\bar{\mathcal{A}}_{\text{even}} = \text{Span}\{\underline{c}_u, \underline{c}_u^* = \underline{c}_{-u} : u = 0, \pm 1, \ldots\}, \text{ and} \\
(\text{Span}\{\underline{P}_u, \underline{P}_u^* : u = 0, \pm 1, \ldots\}, (n+p)^{-1}\text{Tr}) \to (\bar{\mathcal{A}}_{\text{even}}, \bar{\varphi}_{\text{even}}),
\end{gathered}$$

where for all $T \geq 1$ and $i_1, i_2, \ldots, i_T = 0, \pm 1, \pm 2, \ldots$, we have

$$\bar{\varphi}_{\text{even}}(\prod_{j=1}^{T} \underline{c}_{i_j}) = \lim \frac{1}{n+p}\text{Tr}(\prod_{j=1}^{T} \underline{P}_{i_j}) = \frac{1}{1+y} I(\sum_{j=1}^{T} i_j = 0). \tag{7.19}$$

Recall the NCP $(\text{Span}\{s_u\}, \varphi_s)$ of free semicircle variables, defined at the beginning of Section 5.3. Let $s \in \{s_u\}$ be any typical standard semi-circle variable and $(\text{Span}\{s\}, \varphi_s)$ be the NCP generated by s with moment sequence $\{\varphi_s(s^k) = \beta_k\}$ where $\{\beta_k\}$ is given in (4.11).

Recall the NCP $(\mathcal{A}, \bar{\varphi})$ defined in (5.19), where $\{\bar{\eta}_j, \bar{\eta}_j^*\}$, $\{\underline{c}_j, \underline{c}_j^*\}$ and s are free.

Consider the following polynomials in $\mathcal{A}$

$$\bar{\gamma}_{uq} = (1+y)\sum_{j,j'=0}^{q} \bar{\eta}_j s \underline{c}_{j-j'+u} s \bar{\eta}_{j'}^*, \quad \forall u, q \geq 0. \tag{7.20}$$

Then we have the following Theorem (see Bhattacharjee and Bose [2015a]).

Theorem 7.3.1. *Consider the model (3.7) given by*

$$X_t = \sum_{j=0}^{q} \psi_j \varepsilon_{t-j}, \ \forall t. \tag{7.21}$$

Suppose (B1), (B3) and (B) hold and $n, p(n) \to \infty$, $p/n \to y > 0$. *Then the LSD of any symmetric polynomial* $\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)$ *in* $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$ *exists almost surely and it is uniquely determined by the moment sequence*

$$\lim p^{-1} E\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0))^k = \frac{1+y}{y}\bar{\varphi}(\Pi(\bar{\gamma}_{uq}, \bar{\gamma}_{uq}^* : u \geq 0))^k \quad \forall k \geq 1. \tag{7.22}$$

Proof. To prove the above theorem, we use the moment method and Lemma 4.2.1 as described in Section 4.2. Recall $\{\Delta_u\}$ in (5.1):

$$\Delta_u = \frac{1}{n}\sum_{j,j'} \psi_j Z P_{j-j'+u} Z^* \psi_{j'}^*. \tag{7.23}$$

The following lemma describes the approximation of $\{\hat{\Gamma}_u\}$ by $\{\Delta_u\}$ and is useful to establish (M1) and (M4). Proof of this lemma is very technical and is presented in Section 7.4.

Lemma 7.3.2. *Consider the model (3.7) (or (7.21)). Suppose (B1), (B3) and (B) hold and* $n, p(n) \to \infty$, $p/n \to y > 0$. *Then the following statements are true.*

(a) For any polynomial Π,

$$\lim p^{-1} E\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)) = \lim p^{-1} E\mathrm{Tr}(\Pi(\Delta_u, \Delta_u^* : u \geq 0)). \tag{7.24}$$

(b) Let, for $1 \leq i \leq T$, m_i *be polynomials. Let for all* $1 \leq i \leq T$,

$$\begin{aligned}\mathcal{P}_i &= \mathrm{Tr}(m_i(\Delta_u, \Delta_u^* : u \geq 0)), \quad \mathcal{P}_i^0 = E\mathcal{P}_i,\\ \tilde{\mathcal{P}}_i &= \mathrm{Tr}(m_i(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)), \quad \tilde{\mathcal{P}}_i^0 = E\tilde{\mathcal{P}}_i.\end{aligned}$$

Then we have

$$\lim E(\prod_{i=1}^{T}(\tilde{\mathcal{P}}_i - \tilde{\mathcal{P}}_i^0)) = \lim E(\prod_{i=1}^{T}(\mathcal{P}_i - \mathcal{P}_i^0)). \tag{7.25}$$

Now to continue the proof of the theorem, by Theorem 5.3.1, for any polynomial Π,

$$\lim p^{-1} E\mathrm{Tr}(\Pi(\Delta_u, \Delta_u^* : u \geq 0))^k = \frac{1+y}{y}\bar{\varphi}(\Pi(\bar{\gamma}_{uq}, \bar{\gamma}_{uq}^* : u \geq 0))^k \quad \forall k \geq 1. \tag{7.26}$$

Hence by Lemma 7.3.2 (a),

$$\lim p^{-1} E\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0))^k = \frac{1+y}{y}\bar{\varphi}(\Pi(\bar{\gamma}_{uq}, \bar{\gamma}_{uq}^* : u \geq 0))^k \quad \forall k \geq 1. \tag{7.27}$$

This establishes (M1).

By Lemmas 5.4.2 and 7.3.2 (b), we have for any polynomial Π

$$E\left[\frac{1}{p}\mathrm{Tr}\left(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* :\ u \geq 0)\right)^h - E\left(\frac{1}{p}\mathrm{Tr}\left(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* :\ u \geq 0)\right)^h\right)\right]^4 = O(p^{-4}). \tag{7.28}$$

and hence (M4) is established.

Proof of (C) follows the same arguments as in the proof of (C) in Theorem 5.4.1. This completes the proof of Theorem 7.3.1. □

Now we move to the case $q = \infty$. Recall $||\cdot||_2$ defined in (2.4). Consider the following assumption on $\{\psi_j\}$.

(B4) $\sum_{j=0}^{\infty} \sup_p ||\psi_j||_2 < \infty$.

To describe the LSD of any symmetric polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$ for the MA(∞) process defined in (3.2) (or (7.3)), we need the following lemma.

Lemma 7.3.3. *Suppose (B4) holds. Let* $\epsilon_i = 1, * \ \forall i \geq 1$. *Then we have the following result.*

(a) For all $K \geq 1$, *we have*

$$\sum_{\substack{1 \leq j_i, j_i' \leq \infty \\ 1 \leq i \leq K}} |\bar{\varphi}\left(\prod_{i=1}^{K} \bar{\eta}_{j_i} s \underline{c}_{j_i - j_1' + u_i}^{\epsilon_i} s \bar{\eta}_{j_i'}^*\right)| \leq (2C)^{2K}, \quad \textit{where} \ \ C = \sum_{j=0}^{\infty} \sup_p ||\psi_j||_2^2. \tag{7.29}$$

(b) For any polynomial $\Pi(\bar{\gamma}_{uq}, \bar{\gamma}_{uq}^* : u \geq 0)$, $\lim_{q \to \infty} \bar{\varphi}(\Pi(\bar{\gamma}_{uq}, \bar{\gamma}_{uq}^* : u \geq 0))$ *exists and is finite.*

Proof. (a) By Lemma 4.3.6,

$$\bar{\varphi}\left(\prod_{i=1}^{K} \bar{\eta}_{j_i} s \underline{c}_{j_i - j_i' + u_i}^{\epsilon_i} s \bar{\eta}_{j_i'}^*\right) = \sum_{\pi \in NC_2(2K)} \bar{\varphi}_{K(\pi)}[\bar{\eta}_{j_1}, \underline{c}_{j_1 - j_1' + u_1}^{\epsilon_1}, \bar{\eta}_{j_1'}^*, \ldots, \bar{\eta}_{j_k}, \underline{c}_{j_k - j_k' + u_k}^{\epsilon_k}, \bar{\eta}_{j_k'}^*].$$

Therefore, by Lemma 4.3.1 (b) and as $\#NC_2(2K) \leq 2^{2K}$, we have some $h_i, r_i \geq 1$ such that

$$|\bar{\varphi}\left(\prod_{i=1}^{K} \bar{\eta}_{j_i} s \underline{c}_{j_i - j_i' + u_i}^{\epsilon_i} s \bar{\eta}_{j_i'}^*\right)| \leq 2^{2k} \prod_{i=1}^{K} (\bar{\varphi}(\bar{\eta}_{j_i}^* \bar{\eta}_{j_i})^{h_i})^{1/h_i} \prod_{i=1}^{K} (\bar{\varphi}(\underline{c}_{j_i - j_i' + u_i}^* \underline{c}_{j_i - j_i' + u_i})^{r_i})^{1/r_i}. \tag{7.30}$$

Now, by (7.19),

$$\bar{\varphi}(\underline{c}_u^* \underline{c}_u)^r \leq 1, \quad \forall r, u \geq 1. \tag{7.31}$$

Also , for all $j \geq 1$, we have

$$\bar{\varphi}(\bar{\eta}_j^* \bar{\eta}_j) = \lim p^{-1} \mathrm{Tr}(\bar{\psi}_j^* \bar{\psi}_j) \leq \sup_p ||\psi_j||_2^2. \tag{7.32}$$

Hence, by (7.30), (7.31) and (7.32), we have

$$|\bar{\varphi}\left(\prod_{i=1}^{K} \bar{\eta}_{j_i} s\underline{c}^{\epsilon_i}_{j_i - j'_i + u_i} s\bar{\eta}^*_{j'_i}\right)| \leq 2^{2K} \prod_{i=1}^{K} \sup_p ||\psi_{j_i}||_2^2. \tag{7.33}$$

Hence, under Assumption (B4), Lemma 7.3.3 (a) holds by summing both sides of (7.33) over j_i, j'_i for all $1 \leq i \leq K$.

(b) Note that without loss of generality, we can take

$$\Pi(\bar{\gamma}_{uq}, \bar{\gamma}^*_{uq} : u \geq 0) = \sum_{j=1}^{r} m_{l_j}, \quad \text{where } \; m_{l_j} = \prod_{i=1}^{l_j} \bar{\gamma}^{\epsilon_{j,i}}_{u_{j,i}q}, \; \epsilon_{j,i} = 1, *. \tag{7.34}$$

Now, by (7.20)

$$\bar{\varphi}(\Pi(\bar{\gamma}_{uq}, \bar{\gamma}^*_{uq} : u \geq 0)) = \sum_{j=1}^{r} \sum_{\substack{1 \leq k_{i,j}, k'_{i,j} \leq q \\ 1 \leq i \leq l_j}} \bar{\varphi}\bigg(\prod_{i=1}^{l_j} \bar{\eta}^{\epsilon_{j,i}}_{k_{j,i}} s\underline{c}^{\epsilon_{j,i}}_{k_{j,i} - k'_{j,i} + u_{j,i}} s\bar{\eta}^{*(1-\epsilon_{j,i})}_{k'_{i,j}}\bigg).$$

Hence, by Lemma 7.3.3 (a), under Assumption (B4), we have

$$\lim_{q \to \infty} \bar{\varphi}(\Pi(\bar{\gamma}_{uq}, \bar{\gamma}^*_{uq} : u \geq 0)) = \sum_{j=1}^{r} \sum_{\substack{1 \leq k_{i,j}, k'_{i,j} \leq \infty \\ 1 \leq i \leq l_j}} \bar{\varphi}\bigg(\prod_{i=1}^{l_j} \bar{\eta}^{\epsilon_{j,i}}_{k_{j,i}} s\underline{c}^{\epsilon_{j,i}}_{k_{j,i} - k'_{j,i} + u_{j,i}} s\bar{\eta}^{*(1-\epsilon_{j,i})}_{k'_{i,j}}\bigg),$$

which is finite. This completes the proof of Lemma 7.3.3 (b). $\square$

Now consider the NCP $(\mathcal{A}_\infty, \bar{\varphi}_\infty)$ where

$$\mathcal{A}_\infty = \text{Span}\{\bar{\gamma}_{u\infty}, \bar{\gamma}^*_{u\infty} : u \geq 0\} \tag{7.35}$$

and for any polynomial $\Pi(\bar{\gamma}_{u\infty}, \bar{\gamma}^*_{u\infty} : u \geq 0)$,

$$\bar{\varphi}_\infty(\Pi(\bar{\gamma}_{u\infty}, \bar{\gamma}^*_{u\infty} : u \geq 0)) = \lim_{q \to \infty} \bar{\varphi}(\Pi(\bar{\gamma}_{uq}, \bar{\gamma}^*_{uq} : u \geq 0)). \tag{7.36}$$

The existence of the limit at right side of (7.36) is guaranteed by Lemma 7.3.3 (b). Now we have the following theorem (see Bhattacharjee and Bose [2015a]).

Theorem 7.3.4. *Consider the model (3.2) (or (7.3)). Suppose (B1), (B3), (B) and (B4) hold and $n, p(n) \to \infty$, $p/n \to y > 0$. Then the LSD of any symmetric polynomial $\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)$ in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$ exists almost surely and it is uniquely determined by the moment sequence*

$$\lim p^{-1} E\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0))^k = \frac{1+y}{y} \bar{\varphi}_\infty(\Pi(\bar{\gamma}_{u\infty}, \bar{\gamma}_{u\infty}^* : u \geq 0))^k \quad \forall k \geq 1. \tag{7.37}$$

To prove the above theorem, we need the following Lemmas.

Lemma 7.3.5. *Suppose (B4) holds. Then we have the following results.*

(a) For any symmetric polynomial $\Pi(\bar{\gamma}_{uq}, \bar{\gamma}_{uq}^* : u \geq 0)$, there exists a unique probability measure F_q on $\mathbb{R}$ such that

$$\int x^K dF_q = \frac{1+y}{y} \bar{\varphi}(\Pi(\bar{\gamma}_{uq}, \bar{\gamma}_{uq}^* : u \geq 0))^K, \ \forall K \geq 1. \tag{7.38}$$

(b) For any symmetric polynomial $\Pi(\bar{\gamma}_{u\infty}, \bar{\gamma}_{u\infty}^* : u \geq 0)$, there exists a unique probability measure F on $\mathbb{R}$ such that

$$\int x^K dF = \frac{1+y}{y} \bar{\varphi}_\infty(\Pi(\bar{\gamma}_{u\infty}, \bar{\gamma}_{u\infty}^* : u \geq 0))^K, \ \forall K \geq 1. \tag{7.39}$$

(c) F_q converges weakly to F as $q \to \infty$.

Proof. First note that by (7.22), the right side of (7.38) is a moment sequence. By using (7.20) and Lemma 7.3.3 (a), it is easy to see that for all $u_1, u_2, \ldots, u_k \geq 0$, $\epsilon_1, \epsilon_2, \ldots, \epsilon_k = 1, *$ and $K \geq 1$, we have

$$|\bar{\varphi}\Big(\prod_{i=1}^{K} \bar{\gamma}_{u_i q}^{\epsilon_i}\Big)| \leq \ (2C)^K, \tag{7.40}$$

where C is as in Lemma 7.3.3 (a).

Therefore, expressing Π in the form (7.34), we have

$$|\bar{\varphi}(\Pi(\bar{\gamma}_{uq}, \bar{\gamma}_{uq}^* : u \geq 0))^K| = \sum_{1 \leq j_1 \ldots j_K \leq r} |\bar{\varphi}(\prod_{u=1}^{K} m_{l_{j_u}})| \leq \sum_{1 \leq j_1 \ldots j_K \leq r} (2C)^{\sum_{u=1}^{K} l_{j_u}} \leq (C')^K,$$

where $C' > 0$ does not depend on q. Hence, by Lemma 4.2.2 (b), proof of Lemma 7.3.5 (a) is complete.

Now note that by (7.36) and (7.38), the right side of (7.39) is a moment sequence. As $C' > 0$ does not depend on q, by (7.36), we have

$$|\bar{\varphi}_\infty(\Pi(\bar{\gamma}_{u\infty}, \bar{\gamma}_{u\infty}^* : u \geq 0))^K| \leq (C')^K. \tag{7.41}$$

Hence, by Lemma 4.2.2 (a), proof of Lemma 7.3.5 (b) is complete.

Lemma 7.3.5 (c) is trivial by (7.36). $\square$

Lemma 7.3.6. *For any non-commutative variables $\{a_i, b_i : 1 \leq i \leq k\}$ we have*

$$\prod_{i=1}^{k} a_i - \prod_{i=1}^{k} b_i = \sum_{j=1}^{k} \left(\prod_{i=1}^{j-1} a_i\right) (a_j - b_j) \left(\prod_{i=j+1}^{k} b_i\right), \ \forall k \geq 2, \ a_0 = b_{k+1} = 1. \tag{7.42}$$

Proof. We prove this Lemma using induction on k. Note that (7.42) is true for $k = 2$, as $a_1a_2 - b_1b_2 = (a_1 - b_1)b_2 + a_1(a_2 - b_2)$. Suppose (7.42) is true for $k = m$. Then for $k = m + 1$, note that

$$\begin{aligned}
& \left(\prod_{i=1}^{m} a_i\right) a_{m+1} - \left(\prod_{i=1}^{m} b_i\right) b_{m+1} \\
= & \left(\prod_{i=1}^{m} a_i - \prod_{i=1}^{m} b_i\right) b_{m+1} + \left(\prod_{i=1}^{m} a_i\right) (a_{m+1} - b_{m+1}), \quad \text{using (7.42) for } k = 2 \\
= & \sum_{j=1}^{m} \left(\prod_{i=1}^{j-1} a_i\right) (a_j - b_j) \left(\prod_{i=j+1}^{m} b_i\right) b_{m+1} + \left(\prod_{i=1}^{m} a_i\right) (a_{m+1} - b_{m+1}), \\
& \qquad\qquad\qquad\qquad \text{using (7.42) for } k = m
\end{aligned}$$

$$= \sum_{j=1}^{m+1} \left(\prod_{i=1}^{j-1} a_i \right) (a_j - b_j) \left(\prod_{i=j+1}^{m+1} b_i \right).$$

Hence, the proof is complete. □

Let,

$$F_{p,q} = \text{ESD of } \Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0) \text{ for the MA}(q) \text{ process,} \tag{7.43}$$

$$F_{p,\infty} = \text{ESD of } \Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0) \text{ for the MA}(\infty) \text{ process.} \tag{7.44}$$

Let $L(F, G)$ be the Levy distance between distribution functions F and G.

Lemma 7.3.7. $\lim_{q\to\infty} \lim_{p\to\infty} L(F_{p,q}, F_{p,\infty}) = 0$ *almost surely.*

Proof. For convenience, in this proof, let us denote $\hat{\Gamma}_u$ for MA(q) and MA(∞) processes respectively by $\hat{\Gamma}_{uq}$ and $\hat{\Gamma}_{u\infty}$. Let,

$$g_q = \Pi(\hat{\Gamma}_{uq}, \hat{\Gamma}_{uq}^* : u \geq 0), \; g_\infty = \Pi(\hat{\Gamma}_{u\infty}, \hat{\Gamma}_{u\infty}^* : u \geq 0).$$

To prove Lemma 7.3.7, by Corollary A.41 in Bai and Silverstein [2009], it is enough to show

$$\lim_{q\to\infty} \lim_{p\to\infty} \frac{1}{p} \text{Tr}(g_q - g_\infty)(g_q - g_\infty)^* \to 0, \text{ almost surely.} \tag{7.45}$$

Let us first prove (7.45) in the simplest case when $g_q = \hat{\Gamma}_{0q}$ and $g_\infty = \hat{\Gamma}_{0\infty}$. Recall the matrices $\{\Delta_u\}$ in (5.1). For convenience, in this proof, let us denote these matrices respectively for $q < \infty$ and $q = \infty$ by $\{\Delta_{uq}\}$ and $\{\Delta_{u\infty}\}$. Note that there is a $C > 0$ such that

$$\begin{aligned} \frac{1}{p}\text{Tr}(\hat{\Gamma}_{0q} - \hat{\Gamma}_{0\infty})^2 \leq\; & C\Big(\frac{1}{p}\text{Tr}(\hat{\Gamma}_{0q} - \Delta_{0q})^2 + \frac{1}{p}\text{Tr}(\Delta_{0q} - \Delta_{0\infty})^2 \\ & + \frac{1}{p}\text{Tr}(\Delta_{0\infty} - \hat{\Gamma}_{0\infty})^2\Big). \end{aligned} \tag{7.46}$$

Using similar techniques as in the proof of Lemma 7.3.2, it can be proved that as $p \to \infty$ (to prove (7.48), we additionally need Assumption (B4)),

$$\frac{1}{p}\mathrm{Tr}(\hat{\Gamma}_{0q} - \Delta_{0q})^2 \overset{\text{a.s.}}{\to} 0 \ \forall q \geq 0, \text{ and} \tag{7.47}$$

$$\frac{1}{p}\mathrm{Tr}(\Delta_{0\infty} - \hat{\Gamma}_{0\infty})^2 \overset{\text{a.s.}}{\to} 0. \tag{7.48}$$

We omit the details. Therefore, by (7.46), (7.47) and (7.48), proof of (7.45) when $g_q = \hat{\Gamma}_{0q}$ and $g_\infty = \hat{\Gamma}_{0\infty}$ will be completed if we can show

$$\lim_{q\to\infty}\lim_{p\to\infty}\frac{1}{p}\mathrm{Tr}(\Delta_{0q} - \Delta_{0\infty})^2 \overset{\text{a.s.}}{=} 0. \tag{7.49}$$

To show (7.49), now note that

$$\begin{aligned}
&\frac{1}{p}\mathrm{Tr}\Big[\sum_{j=0}^{\infty}\sum_{j'=0}^{\infty}\psi_j Z P_{j'-j} Z^* \psi_{j'}^* - \sum_{j=0}^{q}\sum_{j'=0}^{q}\psi_j Z P_{j'-j} Z^* \psi_{j'}^*\Big]^2 \\
= \ &\frac{1}{p}\mathrm{Tr}\Big[\sum_{j=q+1}^{\infty}\sum_{j'=q+1}^{\infty}\psi_j Z P_{j'-j} Z^* \psi_{j'}^* + \sum_{j=q+1}^{\infty}\sum_{j'=0}^{q}\psi_j Z P_{j'-j} Z^* \psi_{j'}^* \\
&\qquad + \sum_{j=0}^{q}\sum_{j'=q+1}^{\infty}\psi_j Z P_{j'-j} Z^* \psi_{j'}^*\Big]^2 \qquad (7.50) \\
= \ &\frac{1}{p}\mathrm{Tr}\Big[\sum_{j,j'=q+1}^{\infty}\sum_{k,k'=q+1}^{\infty}\psi_j Z P_{j'-j} Z^* \psi_{j'}^* \psi_k Z P_{k'-k} Z^* Z_{k'}^* \\
&\qquad + \sum_{j,k=q+1}^{\infty}\sum_{j',k'=0}^{q}\psi_j Z P_{j'-j} Z^* \psi_{j'}^* \psi_k Z P_{k'-k} Z^* \psi_{k'}^* \\
&\qquad + \sum_{j,k=0}^{q}\sum_{j',k'=q+1}^{\infty}\psi_j Z P_{j'-j} Z \psi_{j'}^* \psi_k Z P_{k'-k} Z^* \psi_{k'}^* \\
&\qquad + \sum_{j,k'=q+1}^{\infty}\sum_{j',k=0}^{q}\psi_j Z P_{j'-j} Z^* \psi_{j'}^* \psi_k Z P_{k'-k} Z^* \psi_{k'}^* \\
&\qquad + \sum_{j,k'=0}^{q}\sum_{j',k=q+1}^{\infty}\psi_j Z P_{j'-j} Z^* \psi_{j'}^* \psi_k Z P_{k'-k} Z^* \psi_{k'}^*
\end{aligned}$$

$$+ \sum_{j,j',k=q+1}^{\infty} \sum_{k'=0}^{q} \psi_j Z P_{j'-j} Z^* \psi_{j'}^* \psi_k Z P_{k'-k} Z^* \psi_{k'}^*$$
$$+ \sum_{j,k,k'=q+1}^{\infty} \sum_{j'=0}^{q} \psi_j Z P_{j'-j} Z^* \psi_{j'}^* \psi_k Z P_{k'-k} Z^* \psi_{k'}^*$$
$$+ \sum_{j,j',k'=q+1}^{\infty} \sum_{k=0}^{q} \psi_j Z P_{j'-j} Z^* \psi_{j'}^* \psi_k Z P_{k'-k} Z^* \psi_{k'}^*$$
$$\left. + \sum_{j=0}^{q} \sum_{j',k,k'=q+1}^{\infty} \psi_j Z P_{j'-j} Z^* \psi_{j'}^* \psi_k Z P_{k'-k} Z^* \psi_{k'}^* \right].$$
$$= \sum_{i=1}^{9} T_i, \quad \text{say.} \tag{7.51}$$

Using the same technique as in the proof of Lemma 5.4.2, under (B4), it can be shown that as $p \to \infty$,

$$E(T_i - ET_i)^4 = O(p^{-4}), \ \forall\ 1 \le i \le 9, \ q \ge 1. \tag{7.52}$$

Moreover, under (B4), one can easily show that

$$\lim_{q\to\infty} \lim_{p\to\infty} E(T_i) \to 0. \tag{7.53}$$

For example, note that

$$\begin{aligned}
\lim_{p\to\infty} E(T_1) &= \Big(\sum_{j,j',k,k'=q+1}^{\infty} \lim_{p\to\infty} \frac{1}{p} \mathrm{Tr}(\psi_j \psi_{j'}^* \psi_k \psi_{k'}^*) \Big) + \Big(\sum_{j,j'=q+1}^{\infty} \lim_{p\to\infty} \frac{1}{p} \mathrm{Tr}(\psi_j \psi_{j'}^*) \Big)^2 \\
&\le 2\big(\sum_{j=q+1}^{\infty} \sup_p ||\psi_j||_2 \big)^4 \to 0, \ (\text{as } q \to \infty)
\end{aligned} \tag{7.54}$$

by (7.54) and (B4). Similar arguments work for $2 \le i \le 9$.

