On eigenvalues of sample covariance matrices based on high-dimensional compositional data*

Qianqian Jiang[†] Jiaxin Qiu[‡] Zeng Li[§]

(v1.0: 2023-Dec-22; v2.0: 2024-Dec-17; v3.0: 2025-Jun-11.)

Abstract

This paper studies the asymptotic spectral properties of the sample covariance matrix for high-dimensional compositional data, including the limiting spectral distribution, the limit of extreme eigenvalues, and the central limit theorem for linear spectral statistics. All asymptotic results are derived under the high-dimensional regime where the data dimension increases to infinity proportionally with the sample size. The findings reveal that the limiting spectral distribution is the well-known Marčenko-Pastur law. The largest (or smallest non-zero) eigenvalue converges almost surely to the left (or right) endpoint of the limiting spectral distribution, respectively. Moreover, the linear spectral statistics demonstrate a Gaussian limit. Based on our CLT result, we investigate a test problem on the population covariance structure of the basis data. Simulation experiments demonstrate the accuracy of theoretical results.

Contents

*arXiv: 2312.14420

[†]Southern University of Science and Technology, email: jqq172515@gmail.com

[‡]The University of Hong Kong, email: jxqiu@hku.hk

[§]Southern University of Science and Technology, email: liz9@sustech.edu.cn. Corresponding author.

1 Introduction

In recent years, there has been increasing interest in the analysis of high-dimensional compositional data (HCD), which arise in various fields including genomics, ecology, finance, and social sciences. Compositional data refers to observations whose sum is a constant, such as proportions or percentages. HCD often involve a large number of variables or features measured for each sample, posing unique challenges for analysis. In the field of genomics, HCD analysis plays a crucial role in studying the composition and abundance of microbial communities, such as the human gut microbiome. Understanding the microbial composition and its relationship with health and disease has significant implications for personalized medicine and therapeutic interventions.

Statistical inference in HCD involves microbial mean tests, covariance matrix structural tests, and linear regression hypothesis testing. These inferences are intricately linked to the statistical properties of the sample covariance matrix. Mean tests typically utilize sum-of-squares-type and maximum-type statistics for dense and sparse alternative hypothesis, respectively. Cao et al. [2018] extended the maximum test framework by Cai et al. [2014] for compositional data. However, there's a gap in having a suitable sum-of-squares-type statistic for dense alternatives in HCD mean tests. Many sum-of-squares-type statistics, like Hotelling's T^2-statistic, rely on the sample covariance matrix. For bacterial species correlation, Faust et al. [2012] introduced the permutation-renormalization bootstrap (ReBoot), directly calculating correlations from compositional components. Shuffling is suggested due to compositional data's closure constraint, introducing negative correlations. Yet, compositional data's unique properties require an additional normalization step within the same sample post-shuffling, potentially impacting the theoretical validity of permutation and resampling methods. Additionally, resampling increases computational complexity for p-value calculation and confidence interval construction. To address these challenges, Wu et al. [2011] developed a covariance matrix element hypothesis testing method, allowing control over false discovery proportion (FDP) and false discovery rate (FDR). All these studies are closely related to the sample covariance matrix of HCD.

Current research predominantly focuses on sparse compositional data. In dense scenarios, researchers often turn to the spectral properties of sample covariance matrices. Despite this, there is a notable gap in the field of random matrices where specific attention to structures resembling compositional data, where row sum of the data matrix is constant, is lacking. Statistical inference for HCD encounters challenges arising not only from constraints but also from high dimensionality. Recognizing the crucial role of spectral theory in sample covariance matrices is also vital for addressing statistical challenges associated with high-dimensional data. Importantly, while previous research on statistical inference for HCD has overlooked studies under the spectral theory of sample covariance matrices, our work takes on these challenges from a Random Matrix Theory perspective. Existing literature extensively covers spectral properties of large-dimensional sample covariance matrices, but most results rely on independent component data structure, i.e. $\mathbf{Z} = \mathbf{\Gamma X}$, where $\mathbf{\Gamma}$ is determined, and $\mathbf{X}$ has independent and identically distributed (i.i.d.) components. Seminal works by Marčenko and Pastur [1967] and Jonsson [1982] established the limiting spectral distribution (LSD) of the sample covariance matrix $n^{-1}\mathbf{X}\mathbf{X}'$, where $\mathbf{X}$ is an i.i.d. data matrix with zero mean, leading to the well-known Marčenko-Pastur law. Subsequent research by Yin and Krishnaiah [1983] and Silverstein and Bai [1995] extended these findings to the sample covariance matrix $n^{-1}\mathbf{X}\mathbf{\Sigma}\mathbf{X}'$ for data with a linear dependence structure. Zhang [2007] extended to the general separable product form $n^{-1}\mathbf{A}^{1/2}\mathbf{X}\mathbf{B}\mathbf{X}'\mathbf{A}^{1/2}$, where $\mathbf{A}$ is nonnegative definite, and $\mathbf{B}$ is Hermitian. Another important area of interest is the investigation of extreme eigenvalues. Johnstone [2001] explored the fluctuation of the extreme eigenvalues of the sample covariance matrix $n^{-1}\mathbf{X}\mathbf{X}'$, proving that the standardized largest eigenvalue follows the Tracy-Widom law. Related extensions include sample covariance matrices with linear dependence structures [El Karoui, 2007], Kendall rank correlation coefficient matrices [Bao, 2019], among others. Considerable attention has also been given to the study of linear functionals of eigenvalues. Bai and Silverstein [2004] established the Central Limit Theorem (CLT) for the Linear Spectral Statistics (LSS) of the sample covariance matrix $n^{-1}\mathbf{A}^{1/2}\mathbf{X}\mathbf{X}'\mathbf{A}^{1/2}$, later extended to sample correlation coefficient matrices [Gao et al., 2017], and separable product matrices [Bai et al., 2019]. To summarize, existing results in spectral theory of large dimensional sample covariance matrix predominantly rely on independent component data structure which, unfortunately, HCD does not fit in.

Specifically, current second-order limit theorems do not apply to HCD, making the exploration of spectral theory for HCD with distinct constraints crucial. This paper delves into spectral theory for sample covariance matrices of HCD, including LSD, extreme eigenvalues, and CLT for LSS. Analyzing HCD faces challenges due to compositional data's specific dependence structure, making existing techniques for i.i.d. observations less applicable. However, we can assume that HCD are generated from unobservable basis data, while the underlying basis data follow independent component model structure. In this way, spectral analysis of the sample covariance matrix of HCD can be approached through the basis data. In fact, the structure of the sample covariance matrix of HCD is similar to that of the Pearson sample correlation matrix in basis data. Therefore, we leverage the analysis methods of the spectral theory of the Pearson sample correlation matrix to study the spectral theory of the sample covariance matrix of HCD. In the field of random matrices, research on the spectral theory of the Pearson sample correlation matrix based on independent data is relatively mature. Jiang [2004] demonstrated that the LSD of sample correlation matrix for i.i.d data is the well-known Marčenko-Pastur law. Gao et al. [2017] derive the CLT for LSS of the Pearson sample correlation matrix. The derivation of spectral theory for the sample covariance matrix of HCD can benefit from methods in this context. The LSD of the sample covariance matrix for HCD in Theorem 2.3 is established following the strategy in Jiang [2004], and we further investigate the extreme eigenvalues in Proposition 2.4. The proof strategy of CLT for LSS in Theorem 2.5 follows the methodologies outlined in Bai and Silverstein [2004] for the sample covariance matrix and Gao et al. [2017] for the sample correlation matrix. However, due to the dependence inherent in HCD, certain tools from these works cannot be directly applied to the sample covariance matrix of HCD. In response, we introduce new techniques. Specifically, we establish concentration inequalities for compositional data. One of the central ideas of the paper, grounded in concentration phenomena, permeates the entire proof (details in Section 5.2 and Section 5.3), where we develop three crucial technique lemmas (see Lemmas 5.3 – 5.5) essential for the proof. Finally, it is noteworthy that the mean and variance-covariance in Theorem 2.5 differ from those in Bai and Silverstein [2004], and additional terms are present in both the mean and variance-covariance.

The paper is organized as follows. Section 2.2 investigates the LSD and extreme eigenvalues of the sample covariance matrix for HCD. Section 2.3 establishes the CLT for LSS of the sample covariance matrix for HCD. Section 3 studies a test problem on the population covariance structure of the basis data based on our CLT result. Section 4 reports numerical studies. Section 5 presents the sketch of proof of our CLT for LSS. Auxiliary lemmas and technical proofs are relegated to the supplementary material.

Before proceeding, we introduce some notations that will be used throughout this paper. We adopt the convention of using regular letters for scalars and boldface letters for vectors or matrices. For any matrix $\mathbf{A}$, we denote its (i,j)-th entry by A_{ij}, its transpose by $\mathbf{A}'$, its trace by $\mathrm{tr}(\mathbf{A})$, its j-th largest eigenvalue by $\lambda_j(\mathbf{A})$, its spectral norm by $\|\mathbf{A}\| = \sqrt{\lambda_1(\mathbf{A}\mathbf{A}')}$. For a set of random variables $\{X_n\}_{n=1}^{\infty}$ and a corresponding set of nonnegative real numbers $\{a_n\}_{n=1}^{\infty}$, we write $X_n = O_P(a_n)$ if, for any $\varepsilon > 0$, there exists a constant $C > 0$ and $N > 0$ such that $\mathbb{P}(|X_n/a_n| \geq C) \leq \varepsilon$ for all $n \geq N$. We write $X_n = o_P(a_n)$ if $\lim_{n\to\infty} \mathbb{P}(|X_n/a_n| \geq \varepsilon) = 0$ for any $\varepsilon > 0$. Furthermore, we write $X_n \xrightarrow{a.s.} X$ ($X_n \xrightarrow{P} X$, $X \xrightarrow{D} X$ respectively) if X_n converges almost surely (in probability, in distribution, respectively) to X. We denote by C and K constants that may vary from line to line.

2 Main Results

2.1 Preliminaries and notations

Let $\mathbf{X}_n = (x_{ij})_{n\times p}$ denote the $n \times p$ observed data matrix, and each row $(x_{i1}, \ldots, x_{ip})$ represents compositions that lie in the $(p-1)$-dimensional simplex $\mathcal{S}^{p-1} = \{(y_1, \ldots, y_p) : \sum_{j=1}^{p} y_j = 1, y_j \geq 0\}$. We assume that the compositional variables arise from a vector of latent variables, which we call the basis. Let $\mathbf{W}_n = (w_{ij})_{n\times p}$ denote the $n \times p$ matrices of unobserved bases, where w_{ij}'s are positive and i.i.d. with mean $\mu > 0$ and variance σ^2. The observed compositional data is generated via the normalization

$$x_{ij} = \frac{w_{ij}}{\sum_{\ell=1}^{p} w_{i\ell}}, \qquad 1 \le i \le n,\ 1 \le j \le p. \tag{1}$$

The unbiased sample covariance matrix of $\mathbf{X}_n$ is defined by $\mathbf{S}_{n,N} = \frac{1}{N}\mathbf{X}_n'\mathbf{C}_n\mathbf{X}_n$, where $\mathbf{C}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$, $\mathbf{1}_n$ is a n-dimensional vector of all ones, and $N = n-1$ is the adjusted sample size. Since $\sum_{\ell=1}^{p} w_{i\ell} = p\mu(1+\varepsilon_i)$ with $\sup_i \varepsilon_i = o_P(1)$, we rescale $\mathbf{S}_{n,N}$ as

$$\mathbf{B}_{p,N} = p^2\mathbf{S}_{n,N} = \frac{1}{N}(p\mathbf{X}_n)'\mathbf{C}_n(p\mathbf{X}_n).$$

For any $p \times p$ Hermitian matrix $\mathbf{A}_p$, its *empirical spectral distribution* (ESD) is defined by

$$F^{\mathbf{A}_p}(x) = \frac{1}{p}\sum_{i=1}^{p} I_{\{\lambda_i(\mathbf{A}_p)\le x\}},$$

where $I_{\{\cdot\}}$ denotes the indicator function. If $F^{\mathbf{A}_p}(x)$ converges to a non-random limit $F(x)$ as $p \to \infty$, we call $F(x)$ the *limiting spectral distribution* of $\mathbf{A}_p$. The LSD of $\mathbf{A}_p$ is described in terms of its Stieltjes transform. The Stieltjes transform of any cumulative distribution function F is defined by

$$m_F(z) = \int \frac{1}{\lambda - z}\,\mathrm{d}F(\lambda), \qquad z \in \mathbb{C}^+ := \{z : \Im(z) > 0\}.$$

Many classes of statistics related to the eigenvalues of $\mathbf{B}_{p,N}$ are important for multivariate inference, particularly functionals of the ESD. To explore this, for any function f defined on $[0,\infty)$, we consider the *linear spectral statistics* (LSS) of $\mathbf{B}_{p,N}$ given by

$$\int f(x)\,\mathrm{d}F^{\mathbf{B}_{p,N}}(x) = \frac{1}{p}\sum_{i=1}^{p} f\Big(\lambda_i(\mathbf{B}_{p,N})\Big).$$

In this paper, we study the asymptotic spectral properties of $\mathbf{B}_{p,N}$, including the LSD (see Theorem 2.3), the behavior of extreme eigenvalues (see Proposition 2.4), and the CLT for LSS (see Theorem 2.5).

2.2 Limiting spectral distribution and extreme eigenvalues

Analyzing HCD poses challenges due to its unique dependence structure, making existing techniques for i.i.d. observations less applicable. To overcome this difficulty, we assume that the compositional data is generated from basis data and the basis data follows the commonly used independent component structure. Specifically, the unbiased sample covariance matrix of $\mathbf{X}_n$ is defined by

$$\mathbf{S}_{n,N} = \frac{1}{N}\mathbf{X}_n^{'}\mathbf{C}_n\mathbf{X}_n = \frac{1}{N}\mathbf{W}_n^{'}\mathbf{\Lambda}_n\mathbf{C}_n\mathbf{\Lambda}_n\mathbf{W}_n,$$

where $\mathbf{X}_n = \mathbf{\Lambda}_n\mathbf{W}_n$, and $\mathbf{\Lambda}_n = \mathrm{diag}(1/\sum_{j=1}^{p} w_{1j}, \ldots, 1/\sum_{j=1}^{p} w_{nj})$. Here, we assume that $\mathbf{W}_n = (\mathbf{w}_1, \ldots, \mathbf{w}_p)$ has i.i.d. components w_{ij}, with $\mathbb{E}(w_{ij}) = \mu > 0$, $\mathrm{Var}(w_{ij}) = \sigma^2$. Recall that the Pearson sample correlation matrix for $\mathbf{W}_n$ is

$$\mathbf{R}_n = \frac{1}{n}\widetilde{\mathbf{X}}_n'\mathbf{C}_n\widetilde{\mathbf{X}}_n = \frac{1}{n}\widetilde{\mathbf{\Lambda}}_p\mathbf{W}_n^{'}\mathbf{C}_n\mathbf{W}_n\widetilde{\mathbf{\Lambda}}_p,$$

where $\widetilde{\mathbf{X}}_n = \mathbf{W}_n\widetilde{\boldsymbol{\Lambda}}_p$, $\widetilde{\boldsymbol{\Lambda}}_p = \mathrm{diag}(\sqrt{n}\|\mathbf{w}_1 - \overline{\overline{\mathbf{w}}}_1\|_2^{-1}, \ldots, \sqrt{n}\|\mathbf{w}_p - \overline{\overline{\mathbf{w}}}_p\|_2^{-1})$, and $\overline{\overline{\mathbf{w}}}_j = n^{-1}\sum_{i=1}^n w_{ij}\mathbf{1}_n$ with $\mathbf{1}_n$ being an n-dimensional vector whose entries are all 1's. It can be seen that the normalizing matrix $\boldsymbol{\Lambda}_n$ of $\mathbf{S}_{n,N}$ is very similar to $\widetilde{\boldsymbol{\Lambda}}_p$ of $\mathbf{R}_n$. The former uses $(\sum_{j=1}^p w_{ij})^{-1}$ for normalization, while the latter utilizes $\sqrt{n}\|\mathbf{w}_j - \overline{\overline{\mathbf{w}}}_j\|_2^{-1}$. This allows us to leverage the techniques from the spectral theory of the Pearson sample correlation matrix in studying the asymptotic spectral properties of the sample covariance matrix for HCD.

Before diving into the LSS of $\mathbf{B}_{p,N}$, we first explore its LSD and extreme eigenvalues. Specifically, suppose the following assumptions hold,

Assumption 2.1. *$\{w_{ij} > 0, i = 1,\ldots,n, j = 1,\ldots,p\}$ are i.i.d. real random variables with $\mathbb{E}w_{11} = \mu > 0$, $\mathrm{Var}(w_{11}) = \sigma^2$ and $\mathbb{E}(w_{11} - \mu)^4 < \infty$. For notational simplicity, we write $\lambda = \sigma^2/\mu^2$ throughout the paper.*

Assumption 2.2. *$c_N = p/N$ tends to a positive $c > 0$ as $p, N \to \infty$.*

Theorem 2.3. *Under Assumptions 2.1 and 2.2, with probability one, the ESD of $\mathbf{B}_{p,N}$ converges weakly to a deterministic probability distribution with a density function*

$$f_{\mathsf{MP}}(x) = \frac{1}{2\pi c\lambda x}\sqrt{[(b-x)(x-a)]_+}, \tag{2}$$

and a point mass $1 - 1/c$ at $x = 0$ if $c > 1$, where $a := \lambda(1-\sqrt{c})^2$, $b := \lambda(1+\sqrt{c})^2$, and $[y]_+ := \max\{y, 0\}$.

Proof of Theorem 2.3. Let $\mathbf{Y}_n = p\mathbf{C}_n\boldsymbol{\Lambda}_n\mathbf{W}_n/\sqrt{N}$ and $\check{\mathbf{Y}}_n = \mathbf{C}_n\mathbf{W}_n/(\sqrt{N}\mu)$. Note that the LSD of $\check{\mathbf{Y}}_n'\check{\mathbf{Y}}_n$ is the well-known Marčenko-Pastur law with the density function given by (2). From Theorem A.47 of Bai and Silverstein [2010] and our Proposition 2.4, it suffices to prove that

$$\|\mathbf{Y}_n - \check{\mathbf{Y}}_n\| \xrightarrow{a.s.} 0. \tag{3}$$

By Lemma S1.3, we have $\max_{1\le i\le n}|\sum_{j=1}^p w_{ij}/(p\mu) - 1| \xrightarrow{a.s.} 0$, which implies that $\|p\mu\boldsymbol{\Lambda}_n - \mathbf{I}_n\| \xrightarrow{a.s.} 0$. Moreover, we get from Theorem 2.9 in Benaych-Georges and Nadakuditi [2012] that $\|\mathbf{W}_n/(\mu\sqrt{N})\|$ is bounded almost surely. Hence, we have

$$\|\mathbf{Y}_n - \check{\mathbf{Y}}_n\| = \left\|\mathbf{C}_n(p\mu\boldsymbol{\Lambda}_n - \mathbf{I}_n)\frac{\mathbf{W}_n}{\mu\sqrt{N}}\right\| \le \|p\mu\boldsymbol{\Lambda}_n - \mathbf{I}_n\| \cdot \left\|\frac{\mathbf{W}_n}{\mu\sqrt{N}}\right\| \xrightarrow{a.s.} 0.$$

This completes the proof. □

We denote the LSD of $\mathbf{B}_{p,N}$ by $F^c(x)$, whose density function is $f_{\mathsf{MP}}(x)$, as defined in (2). The superscript c indicates the dimension-to-sample ratio. For each $z \in \mathbb{C}^+$, by our Theorem 2.3 and Theorem 1.1 in Silverstein and Bai [1995], the Stieltjes transform of F^c, denoted $m(z) \equiv m_{F^c}(z)$, is the unique solution to the equation

$$m = \frac{1}{\lambda(1-c-czm) - z}.$$

in the set $\{m \in \mathbb{C} : -(1-c)/z + cm \in \mathbb{C}^+\}$. Consider a *companion matrix* $\underline{\mathbf{B}}_{p,N} = \frac{p^2}{N}\mathbf{C}_n\mathbf{X}_n\mathbf{X}_n'\mathbf{C}_n$, which is of size $n \times n$ and differs from $\mathbf{B}_{p,N}$ by $|n-p|$ zeros eigenvalues. The LSD of $\underline{\mathbf{B}}_{p,N}$ is given by $\underline{F}^c(x) = (1-c)I_{[0,\infty)}(x) + cF^c(x)$. Its Stieltjes transform

$$\underline{m}(z) \equiv m_{\underline{F}^c}(z) = cm(z) - \frac{1-c}{z}$$

with inverse

$$z = z(\underline{m}) = -\frac{1}{\underline{m}(z)} + \frac{c\lambda}{1+\lambda\underline{m}(z)}.$$

Proposition 2.4. *Under Assumptions 2.1 and 2.2, we have*

$$\lambda_{\max}(\mathbf{B}_{p,N}) \xrightarrow{a.s.} \lambda(1+\sqrt{c})^2 \quad \text{and} \quad \lambda_{\min}(\mathbf{B}_{p,N}) \xrightarrow{a.s.} \lambda(1-\sqrt{c})^2, \tag{4}$$

where $\lambda_{\max}(\mathbf{B}_{p,N})$ is the largest eigenvalue of $\mathbf{B}_{p,N}$, and $\lambda_{\min}(\mathbf{B}_{p,N})$ is the smallest non-zero eigenvalue of $\mathbf{B}_{p,N}$. Furthermore, for any $\ell > 0$, $\eta_1 > \lambda(1+\sqrt{c})^2$ and $0 < \eta_2 < \lambda(1-\sqrt{c})^2 I_{\{0<c<1\}}$, under the condition $|w_{ij} - \mu| < \delta_n\sqrt{n}$ $(i = 1,\dots,n,\ j = 1,\dots,p)$ where $\{\delta_n\}$ is a positive sequence satisfying that $\delta_n \to 0$, $\delta_n n^{1/4} \to \infty$, $\delta_n^{-4}\mathbb{E}|w_1-\mu|^4 I_{\{|w_1-\mu|\geq\delta_n\sqrt{n}\}} \to 0$ as $n \to \infty$, we have

$$\mathbb{P}\Big(\lambda_{\max}(\mathbf{B}_{p,N}) \geq \eta_1\Big) = o(n^{-\ell}) \quad \text{and} \quad \mathbb{P}\Big(\lambda_{\min}(\mathbf{B}_{p,N}) \leq \eta_2\Big) = o(n^{-\ell}). \tag{5}$$

Proof of Proposition 2.4. The convergence (4) is an immediate consequence of Equation (2.7) in Jiang [2004], Theorem 1.4 in Xiao and Zhou [2010], Equation (3) and Lemma S1.1. The proof of (5) is postponed to the supplementary material. □

Remark 1. The results of extreme eigenvalues are useful in locating eigenvalues of the population covariance matrix and proving the CLT for LSS. Proposition 2.4 demonstrates that, with probability one, there are no eigenvalues of $\mathbf{B}_{p,N}$ outside the support of the LSD under Assumptions 2.1 – 2.2. These results are crucial for applying the Cauchy integral formula (see Equation (11)) and proving tightness.

Remark 2. For the special case when $p = n$, the matrix $\mathbf{X}_n = (w_{ij}/\sum_{\ell=1}^n w_{i\ell})$ is a *random Markov matrix*. The work of Bordenave et al. Bordenave et al. [2012] provided key insights into the first-order properties of both eigenvalues and singular values of the $n \times n$ matrix $\mathbf{X}_n$, including the limiting distribution of its singular values and the convergence of its extreme singular values. Our first-order results (see Theorem 2.3 and Proposition 2.4) can be viewed as an extension of their findings regarding singular values when $p, n \to \infty$ and $p/n \to c \in (0,\infty)$.

In contrast to the scope of Bordenave et al. Bordenave et al. [2012], our work focus on a different setting and investigates a different aspect of $\mathbf{X}_n$. Specifically, we examine the centered version $\mathbf{C}_n\mathbf{X}_n$ without requiring $p = n$ and we focus on second-order fluctuations of LSS. Beyond first-order limits, we derive the CLT for LSS of the sample covariance matrix of $\mathbf{X}_n$ (see Theorem 2.5 in the following Section 2.3).

2.3 CLT for LSS

We focus on linear functionals of eigenvalues of $\mathbf{B}_{p,N}$, i.e. $\frac{1}{p}\sum_{i=1}^p f(\lambda_i(\mathbf{B}_{p,N}))$. Naturally, it converges to the functional integration of the LSD of $\mathbf{B}_{p,N}$, that is, $\int f(x)\,\mathrm{d}F^c(x)$. In this section, we explore the second-order fluctuation of $\frac{1}{p}\sum_{i=1}^p f(\lambda_i)$, which describes how such LSS converges to its first-order limit. Define

$$G_{p,N}(f) = p\int f(x)\,\mathrm{d}\{F^{\mathbf{B}_{p,N}}(x) - F^{c_N}(x)\},$$

where $F^{c_N}(x)$ is obtained by substituting c_N for c in $F^c(x)$, the LSD of $\mathbf{B}_{p,N}$. We show that under Assumptions 2.1 – 2.2 and the analyticity of f, the rate of $\int f(x)\,\mathrm{d}\{F^{\mathbf{B}_{p,N}}(x) - F^{c_N}(x)\}$ approaching zero is essentially $1/n$, and $G_{p,N}(f)$ convergence weakly to a Gaussian variable.

Before presenting the main result, we recall some notations. We denote $\lambda = \sigma^2/\mu^2$. Recall that $m(z)$ and $\underline{m}(z)$ are the Stieltjes transforms of the LSD $F^c(x)$ and the companion LSD $\underline{F}^c(x)$, respectively. Furthermore, we define $m'(z)$ as the derivative of $m(z)$ with respect to z throughout the rest of this paper. The main result is stated in the following theorem.

Theorem 2.5. *Under Assumptions 2.1 and 2.2, let $f_1, f_2, \dots, f_k$ be functions on $\mathbb{R}$ and analytic on an open interval containing $[\lambda(1-\sqrt{c})^2 I_{\{0<c<1\}},\ \lambda(1+\sqrt{c})^2]$. Then, the random vector $(G_{p,N}(f_1), \dots, G_{p,N}(f_k))'$ forms a tight sequence in p and converges weakly to a Gaussian vector $(X_{f_1}, \dots, X_{f_k})$ with mean function*

$$
\begin{aligned}
\mathbb{E}X_f = &\frac{c\lambda^2}{2\pi i}\oint_{\mathcal{C}} \frac{f(z)\underline{m}^3(z)\{1+\lambda\underline{m}(z)\}}{[\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)]^2}\,\mathrm{d}z \\
&-\frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{f(z)\underline{m}^2(z)\{1+\lambda\underline{m}(z)\}\{z(h_1+\lambda)m(z)+\lambda\}}{\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)}\,\mathrm{d}z \\
&-\frac{c}{2\pi i}\oint_{\mathcal{C}} \frac{f(z)z^2\underline{m}^3(z)\{1+\lambda\underline{m}(z)\}\{(\alpha_1+\alpha_2)m^2(z)+2\lambda^2 m'(z)\}}{\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)}\,\mathrm{d}z,
\end{aligned}
$$

and covariance function

$$
\begin{aligned}
\operatorname{Cov}(X_f, X_g) = &-\frac{1}{2\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} \frac{f(z_1)g(z_2)}{\{\underline{m}(z_1)-\underline{m}(z_2)\}^2}\,\mathrm{d}\underline{m}(z_1)\,\mathrm{d}\underline{m}(z_2) \\
&-\frac{c(\alpha_1+\alpha_2)}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} \frac{f(z_1)g(z_2)}{\{1+\lambda\underline{m}(z_1)\}^2\{1+\lambda\underline{m}(z_2)\}^2}\,\mathrm{d}\underline{m}(z_1)\,\mathrm{d}\underline{m}(z_2),
\end{aligned}
$$

where $\alpha_1 = \mathbb{E}(w_{11}/\mu - 1)^4 - 3\lambda^2$, $\alpha_2 = -4\lambda\mathbb{E}w_{11}^3/\mu^3 + 4\lambda^3 + 12\lambda^2 + 4\lambda$, $h_1 = -2\mathbb{E}w_{11}^3/\mu^3 + 3\lambda^2 + 5\lambda + 2$. The contours $\mathcal{C}, \mathcal{C}_1, \mathcal{C}_2$ are closed and taken in the positive direction in the complex plane, each enclosing the support of LSD, i.e., $[\lambda(1-\sqrt{c})^2 I_{\{0<c<1\}}, \lambda(1+\sqrt{c})^2]$.

Remark 3. We restrict attention to functions f which are analytic in a region of the complex plane containing the support of LSD. As demonstrated in Najim and Yao [2016], the analyticity requirement for f in the CLT can be relaxed by representing the LSS with the help of Helffer–Sjöstrand's formula instead of the Cauchy integral formula. For now, we focus on analytic cases because analytic functions are sufficient to achieve our current statistical objectives.

Applying Theorem 2.5 to three polynomial functions, we obtain the following corollary. The proof of Theorem 2.5 is postponed to Section 5, and detailed calculations in these applications are postponed to the supplementary material.

Corollary 2.6. *Under the same notations and assumptions as in Theorem 2.5, let $f_r = x^r$ for $r = 1, 2, 3$, we have*

$$
\begin{aligned}
G_{p,N}(f_1) &= \operatorname{tr}(\mathbf{B}_{p,N}) - p\lambda \xrightarrow{D} \mathcal{N}(\mu_1, V_1), \\
G_{p,N}(f_2) &= \operatorname{tr}(\mathbf{B}_{p,N}^2) - p(1+c_N)\lambda^2 \xrightarrow{D} \mathcal{N}(\mu_2, V_2), \\
G_{p,N}(f_3) &= \operatorname{tr}(\mathbf{B}_{p,N}^3) - p(1+3c_N+c_N^2)\lambda^3 \xrightarrow{D} \mathcal{N}(\mu_3, V_3),
\end{aligned}
$$

where $c_N = p/N$, and

$$
\begin{aligned}
&\mu_1 = h_1, \quad \mu_2 = (1+c)\lambda^2 + 2(1+c)\lambda h_1 + c(\alpha_1+\alpha_2), \\
&\mu_3 = (2+6c+3c^2)\lambda^3 + 3(1+3c+c^2)\lambda^2 h_1 + 3c(1+c)\lambda(\alpha_1+\alpha_2), \\
&V_1 = 2c\lambda^2 + c(\alpha_1+\alpha_2), \\
&V_2 = 4c(2+c)(1+2c)\lambda^4 + 4c(1+c)^2\lambda^2(\alpha_1+\alpha_2), \\
&V_3 = 6c(1+6c+3c^2)(3+6c+c^2)\lambda^6 + 9c(1+3c+c^2)^2\lambda^4(\alpha_1+\alpha_2).
\end{aligned}
$$

3 Application

Let $\mathbf{x} = (\frac{w_1}{w}, \frac{w_2}{w}, \ldots, \frac{w_p}{w})' \in \mathbb{R}^p$ be a random vector, where $\{w_i\}_{i=1}^p$ are independent random variables and $w := \sum_{i=1}^p w_i$. We aim to formally test

$$H_0\colon \{w_i\}_{i=1}^p \text{ are i.i.d.,} \quad \text{vs} \quad H_1\colon \{w_i\}_{i=1}^p \text{ are not i.i.d.,}$$

based on n samples under the regime $p, n \to \infty, p/n \to c \in (0, \infty)$. Under the null hypothesis H_0, the vector $\mathbf{x}$ is exchangeable, that is, its distribution is invariant under any permutation of its coordinates, implying identical marginals and symmetric dependence. Its covariance matrix is given by

$$\mathrm{Cov}(\mathbf{x}) = \frac{\nu_2^\circ}{p(p-1)}\left(\mathbf{I}_p - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p'\right) =: \frac{\nu_2^\circ}{p(p-1)}\mathbf{G}_p, \tag{6}$$

where $\nu_2^\circ := \mathbb{E}\{pw_1/(\sum_{k=1}^p w_k) - 1\}^2$. Readers are referred to Section S2.5 of the supplementary material for the proof of (6).

Remark 4. We cannot fully characterize "$H_0 : \{w_i\}_{i=1}^p$ are i.i.d." by $\mathrm{Cov}(\mathbf{x})$ alone because $\mathrm{Cov}(\mathbf{x})$ and $\mathrm{Cov}(\mathbf{w})$ do not have a one-to-one correspondence. In fact, the observed data $\mathbf{x}$ may satisfy (6) even when the underlying components $\{w_i\}_{i=1}^p$ are not independent. For example, consider the case where $p = 2k$ is even for some integer $k \geq 1$, and let

$$\{w_i\}_{i=1}^k \overset{\text{i.i.d.}}{\sim} \mathrm{Uniform}(0,1), \quad w_{k+i} = \begin{cases} w_i, & \text{with probability } \frac{1}{2}, \\ 1 - w_i, & \text{with probability } \frac{1}{2}. \end{cases}$$

It is easy to verify that $\{w_i\}_{i=1}^p$ are *uncorrelated but dependent*, and therefore do not satisfy H_0. However, $\mathrm{Cov}(\mathbf{x})$ still *satisfies* (6).

This example implies that even when $\mathbf{w}$ have entirely different dependency structures, the covariance of $\mathbf{x}$ can still be the same. However, the dependence structure of $\mathbf{w}$ is critical to the behavior of the LSS. When the dependence structure of $\mathbf{w}$ changes, the CLT for the LSS can also change, altering the distribution of the test statistic. In other words, if we were to characterize H_0 in terms of $\mathrm{Cov}(\mathbf{x})$, the distribution of the test statistic under H_0 would be uncertain and unidentifiable, making valid testing infeasible. Therefore, we must characterize H_0 in terms of $\mathbf{w}$, since its dependence structure directly determines the behavior of the LSS and thus the distribution of the test statistic.

3.1 Test statistic and its limiting null distribution

Given i.i.d. samples $\{\mathbf{x}_i\}_{i=1}^n$, we construct a test statistic based on the rescaled sample covariance matrix $\mathbf{B}_{p,N}$, as defined in Section 2.1. We consider the following Frobenius-norm-type test statistic:

$$T := \frac{1}{p}\left\|\mathbf{B}_{p,N} - \mathbf{B}\right\|_F^2, \quad \mathbf{B} := \mathrm{Cov}(p\mathbf{x}) = \frac{p\nu_2^\circ}{p-1}\mathbf{G}_p,$$

where $\|\mathbf{A}\|_F := (\sum_{i,j} A_{ij}^2)^{1/2}$ denotes the Frobenius norm of a matrix $\mathbf{A}$. The statistic measure the distance between $\mathrm{Cov}(p\mathbf{x})$ and its empirical counterpart. We reject H_0 when T is sufficiently large.

The test statistic T is closely related to LSS of $\mathbf{B}_{p,N}$, particularly for test functions $f(x) = x^k$, $k = 1, 2$. Specifically,

$$T = \frac{1}{p}\mathrm{tr}(\mathbf{B}_{p,N}^2) - \frac{2\nu_2^\circ}{p-1}\mathrm{tr}(\mathbf{B}_{p,N}) + \frac{p(\nu_2^\circ)^2}{p-1}.$$

Here, we use the identity $\mathbf{1}_p'\mathbf{B}_{p,N}\mathbf{1}_p = 0$. Using the CLT for $\mathrm{tr}(\mathbf{B}_{p,N})$ and $\mathrm{tr}(\mathbf{B}_{p,N}^2)$ (see Corollary 2.6) under H_0 and the Delta method, the limiting null distribution of T is obtained:

Theorem 3.1. *Suppose that Assumptions 2.1 and 2.2 hold, and $\mathbb{E}|w_{11}-\mu|^{6+s}<\infty$ for any $s>0$, under H_0, we have*

$$p(T-\mu_T)\xrightarrow{D}\mathcal{N}(0,\,\sigma_T^2),$$

where

$$\begin{aligned}\mu_T &= \lambda^2 c_N + \frac{\mu_2}{p} - \frac{2\lambda h_1+\lambda^2}{p-1}, \quad \sigma_T^2 = 4\lambda^2 V_1 - 4\lambda V_{12} + V_2,\\ V_{12} &= 2\lambda c(1+c)(2\lambda^2+\alpha_1+\alpha_2),\end{aligned} \tag{7}$$

h_1,α_1,α_2 are defined in Theorem 2.5, μ_2, V_1, V_2 are defined in Corollary 2.6.

The detailed proof is postponed to the supplementary material. In practical applications, we replace μ_T with its finite sample counterpart $\hat{\mu}_T$. To eliminate correlation between T and $\hat{\mu}_T$, we split the data $\mathbf{X}$ into two parts:

$$\mathbf{X}=\begin{pmatrix}\mathbf{X}^{(1)}_{n_1\times p}\\ \mathbf{X}^{(2)}_{n_2\times p}\end{pmatrix}\equiv\begin{pmatrix}\mathbf{X}^{(1)}\\ \mathbf{X}^{(2)}\end{pmatrix},\quad \frac{p}{n_1}\to c_1\in(0,\infty),\quad \frac{p}{n_2}\to c_2\in(0,\infty),$$

where $\mathbf{X}^{(1)}$ is used to calculate the test statistic, and $\mathbf{X}^{(2)}$ is used to compute $\hat{\mu}_T$. The new test statistic is defined by

$$\widetilde{T}=\frac{1}{p}\left\|\mathbf{B}^{(1)}_{p,N_1}-\frac{p\hat{v}_2^\circ}{p-1}\mathbf{G}_p\right\|_F^2,$$

where $N_1=n_1-1$, and

$$\mathbf{B}^{(1)}_{p,N_1}=\frac{p^2}{N_1}\left(\mathbf{X}^{(1)}\right)'\mathbf{C}_{n_1}\mathbf{X}^{(1)},\quad \hat{v}_2^\circ=\frac{1}{pn_2}\sum_{i=1}^{n_2}\sum_{j=1}^{p}\left(pX_{ij}^{(2)}-1\right)^2.$$

Moreover, we estimate μ_T by $\hat{\mu}_T$, which is obtained by replacing N in μ_T with $N_2=n_2-1$, and substituting the terms λ, $\mathbb{E}(w_{11}/\mu)^3$, and $\mathbb{E}(w_{11}/\mu-1)^4$ with $\hat{\lambda}=\hat{v}_2^\circ$, and

$$\widehat{\mathbb{E}}\left(\frac{w_{11}}{\mu}\right)^3=\frac{1}{pn_2}\sum_{i=1}^{n_2}\sum_{j=1}^{p}\left(pX_{ij}^{(2)}\right)^3,\qquad \widehat{\mathbb{E}}\left(\frac{w_{11}}{\mu}-1\right)^4=\frac{1}{pn_2}\sum_{i=1}^{n_2}\sum_{j=1}^{p}\left(pX_{ij}^{(2)}-1\right)^4.$$

The consistency of $\hat{\lambda}$, $\widehat{\mathbb{E}}(w_{11}/\mu)^3$, and $\widehat{\mathbb{E}}(w_{11}/\mu-1)^4$ follows from the law of large numbers. Using these newly defined notations, we derive the CLT for $\widetilde{T}$ as follows:

Theorem 3.2. *Suppose that Assumptions 2.1 and 2.2 hold, and $\mathbb{E}|w_{11}-\mu|^{6+s}<\infty$ for any $s>0$, under H_0, we have*

$$p(\widetilde{T}-\hat{\mu}_T)\xrightarrow{D}\mathcal{N}(\mu_\lambda,\sigma_{\widetilde{T}}^2+\sigma_\lambda^2),$$

where $\sigma_{\widetilde{T}}^2$ is defined similarly to σ_T^2 in (7), with c replaced by c_1, and

$$\mu_\lambda=-2c_1\lambda h_1,\quad \sigma_\lambda^2=4\lambda^2c_1^2c_2\left\{\mathbb{E}\left(\frac{w_{11}}{\mu}-1\right)^4-\lambda^2+h_2-2\lambda h_1\right\}.$$

Here, h_1 and h_2 are defined in Lemma 5.3.