Therefore, by Borel Cantelli Lemma, (7.52) and (7.53), $T_i \to 0$ almost surely and by (7.51), proof of (7.45) when $g_q = \hat{\Gamma}_{0q}$ and $g_\infty = \hat{\Gamma}_{0\infty}$, is complete. Using similar arguments as above, it is easy to prove for all $u \geq 0$ and $k \geq 1$,

$$\lim_{q\to\infty}\lim_{p\to\infty}\frac{1}{p}\mathrm{Tr}((\hat{\Gamma}_{uq}-\hat{\Gamma}_{u\infty})(\hat{\Gamma}_{uq}-\hat{\Gamma}_{u\infty})^*)^k \overset{\text{a.s.}}{=} 0. \tag{7.55}$$

Now we prove (7.45) when g_q and g_∞ are monomials. Without loss of generality, suppose for some $k \geq 1$,

$$g_q = \prod_{i=0}^{k}\hat{\Gamma}_{u_i q},\ \ g_\infty = \prod_{i=0}^{k}\hat{\Gamma}_{u_i\infty}. \tag{7.56}$$

Then by Lemma 7.3.6, we have

$$\begin{aligned}(g_q-g_\infty)(g_q-g_\infty)^* \;=\; & \sum_{j,j'=1}^{k}\Bigg[\left(\prod_{i=1}^{j-1}\hat{\Gamma}_{u_i q}\right)(\hat{\Gamma}_{u_j q}-\hat{\Gamma}_{u_j\infty})\left(\prod_{i=j+1}^{k}\hat{\Gamma}_{u_i\infty}\right)\\ & \left(\prod_{i=1}^{j'-1}\hat{\Gamma}_{u_i q}\right)(\hat{\Gamma}_{u'_j q}-\hat{\Gamma}_{u'_j\infty})\left(\prod_{i=j'+1}^{k}\hat{\Gamma}_{u_i\infty}\right)\Bigg].\end{aligned}$$

Now, by Lemma 4.3.1 (b) and noting (7.55), to establish (7.45), it is enough to prove for all $u \geq 0$, $k \geq 1$,

$$\lim_{q\to\infty}\lim_{p\to\infty}\frac{1}{p}\mathrm{Tr}(\hat{\Gamma}_{uq}\hat{\Gamma}_{uq}^*)^k < \infty \text{ a.s. and} \tag{7.57}$$

$$\lim_{q\to\infty}\lim_{p\to\infty}\frac{1}{p}\mathrm{Tr}(\hat{\Gamma}_{u\infty}\hat{\Gamma}_{u\infty}^*)^k < \infty \text{ a.s..} \tag{7.58}$$

Using the same arguments as in the proof of (7.28), it can be proved that

$$E\left(\frac{1}{p}\mathrm{Tr}(\hat{\Gamma}_{uq}\hat{\Gamma}_{uq}^*)^k-\frac{1}{p}E\mathrm{Tr}(\hat{\Gamma}_{uq}\hat{\Gamma}_{uq}^*)^k\right)^4 = O(p^{-4}) \text{ and} \tag{7.59}$$

$$E\left(\frac{1}{p}\mathrm{Tr}(\hat{\Gamma}_{u\infty}\hat{\Gamma}_{u\infty}^*)^k-\frac{1}{p}E\mathrm{Tr}(\hat{\Gamma}_{u\infty}\hat{\Gamma}_{u\infty}^*)^k\right)^4 = O(p^{-4}). \tag{7.60}$$

To show (7.60), we additionally need Assumption (B4). We omit the details.

Hence, using (7.59), (7.60) and Borel Cantelli Lemma, we have for all $u \geq 0$ and

$k \geq 1$,

$$\lim_{p\to\infty} (\frac{1}{p}\mathrm{Tr}(\hat{\Gamma}_{uq}\hat{\Gamma}^*_{uq})^k - \frac{1}{p}E\mathrm{Tr}(\hat{\Gamma}_{uq}\hat{\Gamma}^*_{uq})^k) = 0, \text{ a.s. and} \tag{7.61}$$

$$\lim_{p\to\infty} (\frac{1}{p}\mathrm{Tr}(\hat{\Gamma}_{u\infty}\hat{\Gamma}^*_{u\infty})^k - \frac{1}{p}E\mathrm{Tr}(\hat{\Gamma}_{u\infty}\hat{\Gamma}^*_{u\infty})^k) = 0, \text{ a.s..} \tag{7.62}$$

Again by (7.22) and (7.36), we have for all $u \geq 0$ and $k \geq 1$,

$$\lim_{q\to\infty}\lim_{p\to\infty} \frac{1}{p}E\mathrm{Tr}(\hat{\Gamma}_{uq}\hat{\Gamma}^*_{uq})^k < \infty \text{ and} \tag{7.63}$$

$$\lim_{q\to\infty}\lim_{p\to\infty} \frac{1}{p}E\mathrm{Tr}(\hat{\Gamma}_{u\infty}\hat{\Gamma}^*_{u\infty})^k < \infty. \tag{7.64}$$

Therefore, by (7.61) (7.62), (7.63) and (7.64), (7.57) and (7.58) hold.

Hence, (7.45) is proved for monomials.

Similar arguments work for polynomial also. For example, if $g_q = \pi_{1q} + \pi_{2q}$ and $g_q = \pi_{1\infty} + \pi_{2\infty}$ where π_{1q}, π_{2q}, $\pi_{1\infty}$ and $\pi_{2\infty}$ are monomials. Then

$$\begin{aligned}
& |p^{-1}\mathrm{Tr}(g_q - g_\infty)(g_q - g_\infty)^*| \\
\leq\; & p^{-1}\mathrm{Tr}(\pi_{1q} - \pi_{1\infty})(\pi_{1q} - \pi_{1\infty})^* + p^{-1}\mathrm{Tr}(\pi_{2q} - \pi_{2\infty})(\pi_{2q} - \pi_{2\infty})^* \\
& +2\sqrt{p^{-1}\mathrm{Tr}(\pi_{1q} - \pi_{1\infty})(\pi_{1q} - \pi_{1\infty})^* p^{-1}\mathrm{Tr}(\pi_{2q} - \pi_{2\infty})(\pi_{2q} - \pi_{2\infty})^*} \\
\to\; & 0, \quad \text{as (7.45) holds for monomials}
\end{aligned} \tag{7.65}$$

Therefore, (7.45) is proved for any polynomial. This completes the proof of Lemma 7.3.7. □

Proof of Theorem 7.3.4. Recall $F_{p,q}$ and $F_{p,\infty}$ respectively defined in (7.43) and (7.44). Also recall F_q and F_∞ in Lemma 7.3.5. Let $L(F, G)$ be the Levy distance between distribution functions F and G.

To prove this theorem, our goal is to show $\lim_{p\to\infty} L(F_{p,\infty}, F_\infty) = 0$, almost surely.

All the inequalities, equalities and limits below are in almost sure sense.

Note that

$$L(F_{p,\infty}, F_\infty) \leq L(F_{p,\infty}, F_{p,q}) + L(F_{p,q}, F_q) + L(F_q, F_\infty). \tag{7.66}$$

Now, taking limit as $p \to \infty$ on both sides of (7.66), by Theorem 7.3.1, we have

$$\lim_{p\to\infty} L(F_{p,\infty}, F_\infty) \leq \lim_{p\to\infty} L(F_{p,\infty}, F_{p,q}) + L(F_q, F_\infty),\ \forall q \geq 0. \tag{7.67}$$

Finally taking limit as $q \to \infty$ on both sides of (7.67), by Lemmas 7.3.5 and 7.3.7,

$$\lim_{p\to\infty} L(F_{p,\infty}, F_\infty) \leq \lim_{q\to\infty}\lim_{p\to\infty} L(F_{p,\infty}, F_{p,q}) + \lim_{q\to\infty} L(F_q, F_\infty) = 0.$$

Hence the proof of Theorem 7.3.4 is complete. □

7.3.2 Consequences of Theorems 7.3.1 and 7.3.4, $y \neq 0$

We now list some consequences of Theorems 7.3.1 and 7.3.4. In particular we also show how the existing results follow from these theorems under significantly weaker conditions.

The following corollary describes the LSD of $\{\hat{\Gamma}_u + \hat{\Gamma}_u^*\}$ and implies Theorem 7.2.4 (Liu et al. [2015]). Recall the class $U(\delta)$ of independent random variables defined in (7.2).

Corollary 7.3.8. (a) *Consider the model (3.7) (or (7.21)). Suppose (B1), (B3) and (B) hold and* $n, p(n) \to \infty$, $p/n \to y > 0$. *Then the LSD of* $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ *exists almost surely and its Stieltjes transformation is given by (5.76)-(5.79).*

(b) *Under the additional assumption (B4), the above result in (a) holds for the model (3.2) (or (7.3)) once we replace* q *by* ∞.

(c) *The above results in (a) and (b) hold if instead of* $(B3)$ *we assume* $(B2)$ *or* $\{\varepsilon_{i,j} : i, j \geq 1\} \in U(\delta)$ *for some* $\delta \in (0, 2]$.

(d) *Consider the model (3.7) (or (7.21)). Then the almost sure LSD of* $\frac{1}{2}(\hat{\Gamma}_u+\hat{\Gamma}_u^*)$ *are identical whenever* $u > q$ *and are different for* $u \leq q$.

Proof. (a) Existence of the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ is immediate from Theorem 7.3.1. Recall the matrices $\{\Delta_u\}$ in (7.23). By Theorems 5.4.1 and 7.3.1, LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ and $\frac{1}{2}(\Delta_u + \Delta_u^*)$ are identical and therefore their limiting Stieltjes transformations are same. By Corollary 5.4.7, the Stieltjes transformation of the LSD of $\frac{1}{2}(\Delta_u+\Delta_u^*)$ is given by (5.76)-(5.79). Hence, the same is true for $\frac{1}{2}(\hat{\Gamma}_u+\hat{\Gamma}_u^*)$ and the proof of (a) is complete.

(b) This is immediate from Theorem 7.3.4.

(c) This proof needs appropriate truncation on the support of $\{\varepsilon_{i,j}\}$ and is very technical. We defer the proof to Section 7.5.

(d) By Theorem 7.3.1, the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ depends on u only through the distribution of $\{\underline{c}_{j_1-j_2+u}, \underline{c}_{j_2-j_1-u} : 0 \leq j_1, j_2 \leq q\}$. Now by (7.19), note that the distribution of $\{\underline{c}_{j_1-j_2+u}, \underline{c}_{j_2-j_1-u} : 0 \leq j_1, j_2 \leq q\}$ are identical for $u > q$. Therefore, the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ are identical for $u > q$. Moreover, by Theorem 7.3.1 and Lemma 4.3.6 (a), for $u \leq q$ we have

$$\begin{aligned}
\lim \frac{1}{p} E\mathrm{Tr}(\hat{\Gamma}_u + \hat{\Gamma}_u^*) &= y^{-1}(1+y)\bar{\varphi}(\bar{\gamma}_{uq} + \bar{\gamma}_{uq}*) \\
&= y^{-1}(1+y)^2\bar{\varphi}\left(\sum_{j=0}^{q-u} \bar{\eta}_j\bar{\eta}_{j+u}^* + \sum_{j=0}^{q-u} \bar{\eta}_j^*\bar{\eta}_{j+u}\right). \quad (7.68)
\end{aligned}$$

These vary as u varies. Hence LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ are different for $0 \leq u \leq q$. □

Remark 7.3.9. *Under Assumption (WAP), the Stieltjes transform equations given in (5.76)-(5.79) (with* $q = \infty$*) reduce to equations (7.6)-(7.8). This is easy to see once we observe that under (WAP),* $\varphi_{\mathrm{odd}}(\cdot)$ *in (5.76)-(5.79) reduces to* $\int \cdot\, dF$. *Thus Corollary 7.3.8 in conjunction with Corollary 5.4.7 implies Theorem 7.2.4 (Liu et al. [2015]).*

The following corollary describes the LSD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ when $\psi_j = \lambda_j I_p$, $\forall j$ and implies Theorem 7.2.3 (Pfaffel and Schlemm [2011]).

Corollary 7.3.10. *(a) Consider the model (3.7) (or (7.21)). Suppose (B1) and (B3) hold and* $n, p(n) \to \infty$, $p/n \to y > 0$. *Let* $\psi_j = \lambda_j I_p$, $\forall j$ *and* I_p *is as in (2.8). Then for each* $u \geq 1$, *the almost sure LSD of* $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ *exists and its limiting Stieltjes transformation satisfies (only one solution yields a valid Stieltjes transform)*

$$\begin{aligned} z &= -\frac{1}{m_u(z)} + \frac{1}{2\pi}\int_0^{2\pi} \frac{d\theta}{y m_u(z) + f^{-1}(\theta)}, \quad z \in \mathbb{C}^+ \text{ where} && (7.69) \\ f(\theta) &= \cos(u\theta)|\sum_{k=0}^{q} \lambda_k e^{ik\theta}|^2. && (7.70) \end{aligned}$$

(b) Under the additional assumption $\sum_{j=0}^{\infty} |\lambda_j| < \infty$, *the result in (a) holds for the model (3.2) (or (7.3)) once we replace* q *by* ∞.

(c) The results in (a) and (b) continue to hold if instead of (B3) we assume $\{\varepsilon_{i,j} : i, j \geq 1\} \in U(\delta)$ *for some* $\delta > 0$.

Proof. (a) Recall the definition of free cumulant κ in (4.56). Note that by Theorems 5.4.1 and 7.3.1, the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ and $\frac{1}{2}(\Delta_u + \Delta_u^*)$ are identical and therefore, their free cumulants are also same. By Corollary 5.4.8, free cumulants of the LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ are given by

$$\kappa_{ur} = \frac{1}{2y\pi}\int_0^{2\pi} (yf(\theta))^r d\theta, \quad \forall r \geq 1. \tag{7.71}$$

Recall the free cumulant generating function $C(\cdot)$ in (4.60). Note that $|\kappa_{ur}| \leq C^r$, $\forall r \geq 1$ and some $C > 0$. Therefore, for $z \in \mathbb{C}^+$ and $|z|$ small,

$$C(z) = 1 + \sum_{r=1}^{\infty} \kappa_{ur} z^r = 1 + \frac{1}{2y\pi}\int_0^{2\pi} \frac{yzf(\theta)d\theta}{1 - yzf(\theta)}. \tag{7.72}$$

Note that $m_u(z) \to 0$ as $|z| \to \infty$. Hence, by (4.61), for some $K > 0$ and all

$|z| > K$, $z \in \mathbb{C}^+$

$$-zm_u(z) = C(-m_u(z)) = 1 - \frac{1}{2y\pi} \int_0^{2\pi} \frac{ym_u(z)f(\theta)d\theta}{1 + ym_u(z)f(\theta)}. \tag{7.73}$$

Upon simplifying, (7.73) reduces to (7.69) for large $|z|$. Using analyticity, (7.69) continues to hold for $z \in \mathbb{C}^+$. We omit the details. Hence (a) is proved.

(b) follows from Corollary 7.3.8 (b) as under $\psi_j = \lambda_j I_p$, $\forall j$, (B4) reduces to $\sum_{j=0}^{\infty} |\lambda_j| < \infty$. (c) is immediate from Corollary 7.3.8 (c). □

Remark 7.3.11. *Corollary 7.3.10 (c) implies Theorem 7.2.3 (Pfaffel and Schlemm [2011]) as (B2) imples $\{\varepsilon_{i,j} : i, j \geq 1\} \in U(\delta)$ for some $\delta \in (0, 2]$.*

The following corollary describes the LSD of $\hat{\Gamma}_u \hat{\Gamma}_u^*$. This result appears to be completely new in the literature. Recall the classes $U(\delta)$ and $\mathcal{L}_4$ of independent random variables respectively in (7.2) and (7.14).

Corollary 7.3.12. (*a*) *Consider the model (3.7) (or (7.21)). Suppose (B1), (B3) and (B) hold and $n, p(n) \to \infty$, $p/n \to y > 0$. Then the LSD of $\hat{\Gamma}_u \hat{\Gamma}_u^*$ exists almost surely.*

(*b*) *Under the additional assumption (B4), the above result in (a) holds for the model (3.2) (or (7.3)) once we replace q by ∞.*

(*c*) *The above results in (a) and (b) hold if instead of* (*B*3) *we assume* (*B*2) *or $\{\varepsilon_{i,j} : i, j \geq 1\} \in \mathcal{L}_4 \cap U(\delta)$ for some $\delta \in (0, 2]$.*

(*d*) *The almost sure LSD of $\hat{\Gamma}_u \hat{\Gamma}_u^*$ (in (a)) are identical whenever $u > q$ and are different for $u \leq q$.*

Proof. (a) and (b) follow immediately from Theorems 7.3.1 and 7.3.4. To prove (c) we need appropriate truncation on $\{\varepsilon_{i,j}\}$ and the arguments are very technical. We provide the details in Section 7.6. The first part of (d) i.e. 'the LSD of $\hat{\Gamma}_u \hat{\Gamma}_u^*$ are identical whenever $u > q$' is true for the same reason as given in the proof of Corollary 7.3.8 (d). For the second part of (d), by Theorem 7.3.1 and Lemma 4.3.6 (a), for all $u \leq q$ note that

$$\begin{aligned} \lim \frac{1}{p} E\mathrm{Tr}(\hat{\Gamma}_u\hat{\Gamma}_u^*) = \frac{1+y}{y}\bar{\varphi}(\bar{\gamma}_{uq}\bar{\gamma}_{uq}^*) &= y^{-1}(1+y)^3 \sum_{i,j=0}^{q-u} \bar{\varphi}(\bar{\eta}_i\bar{\eta}_{i+u}^*\bar{\eta}_j\bar{\eta}_{j+u}^*) \\ &+ y^{-1}(1+y)^3 \sum_{\substack{0\leq i,j,j'\leq u \\ 0\leq i+j'-j\leq q}} \bar{\varphi}(\bar{\eta}_j\bar{\eta}_{i+j'-j}^*)\bar{\varphi}(\bar{\eta}_{j'}\bar{\eta}_i^*). \end{aligned}$$

These expressions are different for different values of u. Therefore, the LSD of $\hat{\Gamma}_u\hat{\Gamma}_u^*$ are different for $0 \leq u \leq q$. Hence, (d) is proved. This completes the proof of Corollary 7.3.12. $\square$

The following corollary describes the LSD of $\hat{\Gamma}_0$, $\{\hat{\Gamma}_u + \hat{\Gamma}_u^*\}_{u\geq 1}$ and $\{\hat{\Gamma}_u\hat{\Gamma}_u^*\}_{u\geq 1}$ for the MA(0) process and implies Theorems 7.2.1 (Bai and Silverstein [2009]) and 7.2.2 (Jin et al. [2014]). Recall the Marčenko-Pastur law MP_y with parameter $y > 0$ satisfying the moment sequence (4.25).

Corollary 7.3.13. *Consider the model (3.5) (or (7.1)). Suppose (B1) holds and $n, p(n) \to \infty$, $p/n \to y > 0$. Then the following statements are true.*

(a) Suppose (B3) is satisfied or $\{\varepsilon_{i,j} : i,j \geq 1\} \in U(0)$. Then the almost sure LSD of $\hat{\Gamma}_0$ is the MP_y law.

(b) Suppose (B3) is satisfied or $\{\varepsilon_{i,j} : i,j \geq 1\} \in U(\delta)$ for some $\delta \in (0,2]$. Then for each $u \geq 1$, the almost sure LSD of $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ exists and its limiting Stieltjes transformation is given by (5.90).

(c) Suppose (B3) is satisfied or $\{\varepsilon_{i,j} : i,j \geq 1\} \in \mathcal{L}_4 \cap U(\delta)$ for some $\delta \in (0,2]$. Then for each $u \geq 1$, the almost sure LSD of $\hat{\Gamma}_u\hat{\Gamma}_u^*$ exists and its moment sequence is given by (5.106).

Proof. We prove all the above results under (B3). To prove the results under the corresponding alternative assumptions, we need appropriate truncation on $\{\varepsilon_{i,j}\}$ and follow the same arguments as in the proof of Corollary 7.3.8 (c) given later in Section 7.5. We shall not provide the details of the truncation arguments.

Recall the matrix Z in (4.2.3). Now note that, under (B1) and (B3), by Theorems 5.4.1 and 7.3.1, LSD of (a) $n^{-1}ZZ^*$ and $\hat{\Gamma}_0$, (b) $\frac{1}{2n}Z(P_u + P_{-u})Z^*$ and $\frac{1}{2}(\hat{\Gamma}_u + \hat{\Gamma}_u^*)$ and (c) $p^{-2}ZP_uZ^*ZP_{-u}Z^*$ and $\hat{\Gamma}_u\hat{\Gamma}_u^*$ are identical. Therefore, by Corollaries 5.4.9 , 5.4.11 and 5.4.12, Corollary 7.3.13 follows. □

Remark 7.3.14. *Corollary 7.3.13 (a) and (b) respectively imply Theorems 7.2.1 (Bai and Silverstein [2009]) and 7.2.2 (Jin et al. [2014]).*

7.3.3 LSD for $p/n \to 0$

Now we shall discuss the LSD of any symmetric polynomial in $\{\hat{\Gamma}_u\}$ for the case $p, n(p) \to \infty$ such that $p/n \to 0$. Consider the collection of non-commutative variables $\{w_{u,j_1,j_2}\}$, whose free cumulants are as follows:

For all $j_1, j_2, \ldots \geq 1$ and $u_1, u_2, \ldots = 0, \pm 1, \pm 2, \ldots$,

$$
\begin{aligned}
w^*_{u,j_1,j_2} &= w_{-u,j_2,j_1}, &(7.74)\\
\kappa_2(w_{u_1,j_1,j_2}, w_{u_2,j_3,j_4}) &= \frac{1}{2\pi}\int_0^{2\pi} e^{i(j_1-j_2+u_1)\theta} e^{i(j_3-j_4+u_2)\theta}\, d\theta \\
&= \begin{cases} 1, \text{ if } j_1 - j_2 + j_3 - j_4 = -(u_1+u_2) \\ 0, \quad \text{otherwise}, \end{cases} &(7.75)\\
\kappa_r(w_{u_i,j_{2i-1},j_{2i}} : 1 \leq i \leq r) &= 0, \ \forall r \neq 2. &(7.76)
\end{aligned}
$$

As mentioned in Section 6.3, the above sequence of free cumulants naturally define a state, say φ_w, on $\text{Span}\{w_{u,l,i} : u, l, i \geq 1\}$. This is because, the moments and free cumulants are in one-to-one correspondence (see (4.56) and (4.58) in Chapter 4).

Recall $(\mathcal{A}_{\text{odd}}, \varphi_{\text{odd}})$ and $\{\eta_j, \eta_j^*\}$ respectively in (5.13) and (7.17). Also recall the free product $(\mathcal{B}, \varphi_0)$ of $(\mathcal{A}_{\text{odd}}, \varphi_{\text{odd}})$ and $(\text{Span}\{w_{u,l,i} : u, l, i \geq 1\}, \varphi_w)$ in (6.41). Therefore, $\{w_{u,j_1,j_2}\}$ and $\{\eta_j, \eta_j^*\}$ are free in $(\mathcal{B}, \varphi_0)$.

Consider the following polynomial in $\mathcal{B}$

$$S_{uq} = \sum_{j_1,j_2=0}^{q} \eta_{j_1} w_{u,j_1,j_2} \eta_{j_2}^*, \quad \forall u, q \geq 0. \tag{7.77}$$

Recall the population autocovariance matrices $\{\Gamma_u\}$ defined in (3.1):

$$\Gamma_u = \sum_{j=0}^{q-u} \psi_j \psi_{j+u}^*, \quad \forall u \geq 0.$$

Note that, by (7.17),

$$(\text{Span}\{\Gamma_u, \Gamma_u^* : u \geq 0\}, p^{-1}\text{Tr}) \to (\text{Span}\{G_{uq}, G_{uq}^* : u \geq 0\}, \varphi_\eta), \tag{7.78}$$

where

$$G_{uq} = \sum_{j=0}^{q-u} \eta_j \eta_{j+u}^*, \ \forall u, q \geq 0. \tag{7.79}$$

Now we shall state our LSD result for general self-adjoint polynomials $\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)$. To describe the limit write $\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)$ in the form

$$\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0) = \sum_{l=1}^{T} \beta_l \Big(\prod_{i=1}^{k_l} \hat{\Gamma}_{u_{l,i}}^{\epsilon_{l,i}} \Big), \tag{7.80}$$

where $\epsilon_{l,i} \in \{1, *\}$ and $u_{l,i} \in \{0, 1, 2, \ldots\}$.

Then we have the following Theorem (see Bhattacharjee and Bose [2015b]).

Theorem 7.3.15. *Consider the model (3.7) (or (7.21)). Suppose (B1), (B3) and (B) hold and* $p, n(p) \to \infty$*,* $p/n \to 0$*. Then the LSD of the self-adjoint polynomial*

$$\sqrt{np^{-1}} \big(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0) - \Pi(\Gamma_u, \Gamma_u^* : u \geq 0) \big) \tag{7.81}$$

exists almost surely and is distributed as

$$\sum_{l=1}^{T} \beta_l \Big(\sum_{\substack{\nu_{l,i} \in \{0,1\} \\ \sum_i \nu_{l,i} = 1}} \prod_{i=1}^{k_l} G_{u_{l,i}q}^{\epsilon_{l,i}(1-\nu_{l,i})} S_{u_{l,i}q}^{\epsilon_{l,i}\nu_{l,i}} \Big). \tag{7.82}$$

where $\{S_{uq}\}$ *and* $\{G_{uq}\}$ *are as in (7.77) and (7.79).*

Proof. As in the proof of Theorem 7.3.1, to prove the above theorem, we use the moment method and Lemma 4.2.1 as described in Section 4.2. Recall $\{\Delta_u\}$ given in (5.1). The following lemma, describes the approximation of $\{\hat{\Gamma}_u\}$ by $\{\Delta_u\}$ and is an analogue of Lemma 7.3.2 given earlier for the $p/n \to y > 0$ case. Proof of this lemma follows the same arguments as the proof of Lemma 7.3.2 and therefore we omit it.

Lemma 7.3.16. *Consider the model (7.21). Suppose (B1), (B3) and (B) hold and* $p, n(p) \to \infty$, $p/n \to 0$. *Then the following statements are true.*

(a) For any polynomial Π,

$$\begin{aligned} & \lim p^{-1} E\mathrm{Tr}\left(\sqrt{np^{-1}}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0) - \Pi(\Gamma_u, \Gamma_u^* : u \geq 0))\right) \\ = \; & \lim p^{-1} E\mathrm{Tr}\left(\sqrt{np^{-1}}(\Pi(\Delta_u, \Delta_u^* : u \geq 0) - \Pi(\Gamma_u, \Gamma_u^* : u \geq 0))\right). \end{aligned} \tag{7.83}$$

(b) Let, for $1 \leq i \leq T$, m_i *be polynomials. Let for all* $1 \leq i \leq T$,

$$\begin{aligned} \mathcal{P}_i &= \mathrm{Tr}\left(\sqrt{np^{-1}}(m_i(\Delta_u, \Delta_u^* : u \geq 0) - m_i(\Gamma_u, \Gamma_u^* : u \geq 0))\right), \quad \mathcal{P}_i^0 = E\mathcal{P}_i, \\ \tilde{\mathcal{P}}_i &= \mathrm{Tr}\left(\sqrt{np^{-1}}(m_i(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0) - m_i(\Gamma_u, \Gamma_u^* : u \geq 0))\right), \quad \tilde{\mathcal{P}}_i^0 = E\tilde{\mathcal{P}}_i. \end{aligned}$$

Then we have

$$\lim E\Big(\prod_{i=1}^{T} (\tilde{\mathcal{P}}_i - \tilde{\mathcal{P}}_i^0)\Big) = \lim E\Big(\prod_{i=1}^{T} (\mathcal{P}_i - \mathcal{P}_i^0)\Big). \tag{7.84}$$

Now it is easy to see that, exactly like the proof of Theorem 7.3.1, here (M1), (M4) and (C) hold by application of Theorems 6.3.1, 6.4.1 and Lemma 7.3.16. This completes the proof of Theorem 7.3.15. □

We now consider the case $q = \infty$ (see Bhattacharjee and Bose [2015b]). Proof of Theorem 7.3.17 follows the same arguments as in the proof of Theorem 7.3.4 and hence we omit the proof.

Theorem 7.3.17. *Consider the model (3.2) (or (7.3)). Suppose (B1), (B3), (B) and (B4) hold and $p, n(p) \to \infty$, $p/n \to 0$. Let $\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)$ be as in (7.80). Then LSD of the self-adjoint polynomial*

$$\sqrt{np^{-1}}\big(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0) - \Pi(\Gamma_u, \Gamma_u^* : u \geq 0)\big) \tag{7.85}$$

exists almost surely and is distributed as

$$\sum_{l=1}^{T} \beta_l \Bigg(\sum_{\substack{\nu_{l,i} \in \{0,1\} \\ \sum_i \nu_{l,i} = 1}} \prod_{i=1}^{k_l} G_{u_{l,i}\infty}^{\epsilon_{l,i}(1-\nu_{l,i})} S_{u_{l,i}\infty}^{\epsilon_{l,i}\nu_{l,i}} \Bigg). \tag{7.86}$$

7.3.4 Consequence of Theorems 7.3.15 and 7.3.17, $y = 0$

We now list some consequences of Theorems 7.3.15 and 7.3.17. In particular we also show how the existing results follow from these theorems under significantly weaker conditions. The last corollary is apparently new in the literature.

The following corollary describes the LSD of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$ and implies Theorem 7.2.5 (Wang et al. [2015]). Recall the class $U(\delta)$ of independent random variables given in (7.2).

Corollary 7.3.18. (a) *Consider the model (7.21). Suppose (B1), (B3) and (B) hold and $p, n(p) \to \infty$, $p/n \to 0$. Then the LSD g_{uq} of* $\sqrt{np^{-1}}\big(2^{-1}(\hat{\Gamma}_u + \hat{\Gamma}_u^*) -$

$2^{-1}(\Gamma_u + \Gamma_u^*)\big)$ *exists almost surely and*

$$g_{uq} \overset{\mathcal{D}}{=} \frac{1}{2}\Big(\sum_{j_1,j_2=0}^{q} \eta_{j_1} w_{u,j_1,j_2} \eta_{j_2}^* + \sum_{j_1,j_2=0}^{q} \eta_{j_2} w_{-u,j_2,j_1} \eta_{j_1}^* \Big). \tag{7.87}$$

The Stieltjes transformation of g_{uq} *is given by (6.58)-(6.55).*

(*b*) *Under the additional assumption (B4), the above result in (a) holds for the model (3.2) once we replace* q *by* ∞.

(*c*) *The above results in (a) and (b) hold if instead of* $(B3)$ *we assume* $(B2)$ *or* $\{\varepsilon_{i,j} : i, j \geq 1\} \in U(\delta)$ *for some* $\delta \in (0, 2]$.

(*d*) *Consider the model (3.7). Then the almost sure LSD of* $\sqrt{np^{-1}}\big(2^{-1}(\hat{\Gamma}_u + \hat{\Gamma}_u^*) - 2^{-1}(\Gamma_u + \Gamma_u^*)\big)$ *are identical whenever* $u > q$ *and are different for* $u \leq q$.

Proof. (a) Existence of the LSD of $\sqrt{np^{-1}}\big(2^{-1}(\hat{\Gamma}_u+\hat{\Gamma}_u^*)-2^{-1}(\Gamma_u+\Gamma_u^*)\big)$ and (7.87) are immediate from Theorem 7.3.15 once we put $T = 2, k_1 = k_2 = \beta_1 = \beta_2 = 1$, $u_{1,1} = u_{2,1} = u$, $\epsilon_{1,1} = 1, \epsilon_{2,1} = *$. Recall the matrices $\{\Delta_u\}$ in (5.1). By Theorems 6.4.1 and 7.3.15, LSD of $\sqrt{np^{-1}}\big(2^{-1}(\hat{\Gamma}_u + \hat{\Gamma}_u^*) - 2^{-1}(\Gamma_u + \Gamma_u^*)\big)$ and $\sqrt{np^{-1}}\big(2^{-1}(\Delta_u + \Delta_u^*) - 2^{-1}(\Gamma_u + \Gamma_u^*)\big)$ are identical and therefore their limiting Stieltjes transformations are same. By Corollary 6.4.6, the Stieltjes transformation of the LSD of $\sqrt{np^{-1}}\big(2^{-1}(\Delta_u + \Delta_u^*) - 2^{-1}(\Gamma_u + \Gamma_u^*)\big)$ is given by (6.58)-(6.55). Hence, the same is true for $\sqrt{np^{-1}}\big(2^{-1}(\hat{\Gamma}_u + \hat{\Gamma}_u^*) - 2^{-1}(\Gamma_u + \Gamma_u^*)\big)$ and the proof of (a) is complete.

(b) This is immediate from Theorem 7.3.17.

(c) This proof needs appropriate truncation on the support of $\{\varepsilon_{i,j}\}$ and is very technical. We provide the truncation arguments in Section 7.7.

(d) Note that for each fixed u, g_{uq} depends on the distribution of $\{w_{u,j_1,j_2}, w_{-u,j_2,j_1} : 1 \leq j_1, j_2 \leq q\}$ only. By (7.75), the distribution of $\{w_{u,j_1,j_2}, w_{-u,j_2,j_1} : 1 \leq j_1, j_2 \leq$

$q\}$ is characterized by the second order cumulants,

$$\kappa_2(w_{v,j_1,j_2}, w_{w,k_1,k_2}) = \begin{cases} 1, & \text{if } j_1 - j_2 + k_1 - k_2 = \pm 2u \text{ or } 0;\ v, w = \pm u, \\ 0, & \text{otherwise.} \end{cases}$$

Now note that for an MA(q) process, $-2q \leq j_1 - j_2 + k_1 - k_2 \leq 2q$. Hence, for $u > q$, $j_1 - j_2 + k_1 - k_2 = \pm 2u$ can never happen. Therefore, for $u > q$, the distribution of $\{w_{u,j_1,j_2}, w_{-u,j_2,j_1} : 1 \leq j_1, j_2 \leq q\}$ does not depends on u and hence $g_{uq},\ u \geq q$, are identically distributed. Now, note that for $u \leq q$

$$\begin{aligned} \varphi_0(g_{uq}^2) &= \sum_{\substack{0 \leq j_1, j_2, k_1 \leq q \\ 0 \leq j_1 + k_1 - j_2 - 2u \leq q}} \varphi_0(\eta_{j_2}^* \eta_{k_1}) \varphi_0(\eta_{j_1} \eta_{j_1+k_1-j_2-2u}^*) \\ &\quad + \sum_{\substack{0 \leq j_1, j_2, k_1 \leq q \\ 0 \leq j_1 + k_1 - j_2 + 2u \leq q}} \varphi_0(\eta_{j_2}^* \eta_{k_1}) \varphi_0(\eta_{j_1} \eta_{j_1+k_1-j_2+2u}^*) \\ &\quad + \sum_{\substack{0 \leq j_1, j_2, k_1 \leq q \\ 0 \leq j_1 + k_1 - j_2 \leq q}} \varphi_0(\eta_{j_2}^* \eta_{k_1}) \varphi_0(\eta_{j_1} \eta_{j_1+k_1-j_2}^*). \end{aligned}$$

These are different as u varies. Therefore, the distribution of g_{uq} are different for $0 \leq u \leq q$. Hence, (d) is proved. □

Remark 7.3.19. *Using Corollary 7.3.18 (b), it can be shown that under Assumption (WAP), $\varphi_0(\cdot)$ in (6.58)-(6.55) reduces to $\int \cdot\, dF$. Hence, the Stieltjes transform equations given in (6.58)-(6.55) (with $q = \infty$) reduce to equations (7.8)-(7.11). Thus Corollary 7.3.18 in conjunction with Corollary 6.4.6 implies Theorem 7.2.5 (Wang et al. [2015]).*

Thus we have shown that the existing LSD results in the literature (Bai and Silverstein [2009], Jin et al. [2014], Pfaffel and Schlemm [2011], Liu et al. [2015] and Wang et al. [2015]) follow as special cases of our results.

The following result is apparently new in the literature.

Corollary 7.3.20. *Suppose $X_t = \varepsilon_t + \lambda\varepsilon_{t-1}$, $\forall t$, $\lambda \in \mathbb{R}$. Suppose (B1) and (B3) hold. Then the almost sure LSD of $\sqrt{np^{-1}}(\hat{\Gamma}_1\hat{\Gamma}_1^* - \lambda^2 I_p)$ is distributed as $\sqrt{2}\lambda((1+\lambda^2)s_1 + \lambda s_2 + \sqrt{2}\lambda s_3)$ where s_1, s_2 and s_3 are free standard semicircle variables.*

Proof. Recall $\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)$ in (7.80). Note that $\hat{\Gamma}_u\hat{\Gamma}_u^* = \Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)$ with $T = 1, \beta_1 = 1, k_1 = 2, \epsilon_{1,1} = 1, \epsilon_{1,2} = *, u_{1,1} = 1, u_{1,2} = 1$. By (7.17), $\eta_0 = 1_{\mathcal{A}_{\text{odd}}}, \eta_1 = \lambda 1_{\mathcal{A}_{\text{odd}}}, \eta_j = 0\ \forall j > 1$. Therefore, by (7.77),

$$S_{11} = w_{1,0,0} + \lambda^2 w_{1,1,1} + \lambda w_{1,1,0} + \lambda w_{1,0,1}, \tag{7.88}$$

where by (7.74), (7.75) and (7.76),

$$\kappa_2(w_{1,0,0}, w_{1,0,0}^*) = \kappa_2(w_{1,1,1}, w_{1,1,1}^*) = \kappa_2(w_{1,1,0}, w_{1,1,0}^*) = \kappa_2(w_{1,0,1}, w_{1,0,1}^*) = 1, \tag{7.89}$$

$$\kappa_2(w_{1,0,0}^*, w_{1,1,1}) = \kappa_2(w_{1,0,0}, w_{1,1,1}^*) = \kappa_2(w_{1,1,0}, w_{1,1,0}) = \kappa_2(w_{1,0,1}, w_{1,0,1}) = 1, \tag{7.90}$$

and all other joint free cumulants of $(w_{1,0,0}, w_{1,1,1}, w_{1,1,0}, w_{1,0,1})$ are zero.

By (7.79)

$$G_{11} = G_{11}^* = \lambda 1_{\mathcal{A}_{\text{odd}}}. \tag{7.91}$$

Hence, by Theorem 7.3.15, the almost sure LSD of $\sqrt{np^{-1}}(\hat{\Gamma}_1\hat{\Gamma}_1^* - \lambda^2 I_p)$ is distributed as

$$\begin{aligned}
& S_{11}G_{11}^* + G_{11}S_{11}^* \quad \text{(see (7.82))} \\
= \;& \lambda(w_{1,0,0} + \lambda^2 w_{1,1,1} + \lambda w_{1,1,0} + \lambda w_{1,0,1}) + \lambda(w_{1,0,0} + \lambda^2 w_{1,1,1} + \lambda w_{1,1,0} + \lambda w_{1,0,1})^* \\
= \;& g_1 + g_1^*, \text{ say} \qquad (7.92)
\end{aligned}$$

where $g_1 = \lambda(w_{1,0,0} + \lambda^2 w_{1,1,1} + \lambda w_{1,1,0} + \lambda w_{1,0,1})$. Let w_1 and w_2 be circular elements with $\kappa_2(w_1, w_1^*) = \kappa_2(w_2, w_2^*) = 1$ and s_3 be a standard semicircle variable. Moreover, suppose w_1, w_2 and w_3 are free. Then by (7.89) and (7.90), g_1 has same distribution as $\lambda(1+\lambda^2)w_1 + \lambda^2 w_2 + \lambda^2 s_3$. Therefore, $g_1 + g_1^*$ has same distribution as $\lambda(1+\lambda^2)(w_1 + w_1^*) + \lambda^2(w_2 + w_2^*) + 2\lambda^2 s_3$. Now, note that $(w_1 + w_1^*)$ is a self-adjoint element with all marginal cumulants zero except $\kappa_2(w_1 + w_1^*, w_1 + w_1^*) = 2$. Therefore, by Definition 4.3.5, $(w_1 + w_1^*)$ is distributed as $\sqrt{2}s_1$, where s_1 is a standard semi-circle variable. Similarly, $(w_2 + w_2^*)$ is distributed as $\sqrt{2}s_2$, where s_2 is a standard semi-circle variable. Therefore, $g_1 + g_1^*$ has same distribution as $\sqrt{2}\lambda((1+\lambda^2)s_1 + \lambda s_2 + \sqrt{2}\lambda s_3)$ where s_1, s_2 and s_3 are free standard semicircle variables. This completes the proof of Corollary 7.3.20. □

7.4 Proof of Lemma 7.3.2

Here we shall only prove (a). We omit the proof of (b) as similar arguments will go through for (b) also.

Note that, when $X_t \sim MA(q)$ process, we have

$$X_t = (\psi_0 \quad \psi_1 \quad \psi_2 \ldots \psi_q)(\varepsilon_t^* \quad \varepsilon_{t-1}^* \quad \varepsilon_{t-2}^* \ldots \varepsilon_{t-q}^*)^* \quad \forall t, n \geq 1.$$

Therefore, by the definition of autocovariance matrices, the sample autocovariance matrix of order k is given by

$$\begin{aligned} n\hat{\Gamma}_k \quad = \quad & \sum_{j,j'=0}^{q} \psi_j \ \hat{\Gamma}_{j'-j+k}(\varepsilon) \ \psi_{j'}^* - \sum_{\substack{j,j'=0 \\ j-j'\neq k}}^{q} \ \sum_{t=n-j+1}^{n} \psi_j \varepsilon_{t,p} \varepsilon_{t-(j'+k-j)}^* \psi_{j'}^* \\ & + \sum_{\substack{j,j'=0 \\ j-j'\neq k}}^{q} \ \sum_{t=k-j+1}^{j'+k-j} \psi_j \varepsilon_t \varepsilon_{t-(j'+k-j)}^* \psi_{j'}^* + \sum_{j=0}^{q} \ \sum_{t=n-j+1}^{n} \psi_j \varepsilon_t \varepsilon_t^* \psi_{j-k.p}^* \\ & + \sum_{j=0}^{q} \ \sum_{t=k-j+1}^{0} \psi_j^{(n)} \varepsilon_t \varepsilon_t^* \psi_{j-k}^* \end{aligned}$$

$$= \quad n\Delta_k + R_{1n} + R_{2n} + R_{3n} + R_{4n}, \text{ (say)}. \tag{7.93}$$

Note that, for any k matrices $A_1, A_2, \ldots, A_k$ of order p and for some integers $r_1, p_1, r_2, p_2 \ldots, r_k, p_k \geq 1$, $\sum_{j=1}^{k} r_j^{-1} = 2$, $\sum_{j=1}^{k} p_j^{-1} = 1$, we have

$$\begin{aligned} \frac{1}{p} E\mathrm{Tr}(A_1A_2\ldots A_k) &\leq E\Big[\prod_{j=1}^{k}\left(\frac{1}{p}\mathrm{Tr}(A_j^*A_j)^{r_j}\right)^{\frac{1}{2r_j}}\Big] \\ &\leq \prod_{j=1}^{k}\left[E\left(\frac{1}{p}\mathrm{Tr}(A_j^*A_j)^{r_j}\right)^{\frac{p_j}{2r_j}}\right]^{1/p_j}. \end{aligned} \tag{7.94}$$

Moreover, for any polynomial Π, $(\Pi(\hat{\Gamma}_i, \hat{\Gamma}_i^* : i \geq 0) - \Pi(\Delta_i, \Delta_i^* : i \geq 0))$ involves monomials with at least one of R_{1n}, R_{2n}, R_{3n} or R_{4n}. Hence, to show (a), by (7.94), it suffices to show, for all $r, s \geq 1$ and $i = 1, 2, 3, 4$,

$$\lim E\left(p^{-1}\mathrm{Tr}\left(\Delta_u^*\Delta_u\right)^r\right)^s \quad < \quad \infty, \text{ and} \tag{7.95}$$

$$E\left(p^{-1}\mathrm{Tr}\left(n^{-2}R_{in}^*R_{in}\right)^r\right)^s \quad \to \quad 0. \tag{7.96}$$

Now (7.95) follows as

$$\begin{aligned} E\left(p^{-1}\mathrm{Tr}\left(\Delta_u^*\Delta_u\right)^r\right)^s &\leq E\left(p^{-1}\mathrm{Tr}\left(\Delta_u^*\Delta_u\right)^K\right)^{rs/K}, \text{ where } K > rs \\ &\leq \left(p^{-1}E\mathrm{Tr}\left(\Delta_u^*\Delta_u\right)^K\right)^{rs/K} \end{aligned}$$

and as $\lim p^{-1}E\mathrm{Tr}\left(\Delta_u^*\Delta_u\right)^K < \infty$, which we have already proved in the proof of Theorem 5.3.1.

Now we shall prove (7.96). Note that

$$|E\left(\mathrm{Tr}\left(n^{-2}R_{in}^*R_{in}\right)^r\right)^s| \leq E\left(\mathrm{Tr}\left(n^{-2}R_{in}^*R_{in}\right)\right)^{rs}. \tag{7.97}$$

Therefore, to prove (7.96), by (7.97), it is enough to show, for all $r \geq 1$, there is

$C_r > 0$, such that

$$|E\left(\mathrm{Tr}\left(n^{-2}R_{in}^{*}R_{in}\right)\right)^{r}| < C_r,\ \forall n \geq 1. \tag{7.98}$$

Let us first prove (7.98) for $i = 1$. Similar idea will work to prove (7.98) for $i > 1$.

Recall the definition of R_{1n} in (7.93). Note that to prove (7.98) for $i = 1$, it suffices to show, for $r \geq 1$, $s_k > 0$ and $\{A_k\} \in \mathrm{Span}\{\psi_j, \psi_j^{*} : j \geq 0\}$, there is $C_r > 0$ such that for all $n \geq 1$,

$$|E\prod_{k=1}^{r}\left(n^{-2}\mathrm{Tr}(A_{2k-1}\varepsilon_{t_{2k-1}}\varepsilon^{*}_{t_{2k-1}-s_{2k-1}}A_{2k}\varepsilon_{t_{2k}-s_{2k}}\varepsilon^{*}_{t_{2k}})\right)| < C_r. \tag{7.99}$$

To prove (7.99), we use induction method on r. For $r = 1$, by Assumption $(A3)$,

$$\begin{aligned} & |E\left(n^{-2}\mathrm{Tr}(A_1\varepsilon_{t_1}\varepsilon^{*}_{t_1-s_1}A_2\varepsilon_{t_2-s_2}\varepsilon^{*}_{t_2})\right)| \\ \leq\ & |\prod_{i=1}^{2}\frac{1}{n}\mathrm{Tr}(A_1)| < C_1,\ \forall n \geq 1 \text{ and for some } C_1 > 0. \end{aligned}$$

Suppose (7.99) is true for all $r \leq m$. Now for $r = m+1$, consider

$$E\prod_{k=1}^{m+1}\left(n^{-2}\mathrm{Tr}(A_{2k-1}\varepsilon_{t_{2k-1}}\varepsilon^{*}_{t_{2k-1}-s_{2k-1}}A_{2k}\varepsilon_{t_{2k}-s_{2k}}\varepsilon^{*}_{t_{2k}})\right). \tag{7.100}$$

As $\{\varepsilon_{i,j}\}$ are independent and of mean 0, $\{\varepsilon_t\}$ has to be matched for a non-zero contribution. Now the following two cases may happen.

Case 1. If matchings are such that no index in $\{t_{2k_u-1}, t_{2k_u-1}-s_{2k_u-1}, t_{2k_u}, t_{2k_u}-s_{2k_u} : 1 \leq u \leq U < m+1\}$ matches with any index in $\{t_{2k-1}, t_{2k-1}-s_{2k-1}, t_{2k}, t_{2k}-s_{2k} : k \neq k_u,\ \forall 1 \leq u \leq U < m+1\}$, then (7.100) would become

$$|\left(E\prod_{u=1}^{U}\left(n^{-2}\mathrm{Tr}(A_{2k_u-1}\varepsilon_{t_{2k_u-1}}\varepsilon^{*}_{t_{2k_u-1}-s_{2k_u-1}}A_{2k_u}\varepsilon_{t_{2k_u}-s_{2k_u}}\varepsilon^{*}_{t_{2k_u}})\right)\right)$$

$$\left(E \prod_{k \neq k_u} \left(n^{-2} \mathrm{Tr}(A_{2k-1} \varepsilon_{t_{2k-1}} \varepsilon^*_{t_{2k-1}-s_{2k-1}} A_{2k} \varepsilon_{t_{2k}-s_{2k}} \varepsilon^*_{t_{2k}}) \right) \right) |$$
$$\leq \; C, \text{ using (7.99) for } r = U \leq m \text{ and } r = m+1-U \leq m.$$

Case 2. A typical matching which is not covered in Case 1, is of the following form

$$t_{2k} = t_{2k+1}, \; s_{2k} = s_{2k+1}, \; \forall 1 \leq k \leq m. \tag{7.101}$$

Then (7.100) reduces to

$$n^{-2(m+1)} E\mathrm{Tr}\Big[\Big(\prod_{k=1}^{m+1} A_{t_{2k}} \varepsilon_{t_{2k}-s_{2k}} \varepsilon^*_{t_{2k+1}-s_{2k+1}} \Big) \Big(\prod_{k=0}^{m} A_{2(m-k)+1} \varepsilon_{t_{2(m-k)+1}} \varepsilon^*_{t_{2(t-m)}} \Big) \Big], \tag{7.102}$$

where $t_{2m+3} - s_{2m+3} = t_{2m+2}$ and restriction (7.101) holds. Now, using the idea as in the proof of (M1) condition for $\{\Delta_i, \Delta_i^*\}$ in Theorem 5.3.1, it is easy to see that (7.102) is bounded for all $n \geq 1$. Hence (7.99) is established for $r = m+1$. Therefore, proof of (7.99) and hence Lemma 7.3.2 (a) is complete. □

7.5 Proof of Corollary 7.3.8 (c)

This proof is similar to the truncation arguments given in Jin et al. [2014]. Let $X_t \sim \mathrm{MA}(q)$, $q \geq 1$ process and suppose Assumptions $(B1)$, $(B4)$ hold, $\{\varepsilon_{i,j}\} \in U(\delta)$ and $p/n \to y \in (0, \infty)$. Let

$$\tilde{\varepsilon}_{t,i} = \varepsilon_{t,i} I(|\varepsilon_{t,i}| < \eta_n n^{\frac{1}{2+\delta}}), \quad \hat{\varepsilon}_{t,i} = \tilde{\varepsilon}_{t,i} - E(\tilde{\varepsilon}_{t,i}), \; \forall t, i \text{ and some } \eta_n \downarrow 0,$$
$$\sigma^2_{t,i} = E|\hat{\varepsilon}_{t,i}|^2, \; \Delta = n^{-\frac{\delta}{4+2\delta}}, X_{t,i} = 2Ber(0.5) - 1, \text{ i.i.d. for all } t, i,$$

$$\bar{\varepsilon}_{t,i} = \begin{cases} X_{t,i}, \text{ if } \sigma^2_{t,i} < 1-\Delta, \\ \frac{\hat{\varepsilon}_{t,i}}{\sigma_{t,i}}, \text{ otherwise,} \end{cases}$$

$$\hat{\Gamma}_i(\varepsilon), \tilde{\Gamma}_i(\varepsilon), \hat{\hat{\Gamma}}_i(\varepsilon), \bar{\Gamma}_i(\varepsilon) = i\text{-th order sample autocovariance matrix of } \{\varepsilon_{t,i}\}, \{\tilde{\varepsilon}_{t,i}\}, \{\hat{\varepsilon}_{t,i}\}, \{\bar{\varepsilon}_{t,i}\}(\text{respectively}),$$

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

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

Let F^A denote the ESD of the matrix A and L denote the Lévy metric on the space of probability distribution functions. Existence of the LSD of $\{\bar{T}_i + \bar{T}^*_i\}_{i\geq 0}$ follows by Theorem 7.3.1. We will actually show that the LSD of $\{\hat{\Gamma}_i + \hat{\Gamma}^*_i\}_{i\geq 0}$ is same as that of $\{\bar{T}_i + \bar{T}^*_i\}_{i\geq 0}$ by showing that $L(F^{\hat{\Gamma}_i+\hat{\Gamma}^*_i}, F^{\bar{T}_i+\bar{T}^*_i}) \to 0 \ \forall i \geq 0$ a.s.. Note that

$$\begin{aligned} L(F^{\hat{\Gamma}_i+\hat{\Gamma}^*_i}, F^{\bar{T}_i+\bar{T}^*_i}) &\leq L(F^{\hat{\Gamma}_i+\hat{\Gamma}^*_i}, F^{\hat{T}_i+\hat{T}^*_i}) + L(F^{\hat{T}_i+\hat{T}^*_i}, F^{\tilde{T}_i+\tilde{T}^*_i}) \\ &\quad + L(F^{\tilde{T}_i+\tilde{T}^*_i}, F^{\hat{\hat{T}}_i+\hat{\hat{T}}^*_i}) + L(F^{\hat{\hat{T}}_i+\hat{\hat{T}}^*_i}, F^{\bar{T}_i+\bar{T}^*_i}) \\ &= B_1 + B_2 + B_3 + B_4, \ (\text{say}). \end{aligned} \tag{7.103}$$

We will show that, for each $1 \leq i \leq 4$, $B_i \to 0$ almost surely.

Proof of $B_1 \to 0$. By Theorem A.43 in Bai and Silverstein [2009], we have

$$\begin{aligned} B_1 &\leq 2p^{-1}\left(\text{rank}(R_{1n}) + \text{rank}(R_{2n}) + \text{rank}(R_{3n}) + \text{rank}(R_{4n})\right) \\ &\leq \frac{8q}{p} \to 0 \text{ a.s..} \end{aligned} \tag{7.104}$$

where R_{1n}, R_{2n}, R_{3n} and R_{4n} are as in (7.93).

Proof of $B_2 \to 0$. By Theorem A.43 in Bai and Silverstein [2009], we have for some $C > 0$

$$\begin{aligned}
B_2 &\leq \frac{1}{p}\text{rank}(\hat{T}_i + \hat{T}_i^* - \tilde{T}_i - \tilde{T}_i^*) \leq \frac{2}{p}\text{rank}(\hat{T}_i - \tilde{T}_i) \\
&\leq \frac{1}{p}\text{rank}\left(\sum_{j,j'=0}^{q} \psi_j \left(\hat{\Gamma}_{j-j'+i}(\varepsilon) - \tilde{\Gamma}_{j-j'+i}(\varepsilon)\right)\psi_{j'}^*\right) \\
&\leq \frac{C}{p}\text{rank}(\hat{\Gamma}_i(\varepsilon) - \tilde{\Gamma}_i(\varepsilon)) \to 0 \text{ a.s.} \qquad (7.105)
\end{aligned}$$

(see page 1210 in Jin et al. [2014]).

Proof of $B_3 \to 0$. Recall $||\cdot||_2$ in (2.4). By Theorem $A.45$ in Bai and Silverstein [2009], we have for some $C > 0$,

$$\begin{aligned}
B_3 &\leq ||\tilde{T}_i + \tilde{T}_i^* - \hat{\hat{T}}_i - \hat{\hat{T}}_i^*||_2 \\
&\leq C||\tilde{\Gamma}_i(\varepsilon) - \hat{\hat{\Gamma}}_i(\varepsilon)||_2 \to 0 \text{ a.s.} \qquad (7.106)
\end{aligned}$$

(see page 1211 in Jin et al. [2014]).