The detailed proof of this theorem is postponed to the supplementary material. Based on Theorem 3.2, the procedure for testing H_0 is as follows:

$$\text{Reject } H_0 \text{ if } p(\widetilde{T}-\hat{\mu}_T)-\hat{\mu}_\lambda > z_\alpha\sqrt{\hat{\sigma}_{\widetilde{T}}^2+\hat{\sigma}_\lambda^2}, \tag{8}$$

where z_α represents the upper α-quantile of the standard normal distribution corresponding to the nominal level α. The quantities $\hat{\mu}_\lambda$, $\hat{\sigma}_{\widetilde{T}}^2$, and $\hat{\sigma}_\lambda^2$ are finite-sample estimates of μ_λ, $\sigma_{\widetilde{T}}^2$, and σ_λ^2, respectively. These estimates are obtained by substituting λ, $\mathbb{E}(w_{11}/\mu)^3$, and $\mathbb{E}(w_{11}/\mu-1)^4$ with their corresponding empirical estimates as previously defined.

3.2 Power analysis and discussion

In this section, we conduct a power analysis for a specific alternative hypothesis in which $\mathrm{Cov}(\mathbf{w})$ remains diagonal but has non-identical entries. We also briefly discuss the power of the test under a non-diagonal alternative hypothesis using an example.

Consider the following notations:

$$\mathbf{W}_{n\times p}=(w_{ij}),\quad \mathbf{X}_{n\times p}=(x_{ij}),\quad x_{ij}=\frac{w_{ij}}{\sum_{k=1}^p w_{ik}},\quad \mathbf{V}_{n\times p}=(v_{ij}),\quad \mathbf{Y}_{n\times p}=(y_{ij}),\quad y_{ij}=\frac{v_{ij}}{\sum_{k=1}^p v_{ik}}.$$

We assume that $\mathbf{V}=\mathbf{W}\Sigma$, where w_{ij} i.i.d. (μ,σ^2), where Σ is positive definite and normalized by $\mathrm{tr}(\Sigma)=p$. This normalization is without loss of generality, as $\mathbf{Y}$ is invariant under scaling of the basis data $\mathbf{V}$. The matrices $\mathbf{W}$ and $\mathbf{X}$ are for the null hypothesis, and $\mathbf{V}$ and $\mathbf{Y}$ are for the alternative hypothesis. Define the rescaled sample covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$ as follows:

$$\mathbf{B}_0=\frac{p^2}{n}\mathbf{X}'\mathbf{C}_n\mathbf{X},\qquad \mathbf{B}_1=\frac{p^2}{n}\mathbf{Y}'\mathbf{C}_n\mathbf{Y},\qquad \mathbf{C}_n=\mathbf{I}_n-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n'.$$

To guarantee that the proposed method have good power, we need to show that

$$\mathbb{E}\|\mathbf{B}_1-\mathbf{B}\|_F^2 > \mathbb{E}\|\mathbf{B}_0-\mathbf{B}\|_F^2. \tag{9}$$

We prove this inequality under the specific alternative hypothesis where Σ remains diagonal but has non-identical entries. Specifically, the matrix Σ has the form $\Sigma=\mathrm{diag}(d_1,\ldots,d_p)$ with the normalization $\sum_{k=1}^p d_k=p$. By detailed calculations (see Supplementary Material), we have

$$\mathbb{E}\|\mathbf{B}_0-\mathbf{B}\|_F^2=\frac{\lambda^2p^2}{n}+o(p),\quad \mathbb{E}\|\mathbf{B}_1-\mathbf{B}\|_F^2=\frac{\lambda^2(\sum_{k=1}^p d_k^2)^2}{n}+O(p). \tag{10}$$

By the Cauchy-Schwarz inequality, we have $\sum_{k=1}^p d_k^2\geq p$, and $\sum_{k=1}^p d_k^2=p$ only holds when $d_1=\cdots=d_p=1$. Therefore, inequality (9) holds true when $\{d_k\}_{k=1}^p$ are different.

In the following example, we provide a power analysis for a specific alternative where Σ is non-diagonal.

Example 1. Suppose that $w_{ij}\overset{\text{i.i.d.}}{\sim}\mathrm{Exp}(1)$ and Σ has the following form

$$\Sigma=\begin{pmatrix}1 & & & \alpha\\ \alpha & 1 & & \\ & \ddots & \ddots & \\ & & \alpha & 1\end{pmatrix},\qquad \alpha>-1.$$

In this case, we can derive that (see Supplementary Material for details)

$$\mathbb{E}\|\mathbf{B}_0 - \mathbf{B}\|_F^2 = \frac{p^2}{n} + o(p),$$
$$\mathbb{E}\|\mathbf{B}_1 - \mathbf{B}\|_F^2 = \frac{(\alpha^2+1)^2 p^2}{(\alpha+1)^4 n} + \frac{6\alpha^2 p^3}{(1+\alpha)^4 (p+1)^2} + o(p).$$

Hence, the inequality (9) holds if and only if

$$n > \frac{2(\alpha^2+\alpha+1)}{3\alpha} \frac{(p+1)^2}{p}.$$

Note that the RHS of this inequality serves as a critical threshold depending on both α and p. When $-1 < \alpha < 0$, the threshold is negative, allowing the proposed test to detect non-diagonal covariance structure for any sample size n. In contrast, when $\alpha > 0$, the proposed test fails to effectively detect such structure unless n exceeds the critical threshold.

The main challenge in a general theoretical power analysis arises from the complexity of the transformation from $\mathbf{w}$ to $\mathbf{x}$. If $\mathbf{w}$ were directly observable, changes in $\mathrm{Cov}(\mathbf{w})$ would be reflected directly in the LSS of $\mathbf{w}$, resulting in strong testing power. However, because we only observe $\mathbf{x}$, the normalization can weaken or even eliminate the signal from $\mathrm{Cov}(\mathbf{w})$. As shown in Remark 4, different $\mathbf{w}$ can produce identical $\mathrm{Cov}(\mathbf{x})$, causing the test statistic based on the LSS of $\mathbf{x}$ to lose power. This effect is challenging to characterize precisely, making a general power analysis very difficult.

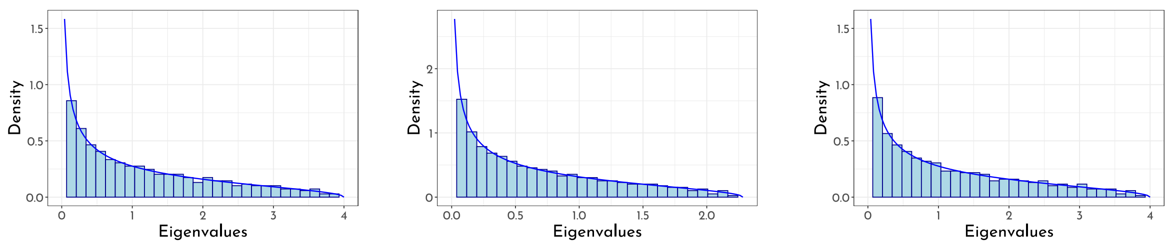
4 Numerical experiments

4.1 Limiting spectral distribution

In this section, simulation experiments are conducted to verify the LSD of the sample covariance matrix $\mathbf{B}_{p,N}$ from compositional data, as stated in Theorem 2.3. Compositional data $\{x_{ij}\}_{1\le i\le n, 1\le j\le p}$ is generated by the normalization $x_{ij} = w_{ij}/\sum_{\ell=1}^{p} w_{i\ell}$. We generate basis data w_{ij} from three populations, drawing histograms of eigenvalues of $\mathbf{B}_{p,N}$ and comparing them with theoretical densities. Specifically, three types of distributions for w_{ij} are considered:

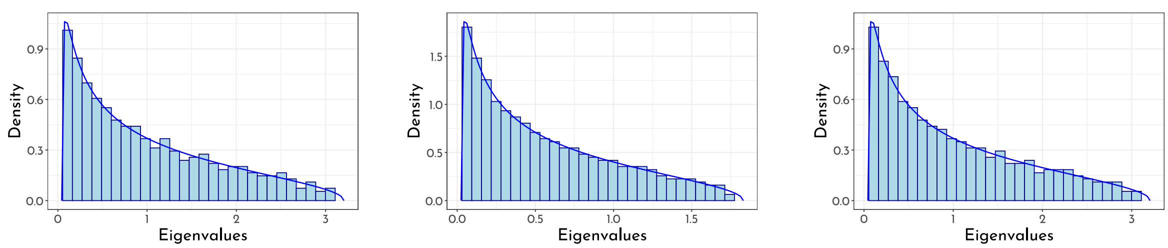
1. w_{ij} follows the exponential distribution with rate parameter 5;
2. w_{ij} follows the truncated standard normal distribution lying within the interval $(0, 10)$, denoted by $\mathrm{TN}(0, 1; 0, 10)$, where the first two parameters (0 and 1) represent the mean and variance of the standard normal distribution;
3. w_{ij} follows the Poisson distribution with parameter 10.

The dimension and sample size pair, (p, n), is set to $(500, 500)$ or $(500, 800)$. We display histograms of eigenvalues of $\mathbf{B}_{p,N}$ generated by three populations under various (p, n) combinations and compare them with their respective limiting densities in Figures 1 – 2. The figures reveal that all histograms align with their theoretical limits, affirming the accuracy of our theoretical results.



(a) Exponential(5) (b) TN(0,1;0,10) (c) Poisson(10)

Figure 1: Histograms of sample eigenvalues of $\mathbf{B}_{p,N}$ with $(p,n) = (500,500)$. The curves are density functions of their corresponding limiting spectral distribution.



(a) Exponential(5) (b) TN(0,1;0,10) (c) Poisson(10)

Figure 2: Histograms of sample eigenvalues of $\mathbf{B}_{p,N}$ with $(p,n) = (500,800)$. The curves are density functions of their corresponding limiting spectral distribution.

4.2 CLT for LSS

In this section, we implement some simulation studies to examine finite-sample properties of some LSS for $\mathbf{B}_{p,N}$ by comparing their empirical means and variances with theoretical limiting values, as stated in Corollary 2.6.

First, we compare the empirical mean and variance of $G_{p,N}(x^r)$, $r = 1,2,3$, with their corresponding theoretical limits in Corollary 2.6. Two types of data distribution of w_{ij} are consider:

1. w_{ij} follows the exponential distribution with rate parameter 5;
2. w_{ij} follows the Chi-squared distribution with degree of freedom 1.

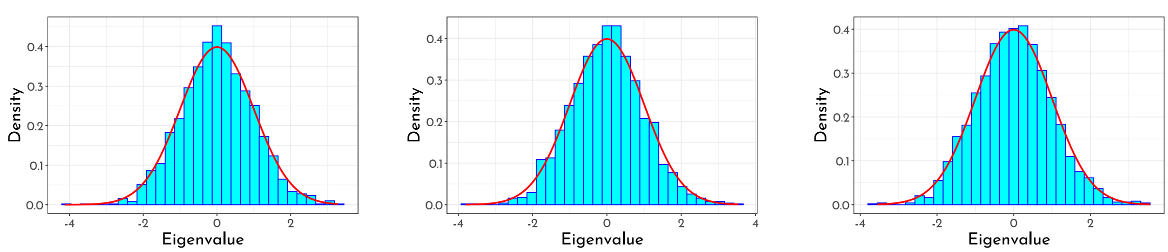
Empirical mean and variance of $G_{p,N}(x^r)$, are calculated for various combinations of (p,n) with $p/n = 3/4$ or $p/n = 1$. For each pair of (p,n), 2000 independent replications are used to obtain the empirical values. Tables 1 – 2 report the empirical results for Exp(5) population and $\chi^2(1)$ population, respectively. As shown in Tables 1 – 2, the empirical mean and variance of $G_{p,N}(x^r)$ closely match their respective theoretical limits under all scenarios. To verify the asymptotic normality of LSS, we draw the histogram of normalized LSS, $\overline{G}_{p,N}(x^r) = \{G_{p,N}(x^r) - \mu_r\}/\sqrt{V_r}$, $r = 1,2,3$, where μ_r and V_r are defined in Corollary 2.6, and compare them with the standard normal density. Figures 3 and 4 depict the histograms of $\overline{G}_{p,N}(x^r)$ for Exp(5) population with $p/n = 1$ and $\chi^2(1)$ population with $p/n = 3/4$, respectively. The histograms for the cases of Exp(5) population with $p/n = 3/4$ and $\chi^2(1)$ population with $p/n = 1$ exhibit similar patterns and are omitted for brevity. It can be seen from Figures 3 – 4 that all the histograms conform to the standard normal density, which fully supports our theoretical results.

Table 1: Empirical mean and variance of $G_{p,N}(x^r)$, $r = 1, 2, 3$, with $w_{ij} \sim \mathrm{Exp}(5)$.

	p/n	n	$G_{p,N}(x)$ mean	$G_{p,N}(x)$ var	$G_{p,N}(x^2)$ mean	$G_{p,N}(x^2)$ var	$G_{p,N}(x^3)$ mean	$G_{p,N}(x^3)$ var
Emp	3/4	100	-2.01	2.63	-4	36.54	-7.82	463.32
		200	-1.99	2.93	-3.85	39.73	-7.23	485.05
		300	-1.93	3.03	-3.57	40.3	-6.32	483.76
		400	-2.04	2.95	-3.98	38.78	-7.67	460.01
Theo			**-2**	**3**	**-3.75**	**39**	**-6.81**	**457**
Emp	1	100	-1.91	3.61	-3.83	64.09	-6.56	1064.75
		200	-1.96	3.89	-3.96	68.37	-6.91	1090.14
		300	-2.01	3.97	-4.06	68.7	-7.16	1082.72
		400	-1.98	3.71	-3.99	64.22	-7.07	1010.09
Theo			**-2**	**4**	**-4**	**68**	**-7**	**1050**

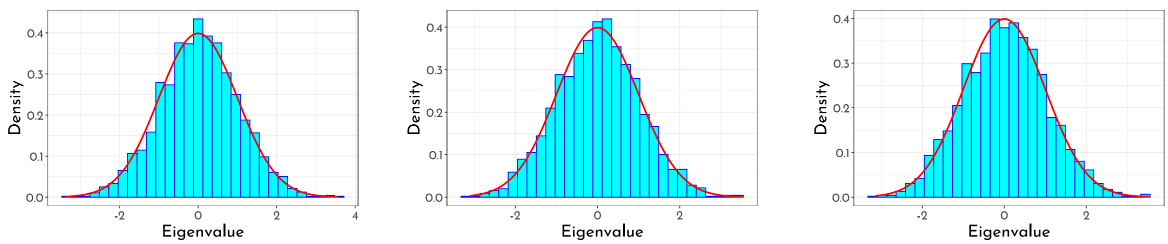
Table 2: Empirical mean and variance of $G_{p,N}(x^r)$, $r = 1, 2, 3$, with $w_{ij} \sim \chi^2(1)$.

	p/n	n	$G_{p,N}(x)$ mean	$G_{p,N}(x)$ var	$G_{p,N}(x^2)$ mean	$G_{p,N}(x^2)$ var	$G_{p,N}(x^3)$ mean	$G_{p,N}(x^3)$ var
Emp	3/4	100	-5.79	15.53	-24.19	888.99	-97.31	46790.03
		200	-5.96	16.74	-24.39	920.63	-96.17	45375.75
		300	-5.94	16.6	-23.75	882.92	-90.59	42487.68
		400	-5.88	17.51	-22.68	912.28	-81.2	42922.06
Theo			**-6**	**18**	**-23**	**918**	**-83**	**41806.12**
Emp	1	100	-5.92	20.81	-26.15	1563.02	-102.73	107846.2
		200	-5.98	23.01	-25.15	1639.95	-90.25	105467.9
		300	-5.81	21.82	-23.16	1526.34	-74.54	96864.11
		400	-6.13	23.18	-25.41	1599.96	-90.31	99475.82
Theo			**-6**	**24**	**-24**	**1600**	**-80**	**96000**



(a) $\overline{G}_{p,N}(x)$ (b) $\overline{G}_{p,N}(x^2)$ (c) $\overline{G}_{p,N}(x^3)$

Figure 3: Histograms of normalized LSS $\overline{G}_{p,N}(x^r)$, $r = 1, 2, 3$, with $w_{ij} \sim \mathrm{Exp}(5)$ and $p = n = 400$. The curves are density functions of the standard normal distribution.



(a) $\overline{G}_{p,N}(x)$ (b) $\overline{G}_{p,N}(x^2)$ (c) $\overline{G}_{p,N}(x^3)$

Figure 4: Histograms of normalized LSS $\overline{G}_{p,N}(x^r)$, $r = 1, 2, 3$, with $w_{ij} \sim \chi^2(1)$ and $(p, n) = (300, 400)$. The curves are density functions of the standard normal distribution.

4.3 Covariance testing for basis data

We conduct numerical simulations to determine the empirical size and power of the proposed test statistic. The nominal significant level is set to be $\alpha = 0.05$. To evaluate the finite-sample performance, we consider a range of (p, n) settings and perform 2000 independent replications for each. To assess empirical size, we consider two scenarios for the basis data matrix $\mathbf{W}_n = (w_{ij})_{n\times p}$:

- Exponential model: $w_{ij} \overset{\text{i.i.d.}}{\sim} \text{Exp}(\text{rate} = 5)$;
- Chi-squared model: $w_{ij} \overset{\text{i.i.d.}}{\sim} \chi^2(1)$.

To examine empirical power, we consider two types of alternatives. For both, an $n \times p$ matrix $\mathbf{Z}_n$ with i.i.d. entries (from either distribution above) is first generated, and then the data matrix is constructed as $\mathbf{W}_n = \mathbf{Z}_n\Sigma$, with two choices of the matrix Σ:

- $\Sigma = \Sigma_1$: The first diagonal entry is ψ_1, the other diagonal entries are 1, and all other entries are zero, with $\psi_1 \in \{1, 3, 3.5, 4, 4.5\}$.
- $\Sigma = \Sigma_2$: Ones on the main diagonal, ψ_2 on the first subdiagonal (below the diagonal), and zeros elsewhere, with $\psi_2 \in \{0, -0.2, -0.25, -0.3, -0.35\}$.

Simulation results are reported in Tables 3 – 4. We conduct simulations under both the estimated and known distribution scenarios for $\{w_{ij}\}$, with the latter (benchmark) results reported in parentheses. In the unknown distribution scenario, the total sample is evenly split, with $n_1 = n_2 = n/2$, where n_2 is used for parameter estimation and n_1 for computing the test statistic.

The empirical power is close to the significance level $\alpha = 0.05$ when (p, n) is large. The empirical power increases as the strength of ψ_1 and ψ_2 increases, and it also grows with larger values of (p, n). The power based on estimated parameters is generally lower than that using the true parameters, which is due to the limited precision of the parameter estimation. These results confirm the theoretical properties of the test and demonstrate its effectiveness in detecting alternatives.

5 Proof of Theorem 2.5

In this section, we first present the difference between the CLT for centralized sample covariance $\mathbf{B}_p^0$ and unbiased sample covariance $\mathbf{B}_{p,N}$ by substitution principle in Section 5.1, where

$$\mathbf{B}_p^0 = p^2\mathbf{S}_n^0 = \frac{p^2}{n}(\mathbf{X}_n - \mathbb{E}\mathbf{X}_n)'(\mathbf{X}_n - \mathbb{E}\mathbf{X}_n) = \frac{1}{n}\mathbf{Y}_n'\mathbf{Y}_n,$$

Table 3: Empirical size and power (2000 replications) for Σ_1; values in parentheses are based on known $\{w_{ij}\}$ distribution and true parameters are used.

				Size	Power			
c	p	n_1	n_2	$\psi_1 = 1$	$\psi_1 = 3$	$\psi_1 = 3.5$	$\psi_1 = 4$	$\psi_1 = 4.5$
				Exp(5) model				
0.5	150	150	150	0.0715 (0.0605)	0.829 (1.0)	0.9175 (1.0)	0.953 (1.0)	0.9765 (1.0)
	300	300	300	0.068 (0.041)	0.9615 (1.0)	0.996 (1.0)	0.999 (1.0)	1.0 (1.0)
	450	450	450	0.0515 (0.055)	0.9895 (1.0)	1.0 (1.0)	1.0 (1.0)	1.0 (1.0)
	600	600	600	0.046 (0.0535)	0.995 (1.0)	1.0 (1.0)	1.0 (1.0)	1.0 (1.0)
1	150	75	75	0.087 (0.052)	0.401 (0.9995)	0.5215 (1.0)	0.613 (1.0)	0.675 (1.0)
	300	150	150	0.0585 (0.052)	0.4765 (1.0)	0.682 (1.0)	0.8205 (1.0)	0.8915 (1.0)
	450	225	225	0.0625 (0.0635)	0.5635 (1.0)	0.787 (1.0)	0.9115 (1.0)	0.9585 (1.0)
	600	300	300	0.0625 (0.0625)	0.6095 (1.0)	0.8525 (1.0)	0.953 (1.0)	0.9885 (1.0)
1.5	150	50	50	0.09 (0.045)	0.241 (0.978)	0.3295 (0.9985)	0.3925 (1.0)	0.443 (1.0)
	300	100	100	0.073 (0.047)	0.2705 (0.995)	0.3975 (1.0)	0.519 (1.0)	0.62 (1.0)
	450	150	150	0.067 (0.053)	0.291 (0.9995)	0.467 (1.0)	0.645 (1.0)	0.771 (1.0)
	600	200	200	0.06 (0.053)	0.306 (1.0)	0.5195 (1.0)	0.7095 (1.0)	0.8275 (1.0)
				$\chi^2(1)$ model				
0.5	150	150	150	0.082 (0.0535)	0.761 (1.0)	0.882 (1.0)	0.935 (1.0)	0.964 (1.0)
	300	300	300	0.0755 (0.0505)	0.9075 (1.0)	0.9765 (1.0)	0.9935 (1.0)	0.997 (1.0)
	450	450	450	0.0675 (0.0505)	0.967 (1.0)	0.998 (1.0)	1.0 (1.0)	1.0 (1.0)
	600	600	600	0.0555 (0.0545)	0.982 (1.0)	1.0 (1.0)	1.0 (1.0)	1.0 (1.0)
1	150	75	75	0.0865 (0.0605)	0.41 (0.984)	0.5315 (1.0)	0.615 (1.0)	0.6915 (1.0)
	300	150	150	0.069 (0.0525)	0.486 (0.999)	0.661 (1.0)	0.7815 (1.0)	0.86 (1.0)
	450	225	225	0.0705 (0.0425)	0.5485 (1.0)	0.7615 (1.0)	0.8745 (1.0)	0.9315 (1.0)
	600	300	300	0.065 (0.047)	0.586 (1.0)	0.8015 (1.0)	0.924 (1.0)	0.965 (1.0)
1.5	150	50	50	0.0895 (0.0545)	0.2745 (0.876)	0.3595 (0.9695)	0.438 (0.9915)	0.5005 (0.9985)
	300	100	100	0.068 (0.0455)	0.2935 (0.9595)	0.4265 (0.998)	0.543 (1.0)	0.645 (1.0)
	450	150	150	0.0635 (0.0575)	0.333 (0.9875)	0.4885 (0.999)	0.6255 (1.0)	0.7345 (1.0)
	600	200	200	0.0595 (0.0535)	0.3265 (0.9905)	0.5275 (1.0)	0.686 (1.0)	0.7935 (1.0)

Table 4: Empirical size and power (2000 replications) for Σ_2; values in parentheses are based on known $\{w_{ij}\}$ distribution and true parameters are used.

				Size	Power			
c	p	n_1	n_2	$\psi_2 = 0$	$\psi_2 = -0.2$	$\psi_2 = -0.25$	$\psi_2 = -0.3$	$\psi_2 = -0.35$
				Exp(5) model				
0.5	150	150	150	0.0715	0.5095	0.73	0.887	0.9575
				(0.0605)	(1.0)	(1.0)	(1.0)	(1.0)
	300	300	300	0.068	0.958	0.9995	1.0	1.0
				(0.041)	(1.0)	(1.0)	(1.0)	(1.0)
	450	450	450	0.0515	0.9995	1.0	1.0	1.0
				(0.055)	(1.0)	(1.0)	(1.0)	(1.0)
	600	600	600	0.046	1.0	1.0	1.0	1.0
				(0.0535)	(1.0)	(1.0)	(1.0)	(1.0)
1	150	75	75	0.087	0.1795	0.2315	0.2615	0.28
				(0.052)	(1.0)	(1.0)	(1.0)	(1.0)
	300	150	150	0.0585	0.334	0.495	0.678	0.79
				(0.052)	(1.0)	(1.0)	(1.0)	(1.0)
	450	225	225	0.0625	0.5785	0.836	0.959	0.9915
				(0.0635)	(1.0)	(1.0)	(1.0)	(1.0)
	600	300	300	0.0625	0.8055	0.9695	0.9985	1.0
				(0.0625)	(1.0)	(1.0)	(1.0)	(1.0)
1.5	150	50	50	0.09	0.112	0.1215	0.1245	0.113
				(0.045)	(1.0)	(1.0)	(1.0)	(1.0)
	300	100	100	0.073	0.176	0.232	0.2885	0.325
				(0.047)	(1.0)	(1.0)	(1.0)	(1.0)
	450	150	150	0.067	0.269	0.407	0.5415	0.6525
				(0.053)	(1.0)	(1.0)	(1.0)	(1.0)
	600	200	200	0.06	0.393	0.6005	0.7795	0.8925
				(0.053)	(1.0)	(1.0)	(1.0)	(1.0)
				$\chi^2(1)$ model				
0.5	150	150	150	0.082	0.3895	0.553	0.687	0.765
				(0.0535)	(1.0)	(1.0)	(1.0)	(1.0)
	300	300	300	0.0755	0.845	0.9815	0.999	1.0
				(0.0505)	(1.0)	(1.0)	(1.0)	(1.0)
	450	450	450	0.0675	0.9955	1.0	1.0	1.0
				(0.0505)	(1.0)	(1.0)	(1.0)	(1.0)
	600	600	600	0.0555	1.0	1.0	1.0	1.0
				(0.0545)	(1.0)	(1.0)	(1.0)	(1.0)
1	150	75	75	0.0865	0.1375	0.1525	0.1475	0.127
				(0.0605)	(1.0)	(1.0)	(1.0)	(1.0)
	300	150	150	0.069	0.246	0.347	0.4395	0.4925
				(0.0525)	(1.0)	(1.0)	(1.0)	(1.0)
	450	225	225	0.0705	0.4275	0.653	0.815	0.896
				(0.0425)	(1.0)	(1.0)	(1.0)	(1.0)
	600	300	300	0.065	0.617	0.8715	0.9745	0.9975
				(0.047)	(1.0)	(1.0)	(1.0)	(1.0)
1.5	150	50	50	0.0895	0.1035	0.098	0.083	0.0625
				(0.0545)	(1.0)	(1.0)	(1.0)	(1.0)
	300	100	100	0.068	0.127	0.153	0.166	0.153
				(0.0455)	(1.0)	(1.0)	(1.0)	(1.0)
	450	150	150	0.0635	0.1915	0.263	0.329	0.363
				(0.0575)	(1.0)	(1.0)	(1.0)	(1.0)
	600	200	200	0.0595	0.284	0.4325	0.5645	0.6595
				(0.0535)	(1.0)	(1.0)	(1.0)	(1.0)

$$\mathbf{B}_{p,N} = p^2 \mathbf{S}_{n,N} = \frac{p^2}{N} \mathbf{X}_n' \mathbf{C}_n \mathbf{X}_n,$$

and $\mathbf{Y}_n = (y_{ij})_{n \times p}$, $y_{ij} = \frac{w_{ij}}{\overline{w}_i} - 1$ and $\overline{w}_i = \frac{1}{p}\sum_{\ell=1}^{p} w_{i\ell}$. By substituting the adjusted sample size $N = n-1$ for the actual sample size n in the centering term, the unbiased sample covariance matrix $\mathbf{B}_{p,N}$ and the centralized sample covariance $\mathbf{B}_p^0$ share the same CLT (see, Section 5.1). The general strategy of the main proof of Theorem 2.5 is explained in the following and three major steps of the general strategy are presented in Section 5.3.

The general strategy of the proof follows the method established in Bai and Silverstein [2004] and Gao et al. [2017], with necessary adjustments for handling the sample covariance matrix of HCD, where conventional tools are not directly applicable. Our novel techniques play a pivotal role in overcoming these challenges. To begin with, we follow the strategy in Jiang [2004] to establish the LSD of $\mathbf{B}_{p,N}$ in Theorem 2.3. Then, we develop Proposition 2.4 to find the extreme eigenvalues of $\mathbf{B}_{p,N}$. Notably, these extreme eigenvalues are highly concentrated around two edges of the support, a crucial aspect for applying the Cauchy integral formula (11) and proving tightness. Given that compositional data $x_{ij} = w_{ij}/\sum_{\ell=1}^{p} w_{i\ell}$ are not i.i.d., dealing with the CLT for LSS of the unbiased sample covariance matrix $\mathbf{B}_{p,N}$ presents challenges. To address this, we employ the substitution principle [Zheng et al., 2015] to reduce the problem to the CLT for LSS of the centralized sample covariance $\mathbf{B}_p^0$. By substituting the adjusted sample size $N = n-1$ for the actual sample size n in the centering term, both the unbiased sample covariance matrix $\mathbf{B}_{p,N}$ and the centralized sample covariance $\mathbf{B}_p^0$ share the same CLT (see Section 5.1). We then leverage the independence of samples to further study the CLT for LSS of $\mathbf{B}_p^0$. Specifically, we exploit the independence of samples to establish independence for $\mathbf{r}_i = \frac{1}{\sqrt{n}}\left(\frac{w_{i1}}{\overline{w}_i} - 1, \ldots, \frac{w_{ip}}{\overline{w}_i} - 1\right)'$, $\overline{w}_i = p^{-1}\sum_{\ell=1}^{p} w_{i\ell}$, $i = 1, 2, \ldots, n$, and express $\mathbf{B}_p^0$ as $\mathbf{B}_p^0 = \frac{1}{n}\mathbf{Y}_n'\mathbf{Y}_n = \sum_{i=1}^{n} \mathbf{r}_i \mathbf{r}_i'$. The ultimate goal is to establish the CLT for LSS of $\mathbf{B}_p^0$.

By the Cauchy integral formula, we have

$$\int f(x)\,\mathrm{d}G(x) = -\frac{1}{2\pi i} \oint_{\mathcal{C}} f(z) m_G(z)\,\mathrm{d}z \tag{11}$$

valid for any cumulative distribution function G and any analytic function f on an open set containing the support of G, where $\oint_{\mathcal{C}}$ is the contour integration in the anti-clockwise direction. In our case, $G(x) := G_p^0(x) := p\{F^{\mathbf{B}_p^0}(x) - F^{c_n}(x)\}$, $c_n = p/n$. Therefore, the problem of finding the limiting distribution reduces to the study of $M_p(z)$ defined as follows:

$$\begin{aligned}
M_p(z) &= p\Big\{m_p(z) - m_p^0(z)\Big\} = n\Big\{\underline{m}_p(z) - \underline{m}_p^0(z)\Big\}, \\
m_p(z) &= m_{F^{\mathbf{B}_p^0}}(z) = \frac{1}{p}\mathrm{tr}(\mathbf{B}_p^0 - z\mathbf{I}_p)^{-1}, \qquad m_p^0(z) = m_{F^{c_n}}(z), \\
\underline{m}_p(z) &= m_{F^{\underline{\mathbf{B}}_p^0}}(z) = \frac{1}{p}\mathrm{tr}(\underline{\mathbf{B}}_p^0 - zI_n)^{-1}, \qquad \underline{m}_p^0(z) = \underline{m}_{F^{c_n}}(z), \\
\underline{\mathbf{B}}_p^0 &= \frac{p^2}{n}(\mathbf{X}_n - \mathbb{E}\mathbf{X}_n)(\mathbf{X}_n - \mathbb{E}\mathbf{X}_n)'.
\end{aligned}$$

Note that the support of $F^{\mathbf{B}_{p,N}}$ is random. Fortunately, we have shown that the extreme eigenvalues of $\mathbf{B}_{p,N}$ are highly concentrated around two edges of the support of the limiting MP law (see, Theorem 2.3, Proposition 2.4). Then the contour $\mathcal{C}$ can be appropriately chosen. Moreover, as in Bai and Silverstein [2004], by Proposition 2.4, we can replace the process $\{M_p(z), z \in \mathcal{C}\}$ by a slightly modified process $\{\widehat{M}_p(z), z \in \mathcal{C}\}$. Below we present the definitions of the contour $\mathcal{C}$ and the modified process $\widehat{M}_p(z)$. Let $v_0 > 0$ be arbitrary. Let x_r be any number greater than $\lambda(1+\sqrt{c})^2$. Let x_l be any negative number if $\lambda(1-\sqrt{c})^2 = 0$. Otherwise we choose $x_l \in (0, \lambda(1-\sqrt{c})^2)$. Now let $\mathcal{C}_u = \{x + iv_0 : x \in [x_l, x_r]\}$. Then we define $\mathcal{C}^+ := \{x_l + iv : v \in [0, v_0]\} \cup \mathcal{C}_u \cup \{x_r + iv : v \in [0, v_0]\}$, and $\mathcal{C} = \mathcal{C}^+ \cup \overline{\mathcal{C}^+}$. Now we define the subsets $\mathcal{C}_n$ of $\mathcal{C}$ on which $M_p(\cdot)$ equals to $\widehat{M}_p(\cdot)$. Choose sequence $\{\varepsilon_n\}$ decreasing to zero satisfying for some $\alpha \in (0,1)$, $\varepsilon_n \geq n^{-\alpha}$. Let

$$\mathcal{C}_l = \begin{cases} \{x_l + iv : v \in [n^{-1}\varepsilon_n, v_0]\}, & \text{if } x_l > 0, \\ \{x_l + iv : v \in [0, v_0]\}, & \text{if } x_l < 0, \end{cases}$$

and $\mathcal{C}_r = \{x_r + iv : v \in [n^{-1}\varepsilon, v_0]\}$ for any $v_0 > 0$. Then $\mathcal{C}_n = \mathcal{C}_l \cup \mathcal{C}_u \cup \mathcal{C}_r$. For $z = x + iv$, we define

$$\widehat{M}_p(z) = \begin{cases} M_p(z), & \text{for } z \in \mathcal{C}_n \\ M_p(x_r + in^{-1}\varepsilon_n), & \text{for } x = x_r, v \in [0, n^{-1}\varepsilon_n], \text{ and if } x_l > 0 \\ M_p(x_l + in^{-1}\varepsilon_n), & \text{for } x = x_l, v \in [0, n^{-1}\varepsilon_n], \end{cases}$$

Most of the paper will deal with proving the following proposition.

Proposition 5.1. *Under the same notations and assumptions as in Theorem 2.5, $\widehat{M}_p(z)$ converges weakly to a two-dimensional Gaussian process $M(z)$ for $z \in \mathcal{C}$, with mean*

$$\begin{aligned} \mathbb{E}M(z) = \frac{\underline{m}^2(z)\{1+\lambda\underline{m}(z)\}}{\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)} \Bigg[&\{z(h_1+\lambda)m(z)+\lambda\} \\ &+ cz^2\underline{m}(z)\Big\{(\alpha_1+\alpha_2)m^2(z) + 2\lambda^2 m'(z)\Big\} - \frac{c\lambda^2\underline{m}(z)}{\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)}\Bigg], \end{aligned} \tag{12}$$

and covariance function

$$\begin{aligned} &\mathrm{Cov}(M(z_1), M(z_2)) \\ &= 2\left[\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{\{\underline{m}(z_1)-\underline{m}(z_2)\}^2} - \frac{1}{(z_1-z_2)^2}\right] + \frac{c(\alpha_1+\alpha_2)\underline{m}'(z_1)\underline{m}'(z_2)}{\{1+\lambda\underline{m}(z_1)\}^2\{1+\lambda\underline{m}(z_2)\}^2}. \end{aligned} \tag{13}$$

Now we explain how Theorem 2.5 follows from the above proposition. As in Bai and Silverstein [2004], with probability one, $|\int f(z)\{M_p(z) - \widehat{M}_p(z)\}\,\mathrm{d}z| \to 0$ as $n \to \infty$. Combining this observation with (11), Theorem 2.5 follows from Proposition 5.1. To prove Proposition 5.1, we decompose $M_p(z)$ into a random part $M_p^{(1)}(z)$ and a deterministic part $M_p^{(2)}(z)$ for $z \in \mathcal{C}_n$, that is, $M_p(z) = M_p^{(1)}(z) + M_p^{(2)}(z)$, where

$$M_p^{(1)}(z) = p\Big\{m_p(z) - \mathbb{E}m_p(z)\Big\} \quad \text{and} \quad M_p^{(2)}(z) = p\Big\{\mathbb{E}m_p(z) - m_p^0(z)\Big\}.$$

The random part contributes to the covariance function and the deterministic part contributes to the mean function. By Theorem 8.1 in Billingsley [1968], the proof of Proposition 5.1 is then complete if we can verify the following three steps:

Step 1 Finite-dimensional convergence of $M_p^{(1)}(z)$ in distribution on $\mathcal{C}_n$ to a centered multivariate Gaussian random vector with covariance function given by (13).

Step 2 Tightness of the $M_p^{(1)}(z)$ for $z \in \mathcal{C}_n$.

Step 3 Convergence of the non-random part $M_p^{(2)}(z)$ to (12) on $z \in \mathcal{C}_n$.