Proof of $B_4 \to 0$. By Corollary $A.41$ in Bai and Silverstein [2009], we have

$$\begin{aligned}
B_4^3 &\leq \frac{1}{p}\text{Tr}\left((\hat{\hat{T}}_i + \hat{\hat{T}}_i^* - \bar{T}_i - \bar{T}_i^*)(\hat{\hat{T}}_i + \hat{\hat{T}}_i^* - \bar{T}_i - \bar{T}_i^*)^*\right) \\
&\leq \frac{4}{p}\text{Tr}\left((\hat{\hat{T}}_i - \bar{T}_i)(\hat{\hat{T}}_i - \bar{T}_i)^*\right) \\
&= \frac{4}{p}\sum_{j,j',k,k'=0}^{q} \text{Tr}\Big(\psi_j \left(\hat{\hat{\Gamma}}_{j-j'+i}(\varepsilon) - \bar{\Gamma}_{j-j'+i}(\varepsilon)\right)\psi_{j'}^* \\
&\qquad\qquad \psi_{k'}\left(\hat{\hat{\Gamma}}_{k-k'+i}(\varepsilon) - \bar{\Gamma}_{k-k'+i}(\varepsilon)\right)^*\psi_k^*\Big) \\
&\leq 4\Bigg(\sum_{j,j'=0}^{q}\Big[p^{-1}\text{Tr}\big(\psi_j\left(\hat{\hat{\Gamma}}_{j-j'+i}(\varepsilon) - \bar{\Gamma}_{j-j'+i}(\varepsilon)\right)\psi_{j'}^* \\
&\qquad\qquad \psi_{j'}\left(\hat{\hat{\Gamma}}_{j-j'+i}(\varepsilon) - \bar{\Gamma}_{j-j'+i}(\varepsilon)\right)^*\psi_j^*\big)\Big]^{1/2}\Bigg)^2.
\end{aligned}$$

Therefore, it is enough to show that

$$p^{-1}\mathrm{Tr}(A(\hat{\hat{\Gamma}}_i(\varepsilon)-\bar{\Gamma}_i(\varepsilon))BB^*(\hat{\hat{\Gamma}}_i(\varepsilon)-\bar{\Gamma}_i(\varepsilon))^*A^*)\to 0,\ \text{a.s.}, \tag{7.107}$$

for any $A, B \in \mathrm{Span}\{\psi_j, \psi_j^* : j \geq 0\}$. The proof of (7.107) given below goes along the same lines as the proof of $p^{-1}\mathrm{Tr}((\hat{\hat{\Gamma}}_i(\varepsilon)-\bar{\Gamma}_i(\varepsilon))(\hat{\hat{\Gamma}}_i(\varepsilon)-\bar{\Gamma}_i(\varepsilon))^*)\to 0$ given in Jin et al. [2014]. In our case we have the extra factors of A, B etc. Let

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

Then,

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

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

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

where,

$$\begin{aligned}
J_1 &= \frac{C}{n^3}\sum_{u,v}\sum_{\substack{l_1,l_2,j_1,j_2\\ k_1>k_2,\ k_1\neq k_2+i}}\Big(a_{ul_1}(\hat{\varepsilon}_{k_1,l_1}-\bar{\varepsilon}_{k_1,l_1})\hat{\varepsilon}^*_{(k_1+i),j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k_2+i),j_2}(\hat{\varepsilon}_{k_2,l_2}-\bar{\varepsilon}_{k_2,l_2})^*a^*_{ul_2}\Big),\\
J_2 &= \frac{C}{n^3}\sum_{u,v}\sum_{\substack{l_1,j_1,l_2,\\ j_2,k_2}}\big(a_{ul_1}(\hat{\varepsilon}_{(k_2+i),l_1}-\bar{\varepsilon}_{(k_2+i),l_1})\hat{\varepsilon}^*_{(k_2+2i),j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k_2+i),j_2}(\hat{\varepsilon}_{k_2,l_2}-\bar{\varepsilon}_{k_2,l_2})^*a^*_{ul_2}\big),\\
J_3 &= \frac{C}{n^3}\sum_{u,v}\sum_{\substack{l_1,l_2,j_1,j_2\\ k_2>k_1,\ k_2\neq k_1+i}}\big(a_{ul_1}(\hat{\varepsilon}_{k_1,l_1}-\bar{\varepsilon}_{k_1,l_1})\hat{\varepsilon}^*_{(k_1+i),j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k_2+i),j_2}(\hat{\varepsilon}_{k_2,l_2}-\bar{\varepsilon}_{k_2,l_2})^*a^*_{ul_2}\big),\\
J_4 &= \frac{C}{n^3}\sum_{u,v}\sum_{\substack{l_1,j_1,l_2,\\ j_2,k_1}}\big(a_{ul_1}(\hat{\varepsilon}_{k_1,l_1}-\bar{\varepsilon}_{k_1,l_1})\hat{\varepsilon}^*_{(k_1+i),j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k_1+2i),j_2}(\hat{\varepsilon}_{(k_1+i),l_2}-\bar{\varepsilon}_{(k_1+i),l_2})^*a^*_{ul_2}\big),\\
J_5 &= \frac{C}{n^3}\sum_{\substack{u,v,l_1,l_2\\ j_1,j_2,k}}\Big[a_{ul_1}(\hat{\varepsilon}_{k,l_1}-\bar{\varepsilon}_{k,l_1})\hat{\varepsilon}^*_{(k+i),j_1}b_{j_1v}b^*_{j_2v}\hat{\varepsilon}_{(k+i),j_2}(\hat{\varepsilon}_{k,l_2}-\bar{\varepsilon}_{k,l_2})^*a^*_{ul_2}\Big].
\end{aligned}$$

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

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

Also for some $C_1,C_2>0$,

$$\mathrm{Var}(J_2)=E(J_2)^2$$

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

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

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

$$1 = \mathrm{Var}(\varepsilon_{t,i}) = \mathrm{Var}\left(\tilde{\varepsilon}_{t,i} - E(\tilde{\varepsilon}_{t,i}) + \tilde{\tilde{\varepsilon}}_{t,i} - E(\tilde{\tilde{\varepsilon}}_{t,i})\right) = \sigma^2_{t,i} + \mathrm{Var}(\tilde{\tilde{\varepsilon}}_{t,i}) + 2(E(\tilde{\tilde{\varepsilon}}_{t,i}))^2.$$

Therefore, as $\{\varepsilon_{i,j}\} \in U(\delta, p(n), n, n)$, for some $C > 0$

$$(1 - \sigma^2_{t,i}) \leq 2E(\tilde{\tilde{\varepsilon}}^2_{t,i}) \leq 2C(P(|\varepsilon_{t,i}| > \eta_n n^{\frac{1}{2+\delta}}))^{\frac{\delta}{2+\delta}} \leq 2C\eta_n^{-\delta} n^{-\frac{\delta}{2+\delta}} \quad (7.108)$$

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

$$(1 - \sigma^{-1}_{t,i})^2 \leq C\eta_n^{-2\delta} n^{-\frac{2\delta}{2+\delta}}. \quad (7.109)$$

Moreover note that if $(t,i) \in E$, then $\frac{1-\sigma^2_{t,i}}{\Delta} > 1$. Then by (7.108) and (7.109), we have for some $C_1, C_2 > 0$,

$$E(J_5) = \frac{C}{n^3} \sum_{u,v,l_1,j_1,k} a_{ul_1} E|\hat{\varepsilon}_{k,l_1} - \bar{\varepsilon}_{k,l_1}|^2 E|\hat{\varepsilon}_{(k+i),j_1}|^2 b_{j_1v} b^*_{j_1v} a^*_{ul_1}$$
$$\leq \frac{C_1}{n^3} \sum_{\substack{u,v,j_1\\(k,l_1)\in E}} a_{ul_1} b_{j_1v} b^*_{j_1v} a^*_{ul_1} \left(\frac{1 - \sigma^2_{kl_1}}{\Delta}\right)$$

$$+\frac{C_2}{n^3} \sum_{\substack{u,v,j_1 \\ (k,l_1) \notin E}} a_{ul_1} b_{j_1 v} b^*_{j_1 v} a^*_{ul_1} \left(1 - \sigma^{-1}_{kl}\right)^2$$
$$= O(n^{\frac{-\delta}{4+2\delta}}) + O(n^{\frac{-2\delta}{2+\delta}}).$$

Therefore,

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

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

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

Hence,

$$p^{-1}\mathrm{Tr}(A(\hat{U} - \bar{U})\hat{V}^* BB^* \hat{V}(\hat{U} - \bar{U})^* A^*) \to 0, a.s.. \tag{7.110}$$

Similarly,

$$p^{-1}\mathrm{Tr}(A\bar{U}(\hat{V} - \bar{V})^* BB^* (\hat{V} - \bar{V})\bar{U}^* A^*) \to 0, a.s.. \tag{7.111}$$

Hence by (7.110) and (7.111), (7.107) is established and $B_4 \to 0$ almost surely. Therefore, proof of Corollary 7.3.8 (c) is complete. □

7.6 Proof of Corollary 7.3.12 (c)

Additionally if we assume $\sup_{t,i} E|\varepsilon_{ti}|^4 < M < \infty$, then to show that the LSD of $\{\hat{\Gamma}_i \hat{\Gamma}_i^*\}_{i \geq 0}$ exists. Proof goes through exactly the same lines as Corollary 7.3.8 (c). Hence we omit the detailed calculations and briefly outline the steps. The convergence below are all in the almost sure sense.

1. $$\begin{aligned} L(F^{\hat{\Gamma}_i \hat{\Gamma}_i^*}, F^{\hat{T}_i \hat{T}_i^*}) &\leq p^{-1}\mathrm{rank}(\hat{\Gamma}_i \hat{\Gamma}_i^* - \hat{T}_i \hat{T}_i^*) \\ &\leq p^{-1}\mathrm{rank}((\hat{\Gamma}_i - \hat{T}_i)\hat{\Gamma}_i^* + \hat{T}_i(\hat{\Gamma}_i - \hat{T}_i)^*) \\ &\leq 2p^{-1}\mathrm{rank}(\hat{\Gamma}_i - \hat{T}_i) \to 0 \text{ (proof similar to } B_1 \to 0). \end{aligned}$$

2. $L(F^{\hat{T}_i\hat{T}_i^*}, F^{\tilde{T}_i\tilde{T}_i^*}) \leq 2p^{-1}\text{rank}(\hat{T}_i - \tilde{T}_i) \to 0$ (proof similar to $B_2 \to 0$).

By Corollary A.42 of Bai and Silverstein [2009]

$$
\begin{aligned}
3.\ L^4(F^{\tilde{T}_i\tilde{T}_i^*}, F^{\hat{\hat{T}}_i\hat{\hat{T}}_i^*}) &\leq 2p^{-2}\text{Tr}(\tilde{T}_i\tilde{T}_i^* + \hat{\hat{T}}_i\hat{\hat{T}}_i^*)\text{Tr}((\tilde{T}_i - \hat{\hat{T}}_i)(\tilde{T}_i - \hat{\hat{T}}_i)^*) \\
&\leq 2p^{-1}\text{Tr}(\tilde{T}_i\tilde{T}_i^* + \hat{\hat{T}}_i\hat{\hat{T}}_i^*)||\tilde{T}_i - \hat{\hat{T}}_i||_2^2, \\
&\to 0 \text{ (proof similar to } B_3 \to 0). \\
4.\ L^4(F^{\bar{T}_i\bar{T}_i^*}, F^{\hat{\hat{T}}_i\hat{\hat{T}}_i^*}) &\leq 2p^{-2}\text{Tr}(\bar{T}_i\bar{T}_i^* + \hat{\hat{T}}_i\hat{\hat{T}}_i^*)\text{Tr}((\bar{T}_i - \hat{\hat{T}}_i)(\bar{T}_i - \hat{\hat{T}}_i)^*) \\
&\to 0 \text{ (proof similar to } B_4 \to 0).
\end{aligned}
$$

This completes the proof of Corollary 7.3.10 (c). □

7.7 Proof of Corollary 7.3.18 (c)

Here we show that Corollary 7.3.18 (a) remains true even if we drop (B3) and use the more relaxed Assumption $\{\varepsilon_{i,j}\} \in U(\delta, p, n(p), p)$ for some $\delta > 0$. This is achieved by truncation arguments similar to those given on pages $1210 - 1217$ of Jin et al. [2014].

Let $X_t \sim \text{MA}(q)$ and suppose Assumptions (B1), (B4) hold, $\{\varepsilon_{i,j}\} \in \mathcal{L}_4 \cap U(\delta)$ for some $\delta > 0$ and $p/n \to 0$. Let

$$
\begin{aligned}
\tilde{\varepsilon}_{t,i} &= \varepsilon_{t,i}I(|\varepsilon_{t,i}| < \eta n^{\frac{1}{2+\delta}}),\ \ \hat{\varepsilon}_{t,i} = \tilde{\varepsilon}_{t,i} - E(\tilde{\varepsilon}_{t,i}),\ \forall t, i \text{ and some } \eta > 0, \\
\sigma_{t,i}^2 &= E|\hat{\varepsilon}_{t,i}|^2,\ \Delta = n^{-\frac{\delta}{4+2\delta}},\ \ B_{t,i} = 2Ber(0.5) - 1, \text{ i.i.d. for all } t, i, \\
\bar{\varepsilon}_{t,i} &= \begin{cases} B_{t,i}, \text{ if } \sigma_{t,i}^2 < 1 - \Delta, \\ \frac{\hat{\varepsilon}_{t,i}}{\sigma_{t,i}}, \text{ otherwise,} \end{cases} \qquad C_p = \Gamma_i + \Gamma_i^*,
\end{aligned}
$$

$$
\begin{aligned}
\hat{\Gamma}_i(\varepsilon), \tilde{\Gamma}_i(\varepsilon), \hat{\hat{\Gamma}}_i(\varepsilon), \bar{\Gamma}_i(\varepsilon) &= i\text{-th order sample autocovariance matrix of} \\
&\quad \{\varepsilon_{t,i}\}, \{\tilde{\varepsilon}_{t,i}\}, \{\hat{\varepsilon}_{t,i}\}, \{\bar{\varepsilon}_{t,i}\}\text{(respectively)},
\end{aligned}
$$

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

Since $\{\bar{\varepsilon}_{t,i}\}$ satisfy the stronger assumption $(B3)$, the existence of the LSD of $\sqrt{np^{-1}}(\bar{T}_i + \bar{T}_i^* - C_p)$ is guaranteed by Corollary 7.3.18 (a).

We will actually show that the LSD of $\sqrt{np^{-1}}(\hat{\Gamma}_i + \hat{\Gamma}_i^* - C_p)$ is same as that of $\sqrt{np^{-1}}(\bar{T}_i + \bar{T}_i^* - C_p)$. Let L be the Levy metric between two distribution functions. For any matrix A, F^A denotes the cumulative distribution function of the ESD of A. Then note that

$$\begin{aligned}
& L(F^{\sqrt{np^{-1}}(\hat{\Gamma}_i+\hat{\Gamma}_i^*-C_p)}, F^{\sqrt{np^{-1}}(\bar{T}_i+\bar{T}_i^*-C_p)}) \leq L(F^{\sqrt{np^{-1}}(\hat{\Gamma}_i+\hat{\Gamma}_i^*-C_p)}, F^{\sqrt{np^{-1}}(\hat{T}_i+\hat{T}_i^*-C_p)}) \\
& + L(F^{\sqrt{np^{-1}}(\hat{T}_i+\hat{T}_i^*-C_p)}, F^{\sqrt{np^{-1}}(\tilde{T}_i+\tilde{T}_i^*-C_p)}) + L(F^{\sqrt{np^{-1}}(\tilde{T}_i+\tilde{T}_i^*-C_p)}, F^{\sqrt{np^{-1}}(\hat{\hat{T}}_i+\hat{\hat{T}}_i^*-C_p)}) \\
& + L(F^{\sqrt{np^{-1}}(\hat{\hat{T}}_i+\hat{\hat{T}}_i^*-C_p)}, F^{\sqrt{np^{-1}}(\bar{T}_i+\bar{T}_i^*-C_p)}). \\
= \ & T_1 + T_2 + T_3 + T_4, \ \text{(say)}.
\end{aligned} \tag{7.112}$$

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

To prove $T_1 \to 0$ almost surely, note that

$$\begin{aligned}
n\hat{\Gamma}_{i.p} &= \sum_{j,j'=0}^{q} \psi_{j.p}^{(n)} \hat{\Gamma}_{j'-j+i}(\varepsilon) \psi_{j'.p}^{(n)*} - \sum_{\substack{j,j'=0 \\ j-j'\neq i}}^{q} \sum_{t=n-j+1}^{n} \psi_{j.p}^{(n)} \varepsilon_{t,p}\varepsilon_{t-(j'+i-j)}^* \psi_{j'.p}^{(n)*} \\
&+ \sum_{\substack{j,j'=0 \\ j-j'\neq i}}^{q} \sum_{t=i-j+1}^{j'+i-j} \psi_{j.p}^{(n)} \varepsilon_{t.p}\varepsilon_{t-(j'+i-j)}^* \psi_{j'.p}^{(n)*} + \sum_{j=0}^{q} \sum_{t=n-j+1}^{n} \psi_{j.p}^{(n)} \varepsilon_{t.p}\varepsilon_{t.p}^* \psi_{j-i.p}^{(n)*} \\
&+ \sum_{j=0}^{q} \sum_{t=i-j+1}^{0} \psi_{j.p}^{(n)} \varepsilon_{t.p}\varepsilon_{t.p}^* \psi_{j-i.p}^{(n)*} \\
&= \hat{T}_i + R_{1p} + R_{2p} + R_{3p} + R_{4p}, \ \text{(say)}.
\end{aligned} \tag{7.113}$$

By Theorem $A.43$ in Bai and Silverstein [2009], we have for some $C > 0$, with

R_{1n}, R_{2n}, R_{3n} and R_{4n} as in (7.113),

$$\begin{aligned} T_1 &\leq p^{-1}\left(\operatorname{rank}(R_{1p}) + \operatorname{rank}(R_{2p}) + \operatorname{rank}(R_{3p}) + \operatorname{rank}(R_{4p})\right) \\ &\leq \frac{4Cq}{p} \to 0 \text{ a.s..} \end{aligned} \tag{7.114}$$

By Theorem $A.43$ in Bai and Silverstein [2009], we have for some $C, C_1 > 0$

$$\begin{aligned} T_2 &\leq \frac{1}{p}\operatorname{rank}(\hat{T}_i + \hat{T}_i^* - \tilde{T}_i - \tilde{T}_i^*) \leq \frac{2}{p}\operatorname{rank}(\hat{T}_i - \tilde{T}_i) \\ &\leq \frac{1}{p}\operatorname{rank}\left(\sum_{j,j'=0}^{q} \psi_j \left(\hat{\Gamma}_{j-j'+i}(\varepsilon) - \tilde{\Gamma}_{j-j'+i}(\varepsilon)\right)\psi_{j'}^*\right) \\ &\leq \frac{C}{p}\operatorname{rank}(\hat{\Gamma}_i(\varepsilon) - \tilde{\Gamma}_i(\varepsilon)) \\ &\leq \frac{C_1}{p}\sum_{j=1}^{p}\sum_{t=1}^{n+i} I(|\varepsilon_{t,j}| \geq \eta p^{1/(2+\delta)}). \end{aligned} \tag{7.115}$$

Also, we have

$$\begin{aligned} &E\left(\frac{1}{p}\sum_{j=1}^{p}\sum_{t=1}^{n+i} I(|\varepsilon_{t,j}| \geq \eta p^{1/(2+\delta)})\right) \\ \leq\ & \frac{1}{\eta^{2+\delta}p^2}\sum_{j=1}^{p}\sum_{t=1}^{n+i} E\left(|\varepsilon_{t,j}|^{(2+\delta)} I(|\varepsilon_{t,j}| \geq \eta p^{1/(2+\delta)})\right) = o(1) \end{aligned} \tag{7.116}$$

and

$$\begin{aligned} &\operatorname{Var}\left(\frac{1}{p}\sum_{j=1}^{p}\sum_{t=1}^{n+i} I(|\varepsilon_{t,j}| \geq \eta p^{1/(2+\delta)})\right) \\ \leq\ & \frac{1}{\eta^{2+\delta}p^3}\sum_{j=1}^{p}\sum_{t=1}^{n+i} E\left(|\varepsilon_{t,j}|^{(2+\delta)} I(|\varepsilon_{t,j}| \geq \eta p^{1/(2+\delta)})\right) = o(p^{-1}). \end{aligned} \tag{7.117}$$

Applying Bernstein's inequality and (7.116), (7.117), 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

$$T_2 \to 0 \text{ a.s..} \tag{7.118}$$

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 [2009], we have for some $C, C_1 > 0$,

$$\begin{aligned}
T_3 &\leq \sqrt{np^{-1}}||\tilde{T}_i + \tilde{T}_i^* - \hat{\hat{T}}_i - \hat{\hat{T}}_i^*||_2 \leq C\sqrt{np^{-1}}||\tilde{\Gamma}_i(\varepsilon) - \hat{\hat{\Gamma}}_i(\varepsilon)||_2 \\
&\leq C_1\sqrt{np^{-1}}||\sum_{k=1}^{n}(\hat{\gamma}_k E\tilde{\gamma}_{k+i}^* + \hat{\gamma}_{k+i}E\tilde{\gamma}_k^*)||_2 \\
&\quad + C_1\sqrt{np^{-1}}||\sum_{k=1}^{n}(E\hat{\gamma}_k E\tilde{\gamma}_{k+i}^* + E\hat{\gamma}_{k+i}E\tilde{\gamma}_k^*)||_2.
\end{aligned} \tag{7.119}$$

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

$$\begin{aligned}
&\sqrt{np^{-1}}||\sum_{k=1}^{n}(E\hat{\gamma}_k E\tilde{\gamma}_{k+i}^* + E\hat{\gamma}_{k+i}E\tilde{\gamma}_k^*)||_2 \\
\leq\ & \sqrt{(np)^{-1}}\sum_{k=1}^{n}\sum_{j=1}^{p}|E(\varepsilon_{k,j}I(|\varepsilon_{k,j}| > \eta p^{1/(2+\delta)}))E(\varepsilon_{k+i,j}I(|\varepsilon_{k+i,j}| > \eta p^{1/(2+\delta)}))| \\
\leq\ & C\frac{p}{\sqrt{np}}p^{-2}\sum_{k=1}^{n+i}\sum_{j=1}^{p}E(|\varepsilon_{k,j}|^{2+\delta}I(|\varepsilon_{k,j}| > \eta p^{1/(2+\delta)})) = o(1).
\end{aligned} \tag{7.120}$$

For the first part, note that

$$\begin{aligned}
&np^{-1}||\sum_{k=1}^{n}(\hat{\gamma}_k E\tilde{\gamma}_{k+i}^* + \hat{\gamma}_{k+i}E\tilde{\gamma}_k^*)||_2^2 \\
\leq\ & 2np^{-1}\left(||\sum_{k=1}^{n}\hat{\gamma}_k E\tilde{\gamma}_{k+i}^*||_2^2 + ||\sum_{k=1}^{n}\hat{\gamma}_{k+i}E\tilde{\gamma}_k^*||_2^2\right).
\end{aligned} \tag{7.121}$$

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

$$\begin{aligned}
np^{-1}||\sum_{k=1}^{n}\hat{\gamma}_k E\tilde{\gamma}^*_{k+i}||_2^2 &\leq C(np)^{-1}\sum_{j=1}^{p}\sum_{l=1}^{p}\left(\sum_{k=1}^{n}\hat{\varepsilon}_{k,j}E\tilde{\varepsilon}_{k+i,l}\right)^2 \\
&= C(np)^{-1}\sum_{j=1}^{p}\sum_{l=1}^{p}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}(\hat{\varepsilon}_{k_1,j}E\tilde{\varepsilon}_{k_1+i,l}\hat{\varepsilon}_{k_2,j}E\tilde{\varepsilon}_{k_2+i,l}) \\
&= C(np)^{-1}\sum_{j=1}^{p}\sum_{l=1}^{p}\left(\sum_{k_1=1}^{n}\hat{\varepsilon}^2_{k_1,j}(E\tilde{\varepsilon}_{k_1+i,l})^2+\sum_{k_1\neq k_2}\hat{\varepsilon}_{k_1,j}E\tilde{\varepsilon}_{k_1+i,l}\hat{\varepsilon}_{k_2,j}E\tilde{\varepsilon}_{k_2+i,l}\right) \\
&= J_{11}+J_{12},\ \text{(say)}. \qquad (7.122)
\end{aligned}$$

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

$$\begin{aligned}
EJ_{11} &= C(np)^{-1}\sum_{j=1}^{p}\sum_{l=1}^{p}\sum_{k_1=1}^{n}\hat{\varepsilon}^2_{k_1,j}(E\tilde{\varepsilon}_{k_1+i,l})^2 \\
&\leq C(np)^{-1}\sum_{j=1}^{p}\sum_{l=1}^{p}\sum_{k_1=1}^{n}(E(|\varepsilon_{k_1,l}|I(|\varepsilon_{k_1,l}|>\eta p^{1/(2+\delta)})))^2 \\
&\leq C(np)^{-1}\eta^{-2(1+\delta)}p^{-2(1+\delta)/(2+\delta)}\sum_{j,l=1}^{p}\sum_{k_1=1}^{n}E(|\varepsilon_{k_1,l}|^{2+\delta}I(|\varepsilon_{k_1,l}|>\eta p^{1/(2+\delta)})))^2 \\
&= O(p^{-\delta/(2+\delta)}). \qquad (7.123)
\end{aligned}$$

and

$$\begin{aligned}
\text{Var}J_{11} &= C^2(np)^{-2}\sum_{j=1}^{p}\sum_{k_1=1}^{n}E(\hat{\varepsilon}^2_{k_1,j}-E\hat{\varepsilon}^2_{k_1,j})^2\left(\sum_{l=1}^{p}(E\tilde{\varepsilon}_{k_1+i,l})^2\right)^2 \\
&\leq C_2(np)^{-2}\sum_{j=1}^{p}\sum_{k_1=1}^{n}E(\tilde{\varepsilon}^4_{k_1,j})\left(p\eta^{-2(1+\delta)}p^{-2(1+\delta)/(2+\delta)}\right)^2 \\
&= O(p^{-1-4\delta/(2+\delta)}). \qquad (7.124)
\end{aligned}$$

Therefore, by (7.123), (7.124) and Borel-Cantelli Lemma, $J_{11} \to 0$ a.s.. Further, we have $E(J_{12}) = 0$ and

$$\begin{aligned}
\mathrm{Var} J_{12} &= C(np)^{-2} \sum_{j=1}^{p} \sum_{k_1 \neq k_2} E\hat{\varepsilon}^2_{k_1,j} E\hat{\varepsilon}^2_{k_2,j} \left(\sum_{l=1}^{p} E\tilde{\varepsilon}_{k_1+i,l} E\tilde{\varepsilon}_{k_2+i,l} \right)^2 \\
&\leq C_3(np)^{-2} \sum_{j=1}^{p} \sum_{k_1 \neq k_2} \left(p\eta^{-2(1+\delta)} p^{-2(1+\delta)/(2+\delta)} \right)^2 \\
&= O(p^{-1-2\delta/(2+\delta)}), \qquad (7.125)
\end{aligned}$$

which by Borel-Cantelli Lemma imply $J_{12} \to 0$, a.s.. Hence, we have $|| \sum_{k=1}^{n} \hat{\gamma}_k E\tilde{\gamma}^*_{k+i} ||_2^2 \to 0$, a.s.. Similarly, $|| \sum_{k=1}^{n} \hat{\gamma}_{k+i} E\tilde{\gamma}^*_k ||_2^2 \to 0$, a.s.. Thus, by (7.119)-(7.121)

$$T_3 \to 0, \text{ a.s.}. \qquad (7.126)$$

We now finally prove $T_4 \to 0$ almost surely. By Corollary $A.41$ in Bai and Silverstein [2009], we have

$$\begin{aligned}
T_4^3 &\leq \frac{n}{p^2} \mathrm{Tr} \left((\hat{\hat{T}}_i + \hat{\hat{T}}_i^* - \bar{T}_i - \bar{T}_i^*)(\hat{\hat{T}}_i + \hat{\hat{T}}_i^* - \bar{T}_i - \bar{T}_i^*)^* \right) \\
&\leq \frac{4n}{p^2} \mathrm{Tr} \left((\hat{\hat{T}}_i - \bar{T}_i)(\hat{\hat{T}}_i - \bar{T}_i)^* \right) \\
&= \frac{4n}{p^2} \sum_{j,j',k,k'=0}^{q} \mathrm{Tr} \Bigg(\psi_j \left(\hat{\hat{\Gamma}}_{j-j'+i}(\varepsilon) - \bar{\Gamma}_{j-j'+i}(\varepsilon) \right) \psi^*_{j'} \\
&\qquad \psi_{k'} \left(\hat{\hat{\Gamma}}_{k-k'+i}(\varepsilon) - \bar{\Gamma}_{k-k'+i}(\varepsilon) \right)^* \psi^*_k \Bigg).
\end{aligned}$$

Therefore, it is enough to show that

$$np^{-2} \mathrm{Tr}(A(\hat{\hat{\Gamma}}_i(\varepsilon) - \bar{\Gamma}_i(\varepsilon)) BB^* (\hat{\hat{\Gamma}}_i(\varepsilon) - \bar{\Gamma}_i(\varepsilon))^* A^*) \to 0, \text{ a.s.} \qquad (7.127)$$

for any $A, B \in \mathrm{Span}\{\psi_j, \psi_j^* : j \geq 0\}$. The proof of (7.127) given below goes along the same lines as the proof of $p^{-1}\mathrm{Tr}((\hat{\hat{\Gamma}}_i(\varepsilon) - \bar{\Gamma}_i(\varepsilon))(\hat{\hat{\Gamma}}_i(\varepsilon) - \bar{\Gamma}_i(\varepsilon))^*) \to 0$ given in page $1210 - 1217$ of Jin et al. [2014]. In our case we have the extra factors of A,

B etc. Let

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

Then,

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

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

$$
\begin{aligned}
& p^{-1}\mathrm{Tr}(A(\hat{U} - \bar{U})\hat{V}^*BB^*\hat{V}(\hat{U} - \bar{U})^*A^*)\\
\leq \;& \frac{Cn}{p^2n^2}\sum_{u,v}|\sum_{l,k,j} a_{ul}(\hat{\varepsilon}_{k,l} - \bar{\varepsilon}_{k,l})\hat{\varepsilon}^*_{(k+i),j}b_{jv}|^2\\
= \;& \frac{Cn}{p^2n^2}\sum_{u,v}\sum_{l_1,k_1,j_1}\sum_{l_2,k_2,j_2}(a_{ul_1}(\hat{\varepsilon}_{k_1,l_1} - \bar{\varepsilon}_{k_1,l_1})\hat{\varepsilon}^*_{(k_1+i),j_1}b_{j_1v}\\
& \qquad b^*_{j_2v}\hat{\varepsilon}_{(k_2+i),j_2}(\hat{\varepsilon}_{k_2,l_2} - \bar{\varepsilon}_{k_2,l_2})^*a^*_{ul_2})\\
= \;& J_1 + J_2 + J_3 + J_4 + J_5,
\end{aligned}
\tag{7.128}
$$

where,

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

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

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

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

Note that $E(J_1)=E(J_2)=E(J_3)=E(J_4)=0$.

Moreover for some $C_1,C_2,C_3>0$,

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

Also for some $C_1,C_2>0$,

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

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

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

$$1 = \mathrm{Var}(\varepsilon_{t,i}) = \mathrm{Var}\left(\tilde{\varepsilon}_{t,i} - E(\tilde{\varepsilon}_{t,i}) + \tilde{\tilde{\varepsilon}}_{t,i} - E(\tilde{\tilde{\varepsilon}}_{t,i})\right) = \sigma_{t,i}^2 + \mathrm{Var}(\tilde{\tilde{\varepsilon}}_{t,i}) + 2(E(\tilde{\tilde{\varepsilon}}_{t,i}))^2.$$

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

$$(1 - \sigma_{t,i}^2) \leq 2E(\tilde{\tilde{\varepsilon}}_{t,i}^2) \leq 2C(P(|\varepsilon_{t,i}| > \eta_n n^{\frac{1}{2+\delta}}))^{\frac{\delta}{2+\delta}} \leq 2C\eta_n^{-\delta} p^{-\frac{\delta}{2+\delta}} \tag{7.129}$$

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

$$(1 - \sigma_{t,i}^{-1})^2 \leq C\eta_n^{-2\delta} p^{-\frac{2\delta}{2+\delta}}. \tag{7.130}$$

Moreover note that if $(t,i) \in E$, then $\frac{1-\sigma_{t,i}^2}{\Delta} > 1$. Then by (7.129) and (7.130), we have for some $C_1, C_2 > 0$,

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

Similarly one can show that for some $\epsilon > 0$, $\mathrm{Var}(J_5) = O(p^{-1-\epsilon})$ Therefore, using (7.128) and the estimate for $E(J_i)$ and $V(J_i)$,

$$\begin{aligned}
E(np^{-2}\mathrm{Tr}(A(\hat{U} - \bar{U})\hat{V}^* BB^* \hat{V}(\hat{U} - \bar{U})^* A^*)) &\to 0, \text{ and} \qquad (7.131) \\
\mathrm{Var}(np^{-2}\mathrm{Tr}(A(\hat{U} - \bar{U})\hat{V}^* BB^* \hat{V}(\hat{U} - \bar{U})^* A^*)) &= O(p^{-1-\epsilon}). \qquad (7.132)
\end{aligned}$$

Hence, by (7.131), (7.132) and Borel-Cantelli Lemma,

$$np^{-2}\text{Tr}(A(\hat{U}-\bar{U})\hat{V}^*BB^*\hat{V}(\hat{U}-\bar{U})^*A^*) \to 0, a.s.. \tag{7.133}$$

Similarly,

$$np^{-2}\text{Tr}(A\bar{U}(\hat{V}-\bar{V})^*BB^*(\hat{V}-\bar{V})\bar{U}^*A^*) \to 0, a.s.. \tag{7.134}$$

Hence by (7.133) and (7.134), (7.127) is proved. Also by (7.114), (7.115), (7.119) and (7.127),

$$T_4 \to 0, a.s.. \tag{7.135}$$

Since we have shown $T_i \to 0$ almost surely for all $i = 1, 2, 3, 4$ in (7.114), (7.118), (7.126) and (7.135), the proof of Corollary 7.3.18 (c) is now complete. □

Chapter 8

Inference in high dimensional time series: order determination and testing

8.1 Introduction

In Chapter 7, we have established the LSD of any symmetric polynomial in sample autocovariance matrices $\{\hat{\Gamma}_u\}$. These results have plenty of potential for application in high-dimensional time series. Liu [2013] used Theorem 7.2.4 to estimate the spectral distribution of the coefficient matrices $\{\psi_j\}$ of the infinite dimensional MA(q) process defined in (3.7). In this chapter, we shall discuss a couple of other applications:

1. a model identification problem, namely determination of the *unknown order* of infinite dimensional moving average (MA) and autoregressive (AR) processes and,

2. *testing* of simple hypotheses by using asymptotic normality of traces of polynomials.

In the univariate set up, a plot of the sample autocovariances provides a method to identify the order of an MA process. If the sample autocovariances are almost equal to zero for order $u > \hat{q}$, then $\hat{q}$ is taken to be an estimator of the unknown order. Similar method is also applicable for AR processes. The only difference is that for an AR process, the sample *partial autocovariances* are plotted instead of the sample autocovariances. The theoretical support for this method is the fact that the population autocovariances of order greater than q are all zero for an MA(q) process. Similarly, for an AR(r) process, the population partial autocovariances of order greater than r vanish. Moreover, since the sample autocovariances are consistent for the population autocovariances, for large enough sample size, the sample autocovariances are close to population autocovariances.

In the high-dimensional setting, no equivalent method seems to be available in the literature to determine the unknown order of MA and AR processes. The above method cannot be extended naively since, as we have seen in Chapter 3, the sample autocovariance matrices $\{\hat{\Gamma}_u\}$ are not consistent for the population autocovariance matrices $\{\Gamma_u\}$. Nevertheless, $\{\hat{\Gamma}_u\}$ provide a graphical method for order determination of high-dimensional processes.

For the infinite dimensional MA(q) process, when $p/n \to y > 0$, Corollary 7.3.12 (c) guarantees that for large sample size, ESD of $\hat{\Gamma}_u \hat{\Gamma}_u^*$ would be close for $u > q$ and different for $0 \leq u \leq q$. When $p/n \to 0$, a similar result for $\sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - \Gamma_u - \Gamma_u)$ is guaranteed by Corollary 7.3.18 (c). Clearly, this property of sample autocovariance matrices provides a clue to identify graphically the unknown order of an MA process. For more details see Section 8.2.

To apply a similar idea to an AR(r) process defined in (3.9) with unknown parameter matrices, we first need consistent estimators of the parameter matrices $\{A_i : 1 \leq i \leq r\}$. Note that we have already obtained such estimators in Chapter 3. Let us denote these consistent estimators by $\{\hat{A}_i^{(r)} : 1 \leq i \leq r\}$. Now, suppose r is unknown. Consider the residual process $\{\hat{\varepsilon}_t^{(s)}\}$ after fitting the AR(s) process using $\{\hat{A}_i^{(s)} : 1 \leq i \leq s\}$. In Theorem 8.3.1 and Remark 8.3.2, we argue that the residual process $\{\hat{\varepsilon}_t^{(s)}\}$ behaves like an MA(0) process if and only if $s = r$, the true order of the AR process. We use these results to identify the unknown order of an AR processes.

Linear spectral statistics of a random matrix M are of the form $\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i)$ where $\{\lambda_i\}$ are eigenvalues of M and f is a "suitable" function. Such statistics have been discussed in Diaconis and Evans [2001], Bai and Silverstein [2004] and Bai et al. [2009]. Asymptotic normality of these statistics is extremely useful in statistical inference. While we do not discuss these statistics in general in this thesis, we deal with a specific spectral linear statistics, namely traces of polynomials in sample autocovariance matrices. In Section 8.4, we establish the asymptotic normality of these statistics and suggest how it may be used for testing problems in high-dimensional time series.

The main material of this chapter is taken from Bhattacharjee and Bose [2015a] and Bhattacharjee and Bose [2015b].

8.2 Order determination of an MA processes

Consider the infinite dimensional MA(q) process defined in (3.7) with unknown coefficient matrices. Suppose $\{X_{t,p} : 1 \leq t \leq n\}$ is a sample of size n from the process (3.7), where q is unknown. Our problem is to estimate q. The graphical method that we suggest to identify q is based on Corollaries 7.3.8 (d), 7.3.12 (d) and 7.3.18 (d). For convenience of the reader, we collect those results along with some more consequences in the following theorem. We omit its proof. Existence of all the LSD below is guaranteed by Theorems 7.3.1 and 7.3.15. Recall the classes $\mathcal{L}_r$ and $U(\delta)$ respectively in (4.14) and (4.17).

For future use, let us denote

$$\begin{aligned}
\Pi_{1u} &= \hat{\Gamma}_u\hat{\Gamma}_u^*,\\
\Pi_{2u} &= \hat{\Gamma}_u\hat{\Gamma}_u^* + \hat{\Gamma}_{u+1}\hat{\Gamma}_{u+1}^*,\\
\Pi_{3u} &= \sqrt{np^{-1}}(\hat{\Gamma}_u + \hat{\Gamma}_u^* - \Gamma_u - \Gamma_u^*),\\
\Pi_{4u} &= \sqrt{np^{-1}}(\hat{\Gamma}_u\hat{\Gamma}_u^* - \Gamma_u\Gamma_u^*).
\end{aligned}$$

Theorem 8.2.1. *Consider the model (3.7). Suppose* $\{\varepsilon_{i,j}\}$ *are independently distributed with* $E(\varepsilon_{i,j}) = 0$, $E|\varepsilon_{i,j}|^2 = 1$, $\forall i, j$ *and* $\{\varepsilon_{i,j}\} \in \mathcal{L}_4 \cap U(\delta)$ *for some* $\delta \in (0, 2]$. *Further* $\{\psi_j\}$ *are all norm bounded and converge jointly.*

(a) If $p/n \to y > 0$, *then*

(i) the LSD of Π_{1u} *are identical for* $u > q$ *and are different for* $0 \le u \le q$, *and*

(ii) the LSD of Π_{2u} *are identical for* $u > q$ *and are different for* $0 \le u \le q$.

(b) If $p/n \to 0$, *then*

(i) the LSD of Π_{3u} *are identical for* $u > q$ *and are different for* $0 \le u \le q$, *and*

(ii) the LSD of Π_{4u} *are identical for* $u > q$ *and are different for* $0 \le u \le q$.

It may be noted that even though the above theorem is stated for four specific polynomials of $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$, the conclusion of Theorem 8.2.1 holds true for other polynomials if we are ready to make appropriate moment assumptions on $\{\varepsilon_{i,j}\}$. We have restricted to the above polynomials only for illustrative purposes.

Identification of q. If $p/n \to y > 0$, we propose to plot the CDF of ESD (call it ECDF) of Π_{1u} (or Π_{2u}) for first few sample autocovariance matrices in the same graph. If $p/n \to 0$, we propose to plot the ECDF of Π_{3u} (or Π_{4u}) for first few sample autocovariance matrices in the same graph. We say that $\hat{q}$ is an estimate of q, if the ECDF of Π_{iu} with order $u > \hat{q}$ empirically coincide with each other. Note that by Theorem 8.2.1, for large enough n, $\hat{q}$ is a reasonable estimator of q.

8.2.1 Simulations

This section presents some simulation results on the above proposed method. As we shall see, it works very well. First we define the models from where we simulate.

Models. Let I_p and J_p be respectively as in (2.8) and (2.9). Let

$$\varepsilon_t \sim \mathcal{N}_p(0, I_p), \forall t. \tag{8.1}$$

Suppose

$$\begin{aligned} A_p &= 0.5I_p, \quad B_p = 0.5(I_p + J_p), \\ C_p &= ((I(1 \le i = j \le [p/2]) - I([p/2] < i = j \le p))), \\ D_p &= ((I(i + j = p + 1))). \end{aligned}$$

We consider the following models.

Model 1 $X_t = \varepsilon_t$.

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

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

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

By Examples 4.2.1-4.2.4, LSD of A_p, B_p, C_p and D_p exist. Moreover, it is easy to see that $\{C_p, D_p\}$ converge jointly. Hence, Theorems 7.3.1 and 7.3.15 are applicable for the above models whenever $p/n \to y > 0$ and $p/n \to 0$, respectively. Incidentally, note that in Model 4, $C_pD_p \neq D_pC_p$, and therefore they are not simultaneously diagonalizable. Hence, Theorems 7.2.4 (Liu et al. [2015]) and 7.2.5 (Wang et al. [2015]) are not applicable for Model 4.

Case a: $p/n \to y > 0$.

Now suppose the original model is unknown to us. Suppose we have random sample S_i each from Model i, $1 \le i \le 4$. For each i, we are asked to determine the order of the corresponding moving average process. For each $1 \le i \le 4$, we plot the ECDF of Π_{1u} for $1 \le u \le 4$, based on sample S_i, in the same graph. See Figure 8.1 for the case $n = 300$ and $p = n = 300$.

Note that ECDF of Π_{1u} coincide – for all $u > 0$ in Model 1, for $u > 1$ in Model 2 and 3 and for $u > 2$ for Model 4. Hence, $\hat{q}$ is 0, 1, 1 and 2 respectively for Models $1 - 4$. Thus the above method identified q accurately.

We repeated the same process with $n = 500$, $p = n = 500$ and $p = n = 1000$. See Figures 8.2 and 8.3. The following observations emerge.

(i) For each of the above models, the ECDF are nearly identical for $n = 300, 500$ and 1000 i.e. convergence has already occurred at $n = 300$. For smaller values of n, convergence did not occur in our simulation. Some modification may improve the situation for smaller sample sizes. We did not investigate such possibilities.

(ii) For the MA(1) process, LSD of Π_{1u} depends on ψ_1 only through its LSD. Since LSD of A_p and B_p are identical (see Examples 4.2.1 and 4.2.2, both have mass 1 at 0.5), the ECDF for Models 2 and 3 are almost identical.

(iii) As noted above, Theorems 7.2.4 and 7.2.5 are not applicable for Model 4. However, by Theorems 7.3.1 and Examples 4.2.3, 4.2.4, the LSD of any self-adjoint polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$ for Model 4 exists when $p/n \to y > 0$. This is supported numerically by Row 2 right panel of Figures 8.1-8.3.

We repeated the same process for the polynomial Π_{2u} (see Figures 8.4-8.6) and we came to the same conclusions as for Π_{1u}.

Case b: $p/n \to 0$.

For each of the above models, we draw three random samples with $n = 300$, $p = n^{0.9} \sim 170$; $n = 500$, $p = n^{0.9} \sim 269$ and; $n = 1000$, $p = n^{0.9} \sim 502$.

For each sample, we plot ECDF of Π_{3u} (and Π_{4u}), for $1 \leq u \leq 4$, in the same graph. See Figures 8.7-8.12. We have the same conclusions as in Case a.

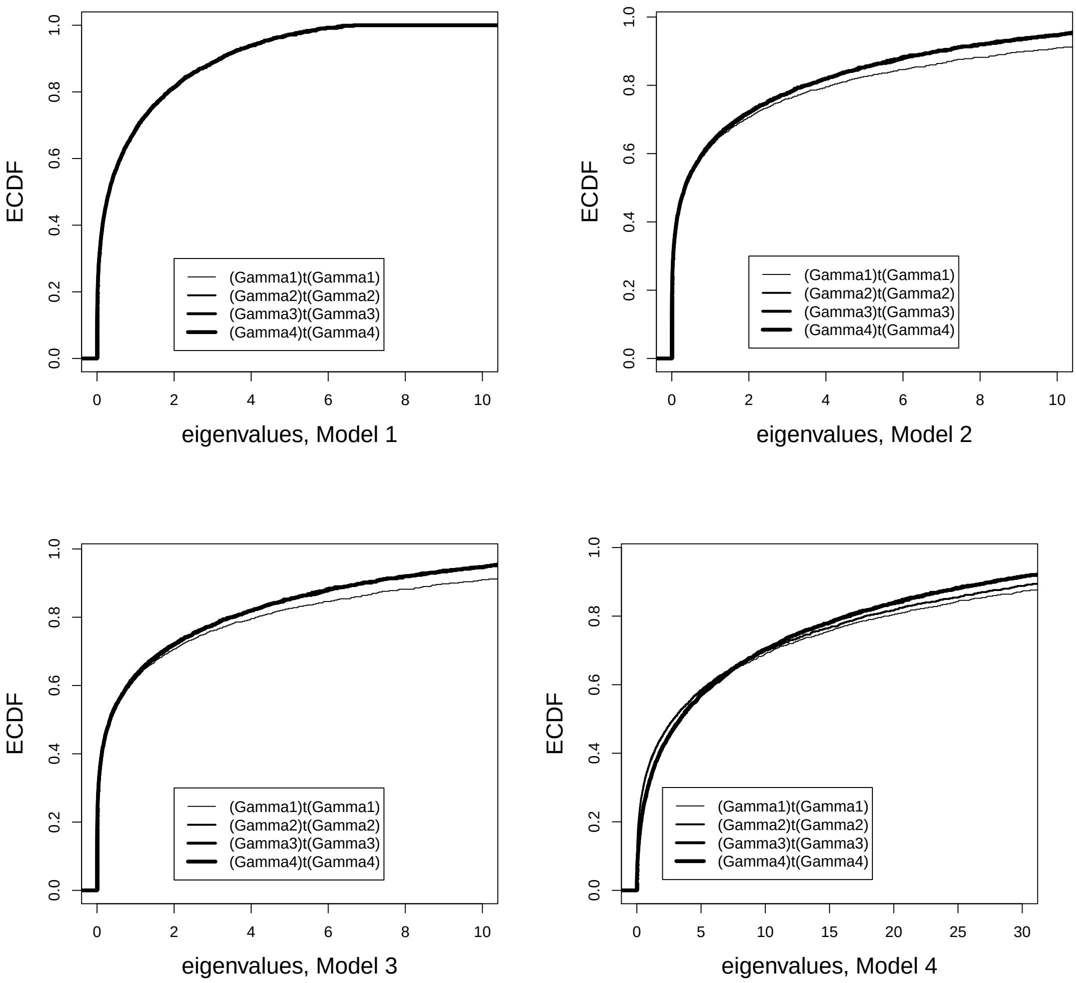


Figure 8.1: ECDF of Π_{1u}, $1 \leq u \leq 4$, $n = p = 300$. For all the figures Gamma u $= \hat{\Gamma}_u$ and t(Gamma u) $= \hat{\Gamma}_u^*$ $\forall u$.

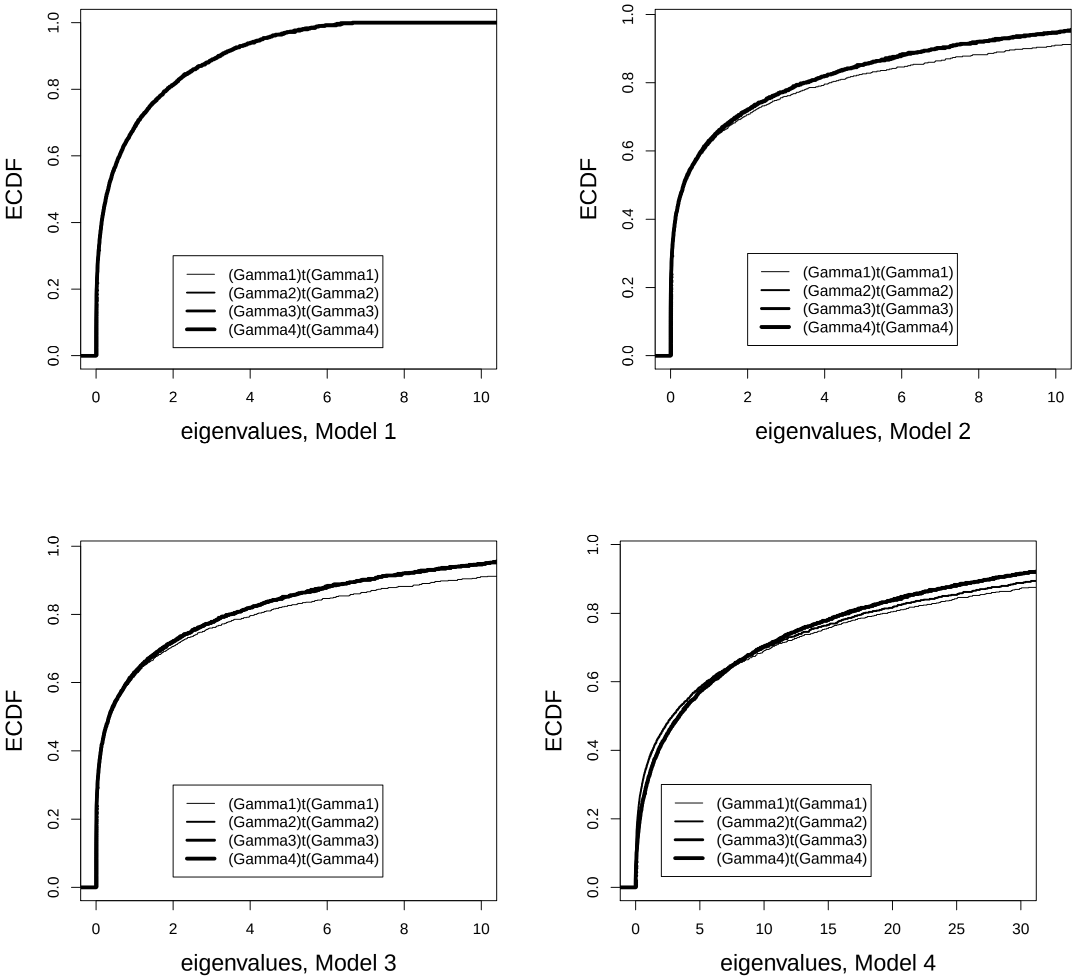


Figure 8.2: ECDF of Π_{1u}, $1 \leq u \leq 4$, $n = p = 500$.

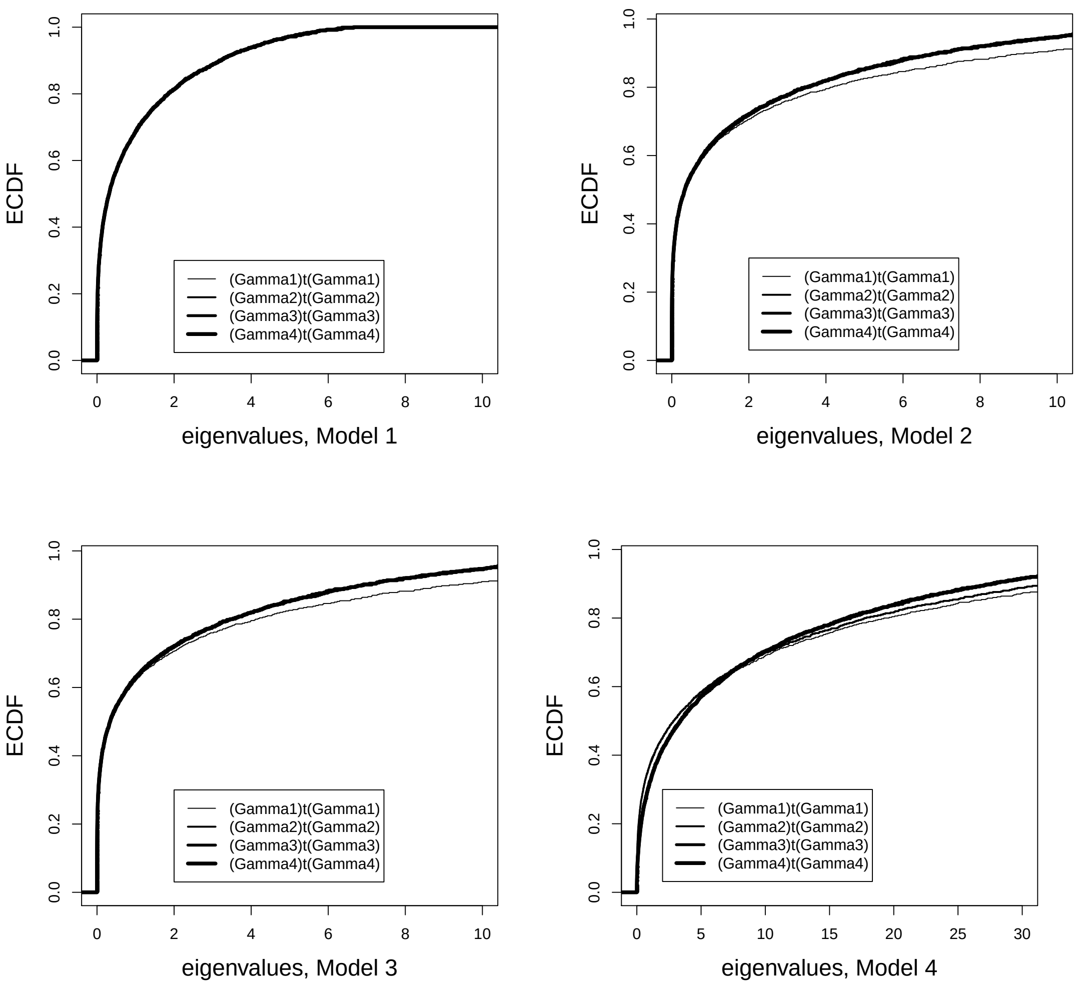


Figure 8.3: ECDF of Π_{1u}, $1 \leq u \leq 4$, $n = p = 1000$.