The proof of these steps is presented in the coming sections. Before that, we introduce the substitution principle, truncation and crucial lemmas in Sections 5.1 and 5.2 respectively. The former explains the reduction of problem of the CLT for LSS of $\mathbf{B}_{p,N}$ to that of $\mathbf{B}_p^0$, while the latter provides truncation and essential lemmas for these three steps in proving the CLT for LSS of $\mathbf{B}_p^0$.

5.1 Substitution principle

By the Cauchy integral formula, we have

$$G_{p,N}(f) = -\frac{1}{2\pi i} \oint_{\mathcal{C}} f(z) \Big\{ \operatorname{tr}(\mathbf{B}_{p,N} - z\mathbf{I}_p)^{-1} - p m_N^0(z) \Big\} \mathrm{d}z$$

valid for any function f analytic on an open set containing the support of $p\{F^{\mathbf{B}_{p,N}}(x) - F^{c_N}(x)\}$, where

$$m_N^0(z) \equiv m_{F^{c_N}}(z) = \frac{1}{\lambda(1 - c_N - c_N z m_N^0) - z},$$

with $c_N = p/N$. To obtain the asymptotic distribution of $G_{p,N}(f)$, it is necessary to find the asymptotic distribution of $\operatorname{tr}(\mathbf{B}_{p,N} - z\mathbf{I}_p)^{-1} - p m_N^0(z)$. To achieve this, we derive the following Lemma 5.2 whose proof is postponed to the supplementary material.

Lemma 5.2. *Under Assumptions 2.1 and 2.2, as $n \to \infty$,*

$$\operatorname{tr}(\mathbf{B}_{p,N} - z\mathbf{I}_p)^{-1} - p m_N^0(z) = \operatorname{tr}(\mathbf{B}_p^0 - z\mathbf{I}_p)^{-1} - p m_n^0(z) + o_P(1).$$

By Lemma 5.2, the asymptotic distribution of $G_{p,N}(f)$ is identical to that of

$$G_p^0(f) = -\frac{1}{2\pi i} \oint_{\mathcal{C}} f(z) \Big\{ \operatorname{tr}(\mathbf{B}_p^0 - z\mathbf{I}_p)^{-1} - p m_n^0(z) \Big\} \mathrm{d}z,$$

where $c_n = p/n$, $m_n^0(z) = m_{F^{c_n}}(z)$ (note that we denote $m_n^0(z)$ as $m_p^0(z)$ in other sections except this subsection).

5.2 Truncation and some important lemmas

We begin the proof of Proposition 5.1 with the replacement of the entries of $\mathbf{W}_n$ with truncated variables. Next, we introduce three pivotal lemmas that are essential for proving Proposition 5.1.

5.2.1 Truncation

In the following, we will show that the limiting distribution for the LSS remains unchanged before and after truncation. Therefore, it suffices to derive the limiting distribution of the LSS after truncation. Specifically, we can choose a positive sequence of $\{\delta_n\}$ such that

$$\delta_n \to 0, \quad \delta_n n^{1/4} \to \infty, \quad \delta_n^{-4} \mathbb{E}|w_{11} - \mu|^4 I_{\{|w_{11} - \mu| \geq \delta_n \sqrt{n}\}} \to 0. \tag{14}$$

Let $\widehat{\mathbf{B}}_p^0 = \frac{p^2}{n}(\widehat{\mathbf{X}}_n - \mathbb{E}\widehat{\mathbf{X}}_n)'(\widehat{\mathbf{X}}_n - \mathbb{E}\widehat{\mathbf{X}}_n)$, where the (i,j)-th entry of $\widehat{\mathbf{X}}_n$ is normalized using truncated variables $\hat{w}_{ij} = w_{ij} I_{\{|w_{ij} - \mu| < \delta_n \sqrt{n}\}}$, as described in (1). We then have

$$\mathbb{P}(\mathbf{B}_p^0 \neq \widehat{\mathbf{B}}_p^0) \leq \mathbb{P}\Big(\cup_{i \leq n, j \leq p} \Big\{ |w_{ij} - \mu| \geq \delta_n \sqrt{n} \Big\}\Big) \leq np \cdot \mathbb{P}\Big(|w_{ij} - \mu| \geq \delta_n \sqrt{n}\Big)$$

$$\leq K\delta_n^{-4}\int_{\{|w_{11}-\mu|\geq\delta_n\sqrt{n}\}}|w_{11}-\mu|^4 = o(1).$$

Let $\widehat{G}_p^0(x)$ be $G_p^0(x) := p\{F^{\mathbf{B}_p^0}(x) - F^{c_n}(x)\}$ with $\mathbf{B}_p^0$ replaced by $\widehat{\mathbf{B}}_p^0$, then $\mathbb{P}(\widehat{G}_p^0(x) \neq G_p^0(x)) \leq o(1)$. In view of the above, we obtain

$$\int f_j(x)\,\mathrm{d}G_p^0(x) = \int f_j(x)\,\mathrm{d}\widehat{G}_p^0(x) + o_P(1).$$

Therefore, in the remaining part of the proof of the CLT for LSS, we assume that the underlying variables are truncated at $\delta_n\sqrt{n}$.

5.2.2 Some important Lemmas

In this section, we introduce three pivotal lemmas, which are crucial for proving the CLT for LSS. These lemmas represent novel contributions of this paper and unveil the concentration phenomena. Lemma 5.3 is crafted to estimate essential parameters, facilitating the derivation of estimates of any order. Concerning ν_2 and ν_{12} (see Lemma 5.3 for their definitions), the terms h_1/p and h_2/p emerge as non-negligible due to the multiplication by p in the CLT. To address these parameters, we establish that the probability of the event $B_p^c(\varepsilon)$ decays polynomially to 0 and leverage Taylor expansion on the event $B_p(\varepsilon) = \{\omega : |\overline{w}_i - \mu| \leq \varepsilon, \overline{w}_i = \sum_{j=1}^p w_{ij}/p\}$ to handle the issue of dependence. The proof of the CLT for LSS relies on two pivotal steps: the moment inequality for random quadratic forms and the precise estimation of the expectation of the product of two random quadratic forms. Lemma 5.4 establishes the former step, essential for converting them into the corresponding traces, while Lemma 5.5 establishes the latter step, enabling the application of CLT for martingale differences. The proof of Lemmas 5.3 – 5.5 are postponed to the supplementary material. In the following, we write $p \asymp n$ if and only if there exist constants $C_1, C_2 > 0$ such that $C_1 n \leq p \leq C_2 n$ for all sufficiently large n.

Lemma 5.3. *Suppose that $\mathbf{w} = \left(w_1,\ldots,w_p\right)'$ has positive i.i.d. entries with $\mathbb{E}w_1 = \mu > 0$, $\mathbb{E}(w_1-\mu)^2 = \sigma^2$, and $\mathbb{E}\left|w_1-\mu\right|^4 < \infty$, and that $p \asymp n$. Then for the truncated data $\hat{w}_j = w_j I_{\{|w_j-\mu|<\delta_n\sqrt{n}\}}$ $(j = 1,\ldots,p)$ where δ_n satisfies condition (14), we have*

$$\nu_2 := \mathbb{E}\left(\frac{\hat{w}_1}{\overline{\hat{w}}} - 1\right)^2 = \lambda + \frac{h_1}{p} + o(p^{-1}),$$
$$\nu_{12} := \mathbb{E}\left(\frac{\hat{w}_1}{\overline{\hat{w}}} - 1\right)^2\left(\frac{\hat{w}_2}{\overline{\hat{w}}} - 1\right)^2 = \lambda^2 + \frac{h_2}{p} + o(p^{-1}),$$
$$\nu_4 := \mathbb{E}\left(\frac{\hat{w}_1}{\overline{\hat{w}}} - 1\right)^4 = \mathbb{E}\left(\frac{w_1}{\mu} - 1\right)^4 + o(1),$$

where $\overline{\hat{w}} = p^{-1}\sum_{j=1}^p \hat{w}_j$, $\lambda = \sigma^2/\mu^2$, and

$$h_1 = -2\frac{\mathbb{E}w_1^3}{\mu^3} + 3\lambda^2 + 5\lambda + 2, \quad h_2 = -8\lambda\frac{\mathbb{E}w_1^3}{\mu^3} + 10\lambda^3 + 22\lambda^2 + 8\lambda.$$

Lemma 5.4. *Suppose that $\mathbf{w} = \left(w_1,\ldots,w_p\right)'$ has positive i.i.d. entries with $\mathbb{E}w_1 = \mu > 0$ and $\mathbb{E}(w_1-\mu)^2 = \sigma^2$, and that $p \asymp n$. For any $p\times p$ matrix $\mathbf{A}$ and $q \geq 2$, then there is a positive constant K_q depending on q such that*

$$\mathbb{E}\left|\mathbf{r}'\mathbf{A}\mathbf{r} - \frac{1}{n}\mathbb{E}\left(w_1/\overline{w} - 1\right)^2 \operatorname{tr}\mathbf{A}\right|^q$$

$$\leq K_q\Big\{n^{-q}\Big[\Big\{\mathbb{E}\Big|w_1-\mu\Big|^4\mathrm{tr}(\mathbf{A}\mathbf{A}')\Big\}^{q/2}+\mathbb{E}|w_1-\mu|^{2q}\mathrm{tr}(\mathbf{A}\mathbf{A}')^{q/2}\Big]$$
$$+n^q\mathbb{P}\Big(B_p^c(\varepsilon)\Big)\|\mathbf{A}\|^q+|\mathbb{E}\,(w_1/\overline{w}-1)^2-\lambda|^q\|\mathbf{A}\|^q\Big\},$$

where $\mathbf{r}=n^{-1/2}(w_1/\overline{w}-1,\ldots,w_p/\overline{w}-1)'$, $\lambda=\sigma^2/\mu^2$, $B_p(\varepsilon)=\{\omega:|\overline{w}-\mu|\leq\varepsilon,\overline{w}=\sum_{j=1}^p w_j/p\}$, *and*

$$\mathbb{P}\Big(B_p^c(\varepsilon)\Big)\leq C_t\varepsilon^{-t}\Big\{p^{-t/2}(\mathbb{E}|w_1-\mu|^2)^{t/2}+p^{-t+1}\mathbb{E}\Big|w_1-\mu\Big|^t\Big\},\tag{15}$$

in which ε, t, C_t *(which depends on* t*) are positive constants. Furthermore, if* $\mathbb{E}\Big|w_1-\mu\Big|^4<\infty$ *and* $\|\mathbf{A}\|$ *is bounded, then for the truncated data* $\hat{w}_j=w_jI_{\{|w_j-\mu|<\delta_n\sqrt{n}\}}$ $(j=1,\ldots,p)$ *where* δ_n *satisfies condition* (14), *we have*

$$\mathbb{E}\Big|\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{1}{n}\mathbb{E}\Big(\hat{w}_1/\overline{\hat{w}}-1\Big)^2\mathrm{tr}\mathbf{A}\Big|^q\leq K_qn^{-1}\delta_n^{2q-4},\quad \textit{for any } q\geq 2,$$

where $\mathbf{r}=n^{-1/2}(\hat{w}_1/\overline{\hat{w}}-1,\ldots,\hat{w}_p/\overline{\hat{w}}-1)'$, *and* $\overline{\hat{w}}=p^{-1}\sum_{j=1}^p\hat{w}_j$.

Lemma 5.5. *Suppose that* $\mathbf{w}=\Big(w_1,\ldots,w_p\Big)'$ *has positive i.i.d. entries with* $\mathbb{E}w_1=\mu>0$ *and* $\mathbb{E}(w_1-\mu)^2=\sigma^2$, $p\asymp n$, $\mathbf{A}$ *and* $\mathbf{B}$ *are* $p\times p$ *matrices, if* $\mathbb{E}\Big|w_1-\mu\Big|^4<\infty$, $\|\mathbf{A}\|$ *and* $\|\mathbf{B}\|$ *are bounded, then for the truncated data* $\hat{w}_j=w_jI_{\{|w_j-\mu|<\delta_n\sqrt{n}\}}$ $(j=1,\ldots,p)$ *where* δ_n *satisfies condition* (14), *we have*

$$\mathbb{E}\Big(\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\nu_2}{n}\mathrm{tr}\mathbf{A}\Big)\Big(\mathbf{r}'\mathbf{B}\mathbf{r}-\frac{\nu_2}{n}\mathrm{tr}\mathbf{B}\Big)$$
$$=\frac{\nu_4-3\nu_{12}}{n^2}\sum_{i=1}^pA_{ii}B_{ii}+\frac{\nu_{12}}{n^2}\Big\{\mathrm{tr}(\mathbf{A}\mathbf{B}')+\mathrm{tr}(\mathbf{A}\mathbf{B})\Big\}+\frac{\nu_{12}-\nu_2^2}{n^2}\mathrm{tr}\mathbf{A}\mathrm{tr}\mathbf{B}+o(n^{-1}),$$

where $\mathbf{r}=n^{-1/2}(\hat{w}_1/\overline{\hat{w}}-1,\ldots,\hat{w}_p/\overline{\hat{w}}-1)'$, $\nu_2=\mathbb{E}\Big(\frac{\hat{w}_1}{\overline{\hat{w}}}-1\Big)^2$, $\nu_4=\mathbb{E}\Big(\frac{\hat{w}_1}{\overline{\hat{w}}}-1\Big)^4$, $\nu_{12}=\mathbb{E}\Big(\frac{\hat{w}_1}{\overline{\hat{w}}}-1\Big)^2\Big(\frac{\hat{w}_2}{\overline{\hat{w}}}-1\Big)^2$, *and* $\overline{\hat{w}}=p^{-1}\sum_{j=1}^p\hat{w}_j$.

5.3 CLT for LSS of the centralized sample covariance matrix B_p^0

Recall that in the following, we assume that the underlying variables are truncated at $\delta_n\sqrt{n}$. To simplify notation, we suppress the superscripts on the variables $\hat{w}_{ij}$; that is, we use w_{ij} in place of $\hat{w}_{ij}$ below.

5.3.1 Step 1: Finite dimensional convergence of $M_p^{(1)}(z)$ in distribution

Lemma 5.6. *Under Assumptions 2.1 and 2.2, as* $p\to\infty$, *for any set of* r *points* $\{z_1,z_2,\ldots,z_r\}\cup\mathcal{C}$, *the random vector* $(M_p^{(1)}(z_1),\ldots,M_p^{(1)}(z_r))$ *converges weakly to a* r*-dimensional centered Gaussian distribution with covariance function* (13).

We now proceed to the proof of this lemma. By the fact that a random vector is multivariate normally distributed if and only if every linear combination of its components is normally distributed, we need only show that, for any positive integer r and any complex sequence $\{\alpha_j\}_{j=1}^r$, the sum $\sum_{j=1}^r\alpha_jM_p^{(1)}(z_j)$ converges weakly to a Gaussian random variable. To this end, we first approximate $M_p^{(1)}(z)$ by a sum of martingale difference, which is given in (17). Then, we apply the martingale CLT (Theorem 35.12 in Billingsley [1995]) to obtain the asymptotic distribution of $M_p^{(1)}(z)$. Details of these two steps are provided in the following two parts.

Part 1: Martingale difference decomposition of $M_p^{(1)}(z)$.

First, we introduce some notations. In the following proof, we assume that $v = \Im z \geq v_0 > 0$. Moreover, for $j = 1, 2, \ldots, n$, let $\overline{w}_j = p^{-1}\sum_{\ell=1}^{p} w_{j\ell}$, $\nu_2 = \mathbb{E}\Big(\frac{w_{11}}{\overline{w}_1} - 1\Big)^2$, $\nu_{12} = \mathbb{E}\Big(\frac{w_{11}}{\overline{w}_1} - 1\Big)^2\Big(\frac{w_{12}}{\overline{w}_1} - 1\Big)^2$, $\nu_4 = \mathbb{E}\Big(\frac{w_{11}}{\overline{w}_1} - 1\Big)^4$,

$$\begin{aligned}
\mathbf{r}_j &= \frac{1}{\sqrt{n}}\left(\frac{w_{j1}}{\overline{w}_j} - 1, \ldots, \frac{w_{jp}}{\overline{w}_j} - 1\right)', \quad \mathbf{D}(z) = \mathbf{B}_p^0 - z\mathbf{I}_p, \\
\mathbf{D}_j(z) &= \mathbf{D}(z) - \mathbf{r}_j\mathbf{r}_j', \quad \beta_j(z) = \frac{1}{1 + \mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{r}_j}, \\
\overline{\beta}_j(z) &= \frac{1}{1 + n^{-1}\nu_2\mathrm{tr}\mathbf{D}_j^{-1}(z)}, \quad b_p(z) = \frac{1}{1 + n^{-1}\nu_2\mathbb{E}\mathrm{tr}\mathbf{D}_1^{-1}(z)}, \\
\varepsilon_j(z) &= \mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_j^{-1}(z), \\
\tilde{\varepsilon}_j(z) &= \mathbf{r}_j'\mathbf{D}_j^{-2}(z)\mathbf{r}_j - \frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_j^{-2}(z) = \frac{\mathrm{d}}{\mathrm{d}z}\varepsilon_j(z).
\end{aligned}$$

By Lemma 5.4, we have, for any $q \geq 2$,

$$\mathbb{E}|\varepsilon_j(z)|^q \leq \frac{K}{v^{2q}} n^{-1}\delta_n^{2q-4} \quad \text{and} \quad \mathbb{E}|\tilde{\varepsilon}_j(z)|^q \leq \frac{K}{v^{2q}} n^{-1}\delta_n^{2q-4}. \tag{16}$$

It is easy to see that

$$\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z) = -\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j'\mathbf{D}_j^{-1}(z)\beta_j(z),$$

where we use the formula that $\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1} = \mathbf{A}_2^{-1}(\mathbf{A}_2 - \mathbf{A}_1)\mathbf{A}_1^{-1}$ holds for any two invertible matrices $\mathbf{A}_1$ and $\mathbf{A}_2$. Note that $|\beta_j(z)|$, $|\overline{\beta}_j(z)|$ and $|b_n(z)|$ are bounded by $\frac{|z|}{v}$. Let $\mathbb{E}_j(\cdot)$ denote conditional expectation with respect to the σ-field generated by $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_j\}$, where $j = 1, 2, \ldots, n$. By convention, we use $\mathbb{E}_0 = \mathbb{E}$ to denote expectation. By using the above identity, we write

$$M_p^{(1)}(z) = \sum_{j=1}^{n}\mathrm{tr}\Big\{(\mathbb{E}_j - \mathbb{E}_{j-1})\mathbf{D}^{-1}(z)\Big\} = -\sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j'\mathbf{D}_j^{-2}(z)\mathbf{r}_j.$$

From the identity $\beta_j(z) = \overline{\beta}_j(z) - \beta_j(z)\overline{\beta}_j(z)\varepsilon_j(z) = \overline{\beta}_j(z) - \overline{\beta}_j^2(z)\varepsilon_j(z) + \overline{\beta}_j^2(z)\beta_j(z)\varepsilon_j^2(z)$ and the definition of $\tilde{\varepsilon}_j(z)$, we obtain that

$$\begin{aligned}
&(\mathbb{E}_j - \mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j'\mathbf{D}_j^{-2}(z)\mathbf{r}_j \\
&= (\mathbb{E}_j - \mathbb{E}_{j-1})\Big[\Big\{\overline{\beta}_j(z) - \overline{\beta}_j^2(z)\varepsilon_j(z) + \overline{\beta}_j^2(z)\beta_j(z)\varepsilon_j^2(z)\Big\}\Big\{\tilde{\varepsilon}_j(z) + \frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_j^{-2}(z)\Big\}\Big] \\
&= -Y_j(z) + \mathbb{E}_{j-1}Y_j(z) - (\mathbb{E}_j - \mathbb{E}_{j-1})\Big[\overline{\beta}_j^2(z)\Big\{\varepsilon_j(z)\tilde{\varepsilon}_j(z) - \beta_j(z)\varepsilon_j^2(z)\mathbf{r}_j'\mathbf{D}_j^{-2}(z)\mathbf{r}_j\Big\}\Big],
\end{aligned}$$

where

$$Y_j(z) := -\mathbb{E}_j\Big\{\overline{\beta}_j(z)\tilde{\varepsilon}_j(z) - \overline{\beta}_j^2(z)\varepsilon_j(z)\frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_j^{-2}(z)\Big\},$$

and the second equality follows from $(\mathbb{E}_j - \mathbb{E}_{j-1})\overline{\beta}_j(z)\mathrm{tr}\mathbf{D}_j^{-2}(z) = 0$. By using (16), we have

$$\mathbb{E}\left|\sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\overline{\beta}_j^2(z)\varepsilon_j(z)\tilde{\varepsilon}_j(z)\right|^2 \le 4\sum_{j=1}^{n}\mathbb{E}\left|\overline{\beta}_j^2(z)\varepsilon_j(z)\tilde{\varepsilon}_j(z)\right|^2 = o(1),$$

here we use the martingale difference property of $(\mathbb{E}_j - \mathbb{E}_{j-1})\overline{\beta}_j^2(z)\varepsilon_j(z)\tilde{\varepsilon}_j(z)$. Thus, $\sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\overline{\beta}_j^2(z)\varepsilon_j(z)\tilde{\varepsilon}_j(z) \xrightarrow{P} 0$. By the same argument, we have

$$\sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\overline{\beta}_j^2(z)\beta_j(z)\mathbf{r}_j'\mathbf{D}_j^{-2}(z)\mathbf{r}_j\varepsilon_j^2(z) \xrightarrow{P} 0.$$

The estimates above imply that

$$M_p^{(1)}(z) = \sum_{j=1}^{n}\{Y_j(z) - \mathbb{E}_{j-1}Y_j(z)\} + o_P(1), \tag{17}$$

where $\{Y_j(z) - \mathbb{E}_{j-1}Y_j(z)\}_{j=1}^n$ is a sequence of martingale difference.
Part 2: Application of martingales CLT to (17).

To prove finite-dimensional convergence of $M_p^{(1)}(z)$, $z \in \mathcal{C}$, we need only to consider the limit of the following martingale difference decomposition:

$$\sum_{i=1}^{r}\alpha_i M_p^{(1)}(z_i) = \sum_{j=1}^{n}\sum_{i=1}^{r}\alpha_i\Big\{Y_j(z_i) - \mathbb{E}_{j-1}Y_j(z_i)\Big\} + o_P(1), \tag{18}$$

where $\Im(z_i) \neq 0$, and $\{\alpha_i\}_{i=1}^r$ are constants. We apply the martingale CLT [Billingsley, 1995, Theorem 35.12] to this martingale difference decomposition (18). To this end, we need to check two conditions: as $n \to \infty$, for any $\varepsilon > 0$

$$\sum_{j=1}^{n}\mathbb{E}\left|\sum_{i=1}^{r}\alpha_i\{Y_j(z_i) - \mathbb{E}_{j-1}Y_j(z_i)\}\right|^2 I_{\{|\sum_{i=1}^{r}\alpha_i(Y_j(z_i) - \mathbb{E}_{j-1}Y_j(z_i))| \ge \varepsilon\}} \to 0, \tag{19}$$

$$\sum_{j=1}^{n}\mathbb{E}_{j-1}\Big[\Big\{Y_j(z_1) - \mathbb{E}_{j-1}Y_j(z_1)\Big\}\Big\{Y_j(z_2) - \mathbb{E}_{j-1}Y_j(z_2)\Big\}\Big] \xrightarrow{P} (13). \tag{20}$$

First, we verify (19). By Lemma 5.4, we obtain

$$\mathbb{E}|Y_j(z)|^4 \le K\mathbb{E}|\varepsilon_j(z)|^4 = o(p^{-1}),$$

which, together with Jensen's inequality, implies that

$$\mathbb{E}|\mathbb{E}_{j-1}Y_j(z)|^4 \le \mathbb{E}(\mathbb{E}_{j-1}|Y_j(z)|^4) = \mathbb{E}|Y_j(z)|^4 = o(p^{-1}).$$

It follows from the above two equations that

$$\text{LHS of (19)} \le \frac{K}{\varepsilon^2}\sum_{j=1}^{n}\mathbb{E}\left|\sum_{i=1}^{r}\alpha_i Y_j(z_i)\right|^4 + \frac{K}{\varepsilon^2}\sum_{j=1}^{n}\mathbb{E}\left|\sum_{i=1}^{r}\alpha_i\mathbb{E}_{j-1}Y_j(z_i)\right|^4 \to 0.$$

Then, we verify (20). Since $Y_j(z) = -\mathbb{E}_j \frac{\mathrm{d}}{\mathrm{d}z}\left\{\overline{\beta}_j(z)\varepsilon_j(z)\right\}$, we have

$$\text{LHS of (20)} = \frac{\partial^2}{\partial z_1 \partial z_2}\left\{\mathcal{Y}_1(z_1,z_2) - \mathcal{Y}_2(z_1,z_2)\right\}, \tag{21}$$
$$\mathcal{Y}_1(z_1,z_2) := \sum_{j=1}^{n} \mathbb{E}_{j-1}\left[\mathbb{E}_j\left\{\overline{\beta}_j(z_1)\varepsilon_j(z_1)\right\}\mathbb{E}_j\left\{\overline{\beta}_j(z_2)\varepsilon_j(z_2)\right\}\right],$$
$$\mathcal{Y}_2(z_1,z_2) := \sum_{j=1}^{n} \mathbb{E}_{j-1}\left\{\overline{\beta}_j(z_1)\varepsilon_j(z_1)\right\}\mathbb{E}_{j-1}\left\{\overline{\beta}_j(z_2)\varepsilon_j(z_2)\right\}.$$

Thus, it is enough to consider the limits of $\mathcal{Y}_i(z_1,z_2), i = 1,2$, which are provided in the following lemma.

Lemma 5.7. *Under Assumptions 2.1 and 2.2, as $n \to \infty$, we have*

$$\frac{\partial^2}{\partial z_1 \partial z_2}\mathcal{Y}_1(z_1,z_2) \xrightarrow{P} (13), \qquad \frac{\partial^2}{\partial z_1 \partial z_2}\mathcal{Y}_2(z_1,z_2) \xrightarrow{P} 0.$$

The proof of Lemma 5.7 is postponed to the supplementary material. This lemma and Equation (21) complete the proof of (20).

5.3.2 Step 2: Tightness of $M_p^{(1)}(z)$

Tightness of $M_p^{(1)}(z)$ can be established by Theorem 12.3 of Billingsley [1968]. It is sufficient to prove the moment condition of Billingsley [1968], i.e., $\sup_{n;z_1,z_2\in\mathcal{C}_n} \frac{\mathbb{E}|M_p^{(1)}(z_1)-M_p^{(1)}(z_2)|^2}{|z_1-z_2|^2}$ is finite. Its proof exactly follows Bai and Silverstein [2004], and is postponed to the supplementary material.

5.3.3 Step 3: Convergence of $M_p^{(2)}(z)$

Recalling that $M_p^{(2)}(z) = p\{\mathbb{E}m_p(z) - m_p^0(z)\} = n\{\mathbb{E}\underline{m}_p(z) - \underline{m}_p^0(z)\}$, where $m_p(z) = m_{F^{\mathbf{B}_p^0}}(z)$, $m_p^0(z) = m_{F^{c_n}}(z)$, $\underline{m}_p(z) = m_{F^{\underline{\mathbf{B}}_p^0}}(z)$, $\underline{m}_p^0(z) = \underline{m}_{F^{c_n}}(z)$, $c_n = p/n$. From

$$\frac{1}{\mathbb{E}\underline{m}_p(z)} + z - \frac{c_n\lambda}{1+\lambda\mathbb{E}\underline{m}_p(z)} = \frac{1}{\mathbb{E}\underline{m}_p(z)}\left\{1 - c_n + z\mathbb{E}\underline{m}_p(z) + \frac{c_n}{1+\lambda\mathbb{E}\underline{m}_p(z)}\right\},$$

we have

$$\mathbb{E}\underline{m}_p(z) = \left\{-z + \frac{\lambda c_n}{1+\lambda\mathbb{E}\underline{m}_p(z)} + \frac{A_p(z)}{\mathbb{E}\underline{m}_p(z)}\right\}^{-1},$$

where $A_p(z) := \frac{c_n}{1+\lambda\mathbb{E}\underline{m}_p(z)} + zc_n\mathbb{E}m_p(z)$. From this equation and the identity $\underline{m}_p^0 = (-z + \frac{\lambda c_n}{1+\lambda\underline{m}_p^0})^{-1}$, we get

$$\mathbb{E}\underline{m}_p(z) - \underline{m}_p^0(z) = -\underline{m}_p^0(z)A_p(z)\left[1 - \frac{c_n\lambda^2\underline{m}_p^0(z)\mathbb{E}\underline{m}_p(z)}{\{1+\lambda\mathbb{E}\underline{m}_p(z)\}\{1+\lambda\underline{m}_p^0(z)\}}\right]^{-1}. \tag{22}$$

Note that $\mathbb{E}\underline{m}_p(z) \to \underline{m}(z)$, $\underline{m}_p^0(z) \to \underline{m}(z)$. It suffices to derive the limit of $nA_p(z)$, which is provided in the following lemma.

Lemma 5.8. *Under Assumptions 2.1 and 2.2, as $n \to \infty$, we have*

$$nA_p(z) \to -\frac{\underline{m}(z)\{z(\lambda+h_1)m(z)+\lambda\}}{1+\lambda\underline{m}(z)} - \frac{cz^2\underline{m}^2(z)\{(\alpha_1+\alpha_2)m^2(z)+2\lambda^2 m'(z)\}}{1+\lambda\underline{m}(z)} + \frac{c\lambda^2\underline{m}^2(z)}{\{1+\lambda\underline{m}(z)\}[\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)]}.$$

The proof of Lemma 5.8 is postponed to supplementary material. By (22) and Lemma 5.8, we have $M_p^{(2)}(z) \to$ (12) as $n \to \infty$. Combining two parts above yields Lemma 5.6.

References

Zhidong Bai and Jack W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, 32(1A):553–605, 2004. doi: 10.1214/aop/1078415845.

Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer, 2nd edition, 2010.

Zhidong Bai and Yongquan Yin. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *The Annals of Probability*, 21(3):1275–1294, 1993. doi: 10.1214/aop/1176989118.

Zhidong Bai, Baiqi Miao, and Guangming Pan. On asymptotics of eigenvectors of large sample covariance matrix. *The Annals of Probability*, 35(4):1532 – 1572, 2007. doi: 10.1214/009117906000001079.

Zhidong Bai, Huiqin Li, and Guangming Pan. Central limit theorem for linear spectral statistics of large dimensional separable sample covariance matrices. *Bernoulli*, 25(3):1838–1869, 2019. doi: 10.3150/18-bej1038.

Zhigang Bao. Tracy-widom limit for kendall's tau. *The Annals of Statistics*, 47(6):3504–3532, 2019. doi: 10.1214/18-aos1786.

Florent Benaych-Georges and Raj Rao Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012. doi: 10.1016/j.jmva.2012.04.019.

Patrick Billingsley. *Convergence of probability measures*. New York: Wiley, 1968.

Patrick Billingsley. *Probability and Measure*. New York: Wiley, 1995.

Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Circular law theorem for random markov matrices. *Probability Theory and Related Fields*, 152(3):751–779, 2012. doi: 10.1007/s00440-010-0336-1.

Donald L. Burkholder. Distribution function inequalities for martingales. *The Annals of Probability*, 1(1):19–42, 1973. doi: 10.1214/aop/1176997023.

Tony Cai, Weidong Liu, and Yin Xia. Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):349 – 372, 2014. doi: 10.1111/rssb.12034.

Yuanpei Cao, Wei Lin, and Hongzhe Li. Two-sample tests of high-dimensional means for compositional data. *Biometrika*, 105(1):115 – 132, 2018. doi: 10.1093/biomet/asx060.

Noureddine El Karoui. Tracy–widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, 35(2):470 – 490, 2007. doi: 10.1214/009117906000000917.

Karoline Faust, Jarupon Fah Sathirapongsasuti, Jacques Izard, N. Segata, Dirk Gevers, Jeroen Raes, and Curtis Huttenhower. Microbial co-occurrence relationships in the human microbiome. *PLoS Computational Biology*, 8(7):e1002606, 2012. doi: 10.1371/journal.pcbi.1002606.

Jiti Gao, Xiao Han, Guangming Pan, and Yanrong Yang. High dimensional correlation matrices: The central limit theorem and its applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):677–693, 2017. doi: 10.1111/rssb.12189.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge University Press, 2012.

Tiefeng Jiang. The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā: The Indian Journal of Statistics*, 66(1):35–48, 2004. doi: 10.2307/25053330.

Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295 – 327, 2001. doi: 10.1214/aos/1009210544.

Dag Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis*, 12(1):1–38, 1982. doi: 10.1016/0047-259x(82)90080-x.

V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483, 1967. doi: 10.1070/sm1967v001n04abeh001994.

Jamal Najim and Jianfeng Yao. Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *The Annals of Applied Probability*, 26(3), 2016. doi: 10.1214/15-aap1135.

Jack W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55(2):331–339, 1995. doi: 10.1006/jmva.1995.1083.

Jack W. Silverstein and Zhidong Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):175–192, 1995. doi: 10.1006/jmva.1995.1051.

Gary D. Wu, Jun Chen, Christian Hoffmann, Kyle Bittinger, Ying-Yu Chen, Sue A. Keilbaugh, Meenakshi Bewtra, Dan Knights, William A. Walters, Rob Knight, Rohini Sinha, Erin Gilroy, Kernika Gupta, Robert N. Baldassano, Lisa C. Nessel, Hongzhe Li, Frederic D. Bushman, and James D. Lewis. Linking long-term dietary patterns with gut microbial enterotypes. *Science*, 334 (6052):105 – 108, 2011. doi: 10.1126/science.1208344.

Han Xiao and Wang Zhou. Almost sure limit of the smallest eigenvalue of some sample correlation matrices. *Journal of Theoretical Probability*, 23(1):1–20, 2010. doi: 10.1007/s10959-009-0270-2.

Yongquan Yin and Paruchuri R. Krishnaiah. A limit theorem for the eigenvalues of product of two random matrices. *Journal of Multivariate Analysis*, 13(4):489–507, 1983. doi: 10.1016/0047-259x(83)90035-0.

Lixin Zhang. *Spectral analysis of large dimentional random matrices*. Phd thesis, National University of Singapore, 2007.

Shurong Zheng, Zhidong Bai, and Jianfeng Yao. Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *The Annals of Statistics*, 43(2):546–591, 2015. doi: 10.1214/14-AOS1292.

Supplementary Material for "On eigenvalues of sample covariance matrices based on high-dimensional compositional data"

S1 Auxiliary lemmas

This section introduces several auxiliary lemmas used in the technical proofs of our theoretical results. Lemmas S1.1 – S1.4 are from existing literature, while Lemma S1.5 is our original contribution, and its proof is provided in Section S2.13.

Lemma S1.1 (Weyl's inequality, Corollary 7.3.5 of Horn and Johnson [2012]). *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $p\times n$ *matrices and let* $r=\min\{p,n\}$. *Let* $s_1(\mathbf{A})\geq\cdots\geq s_r(\mathbf{A})$ *and* $s_1(\mathbf{B})\geq\cdots\geq s_r(\mathbf{B})$ *be the nonincreasingly ordered singular values of* $\mathbf{A}$ *and* $\mathbf{B}$, *respectively. Then*

$$\max_{1\leq i\leq r}|s_i(\mathbf{A})-s_i(\mathbf{B})|\leq\|\mathbf{A}-\mathbf{B}\|,$$

where $\|\mathbf{A}-\mathbf{B}\|$ *denotes the spectral norm of* $\mathbf{A}-\mathbf{B}$.

Lemma S1.2 (Burkholder's inequality, Burkholder [1973]). *Let* $\{X_k\}$ *be a complex martingale difference sequence with respect to the increasing* σ*-field* $\{\mathcal{F}_k\}$, *and let* $\mathbb{E}_k$ *denote conditional expectation with respect to* $\mathcal{F}_k$. *Then, for* $q>1$,

$$\mathbb{E}\left|\sum X_k\right|^q\leq K_q\left\{\mathbb{E}\left(\sum\mathbb{E}_{k-1}|X_k|^2\right)^{q/2}+\mathbb{E}\sum|X_k|^q\right\}.$$

Lemma S1.3 (Uniform law of large numbers, Lemma 2 of Bai and Yin [1993]). *Let* $\{X_{ij},i,j=1,2,\ldots\}$ *be a double array of i.i.d. random variables and let* $\alpha>1/2$, $\beta\geq 0$ *and* $M>0$ *be constants. Then as* $n\to\infty$,

$$\max_{j\leq Mn^\beta}\left|n^{-\alpha}\sum_{i=1}^n(X_{ij}-c)\right|\xrightarrow{a.s.}c,$$

if and only if the following hold:

$$\mathbb{E}|X_{11}|^{(1+\beta)/\alpha}<\infty,\qquad c=\begin{cases}\mathbb{E}X_{11}, & \text{if } \alpha\leq 1,\\ \text{any number}, & \text{if } \alpha>1.\end{cases}$$

Lemma S1.4 (Martingale CLT, Theorem 35.12 of Billingsley [1995]). *Suppose that for each* n, *the sequence* $\{Y_{n1},\ldots,Y_{nr_n}\}$ *is a real martingale difference sequence with respect to the increasing* σ*-field* $\{\mathcal{F}_{nj}\}$ *having second moments. If as* $n\to\infty$,

$$\sum_{j=1}^{r_n}\mathbb{E}(Y_{nj}^2|\mathcal{F}_{n,j-1})\xrightarrow{P}\sigma^2,$$

where σ^2 *is positive constant, and for each* $\varepsilon>0$,

$$\sum_{j=1}^{r_n}\mathbb{E}(Y_{nj}^2 I_{\{|Y_{nj}|\geq\varepsilon\}})\to 0,$$

then $\sum_{j=1}^{r_n}Y_{nj}\xrightarrow{D}\mathcal{N}(0,\sigma^2)$.

Lemma S1.5. *Suppose that $\mathbf{x}_p = \frac{1}{\sqrt{p}}(1,1,\ldots,1)'$ is a p-dimensional normalized all-one vector, then for the random variables $\{w_{ij}, i=1,\ldots,n, j=1\ldots,p\}$ satisfying Assumptions 2.1 and 2.2 and the truncation condition $|w_{ij}-\mu| \le \delta_n\sqrt{n}$, where δ_n satisfies condition (14), we have*

$$\mathbb{E}|\mathbf{x}_p'\mathbf{D}^{-1}(z)\mathbf{x}_p + 1/z|^2 \to 0,$$

for $z = u+iv$, $v>0$.

S2 Proofs

S2.1 Proof of Proposition 2.4

The proof of the first part of Proposition 2.4 is given in the main text. In the follows, we present the proof of the second part. Recall that $\mathbf{Y}_n = p\mathbf{C}_n\mathbf{\Lambda}_n\mathbf{W}_n/\sqrt{N}$, $\check{\mathbf{Y}}_n = \mathbf{C}_n\mathbf{W}_n/(\sqrt{N}\mu)$, and $\mathbf{B}_{p,N} = \mathbf{Y}_n'\mathbf{Y}_n$. Let $\check{\mathbf{B}}_{p,N} = \check{\mathbf{Y}}_n'\check{\mathbf{Y}}_n$. For any positive constant ε small enough such that

$$\eta_1 - \varepsilon > \lambda(1+\sqrt{c})^2, \qquad \eta_2 + 2\varepsilon < \lambda(1-\sqrt{c})^2 I_{\{0<c<1\}}, \tag{S2.1}$$

we have

$$\begin{aligned}
&\mathbb{P}\Big(\lambda_{\max}(\mathbf{B}_{p,N}) \ge \eta_1\Big)\\
=\ &\mathbb{P}\Big(\lambda_{\max}(\mathbf{B}_{p,N}) \ge \eta_1, \lambda_{\max}(\check{\mathbf{B}}_{p,N}) \ge \eta_1 - \varepsilon\Big)\\
&\quad + \mathbb{P}\Big(\lambda_{\max}(\mathbf{B}_{p,N}) \ge \eta_1, \lambda_{\max}(\check{\mathbf{B}}_{p,N}) < \eta_1 - \varepsilon\Big)\\
\le\ &\mathbb{P}\Big(\lambda_{\max}(\check{\mathbf{B}}_{p,N}) \ge \eta_1 - \varepsilon\Big) + \mathbb{P}\Big(|\lambda_{\max}(\mathbf{B}_{p,N}) - \lambda_{\max}(\check{\mathbf{B}}_{p,N})| \ge \varepsilon\Big)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}\Big(\lambda_{\min}(\mathbf{B}_{p,N}) \le \eta_2\Big)\\
=\ &\mathbb{P}\Big(\lambda_{\min}(\mathbf{B}_{p,N}) \le \eta_2, \lambda_{\min}(\check{\mathbf{B}}_{p,N}) \le \eta_2 + \varepsilon\Big)\\
&\quad + \mathbb{P}\Big(\lambda_{\min}(\mathbf{B}_{p,N}) \le \eta_2, \lambda_{\min}(\check{\mathbf{B}}_{p,N}) > \eta_2 + \varepsilon\Big)\\
\le\ &\mathbb{P}\Big(\lambda_{\min}(\check{\mathbf{B}}_{p,N}) \le \eta_2 + \varepsilon\Big) + \mathbb{P}\Big(|\lambda_{\min}(\mathbf{B}_{p,N}) - \lambda_{\min}(\check{\mathbf{B}}_{p,N})| \ge \varepsilon\Big).
\end{aligned}$$

To prove the second part of Proposition 2.4, it suffices to give the following three estimations:

$$\mathbb{P}\Big(\max_{1\le i\le p}|\lambda_i(\mathbf{B}_{p,N}) - \lambda_i(\check{\mathbf{B}}_{p,N})| \ge \varepsilon\Big) = o(n^{-\ell}), \tag{S2.2}$$

$$\mathbb{P}\Big(\lambda_{\max}(\check{\mathbf{B}}_{p,N}) \ge \eta_1 - \varepsilon\Big) = o(n^{-\ell}), \tag{S2.3}$$

$$\mathbb{P}\Big(\lambda_{\min}(\check{\mathbf{B}}_{p,N}) \le \eta_2 + \varepsilon\Big) = o(n^{-\ell}). \tag{S2.4}$$

The proof of these estimates are provided as follows:
Proof of (S2.2): By Lemma S1.1, we have

$$\max_{1\le i\le p}|\lambda_i(\mathbf{B}_{p,N}) - \lambda_i(\check{\mathbf{B}}_{p,N})| \le \|\mathbf{B}_{p,N} - \check{\mathbf{B}}_{p,N}\| \le \|\mathbf{Y}_n - \check{\mathbf{Y}}_n\|^2 + 2\|\mathbf{Y}_n - \check{\mathbf{Y}}_n\|\|\check{\mathbf{Y}}_n\|.$$

Note that $\|\mathbf{Y}_n - \check{\mathbf{Y}}_n\| \le \|p\mu\mathbf{\Lambda}_n - \mathbf{I}_n\|\|\mathbf{W}_n/(\mu\sqrt{N})\|$. We get from Theorem 2.9 in Benaych-Georges and Nadakuditi [2012] that $\|\mathbf{W}_n/(\mu\sqrt{N})\|$ is bounded almost surely. In view of the above inequalities and (S2.3) (will be proved below), it suffices to show that, for any $\ell > 0$ and $\varepsilon > 0$, $\mathbb{P}(\|p\mu\mathbf{\Lambda}_n - \mathbf{I}_n\| \ge \varepsilon) = o(n^{-\ell})$, which is guaranteed by

$$\mathbb{P}\left(\max_{1\le i\le n}\left|\frac{\sum_{j=1}^{p} w_{ij}/p}{\mu} - 1\right| \ge \varepsilon\right) = o(n^{-\ell}). \tag{S2.5}$$

This inequality follows from Equation (B.115) in Gao et al. [2017], and thus we complete the proof of (S2.2).

Proof of (S2.3): Let $\mathbf{B}^{\circ}_{p,N} = (\mathbf{Y}^{\circ}_n)'\mathbf{Y}^{\circ}_n$, where $\mathbf{Y}^{\circ}_n = \frac{\mathbf{W}_n - \mathbb{E}\mathbf{W}_n}{\sqrt{N}\mu}$. From Bai and Silverstein [2004], we have

$$\mathbb{P}\Big(\lambda_{\max}(\mathbf{B}^{\circ}_{p,N}) \ge \eta_1 - \varepsilon\Big) = o(n^{-\ell}), \tag{S2.6}$$

$$\mathbb{P}\Big(\lambda_{\min}(\mathbf{B}^{\circ}_{p,N}) \le \eta_2 + \varepsilon\Big) = o(n^{-\ell}). \tag{S2.7}$$

By the identity $\check{\mathbf{Y}}_n = \mathbf{C}_n\mathbf{Y}^{\circ}_n$, we have

$$\check{\mathbf{B}}_{p,N} = \mathbf{B}^{\circ}_{p,N} - \frac{1}{n}(\mathbf{Y}^{\circ}_n)'\mathbf{1}_n\mathbf{1}_n'\mathbf{Y}^{\circ}_n. \tag{S2.8}$$

This, together with Cauchy interlacing theorem, implies that

$$\lambda_1(\mathbf{B}^{\circ}_{p,N}) \ge \lambda_1(\check{\mathbf{B}}_{p,N}) \ge \lambda_2(\mathbf{B}^{\circ}_{p,N}) \ge \lambda_2(\check{\mathbf{B}}_{p,N}) \ge \cdots \ge \lambda_p(\mathbf{B}^{\circ}_{p,N}) \ge \lambda_p(\check{\mathbf{B}}_{p,N}). \tag{S2.9}$$

For the largest eigenvalue, we have

$$\lambda_{\max}(\mathbf{B}^{\circ}_{p,N}) = \lambda_1(\mathbf{B}^{\circ}_{p,N}) \ge \lambda_1(\check{\mathbf{B}}_{p,N}) = \lambda_{\max}(\check{\mathbf{B}}_{p,N}),$$

which, together with (S2.6), implies (S2.3).

Proof of (S2.4): When $p \ge n$, the smallest eigenvalue of $\check{\mathbf{B}}_{p,N}$ is its $(n-1)$-th largest eigenvalue. By using (S2.9), we have

$$\lambda_{\min}(\check{\mathbf{B}}_{p,N}) = \lambda_{n-1}(\check{\mathbf{B}}_{p,N}) \ge \lambda_n(\mathbf{B}^{\circ}_{p,N}) = \lambda_{\min}(\mathbf{B}^{\circ}_{p,N}),$$

which, together with (S2.7), implies (S2.4). When $p < n$, the smallest eigenvalue of $\check{\mathbf{B}}_{p,N}$ is its p-th largest eigenvalue, and all eigenvalues of $\check{\mathbf{B}}_{p,N}$ and $\mathbf{B}^{\circ}_{p,N}$ are interlaced each other as in (S2.9). From (S2.8), we have

$$\mathrm{tr}(\mathbf{B}^{\circ}_{p,N}) = \mathrm{tr}(\check{\mathbf{B}}_{p,N}) + \frac{n}{N}\sum_{j=1}^{p}\Delta^2_{j,n}, \qquad \Delta_{j,n} = \frac{\sum_{i=1}^{n} w_{ij}/n}{\mu} - 1.$$

Hence, there exists some constant C such that $\lambda_{\min}(\check{\mathbf{B}}_{p,N}) = \lambda_{\min}(\mathbf{B}^{\circ}_{p,N}) - \frac{C}{p}\frac{n}{N}\sum_{j=1}^{p}\Delta^2_{j,n}$, and thus

$$\begin{aligned}
&\Pr\Big(\lambda_{\min}(\check{\mathbf{B}}_{p,N}) \le \eta_2 + \varepsilon\Big) \\
=&\Pr\left(\lambda_{\min}(\check{\mathbf{B}}_{p,N}) \le \eta_2 + \varepsilon,\ \frac{C}{p}\frac{n}{N}\sum_{j=1}^{p}\Delta^2_{j,n} \le \varepsilon\right)
\end{aligned}$$

$$
\begin{aligned}
&+\Pr\left(\lambda_{\min}(\breve{\mathbf{B}}_{p,N}) \le \eta_2+\varepsilon,\ \frac{C}{p}\frac{n}{N}\sum_{j=1}^{p}\Delta_{j,n}^2 \ge \varepsilon\right)\\
\le&\Pr\left(\lambda_{\min}(\mathbf{B}_{p,N}^{\circ}) \le \eta_2+2\varepsilon\right)+\Pr\left(\max_{1\le j\le p}|\Delta_{j,n}|^2 \ge \varepsilon/C\right).
\end{aligned}
\tag{S2.10}
$$

From (S2.1) and Bai and Silverstein [2004], the first term in (S2.10) is of order $o(n^{-\ell})$ for any $\ell > 0$. Similar to (S2.5), for any $\ell > 0$ and $\varepsilon > 0$, we have $\mathbb{P}(\max_{1\le j\le p}|\Delta_{j,n}|^2 \ge \varepsilon) = o(n^{-\ell})$. Therefore, we conclude that (S2.4) still holds true when $p < n$.

S2.2 Proof of Lemma 5.2

The proof of this lemma is quite similar to Sections 5.3.1, 5.3.2, and 5.5 of Zheng et al. [2015], it is then omitted. For readers' convenience, we present the outline of the proof for this lemma. In this situation, $\mathbf{B}_p^0 = \frac{1}{n}\mathbf{Y}_n'\mathbf{Y}_n = \sum_{i=1}^n \mathbf{r}_i\mathbf{r}_i'$, $\mathbf{r}_i = \frac{1}{\sqrt{n}}(\frac{w_{i1}}{\overline{w}_i}-1,\dots,\frac{w_{ip}}{\overline{w}_i}-1)' = \frac{1}{\sqrt{n}}(y_{i1},\dots,y_{ip})'$, $\overline{w}_i = p^{-1}\sum_{\ell=1}^p w_{i\ell}$. As for moments of y_{ij}, by Lemma 5.4, for any $q > 0$, we have

$$
\mathbb{E}y_{ij}^q = \mathbb{E}\left[\left(\frac{w_{ij}}{\overline{w}_i}-1\right)^q I_{B_p(\varepsilon)}\right] + \mathbb{E}\left[\left(\frac{w_{ij}}{\overline{w}_i}-1\right)^q I_{B_p^c(\varepsilon)}\right] \le K\mathbb{E}\left(w_{ij}-\mu\right)^q,
$$

where $B_p(\varepsilon) = \{\omega : |\overline{w}_i - u| \le \varepsilon, \overline{w}_i = p^{-1}\sum_{\ell=1}^p w_{i\ell}\}$, in which $\varepsilon > 0$ is a constant. Therefore, in the following proof, the requirement of truncation of y_{ij} reduces to truncation of w_{ij}. First, we get that

$$
\begin{aligned}
&\operatorname{tr}(\mathbf{B}_{p,N} - z\mathbf{I}_p)^{-1} - pm_N^0(z)\\
=&\operatorname{tr}\{\mathbf{A}^{-1}(z)\} - pm_n^0(z) + p\{m_n^0(z) - m_N^0(z)\} + \operatorname{tr}\{\mathbf{A}^{-2}(z)\mathbf{\Delta}\}\\
&+\operatorname{tr}\left[\mathbf{A}^{-1}(z)\{\mathbf{\Delta}\mathbf{A}^{-1}(z)\}^2\right] + \operatorname{tr}\left[\{\mathbf{A}(z)-\mathbf{\Delta}\}^{-1}\{\mathbf{\Delta}\mathbf{A}^{-1}(z)\}^3\right],
\end{aligned}
$$

where $\mathbf{A}(z) = \mathbf{B}_p^0 - z\mathbf{I}_p$ and $\mathbf{\Delta} = \mathbf{B}_p^0 - \mathbf{B}_{p,N}$. Moreover, after truncation and normalization, for every $z \in \mathbb{C}^+ = \{z : \mathbb{I}\mathrm{m}\, z > 0\}$,

$$
p\{m_n^0(z) - m_N^0(z)\} = \{1 + z\underline{m}(z)\}\frac{\underline{m}(z) + z\underline{m}'(z)}{z\underline{m}(z)} + o_P(1), \quad \operatorname{tr}\{\mathbf{A}^{-2}(z)\mathbf{\Delta}\} = o_P(1), \tag{S2.11}
$$

$$
\operatorname{tr}\{\mathbf{A}^{-2}(z)\mathbf{\Delta}\mathbf{A}^{-1}(z)\mathbf{\Delta}\} = \{\underline{m}(z) + z\underline{m}'(z)\}\{1 + z\underline{m}(z)\} + o_P(1), \tag{S2.12}
$$

$$
\operatorname{tr}\left[\{\mathbf{\Delta}\mathbf{A}^{-1}(z)\}^3\{\mathbf{A}(z)-\mathbf{\Delta}\}^{-1}\right] = \frac{\{1 + z\underline{m}(z)\}^2\{\underline{m}(z) + z\underline{m}'(z)\}}{-z\underline{m}(z)} + o_P(1). \tag{S2.13}
$$

Note that, we also need to check the tightness of $\operatorname{tr}(\mathbf{B}_{p,N} - z\mathbf{I}_p)^{-1} - pm_N^0(z)$. Since

$$
\begin{aligned}
&\operatorname{tr}(\mathbf{B}_{p,N} - z\mathbf{I}_p)^{-1} - pm_N^0(z)\\
=&\operatorname{tr}(\mathbf{B}_{p,N} - z\mathbf{I}_p)^{-1} - \operatorname{tr}\{\mathbf{A}^{-1}(z)\} + \operatorname{tr}\{\mathbf{A}^{-1}(z)\} - pm_n^0(z) + p\{m_n^0(z) - m_N^0(z)\},
\end{aligned}
$$

and the tightness of $\operatorname{tr}\{\mathbf{A}^{-1}(z)\} - pm_n^0(z)$ is proved in Step 2 of Section 5.3, it suffices to prove tightness of $\operatorname{tr}(\mathbf{B}_{p,N} - z\mathbf{I}_p)^{-1} - \operatorname{tr}\{\mathbf{A}^{-1}(z)\}$. It can be obtained from similar arguments in Section 5.3.2 of Zheng et al. [2015] and we omit the details. Finally, the proof is completed.

S2.3 Proof of Lemma 5.3

Note that, by Taylor expansion, there exist $C_1 > 0$ such that, for any $-1/2 \le x \le 1/2$,

$$\frac{1}{(1+x)^2} = 1 - 2x + 3x^2 + a(x), \quad |a(x)| \le C_1|x|^3.$$

Hence, there exist $C_1 > 0$ such that, for any $0 < \varepsilon < 1/2$, on the event $\hat{B}_p(\varepsilon) = \{\omega : |\overline{\hat{w}} - \mu_n| \le \varepsilon, \overline{\hat{w}} = \sum_{j=1}^p \hat{w}_j/p, \mu_n = \mathbb{E}\hat{w}_1\}$,

$$\frac{1}{\overline{\hat{w}}^2} = \frac{1}{\mu_n^2(\frac{\overline{\hat{w}}-\mu_n}{\mu_n}+1)^2} = \frac{1}{\mu_n^2}\Big[1 - \frac{2(\overline{\hat{w}}-\mu_n)}{\mu_n} + \frac{3(\overline{\hat{w}}-\mu_n)^2}{\mu_n^2} + a\Big(\frac{\overline{\hat{w}}-\mu_n}{\mu_n}\Big)\Big],$$

where $|a((\overline{\hat{w}}-\mu_n)/\mu_n)| < C_1\varepsilon^3$. Hence, we have

$$\frac{\hat{w}_1^2}{\overline{\hat{w}}^2} I_{\hat{B}_p(\varepsilon)} = \Big\{\frac{\hat{w}_1^2}{\mu_n^2} - \frac{2\hat{w}_1^2(\overline{\hat{w}}-\mu_n)}{\mu_n^3} + \frac{3\hat{w}_1^2(\overline{\hat{w}}-\mu_n)^2}{\mu_n^4}\Big\} I_{\hat{B}_p(\varepsilon)} + \frac{\hat{w}_1^2}{\mu_n^2} a\Big(\frac{\overline{\hat{w}}-\mu_n}{\mu_n}\Big) I_{\hat{B}_p(\varepsilon)}. \tag{S2.14}$$

This, together with the fact $I_{\hat{B}_p(\varepsilon)} = 1 - I_{\hat{B}_p^c(\varepsilon)}$, implies that

$$\begin{aligned}\frac{\hat{w}_1^2}{\overline{\hat{w}}^2} - \frac{\hat{w}_1^2}{\mu_n^2} &= -\frac{2\hat{w}_1^2(\overline{\hat{w}}-\mu_n)}{\mu_n^3} + \frac{3\hat{w}_1^2(\overline{\hat{w}}-\mu_n)^2}{\mu_n^4} \\ &\quad - \Big\{\frac{\hat{w}_1^2}{\mu_n^2} - \frac{2\hat{w}_1^2(\overline{\hat{w}}-\mu_n)}{\mu_n^3} + \frac{3\hat{w}_1^2(\overline{\hat{w}}-\mu_n)^2}{\mu_n^4}\Big\} I_{\hat{B}_p^c(\varepsilon)} + a + \frac{\hat{w}_1^2}{\overline{\hat{w}}^2} I_{\hat{B}_p^c(\varepsilon)},\end{aligned} \tag{S2.15}$$

where $a := \frac{\hat{w}_1^2}{\mu_n^2} a\Big(\frac{\overline{\hat{w}}-\mu_n}{\mu_n}\Big) I_{\hat{B}_p(\varepsilon)}$ and $|a| \le C_1 \frac{\hat{w}_1^2}{\mu_n^2}\varepsilon^3$. Taking expectation for (S2.15) yields that

$$\mathbb{E}\frac{\hat{w}_1^2}{\overline{\hat{w}}^2} - \mathbb{E}\frac{\hat{w}_1^2}{\mu_n^2} = -\frac{2\mathbb{E}\hat{w}_1^2(\overline{\hat{w}}-\mu_n)}{\mu_n^3} + \frac{3\mathbb{E}\hat{w}_1^2(\overline{\hat{w}}-\mu_n)^2}{\mu_n^4} - \mathbb{E}b + \mathbb{E}a + \mathbb{E}c. \tag{S2.16}$$

where $b := \{\hat{w}_1^2/\mu_n^2 - 2\hat{w}_1^2(\overline{\hat{w}}-\mu_n)/\mu_n^3 + 3\hat{w}_1^2(\overline{\hat{w}}-\mu_n)^2/\mu_n^4\} I_{\hat{B}_p^c(\varepsilon)}$ and $c := (\hat{w}_1^2/\overline{\hat{w}}^2) I_{\hat{B}_p^c(\varepsilon)}$. Note that

$$|\mathbb{E}c| \le p^2 \mathbb{P}(\hat{B}_p^c(\varepsilon)). \tag{S2.17}$$

Next, we derive a bound for $\mathbb{E}b$. In save of notation, we denote by

$$y_1 = \frac{\hat{w}_1^2}{\mu_n^2}, \qquad y_2 = \frac{2\hat{w}_1^2(\overline{\hat{w}}-\mu_n)}{\mu_n^3}, \qquad y_3 = \frac{3\hat{w}_1^2(\overline{\hat{w}}-\mu_n)^2}{\mu_n^4}.$$

It is obvious that

$$\Big|\mathbb{E}\Big(y_1 I_{\hat{B}_p^c(\varepsilon)}\Big)\Big| \le C_2 \Big|\mathbb{E}\hat{w}_1^2 I_{\hat{B}_p^c(\varepsilon)}\Big| \le C_2 \mathbb{P}^{1/2}(\hat{B}_p^c(\varepsilon)), \tag{S2.18}$$

where $C_2 > 0$ is a constant. Note that

$$\frac{\mathbb{E}\hat{w}_1^2(\overline{\hat{w}}-\mu_n)}{\mu_n^3} = \frac{1}{p}\frac{\mathbb{E}\hat{w}_1^2(\hat{w}_1-\mu_n)}{\mu_n^3}, \tag{S2.19}$$

$$\frac{\mathbb{E}\hat{w}_1^2(\overline{\hat{w}}-\mu_n)^2}{\mu_n^4}=\frac{1}{p^2}\frac{\mathbb{E}\hat{w}_1^2(\hat{w}_1-\mu_n)^2}{\mu_n^4}+\frac{p-1}{p^2}\frac{\mathbb{E}\hat{w}_1^2\mathbb{E}(\hat{w}_1-\mu_n)^2}{\mu_n^4}. \tag{S2.20}$$

By (S2.19), we get

$$\begin{aligned}\left|\mathbb{E}\Big(y_2 I_{\hat{B}_p^c(\varepsilon)}\Big)\right| &= \frac{2}{p\mu_n^3}\left|\mathbb{E}\hat{w}_1^2(\hat{w}_1-\mu_n)I_{\hat{B}_p^c(\varepsilon)}\right| \\ &\le \frac{C_3}{p}\Big(\left|\mathbb{E}\hat{w}_1^2 I_{\hat{B}_p^c(\varepsilon)}\right|+\left|\mathbb{E}\hat{w}_1^3 I_{\hat{B}_p^c(\varepsilon)}\right|\Big) \\ &\le \frac{C_3}{p}\left\{\mathbb{P}^{1/2}(\hat{B}_p^c(\varepsilon))+\left|(\mathbb{E}\hat{w}_1^4)^{1/2}\Big(\mathbb{E}\hat{w}_1^2 I_{\hat{B}_p^c(\varepsilon)}\Big)^{1/2}\right|\right\} \\ &\le \frac{C_3}{p}\left\{\mathbb{P}^{1/2}(\hat{B}_p^c(\varepsilon))+\mathbb{P}^{1/4}(\hat{B}_p^c(\varepsilon))\right\}, \end{aligned} \tag{S2.21}$$

where $C_3>0$ is a constant. By (S2.20), we get

$$\begin{aligned}\left|\mathbb{E}\Big(y_3 I_{\hat{B}_p^c(\varepsilon)}\Big)\right| &\le \frac{1}{p^2\mu_n^4}\left|\mathbb{E}\hat{w}_1^2(\hat{w}_1-\mu_n)^2 I_{\hat{B}_p^c(\varepsilon)}\right|+\frac{p-1}{p^2\mu_n^4}\left|\Big\{\mathbb{E}\hat{w}_1^2 I_{\hat{B}_p^c(\varepsilon)}\Big\}\Big\{\mathbb{E}(\hat{w}_1-\mu_n)^2 I_{\hat{B}_p^c(\varepsilon)}\Big\}\right| \\ &\le \frac{C_4}{p^2}+\frac{C_4}{p}\left\{\mathbb{P}^{1/2}(\hat{B}_p^c(\varepsilon))\cdot\mathbb{P}^{1/2}(\hat{B}_p^c(\varepsilon))\right\} \\ &\le C_4\left\{\frac{1}{p^2}+\frac{1}{p}\mathbb{P}(\hat{B}_p^c(\varepsilon))\right\}, \end{aligned} \tag{S2.22}$$

where $C_4>0$ is a constant. By (S2.18) – (S2.22), we have

$$|\mathbb{E}b|\le C_5\left\{\mathbb{P}^{1/2}(\hat{B}_p^c(\varepsilon))+\frac{1}{p}\mathbb{P}^{1/4}(\hat{B}_p^c(\varepsilon))+\frac{1}{p^2}\right\}, \tag{S2.23}$$

where $C_5>0$ is a constant. Now, we provide an estimate for $\mathbb{P}(\hat{B}_p^c(\varepsilon))$. By Markov's inequality and Burkhölder inequality, we get

$$\begin{aligned}\mathbb{P}(\hat{B}_p^c(\varepsilon)) &= \mathbb{P}(|\overline{\hat{w}}-\mu_n|\ge\varepsilon) \\ &\le \varepsilon^{-t}\mathbb{E}\left|\frac{1}{p}\sum_{j=1}^p(\hat{w}_j-\mu_n)\right|^t \\ &\le C_t\varepsilon^{-t}p^{-t}\left\{\mathbb{E}\left(\sum_{j=1}^p\mathbb{E}_{j-1}|\hat{w}_j-\mu_n|^2\right)^{t/2}+\mathbb{E}\sum_{j=1}^p|\hat{w}_j-\mu_n|^t\right\} \\ &= C_t\varepsilon^{-t}p^{-t}\left\{\left(\sum_{j=1}^p\mathbb{E}|\hat{w}_j-\mu_n|^2\right)^{t/2}+\mathbb{E}\sum_{j=1}^p|\hat{w}_j-\mu_n|^t\right\} \\ &= C_t\varepsilon^{-t}p^{-t}\left\{\left(p\mathbb{E}|\hat{w}_1-\mu_n|^2\right)^{t/2}+p\mathbb{E}|\hat{w}_1-\mu_n|^t\right\} \end{aligned}$$

where ε, t, and C_t (which depends on t) are positive constants. Since $\mathbb{E}\left|w_1-\mu\right|^4<\infty$, and the truncation $|w_j-\mu|<\delta_n\sqrt{n}$, where $\{\delta_n\}$ is a positive sequence satisfying $\delta_n\to 0$, $\delta_n n^{1/4}\to\infty$, $\delta_n^{-4}\mathbb{E}|w_1-\mu|^4 I_{\{|w_1-\mu|\ge\delta_n\sqrt{n}\}}\to 0$ as $n\to\infty$, we have $\sup_n\mathbb{E}\left|\hat{w}_1-\mu_n\right|^4<\infty$ and $|\hat{w}_j-\mu_n|<4\delta_n\sqrt{n}$ for sufficiently large n (note that $\mu\le\delta_n\sqrt{n}$ for sufficiently large n). Based on the estimate for $\mathbb{P}(\hat{B}_p^c(\varepsilon))$,

$\sup_n \mathbb{E}\left|\hat{w}_1-\mu_n\right|^4<\infty$ and $|\hat{w}_j-\mu_n|<4\delta_n\sqrt{n}$ where $\{\delta_n\}$ is a positive sequence satisfying $\delta_n\to 0$, $\delta_n n^{1/4}\to\infty$, $\delta_n^{-4}\mathbb{E}|w_1-\mu|^4 I_{\{|w_1-\mu|\ge\delta_n\sqrt{n}\}}\to 0$ as $n\to\infty$, we have $\mathbb{E}a-\mathbb{E}b+\mathbb{E}c=o(p^{-1})$. From (S2.16), (S2.17) and (S2.23), we get

$$\mathbb{E}\frac{\hat{w}_1^2}{\overline{\hat{w}}^2}-\mathbb{E}\frac{\hat{w}_1^2}{\mu_n^2}=-\frac{2\mathbb{E}\hat{w}_1^2(\overline{\hat{w}}-\mu_n)}{\mu_n^3}+\frac{3\mathbb{E}\hat{w}_1^2(\overline{\hat{w}}-\mu_n)^2}{\mu_n^4}+o(p^{-1}). \tag{S2.24}$$

Plugging (S2.19) – (S2.20) into (S2.24), we get

$$\mathbb{E}\frac{\hat{w}_1^2}{\overline{\hat{w}}^2}-\mathbb{E}\frac{\hat{w}_1^2}{\mu_n^2}=\frac{h_{1,n}}{p}+o(p^{-1}),\quad h_{1,n}=-2\frac{\mathbb{E}\hat{w}_1^3}{\mu_n^3}+3\lambda_n^2+5\lambda_n+2,\quad \lambda_n=\mathbb{E}\Big(\frac{\hat{w}_1}{\mu_n}-1\Big)^2,$$

and hence

$$\mathbb{E}\frac{\hat{w}_1^2}{\overline{\hat{w}}^2}-\mathbb{E}\frac{w_1^2}{\mu^2}=\frac{h_1}{p}+o(p^{-1}),\quad h_1=-2\frac{\mathbb{E}w_1^3}{\mu^3}+3\lambda^2+5\lambda+2,$$

which implies the first equation in Lemma 5.3.

Similar to the previous calculation, we obtain

$$\begin{aligned}
&\Big(\frac{\hat{w}_1}{\overline{\hat{w}}}-1\Big)^2\Big(\frac{\hat{w}_2}{\overline{\hat{w}}}-1\Big)^2 I_{\hat{B}_p(\varepsilon)}\\
&=\Big(\frac{\hat{w}_1}{\mu_n}-1\Big)^2\Big(\frac{\hat{w}_2}{\mu_n}-1\Big)^2 I_{\hat{B}_p(\varepsilon)}+\Big(\frac{\overline{\hat{w}}}{\mu_n}-1\Big)f_1(\hat{w}_1,\hat{w}_2,\mu_n)I_{\hat{B}_p(\varepsilon)}\\
&\qquad+\Big(\frac{\overline{\hat{w}}}{\mu_n}-1\Big)^2 f_2(\hat{w}_1,\hat{w}_2,\mu_n)I_{\hat{B}_p(\varepsilon)}+\tilde{a},
\end{aligned}$$

where $|\mathbb{E}\tilde{a}|=o(p^{-1})$,

$$\begin{aligned}
f_1(\hat{w}_1,\hat{w}_2,\mu_n)&=-\frac{2\hat{w}_1}{\mu_n}\Big(\frac{\hat{w}_1}{\mu_n}-1\Big)\Big(\frac{\hat{w}_2}{\mu_n}-1\Big)^2-\frac{2\hat{w}_2}{\mu_n}\Big(\frac{\hat{w}_2}{\mu_n}-1\Big)\Big(\frac{\hat{w}_1}{\mu_n}-1\Big)^2,\\
f_2(\hat{w}_1,\hat{w}_2,\mu_n)&=4\frac{\hat{w}_1}{\mu_n}\frac{\hat{w}_2}{\mu_n}\Big(\frac{\hat{w}_1}{\mu_n}-1\Big)\Big(\frac{\hat{w}_2}{\mu_n}-1\Big)\\
&\quad+\Big(\frac{\hat{w}_1}{\mu_n}-1\Big)^2\Big(\frac{\hat{w}_2}{\mu_n}-1\Big)^2\Big\{\frac{\hat{w}_1^2/\mu_n^2}{(\hat{w}_1/\mu_n-1)^2}+\frac{2\hat{w}_1/\mu_n}{\hat{w}_1/\mu_n-1}\Big\}\\
&\quad+\Big(\frac{\hat{w}_1}{\mu_n}-1\Big)^2\Big(\frac{\hat{w}_2}{\mu_n}-1\Big)^2\Big\{\frac{\hat{w}_2^2/\mu_n^2}{(\hat{w}_2/\mu_n-1)^2}+\frac{2\hat{w}_2/\mu_n}{\hat{w}_2/\mu_n-1}\Big\}.
\end{aligned}\tag{S2.25}$$

Similar to (S2.14) – (S2.24), we get

$$\begin{aligned}
&\mathbb{E}\Big(\frac{\hat{w}_1}{\overline{\hat{w}}}-1\Big)^2\Big(\frac{\hat{w}_2}{\overline{\hat{w}}}-1\Big)^2\\
&=\mathbb{E}\Big(\frac{\hat{w}_1}{\mu_n}-1\Big)^2\Big(\frac{\hat{w}_2}{\mu_n}-1\Big)^2+\mathbb{E}\Big(\frac{\overline{\hat{w}}}{\mu_n}-1\Big)f_1(\hat{w}_1,\hat{w}_2,\mu_n)+\mathbb{E}\Big(\frac{\overline{\hat{w}}}{\mu_n}-1\Big)^2 f_2(\hat{w}_1,\hat{w}_2,\mu_n)+o(p^{-1})\\
&=:T_1+T_2+T_3+o(p^{-1}).
\end{aligned}\tag{S2.26}$$

Similar to (S2.19) and (S2.20), we obtain

$$T_2=\frac{1}{p}\mathbb{E}\Big(\frac{\hat{w}_1}{\mu_n}+\frac{\hat{w}_2}{\mu_n}-2\Big)f_1(\hat{w}_1,\hat{w}_2,\mu_n) \tag{S2.27}$$

and

$$T_3 = \frac{1}{p^2}\sum_{i=1}^{2}\mathbb{E}\Big(\frac{\hat{w}_i}{\mu_n}-1\Big)^2 f_2(\hat{w}_1,\hat{w}_2,\mu_n)+\frac{p-2}{p^2}\mathbb{E}\Big(\frac{\hat{w}_1}{\mu_n}-1\Big)^2\mathbb{E}f_2(\hat{w}_1,\hat{w}_2,\mu_n). \tag{S2.28}$$

Thus, by (S2.26) – (S2.28), we get

$$\begin{aligned}\mathbb{E}\Big(\frac{\hat{w}_1}{\bar{\hat{w}}}-1\Big)^2\Big(\frac{\hat{w}_2}{\bar{\hat{w}}}-1\Big)^2 &= \mathbb{E}\Big(\frac{\hat{w}_1}{\mu_n}-1\Big)^2\Big(\frac{\hat{w}_2}{\mu_n}-1\Big)^2+\frac{1}{p}\mathbb{E}\Big(\frac{\hat{w}_1}{\mu_n}+\frac{\hat{w}_2}{\mu_n}-2\Big)f_1(\hat{w}_1,\hat{w}_2,\mu_n)\\ &\quad+\frac{\lambda_n}{p}\mathbb{E}f_2(\hat{w}_1,\hat{w}_2,\mu_n)+o(p^{-1})\\ &= \mathbb{E}\Big(\frac{\hat{w}_1}{\mu_n}-1\Big)^2\Big(\frac{\hat{w}_2}{\mu_n}-1\Big)^2+\frac{1}{p}h_{2,n}+o(p^{-1}),\end{aligned} \tag{S2.29}$$

where $h_{2,n}=-8\lambda_n\frac{\mathbb{E}\hat{w}_1^3}{\mu_n^3}+10\lambda_n^3+22\lambda_n^2+8\lambda_n$, and hence

$$\mathbb{E}\Big(\frac{\hat{w}_1}{\bar{\hat{w}}}-1\Big)^2\Big(\frac{\hat{w}_2}{\bar{\hat{w}}}-1\Big)^2=\mathbb{E}\Big(\frac{w_1}{\mu}-1\Big)^2\Big(\frac{w_2}{\mu}-1\Big)^2+\frac{1}{p}h_2+o(p^{-1}),$$

where $h_2=-8\lambda\frac{\mathbb{E}w_1^3}{\mu^3}+10\lambda^3+22\lambda^2+8\lambda$, which implies the second equation in Lemma 5.3.

Similarly, we get

$$\mathbb{E}\Big(\frac{\hat{w}_1}{\bar{\hat{w}}}-1\Big)^4=\mathbb{E}\Big(\frac{w_1}{\mu}-1\Big)^4+o(1),$$

which implies the third equation in Lemma 5.3.