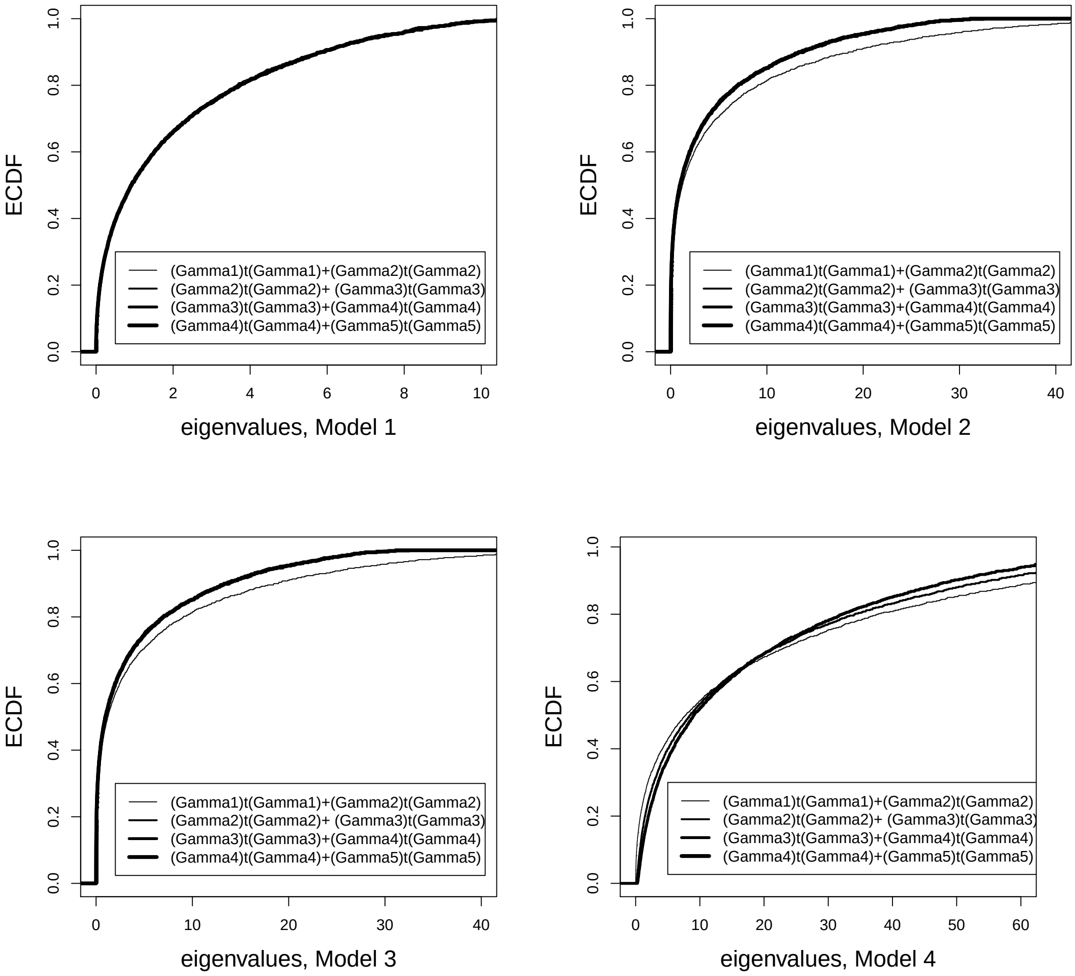


Figure 8.4: ECDF of Π_{2u}, $1 \leq u \leq 4$, $n = p = 300$.

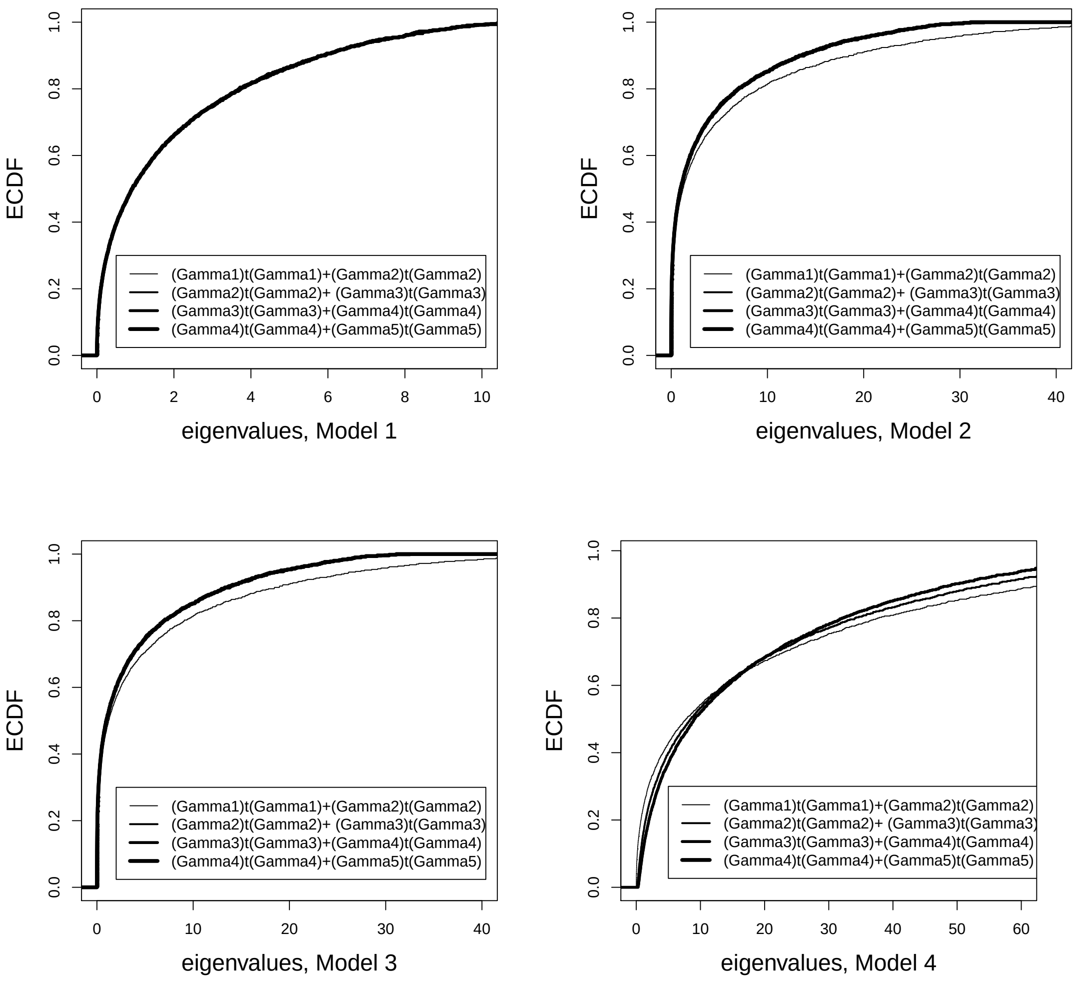


Figure 8.5: ECDF of Π_{2u}, $1 \leq u \leq 4$, $n = p = 500$.

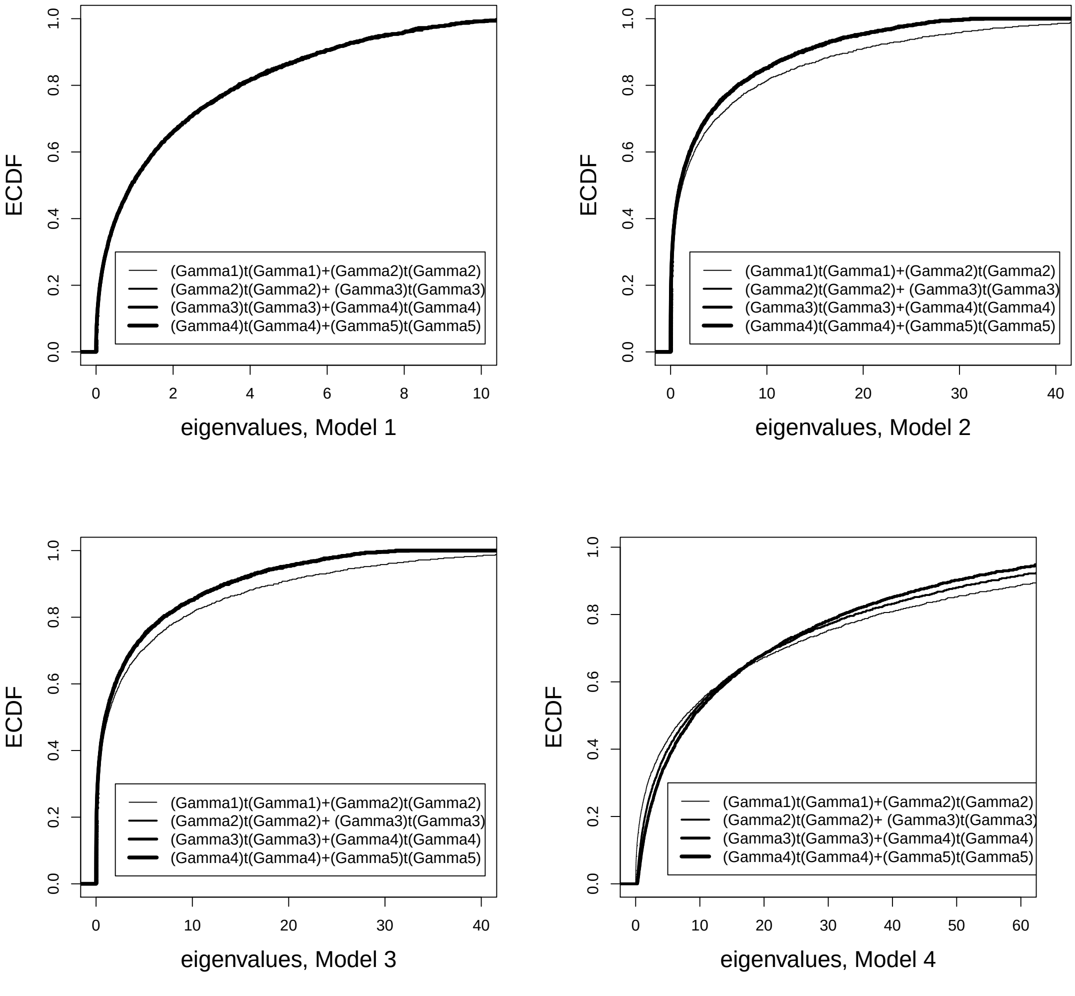


Figure 8.6: ECDF of Π_{2u}, $1 \leq u \leq 4$, $n = p = 1000$.

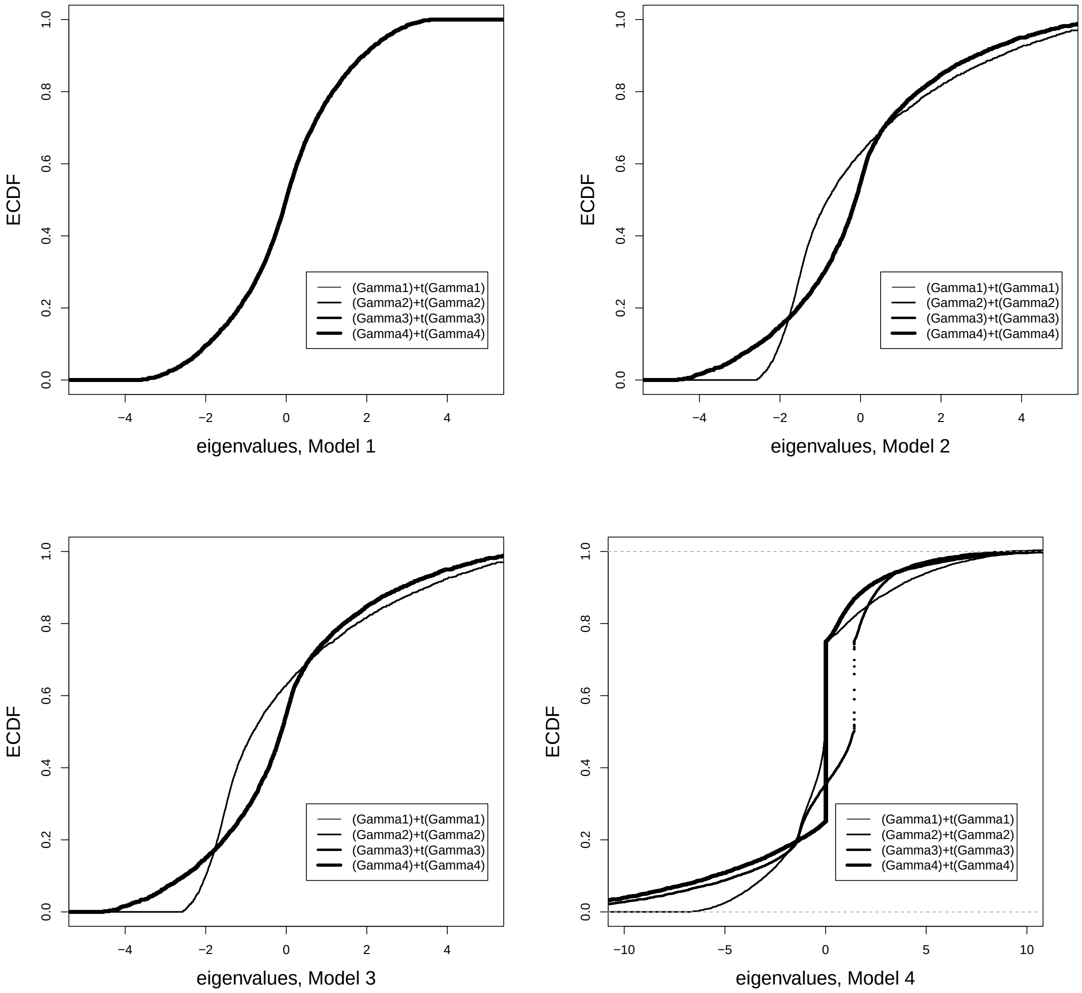


Figure 8.7: ECDF of Π_{3u}, $1 \leq u \leq 4$, $n = 300, p = n^{0.9}$.

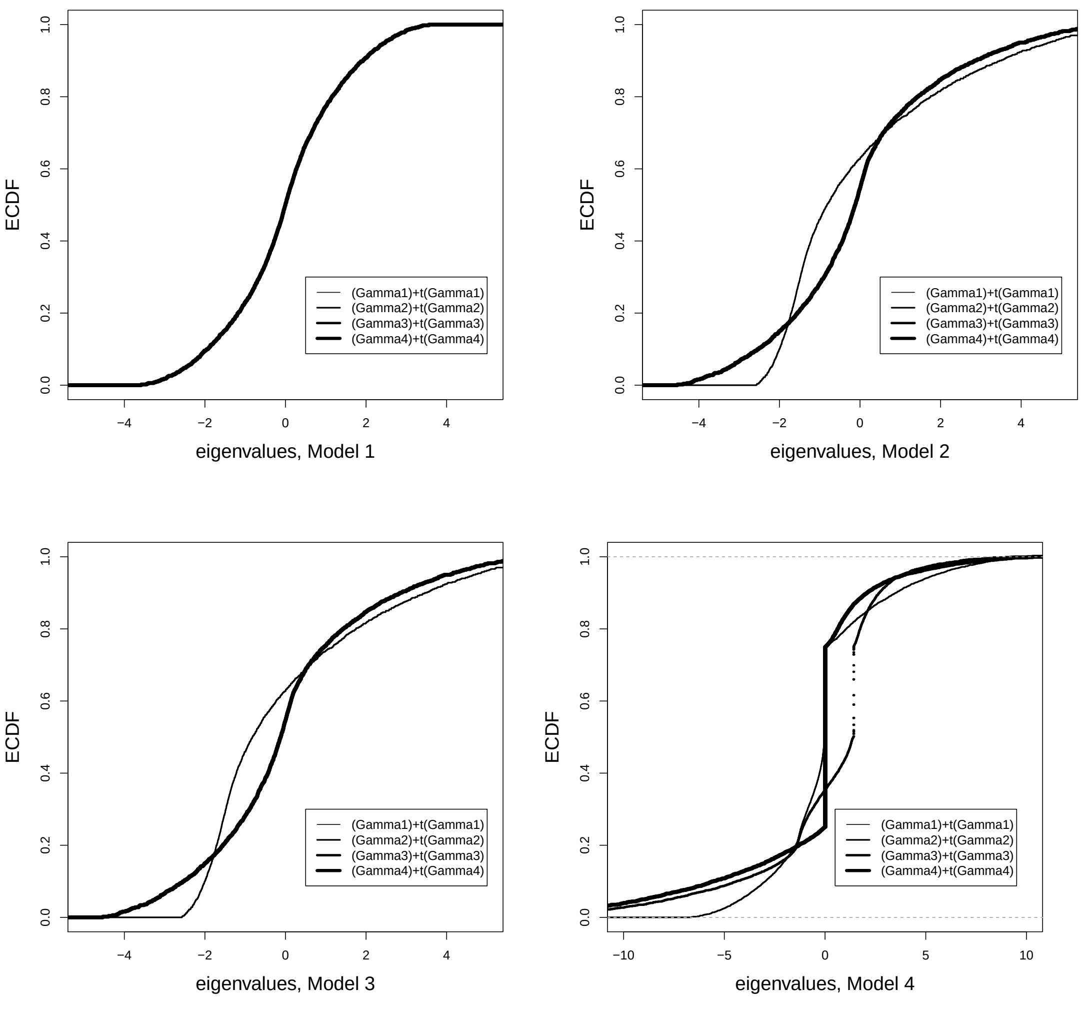


Figure 8.8: ECDF of Π_{3u}, $1 \leq u \leq 4$, $n = 500, p = n^{0.9}$.

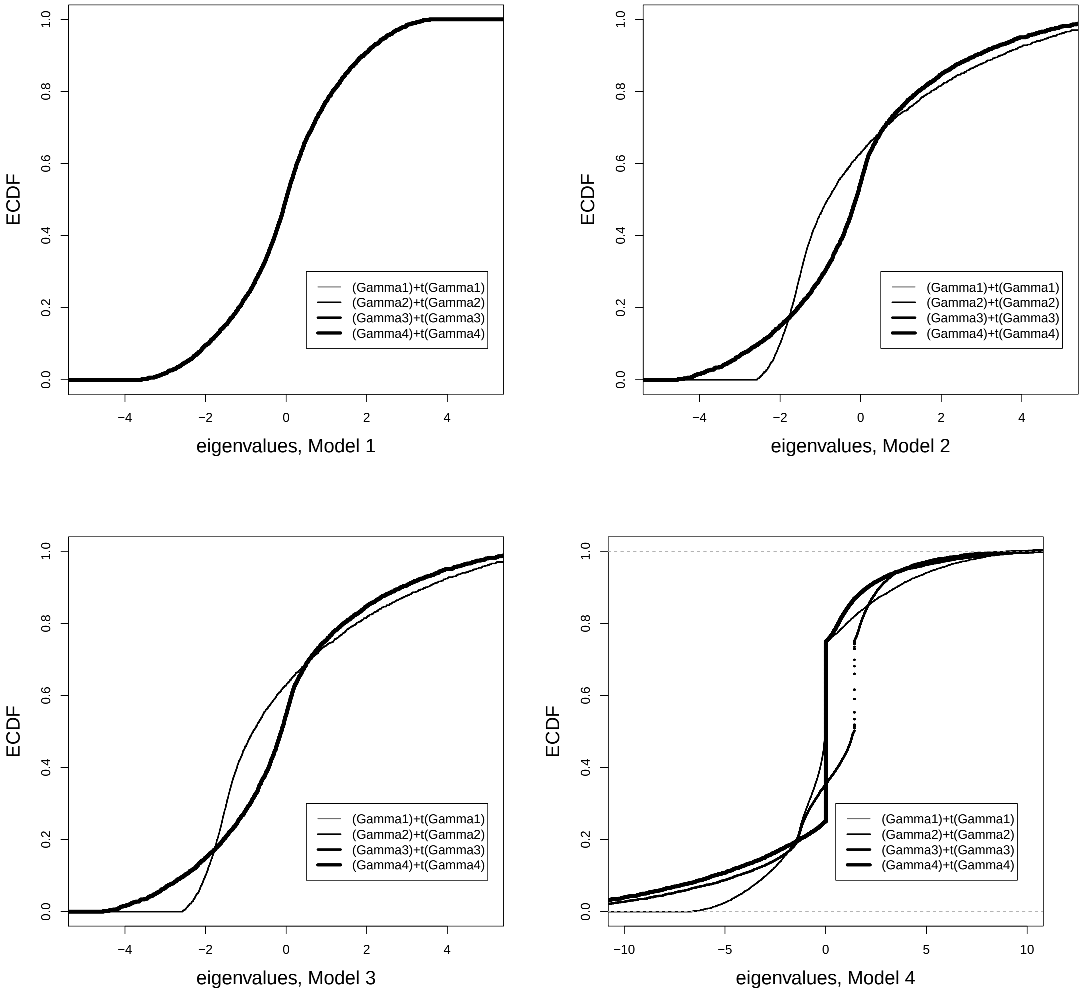


Figure 8.9: ECDF of Π_{3u}, $1 \leq u \leq 4$, $n = 1000, p = n^{0.9}$.

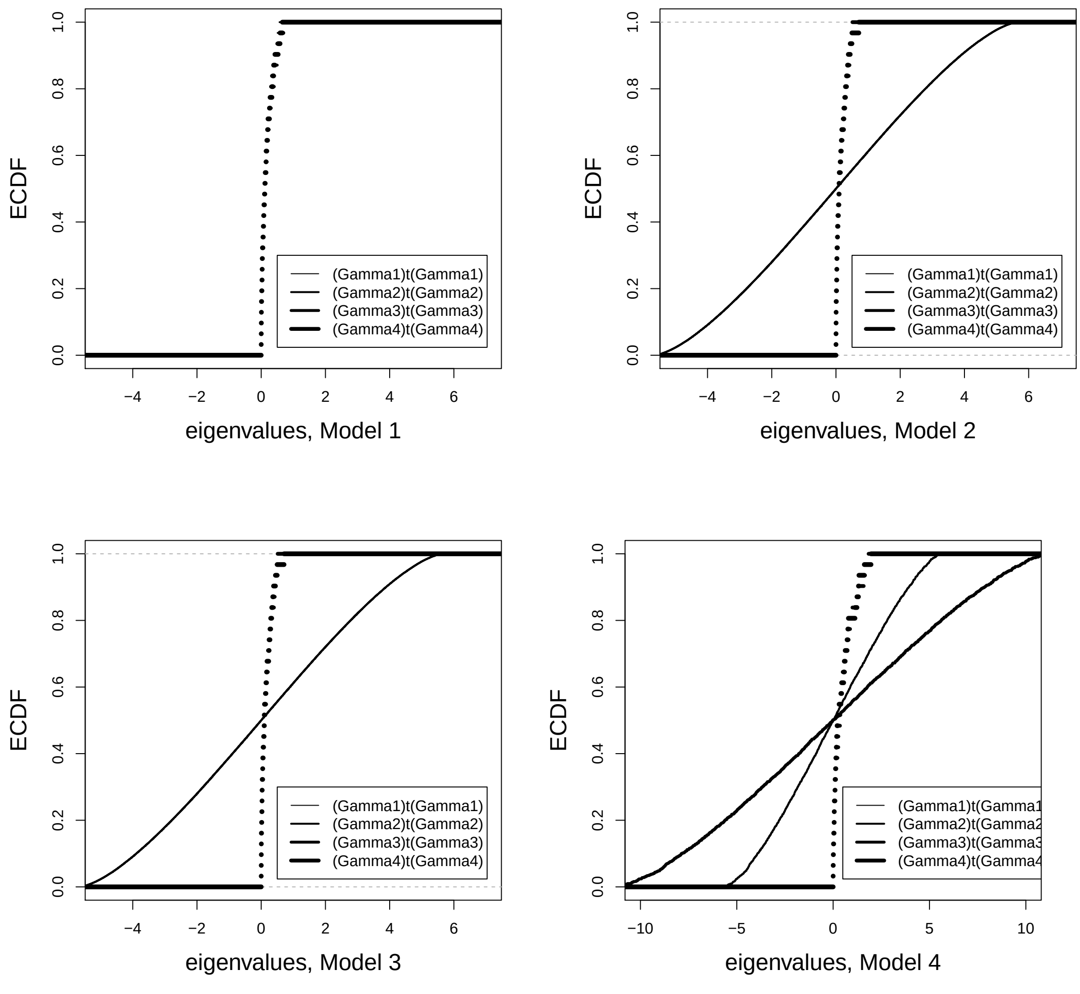


Figure 8.10: ECDF of Π_{4u}, $1 \leq u \leq 4$, $n = 300, p = n^{0.9}$.

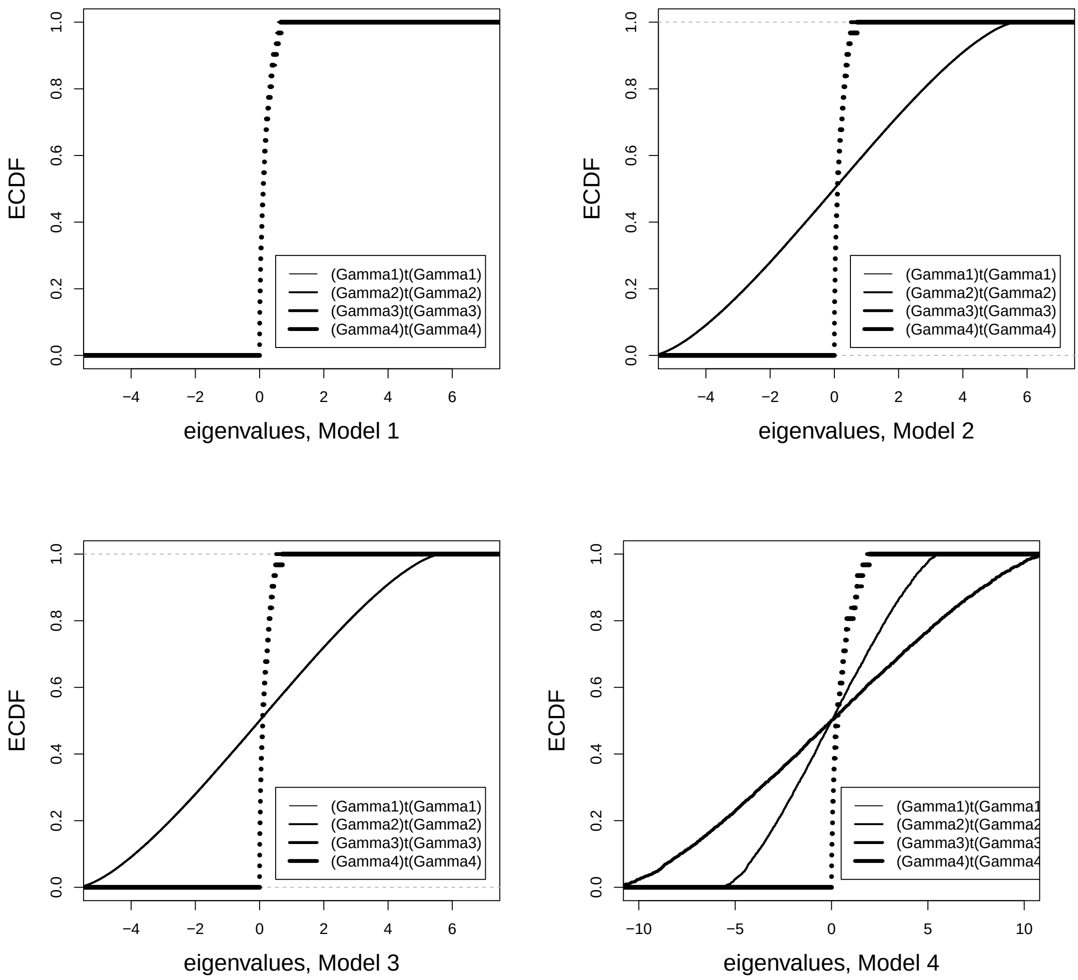


Figure 8.11: ECDF of Π_{4u}, $1 \leq u \leq 4$, $n = 500, p = n^{0.9}$.

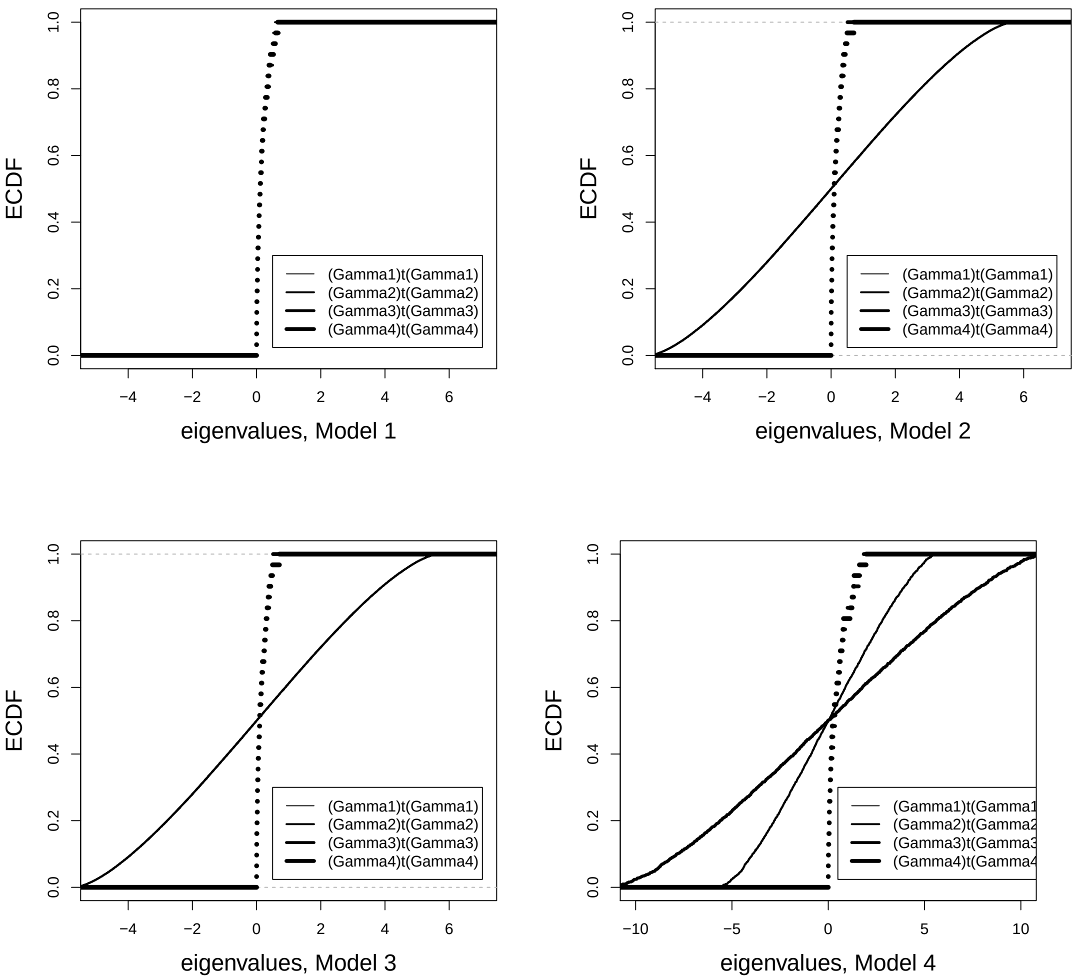


Figure 8.12: ECDF of Π_{4u}, $1 \leq u \leq 4$, $n = 1000, p = n^{0.9}$.

8.3 Order determination of AR processes

Now suppose we have an infinite dimensional AR process with unknown parameter matrices. We wish to estimate the order of the process. Since the parameter matrices are assumed to be unknown, we first obtain consistent estimators for these. Then we use these consistent estimators in conjunction with the ideas of the previous section.

First recall some results from Chapter 3. Consider the IVAR(r) process defined in (3.9):

$$X_{t,p} = \sum_{i=1}^{r} A_{i,p} X_{t-i,p} + \varepsilon_t, \ \forall t, \tag{8.2}$$

where $X_{t,p}$ and ε_t are p-dimensional vectors, $\{\varepsilon_t = (\varepsilon_{t,1}, \varepsilon_{t,2}, \ldots, \varepsilon_{t,p})'\}$ are i.i.d. with mean 0 and variance-covariance matrix I_p and, $\{A_{i,p} : 1 \leq i \leq r\}$ are the $p \times p$ *parameter matrices*. For convenience, we write X_t and A_i respectively for $X_{t,p}$ and $A_{i,p}$. Consider the following assumption on $\{\varepsilon_{i,j}\}$.

(C1) $\varepsilon_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \quad \forall i, j \geq 1.$

For any matrix $M = ((m_{ij}))$, recall the $||\cdot||_{(1,1)}$ norm in (2.26):

$$||M||_{(1,1)} = \sup_j \sum_i |m_{ij}|.$$

Recall the set of all matrices with polynomially decaying corners in (2.20):

$$\mathcal{X}(\alpha, C) \ = \ \{M : T(M,k) \leq Ck^{-\alpha}, \ \forall k \geq 1\}, \ \forall \alpha, C > 0,$$

where for any matrix M, the corner measure $T(M,k)$ is as given in (2.18):

$$T(M,k) = \sup_j \sum_{i:|i-j|>k} |M_{ij}|.$$

Let

$$||A_{i,p}||_{(1,1)} = \theta_{i,n} \text{ and } ||A^*_{i,p}||_{(1,1)} = \theta'_{i,n}, 1 \leq i \leq r. \tag{8.3}$$

Also let $\{\alpha_{i,n} : i = 1, 2, \ldots, r\}$ and $\{\alpha'_{i,n} : i = 1, 2, \ldots, r\}$ respectively be the roots of the following polynomials.

$$\begin{aligned} 1 - \theta_{1,n}z - \theta_{2,n}z^2 \ldots \theta_{r,n}z^r &= 0, \\ 1 - \theta'_{1,n}z - \theta'_{2,n}z^2 \ldots \theta'_{r,n}z^r &= 0. \end{aligned}$$

For each $1 \leq i \leq r$, let $A_{i,\infty}$ be the $\infty \times \infty$ extension of the sequence of matrices $\{A_{i,p(n)}\}_{n\geq 1}$. Consider the parameter space given in (3.30) for $\{A_{i,\infty}\}_{i=1}^r$ as,

$$\mathcal{P}(C, \alpha, \epsilon) = \left\{ \{A_{i,\infty}\}_{i=1}^r : \inf_p \min_{1\leq i\leq r} (|\alpha_{i,p}|, |\alpha'_{i,p}|) > 1 + \epsilon, \text{ and } A_{i,\infty} \in \mathcal{X}(C, \alpha) \ \forall i \right\}.$$

We have the following assumption on the parameter matrices $\{A_i\}$.

(C2) $\{A_{i,\infty}\} \in \mathcal{P}(C, \alpha, \epsilon)$ for some $C, \epsilon, \alpha > 0$.

For any matrix $M = ((m_{ij}))$, recall the k-banded version of M in (2.25):

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

Recall $||\cdot||_2$ in (2.4). Also recall from (2.5) that by consistent estimator $\hat{M}_n$ (based on a sample of size n) of M, we mean

$$||\hat{M}_n - M||_2 \xrightarrow{P} 0, \text{ as } n \to \infty. \tag{8.4}$$

By Theorem 3.5.6, we can say that if (C1) and (C2) hold, then $B_{k_n}(\hat{\Gamma}_u)$ with $k_n = (n^{-1} \log p)^{-1/4}$, is a consistent estimator of Γ_u for each u. Moreover, in Section 3.5.1, we argued that $(B_{k_n}(\hat{\Gamma}_u))^{-1}$ with $k_n = (n^{-1} \log p)^{-1/4}$, is a consistent estimator of Γ_u^{-1} for each u, provided Γ_u^{-1} exists.

Let

$$\mathcal{Y}_r = (\Gamma_1, \Gamma_2, \ldots, \Gamma_r)^*, \quad \mathcal{A}_r = (A_1^*, A_2^*, \ldots, A_r^*)^*.$$

Let G_r be a block matrix with r^2 many $p \times p$ blocks where the (i,j)-th block is given by

$$G_r(i,j) = \Gamma_{|i-j|} I(i < j) + \Gamma^*_{|i-j|} I(i \geq j), \ 1 \leq i, j \leq r. \tag{8.5}$$

Consider the Yule-Walker equation,

$$\mathcal{Y}_r = G_r \mathcal{A}_r. \tag{8.6}$$

Consider the following assumption on $\{A_i\}$.

(C3) Γ_0 is non-singular.

By Lemma 3.5.7, if (C1)-(C3) hold, then

$$\mathcal{A}_r = G_r^{-1} \mathcal{Y}_r \tag{8.7}$$

i.e., each A_i is the finite sum of the finite products of $\{\Gamma_u, \Gamma_u^{-1}, \Gamma_u^*, \Gamma_u^{*-1} \ 1 \leq u \leq r\}$. Moreover, (8.7) provides consistent estimates of A_i, once we replace the population autocovariance matrices $\{\Gamma_u\}$ by their above mentioned consistent estimates $\{B_{k_n}(\hat{\Gamma}_u)\}$ with $k_n = (n^{-1} \log p)^{-1/4}$. Let us denote these estimators of $\{A_i : 1 \leq i \leq r\}$ by $\{\hat{A}_i^{(r)} : 1 \leq i \leq r\}$.

We need the following assumption to guarantee the LSD of symmetric polynomials in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$.

(C4) $\{A_i\}$ converge jointly.

Let

$$\hat{\varepsilon}_t^{(s)} = X_t - \sum_{i=1}^{s} \hat{A}_i^{(s)} X_{t-i}, \quad \forall t, s.$$

Proof of the following theorem is given in Section 8.3.2.

Theorem 8.3.1. *Consider the IVAR*(r) *process defined in (3.9) (or (8.2)). Suppose (C1)-(C4) hold.*

(a) Suppose $p/n \to y > 0$. *Then for each* $u \geq 1$, *the LSD (almost sure) of* Π_{1u} *for the process* $\{\varepsilon_t\}$ *(i.e. for the MA*(0) *process), coincides with the LSD (in probability) of* Π_{1u} *for* $\{\hat{\varepsilon}_t^{(s)}\}$ *if* $s = r$.

(b) Suppose $p/n \to 0$. *Then for each* $u \geq 1$, *the LSD (almost sure) of* Π_{3u} *for the process* $\{\varepsilon_t\}$, *coincides with the LSD (in probability) of* Π_{3u} *for* $\{\hat{\varepsilon}_t^{(s)}\}$ *if* $s = r$.

It may be noted that even though the above theorem is stated for Π_{1u} and Π_{3u}, the conclusion of Theorem 8.3.1 holds true for other polynomials if we are ready to make appropriate moment assumptions on $\{\varepsilon_{i,j}\}$. We have restricted to the above polynomials only for illustrative purposes.

Remark 8.3.2. *Instead of* r, *if we use any other positive integer* $s < r$, *then the residual process* $\{\hat{\varepsilon}_t^{(s)}\}$ *does not behave like the MA*(0) *process. This can be proved rigorously under appropriate assumptions. However in this thesis we limit ourselves to a heuristic idea to show this, as follows. Suppose* $\{X_t\}$ *is an IVAR*(2) *process i.e.*

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + \varepsilon_t, \quad \forall t \tag{8.8}$$

and we fit the IVAR(1) *process using* $\hat{A}_1^{(1)}$. *Let* $\hat{B} = A_1 - \hat{A}_1^{(1)}$. *Therefore,*

$$\hat{\varepsilon}_t^{(1)} = X_t - \hat{A}_1^{(1)} X_{t-1} = \hat{B} X_{t-1} + A_2 X_{t-2} + \varepsilon_t. \tag{8.9}$$

Let $B = A_1 - \Gamma_0^{-1}\Gamma_1$. *Using the fact that* $||\hat{A}_1^{(1)} - \Gamma_0^{-1}\Gamma_1||_2 \xrightarrow{P} 0$ *(by Theorem 3.5.6), it is easy to see that the LSD of* Π_{1u} *(for* $p/n \to y > 0$*) and* Π_{3u} *(for* $p/n \to 0$*) for the process* $\{\hat{\varepsilon}_t^{(1)}\}$ *coincides with the corresponding LSD (in probability) for*

$$\tilde{\varepsilon}_t^{(1)} = B X_{t-1} + A_2 X_{t-2} + \varepsilon_t. \tag{8.10}$$

Note that under (C2), by Theorem 3.4.6, $\{X_t\}$ *can be expressed as*

$$X_t = \varepsilon_t + \sum_{j=1}^{\infty} \phi_j \varepsilon_{t-j}, \ \forall t, \tag{8.11}$$

where $\{\phi_j\}$ *are functions of* A_2 *and* B*. Therefore,*

$$\tilde{\varepsilon}_t^{(1)} = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}, \quad where \ \ \theta_0 = I_p, \ \theta_1 = B, \ \theta_{j+2} = B\phi_{j+1} + A_2\phi_j, \ j \geq 2. \tag{8.12}$$

Note that $\{\tilde{\varepsilon}_t^{(1)}\}$ *is an MA*(∞) *process. Then using similar idea as in the proofs of Corollaries 7.3.12 (d) and 7.3.18 (d),* ***if*** $\{\theta_j\}$ *are norm bounded and converge jointly, it is easy to prove that the LSD of* Π_{1u} *(for* $p/n \to y > 0$*) and* Π_{3u} *(for* $p/n \to 0$*) for the process* $\{\varepsilon_t\}$ *do not coincide with the corresponding LSD (in probability) for* $\{\tilde{\varepsilon}_t^{(1)}\}$*.*

Under suitable conditions, $\{A_1, A_2\}$ *and* $\{\theta_j\}$ *do indeed converge jointly. This needs more work and we did not pursue it in this thesis.*

Therefore, for each $u \geq 0$, the LSD of Π_{1u} (for $p/n \to y > 0$) and Π_{3u} (for $p/n \to 0$) for the process $\{\varepsilon_t\}$ coincides with the LSD (in probability) for $\{\hat{\varepsilon}_t^{(r)}\}$. Instead of r, if we use any other positive integer $s < r$, then the residual process $\{\hat{\varepsilon}_t^{(s)}\}$ does not behave like the MA(0) process. As by Theorem 8.2.1, ECDF of Π_{1u} (for $p/n \to y > 0$) or Π_{3u} (for $p/n \to 0$) for $u = 1, 2$ coincide (almost surely) under MA(0) process, to determine the order of the IVAR process, it is enough to check whether the ECDF of Π_{1u} (for $p/n \to y > 0$) or Π_{3u} (for $p/n \to 0$) of $\{\hat{\varepsilon}_t^{(r)}\}$ for $u = 1, 2$ coincide or not. Therefore, if we plot the ECDF of Π_{1u} (for $p/n \to y > 0$) or Π_{3u} (for $p/n \to 0$), $u = 1, 2$ for the residual process $\{\hat{\varepsilon}_t^{(s)}\}$ in the same graph, the two distribution functions will coincide only when $s = r$. Hence, we have the following method.

Identification of unknown order r. Successively fit an IVAR(s) process for $s = 0, 1, 2, \ldots$ and for each s, plot the ECDF of Π_{1u} (for $p/n \to y > 0$) or Π_{3u} (for $p/n \to 0$), $u = 1, 2$ for residuals $\{\hat{\varepsilon}_t^{(s)}\}$ in the same graph.

We say that $\hat{r}$ is an estimate of the unknown order r of the IVAR process, if the ECDF of Π_{1u} (for $p/n \to y > 0$) or Π_{3u} (for $p/n \to 0$), $u = 1, 2$ do not coincide for all $s < \hat{r}$ and coincide for $s = \hat{r}$.

Note that the polynomial Π_{2u} and Π_{4u} could also be used instead Π_{1u} and Π_{3u}.

8.3.1 Simulations

Consider the following IVAR processes. Suppose $\{\varepsilon_t\}$ is as in (8.1).

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

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

Assuming that we do not know the parameter matrices, we use their consistent estimators discussed above.

For Model 5, we plot the ECDF of Π_{1u} (or Π_{3u}), $u = 1, 2$ for the residual process $\{\hat{\varepsilon}_t^{(1)}\}$ and for $n = p = 500$ (or $n = 500$, $p = n^{0.9} = 269$) in the same graph and observe that they coincide. See Row 1 left panel in Figure 8.14 (or Figure 8.13). Therefore, 1 is an estimate of the order of Model 5.

For Model 6, we do the same but the two ECDF do not coincide (see Row 1 right panel in Figures 8.14 and 8.13). In Row 2 of Figures 8.14 and 8.13, the same two ECDF are plotted for $\{\hat{\varepsilon}_t^{(2)}\}$ and they coincide and hence 2 is an estimate of the order for Model 6.

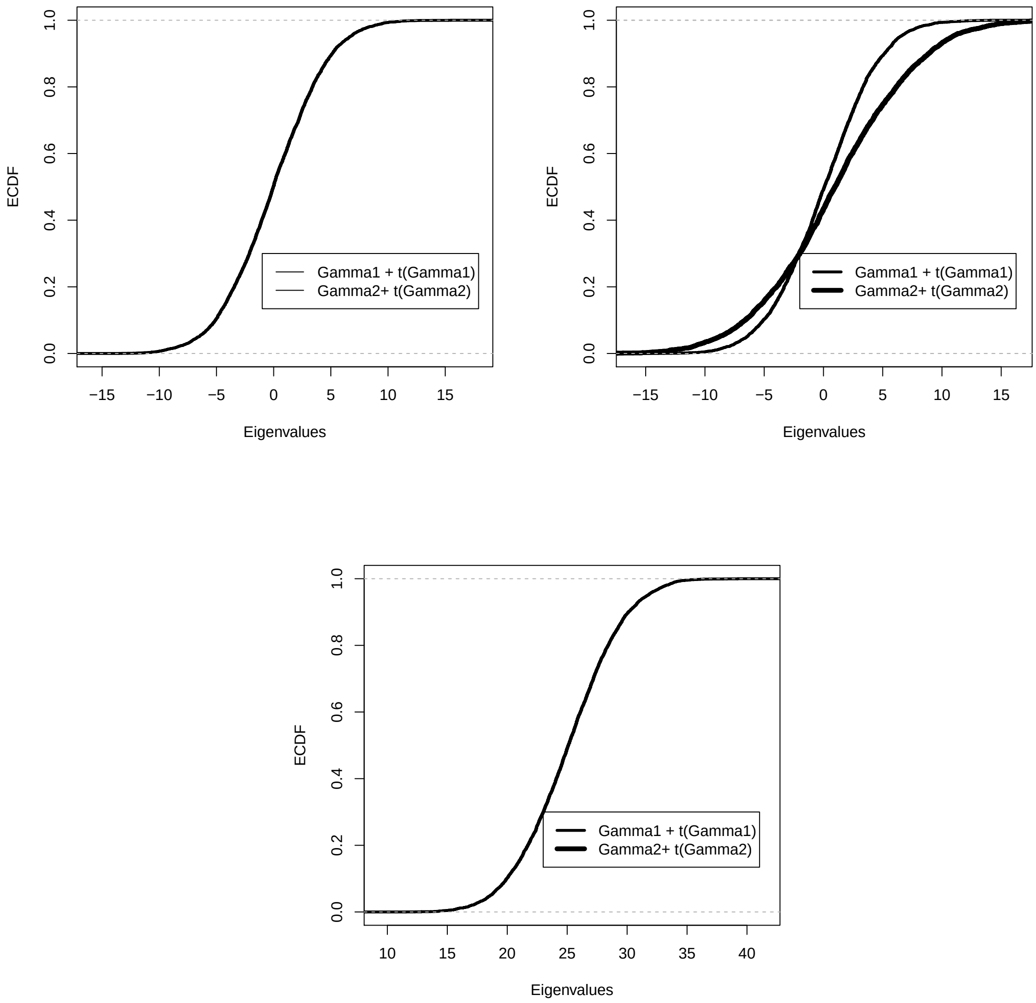


Figure 8.13: $n = 500$, $p = n^{0.9}$. Row 1 left: ECDF of Π_{31} and Π_{32} for residuals after fitting AR(1) in Model 5. Row 1 right: same after fitting AR(1) in Model 6. Row 2: same after fitting AR(2) in Model 6.

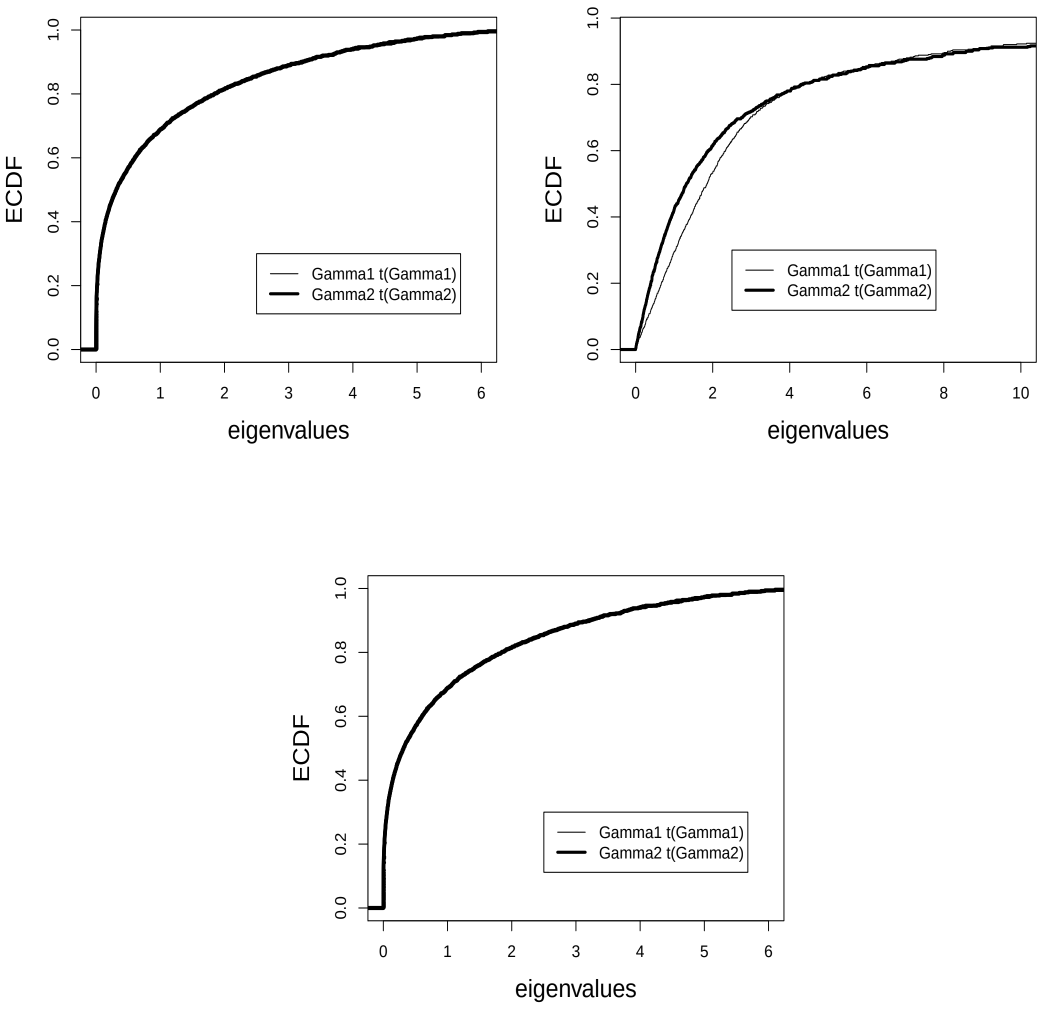


Figure 8.14: $n = p = 500$. Row 1 left: ECDF of Π_{11} and Π_{12} for residuals after fitting AR(1) in Model 5. Row 1 right: same after fitting AR(1) in Model 6. Row 2: same after fitting AR(2) in Model 6.

8.3.2 Proof of Theorem 8.3.1

For simplicity, we consider only the IVAR(1) process. Let $\{X_t : 1 \le t \le n\}$ be a sample of size n from the IVAR(1) process satisfying

$$X_{t+1} = \varepsilon_t + AX_{t-1}, \ \forall t, \tag{8.13}$$

where $\varepsilon_t \sim IID(0, I)$. Recall $||\cdot||_2$ in (2.4). By Theorem 3.5.8, under (C2), the consistent estimator $\hat{A}$ of A satisfies

$$||\hat{A} - A|| \xrightarrow{P} 0. \tag{8.14}$$

Consider the process $\{\hat{\varepsilon}_t^{(1)} = X_t - \hat{A}X_{t-1} = \varepsilon_t + (A - \hat{A})X_{t-1}\}$. Let, for all $k \ge 0$,

$$\begin{aligned}
B_k &= \big(n^{-1}\sum_{t=1}^{n-k} \varepsilon_t^{(1)}\varepsilon_{t+k}^{(1)*}\big)\big(n^{-1}\sum_{t=1}^{n-k} \varepsilon_{t+k}^{(1)}\varepsilon_t^{(1)*}\big),\\
D_k &= \big(n^{-1}\sum_{t=1}^{n-k} \varepsilon_t\varepsilon_{t+k}^{*}\big)\big(n^{-1}\sum_{t=1}^{n-k} \varepsilon_{t+k}\varepsilon_t^{*}\big),\\
E_k &= \big(n^{-1}\sum_{t=1}^{n-k} \hat{\varepsilon}_t^{(1)}\hat{\varepsilon}_{t+k}^{(1)*}\big) + \big(n^{-1}\sum_{t=1}^{n-k} \hat{\varepsilon}_{t+k}^{(1)}\hat{\varepsilon}_t^{(1)*}\big),\\
F_k &= \big(n^{-1}\sum_{t=1}^{n-k} \varepsilon_t\varepsilon_{t+k}^{*}\big) + \big(n^{-1}\sum_{t=1}^{n-k} \varepsilon_{t+k}\varepsilon_t^{*}\big).
\end{aligned}$$

In Theorem 8.3.1 (a) and (b), our two claims are respectively that, for any fixed $k \ge 0$, the LSD of (a) B_k and D_k, (b) E_k and F_k, are identical in probability. To prove (a) and (b), By Corollary $A.41$ in Bai and Silverstein [2009], it is respectively enough to show that

$$\begin{aligned}
n^{-1}\mathrm{Tr}(B_k - D_k)^2 &\xrightarrow{P} 0, \quad \text{as } p/n \to y > 0 \text{ and} &(8.15)\\
np^{-2}\mathrm{Tr}(E_k - F_k)^2 &\xrightarrow{P} 0, \quad \text{as } p/n \to 0. &(8.16)
\end{aligned}$$

Proof of (8.15). Note that

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n-k}\varepsilon_t^{(1)}\varepsilon_{t+k}^{(1)*}\\
=&\left(\frac{1}{n}\sum_{t=1}^{n-k}\varepsilon_t\varepsilon_{t+k}^*\right)+(A-\hat{A})\left(\frac{1}{n}\sum_{t=1}^{n-k}X_{t-1}\varepsilon_{t+k}^*\right)+\left(\frac{1}{n}\sum_{t=1}^{n-k}\varepsilon_tX_{t+k-1}^*\right)(A-\hat{A})^*\\
&+(A-\hat{A})\left(\frac{1}{n}\sum_{t=1}^{n-k}X_{t-1}X_{t+k-1}^*\right)(A-\hat{A})^*\\
=&\left(\frac{1}{n}\sum_{t=1}^{n-k}\varepsilon_t\varepsilon_{t+k}^*\right)+(A-\hat{A})\left(\frac{1}{n}\sum_{t=1}^{n-k}X_{t-1}X_{t+k}^*\right)-(A-\hat{A})\left(\frac{1}{n}\sum_{t=1}^{n-k}X_{t-1}X_{t+k-1}^*\right)A^*\\
&+\left(\frac{1}{n}\sum_{t=1}^{n-k}X_tX_{t+k-1}^*\right)(A-\hat{A})^*-A\left(\frac{1}{n}\sum_{t=1}^{n-k}X_{t-1}X_{t+k-1}^*\right)(A-\hat{A})^*\\
&+(A-\hat{A})\left(\frac{1}{n}\sum_{t=1}^{n-k}X_{t-1}X_{t+k-1}^*\right)(A-\hat{A})^*\\
=&\ G_1+G_2+G_3+G_4+G_5+G_6,\ \text{(say)}.
\end{aligned}
\tag{8.17}
$$

Therefore, $B_k=\sum_{j,l=1}^{6}G_jG_l^*$. Note that $D_k=G_1G_1^*$. Hence, by Hölder's inequality,

$$
\begin{aligned}
n^{-1}\mathrm{Tr}(B_k-D_k)^2 &= \sum_{\substack{1\le j_1,j_2,l_1,l_2\le 6\\(j_1,l_1),(j_2,l_2)\ne(1,1)}} n^{-1}\mathrm{Tr}(G_{j_1}G_{l_1}^*G_{j_2}G_{l_2}^*),\\
&\le \sum_{\substack{1\le j_1,j_2,l_1,l_2\le 6\\(j_1,l_1),(j_2,l_2)\ne(1,1)}}\left(\prod_{s=1}^{2}\frac{1}{n}\mathrm{Tr}(G_{j_s}^*G_{j_s})^2\frac{1}{n}\mathrm{Tr}(G_{l_s}^*G_{l_s})^2\right)^{1/4}.
\end{aligned}
\tag{8.18}
$$

Therefore, to show (8.15), it is enough to prove

(i) $\frac{1}{n}\mathrm{Tr}(G_1^*G_1)^2=O_P(1)$ and

(ii) $\frac{1}{n}\mathrm{Tr}(G_i^*G_i)^2=o_P(1),\ \forall i=2,3,4,5,6.$

To prove (i) and (ii), we need the following lemma.