S2.4 Proof of Lemma 5.4

First, we establish an estimate for $\mathbb{P}(B_p^c(\varepsilon))$, where $B_p(\varepsilon)=\{\omega:|\overline{w}-u|\le\varepsilon,\overline{w}=\sum_{j=1}^p w_j/p\}$. By Markov's inequality and Burkhölder inequality, we get

$$\begin{aligned}\mathbb{P}(B_p^c(\varepsilon)) &= \mathbb{P}(|\overline{w}-\mu|\ge\varepsilon)\\ &\le \varepsilon^{-t}\mathbb{E}\Big|\frac{1}{p}\sum_{j=1}^{p}(w_j-\mu)\Big|^t\\ &\le K_t\varepsilon^{-t}p^{-t}\Big\{\mathbb{E}\Big(\sum_{j=1}^{p}\mathbb{E}_{j-1}|w_j-\mu|^2\Big)^{t/2}+\mathbb{E}\sum_{j=1}^{p}|w_j-\mu|^t\Big\}\\ &= K_t\varepsilon^{-t}p^{-t}\Big\{\Big(\sum_{j=1}^{p}\mathbb{E}|w_j-\mu|^2\Big)^{t/2}+\mathbb{E}\sum_{j=1}^{p}|w_j-\mu|^t\Big\}\\ &= K_t\varepsilon^{-t}\Big\{p^{-t/2}\Big(\mathbb{E}|w_1-\mu|^2\Big)^{t/2}+p^{-t+1}\mathbb{E}|w_1-\mu|^t\Big\},\end{aligned} \tag{S2.30}$$

where ε, t, and K_t (which depends on t) are positive constants. We aim to derive an estimate for the q-th moment of $\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{1}{n}\mathbb{E}\Big(\frac{w_1}{\overline{w}}-1\Big)^2\mathrm{tr}\mathbf{A}$. For any $q\ge2$, there exists a positive constant K_q such that

$$\mathbb{E}\Big|\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{1}{n}\mathbb{E}\Big(\frac{w_1}{\overline{w}}-1\Big)^2\mathrm{tr}\mathbf{A}\Big|^q\le K_q\Big(\mathbb{E}\Big|\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\lambda}{n}\mathrm{tr}\mathbf{A}\Big|^q+\mathbb{E}\Big|\frac{\lambda}{n}\mathrm{tr}\mathbf{A}-\frac{1}{n}\mathbb{E}\Big(\frac{w_1}{\overline{w}}-1\Big)^2\mathrm{tr}\mathbf{A}\Big|^q\Big).$$

Note that

$$\mathbb{E}\left|\frac{\lambda}{n}\mathrm{tr}\mathbf{A}-\frac{1}{n}\mathbb{E}\Big(\frac{w_1}{\overline{w}}-1\Big)^2\mathrm{tr}\mathbf{A}\right|^q \le K_q\left|\mathbb{E}\Big(\frac{w_1}{\overline{w}}-1\Big)^2-\lambda\right|^q\|\mathbf{A}\|^q. \tag{S2.31}$$

Now, we consider $\mathbb{E}\left|\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\lambda}{n}\mathrm{tr}\mathbf{A}\right|^q$. There exists a positive constant K_q such that

$$\mathbb{E}\left|\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\lambda}{n}\mathrm{tr}\mathbf{A}\right|^q \le K_q\left\{\mathbb{E}\left|\Big(\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\lambda}{n}\mathrm{tr}\mathbf{A}\Big)I_{B_p^c(\varepsilon)}\right|^q+\mathbb{E}\left|\Big(\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\lambda}{n}\mathrm{tr}\mathbf{A}\Big)I_{B_p(\varepsilon)}\right|^q\right\}. \tag{S2.32}$$

Estimating $\mathbb{E}|(\mathbf{r}'\mathbf{A}\mathbf{r}-n^{-1}\lambda\mathrm{tr}\mathbf{A})I_{B_p^c(\varepsilon)}|^q$: Since $\left|\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\lambda}{n}\mathrm{tr}\mathbf{A}\right|\le\|\mathbf{r}\|\|\mathbf{A}\|+C\|\mathbf{A}\|\le C\cdot p\|\mathbf{A}\|$, C is a positive constant, we have

$$\mathbb{E}\left|\Big(\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\lambda}{n}\mathrm{tr}\mathbf{A}\Big)I_{B_p^c(\varepsilon)}\right|^q \le C_q p^q\|\mathbf{A}\|^q\mathbb{P}(B_p^c(\varepsilon)), \tag{S2.33}$$

where C_q denotes a positive constant depending on q.
Estimating $\mathbb{E}|(\mathbf{r}'\mathbf{A}\mathbf{r}-n^{-1}\lambda\mathrm{tr}\mathbf{A})I_{B_p(\varepsilon)}|^q$: Write

$$\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\lambda}{n}\mathrm{tr}\mathbf{A}=\frac{\sigma^2}{n\overline{w}^2}\left\{\frac{(\mathbf{w}-\overline{w}\mathbf{1}_p)'\mathbf{A}(\mathbf{w}-\overline{w}\mathbf{1}_p)}{\sigma^2}-\mathrm{tr}\mathbf{A}\right\}+\frac{\sigma^2/\overline{w}^2-\lambda}{n}\mathrm{tr}\mathbf{A}=:v_1+v_2, \tag{S2.34}$$

where $\mathbf{1}_p=(1,1,\dots,1)'\in\mathbb{R}^p$. For $0<\varepsilon<1/2$, there exists a positive constant K_q such that

$$\mathbb{E}\left|\Big(\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\lambda}{n}\mathrm{tr}\mathbf{A}\Big)I_{B_p(\varepsilon)}\right|^q \le K_q\Big(\mathbb{E}|v_1I_{B_p(\varepsilon)}|^q+\mathbb{E}|v_2I_{B_p(\varepsilon)}|^q\Big).$$

On the event $B_p(\varepsilon)$, we have $-\varepsilon\le\overline{w}-\mu\le\varepsilon$, and

$$\begin{aligned}\mathbb{E}|v_2I_{B_p(\varepsilon)}|^q &= \sigma^{2q}\left|\frac{\mathrm{tr}\mathbf{A}}{n}\right|^q\mathbb{E}\left|\Big(\frac{1}{\overline{w}^2}-\frac{1}{\mu^2}\Big)I_{B_p(\varepsilon)}\right|^q\\ &\le \sigma^{2q}\|\mathbf{A}\|^q\mathbb{E}\left|\frac{(\overline{w}-\mu)(\overline{w}+\mu)}{\overline{w}^2\mu^2}I_{B_p(\varepsilon)}\right|^q\\ &\le K_q\|\mathbf{A}\|^q\mathbb{E}|\overline{w}-\mu|^q\\ &\le K_q\|\mathbf{A}\|^q\Big(p^{-q/2}+p^{-q+1}\mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^q\Big),\end{aligned} \tag{S2.35}$$

where the last inequality follows from the same argument in the proof of (S2.30). By using $\mathbf{w}-\overline{w}\mathbf{1}_p=\mathbf{w}-\mu\mathbf{1}_p-(\overline{w}\mathbf{1}_p-\mu\mathbf{1}_p)$, we get

$$\begin{aligned}\mathbb{E}|v_1I_{B_p(\varepsilon)}|^q &\le K_qn^{-q}\mathbb{E}\left|\left\{\frac{(\mathbf{w}-\overline{w}\mathbf{1}_p)'}{\sigma}\mathbf{A}\frac{(\mathbf{w}-\overline{w}\mathbf{1}_p)}{\sigma}-\mathrm{tr}\mathbf{A}\right\}I_{B_p(\varepsilon)}\right|^q\\ &\le K_qn^{-q}\mathbb{E}\left|\left\{\frac{(\mathbf{w}-\mu\mathbf{1}_p)'}{\sigma}\mathbf{A}\frac{(\mathbf{w}-\mu\mathbf{1}_p)}{\sigma}-\mathrm{tr}\mathbf{A}\right\}I_{B_p(\varepsilon)}\right|^q\\ &\quad+K_qn^{-q}\mathbb{E}\left|\left\{\frac{(\overline{w}\mathbf{1}_p-\mu\mathbf{1}_p)'}{\sigma}\mathbf{A}\frac{(\mathbf{w}-\mu\mathbf{1}_p)}{\sigma}\right\}I_{B_p(\varepsilon)}\right|^q\\ &\quad+K_qn^{-q}\mathbb{E}\left|\left\{\frac{(\mathbf{w}-\mu\mathbf{1}_p)'}{\sigma}\mathbf{A}\frac{(\overline{w}\mathbf{1}_p-\mu\mathbf{1}_p)}{\sigma}\right\}I_{B_p(\varepsilon)}\right|^q\\ &\quad+K_qn^{-q}\mathbb{E}\left|\left\{\frac{(\overline{w}\mathbf{1}_p-\mu\mathbf{1}_p)'}{\sigma}\mathbf{A}\frac{(\overline{w}\mathbf{1}_p-\mu\mathbf{1}_p)}{\sigma}\right\}I_{B_p(\varepsilon)}\right|^q\end{aligned}$$

$$=: K_q n^{-q}(V_{11}+V_{12}+V_{13}+V_{14}). \tag{S2.36}$$

By Lemma 2.2 in Bai and Silverstein [2004], we have

$$V_{11} \le K_q\left[\left\{\mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^4 \operatorname{tr}(\mathbf{A}\mathbf{A}')\right\}^{q/2} + \mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^{2q} \operatorname{tr}(\mathbf{A}\mathbf{A}')^{q/2}\right]. \tag{S2.37}$$

From $(\overline{w}-\mu)\mathbf{1}_p' = \frac{1}{p}\mathbf{1}_p\mathbf{1}_p'(\mathbf{w}-\mu\mathbf{1}_p)$ and $p^{-1}\mathbb{E}\operatorname{tr}(\mathbf{1}_p\mathbf{1}_p'\mathbf{A}) = p^{-2}\mathbb{E}\operatorname{tr}(\mathbf{1}_p\mathbf{1}_p'\mathbf{A}\mathbf{1}_p\mathbf{1}_p') \le \|\mathbf{A}\|$, we get

$$\begin{aligned}
V_{12} &= \mathbb{E}\left|\left\{\frac{1}{p}\frac{(\mathbf{w}-\mu\mathbf{1}_p)'}{\sigma}\mathbf{1}_p\mathbf{1}_p'\mathbf{A}\frac{(\mathbf{w}-\mu\mathbf{1}_p)}{\sigma}\right\}I_{B_p(\varepsilon)}\right|^q \\
&\le K_q\left\{\mathbb{E}\left|\frac{(\mathbf{w}-\mu\mathbf{1}_p)'}{\sigma}\left(\frac{1}{p}\mathbf{1}_p\mathbf{1}_p'\mathbf{A}\right)\frac{(\mathbf{w}-\mu\mathbf{1}_p)}{\sigma} - \operatorname{tr}\left(\frac{1}{p}\mathbf{1}_p\mathbf{1}_p'\mathbf{A}\right)\right|^q + \mathbb{E}\left|\operatorname{tr}\left(\frac{1}{p}\mathbf{1}_p\mathbf{1}_p'\mathbf{A}\right)\right|^q\right\} \\
&\le K_q\left\{\mathbb{E}\left|\frac{(\mathbf{w}-\mu\mathbf{1}_p)'}{\sigma}\left(\frac{1}{p}\mathbf{1}_p\mathbf{1}_p'\mathbf{A}\right)\frac{(\mathbf{w}-\mu\mathbf{1}_p)}{\sigma} - \operatorname{tr}\left(\frac{1}{p}\mathbf{1}_p\mathbf{1}_p'\mathbf{A}\right)\right|^q + \|\mathbf{A}\|^q\right\} \\
&\le K_q\left[\left\{\mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^4 \operatorname{tr}\left(\frac{1}{p^2}\mathbf{1}_p\mathbf{1}_p'\mathbf{A}\mathbf{A}'\mathbf{1}_p\mathbf{1}_p'\right)\right\}^{q/2} + \mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^{2q}\operatorname{tr}\left(\frac{1}{p^2}\mathbf{1}_p\mathbf{1}_p'\mathbf{A}\mathbf{A}'\mathbf{1}_p\mathbf{1}_p'\right)^{q/2}\right] + K_q\|\mathbf{A}\|^q \\
&\le V_{11}.
\end{aligned} \tag{S2.38}$$

Similarly, we get

$$V_{13} \le V_{11}, \qquad V_{14} \le V_{11}. \tag{S2.39}$$

By (S2.36) – (S2.39), we get

$$\mathbb{E}|v_1 I_{B_p(\varepsilon)}|^q \le K_q n^{-q}\left[\left\{\mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^4 \operatorname{tr}(\mathbf{A}\mathbf{A}')\right\}^{q/2} + \mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^{2q} \operatorname{tr}(\mathbf{A}\mathbf{A}')^{q/2}\right]. \tag{S2.40}$$

From (S2.32) – (S2.35) and (S2.40), we have

$$\begin{aligned}
&\mathbb{E}\left|\mathbf{r}'\mathbf{A}\mathbf{r} - \frac{\lambda}{n}\operatorname{tr}\mathbf{A}\right|^q \\
&\le K_q\left\{n^{-q}\left[\left\{\mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^4 \operatorname{tr}(\mathbf{A}\mathbf{A}')\right\}^{q/2} + \mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^{2q} \operatorname{tr}(\mathbf{A}\mathbf{A}')^{q/2}\right]\right. \\
&\qquad\left. + \|\mathbf{A}\|^q\left(p^{-q/2} + p^{-q+1}\mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^q\right) + p^q\|\mathbf{A}\|^q\mathbb{P}(B_p^c(\varepsilon))\right\} \\
&\le K_q\left(n^{-q}\left[\left\{\mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^4 \operatorname{tr}(\mathbf{A}\mathbf{A}')\right\}^{q/2} + \mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^{2q} \operatorname{tr}(\mathbf{A}\mathbf{A}')^{q/2}\right] + n^q\|\mathbf{A}\|^q\mathbb{P}(B_p^c(\varepsilon))\right).
\end{aligned}$$

By this inequality and (S2.31), we get

$$\begin{aligned}
\mathbb{E}\left|\mathbf{r}'\mathbf{A}r - \frac{1}{n}\mathbb{E}\left(\frac{w_1}{\overline{w}} - 1\right)^2 \operatorname{tr}\mathbf{A}\right|^q &\le K_q\left(n^{-q}\left[\left\{\mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^4 \operatorname{tr}(\mathbf{A}\mathbf{A}')\right\}^{q/2} + \mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^{2q} \operatorname{tr}(\mathbf{A}\mathbf{A}')^{q/2}\right]\right. \\
&\qquad\left. + n^q\|\mathbf{A}\|^q\mathbb{P}(B_p^c(\varepsilon)) + \left|\mathbb{E}\left(\frac{w_1}{\overline{w}} - 1\right)^2 - \lambda\right|^q \|\mathbf{A}\|^q\right).
\end{aligned}$$

Finally, we prove the last inequality in Lemma 5.4. Furthermore, if $\|\mathbf{A}\|$ is bounded, $\mathbb{E}\left|w_1-\mu\right|^4 < \infty$, $|w_j-\mu| < \delta_n\sqrt{n}$ $(j = 1,\ldots,p)$ where $\{\delta_n\}$ is a positive sequence satisfying $\delta_n \to 0$, $\delta_n n^{1/4} \to \infty$, $\delta_n^{-4}\mathbb{E}|w_1-\mu|^4 I_{\{|w_1-\mu|\ge\delta_n\sqrt{n}\}} \to 0$ as $n \to \infty$, then, for any $q \ge 2$,

$$n^{-q}\left[\left\{\mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^4 \operatorname{tr}(\mathbf{A}\mathbf{A}')\right\}^{q/2} + \mathbb{E}\left|\frac{w_1-\mu}{\sigma}\right|^{2q} \operatorname{tr}(\mathbf{A}\mathbf{A}')^{q/2}\right] \le K_q n^{-1}\delta_n^{2q-4}$$

and

$$\left|\mathbb{E}\Big(\frac{w_1}{\bar{w}}-1\Big)^2-\lambda\right|^q\|\mathbf{A}\|^q \le K_q n^{-q}\|\mathbf{A}\|^q \le K_q n^{-1}\delta_n^{2q-4}.$$

Taking $\varepsilon = n^{-\alpha}$, $0<\alpha<1/2$, and $t>\frac{3q}{1-2\alpha}$ yields that

$$n^q\|\mathbf{A}\|^q\mathbb{P}(B_p^c(\varepsilon)) \le K_q n^{-1}\delta_n^{2q-4}.$$

In fact, from (S2.30), we write

$$\mathbb{P}(B_p^c(\varepsilon)) \le K_t\varepsilon^{-t}\Big\{p^{-t/2}\Big(\mathbb{E}|w_1-\mu|^2\Big)^{t/2}+p^{-t+1}\mathbb{E}|w_1-\mu|^t\Big\} =: P_1+P_2.$$

Since $P_2 \le K_t\varepsilon^{-t}p^{-t+1}(n^{1/2}\delta_n)^{t-4} \le K_t\delta_n^{t-4}\varepsilon^{-t}n^{-t/2-1} \le P_1$, we obtain that

$$\mathbb{P}(B_p^c(\varepsilon)) \le 2P_1. \tag{S2.41}$$

Take $\varepsilon = n^{-\alpha}$ $(0<\alpha<1/2)$ and $t>\frac{3q}{1-2\alpha}$ into (S2.41), we have

$$n^q\mathbb{P}(B_p^c(\varepsilon)) \le K_q n^q n^{-(\frac{1}{2}-\alpha)t} \le K_q n^{-q/2} \le K_q n^{-1}\delta_n^{2q-4}.$$

Combining all these estimates, we obtain the last inequality in the Lemma.

S2.5 Proof of Lemma 5.5

First, we denote $R_j = \frac{1}{\sqrt{n}}\Big(\frac{\hat{w}_j}{\bar{\hat{w}}}-1\Big)$ and derive some identities that will be used in the proof. It is obvious that $\sum_{j=1}^p R_j = 0$ and $\sum_{j=1}^p R_j^2+\sum_{j_1\neq j_2}R_{j_1}R_{j_2}=0$. Since $\{\hat{w}_j\}_{j=1}^p$ are i.i.d., taking expectation on the above two identities yields that, for any $1\le i\neq j\le p$,

$$\mathbb{E}R_j = 0, \quad \mathbb{E}R_j^2 = \frac{\nu_2}{n}, \quad \mathbb{E}R_iR_j = -\frac{\nu_2}{n(p-1)}, \quad \mathbb{E}R_j^4 = \frac{\nu_4}{n^2}, \quad \mathbb{E}R_i^2R_j^2 = \frac{\nu_{12}}{n^2}. \tag{S2.42}$$

Recall that $\mathbf{r} = (R_1,\ldots,R_p)'$. From (S2.42), we have

$$\mathbb{E}(\mathbf{r}\mathbf{r}') = \frac{p\nu_2}{n(p-1)}\Big(\mathbf{I}_p-\frac{1}{p}\mathbf{1}_p\mathbf{1}_p'\Big), \tag{S2.43}$$

where $\mathbf{1}_p = (1,1,\ldots,1)'\in\mathbb{R}^p$. It is obvious that

$$\begin{aligned}&\mathbb{E}\Big(\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\nu_2}{n}\mathrm{tr}\mathbf{A}\Big)\Big(\mathbf{r}'\mathbf{B}\mathbf{r}-\frac{\nu_2}{n}\mathrm{tr}\mathbf{B}\Big)\\ &= \mathbb{E}\Big(\mathbf{r}'\mathbf{A}\mathbf{r}\mathbf{r}'\mathbf{B}\mathbf{r}\Big)-\frac{\nu_2\mathrm{tr}\mathbf{B}}{n}\mathbb{E}\Big(\mathbf{r}'\mathbf{A}\mathbf{r}-\frac{\nu_2}{n}\mathrm{tr}\mathbf{A}\Big)-\frac{\nu_2\mathrm{tr}\mathbf{A}}{n}\mathbb{E}\Big(\mathbf{r}'\mathbf{B}\mathbf{r}-\frac{\nu_2}{n}\mathrm{tr}\mathbf{B}\Big)-\frac{\nu_2^2}{n^2}\mathrm{tr}\mathbf{A}\mathrm{tr}\mathbf{B}.\end{aligned} \tag{S2.44}$$

We first estimate each terms in the RHS of the identity above, and finally prove the equation given in Lemma 5.5. The details are provided in the following three steps.
Step 1: Estimate the first term. It follows from (S2.42) and the identity $\sum_{j=1}^p R_j = 0$ that

$$\mathbb{E}R_1^3R_2 = \frac{1}{p-1}\mathbb{E}\Big[R_1^3\Big(\sum_{j=1}^p R_j - R_1\Big)\Big] = -\frac{\nu_4}{n^2(p-1)}, \tag{S2.45}$$

$$
\begin{aligned}
\mathbb{E}R_1^2R_2R_3 &= \frac{1}{p-2}\mathbb{E}\left[R_1^2R_2\left(\sum_{j=1}^{p}R_j - R_1 - R_2\right)\right] \\
&= -\frac{1}{p-2}\mathbb{E}(R_1^3R_2) - \frac{1}{p-2}\mathbb{E}(R_1^2R_2^2) \\
&\overset{(S2.45)}{=} -\frac{\nu_4}{n^2(p-1)(p-2)} - \frac{\nu_{12}}{n^2(p-2)}, \qquad \text{(S2.46)}\\
\mathbb{E}R_1R_2R_3R_4 &= \frac{1}{p-3}\mathbb{E}\left[R_1R_2R_3\left(\sum_{j=1}^{p}R_j - R_1 - R_2 - R_3\right)\right] \\
&= -\frac{3}{p-3}\mathbb{E}(R_1^2R_2R_3) \\
&\overset{(S2.46)}{=} -\frac{3\nu_4}{n^2(p-1)(p-2)(p-3)} + \frac{3\nu_{12}}{n^2(p-2)(p-3)}. \qquad \text{(S2.47)}
\end{aligned}
$$

To calculate $\mathbb{E}(\mathbf{r}'\mathbf{A}\mathbf{r}\mathbf{r}'\mathbf{B}\mathbf{r})$, we expand it as

$$
\mathbb{E}(\mathbf{r}'\mathbf{A}\mathbf{r}\mathbf{r}'\mathbf{B}\mathbf{r}) = \mathbb{E}\left(\sum_{i,j}R_iA_{ij}R_j\sum_{k,\ell}R_kB_{k\ell}R_\ell\right) = \sum_{i,j,k,\ell}\mathbb{E}(R_iR_jR_kR_\ell A_{ij}B_{k\ell}). \qquad \text{(S2.48)}
$$

To calculate (S2.48), we split it into the following 11 cases:

1. $i=j=k=\ell$, $\sum_i(R_i^4)A_{ii}B_{ii}$;
2. $i=j, k=\ell, i\neq k, \sum_{\substack{i,k\\ i\neq k}}(R_i^2R_k^2)A_{ii}B_{kk}$;
3. $i=j, k\neq\ell, \sum_{\substack{i,k,\ell\\ k\neq\ell}}(R_i^2R_kR_\ell)A_{ii}B_{k\ell}$;
4. $i\neq j, k=\ell, \sum_{\substack{i,j,k\\ i\neq j}}(R_iR_jR_k^2)A_{ij}B_{kk}$;
5. $i\neq j, k\neq\ell, i=k, j=\ell, \sum_{\substack{i,j\\ i\neq j}}(R_i^2R_j^2)A_{ij}B_{ij}$;
6. $i\neq j, k\neq\ell, i=\ell, j=k, \sum_{\substack{i,j\\ i\neq j}}(R_i^2R_j^2)A_{ij}B_{ji}$;
7. $i\neq j, k\neq\ell, i=k, \ell\neq j, \sum_{\substack{i,j,\ell\\ i\neq j\neq\ell}}(R_i^2R_jR_\ell)A_{ij}B_{i\ell}$;
8. $i\neq j, k\neq\ell, \ell=j, i\neq k, \sum_{\substack{i,j,k\\ i\neq j\neq k}}(R_iR_j^2R_k)A_{ij}B_{kj}$;
9. $i\neq j, k\neq\ell, k=j, i\neq\ell, \sum_{\substack{i,j,\ell\\ i\neq j\neq\ell}}(R_iR_j^2R_\ell)A_{ij}B_{j\ell}$;
10. $i\neq j, k\neq\ell, i=\ell, k\neq j, \sum_{\substack{i,j,k\\ i\neq j\neq k}}(R_i^2R_jR_k)A_{ij}B_{ki}$;
11. $i\neq j, k\neq\ell, \ell\neq j, i\neq k, \sum_{\substack{i,j,k,\ell\\ i\neq j\neq k\neq\ell}}(R_iR_jR_kR_\ell)A_{ij}B_{k\ell}$.

For ease of presentation, we still keep ν_4 in the expectations although we have obtained its value. The expectations of all cases are listed as follows.

Case 1: From (S2.42), we have

$$\mathbb{E}\sum_{i} R_i^4 A_{ii}B_{ii} = \frac{\nu_4}{n^2}\sum_{i} A_{ii}B_{ii}.$$

Case 2: From (S2.42), we have

$$\mathbb{E}\sum_{\substack{i,k\\ i\neq k}} R_i^2 R_k^2 A_{ii}B_{kk} = \frac{\nu_{12}}{n^2}\sum_{\substack{i,k\\ i\neq k}} A_{ii}B_{kk} = \frac{\nu_{12}}{n^2}\Big(\mathrm{tr}\mathbf{A}\mathrm{tr}\mathbf{B} - \sum_{i} A_{ii}B_{ii}\Big).$$

Case 3: Note that

$$\mathbb{E}\sum_{\substack{i,k,\ell\\ k\neq \ell}} R_i^2 R_k R_\ell A_{ii}B_{k\ell} = \mathbb{E}R_1^2R_2R_3 \sum_{\substack{i,k,\ell\\ i\neq k\neq \ell}} A_{ii}B_{k\ell} + \mathbb{E}R_1^3R_2\Big(\sum_{\substack{i,\ell\\ \ell\neq i}} A_{ii}B_{i\ell} + \sum_{\substack{i,k\\ k\neq i}} A_{ii}B_{ki}\Big). \tag{S2.49}$$

Now, we estimate the magnitude of the summation terms on the RHS of (S2.49). Recall that for any $p\times p$ matrix, we have $\mathbb{E}\mathrm{tr}(\mathbf{M}) \leq p\mathbb{E}\|\mathbf{M}\|$ and $\max_{1\leq i\leq p} M_{ii} \leq \|\mathbf{M}\|$. By using these facts and the Hölder's inequality, we obtain that

$$\mathbb{E}\Big|\sum_{i,k,\ell} A_{ii}B_{k\ell}\Big| = \mathbb{E}\Big|(\mathrm{tr}\mathbf{A})\mathbf{1}_p'\mathbf{B}\mathbf{1}_p\Big| \leq p\mathbb{E}(\|\mathbf{A}\|\cdot|\mathbf{1}_p'\mathbf{B}\mathbf{1}_p|) \leq p^2\mathbb{E}(\|\mathbf{A}\|\cdot\|\mathbf{B}\|) = O(p^2), \tag{S2.50}$$

$$\mathbb{E}\sum_{i,l} A_{ii}B_{\ell\ell} = \mathbb{E}\mathrm{tr}\mathbf{A}\mathrm{tr}\mathbf{B} \leq p^2\mathbb{E}(\|\mathbf{A}\|\cdot\|\mathbf{B}\|) = O(p^2), \tag{S2.51}$$

$$\mathbb{E}\sum_{i} A_{ii}B_{ii} \leq p\mathbb{E}\|\mathrm{diag}(\mathbf{A})\mathrm{diag}(\mathbf{B})\| \leq p\mathbb{E}(\|\mathbf{A}\|\cdot\|\mathbf{B}\|) = O(p). \tag{S2.52}$$

Let $\mathbf{1}_p^i$ be the p-dimensional vector with all components being 0 except for the i-th component being 1, then we have

$$\mathbb{E}\Big|\sum_{i,\ell} A_{ii}B_{i\ell}\Big| = \mathbb{E}\Big|\sum_{i} A_{ii}(\mathbf{1}_p^i)'\mathbf{B}\mathbf{1}_p\Big| \leq \mathbb{E}\Big(\sum_{i} A_{ii}^2\Big)^{1/2}\Big\{\sum_{i}(\mathbf{1}_p^i)'\mathbf{B}\mathbf{1}_p\mathbf{1}_p'\mathbf{B}\mathbf{1}_p^i\Big\}^{1/2} = O(p^{3/2}). \tag{S2.53}$$

From (S2.50) – (S2.52), we have

$$\mathbb{E}\sum_{\substack{i,k,\ell\\ i\neq k\neq \ell}} A_{ii}B_{k\ell} = \mathbb{E}\Big(\sum_{i,k,\ell} A_{ii}B_{k\ell} - \sum_{i,\ell} A_{ii}B_{i\ell} - \sum_{i,k} A_{ii}B_{ki} - \sum_{i,\ell} A_{ii}B_{\ell\ell} + 2\sum_{i} A_{ii}B_{ii}\Big) = O(p^2). \tag{S2.54}$$

It follows from (S2.45), (S2.46), (S2.49), (S2.53), and (S2.54), that

$$\mathbb{E}\sum_{\substack{i,k,\ell\\ k\neq \ell}} (R_i^2R_kR_\ell)A_{ii}B_{k\ell} = -\frac{\nu_{12}}{n^2(p-2)}\sum_{\substack{i,k,\ell\\ i\neq k\neq \ell}} A_{ii}B_{k\ell} + o(n^{-1}).$$

Case 4: Similarly to Case 3, one can conclude that

$$\mathbb{E}\sum_{\substack{i,j,k\\ i\neq j}} R_iR_jR_k^2 A_{ij}B_{kk} = -\frac{\nu_{12}}{n^2(p-2)}\sum_{\substack{i,j,k\\ i\neq j\neq k}} A_{ij}B_{kk} + o(n^{-1}).$$

Case 5:
$$\mathbb{E}\sum_{\substack{i,j\\ i\neq j}} R_i^2R_j^2A_{ij}B_{ij} = \mathbb{E}\left(R_1^2R_2^2\right)\sum_{\substack{i,j\\ i\neq j}} A_{ij}B_{ij} = \frac{\nu_{12}}{n^2}\left\{\mathrm{tr}(\mathbf{A}\mathbf{B}') - \sum_i A_{ii}B_{ii}\right\}.$$

Case 6:
$$\mathbb{E}\sum_{\substack{i,j\\ i\neq j}} R_i^2R_j^2A_{ij}B_{ji} = \mathbb{E}\left(R_1^2R_2^2\right)\sum_{\substack{i,j\\ i\neq j}} A_{ij}B_{ji} = \frac{\nu_{12}}{n^2}\left\{\mathrm{tr}(\mathbf{A}\mathbf{B}) - \sum_i A_{ii}B_{ii}\right\}.$$

Case 7: By (S2.46), we have
$$\mathbb{E}\sum_{\substack{i,j,\ell\\ i\neq j\neq \ell}} R_i^2R_jR_\ell A_{ij}B_{i\ell} = O(n^{-3})\times \sum_{\substack{i,j,\ell\\ i\neq j\neq \ell}} A_{ij}B_{i\ell}. \tag{S2.55}$$

Note that, $\mathbf{1}_p'\mathbf{A}'\mathbf{B}\mathbf{1}_p \leq p\|\mathbf{A}\|\cdot\|\mathbf{B}\| = O(p)$, $\mathrm{tr}(\mathbf{A}'\mathbf{B}) \leq p\|\mathbf{A}\|\cdot\|\mathbf{B}\| = O(p)$, and by (S2.52) and (S2.53), we have
$$\mathbb{E}\sum_{i\neq \ell} A_{ii}B_{i\ell} = \mathbb{E}\sum_{i,\ell} A_{ii}B_{i\ell} - \mathbb{E}\sum_i A_{ii}B_{ii} = O(p^{3/2}), \tag{S2.56}$$

thus,
$$\mathbb{E}\sum_{\substack{i,j,\ell\\ i\neq j\neq \ell}} A_{ij}B_{i\ell} = \mathbb{E}\left\{\mathbf{1}_p'\mathbf{A}'\mathbf{B}\mathbf{1}_p - \mathrm{tr}(\mathbf{A}'\mathbf{B}) - \sum_{i\neq\ell} A_{ii}B_{i\ell} - \sum_{i\neq j} A_{ij}B_{ii}\right\} = O(p^{3/2}). \tag{S2.57}$$

It follows from (S2.55) and (S2.57) that
$$\mathbb{E}\sum_{\substack{i,j,\ell\\ i\neq j\neq \ell}} R_i^2R_jR_\ell A_{ij}B_{i\ell} = o(n^{-1}).$$

Case 8: Similarly to Case 7, we have
$$\mathbb{E}\sum_{\substack{i,j,k\\ i\neq j\neq k}} R_iR_j^2R_kA_{ij}B_{kj} = o(n^{-1}).$$

Case 9: Similarly to Case 7, we have
$$\mathbb{E}\sum_{\substack{i,j,\ell\\ i\neq j\neq \ell}} R_iR_j^2R_\ell A_{ij}B_{j\ell} = o(n^{-1}).$$

Case 10: Similarly to Case 7, we have
$$\mathbb{E}\sum_{\substack{i,j,k\\ i\neq j\neq k}} (R_i^2R_jR_k)A_{ij}B_{ki} = o(n^{-1}).$$

Case 11: By (S2.47), we have

$$\mathbb{E}\sum_{\substack{i,j,k,\ell\\ i\neq j\neq k\neq \ell}} R_iR_jR_kR_\ell A_{ij}B_{k\ell} = O(n^{-4})\sum_{\substack{i,j,k,\ell\\ i\neq j\neq k\neq \ell}} A_{ij}B_{k\ell}.$$

Note that, by (S2.57) we have

$$\begin{aligned}
&\mathbb{E}\sum_{\substack{i,j,k,\ell\\ i\neq j\neq k\neq \ell}} A_{ij}B_{k\ell}\\
&= \mathbb{E}(\mathbf{1}_p'\mathbf{A}\mathbf{1}_p - \mathrm{tr}\mathbf{A})(\mathbf{1}_p'\mathbf{B}\mathbf{1}_p - \mathrm{tr}\mathbf{B}) - \mathbb{E}\sum_{\substack{i,j,\ell\\ i\neq j\neq \ell}} A_{ij}B_{i\ell} - \mathbb{E}\sum_{\substack{i,j,k\\ i\neq j\neq k}} A_{ij}B_{ki} - \mathbb{E}\sum_{\substack{i,j\\ i\neq j}} A_{ij}B_{ij}\\
&= O(p^2).
\end{aligned}$$

Thus,

$$\mathbb{E}\sum_{\substack{i,j,k,\ell\\ i\neq j\neq k\neq \ell}} R_iR_jR_kR_\ell A_{ij}B_{k\ell} = o(n^{-1}).$$

Combining (S2.48) and Cases 1 – 11 gives us

$$\begin{aligned}
\mathbb{E}(\mathbf{r}'\mathbf{A}\mathbf{r}\mathbf{r}'\mathbf{B}\mathbf{r}) =& \frac{\nu_4 - 3\nu_{12}}{n^2}\sum_{i=1}^{p} A_{ii}B_{ii} + \frac{\nu_{12}}{n^2}\Big\{\mathrm{tr}\mathbf{A}\mathrm{tr}\mathbf{B} + \mathrm{tr}(\mathbf{A}\mathbf{B}') + \mathrm{tr}(\mathbf{A}\mathbf{B})\Big\}\\
&- \frac{\nu_{12}}{n^2(p-2)}\Bigg(\sum_{\substack{i,k,\ell\\ i\neq k\neq \ell}} A_{ii}B_{k\ell} + \sum_{\substack{i,j,k\\ i\neq j\neq k}} A_{ij}B_{kk}\Bigg) + o(n^{-1}).
\end{aligned}\tag{S2.58}$$

Step 2: <u>Estimate the second and third terms.</u> From (S2.43) and (S2.56), we have

$$\begin{aligned}
\frac{\nu_2\mathrm{tr}\mathbf{B}}{n}\mathbb{E}\left(\mathbf{r}'\mathbf{A}\mathbf{r} - \frac{\nu_2}{n}\mathrm{tr}\mathbf{A}\right) &= \frac{\nu_2\mathrm{tr}\mathbf{B}}{n}\left\{\mathbb{E}\mathrm{tr}(\mathbf{A}\mathbf{r}\mathbf{r}') - \frac{\nu_2}{n}\mathrm{tr}\mathbf{A}\right\} = -\frac{\nu_2^2\mathrm{tr}\mathbf{B}}{n^2(p-1)}\sum_{k\neq\ell} A_{k\ell}\\
&= -\frac{\nu_2^2}{n^2(p-1)}\Bigg(\sum_{\substack{k,\ell\\ k\neq\ell}} B_{kk}A_{k\ell} + \sum_{\substack{k,\ell\\ k\neq\ell}} B_{\ell\ell}A_{k\ell} + \sum_{\substack{i,k,\ell\\ i\neq k\neq\ell}} B_{ii}A_{k\ell}\Bigg)\\
&= -\frac{\nu_2^2}{n^2(p-1)}\sum_{\substack{i,k,\ell\\ i\neq k\neq\ell}} B_{ii}A_{k\ell} + o(n^{-1}).
\end{aligned}\tag{S2.59}$$

Similarly, we obtain

$$\frac{\nu_2\mathrm{tr}\mathbf{A}}{n}\mathbb{E}\left(\mathbf{r}'\mathbf{B}\mathbf{r} - \frac{\nu_2}{n}\mathrm{tr}\mathbf{B}\right) = -\frac{\nu_2^2}{n^2(p-1)}\sum_{\substack{i,k,\ell\\ i\neq k\neq\ell}} A_{ii}B_{k\ell} + o(n^{-1}).\tag{S2.60}$$

Step 3: From (S2.44) and (S2.58) – (S2.60), we have

$$\mathbb{E}\left(\mathbf{r}'\mathbf{A}\mathbf{r} - \frac{\nu_2}{n}\mathrm{tr}\mathbf{A}\right)\left(\mathbf{r}'\mathbf{B}\mathbf{r} - \frac{\nu_2}{n}\mathrm{tr}\mathbf{B}\right)$$

$$
\begin{aligned}
=&\frac{\nu_4-3\nu_{12}}{n^2}\sum_{i=1}^{p}A_{ii}B_{ii}+\frac{\nu_{12}}{n^2}\Big\{\mathrm{tr}\mathbf{A}\mathrm{tr}\mathbf{B}+\mathrm{tr}(\mathbf{A}\mathbf{B}')+\mathrm{tr}(\mathbf{A}\mathbf{B})\Big\}\\
&-\frac{\nu_{12}}{n^2(p-2)}\Bigg(\sum_{\substack{i,k,\ell\\ i\neq k\neq \ell}}A_{ii}B_{k\ell}+\sum_{\substack{i,j,k\\ i\neq j\neq k}}A_{ij}B_{kk}\Bigg)+\frac{\nu_2^2}{n^2(p-1)}\Bigg(\sum_{\substack{i,k,\ell\\ i\neq k\neq \ell}}B_{ii}A_{k\ell}+\sum_{\substack{i,k,\ell\\ i\neq k\neq \ell}}A_{ii}B_{k\ell}\Bigg)\\
&-\frac{\nu_2^2}{n^2}\mathrm{tr}\mathbf{A}\mathrm{tr}\mathbf{B}+o(n^{-1})\\
=&\frac{\nu_4-3\nu_{12}}{n^2}\sum_{i=1}^{p}A_{ii}B_{ii}+\frac{\nu_{12}}{n^2}\Big\{\mathrm{tr}(\mathbf{A}\mathbf{B}')+\mathrm{tr}(\mathbf{A}\mathbf{B})\Big\}+\frac{\nu_{12}-\nu_2^2}{n^2}\mathrm{tr}\mathbf{A}\mathrm{tr}\mathbf{B}\\
&+\frac{\nu_2^2-\nu_{12}}{n^2(p-2)}\Bigg(\sum_{\substack{i,k,\ell\\ i\neq k\neq \ell}}A_{ii}B_{k\ell}+\sum_{\substack{i,j,k\\ i\neq j\neq k}}A_{ij}B_{kk}\Bigg)+o(n^{-1})\\
=&\frac{\nu_4-3\nu_{12}}{n^2}\sum_{i=1}^{p}A_{ii}B_{ii}+\frac{\nu_{12}}{n^2}\Big\{\mathrm{tr}(\mathbf{A}\mathbf{B}')+\mathrm{tr}(\mathbf{A}\mathbf{B})\Big\}+\frac{\nu_{12}-\nu_2^2}{n^2}\mathrm{tr}\mathbf{A}\mathrm{tr}\mathbf{B}+O(\frac{\nu_2^2-\nu_{12}}{n})+o(n^{-1}),\\
=&\frac{\nu_4-3\nu_{12}}{n^2}\sum_{i=1}^{p}A_{ii}B_{ii}+\frac{\nu_{12}}{n^2}\Big\{\mathrm{tr}(\mathbf{A}\mathbf{B}')+\mathrm{tr}(\mathbf{A}\mathbf{B})\Big\}+\frac{\nu_{12}-\nu_2^2}{n^2}\mathrm{tr}\mathbf{A}\mathrm{tr}\mathbf{B}+o(n^{-1}),
\end{aligned}
$$

where in the third "=" we use Equation (S2.54), and in the last "=" we use Lemma 5.3.

S2.6 Proof of Lemma 5.7

Limit of $\partial^2\mathcal{Y}_1(z_1,z_2)/(\partial z_1\partial z_2)$:

Since $\mathbb{E}\big|\overline{\beta}_j(z)-b_p(z)\big|^2\le K|z|^4/(nv_0^6)$, it is enough to prove that

$$
\begin{aligned}
&\frac{\partial^2}{\partial z_1\partial z_2}\Bigg[b_p(z_1)b_p(z_2)\sum_{j=1}^{n}\mathbb{E}_{j-1}\Big\{\mathbb{E}_j\varepsilon_j(z_1)\mathbb{E}_j\varepsilon_j(z_2)\Big\}\Bigg]\\
&\xrightarrow{P}2\Bigg[\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{\{\underline{m}(z_1)-\underline{m}(z_2)\}^2}-\frac{1}{(z_1-z_2)^2}\Bigg]+\frac{c(\alpha_1+\alpha_2)\underline{m}'(z_1)\underline{m}'(z_2)}{\{1+\lambda\underline{m}(z_1)\}^2\{1+\lambda\underline{m}(z_2)\}^2}.
\end{aligned}
\tag{S2.61}
$$

By Lemma 5.5, we have

$$
\text{LHS of (S2.61)}=\frac{\partial^2}{\partial z_1\partial z_2}\Big\{\mathcal{Y}_{11}(z_1,z_2)+\mathcal{Y}_{12}(z_1,z_2)+\mathcal{Y}_{13}(z_1,z_2)\Big\}+o_P(1),
\tag{S2.62}
$$

where

$$
\begin{aligned}
\mathcal{Y}_{11}(z_1,z_2)&=\frac{\nu_4-3\nu_{12}}{n^2}b_p(z_1)b_p(z_2)\sum_{j=1}^{n}\sum_{i=1}^{p}\mathbb{E}_j[\mathbf{D}_j^{-1}(z_1)]_{ii}\mathbb{E}_j[\mathbf{D}_j^{-1}(z_2)]_{ii},\\
\mathcal{Y}_{12}(z_1,z_2)&=\frac{2\nu_{12}}{n^2}b_p(z_1)b_p(z_2)\sum_{j=1}^{n}\mathrm{tr}\Big\{\mathbb{E}_j\mathbf{D}_j^{-1}(z_1)\mathbb{E}_j\mathbf{D}_j^{-1}(z_2)\Big\},\\
\mathcal{Y}_{13}(z_1,z_2)&=\frac{\nu_{12}-\nu_2^2}{n^2}b_p(z_1)b_p(z_2)\sum_{j=1}^{n}\mathrm{tr}\Big\{\mathbb{E}_j\mathbf{D}_j^{-1}(z_1)\Big\}\mathrm{tr}\Big\{\mathbb{E}_j\mathbf{D}_j^{-1}(z_2)\Big\},
\end{aligned}
$$

where $\nu_2 = \mathbb{E}\Big(\frac{w_{11}}{\overline{w}_1} - 1\Big)^2$, $\nu_4 = \mathbb{E}\Big(\frac{w_{11}}{\overline{w}_1} - 1\Big)^4$, $\nu_{12} = \mathbb{E}\Big(\frac{w_{11}}{\overline{w}_1} - 1\Big)^2\Big(\frac{w_{12}}{\overline{w}_1} - 1\Big)^2$, $\overline{w}_1 = \sum_{\ell=1}^p w_{1\ell}/p$. We claim that the following statements hold true as $n, p \to \infty$ (to be proven later):

$$\frac{\partial^2 \mathcal{Y}_{11}(z_1, z_2)}{\partial z_1 \partial z_2} \xrightarrow{P} \frac{c\alpha_1 \underline{m}'(z_1)\underline{m}'(z_2)}{\{1+\lambda \underline{m}(z_1)\}^2\{1+\lambda \underline{m}(z_2)\}^2}, \tag{S2.63}$$

$$\frac{\partial^2 \mathcal{Y}_{12}(z_1, z_2)}{\partial z_1 \partial z_2} \xrightarrow{P} \frac{2\underline{m}'(z_1)\underline{m}'(z_2)}{\{\underline{m}(z_1) - \underline{m}(z_2)\}^2} - \frac{2}{(z_1 - z_2)^2}, \tag{S2.64}$$

$$\frac{\partial^2 \mathcal{Y}_{13}(z_1, z_2)}{\partial z_1 \partial z_2} \xrightarrow{P} \frac{c\alpha_2 \underline{m}'(z_1)\underline{m}'(z_2)}{\{1+\lambda \underline{m}(z_1)\}^2\{1+\lambda \underline{m}(z_2)\}^2}. \tag{S2.65}$$

By (S2.61) – (S2.65), we obtain the limit of $\frac{\partial^2}{\partial z_1 \partial z_2}\mathcal{Y}_1(z_1, z_2)$.

Now, we provide the proofs of (S2.63) – (S2.65) as follows:

Proof of (S2.63): It is enough to find the limit of

$$\frac{1}{n^2}\sum_{j=1}^n \sum_{i=1}^p \mathbb{E}_j\Big[\mathbf{D}_j^{-1}(z_1)\Big]_{ii}\mathbb{E}_j\Big[\mathbf{D}_j^{-1}(z_2)\Big]_{ii}.$$

By similar calculation of Gao et al. [2017], we get the following lemma and its proof is postponed to Section S2.7.

Lemma S2.1. *Under Assumptions 2.1 and 2.2, for any $1 \le j \le n$, we have as $p \to \infty$*

$$\frac{1}{p}\sum_{i=1}^p \mathbb{E}_j\Big[\mathbf{D}_j^{-1}(z_1)\Big]_{ii}\mathbb{E}_j\Big[\mathbf{D}_j^{-1}(z_2)\Big]_{ii} \xrightarrow{P} m(z_1)m(z_2).$$

By $\mathbb{E}|\frac{\nu_2}{n}\mathrm{tr}\mathbf{D}^{-1}(z) - \frac{\nu_2}{n}\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}(z)|^q \le K_q n^{-q/2} v_0^{-q}$ (K_q is a positive constant), the formula (2.2) of Silverstein [1995], $\underline{m}_p(z) = -\frac{1}{zn}\sum_{j=1}^n \beta_j(z)$, and Lemma 5.4, we have

$$|b_p(z) - \mathbb{E}\beta_1(z)| \le \frac{K}{\sqrt{n}}, \ \mathbb{E}\beta_1(z) = -z\mathbb{E}\underline{m}_p(z), \ |b_p(z) + z\underline{m}_p^0(z)| \le \frac{K}{\sqrt{n}}. \tag{S2.66}$$

Thus, by (S2.66), Lemmas 5.3 and S2.1, we have as $n, p \to \infty$

$$\mathcal{Y}_{11}(z_1, z_2) \xrightarrow{P} c\alpha_1 z_1 z_2 m(z_1)m(z_2)\underline{m}(z_1)\underline{m}(z_2) = \frac{c\alpha_1 \underline{m}(z_1)\underline{m}(z_2)}{\{1+\lambda \underline{m}(z_1)\}\{1+\lambda \underline{m}(z_2)\}},$$

where the equality above follows from $m(z) = -z^{-1}\{1+\lambda \underline{m}(z)\}^{-1}$. Thus, the in probability (i.p.) limit of $\frac{\partial^2 \mathcal{Y}_{11}(z_1,z_2)}{\partial z_1 \partial z_2}$ is in (S2.63).

Proof of (S2.64): By similar calculation of Bai and Silverstein [2004], we get the following lemma and its proof is postponed to Section S2.8.

Lemma S2.2. *Under Assumptions 2.1 and 2.2, for any $1 \le j \le n$, we have*

$$\mathrm{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_j^{-1}(z_1)\Big\}\mathbf{D}_j^{-1}(z_2)\Big]\Bigg[1 - \frac{\frac{j-1}{n}c_n \nu_2^2 \underline{m}_p^0(z_1)\underline{m}_p^0(z_2)}{\{1+\frac{n-1}{n}\nu_2 \underline{m}_p^0(z_1)\}\{1+\frac{n-1}{n}\nu_2 \underline{m}_p^0(z_2)\}}\Bigg]$$
$$= \frac{nc_n}{z_1 z_2}\frac{1}{\{1+\frac{n-1}{n}\nu_2 \underline{m}_p^0(z_1)\}\{1+\frac{n-1}{n}\nu_2 \underline{m}_p^0(z_2)\}} + O_P(n^{1/2}).$$

By using (S2.66) and this lemma, $\mathcal{Y}_{12}(z_1,z_2)$ can be written as

$$\mathcal{Y}_{12}(z_1,z_2)=\frac{a_p(z_1,z_2)\nu_{12}}{n\nu_2^2}\sum_{j=1}^{n}\frac{2}{1-\frac{j-1}{n}a_p(z_1,z_2)}+O_P(n^{-1/2}),$$

where $a_p(z_1,z_2)=\frac{\nu_2^2 c_n \underline{m}_p^0(z_1)\underline{m}_p^0(z_2)}{\{1+\frac{n-1}{n}\nu_2\underline{m}_p^0(z_1)\}\{1+\frac{n-1}{n}\nu_2\underline{m}_p^0(z_2)\}}$. By Lemma 5.3, the limit of $a_p(z_1,z_2)$ is $a(z_1,z_2)=\frac{c\lambda^2\underline{m}(z_1)\underline{m}(z_2)}{\{1+\lambda\underline{m}(z_1)\}\{1+\lambda\underline{m}(z_2)\}}$, and thus the i.p. limit of $\frac{\partial^2}{\partial z_1\partial z_2}\mathcal{Y}_{12}(z_1,z_2)$ is (S2.64).

Proof of (S2.65): We have $\mathbb{E}\left|\frac{1}{p}\mathrm{tr}\mathbb{E}_j\mathbf{D}_j^{-1}(z_1)\frac{1}{p}\mathrm{tr}\mathbb{E}_j\mathbf{D}_j^{-1}(z_2)-m_p^0(z_1)m_p^0(z_2)\right|=o(1)$. By Lemma 5.3, we get $\lim_{p\to\infty}p(\nu_{12}-\nu_2^2)=\alpha_2$. This, together with (S2.66), implies that as $n,p\to\infty$

$$\mathcal{Y}_{13}(z_1,z_2)\xrightarrow{P}\frac{c\alpha_2\underline{m}(z_1)\underline{m}(z_2)}{\{1+\lambda\underline{m}(z_1)\}\{1+\lambda\underline{m}(z_2)\}}.$$

Thus, the i.p. limit of $\frac{\partial^2}{\partial z_1\partial z_2}\mathcal{Y}_{13}(z_1,z_2)$ is (S2.65).

Limit of $\partial^2\mathcal{Y}_2(z_1,z_2)/(\partial z_1\partial z_2)$:

For any $p\times p$ matrix $\mathbf{A}$, we have

$$\left|\mathrm{tr}\left\{\mathbf{D}^{-1}(z)-\mathbf{D}_j^{-1}(z)\right\}\mathbf{A}\right|\le\frac{\|\mathbf{A}\|}{\Im(z)}. \tag{S2.67}$$

By Lemma S1.2 and (S2.67), we have

$$\mathbb{E}\left|\frac{\nu_2}{n}\mathrm{tr}\mathbf{D}^{-1}(z)-\frac{\nu_2}{n}\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}(z)\right|^q\le\frac{K_q}{n^{q/2}v_0^q}, \tag{S2.68}$$

where $K_q>0$ is a constant depending on q, which implies that

$$\mathbb{E}\left|\overline{\beta}_j(z)-b_p(z)\right|^2\le\frac{K|z|^4}{nv_0^6}. \tag{S2.69}$$

The above inequalities will be used in the following proof.

Denoting $\mathcal{Y}_{2,j}(z_1,z_2):=\mathbb{E}_{j-1}\left\{\overline{\beta}_j(z_1)\varepsilon_j(z_1)\right\}\mathbb{E}_{j-1}\left\{\overline{\beta}_j(z_2)\varepsilon_j(z_2)\right\}$. To prove the second part of the lemma, it is enough to show that, for each $1\le j\le n$,

$$\mathcal{Y}_{2,j}(z_1,z_2)-b_p(z_1)b_p(z_2)\left\{\mathbb{E}_{j-1}\varepsilon_j(z_1)\right\}\left\{\mathbb{E}_{j-1}\varepsilon_j(z_2)\right\}=o_P(n^{-1}), \tag{S2.70}$$

and

$$b_p(z_1)b_p(z_2)\{\mathbb{E}_{j-1}\varepsilon_j(z_1)\}\{\mathbb{E}_{j-1}\varepsilon_j(z_2)\}=o_P(n^{-1}). \tag{S2.71}$$

Proof of (S2.70): We write

$$\begin{aligned}\text{LHS of (S2.70)}&=\mathbb{E}_{j-1}\left[\left\{\overline{\beta}_j(z_1)-b_p(z_1)\right\}\varepsilon_j(z_1)\right]\mathbb{E}_{j-1}\left\{\overline{\beta}_j(z_2)\varepsilon_j(z_2)\right\}\\&\quad+\mathbb{E}_{j-1}\left\{b_p(z_1)\varepsilon_j(z_1)\right\}\mathbb{E}_{j-1}\left[\left\{\overline{\beta}_j(z_2)-b_p(z_2)\right\}\varepsilon_j(z_2)\right]\\&=:\Delta\mathcal{Y}_{2,j}^{(1)}(z_1,z_2)+\Delta\mathcal{Y}_{2,j}^{(2)}(z_1,z_2).\end{aligned}$$

Note that, for any $q_1, q_2 \geq 1$ with $1/q_1 + 1/q_2 = 1$, we have

$$\begin{aligned}\mathbb{E}\left|\Delta\mathcal{Y}_{2,j}^{(1)}(z_1,z_2)\right| &\leq \left(\mathbb{E}\left|\mathbb{E}_{j-1}\{\overline{\beta}_j(z_1)-b_p(z_1)\}\varepsilon_j(z_1)\right|^{q_1}\right)^{1/q_1}\left(\mathbb{E}\left|\mathbb{E}_{j-1}\overline{\beta}_j(z_2)\varepsilon_j(z_2)\right|^{q_2}\right)^{1/q_2}\\ &\leq \left(\mathbb{E}\left|\{\overline{\beta}_j(z_1)-b_p(z_1)\}\varepsilon_j(z_1)\right|^{q_1}\right)^{1/q_1}\left(\mathbb{E}\left|\overline{\beta}_j(z_2)\varepsilon_j(z_2)\right|^{q_2}\right)^{1/q_2}\end{aligned}$$

From Lemma 5.4 and (S2.69), for any $1 < q_1 < 2$ and $q_2 > 2$, we have

$$\begin{aligned}\mathbb{E}\left|\left\{\overline{\beta}_j(z_1)-b_p(z_1)\right\}\varepsilon_j(z_1)\right|^{q_1} &\leq \left(\mathbb{E}\left|\overline{\beta}_j(z_1)-b_p(z_1)\right|^{q_1\frac{2}{q_1}}\right)^{q_1/2}\left(\mathbb{E}\left|\varepsilon_j(z_1)\right|^{q_1\frac{2}{2-q_1}}\right)^{\frac{2-q_1}{2}}\\ &\leq Kn^{-\frac{q_1}{2}}n^{-\frac{2-q_1}{2}}\delta_n^{4q_1-4}\\ &= Kn^{-1}\delta_n^{4q_1-4},\end{aligned}$$

and

$$\mathbb{E}\left|\overline{\beta}_j(z_2)\varepsilon_j(z_2)\right|^{q_2} \leq K\frac{|z|^{q_2}}{v_0^{q_2}}n^{-1}\delta_n^{2q_2-4}.$$

Thus,

$$\mathbb{E}\left|\Delta\mathcal{Y}_{2,j}^{(1)}(z_1,z_2)\right| \leq Kn^{-1}\delta_n^{\frac{4q_1-4}{q_1}+\frac{2q_2-4}{q_2}} = o(n^{-1}).$$

Similarly, we can obtain $\mathbb{E}\left|\Delta\mathcal{Y}_{2,j}^{(2)}(z_1,z_2)\right| = o(n^{-1})$. Thus, we complete the proof of (S2.70).
Proof of (S2.71): By (S2.43), we get

$$\begin{aligned}\mathbb{E}_{j-1}\varepsilon_j(z) &= \mathbb{E}_{j-1}\left\{\mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_j^{-1}(z)\right\}\\ &= \mathrm{tr}\left\{\mathbb{E}_{j-1}\mathbf{D}_j^{-1}(z)\mathbb{E}\mathbf{r}_j\mathbf{r}_j'\right\} - \frac{\nu_2}{n}\mathbb{E}_{j-1}\mathrm{tr}\mathbf{D}_j^{-1}(z)\\ &= \frac{p\nu_2}{n(p-1)}\mathrm{tr}\left\{\mathbb{E}_{j-1}\mathbf{D}_j^{-1}(z)\left(\mathbf{I}_p - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p'\right)\right\} - \frac{\nu_2}{n}\mathbb{E}_{j-1}\mathrm{tr}\mathbf{D}_j^{-1}(z)\\ &= \frac{\nu_2}{n(p-1)}\mathbb{E}_{j-1}\left\{\mathrm{tr}\mathbf{D}_j^{-1}(z) - \mathbf{1}_p'\mathbf{D}_j^{-1}(z)\mathbf{1}_p\right\}.\end{aligned}$$

By Lemma 2.3 in Bai and Silverstein [2004] and our Lemma S1.5, we have as $p \to \infty$

$$\frac{1}{p}\mathrm{tr}\mathbf{D}_j^{-1}(z) \xrightarrow{P} m(z), \qquad \frac{1}{p}\mathbf{1}_p'\mathbf{D}_j^{-1}(z)\mathbf{1}_p \xrightarrow{P} -\frac{1}{z}.$$

By (S2.68) and the identity $\underline{m}_p(z) = -\frac{1}{zn}\sum_{j=1}^n \beta_j(z)$, we have

$$|b_p(z) + \mathbb{E}\underline{m}_p(z)| \leq Kn^{-1/2}.$$

This, together with Lemma 5.4, yields that

$$|b_p(z) + z\underline{m}_p^0(z)| \leq Kn^{-1/2}. \tag{S2.72}$$

Equation (S2.71) follows from the above estimates.

S2.7 Proof of Lemma S2.1

By Lemma S1.2, the inequality $|\beta_{ij}(z)| \le \frac{|z|}{v_0}$, and Lemma 5.4, we get

$$
\begin{aligned}
&\mathbb{E}\Big|(\mathbf{1}_p^i)'\Big\{\mathbf{D}_1^{-1}(z_1) - \mathbb{E}\mathbf{D}_1^{-1}(z_1)\Big\}\mathbf{1}_p^i\Big|^2 \\
&= \mathbb{E}\Big|\sum_{j=1}^n (\mathbb{E}_j - \mathbb{E}_{j-1})(\mathbf{1}_p^i)'\Big\{\mathbf{D}_1^{-1}(z_1) - \mathbf{D}_{1j}^{-1}(z_1)\Big\}\mathbf{1}_p^i\Big|^2 \\
&\le K\sum_{j=1}^n \mathbb{E}\Big|(\mathbb{E}_j - \mathbb{E}_{j-1})\beta_{1j}(z_1)\mathbf{r}_j'\mathbf{D}_{1j}^{-1}\mathbf{1}_p^i(\mathbf{1}_p^i)'\mathbf{D}_{1j}^{-1}\mathbf{r}_j\Big|^2 \\
&\le K\sum_{j=1}^n \mathbb{E}\Big|\beta_{1j}(z_1)\mathbf{r}_j'\mathbf{D}_{1j}^{-1}\mathbf{1}_p^i(\mathbf{1}_p^i)'\mathbf{D}_{1j}^{-1}\mathbf{r}_j\Big|^2 \\
&\le Kn^{-1},
\end{aligned} \tag{S2.73}
$$

where $\mathbf{1}_p^i$ is the p-dimensional vector with all components being 0 except for the i-th component being 1. Hence, we have

$$
\begin{aligned}
&\mathbb{E}\Big|\frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^p \mathbb{E}_j\Big[\mathbf{D}_j^{-1}(z_1) - \mathbb{E}\mathbf{D}_j^{-1}(z_1)\Big]_{ii}\mathbb{E}_j\Big[\mathbf{D}_j^{-1}(z_2)\Big]_{ii}\Big| \\
\le\ &\frac{nK}{n^2 v_0}\sum_{i=1}^p \mathbb{E}\Big|(\mathbf{1}_p^i)'\{\mathbf{D}_1^{-1}(z_1) - \mathbb{E}\mathbf{D}_1^{-1}(z_1)\}\mathbf{1}_p^i\Big| \le Kn^{-1/2},
\end{aligned}
$$

and thus

$$
\frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^p \mathbb{E}_j\Big[\mathbf{D}_j^{-1}(z_1) - \mathbb{E}\mathbf{D}_j^{-1}(z_1)\Big]_{ii}\mathbb{E}_j\Big[\mathbf{D}_j^{-1}(z_2)\Big]_{ii} = O_P(n^{-1/2}).
$$

Similarly, we have

$$
\frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^p \mathbb{E}_j\Big[\mathbf{D}_j^{-1}(z_2) - \mathbb{E}\mathbf{D}_j^{-1}(z_2)\Big]_{ii}\mathbb{E}\Big[\mathbf{D}_j^{-1}(z_1)\Big]_{ii} = O_P(n^{-1/2}).
$$

With the above two inequalities, it remains to find the limit of

$$
\frac{1}{p}\sum_{i=1}^p \mathbb{E}\Big[\mathbf{D}_j^{-1}(z_1)\Big]_{ii}\mathbb{E}\Big[\mathbf{D}_j^{-1}(z_2)\Big]_{ii}. \tag{S2.74}
$$

It is easy to see that the sum of expectations in (S2.74) is exactly the same for any j. Moreover, we have

$$
\frac{1}{p}\sum_{i=1}^p \mathbb{E}\Big[\mathbf{D}_j^{-1}(z_1)\Big]_{ii}\mathbb{E}\Big[\mathbf{D}_j^{-1}(z_2)\Big]_{ii} \xrightarrow{P} m(z_1)m(z_2), \quad \text{as } p\to\infty.
$$

This completes the proof of Lemma S2.1.

S2.8 Proof of Lemma S2.2

Let

$$\mathbf{D}_{ij}(z) = \mathbf{D}(z) - \mathbf{r}_i\mathbf{r}_i' - \mathbf{r}_j\mathbf{r}_j', \quad b_1(z) = \frac{1}{1 + n^{-1}\nu_2\mathbb{E}\mathrm{tr}\mathbf{D}_{12}^{-1}(z)},$$
$$\beta_{ij}(z) = \frac{1}{1 + \mathbf{r}_i'\mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i}.$$

We have the equality

$$\mathbf{D}_j(z_1) + z_1\mathbf{I}_p - \frac{n-1}{n}\nu_2 b_1(z_1)\mathbf{I}_p = \sum_{i\neq j}^{n} \mathbf{r}_i\mathbf{r}_i' - \frac{n-1}{n}\nu_2 b_1(z_1)\mathbf{I}_p.$$

Multiplying by $\mathbf{Q}_p(z_1) := \{z_1\mathbf{I}_p - \frac{n-1}{n}\nu_2 b_1(z_1)\mathbf{I}_p\}^{-1}$ on the LHS and $\mathbf{D}_j^{-1}(z_1)$ on the RHS, and using $\mathbf{r}_i'\mathbf{D}_j^{-1}(z_1) = \beta_{ij}(z_1)\mathbf{r}_i'\mathbf{D}_{ij}^{-1}(z_1)$, we get

$$\mathbf{D}_j^{-1}(z_1) = -\mathbf{Q}_p(z_1) + b_1(z_1)\mathbf{A}(z_1) + \mathbf{B}(z_1) + \mathbf{C}(z_1), \tag{S2.75}$$

where

$$\mathbf{A}(z_1) = \sum_{i\neq j}^{n} \mathbf{Q}_p(z_1)\Big(\mathbf{r}_i\mathbf{r}_i' - \frac{\nu_2}{n}\mathbf{I}_p\Big)\mathbf{D}_{ij}^{-1}(z_1),$$
$$\mathbf{B}(z_1) = \sum_{i\neq j}^{n} \Big\{\beta_{ij}(z_1) - b_1(z_1)\Big\}\mathbf{Q}_p(z_1)\mathbf{r}_i\mathbf{r}_i'\mathbf{D}_{ij}^{-1}(z_1),$$
$$\mathbf{C}(z_1) = \frac{\nu_2}{n} b_1(z_1)\mathbf{Q}_p(z_1)\sum_{i\neq j}^{n}\Big\{\mathbf{D}_{ij}^{-1}(z_1) - \mathbf{D}_j^{-1}(z_1)\Big\}.$$

For any $t \in \mathbb{R}$, $|1 - tb_1(z)/z|^{-1} \leq \frac{|z/b_1(z)|}{\mathrm{Im}\{z/b_1(z)\}} \leq \frac{|z|\{1+p/(nv_0)\}}{v_0}$. Thus,

$$\|\mathbf{Q}_p(z_1)\| \leq \frac{1 + p/(nv_0)}{v_0}. \tag{S2.76}$$

For any random matrix $\mathbf{M}$, denote its nonrandom bound on the spectrum norm of $\mathbf{M}$ by $|||\mathbf{M}|||$. Since the same argument in (S2.68) holds for $\mathbf{D}_{12}^{-1}$, and from (S2.76), Lemma 5.4, we get

$$\mathbb{E}\Big|\mathrm{tr}\mathbf{B}(z_1)\mathbf{M}\Big| \leq K|||\mathbf{M}|||\frac{|z_1|^2\{1 + p/(nv_0)\}}{v_0^5}n^{1/2}. \tag{S2.77}$$

From (S2.67), we have

$$\Big|\mathrm{tr}\mathbf{C}(z_1)\mathbf{M}\Big| \leq |||\mathbf{M}|||\frac{|z_1|\{1 + p/(nv_0)\}}{v_0^3}. \tag{S2.78}$$

From (S2.76) and Lemma 5.4, we get, for $\mathbf{M}$ nonrandom,

$$\mathbb{E}\Big|\mathrm{tr}\mathbf{A}(z_1)\mathbf{M}\Big| \leq K\|\mathbf{M}\|\frac{1 + p/(nv_0)}{v_0^2}n^{1/2}. \tag{S2.79}$$

Note that

$$
\begin{aligned}
&\operatorname{tr}\Big[\mathbb{E}_j\Big\{\mathbf{A}(z_1)\Big\}\mathbf{D}_j^{-1}(z_2)\Big] \\
&= \operatorname{tr}\sum_{i<j}^{n}\mathbf{Q}_p(z_1)\Big(\mathbf{r}_i\mathbf{r}_i' - \frac{\nu_2}{n}\mathbf{I}_p\Big)\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2) \\
&\quad + \operatorname{tr}\sum_{i<j}^{n}\mathbf{Q}_p(z_1)\Big(\mathbf{r}_i\mathbf{r}_i' - \frac{\nu_2}{n}\mathbf{I}_p\Big)\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\{\mathbf{D}_j^{-1}(z_2) - \mathbf{D}_{ij}^{-1}(z_2)\} \\
&\quad + \operatorname{tr}\mathbb{E}_j\Bigg\{\sum_{i>j}^{n}\mathbf{Q}_p(z_1)\Big(\mathbf{r}_i\mathbf{r}_i' - \frac{\nu_2}{n}\mathbf{I}_p\Big)\mathbf{D}_{ij}^{-1}(z_1)\Bigg\}\mathbf{D}_j^{-1}(z_2),
\end{aligned}
$$

thus, by using $\mathbf{D}_j^{-1}(z) - \mathbf{D}_{ij}^{-1}(z) = -\beta_{ij}(z)\mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i\mathbf{r}_i'\mathbf{D}_{ij}^{-1}(z)$, we can write

$$
\operatorname{tr}\Big[\mathbb{E}_j\Big\{\mathbf{A}(z_1)\Big\}\mathbf{D}_j^{-1}(z_2)\Big] = A_1(z_1,z_2) + A_2(z_1,z_2) + A_3(z_1,z_2) + R(z_1,z_2), \tag{S2.80}
$$

where

$$
A_1(z_1,z_2) = -\sum_{i<j}^{n}\beta_{ij}(z_2)\mathbf{r}_i'\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i'\mathbf{D}_{ij}^{-1}(z_2)\mathbf{Q}_p(z_1)\mathbf{r}_i, \tag{S2.81}
$$

$$
\begin{aligned}
A_2(z_1,z_2) &= -\frac{\nu_2}{n}\operatorname{tr}\sum_{i<j}^{n}\mathbf{Q}_p(z_1)\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\Big\{\mathbf{D}_j^{-1}(z_2) - \mathbf{D}_{ij}^{-1}(z_2)\Big\}, \\
A_3(z_1,z_2) &= \operatorname{tr}\sum_{i<j}^{n}\mathbf{Q}_p(z_1)\Big(\mathbf{r}_i\mathbf{r}_i' - \frac{\nu_2}{n}\mathbf{I}_p\Big)\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2),
\end{aligned}
$$

$$
R(z_1,z_2) = \operatorname{tr}\mathbb{E}_j\Bigg\{\sum_{i>j}^{n}\mathbf{Q}_p(z_1)\Big(\mathbf{r}_i\mathbf{r}_i' - \frac{\nu_2}{n}\mathbf{I}_p\Big)\mathbf{D}_{ij}^{-1}(z_1)\Bigg\}\mathbf{D}_j^{-1}(z_2). \tag{S2.82}
$$

It is easy to see that $R(z_1,z_2) = O_P(1)$. We get from (S2.67) and (S2.76) that $|A_2(z_1,z_2)| \le \frac{1+p/(nv_0)}{v_0^2}$. Similar to (S2.79), we have $\mathbb{E}|A_3(z_1,z_2)| \le \frac{1+p/(nv_0)}{v_0^3}n^{1/2}$. It remains to derive the limit of $A_1(z_1,z_2)$. By using Lemma 5.4 and similar argument in (S2.92), we have, for $i<j$,

$$
\begin{aligned}
&\mathbb{E}\Big|\beta_{ij}(z_2)\mathbf{r}_i'\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i'\mathbf{D}_{ij}^{-1}(z_2)\mathbf{Q}_p(z_1)\mathbf{r}_i \\
&\qquad - \frac{\nu_2^2}{n^2}b_1(z_2)\operatorname{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2)\Big]\operatorname{tr}\Big\{\mathbf{D}_{ij}^{-1}(z_2)\mathbf{Q}_p(z_1)\Big\}\Big| \\
&\le \mathbb{E}\Big|\beta_{ij}(z_2)\mathbf{r}_i'\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i'\mathbf{D}_{ij}^{-1}(z_2)\mathbf{Q}_p(z_1)\mathbf{r}_i \\
&\qquad - \frac{\nu_2^2}{n^2}\beta_{ij}(z_2)\operatorname{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2)\Big]\operatorname{tr}\Big\{\mathbf{D}_{ij}^{-1}(z_2)\mathbf{Q}_p(z_1)\Big\}\Big| \\
&\quad + \mathbb{E}\Big|\frac{\nu_2^2}{n^2}\Big\{\beta_{ij}(z_2) - b_1(z_2)\Big\}\operatorname{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2)\Big]\operatorname{tr}\Big\{\mathbf{D}_{ij}^{-1}(z_2)\mathbf{Q}_p(z_1)\Big\}\Big| \\
&\le \mathbb{E}\Big|\beta_{ij}(z_2)\Big(\mathbf{r}_i'\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i - \frac{\nu_2}{n}\operatorname{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2)\Big]\Big)\mathbf{r}_i'\mathbf{D}_{ij}^{-1}(z_2)\mathbf{Q}_p(z_1)\mathbf{r}_i
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{\nu_2}{n}\beta_{ij}(z_2)\mathrm{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big)\Big\}\mathbf{D}_{ij}^{-1}(z_2)\Big]\Big[\mathbf{r}_i'\mathbf{D}_{ij}^{-1}(z_2)\mathbf{Q}_p(z_1)\mathbf{r}_i-\frac{\nu_2}{n}\mathrm{tr}\Big\{\mathbf{D}_{ij}^{-1}(z_2)\mathbf{Q}_p(z_1)\Big\}\Big]\Big|\\
&+\mathbb{E}\Big|\frac{\nu_2^2}{n^2}\Big\{\beta_{ij}(z_2)-b_1(z_2)\Big\}\mathrm{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2)\Big]\mathrm{tr}\Big\{\mathbf{D}_{ij}^{-1}(z_2)\mathbf{Q}_p(z_1)\Big\}\Big|\\
&\le Kn^{-1/2}. \qquad \text{(S2.83)}
\end{aligned}
$$

By (S2.67), we have

$$
\begin{aligned}
\mathbb{E}\Big|&\mathrm{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{ij}^{-1}(z_1)\Big\}\mathbf{D}_{ij}^{-1}(z_2)\Big]\mathrm{tr}\Big\{\mathbf{D}_{ij}^{-1}(z_2)\mathbf{Q}_p(z_1)\Big\}\\
&-\mathrm{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{j}^{-1}(z_1)\Big\}\mathbf{D}_{j}^{-1}(z_2)\Big]\mathrm{tr}\Big\{\mathbf{D}_{j}^{-1}(z_2)\mathbf{Q}_p(z_1)\Big\}\Big|\le Kn. \qquad \text{(S2.84)}
\end{aligned}
$$

It follows from (S2.83) and (S2.84) that

$$
\mathbb{E}\Big|A_1(z_1,z_2)+\frac{j-1}{n^2}\nu_2^2 b_1(z_2)\mathrm{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{j}^{-1}(z_1)\Big\}\mathbf{D}_{j}^{-1}(z_2)\Big]\mathrm{tr}\Big\{\mathbf{D}_{j}^{-1}(z_2)\mathbf{Q}_p(z_1)\Big\}\Big|\le Kn^{1/2}. \qquad \text{(S2.85)}
$$

By using (S2.75) – (S2.80), we have

$$
\begin{aligned}
\mathrm{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{j}^{-1}(z_1)\Big\}\mathbf{D}_{j}^{-1}(z_2)\Big]&=\mathrm{tr}\Big[\mathbb{E}_j\Big\{-\mathbf{Q}_p(z_1)+b_1(z)\mathbf{A}(z_1)\Big\}\mathbf{D}_{j}^{-1}(z_2)\Big]+O_P(n^{1/2})\\
&=-\mathrm{tr}\Big\{\mathbf{Q}_p(z_1)\mathbf{D}_{j}^{-1}(z_2)\Big\}+b_1(z_1)\mathrm{tr}\mathbb{E}_j\Big\{\mathbf{A}(z_1)\Big\}\mathbf{D}_{j}^{-1}(z_2)+O_P(n^{1/2})\\
&=-\mathrm{tr}\Big\{\mathbf{Q}_p(z_1)\mathbf{D}_{j}^{-1}(z_2)\Big\}+b_1(z_1)A_1(z_1,z_2)+O_P(n^{1/2}).
\end{aligned}
$$

This, together with (S2.85), implies that

$$
\begin{aligned}
&\mathrm{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{j}^{-1}(z_1)\Big\}\mathbf{D}_{j}^{-1}(z_2)\Big]\Big[1+\frac{j-1}{n^2}\nu_2^2 b_1(z_1)b_1(z_2)\mathrm{tr}\Big\{\mathbf{D}_{j}^{-1}(z_2)\mathbf{Q}_p(z_1)\Big\}\Big]\\
&=-\mathrm{tr}\Big\{\mathbf{Q}_p(z_1)\mathbf{D}_{j}^{-1}(z_2)\Big\}+O_P(n^{1/2}).
\end{aligned}
$$

By using (S2.75) – (S2.79) and (S2.67), we have

$$
\begin{aligned}
&\mathrm{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{j}^{-1}(z_1)\Big\}\mathbf{D}_{j}^{-1}(z_2)\Big]\Big[1-\frac{j-1}{n^2}\nu_2^2 b_1(z_1)b_1(z_2)\mathrm{tr}\Big\{\mathbf{Q}_p(z_2)\mathbf{Q}_p(z_1)\Big\}\Big]\\
&=\mathrm{tr}\Big\{\mathbf{Q}_p(z_2)\mathbf{Q}_p(z_1)\Big\}+O_P(n^{1/2}). \qquad \text{(S2.86)}
\end{aligned}
$$

From (S2.67), we have $|b_1(z)-b_p(z)|=O(n^{-1})$. This, together with (S2.72) and (S2.86), implies that

$$
\begin{aligned}
&\mathrm{tr}\Big[\mathbb{E}_j\Big\{\mathbf{D}_{j}^{-1}(z_1)\Big\}\mathbf{D}_{j}^{-1}(z_2)\Big]\Big[1-\frac{\frac{j-1}{n}c_n\nu_2^2\underline{m}_p^0(z_1)\underline{m}_p^0(z_2)}{\{1+\frac{n-1}{n}\nu_2\underline{m}_p^0(z_1)\}\{1+\frac{n-1}{n}\nu_2\underline{m}_p^0(z_2)\}}\Big]\\
&=\frac{nc_n}{z_1z_2}\frac{1}{\{1+\frac{n-1}{n}\nu_2\underline{m}_p^0(z_1)\}\{1+\frac{n-1}{n}\nu_2\underline{m}_p^0(z_2)\}}+O_P(n^{1/2}).
\end{aligned}
$$

This completes the proof of Lemma S2.2.

S2.9 Tightness of $M_p^{(1)}(z)$

The proof of the tightness of $M_p^{(1)}(z)$ is similar to that provided in Bai and Silverstein [2004]. It is sufficient to prove the moment condition (12.51) of Billingsley [1968], i.e.

$$\sup_{n;z_1,z_2\in\mathcal{C}_n} \frac{\mathbb{E}|M_p^{(1)}(z_1)-M_p^{(1)}(z_2)|^2}{|z_1-z_2|^2} \tag{S2.87}$$

is finite.