Lemma 8.3.3. *Consider MA(q) or MA(∞) processes respectively defined in (3.7) and (3.2). Suppose (C1) and (C4) hold and $p/n\to y>0$. Then for any symmetric* *polynomial* Π,

$$\frac{1}{n}\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)) \overset{a.s.}{\to} \lim E\frac{1}{n}\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)), \tag{8.19}$$

and hence

$$\frac{1}{n}\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)) = O_P(1).$$

Proof. By (M1) and (M4) in the proof of Theorems 7.3.1 and 7.3.4, under (C1) and (C4) and as $p/n \to y > 0$, we have

$$\begin{aligned}
&\frac{1}{n}E\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)) \to \lim E\frac{1}{n}\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)) \\
&E(n^{-1}\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)) - En^{-1}\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)))^4 = O(n^{-4}).
\end{aligned}$$

Hence, by Borel-Cantelli lemma, the first part of Lemma 8.3.3 follows. The second part is trivial. □

Note that G_1 is $\hat{\Gamma}_k$ for the MA(0) process $\{\varepsilon_t\}$. Therefore, (i) follows immediately by Lemma 8.3.3.

We now prove (ii), first for $i = 6$. Recall $||\cdot||_2$ in (2.4). Note that for any $n \times n$ matrix A,

$$n^{-1}\mathrm{Tr}(A^*A) \leq ||A||_2^2. \tag{8.20}$$

Note that $G_6 = (A - \hat{A})\hat{\Gamma}_k(A - \hat{A})^*$. Therefore, by Hölder's inequality,

$$\begin{aligned}
\frac{1}{n}\mathrm{Tr}(G_6^*G_6)^2 &\leq \left(\frac{1}{n}\mathrm{Tr}((A-\hat{A})(A-\hat{A})^*)^8 \frac{1}{n}\mathrm{Tr}(\hat{\Gamma}_k^*\hat{\Gamma}_k)^8\right)^{1/2} \\
&\leq ||A - \hat{A}||_2^8 \left(\frac{1}{n}\mathrm{Tr}(\hat{\Gamma}_k^*\hat{\Gamma}_k)^8\right)^{1/2}, \quad \text{by (8.25)}. \qquad (8.21)
\end{aligned}$$

Now note that by (8.14), $||A - \hat{A}||_2 = o_P(1)$ and by Lemma 8.3.3, $\frac{1}{n}\mathrm{Tr}(\hat{\Gamma}_k^*\hat{\Gamma}_k)^8 =$

$O_P(1)$. Hence, $\frac{1}{n}\text{Tr}(G_6^*G_6)^2 = o_P(1)$. therefore, (ii) is proved for $i = 6$.

Similar arguments will go through for $i = 2, 3, 4, 5$. Hence the proof of (8.16) is complete. Therefore, (8.15) is proved and hence (a) is proved for the IVAR(1) process. To prove (b), it remains to prove (8.16).

Proof of (8.16). Note that by (8.17), $E_k = \sum_{j=1}^{6} G_j + \sum_{j=1}^{6} G_j^*$ and $F_k = G_1 + G_1^*$. Hence,

$$\frac{n}{p^2}\text{Tr}(E_k - F_k)^2 = \sum_{j,l=2}^{6} \frac{n}{p^2}\text{Tr}(G_j + G_j^*)(G_l + G_l^*). \tag{8.22}$$

By Hölder's inequality for all $2 \leq j, l \leq 6$

$$|\frac{n}{p^2}\text{Tr}(G_j + G_j^*)(G_l + G_l^*)| \leq \frac{4n}{p}\big(p^{-1}\text{Tr}(G_j^*G_j)p^{-1}\text{Tr}(G_l^*G_l)\big)^{1/2}. \tag{8.23}$$

Now the proof of (8.16) for IVAR(1) is completed using similar arguments as in the proof of (a) for IVAR(1). This establishes (b).

One can deal with the IVAR(r), $r \geq 2$, using the exactly same idea as above.

8.4 Asymptotic normality of traces

Linear spectral statistics of a random matrix M are of the form $\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i)$ where $\{\lambda_i\}$ are eigenvalues of M and f is a "suitable" function. Such statistics have been discussed in Diaconis and Evans [2001], Bai and Silverstein [2004] and Bai et al. [2009]. Asymptotic normality of these statistics is extremely useful in statistical inference. While we do not discuss these statistics in general, in this thesis, we deal with a specific class of spectral linear statistics namely traces of polynomials in sample autocovariance matrices.

Consider the infinite dimensional MA process in (3.7). In this section, we shall discuss asymptotic normality of the trace of any symmetric polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^*\}$. We will then show how this can be applied in statistical testing problems in high-dimensional time series.

Recall the assumptions (B1), (B3) and (B) which appeared in Chapter 7. For convenience of the reader, we state them again. Recall the classes $\mathcal{L}$ and $C(\delta, p)$ respectively in (4.14) and (4.16).

(B1) $\{\varepsilon_{i,j}\}$ are independently distributed with mean 0 and variance 1.

(B2) $\{\varepsilon_{i,j} : 1 \leq i \leq p,\ 1 \leq j \leq n\} \in \mathcal{L} \cap \mathcal{C}(\delta, p)$, for all $p \geq 1$.

(B) $\{\psi_j, \psi_j^*\}$ are norm bounded and jointly converge.

Let $\Pi := \Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)$ be a symmetric polynomial in $\{\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0\}$ and

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

Let

$$\sigma_\Pi^2 = \lim E(\mathrm{Tr}(\Pi) - E\mathrm{Tr}(\Pi))^2, \quad \sigma_R^2 = E(\mathrm{Tr}(R_\Pi))^2.$$

Note that σ_Π^2 and σ_R^2 are finite under our assumptions. Then we have the following theorem

Theorem 8.4.1. *Suppose (B1), (B3), (B) hold.*

(a) Suppose $p/n \to y > 0$. Then $\mathrm{Tr}(\Pi) - E\mathrm{Tr}(\Pi) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_\Pi^2)$.

(b) Suppose $p/n \to 0$. Then $\mathrm{Tr}(R_\Pi) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_R^2)$.

Proof. (a) We use Lemmas 5.4.2 and 7.3.2. In these lemmas put

$$\begin{aligned} \mathcal{P}_i &= \mathrm{Tr}(\Pi(\Delta_u, \Delta_u^* : u \geq 0)),\ \forall i \geq 1 \text{ and} && (8.24)\\ \tilde{\mathcal{P}}_i &= \mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)),\ \forall i \geq 1. && (8.25) \end{aligned}$$

Therefore,

$$\begin{aligned} \mathcal{P}_i^0 &= E\mathrm{Tr}(\Pi(\Delta_u, \Delta_u^* : u \geq 0)),\ \forall i \geq 1 \text{ and} && (8.26)\\ \tilde{\mathcal{P}}_i^0 &= E\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)),\ \forall i \geq 1. && (8.27) \end{aligned}$$

Also note that, by (8.25) and (8.27), we have

$$\lim E\big[\big(\tilde{\mathcal{P}}_i - \tilde{\mathcal{P}}_i^0\big)\big(\tilde{\mathcal{P}}_j - \tilde{\mathcal{P}}_j^0\big)\big] = \sigma_\Pi^2,\ \forall i,j \geq 1. \tag{8.28}$$

Therefore,

$$\begin{aligned}
& \lim E(\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)) - E\mathrm{Tr}(\Pi(\hat{\Gamma}_u, \hat{\Gamma}_u^* : u \geq 0)))^T \\
= \ & \lim E(\prod_{i=1}^{T}(\tilde{\mathcal{P}}_i - \tilde{\mathcal{P}}_i^0)),\ \text{(by (8.25) and (8.27)} \\
= \ & \lim E(\prod_{i=1}^{T}(\mathcal{P}_i - \mathcal{P}_i^0)),\ \text{(by Lemma 7.3.2(b))} \\
= \ & \begin{cases} 0 \text{ if } T = 2d-1, \\ \sum_{\mathcal{S}_d}\prod_{k=1}^{d}\lim E\big[\big(\mathcal{P}_{i_{2k-1}} - \mathcal{P}_{i_{2k-1}}^0\big)\big(\mathcal{P}_{i_{2k}} - \mathcal{P}_{i_{2k}}^0\big)\big],\ \text{if } T = 2d. \end{cases},\quad \text{(by Lemma 5.4.2)} \\
= \ & \begin{cases} 0 \text{ if } T = 2d-1, \\ \sum_{\mathcal{S}_d}\prod_{k=1}^{d}\lim E\big[\big(\tilde{\mathcal{P}}_{i_{2k-1}} - \tilde{\mathcal{P}}_{i_{2k-1}}^0\big)\big(\tilde{\mathcal{P}}_{i_{2k}} - \tilde{\mathcal{P}}_{i_{2k}}^0\big)\big],\ \text{if } T = 2d. \end{cases},\quad \text{(by Lemma 7.3.2 (b))} \\
= \ & \begin{cases} 0, \quad \text{if } T = 2d-1 \\ (\#\mathcal{S}_d)\sigma_\Pi^{2d}, \quad \text{if } T = 2d \end{cases},\quad \text{(by (8.28))} \\
= \ & \begin{cases} 0, \quad \text{if } T = 2d-1 \\ (\text{total number of pair partitions of } \{1, 2, \ldots, 2d\})\sigma_\Pi^{2d}, \quad \text{if } T = 2d. \end{cases}
\end{aligned} \tag{8.29}$$

which is nothing but the T-th order raw moment of $\mathcal{N}(0, \sigma_\Pi^2)$. This completes the proof of (a).

(b) Similar arguments as in (a), works for (b) also. Hence we omit its proof. □

Following are some examples and simulations to support Theorem 8.4.1.

Example 8.4.1. *Let $X_t = \varepsilon_t$, $\forall t$. We consider $\varepsilon_t \sim \mathcal{N}_p(0, I_p)$, where ε_t's are independent. Then using Theorem 8.4.1 (a), it is easy to see that, when $p/n \to 1$,*

$$\begin{aligned}
\mathrm{Tr}(\hat{\Gamma}_0) - n &\stackrel{\mathcal{D}}{\to} \mathcal{N}(0, 2), \\
\mathrm{Tr}(\hat{\Gamma}_1\hat{\Gamma}_1^*) - n + 1 &\stackrel{\mathcal{D}}{\to} \mathcal{N}(0, 10), \\
\mathrm{Tr}(\hat{\Gamma}_1 + \hat{\Gamma}_1^*) &\stackrel{\mathcal{D}}{\to} \mathcal{N}(0, 4).
\end{aligned} \tag{8.30}$$

Moreover, using Theorem 8.4.1 (b), it is easy to see that, when $p/n \to 0$, we have

$$\begin{aligned}
\mathrm{Tr}\sqrt{np^{-1}}(\hat{\Gamma}_0 - I) &\stackrel{\mathcal{D}}{\to} \mathcal{N}(0, 1), \\
\mathrm{Tr}\sqrt{np^{-1}}(\hat{\Gamma}_1 + \hat{\Gamma}_1^*) &\stackrel{\mathcal{D}}{\to} \mathcal{N}(0, 2).
\end{aligned}$$

Simulation results given in Row 1 (left and right panels, Figure 8.15), Row 2 (left panel Figure 8.15) and Row 1 (left and right panels, Figure 8.16) support the above convergences.

Example 8.4.2. *Let $X_t = \varepsilon_t + \varepsilon_{t-1}$. Then using Theorem 8.4.1 (a), it is easy to see that, when $p/n \to 1$, we have*

$$\mathrm{Tr}(\hat{\Gamma}_0) - 2(n-1) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, 8).$$

Moreover, using Theorem 8.4.1 (b), it is easy to see that, when $p/n \to 0$, we have

$$\begin{aligned}
\mathrm{Tr}\sqrt{np^{-1}}(\hat{\Gamma}_0 - 2I_p) &\stackrel{\mathcal{D}}{\to} \mathcal{N}(0, 6), \\
\mathrm{Tr}\sqrt{np^{-1}}(\hat{\Gamma}_1 + \hat{\Gamma}_1^* - 2I_p) &\stackrel{\mathcal{D}}{\to} \mathcal{N}(0, 12).
\end{aligned}$$

Simulation result given in Row 2 (right panel, Figure 8.15) and Row 2 (left and right panels, Figure 8.16) support the above convergences.

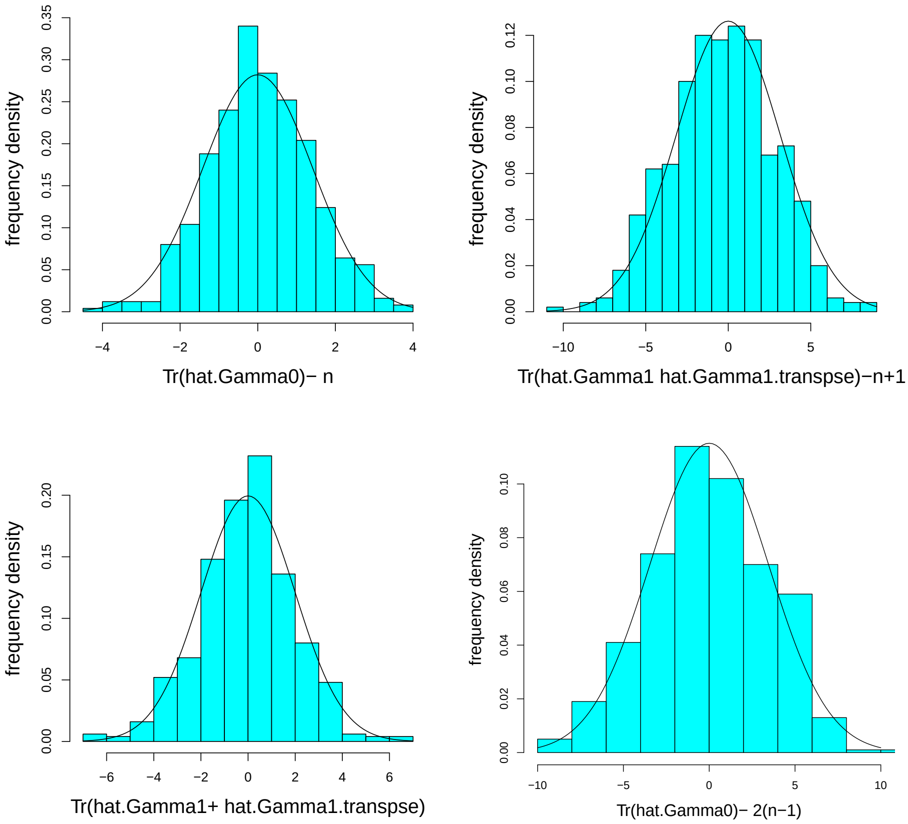


Figure 8.15: $n = p = 500$ and 500 replications. Row (1) left, Row (1) right and Row (2) left represent respectively the histogram of $(\mathrm{Tr}(\hat{\Gamma}_0) - n)$, $(\mathrm{Tr}(\hat{\Gamma}_1\hat{\Gamma}_1^*) - n + 1)$ and $\mathrm{Tr}(\hat{\Gamma}_1 + \hat{\Gamma}_1^*)$, when $X_t = \varepsilon_t$. Row (2) right represents the histogram of $(\mathrm{Tr}(\hat{\Gamma}_0) - 2(n-1))$, when $X_t = \varepsilon_t + \varepsilon_{t-1}$.

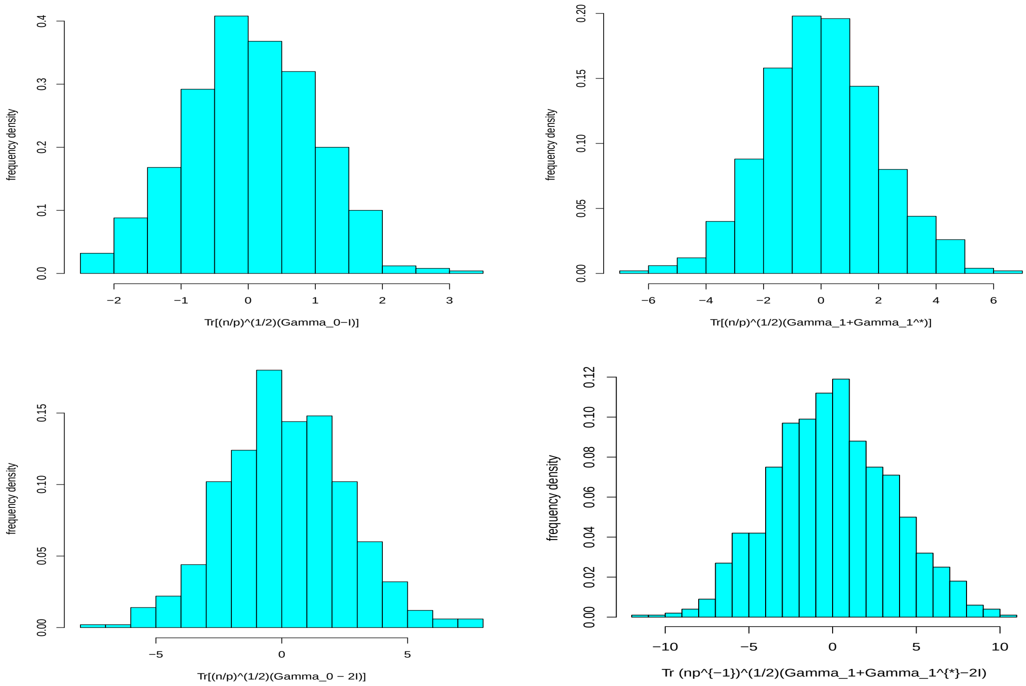


Figure 8.16: $p = n^{0.9}$ and 500 replications. Row (1) left and Row (1) right represent respectively the histogram of $(\mathrm{Tr}\sqrt{np^{-1}}(\hat{\Gamma}_0 - I))$ and $\mathrm{Tr}\sqrt{np^{-1}}(\hat{\Gamma}_1 + \hat{\Gamma}_1^*)$, when $X_t = \varepsilon_t$. Row (2) left and Row (2) right respectively represent the histogram of $(\mathrm{Tr}\sqrt{np^{-1}}(\hat{\Gamma}_0 - 2I))$ and $\mathrm{Tr}\sqrt{np^{-1}}(\hat{\Gamma}_1 + \hat{\Gamma}_1^* - 2I)$, when $X_t = \varepsilon_t + \varepsilon_{t-1}$.

Now we point out briefly how the above results can be used for testing.

Application to testing. Suppose we wish to test

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

If $p/n \to 1$, then by Example 8.4.1, under H_0, $(\mathrm{Tr}(\hat{\Gamma}_0) - n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2)$ and under H_1, $(\mathrm{Tr}(\hat{\Gamma}_0) - 2(n-1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 8)$. Similarly, if $p/n \to 0$, then by Example 8.4.2, under H_0, $\mathrm{Tr}\sqrt{np^{-1}}(\hat{\Gamma}_0 - I_p) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 6)$ and under H_1, $\mathrm{Tr}\sqrt{np^{-1}}(\hat{\Gamma}_0 - 2I_p) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 6)$.

Then $(\mathrm{Tr}(\hat{\Gamma}_0) - n)$ (when $p/n \to 1$) and $\mathrm{Tr}\sqrt{np^{-1}}(\hat{\Gamma}_0 - I_p)$ (when $p/n \to 0$) can be used as test statistics and large value of the test statistic will indicate rejection of H_0. Clearly this idea can be extended to test other pairs of simple null and alternative hypotheses for Model (3.7) and also when $p/n \to y \neq 1$.

We have not pursued this idea further in this thesis. We believe this idea can be developed further and will be extremely useful in the statistical analysis of high-dimensional time series.

Bibliography

G. I. Allen and R. Tibshirani. Transposable regularized covariance models with an application to missing data imputation. *Ann. Appl. Stat.*, 4(2):764–790, 2010.

G. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random matrices.* Cambridge University Press, Cambridge, UK, 2009.

L. Arnold. On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.*, 20(2):262–268, 1967.

L. Arnold. On Wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahr. und Verw. Gebiete*, 19(3):191–198, 1971.

Z. Bai, X. Wang, and W. Zhou. CLT for linear spectral statistics of Wigner matrices. *Electron. J. Probab.*, 14(83):2391–2417, 2009.

Z. D. Bai and J. W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, 32(1A):553–605, 2004.

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

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.

T. Banica, S. T. Belinschi, M. Capitaine, and B. Collins. Free Bessel laws. *Canad. J. Math.*, 63(1):3–37, 2011.

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.

F. Benaych-Georges. Rectangular random matrices, related convolution. *Probab. Theory Related Fields*, 144(3-4):471–515, 2009.

F. Benaych-Georges. On a surprising relation between the Marčenko–Pastur law, rectangular and square free convolutions. *Ann. de l'institut Henri Poincaré (B)*, 46(3):644–652, 2010.

F. Benaych-Georges and R. R. Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Mult. Anal.*, 111:120–135, 2012.

J. Bennett and S. Lanning. The netflix prize. *Proceedings of KDD Cup and Workshop 2007, San Jose, California, USA*, 2007.

R. Bhatia. *Notes on Functional Analysis.* Hindustan Book Agency, India, 2009.

M. Bhattacharjee and A. Bose. Consistency of large dimensional sample covariance matrix under weak dependence. *Statistical Methodology*, 20:11–26, 2014a.

M. Bhattacharjee and A. Bose. Estimation of autocovariance matrices for infinite dimensional vector linear process. *J. Time Series Anal.*, 35(3):262–281, 2014b.

M. Bhattacharjee and A. Bose. Large sample behaviour of high dimensional autocovariance matrices. *To appear in Ann. Statist.*, (http://www.e-publications.org/ims/submission/AOS/user/submissionFile/22891?confirm=b1fb1867), 2015a.

M. Bhattacharjee and A. Bose. Polynomial generalizations of sample variance-covariance matrices when $pn^{-1} \to 0$. *Submitted for publication*, (http://www.isical.ac.in/~statmath/html/publication/y0.pdf), 2015b.

P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227, 2008.

A. Bose, R. Subhra Hazra, and K. Saha. Patterned random matrices and method of moments. *Proceedings of the International Congress of Mathematicians. Hyderabad, India*, World Scientific, Singapore and Imperial College Press, UK: 2203–2230, 2010.

P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer, 2009.

A. Chudik and M. H. Pesaran. Infinite-dimensional VARs and factor models. *J. Econometrics*, 163(1):4–22, 2011.

P. Diaconis and S. N. Evans. Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.*, 353(7):2615–2633, 2001.

S. Dudoit, J. Fridlyand, and T. P. Speed. Comparison of discrimination methods for the classification of tumor using gene expression data. *J. Amer. Statist. Assoc.*, 97(457):77–87, 2002.

P. H. Edelman. Chain enumeration and non-crossing partitions. *Discrete Mathematics*, 31(2):171–180, 1980.

B. Efron. Are a set of microarrays independent of each other? *Ann. Appl. Stat.*, 3(3):922–942, 2009.

M. Forni and M. Lippi. The generalized dynamic factor model: representation theory. *Econometric Theory*, 17(6):1113–1141, 2001.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: consistency and rates. *J. Econometrics*, 119(2):231–235, 2004.

G. H. Golub and C. F. van Loan. *Matrix Computations, 3rd ed.* The Johns Hopkins University Press, Baltimore, 1996.

U. Grenander and G. Szegö. *Toeplitz Forms and Their Applications.* University of California Press, 1958.

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.

I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.

I. M. Johnstone and A. Y. Lu. Sparse principal components analysis. *Technical Report, Stanford University*, 2004.

I. M. Johnstone and A. Y. Lu. On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.*, 104(486):682–693, 2009.

L. Klebanov and A. Yakovlev. Diverse correlation structures in gene expression data and their utility in improving statistical inference. *Ann. Appl. Stat.*, 1(2): 538–559, 2007.

J. Leek and J. Storey. A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences*, 105(48):18718–18723, 2008.

H. Liu. *Spectral Analysis of High Dimensional Time Series.* PhD Thesis, Applied Mathematics, University of California, Davis, 2013.

H. Liu, A. Aue, and D. Paul. On the Marčenko-Pastur law for linear time series. *Ann. Statist.*, 43(2):675–712, 2015.

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

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

A. B. Owen. Variance of the number of false discoveries. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(3):411–426, 2005.

D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica*, 17(4):1617–1642, 2007.

O. Pfaffel and E. Schlemm. Eigenvalue distribution of large sample covariance matrices of linear processes. *Probab. Math. Statist.*, 31(2):313–329, 2011.

L. Saulis and V. A. Statulevičius. *Limit Theorems for Large Deviations*. Kluwer Academic Publishers, Dordrecht, 1991.

X. Shao. Testing for white noise under unknown dependence and its application to diagnostic checking for time series models. *Econometric Theory*, 27(2):312–343, 2011.

K. W. Wachter. The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. probab.*, 6(1):1–18, 1978.

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.

L. Wang, A. Aue, and D. Paul. Spectral analysis of linear time series in moderately high dimensions. (http://arxiv.org/abs/1504.06360), 2015.

E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. Math.*, 67(2):325–328, 1958.

H. Xiao and W. B. Wu. Portmanteau test and simultaneous inference for serial covariances. *Statist. Sinica*, 24(2):577–600, 2014.

Y. Q. Yin. Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.*, 20(1):50–68, 1986.