Before proceeding, we provide some results needed in the proof later. First, moments of $\|\mathbf{D}^{-1}(z)\|$, $\|\mathbf{D}_j^{-1}(z)\|$ and $\|\mathbf{D}_{ij}^{-1}(z)\|$ are bounded in p and $z\in\mathcal{C}_n$. It is easy to see that it is true for $z\in\mathcal{C}_u$ and for $z\in\mathcal{C}_\ell$ if $x_\ell<0$. For $z\in\mathcal{C}_r$ or, if $x_\ell>0$, $z\in\mathcal{C}_\ell$, we have from Proposition 2.4 that

$$\begin{aligned}\mathbb{E}\|\mathbf{D}_j^{-1}(z)\|^m &\leq K_1+v^{-m}P(\|\mathbf{B}_{(j)}\|\geq\eta_r \text{ or } \lambda_{\min}(\mathbf{B}_{(j)})\leq\eta_\ell)\\ &\leq K_1+K_2n^m\varepsilon^{-m}n^{-\ell}\leq K\end{aligned}$$

for large ℓ, where $\mathbf{B}_{(j)}=\mathbf{B}_p^0-\mathbf{r}_j\mathbf{r}_j'$, K_1,K_2,K are positive constants. Here η_r is any number between $\lambda(1+\sqrt{c})^2$ and x_r; if $x_\ell>0$, η_ℓ is any number between x_ℓ and $\lambda(1-\sqrt{c})^2$ and if $x_\ell<0$, η_ℓ can be any negative number. So for any positive integer m,

$$\max\left(\mathbb{E}\|\mathbf{D}^{-1}(z)\|^m,\mathbb{E}\|\mathbf{D}_j^{-1}(z)\|^m,\mathbb{E}\|\mathbf{D}_{ij}^{-1}(z)\|^m\right)\leq K. \tag{S2.88}$$

By the argument above, we can extend Lemma 5.4 and get

$$\left|\mathbb{E}\left(a(v)\prod_{\ell=1}^{q}\left(\mathbf{r}_1'\mathbf{B}_{(\ell)}(v)\mathbf{r}_1-n^{-1}\mathrm{tr}\mathbf{B}_{(\ell)}(v)\right)\right)\right|\leq Kn^{-1}\delta_n^{2q-4}, \tag{S2.89}$$

where $\mathbf{B}_\ell(v)$ is independent of $\mathbf{r}_1$ and

$$\max\left(|a(v)|,\|\mathbf{B}_{(\ell)}(v)\|\right)\leq K\left(1+n^sI_{\{\|\mathbf{B}_p^0\|\geq\eta_r \text{ or } \lambda_{\min}(\widetilde{\mathbf{B}})\leq\eta_\ell\}}\right),$$

for some positive s, with $\widetilde{\mathbf{B}}$ being $\mathbf{B}_{(j)}$ or $\mathbf{B}_p^0$. By (S2.89), we have

$$\mathbb{E}|\varepsilon_j(z)|^m\leq K_mn^{-1}\delta_n^{2m-4}. \tag{S2.90}$$

where K_m is a positive constant. Let $\gamma_j(z)=\mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{r}_j-n^{-1}\nu_2\mathbb{E}\mathrm{tr}\mathbf{D}_j^{-1}(z)$. By Lemma S1.2, (S2.89) and Hölder's inequality, with similar derivation on page 580 of Bai and Silverstein [2004], we have

$$\mathbb{E}|\gamma_j(z)-\varepsilon_j(z)|^m\leq\frac{K_m}{n^{m/2}}. \tag{S2.91}$$

It follows from (S2.90) and (S2.91) that

$$\mathbb{E}|\gamma_j(z)|^m\leq K_mn^{-1}\delta_n^{2m-4},\quad m\geq 2. \tag{S2.92}$$

Next, we prove that $b_p(z)$ is bounded. With (S2.89), we have for any $m\geq 1$,

$$\mathbb{E}|\beta_1(z)|^m\leq K_m. \tag{S2.93}$$

Since $b_p(z)=\beta_1(z)+\beta_1(z)b_p(z)\gamma_1(z)$, it is derived from (S2.92) and (S2.93) that $|b_p(z)|\leq K_1+K_2|b_p(z)|n^{-1/2}$. Hence, we have

$$|b_p(z)|\leq\frac{K_1}{1-K_2n^{-1/2}}\leq K. \tag{S2.94}$$

With (S2.89) – (S2.94) and the same approach on Page 581 – 583 of Bai and Silverstein [2004], we can obtain that (S2.87) is finite.

S2.10 Proof of Lemma 5.8

Let $\widetilde{\mathbf{Q}}_p(z) = \mathbf{I}_p + \lambda \mathbb{E}\underline{m}_p(z)\mathbf{I}_p$, then

$$nA_p = \frac{p}{1+\lambda\mathbb{E}\underline{m}_p(z)} + pz\mathbb{E}m_p(z) = \mathbb{E}\{\beta_1(z)\mathcal{P}_1(z)\} + \mathbb{E}\{\beta_1(z)\mathcal{P}_2(z)\}, \tag{S2.95}$$

where

$$\mathcal{P}_1(z) = n\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1 - \lambda\mathrm{tr}\Big\{\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbb{E}\mathbf{D}_1^{-1}(z)\Big\},$$
$$\mathcal{P}_2(z) = \lambda\mathrm{tr}\Big\{\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbb{E}\mathbf{D}_1^{-1}(z)\Big\} - \lambda\mathrm{tr}\Big\{\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbb{E}\mathbf{D}^{-1}(z)\Big\}.$$

Since $\beta_1 = b_p - b_p^2\gamma_1 + \beta_1 b_p^2\gamma_1^2$, where $\gamma_1 \equiv \gamma_1(z) := \mathbf{r}_1'\mathbf{D}_1^{-1}(z)\mathbf{r}_1 - n^{-1}\nu_2\mathbb{E}\mathrm{tr}\mathbf{D}_1^{-1}(z)$, we have

$$\mathbb{E}\{\beta_1(z)\mathcal{P}_1(z)\} = b_p(z)\mathbb{E}\mathcal{P}_1(z) - b_p^2(z)\mathbb{E}\{\gamma_1(z)\mathcal{P}_1(z)\} + b_p^2(z)\mathbb{E}\{\beta_1(z)\gamma_1^2(z)\mathcal{P}_1(z)\}. \tag{S2.96}$$

The estimates for $\mathbb{E}\mathcal{P}_1(z)$, $\mathbb{E}\{\gamma_1(z)\mathcal{P}_1(z)\}$, $\mathbb{E}\{\beta_1(z)\gamma_1^2(z)\mathcal{P}_1(z)\}$, and $\mathbb{E}\{\beta_1(z)\mathcal{P}_2(z)\}$ are provided in the following lemma, and its proof is postponed to Section S2.11.

Lemma S2.3. *Under Assumptions 2.1 and 2.2, we have*

$$\mathbb{E}\mathcal{P}_1(z) = \frac{n}{1+\lambda\mathbb{E}\underline{m}_p(z)}\Big\{\mathbb{E}\gamma_1(z) + \frac{\nu_2-\lambda}{n}\mathbb{E}\mathrm{tr}\mathbf{D}_1^{-1}(z)\Big\}, \tag{S2.97}$$

$$\begin{aligned}
&\mathbb{E}\{\gamma_1(z)\mathcal{P}_1(z)\}\\
&= n\mathbb{E}\bigg(\Big\{\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\mathbf{r}_1 - \frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_1^{-1}(z)\Big\}\Big[\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1 - \frac{\nu_2}{n}\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big]\bigg)\\
&\quad + \frac{\nu_2^2}{n(p-1)}\mathbb{E}\Big[\Big\{\mathrm{tr}\mathbf{D}_1^{-1}(z) - \mathbf{1}_p'\mathbf{D}_1^{-1}(z)\mathbf{1}_p\Big\}\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big]\\
&\quad - \frac{\lambda}{1+\lambda\mathbb{E}\underline{m}_p(z)}\mathbb{E}\mathrm{tr}\{\mathbf{D}_1^{-1}(z)\}\mathbb{E}\gamma_1(z) + o(1),
\end{aligned} \tag{S2.98}$$

$$\begin{aligned}
&\mathbb{E}\Big\{\beta_1(z)\gamma_1^2(z)\mathcal{P}_1(z)\Big\}\\
&= \mathbb{E}\Big\{n\beta_1(z)\gamma_1^2(z)\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1\Big\} - \mathbb{E}\Big[\beta_1(z)\gamma_1^2(z)\mathrm{tr}\Big\{\lambda\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\Big\}\Big]\\
&\quad + \mathrm{Cov}\Big(\beta_1(z)\gamma_1^2(z),\ \mathrm{tr}\Big\{\lambda\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\Big\}\Big)\\
&= O(\delta_n^2),
\end{aligned} \tag{S2.99}$$

$$\mathbb{E}\{\beta_1(z)\mathcal{P}_2(z)\} = \frac{p\nu_2\lambda b_p^2(z)}{n(p-1)}\mathbb{E}\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\Big\} + O(n^{-1/2}), \tag{S2.100}$$

where δ_n is as defined in the truncation condition (14).

From this lemma and (S2.95), (S2.96), we get

$$nA_p = J_1 + J_2 + J_3 + o(1), \tag{S2.101}$$

where

$$\begin{aligned}
J_1 &= \frac{nb_p(z)}{1+\lambda\mathbb{E}\underline{m}_p(z)}\left\{\mathbb{E}\gamma_1(z) + \frac{\nu_2-\lambda}{n}\mathbb{E}\mathrm{tr}\mathbf{D}_1^{-1}(z)\right\} + \frac{b_p^2(z)\lambda\mathrm{tr}\{\mathbb{E}\mathbf{D}_1^{-1}(z)\}\mathbb{E}\gamma_1(z)}{1+\lambda\mathbb{E}\underline{m}_p(z)} \\
&\quad - \frac{b_p^2(z)\nu_2^2}{n(p-1)}\mathbb{E}\left[\left\{\mathrm{tr}\mathbf{D}_1^{-1}(z) - \mathbf{1}_p'\mathbf{D}_1^{-1}(z)\mathbf{1}_p\right\}\mathrm{tr}\left\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\right\}\right], \\
J_2 &= -nb_p^2(z)\mathbb{E}\left(\left\{\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\mathbf{r}_1 - \frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_1^{-1}(z)\right\}\left[\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1 - \frac{\nu_2}{n}\mathrm{tr}\left\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\right\}\right]\right), \\
J_3 &= \frac{pb_p^2(z)\lambda\nu_2}{n(p-1)}\mathbb{E}\,\mathrm{tr}\left\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\}.
\end{aligned}$$

The limits of J_1, J_2 and J_3 are provided in the following lemma, whose proof is postponed to Section S2.12.

Lemma S2.4. *Under Assumptions 2.1 and 2.2, as $n, p \to \infty$,*

$$\begin{aligned}
J_1 &\to -\frac{\underline{m}(z)\{z(\lambda+h_1)m(z)+\lambda\}}{1+\lambda\underline{m}(z)}, \\
J_2 &\to -\frac{cz^2\underline{m}^2(z)\{(\alpha_1+\alpha_2)m^2(z)+2\lambda^2 m'(z)\}}{1+\lambda\underline{m}(z)}, \\
J_3 &\to \frac{c\lambda^2\underline{m}^2(z)}{\{1+\lambda\underline{m}(z)\}[\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)]}.
\end{aligned}$$

By this Lemma and (S2.101), we get the limit of nA_p.

S2.11 Proof of Lemma S2.3

Proof of (S2.97): this equation follows from the definition of $\gamma_1(z)$.
Proof of (S2.98): For $\mathbb{E}\{\gamma_1(z)\mathcal{P}_1(z)\}$, we have

$$\begin{aligned}
&\mathbb{E}\{\gamma_1(z)\mathcal{P}_1(z)\} \\
&= n\mathbb{E}\left(\left\{\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\mathbf{r}_1 - \frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_1^{-1}(z) + \frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_1^{-1}(z) - \frac{\nu_2}{n}\mathbb{E}\mathrm{tr}\mathbf{D}_1^{-1}(z)\right\}\right. \\
&\qquad \left.\times\left[\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1 - \frac{\nu_2}{n}\mathrm{tr}\left\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\right\} + \frac{\nu_2}{n}\mathrm{tr}\left\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\right\}\right]\right) \\
&\quad - \frac{\lambda}{1+\lambda\mathbb{E}\underline{m}_p(z)}\mathbb{E}\mathrm{tr}\{\mathbf{D}_1^{-1}(z)\}\mathbb{E}\gamma_1(z) \\
&= n\mathbb{E}\left(\left\{\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\mathbf{r}_1 - \frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_1^{-1}(z)\right\}\left[\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1 - \frac{\nu_2}{n}\mathrm{tr}\left\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\right\}\right]\right) \\
&\quad + \frac{\nu_2^2}{n(p-1)}\mathbb{E}\left[\left\{\mathrm{tr}\mathbf{D}_1^{-1}(z) - \mathbf{1}_p'\mathbf{D}_1^{-1}(z)\mathbf{1}_p\right\}\mathrm{tr}\left\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\right\}\right] \\
&\quad + \frac{p\nu_2^2}{n(p-1)}\mathrm{Cov}\left(\mathrm{tr}\mathbf{D}_1^{-1}(z), \mathrm{tr}\left\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\right\}\right)
\end{aligned}$$

$$
\begin{aligned}
&-\frac{\nu_2^2}{n(p-1)}\mathrm{Cov}\Big(\mathrm{tr}\mathbf{D}_1^{-1}(z),\ \mathbf{1}_p'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{1}_p\Big)\\
&-\frac{\lambda}{1+\lambda\mathbb{E}\underline{m}_p(z)}\mathbb{E}\mathrm{tr}\{\mathbf{D}_1^{-1}(z)\}\mathbb{E}\gamma_1(z)\\
=\ & n\mathbb{E}\Big(\Big\{\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\mathbf{r}_1-\frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_1^{-1}(z)\Big\}\Big[\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1-\frac{\nu_2}{n}\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big]\Big)\\
&+\frac{\nu_2^2}{n(p-1)}\mathbb{E}\Big[\Big\{\mathrm{tr}\mathbf{D}_1^{-1}(z)-\mathbf{1}_p'\mathbf{D}_1^{-1}(z)\mathbf{1}_p\Big\}\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big]\\
&-\frac{\lambda}{1+\lambda\mathbb{E}\underline{m}_p(z)}\mathbb{E}\mathrm{tr}\{\mathbf{D}_1^{-1}(z)\}\mathbb{E}\gamma_1(z)+O(n^{-1}),
\end{aligned}
$$

which is the second equation in Lemma S2.3. Below are some interpretations of the above equalities:

1. The second equality uses the following derivation: By (S2.43), we get

$$
\begin{aligned}
& n\mathbb{E}\Big[\Big\{\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\mathbf{r}_1-\frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_1^{-1}(z)\Big\}\cdot\frac{\nu_2}{n}\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big]\\
&+n\mathbb{E}\Big[\Big\{\frac{\nu_2}{n}\mathrm{tr}\mathbf{D}_1^{-1}(z)-\frac{\nu_2}{n}\mathbb{E}\mathrm{tr}\mathbf{D}_1^{-1}(z)\Big\}\cdot\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1\Big]\\
=\ & \nu_2\mathrm{tr}\Big(\mathbb{E}(\mathbf{r}_1\mathbf{r}_1')\mathbb{E}\Big[\mathbf{D}_1^{-1}(z)\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big]\Big)-\frac{\nu_2^2}{n}\mathbb{E}\Big[\mathrm{tr}\mathbf{D}_1^{-1}(z)\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big]\\
&+\nu_2\mathrm{tr}\Big[\mathbb{E}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathrm{tr}\mathbf{D}_1^{-1}(z)\Big\}\mathbb{E}(\mathbf{r}_1\mathbf{r}_1')\Big]-\nu_2\mathbb{E}\Big\{\mathrm{tr}\mathbf{D}_1^{-1}(z)\Big\}\mathrm{tr}\Big[\mathbb{E}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\mathbb{E}(\mathbf{r}_1\mathbf{r}_1')\Big]\\
=\ & \frac{\nu_2^2}{n(p-1)}\mathbb{E}\Big[\Big\{\mathrm{tr}\mathbf{D}_1^{-1}(z)-\mathbf{1}_p'\mathbf{D}_1^{-1}(z)\mathbf{1}_p\Big\}\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big]\\
&+\frac{p\nu_2^2}{n(p-1)}\mathrm{Cov}\Big(\mathrm{tr}\mathbf{D}_1^{-1}(z),\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big)\\
&-\frac{\nu_2^2}{n(p-1)}\mathrm{Cov}\Big(\mathrm{tr}\mathbf{D}_1^{-1}(z),\ \mathbf{1}_p'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{1}_p\Big).
\end{aligned}
$$

2. The last equality is due to

$$
\frac{1}{n}\mathrm{Cov}\Big(\mathrm{tr}\mathbf{D}_1^{-1}(z),\ \mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big)=O(n^{-1}), \tag{S2.102}
$$

$$
-\frac{1}{n(p-1)}\mathrm{Cov}\Big(\mathrm{tr}\mathbf{D}_1^{-1}(z),\ \mathbf{1}_p'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{1}_p\Big)=O(n^{-1}). \tag{S2.103}
$$

   The equation (S2.102) follows from the inequality

$$
\mathbb{E}\Big|\mathrm{tr}\mathbf{D}_1^{-1}(z)\mathbf{M}-\mathbb{E}\mathrm{tr}\mathbf{D}_1^{-1}(z)\mathbf{M}\Big|^2\le K\|\mathbf{M}\|^2, \tag{S2.104}
$$

   where $\mathbf{M}$ is any deterministic $p\times p$ matrix. The proof of (S2.104) is similar to Equation (4.7) of Bai and Silverstein [2004]. The equation (S2.103) follows from (S2.104) and Lemma S1.5.

Proof of (S2.99): For $\mathbb{E}\{\beta_1(z)\gamma_1^2(z)\mathcal{P}_1(z)\}$, we have

$$
\mathbb{E}\Big\{\beta_1(z)\gamma_1^2(z)\mathcal{P}_1(z)\Big\}
$$

$$
\begin{aligned}
&= \mathbb{E}\left\{n\beta_1(z)\gamma_1^2(z)\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1\right\} - \mathbb{E}\left\{\beta_1(z)\gamma_1^2(z)\right\}\mathbb{E}\left[\lambda \mathrm{tr}\left\{\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\}\right] \\
&= \mathbb{E}\left\{n\beta_1(z)\gamma_1^2(z)\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1\right\} - \mathbb{E}\left[\beta_1(z)\gamma_1^2(z)\mathrm{tr}\left\{\lambda\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\}\right] \\
&\qquad + \mathrm{Cov}\left(\beta_1(z)\gamma_1^2(z),\ \mathrm{tr}\left\{\lambda\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\}\right).
\end{aligned}
$$

From Lemma 5.4 and equation (S2.104), we have

$$
\begin{aligned}
&\mathbb{E}\left\{n\beta_1(z)\gamma_1^2(z)\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1\right\} - \mathbb{E}\left[\beta_1(z)\gamma_1^2(z)\mathrm{tr}\left\{\lambda\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\}\right] \\
&\le n\left\{\mathbb{E}\left|\gamma_1^2(z)\beta_1(z)\right|^2\right\}^{1/2}\left[\mathbb{E}\left|\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1 - \frac{\lambda}{n}\mathrm{tr}\left\{\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\}\right|^2\right]^{1/2} \\
&\le Kn(n^{-1}\delta_n^4)^{1/2}n^{-1/2} = K\delta_n^2,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{Cov}\left(\beta_1(z)\gamma_1^2(z),\ \mathrm{tr}\left\{\lambda\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\}\right) \\
&\le \left\{\mathbb{E}|\beta_1(z)|^4\right\}^{1/4}\left\{\mathbb{E}|\gamma_1^2(z)|^4\right\}^{1/4}\left(\mathbb{E}\left|\mathrm{tr}\left\{\lambda\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\} - \mathbb{E}\mathrm{tr}\left\{\lambda\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\}\right|^2\right)^{1/2} \\
&\le Kn^{-1/4}\delta_n^3.
\end{aligned}
$$

These estimates yield (S2.99).
Proof of (S2.100): From (S2.43), and

$$
\begin{aligned}
\mathbf{D}_1^{-1}(z) - \mathbf{D}^{-1}(z) &= \beta_1(z)\mathbf{D}_1^{-1}(z)\mathbf{r}_1\mathbf{r}_1'\mathbf{D}_1^{-1}(z), \\
\beta_1(z) &= b_p(z) - b_p(z)\gamma_1(z)\beta_1(z), \\
\mathbb{E}\beta_1(z) &= b_p(z) + o(n^{-1/2}),
\end{aligned}
$$

we have

$$
\begin{aligned}
&\mathbb{E}\{\beta_1(z)\mathcal{P}_2(z)\} \\
&= \lambda\mathbb{E}\{\beta_1(z)\}\mathrm{tr}\left[\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbb{E}\left\{\mathbf{D}_1^{-1}(z) - \mathbf{D}^{-1}(z)\right\}\right] \\
&= \lambda\mathbb{E}\{\beta_1(z)\}\mathbb{E}\left[\left\{b_p(z) - b_p(z)\beta_1(z)\gamma_1(z)\right\}\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\mathbf{r}_1\right] \\
&= \lambda b_p(z)\mathbb{E}\{\beta_1(z)\}\left[\mathbb{E}\left\{\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\mathbf{r}_1\right\} - \mathbb{E}\left\{\beta_1(z)\gamma_1(z)\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\mathbf{r}_1\right\}\right] \\
&= \lambda b_p^2(z)\mathbb{E}\left\{\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\mathbf{r}_1\right\} + O(n^{-1/2}) \\
&= \lambda b_p^2(z)\mathrm{tr}\left[\mathbb{E}\left\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\}\mathbb{E}(\mathbf{r}_1\mathbf{r}_1')\right] + O(n^{-1/2}) \\
&= \frac{\lambda b_p^2(z)p\nu_2}{n(p-1)}\mathrm{tr}\left[\mathbb{E}\left\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\}\left(\mathbf{I}_p - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p'\right)\right] + O(n^{-1/2}) \\
&= \frac{\lambda b_p^2(z)p\nu_2}{n(p-1)}\mathbb{E}\mathrm{tr}\left\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\right\} + O(n^{-1/2}),
\end{aligned}
$$

which is (S2.100). Below are some interpretations of the above equalities:

1. The fourth equality follows from

$$\mathbb{E}\Big\{\beta_1(z)\gamma_1(z)\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\mathbf{r}_1\Big\} = O(n^{-1/2}),$$

   which can be proved by using Lemma 5.4.

2. The last equality follows from

$$\frac{\lambda b_p^2(z)\nu_2}{n(p-1)}\mathbb{E}\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\mathbf{1}_p\mathbf{1}_p'\Big\} = O(n^{-1}).$$

   This can be proved by using Lemma 5.3 and Lemma S1.5.

S2.12 Proof of Lemma S2.4

Step 1: Consider J_1. By Lemma 2.3 in Bai and Silverstein [2004], we have

$$\frac{1}{p}\mathrm{tr}\mathbf{D}_j^{-1}(z) \xrightarrow{P} m(z).$$

By this estimate, Equation (S2.43) and Lemma S1.5, we get

$$\begin{aligned}
n\mathbb{E}\gamma_1(z) &= \frac{p\nu_2}{p-1}\mathbb{E}\mathrm{tr}\Big\{\Big(\mathbf{I}_p - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p'\Big)\mathbf{D}_1^{-1}(z)\Big\} - \nu_2\mathbb{E}\mathrm{tr}\mathbf{D}_1^{-1}(z)\\
&= -\frac{\nu_2}{p-1}\mathbb{E}\mathrm{tr}\Big\{\mathbf{1}_p'\mathbf{D}_1^{-1}(z)\mathbf{1}_p\Big\} + \frac{\nu_2}{p-1}\mathbb{E}\mathrm{tr}\mathbf{D}_1^{-1}(z)\\
&\to \lambda\{m(z) + 1/z\}.
\end{aligned}$$

By Lemma S1.5, we have

$$\begin{aligned}
&-\frac{b_p^2(z)\nu_2^2}{n(p-1)}\mathbb{E}\Big[\Big\{\mathrm{tr}\mathbf{D}_1^{-1}(z) - \mathbf{1}_p'\mathbf{D}_1^{-1}(z)\mathbf{1}_p\Big\}\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big]\\
\to &-\frac{cz^2\underline{m}^2(z)\lambda^2}{1+\lambda\underline{m}(z)}\Big\{m^2(z) + \frac{m(z)}{z}\Big\}.
\end{aligned}$$

By Lemma 5.3, Equation (S2.66) and the above estimates, we have

$$J_1 \to -\frac{z\underline{m}(z)}{1+\lambda\underline{m}(z)}\Big\{(\lambda+h_1)m(z) + \frac{\lambda}{z}\Big\} = -\frac{\underline{m}(z)\{z(\lambda+h_1)m(z)+\lambda\}}{1+\lambda\underline{m}(z)}.$$

Step 2: Consider J_2. By Lemma 5.5, we have

$$J_2 = J_{21} + 2J_{22} + J_{23} + o(1),$$

where

$$\begin{aligned}
J_{21} &= -\frac{(\nu_4 - 3\nu_{12})b_p^2(z)}{n}\sum_{i=1}^{p}\mathbb{E}\Big(\Big[\mathbf{D}_1^{-1}(z)\Big]_{ii}\Big[\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big]_{ii}\Big),\\
J_{22} &= -\frac{\nu_{12}b_p^2(z)}{n}\mathbb{E}\mathrm{tr}\Big\{\mathbf{D}_1^{-2}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\},
\end{aligned}$$

$$J_{23} = -\frac{(\nu_{12}-\nu_2^2)b_p^2(z)}{n}\mathbb{E}\Big[\mathrm{tr}\mathbf{D}_1^{-1}(z)\,\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\}\Big].$$

Since $\frac{1}{p}\sum_{i=1}^{p}\mathbb{E}\Big(\Big[\mathbf{D}_1^{-1}(z)\Big]_{ii}\Big[\mathbf{D}_1^{-1}(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\Big]_{ii}\Big) \to \frac{m^2(z)}{1+\lambda\underline{m}(z)}$, we have

$$J_{21} \to \frac{-c\alpha_1 z^2 m^2(z)\underline{m}^2(z)}{1+\lambda\underline{m}(z)}.$$

Note that $\frac{1}{p}\mathbb{E}\mathrm{tr}\mathbf{D}_1^{-2}(z) \to m'(z)$, thus we get

$$J_{22} \to -\frac{c\lambda^2 z^2 m'(z)\underline{m}^2(z)}{1+\lambda\underline{m}(z)}.$$

By Lemma 5.3, we get

$$J_{23} \to -\frac{c\alpha_2 z^2 m^2(z)\underline{m}^2(z)}{1+\lambda\underline{m}(z)}.$$

From these estimates, we have

$$J_2 \to -\frac{cz^2\underline{m}^2(z)\{(\alpha_1+\alpha_2)m^2(z)+2\lambda^2 m'(z)\}}{1+\lambda\underline{m}(z)}.$$

Step 3: Consider J_3. To calculate the limit of J_3, we can expand $\mathbf{D}_1^{-1}(z)$ like (S2.75) and find the limit of J_3 using the method similarly to Bai and Silverstein [2004]. The limit of J_3 is

$$\frac{c\lambda^2\underline{m}^2(z)}{\{1+\lambda\underline{m}(z)\}[\{1+\lambda\underline{m}(z)\}^2-c\lambda^2\underline{m}^2(z)]}.$$

S2.13 Proof of Lemma S1.5

By Lemma 5.4, we obtain, for any $2 \le q \in \mathbb{N}^+$, there exists a positive constant K_q such that

$$\begin{aligned}
&\mathbb{E}|\mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p'\mathbf{D}_j^{-1}(z)\mathbf{r}_j|^q \\
&\le K_q\Big(\mathbb{E}\Big|\mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p'\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{\nu_2}{n}\mathrm{tr}\Big\{\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p'\mathbf{D}_j^{-1}(z)\Big\}\Big|^q \\
&\qquad + \Big|\frac{\nu_2}{n}\mathbb{E}\mathrm{tr}\Big\{\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p'\mathbf{D}_j^{-1}(z)\Big\}\Big|^q\Big) \\
&\le K_q\Big(n^{-2}\delta_n^{2q-4}+n^{-q}\Big) = O(n^{-2}),
\end{aligned} \tag{S2.105}$$

where $\frac{1}{n}\mathrm{tr}\Big\{\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p'\mathbf{D}_j^{-1}(z)\Big\} = \frac{1}{np}\mathbf{1}_p'\mathbf{D}_j^{-2}(z)\mathbf{1}_p = O(n^{-1})$. Write

$$\begin{aligned}
&\mathbf{x}_p'\mathbf{D}^{-1}(z)\mathbf{x}_p - \mathbf{x}_p'\mathbb{E}\mathbf{D}^{-1}(z)\mathbf{x}_p \\
&= \sum_{j=1}^{n}\mathbf{x}_p'\Big\{(\mathbb{E}_j-\mathbb{E}_{j-1})\mathbf{D}^{-1}(z)\Big\}\mathbf{x}_p \\
&= \sum_{j=1}^{n}\mathbf{x}_p'\Big[(\mathbb{E}_j-\mathbb{E}_{j-1})\Big\{\mathbf{D}^{-1}(z)-\mathbf{D}_j^{-1}(z)\Big\}\Big]\mathbf{x}_p
\end{aligned}$$

$$= -\sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p'\mathbf{D}_j^{-1}(z)\mathbf{r}_j.$$

By Lemma S1.2, (S2.105), and $|\beta_j(z)| \le |z|/v$, we have

$$\begin{aligned}
&\mathbb{E}\Big|\mathbf{x}_p'\mathbf{D}^{-1}(z)\mathbf{x}_p - \mathbf{x}_p'\mathbb{E}\mathbf{D}^{-1}(z)\mathbf{x}_p\Big|^2 \\
&\le K\sum_{j=1}^{n}\mathbb{E}|(\mathbb{E}_j - \mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p'\mathbf{D}_j^{-1}(z)\mathbf{r}_j|^2 \\
&\le K\sum_{j=1}^{n}\mathbb{E}|\beta_j(z)\mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p'\mathbf{D}_j^{-1}(z)\mathbf{r}_j|^2 \\
&\le Kn^{-1}.
\end{aligned}$$

Thus, we have

$$\mathbb{E}|\mathbf{x}_p'\mathbf{D}^{-1}(z)\mathbf{x}_p - \mathbf{x}_p'\mathbb{E}\mathbf{D}^{-1}(z)\mathbf{x}_p|^2 \to 0. \tag{S2.106}$$

Recall that $\widetilde{\mathbf{Q}}_p(z) = \mathbf{I}_p + \lambda\mathbb{E}\underline{m}_p(z)\mathbf{I}_p$. From $\underline{m}_p(z) = -\frac{1}{nz}\sum_{j=1}^{n}\beta_j(z)$ and $\mathbf{r}_j'\mathbf{D}^{-1}(z) = \beta_j(z)\mathbf{r}_j'\mathbf{D}_j^{-1}(z)$, we obtain

$$\begin{aligned}
&\Big\{-z\widetilde{\mathbf{Q}}_p(z)\Big\}^{-1} - \mathbf{D}^{-1}(z) \\
&= -\frac{1}{z}\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\{\mathbf{D}(z) + z\widetilde{\mathbf{Q}}_p(z)\Big\}\mathbf{D}^{-1}(z) \\
&= -\frac{1}{z}\widetilde{\mathbf{Q}}_p^{-1}(z)\left\{\sum_{j=1}^{n}\mathbf{r}_j\mathbf{r}_j' + z\lambda\mathbb{E}\underline{m}_p(z)\mathbf{I}_p\right\}\mathbf{D}^{-1}(z) \\
&= -\frac{1}{z}\sum_{j=1}^{n}\beta_j(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_j\mathbf{r}_j'\mathbf{D}_j^{-1}(z) + \frac{\lambda\mathbb{E}\{\beta_1(z)\}}{z}\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}^{-1}(z).
\end{aligned}$$

Taking expectation of the above identity yields that

$$\begin{aligned}
&\Big\{-z\widetilde{\mathbf{Q}}_p(z)\Big\}^{-1} - \mathbb{E}\mathbf{D}^{-1}(z) \\
&= -\frac{n}{z}\mathbb{E}\Big\{\beta_1(z)\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1\mathbf{r}_1'\mathbf{D}_j^{-1}(z)\Big\} + \frac{\lambda\mathbb{E}\{\beta_1(z)\}}{z}\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbb{E}\mathbf{D}^{-1}(z).
\end{aligned}$$

Multiplying by $-\mathbf{x}_p'$ on the left and $\mathbf{x}_p$ on the right, we have

$$\mathbf{x}_p'\mathbb{E}\mathbf{D}^{-1}(z)\mathbf{x}_p + \frac{1}{z\{1 + \lambda\mathbb{E}\underline{m}_p(z)\}} = \varrho_1(z) + \varrho_2(z) + \varrho_3(z),$$

where

$$\begin{aligned}
\varrho_1(z) &:= \frac{n}{z}\mathbb{E}\{\beta_1(z)\varrho_{11}(z)\}, \\
\varrho_{11}(z) &:= \mathbf{x}_p'\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{r}_1\mathbf{r}_1'\mathbf{D}_1^{-1}(z)\mathbf{x}_p - \frac{\lambda}{n}\mathbf{x}_p'\widetilde{\mathbf{Q}}_p^{-1}(z)\mathbf{D}_1^{-1}(z)\mathbf{x}_p, \\
\varrho_2(z) &:= \frac{\lambda}{z}\mathbb{E}\Big[\beta_1(z)\mathbf{x}_p'\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\{\mathbf{D}_1^{-1}(z) - \mathbf{D}^{-1}(z)\Big\}\mathbf{x}_p\Big], \\
\varrho_3(z) &:= \frac{\lambda}{z}\mathbb{E}\Big[\beta_1(z)\mathbf{x}_p'\widetilde{\mathbf{Q}}_p^{-1}(z)\Big\{\mathbf{D}^{-1}(z) - \mathbb{E}\mathbf{D}^{-1}(z)\Big\}\mathbf{x}_p\Big].
\end{aligned}$$

Recalling the notations defined above and the following equalities:

$$\varrho_1(z) = \frac{n}{z}\mathbb{E}\Big\{\overline{\beta}_1(z)\varrho_{11}(z)\Big\} - \frac{n}{z}\mathbb{E}\Big\{\beta_1(z)\overline{\beta}_1(z)\varepsilon_1(z)\varrho_{11}(z)\Big\}, \tag{S2.107}$$

$$\overline{\beta}_1(z) = b_p(z) - \frac{\nu_2}{n}b_p(z)\overline{\beta}_1(z)\mathrm{tr}\Big\{\mathbf{D}_1^{-1}(z) - \mathbb{E}\mathbf{D}_1^{-1}(z)\Big\}. \tag{S2.108}$$

From (S2.107) – (S2.108), Lemma 5.3 and Lemma 5.4, it is easy to see that

$$\varrho_1(z) = \frac{n\mathbb{E}\varrho_{11}(z)}{z}\left\{\frac{1}{1+c\lambda m(z)} + o(1)\right\}. \tag{S2.109}$$

From (S2.43) and Lemma 5.3, we have

$$\mathbb{E}\varrho_{11}(z) = -\frac{1}{\lambda\underline{m}(z)+1}\frac{\lambda}{n}\Big\{\mathbf{x}_p'\mathbb{E}\mathbf{D}_1^{-1}(z)\mathbf{x}_p + o(1)\Big\}. \tag{S2.110}$$

Therefore, by (S2.109) – (S2.110), and Lemma 5.3 we have

$$\varrho_1(z) = \frac{\lambda\underline{m}(z)}{\lambda\underline{m}(z)+1}\mathbf{x}_p'\mathbb{E}\mathbf{D}_1^{-1}(z)\mathbf{x}_p + o(1).$$

Similarly to Bai et al. [2007], one may have $\varrho_2(z) = o(1)$ and $\varrho_3(z) = o(1)$. Hence, we obtain

$$\frac{\mathbf{x}_p'\mathbb{E}\mathbf{D}^{-1}(z)\mathbf{x}_p}{1+\lambda\underline{m}(z)} + \frac{1}{z\{1+\mathbb{E}\lambda\underline{m}(z)\}} = o(1),$$

which implies that

$$\mathbf{x}_p'\mathbb{E}\mathbf{D}^{-1}(z)\mathbf{x}_p \to -\frac{1}{z}.$$

This, together with (S2.106), completes the proof of Lemma S1.5.

S2.14 Proof of Corollary 2.6

We begin by presenting some expressions that will be used in the subsequent calculations:

$$z = -\frac{1}{\underline{m}} + \frac{c\lambda}{1+\lambda\underline{m}}, \tag{S2.111}$$

$$\mathrm{d}z = \frac{(1+\lambda\underline{m})^2 - c\lambda^2\underline{m}^2}{\underline{m}^2(1+\lambda\underline{m})^2}\,\mathrm{d}\underline{m}, \tag{S2.112}$$

$$m = \frac{-\underline{m}}{-1-\lambda\underline{m}+c\lambda\underline{m}}, \tag{S2.113}$$

$$\frac{\mathrm{d}m}{\mathrm{d}z} = \frac{\underline{m}^2(1+\lambda\underline{m})^2}{\{(1+\lambda\underline{m})^2 - c\lambda^2\underline{m}^2\}(-1-\lambda\underline{m}+c\lambda\underline{m})^2}. \tag{S2.114}$$

As their derivations are straightforward, the proofs are omitted.

S2.14.1 Calculation of expectation

The contours $\mathcal{C}$ is closed and taken in the positive direction in the complex plane, enclosing the support of F^c. Let $\tilde{C}$ be $\underline{m}(\mathcal{C})$.

For $f_1 = x$,

$$
\begin{aligned}
\mathbb{E}X_x = &\frac{c\lambda^2}{2\pi i}\oint_{\mathcal{C}} \frac{z\underline{m}^3(z)\{1+\lambda\underline{m}(z)\}}{[\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)]^2}\,\mathrm{d}z \\
&- \frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{z^2\underline{m}^2(z)\{h_1 m(z)+\lambda m(z)+\lambda/z\}\{1+\lambda\underline{m}(z)\}}{\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)}\,\mathrm{d}z \\
&- \frac{c}{2\pi i}\oint_{\mathcal{C}} \frac{z^3\underline{m}^3(z)\{(\alpha_1+\alpha_2)m^2(z)+2\lambda^2 m'(z)\}\{1+\lambda\underline{m}(z)\}}{\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)}\,\mathrm{d}z, \\
=: &\, I_1(f_1)+I_2(f_1)+I_3(f_1).
\end{aligned}
$$

For $I_1(f_1)$, by using (S2.111) and (S2.112), we get

$$
\begin{aligned}
I_1(f_1) &= \frac{c\lambda^2}{2\pi i}\oint_{\mathcal{C}} \frac{z\underline{m}^3(z)\{1+\lambda\underline{m}(z)\}}{[\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)]^2}\,\mathrm{d}z \\
&= \frac{c\lambda^2}{2\pi i}\oint_{\tilde{C}} \frac{\underline{m}^3(1+\lambda\underline{m})}{\{(1+\lambda\underline{m})^2 - c\lambda^2\underline{m}^2\}^2}\,\frac{(-1-\lambda\underline{m}+c\lambda\underline{m})\{(1+\lambda\underline{m})^2-c\lambda^2\underline{m}^2\}}{\underline{m}^3(1+\lambda\underline{m})^3}\,\mathrm{d}\underline{m} \\
&= \frac{1}{2\pi i}\oint_{\tilde{C}} \frac{c\lambda^2(1+\lambda\underline{m}-c\lambda\underline{m})}{(1+\lambda\underline{m})^2\{c\lambda^2\underline{m}^2-(1+\lambda\underline{m})^2\}}\,\mathrm{d}\underline{m}.
\end{aligned}
$$

The poles of $I_1(f_1)$ are $-\lambda^{-1}, -\frac{1}{(1\pm\sqrt{c})\lambda}$, we have by the residue theorem

$$
I_1(f_1) = \lambda(1+c) - \frac{\lambda}{2}(1+\sqrt{c})^2 - \frac{\lambda}{2}(1-\sqrt{c})^2 = 0.
$$

For the second integral $I_2(f_1)$, by using (S2.111) – (S2.113), we have,

$$
\begin{aligned}
I_2(f_1) &= -\frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{z^2\underline{m}^2(z)\{1+\lambda\underline{m}(z)\}\{h_1 m(z)+\lambda m(z)+\lambda/z\}}{\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)}\,\mathrm{d}z \\
&= -\frac{1}{2\pi i}\oint_{\tilde{C}} \frac{\underline{m}^2(1+\lambda\underline{m})}{(1+\lambda\underline{m})^2-c\lambda^2\underline{m}^2}\,\frac{1}{c}\left\{(h_1+\lambda)\underline{m}+(h_1+\lambda-h_1c)\frac{\underline{m}(1+\lambda\underline{m})}{-1-\lambda\underline{m}+c\lambda\underline{m}}\right\} \\
&\qquad\qquad \times \frac{(-1-\lambda\underline{m}+c\lambda\underline{m})^2\{(1+\lambda\underline{m})^2-c\lambda^2\underline{m}^2\}}{\underline{m}^4(1+\lambda\underline{m})^4}\,\mathrm{d}\underline{m} \\
&= -\frac{1}{2\pi i}\oint_{\tilde{C}} \frac{(\lambda^2\underline{m}-h_1)(-1-\lambda\underline{m}+c\lambda\underline{m})}{\underline{m}(1+\lambda\underline{m})^3}\,\mathrm{d}\underline{m}.
\end{aligned}
$$

The pole of $I_2(f_1)$ is $-\lambda^{-1}$, we have by the residue theorem

$$
I_2(f_1) = h_1.
$$

For $I_3(f_1)$, by using (S2.111) – (S2.114), we get

$$
\begin{aligned}
I_3(f_1) &= -\frac{c}{2\pi i}\oint_{\mathcal{C}} \frac{z^3\underline{m}^3(z)\{(\alpha_1+\alpha_2)m^2(z)+2\lambda^2 m'(z)\}\{1+\lambda\underline{m}(z)\}}{\{1+\lambda\underline{m}(z)\}^2 - c\lambda^2\underline{m}^2(z)}\,\mathrm{d}z \\
&= -\frac{1}{2\pi i}\oint_{\tilde{C}}\left[(\alpha_1+\alpha_2)\left(\frac{-\underline{m}}{-1-\lambda\underline{m}+c\lambda\underline{m}}\right)^2 + \frac{2\lambda^2\underline{m}^2(1+\lambda\underline{m})^2}{\{(1+\lambda\underline{m})^2-c\lambda^2\underline{m}^2\}(-1-\lambda\underline{m}+c\lambda\underline{m})^2}\right]
\end{aligned}
$$

$$\times \frac{c\underline{m}^3(1+\lambda\underline{m})}{(1+\lambda\underline{m})^2-c\lambda^2\underline{m}^2} \times \frac{(-1-\lambda\underline{m}+c\lambda\underline{m})^3\{(1+\lambda\underline{m})^2-c\lambda^2\underline{m}^2\}}{\underline{m}^5(1+\lambda\underline{m})^5}\,\mathrm{d}\underline{m}$$

$$= -\frac{1}{2\pi i}\oint_{\tilde{\mathcal{C}}}\left\{\frac{c(\alpha_1+\alpha_2)(-1-\lambda\underline{m}+c\lambda\underline{m})}{(1+\lambda\underline{m})^4}\,\mathrm{d}\underline{m} + \frac{2\lambda^2c(-1-\lambda\underline{m}+c\lambda\underline{m})}{(1+\lambda\underline{m})^2\{(1+\lambda\underline{m})^2-c\lambda^2\underline{m}^2\}}\right\}\mathrm{d}\underline{m}.$$

The poles of $I_3(f_1)$ are $-\frac{1}{\lambda}$, $-\frac{1}{(1\pm\sqrt{c})\lambda}$, we have by the residue theorem

$$I_3(f_1) = -2\lambda(1+c) + \lambda(1+\sqrt{c})^2 + \lambda(1-\sqrt{c})^2 = 0.$$

Thus, we get

$$\mathbb{E}X_x = h_1.$$

For $f_2 = x^2$, we have

$$\begin{aligned}
\mathbb{E}X_{x^2} &= \frac{c\lambda^2}{2\pi i}\oint_{\mathcal{C}}\frac{z^2\underline{m}^3(z)\{1+\lambda\underline{m}(z)\}}{[\{1+\lambda\underline{m}(z)\}^2-c\lambda^2\underline{m}^2(z)]^2}\,\mathrm{d}z \\
&\quad -\frac{1}{2\pi i}\oint_{\mathcal{C}}\frac{z^3\underline{m}^2(z)\{h_1m(z)+\lambda m(z)+\lambda/z\}\{1+\lambda\underline{m}(z)\}}{\{1+\lambda\underline{m}(z)\}^2-c\lambda^2\underline{m}^2(z)}\,\mathrm{d}z \\
&\quad -\frac{c}{2\pi i}\oint_{\mathcal{C}}\frac{z^4\underline{m}^3(z)\{(\alpha_1+\alpha_2)m^2(z)+2\lambda^2m'(z)\}\{1+\lambda\underline{m}(z)\}}{\{1+\lambda\underline{m}(z)\}^2-c\lambda^2\underline{m}^2(z)}\,\mathrm{d}z \\
&= \frac{1}{2\pi i}\oint_{\tilde{\mathcal{C}}}\frac{c\lambda^2(-c\lambda\underline{m}+\lambda\underline{m}+1)^2}{\underline{m}(1+\lambda\underline{m})^3\{(1+\lambda\underline{m})^2-c\lambda^2\underline{m}^2\}}\,\mathrm{d}\underline{m} \\
&\quad -\frac{1}{2\pi i}\oint_{\tilde{\mathcal{C}}}\frac{(\lambda^2\underline{m}-h_1)(-1-\lambda\underline{m}+c\lambda\underline{m})^2}{\underline{m}^2(1+\lambda\underline{m})^4}\,\mathrm{d}\underline{m} \\
&\quad -\frac{1}{2\pi i}\oint_{\tilde{\mathcal{C}}}\left\{\frac{c(\alpha_1+\alpha_2)(-1-\lambda\underline{m}+c\lambda\underline{m})^2}{\underline{m}(1+\lambda\underline{m})^5}\,\mathrm{d}\underline{m} + \frac{2\lambda^2c(-1-\lambda\underline{m}+c\lambda\underline{m})^2}{\underline{m}(1+\lambda\underline{m})^3\{(1+\lambda\underline{m})^2-c\lambda^2\underline{m}^2\}}\right\}\mathrm{d}\underline{m} \\
&=: I_1(f_2)+I_2(f_2)+I_3(f_2).
\end{aligned}$$

For the first integral $I_1(f_2)$, the poles are $-\lambda^{-1}$, $-\frac{1}{(1\pm\sqrt{c})\lambda}$, we have by the residue theorem,

$$I_1(f_2) = -c\lambda^2.$$

For the second integral $I_2(f_2)$, the pole is $-\lambda^{-1}$, we have by the residue theorem,

$$I_2(f_2) = \lambda(\lambda+2ch_1+2h_1).$$

For the third integral $I_3(f_2)$, the poles are $-\lambda^{-1}$, $-\frac{1}{(1\pm\sqrt{c})\lambda}$, we have by the residue theorem,

$$I_3(f_2) = c(\alpha_1+\alpha_2)+2\lambda^2c.$$

Thus,

$$\mathbb{E}X_{x^2} = (c+1)\lambda^2 + 2(c+1)\lambda h_1 + c(\alpha_1+\alpha_2).$$

For $f_3 = x^3$, we have

$$\mathbb{E}X_{x^3} = \frac{c\lambda^2}{2\pi i}\oint_{\mathcal{C}}\frac{z^3\underline{m}^3(z)\{1+\lambda\underline{m}(z)\}}{[\{1+\lambda\underline{m}(z)\}^2-c\lambda^2\underline{m}^2(z)]^2}\,\mathrm{d}z$$

$$
\begin{aligned}
&- \frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{z^4\underline{m}^2(z)\{h_1 m(z)+\lambda m(z)+\lambda/z\}\{1+\lambda\underline{m}(z)\}}{\{1+\lambda\underline{m}(z)\}^2-c\lambda^2\underline{m}^2(z)}\,\mathrm{d}z\\
&- \frac{c}{2\pi i}\oint_{\mathcal{C}} \frac{z^5\underline{m}^3(z)\{(\alpha_1+\alpha_2)m^2(z)+2\lambda^2 m'(z)\}\{1+\lambda\underline{m}(z)\}}{\{1+\lambda\underline{m}(z)\}^2-c\lambda^2\underline{m}^2(z)}\,\mathrm{d}z\\
=\;& \frac{1}{2\pi i}\oint_{\tilde{\mathcal{C}}} \frac{c\lambda^2(-1-\lambda\underline{m}+c\lambda\underline{m})^3}{\underline{m}^2(1+\lambda\underline{m})^4\{(1+\lambda\underline{m})^2-c\lambda^2\underline{m}^2\}}\,\mathrm{d}\underline{m}\\
&- \frac{1}{2\pi i}\oint_{\tilde{\mathcal{C}}} \frac{(\lambda^2\underline{m}-h_1)(-1-\lambda\underline{m}+c\lambda\underline{m})^3}{\underline{m}^3(1+\lambda\underline{m})^5}\,\mathrm{d}\underline{m}\\
&- \frac{1}{2\pi i}\oint_{\tilde{\mathcal{C}}} \left\{\frac{c(\alpha_1+\alpha_2)(-1-\lambda\underline{m}+c\lambda\underline{m})^3}{\underline{m}^2(1+\lambda\underline{m})^6}\,\mathrm{d}\underline{m} + \frac{2\lambda^2 c(-1-\lambda\underline{m}+c\lambda\underline{m})^3}{\underline{m}^2(1+\lambda\underline{m})^4\{(1+\lambda\underline{m})^2-c\lambda^2\underline{m}^2\}}\right\}\mathrm{d}\underline{m}\\
=:\;& I_1(f_3)+I_2(f_3)+I_3(f_3).
\end{aligned}
$$

For the first integral $I_1(f_3)$, the poles are $-\lambda^{-1}$, $-\frac{1}{(1\pm\sqrt{c})\lambda}$, we have by the residue theorem

$$I_1(f_3) = -3c(1+c)\lambda^3.$$

For the second integral $I_2(f_3)$, the pole is $-\lambda^{-1}$, we have by the residue theorem

$$I_2(f_3) = \lambda^2\{(2+3c)\lambda + 3(1+3c+c^2)h_1\}.$$

For the third integral $I_3(f_3)$, the poles are $-\lambda^{-1}$, $-\frac{1}{(1\pm\sqrt{c})\lambda}$, we have by the residue theorem

$$I_3(f_3) = 3c(1+c)\lambda(\alpha_1+\alpha_2) + 6c(1+c)\lambda^3.$$

Thus,

$$\mathbb{E}X_{x^3} = (3c^2+6c+2)\lambda^3 + 3(c^2+3c+1)\lambda^2 h_1 + 3c(c+1)\lambda(\alpha_1+\alpha_2).$$

S2.14.2 Calculation of variance

We claim that

$$
\begin{aligned}
&\mathrm{Cov}(X_{x^{r_1}}, X_{x^{r_2}})\\
&= 2(\lambda c)^{r_1+r_2}\sum_{k_1=0}^{r_1-1}\sum_{k_2=0}^{r_2}\binom{r_1}{k_1}\binom{r_2}{k_2}\left(\frac{1-c}{c}\right)^{k_1+k_2}\sum_{\ell=1}^{r_1-k_1}\ell\binom{2r_1-1-k_1-\ell}{r_1-1}\binom{2r_2-1-k_2+\ell}{r_2-1} \qquad \text{(S2.115)}\\
&\quad + \frac{c}{\lambda^2}(\alpha_1+\alpha_2)(\lambda c)^{r_1+r_2}\sum_{k_1=0}^{r_1}\sum_{k_2=0}^{r_2}\binom{r_1}{k_1}\binom{r_2}{k_2}\left(\frac{1-c}{c}\right)^{k_1+k_2}\binom{2r_1-k_1}{r_1-1}\binom{2r_2-k_2}{r_2-1}, \qquad \text{(S2.116)}
\end{aligned}
$$

where $r_1, r_2 \in \mathbb{N}^+$. By using this result, we obtain the variances in Corollary 2.6. It suffices to prove the above equation. The contours $\mathcal{C}_1, \mathcal{C}_2$ are closed and taken in the positive direction in the complex plane, each enclosing the support of F^c. Let $\tilde{\mathcal{C}}_i$ be $\underline{m}(\mathcal{C}_i)$ for $i=1,2$. From Theorem 2.5, we have

$$
\begin{aligned}
\mathrm{Cov}(X_{x^{r_1}}, X_{x^{r_2}}) = &-\frac{1}{2\pi^2}\oint_{\tilde{\mathcal{C}}_1}\oint_{\tilde{\mathcal{C}}_2}\frac{z_1^{r_1}z_2^{r_2}}{(\underline{m}_1-\underline{m}_2)^2}\,\mathrm{d}\underline{m}_1\,\mathrm{d}\underline{m}_2\\
&-\frac{c(\alpha_1+\alpha_2)}{4\pi^2}\oint_{\tilde{\mathcal{C}}_1}\oint_{\tilde{\mathcal{C}}_2}\frac{z_1^{r_1}z_2^{r_2}}{(1+\lambda\underline{m}_1)^2(1+\lambda\underline{m}_2)^2}\,\mathrm{d}\underline{m}_1\,\mathrm{d}\underline{m}_2,
\end{aligned}
$$

$$=: \mathsf{Cov}_{r_1,r_2,1} + \mathsf{Cov}_{r_1,r_2,2}.$$

The proof of "$\mathsf{Cov}_{r_1,r_2,1} =$ (S2.115)" is exactly analogous with Bai and Silverstein [2004], it is then omitted. Now, we prove that "$\mathsf{Cov}_{r_1,r_2,2} =$ (S2.116)". Note that

$$\oint_{\tilde{\mathcal{C}}_1}\oint_{\tilde{\mathcal{C}}_2} \frac{z_1^{r_1} z_2^{r_2}}{(1+\lambda \underline{m}_1)^2(1+\lambda \underline{m}_2)^2}\,\mathrm{d}\underline{m}_1\,\mathrm{d}\underline{m}_2 = \oint_{\tilde{\mathcal{C}}_1} \frac{z_1^{r_1}}{(1+\lambda \underline{m}_1)^2}\,\mathrm{d}\underline{m}_1 \times \oint_{\tilde{\mathcal{C}}_2} \frac{z_2^{r_2}}{(1+\lambda \underline{m}_2)^2}\,\mathrm{d}\underline{m}_2.$$

By (S2.111), we have

$$\begin{aligned}
&\oint_{\tilde{\mathcal{C}}_1} \frac{z_1^{r_1}}{(1+\lambda \underline{m}_1)^2}\,\mathrm{d}\underline{m}_1 \\
&= \oint_{\tilde{\mathcal{C}}_1} \frac{\left(-\frac{1}{\underline{m}_1} + \frac{c\lambda}{1+\lambda \underline{m}_1}\right)^{r_1}}{(1+\lambda \underline{m}_1)^2}\,\mathrm{d}\underline{m}_1 \\
&= (\lambda c)^{r_1} \oint_{\tilde{\mathcal{C}}_1} \left(\frac{1}{1+\lambda \underline{m}_1} + \frac{1-c}{c}\right)^{r_1} \{1-(1+\lambda \underline{m}_1)\}^{-r_1}(1+\lambda \underline{m}_1)^{-2}\,\mathrm{d}\underline{m}_1 \\
&= (\lambda c)^{r_1} \oint_{\tilde{\mathcal{C}}_1} \sum_{k_1=0}^{r_1} \binom{r_1}{k_1}\left(\frac{1-c}{c}\right)^{k_1}(1+\lambda \underline{m}_1)^{k_1-r_1} \sum_{j=0}^{\infty}\binom{r_1+j-1}{j}(1+\lambda \underline{m}_1)^j(1+\lambda \underline{m}_1)^{-2}\,\mathrm{d}\underline{m}_1 \\
&= (\lambda c)^{r_1} \sum_{k_1=0}^{r_1} \binom{r_1}{k_1}\left(\frac{1-c}{c}\right)^{k_1} \oint_{\tilde{\mathcal{C}}_1} \sum_{j=0}^{\infty}\binom{r_1+j-1}{j}(1+\lambda \underline{m}_1)^{k_1-r_1+j-2}\,\mathrm{d}\underline{m}_1,
\end{aligned}$$

by substitution $\widetilde{\underline{m}}_1 = \lambda \underline{m}_1$, we get

$$\oint_{\tilde{\mathcal{C}}_1} \frac{z_1^{r_1}\,\mathrm{d}\underline{m}_1}{(1+\lambda \underline{m}_1)^2} = \frac{(\lambda c)^{r_1}}{\lambda} \sum_{k_1=0}^{r_1} \binom{r_1}{k_1}\left(\frac{1-c}{c}\right)^{k_1} \oint_{\tilde{\tilde{\mathcal{C}}}_1} \sum_{j=0}^{\infty}\binom{r_1+j-1}{j}(1+\widetilde{\underline{m}}_1)^{k_1-r_1+j-2}\,\mathrm{d}\widetilde{\underline{m}}_1,$$

where $\tilde{\tilde{\mathcal{C}}}_1$ is the $\widetilde{\underline{m}}_1$ contour. For this integral, the pole is -1, we have by residual theorem

$$\oint_{\tilde{\mathcal{C}}_1} \frac{z_1^{r_1}}{(1+\lambda \underline{m}_1)^2}\,\mathrm{d}\underline{m}_1 = \frac{2\pi i}{\lambda}(\lambda c)^{r_1} \sum_{k_1=0}^{r_1} \binom{r_1}{k_1}\left(\frac{1-c}{c}\right)^{k_1}\binom{2r_1-k_1}{r_1-1}.$$

Similarly, we get

$$\oint_{\tilde{\mathcal{C}}_2} \frac{z_2^{r_2}}{(1+\lambda \underline{m}_2)^2}\,\mathrm{d}\underline{m}_2 = \frac{2\pi i}{\lambda}(\lambda c)^{r_2} \sum_{k_2=0}^{r_2} \binom{r_2}{k_2}\left(\frac{1-c}{c}\right)^{k_2}\binom{2r_2-k_2}{r_2-1}.$$

Using the two equations above, we derive (S2.116).

S2.15 Proof of Theorem 3.1

Before moving to the proof of Theorem 3.1, we first establish a lemma of $\nu_2^\circ := \mathbb{E}\{pw_1/(\sum_{k=1}^p w_k)-1\}^2$ as follows. The proof of this lemma is provided in Section S2.17.

Lemma S2.5. *Suppose that $p \asymp n$ and $\mathbf{w} = \left(w_1,\ldots,w_p\right)'$ has positive i.i.d. entries with $\mathbb{E}w_1 = \mu > 0$, $\mathbb{E}(w_1-\mu)^2 = \sigma^2$, and $\mathbb{E}\left|w_1-\mu\right|^{6+s} < \infty$ for any $s > 0$, then we have*

$$\nu_2^\circ = \lambda + \frac{h_1}{p} + o(p^{-1}),$$

where $\lambda = \sigma^2/\mu^2$, and $h_1 = -2\mathbb{E}w_1^3/\mu^3 + 3\lambda^2 + 5\lambda + 2$.

Recall that

$$T = \frac{1}{p}\mathrm{tr}(\mathbf{B}_{p,N}^2) - \frac{2\nu_2^\circ}{p-1}\mathrm{tr}(\mathbf{B}_{p,N}) + \frac{p(\nu_2^\circ)^2}{p-1}.$$

Taking $r_1 = 1$ and $r_2 = 2$ in (S2.115) – (S2.116), we obtain that

$$V_{12} := \mathrm{Cov}\Big(\mathrm{tr}(\mathbf{B}_{p,N}),\ \mathrm{tr}(\mathbf{B}_{p,N}^2)\Big) = 2\lambda c(1+c)(2\lambda^2 + \alpha_1 + \alpha_2).$$

From Theorem 2.5 and Corollary 2.6, we have the following joint CLT:

$$p\begin{pmatrix} \frac{1}{p}\mathrm{tr}(\mathbf{B}_{p,N}^2) - \lambda^2(1+c_N) - \frac{\mu_2}{p} \\ \frac{1}{p}\mathrm{tr}(\mathbf{B}_{p,N}) - \lambda - \frac{\mu_1}{p} \end{pmatrix} \xrightarrow{D} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} V_2 & V_{12} \\ V_{12} & V_1 \end{pmatrix}\right).$$

Define a function $f(x,y) = x - \frac{2p\nu_2^\circ}{p-1}y + \frac{p(\nu_2^\circ)^2}{p-1}$, then we have

$$\begin{aligned}
&T = f\left(\frac{1}{p}\mathrm{tr}(\mathbf{B}_{p,N}^2), \frac{1}{p}\mathrm{tr}(\mathbf{B}_{p,N})\right),\\
&\nabla f\left(\lambda^2(1+c_N) + \frac{\mu_2}{p},\ \lambda + \frac{\mu_1}{p}\right) = \left(1,\ -\frac{2p\nu_2^\circ}{p-1}\right)',\\
&f\left(\lambda^2(1+c_N) + \frac{\mu_2}{p},\ \lambda + \frac{\mu_1}{p}\right) = \underbrace{\lambda^2 c_N + \frac{\mu_2}{p} - \frac{2\lambda h_1 + \lambda^2}{p-1}}_{=:\mu_T} + O(p^{-2}),
\end{aligned}$$

where in the last equation we use Lemma S2.5. By the Delta method, we obtain that $p(T-\mu_T)$ is asymptotically Gaussian with mean zero and variance

$$\sigma_T^2 = \lim_{p\to\infty}\left\{\left(1,\ -\frac{2p\nu_2^\circ}{p-1}\right)\begin{pmatrix} V_2 & V_{12} \\ V_{12} & V_1 \end{pmatrix}\begin{pmatrix} 1 \\ -\frac{2p\nu_2^\circ}{p-1}\end{pmatrix}\right\} = 4\lambda^2 V_1 - 4\lambda V_{12} + V_2.$$

This completes the proof of Theorem 3.1.

S2.16 Proof of Theorem 3.2

Define

$$T^* = \frac{1}{p}\left\|\mathbf{B}_{p,N_1}^{(1)} - \frac{p\nu_2^\circ}{p-1}\mathbf{G}_p\right\|_F^2,$$

then we have

$$\begin{aligned}
\widetilde{T} - T^* &= \frac{2(\hat{\nu}_2^\circ - \nu_2^\circ)}{p-1}\mathrm{tr}\Big(\mathbf{B}_{p,N_1}^{(1)}\Big) + \frac{p\{(\nu_2^\circ)^2 - (\hat{\nu}_2^\circ)^2\}}{p-1}\\
&= \frac{2\{\hat{\lambda} - \lambda - h_1/p - o(p^{-1})\}}{p-1}\Big\{p\lambda + O_P(1)\Big\} + \frac{p\{\lambda^2 + 2\lambda h_1/p + O(p^{-2}) - \hat{\lambda}^2\}}{p-1}\\
&= -\frac{p(\lambda - \hat{\lambda})^2}{p-1} + o_P(p^{-1})\\
&= O_P(p^{-2}) + o_P(p^{-1}) = o_P(p^{-1}).
\end{aligned}$$

The second equality follows from Corollary 2.6 and Lemma S2.5. In the fourth inequalities we use the CLT of $\hat{\lambda}$ as shown in (S2.118). Let

$$\tilde{\mu}_T = \frac{p\lambda^2}{n_1 - 1} + \frac{\hat{\mu}_2}{p} - \frac{2\hat{\lambda}\hat{h}_1 + \hat{\lambda}^2}{p-1},$$

then we have

$$\begin{aligned} p(\widetilde{T} - \hat{\mu}_T) &= p(\widetilde{T} - T^*) + p(T^* - \tilde{\mu}_T) + \frac{p^2(\lambda^2 - \hat{\lambda}^2)}{n_1 - 1} \\ &= p(T^* - \tilde{\mu}_T) + pc_1(\lambda^2 - \hat{\lambda}^2) + o_P(1). \end{aligned} \tag{S2.117}$$

The two leading terms in (S2.117) are independent, so we consider their asymptotic distributions separately. From Theorem 3.1, we obtain the asymptotic distribution of the first term in (S2.117) as follows:

$$p(T^* - \tilde{\mu}_T) \xrightarrow{D} \mathcal{N}(0, \sigma^2_{\widetilde{T}}),$$

where $\sigma^2_{\widetilde{T}}$ is defined similarly to σ^2_T in Theorem 3.1, with the limiting value c replaced by c_1. We claim that the asymptotic distribution of $\hat{\lambda}$ (will be proven later) is

$$p(\hat{\lambda} - \lambda) \xrightarrow{D} \mathcal{N}\Big(h_1, c_2\{\mathbb{E}(\tilde{w}_{11} - 1)^4 - \lambda^2 + h_2 - 2\lambda h_1\}\Big), \tag{S2.118}$$

where $\tilde{w}_{11}$ is defined below in (S2.119). Using the Delta method again, we have

$$pc_1(\hat{\lambda}^2 - \lambda^2) \xrightarrow{D} \mathcal{N}\Big(2c_1\lambda h_1, 4\lambda^2 c_1^2 c_2\{\mathbb{E}(\tilde{w}_{11} - 1)^4 - \lambda^2 + h_2 - 2\lambda h_1\}\Big).$$

This, together with (S2.117), Theorem 3.1, and the fact that $\widetilde{T}$ and $\hat{\lambda}$ are independent, implies the asymptotic distribution of $\widetilde{T}$.

Finally, we prove (S2.118). Let

$$\tilde{w}_{ij} := \frac{w_{ij}}{\mu}, \qquad w_i^{[k]} \equiv w_i^{[k,p]} := \frac{1}{p}\sum_{j=1}^{p} \tilde{w}_{ij}^k, \qquad k = 1, 2. \tag{S2.119}$$

By the Lindeberg-Feller CLT, we have

$$\sqrt{p}\begin{pmatrix} w_i^{(2)} - (\lambda+1) \\ w_i^{(1)} - 1 \end{pmatrix} \xrightarrow{D} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbb{E}\tilde{w}_{11}^4 - (\lambda+1)^2 & \mathbb{E}\tilde{w}_{11}^3 - \lambda - 1 \\ \mathbb{E}\tilde{w}_{11}^3 - \lambda - 1 & \lambda \end{pmatrix}\right).$$

By Taylor's theorem and Lindeberg's CLT, we have the following approximation:

$$\hat{\lambda} = \frac{1}{n_2}\sum_{i=1}^{n_2} \hat{\lambda}_i + O_P\left(\frac{1}{p^{3/2}}\right),$$

where

$$\hat{\lambda}_i := w_i^{[2]} - 2w_i^{[1]} + 1 - 2\Big(w_i^{[2]} - w_i^{[1]}\Big)\Big(w_i^{[1]} - 1\Big) + \Big(3w_i^{[2]} - 2w_i^{[1]}\Big)\Big(w_i^{[1]} - 1\Big)^2.$$

Define a function $f(x,y) = x - 2y + 1 - 2(x-y)(y-1) + (3x - 2y)(y-1)^2$. Then we have

$$\hat{\lambda}_i = f\Big(w_i^{[2]}, w_i^{[1]}\Big), \quad f(\lambda+1, 1) = \lambda, \quad \nabla f(\lambda+1, 1) = (1, -2\lambda - 2)'.$$

Note that

$$
\begin{aligned}
&\{\nabla f(\lambda+1,1)\}'\begin{pmatrix}\mathbb{E}\tilde{w}_{11}^4-(\lambda+1)^2 & \mathbb{E}\tilde{w}_{11}^3-\lambda-1\\ \mathbb{E}\tilde{w}_{11}^3-\lambda-1 & \lambda\end{pmatrix}\nabla f(\lambda+1,1)\\
&=\mathbb{E}\tilde{w}_{11}^4-4(\lambda+1)\mathbb{E}\tilde{w}_{11}^3+(\lambda+1)^2(4\lambda+3)\\
&=\mathbb{E}(\tilde{w}_{11}-1)^4-\lambda^2+h_2-2\lambda h_1.
\end{aligned}
$$

By the Delta method, we have

$$
\sqrt{p}\Big(\hat{\lambda}_i-\lambda\Big)\xrightarrow{D}\mathcal{N}\Big(0,\mathbb{E}(\tilde{w}_{11}-1)^4-\lambda^2+h_2-2\lambda h_1\Big). \tag{S2.120}
$$

Since

$$
p(\hat{\lambda}-\lambda)=\sqrt{\frac{p}{n_2}}\frac{1}{\sqrt{n_2}}\sum_{i=1}^{n_2}\sqrt{p}\Big(\hat{\lambda}_i-\lambda\Big)+o_P[1], \tag{S2.121}
$$

it is necessary to expand the expectation of $\hat{\lambda}_i$ up to the order $O(p^{-1})$. From direct calculations, we obtain that

$$
\begin{aligned}
&\mathbb{E}\tilde{w}_i^{[1]}=1,\quad \mathbb{E}\tilde{w}_i^{[2]}=\lambda+1,\\
&\mathbb{E}\{\tilde{w}_i^{[1]}\Big(\tilde{w}_i^{[1]}-1\Big)\}=\frac{\lambda}{p},\quad \mathbb{E}\{\tilde{w}_i^{[2]}\Big(\tilde{w}_i^{[1]}-1\Big)\}=\frac{\mathbb{E}\tilde{w}_{11}^3-\lambda-1}{p},\\
&\mathbb{E}\{\tilde{w}_i^{[1]}\Big(\tilde{w}_i^{[1]}-1\Big)^2\}=\frac{\lambda}{p}+O(p^{-2}),\quad \mathbb{E}\{\tilde{w}_i^{[2]}\Big(\tilde{w}_i^{[1]}-1\Big)^2\}=\frac{\lambda(\lambda+1)}{p}+O(p^{-2}),
\end{aligned}
$$

and thus,

$$
\mathbb{E}\hat{\lambda}_i=\lambda+\frac{-2\mathbb{E}\tilde{w}_{11}^3+3\lambda^2+5\lambda+2}{p}+o_P(p^{-1})=\lambda+\frac{h_1}{p}+o_P(p^{-1}). \tag{S2.122}
$$

From (S2.120), (S2.121), (S2.122) and the fact that $\{\hat{\lambda}_i\}_{i=1}^{n_2}$ are independent, we obtain (S2.118). This completes the proof of Theorem 3.2.

S2.17 Proof of Lemma S2.5

Similar to the derivation of equation (S2.24), we have

$$
\mathbb{E}\frac{w_1^2}{\overline{w}^2}-\mathbb{E}\frac{w_1^2}{\mu^2}=-\frac{2\mathbb{E}w_1^2(\overline{w}-\mu)}{\mu^3}+\frac{3\mathbb{E}w_1^2(\overline{w}-\mu)^2}{\mu^4}+\mathbb{E}a^\circ-\mathbb{E}b^\circ+\mathbb{E}c^\circ, \tag{S2.123}
$$

where $\overline{w}=p^{-1}\sum_{j=1}^p w_j$, $a^\circ:=\frac{w_1^2}{\mu^2}a\Big(\frac{\overline{w}-\mu}{\mu}\Big)I_{B_p(\varepsilon)}$ with $|a\Big(\frac{\overline{w}-\mu}{\mu}\Big)|\le C_1\Big|\frac{\overline{w}-\mu}{\mu}\Big|^3$, $b^\circ:=\{w_1^2/\mu^2-2w_1^2(\overline{w}-\mu)/\mu^3+3w_1^2(\overline{w}-\mu)^2/\mu^4\}I_{B_p^c(\varepsilon)}$, $c^\circ:=(w_1^2/\overline{w}^2)I_{B_p^c(\varepsilon)}$, $B_p(\varepsilon)=\{\omega:|\overline{w}-\mu|\le\varepsilon,\overline{w}=\sum_{j=1}^p w_j/p,\mu=\mathbb{E}w_1\}$, and the probability of its complement admits the following bound

$$
\mathbb{P}\Big(B_p^c(\varepsilon)\Big)\le C_t\varepsilon^{-t}p^{-t}\Big\{\Big(p\mathbb{E}|w_1-\mu|^2\Big)^{t/2}+p\mathbb{E}|w_1-\mu|^t\Big\}, \tag{S2.124}
$$

where ε, t, and C_t (which depends on t) are positive constants. By straightforward calculations, we have

$$\mathbb{E}|a^\circ| \le C\varepsilon\mathbb{E}|w_1^2(\overline{w}-\mu)^2 I_{B_p(\varepsilon)}| \le C\varepsilon\mathbb{E}|w_1^2(\overline{w}-\mu)^2| = \frac{C\varepsilon}{p^2}\sum_{i,j=1}^{p}\mathbb{E}w_1^2(w_i-\mu)(w_j-\mu)$$
$$= \frac{C\varepsilon}{p^2}\sum_{i=1}^{p}\mathbb{E}w_1^2(w_i-\mu)^2 = \frac{C\varepsilon}{p^2}\mathbb{E}w_1^2(w_1-\mu)^2 + \frac{C\varepsilon(p-1)}{p^2}\mathbb{E}w_1^2\mathbb{E}(w_1-\mu)^2 \le C\varepsilon p^{-1},$$
$$|\mathbb{E}b^\circ| \le C\left\{\mathbb{P}^{1/2}\left(B_p^c(\varepsilon)\right) + \frac{1}{p}\mathbb{P}^{1/4}\left(B_p^c(\varepsilon)\right) + \frac{1}{p^2}\right\},$$
$$|\mathbb{E}c^\circ| \le p^2\mathbb{P}\left(B_p^c(\varepsilon)\right),$$

where $C > 0$ is a constant. Given that $\mathbb{E}\left|w_1-\mu\right|^{6+s} < \infty$ for any $s > 0$, we use this moment condition to estimate $\mathbb{P}\left(B_p^c(\varepsilon)\right)$ by taking $t = 6+s$ in (S2.124). Consequently, we obtain $\mathbb{E}a^\circ - \mathbb{E}b^\circ + \mathbb{E}c^\circ = o(p^{-1})$. Combining this with equation (S2.123), we have

$$\mathbb{E}\left(\frac{w_1}{\overline{w}} - 1\right)^2 = \lambda + \frac{h_1}{p} + o(p^{-1}),$$

which corresponds to the result stated in Lemma S2.5.

S2.18 Proof of Example 1 in Section 3.2

Consider the following notations:

$$\mathbf{W} = (w_{ij}), \quad \mathbf{X} = (x_{ij}), \quad x_{ij} = \frac{w_{ij}}{\sum_{k=1}^{p} w_{ik}}, \quad \mathbf{V} = (v_{ij}), \quad \mathbf{Y} = (y_{ij}), \quad y_{ij} = \frac{v_{ij}}{\sum_{k=1}^{p} v_{ik}},$$
$$\tilde{x}_{ij} := x_{ij} - \mathbb{E}x_{ij}, \quad \tilde{y}_{ij} := y_{ij} - \mathbb{E}y_{ij}.$$

We assume that $\mathbf{V} = \mathbf{W}\Sigma$, where Σ is positive definite and normalized by $\mathrm{tr}(\Sigma) = p$. This normalization is without loss of generality, as $\mathbf{Y}$ is invariant under scaling of the basis data $\mathbf{V}$. The matrices $\mathbf{W}$ and $\mathbf{X}$ are for the null hypothesis, and $\mathbf{V}$ and $\mathbf{Y}$ are for the alternative hypothesis. Define the rescaled sample covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$ as follows:

$$\mathbf{B}_0 = \frac{p^2}{n}\mathbf{X}'\mathbf{C}_n\mathbf{X}, \qquad \mathbf{B}_1 = \frac{p^2}{n}\mathbf{Y}'\mathbf{C}_n\mathbf{Y}, \qquad \mathbf{C}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n',$$

The goal is to show that

$$\mathbb{E}\|\mathbf{B}_0 - \mathbf{B}\|_F^2 = \frac{p^2}{n} + o(p), \qquad \mathbb{E}\|\mathbf{B}_1 - \mathbf{B}\|_F^2 = \frac{(1+\alpha^2)^2}{(1+\alpha)^4}\frac{p^2}{n} + \frac{6\alpha^2}{(1+\alpha)^4}\frac{p^3}{(p+1)^2} + o(p). \tag{S2.125}$$

Recall the assumption

$$\{w_{ij}\} \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(1), \qquad \Sigma = \begin{pmatrix} 1 & & & \alpha \\ \alpha & 1 & & \\ & \ddots & \ddots & \\ & & \alpha & 1 \end{pmatrix}, \qquad \alpha > -1. \tag{S2.126}$$

From this assumption, we have

$$(x_{11}, x_{12}, \ldots, x_{1p}) \sim \text{Dirichlet}(1, 1, \ldots, 1),$$
$$y_{ij} = \frac{x_{ij}}{1+\alpha} + \frac{\alpha x_{i,j+1}}{1+\alpha} \quad (x_{i,p+1} \equiv x_{i1}).$$

We summarize several moments of the Dirichlet distribution. All results below are derived from following Dirichlet moment identity

$$\mathbb{E}\left(\prod_{j=1}^{p} x_{1j}^{k_j}\right) = \frac{\Gamma(p)\prod_{j=1}^{p} k_j!}{\Gamma(p + \sum_{j=1}^{p} k_j)}, \quad k_j \in \{0, 1, 2, \ldots\}.$$

Using this identity and direct calculation, we obtain the following results:

$$\mathbb{E}x_{11} = \mathbb{E}y_{11} = \frac{1}{p},$$
$$\mathbb{E}\tilde{x}_{11}^2 = \mathrm{Var}(\tilde{x}_{11}) = \frac{p-1}{p^2(p+1)}, \quad \mathbb{E}\tilde{x}_{11}\tilde{x}_{12} = \mathrm{Cov}(\tilde{x}_{11}, \tilde{x}_{12}) = \frac{-1}{p^2(p+1)}, \tag{S2.127}$$
$$\mathbb{E}\tilde{x}_{11}^2\tilde{x}_{12}^2 = O(p^{-4}), \quad \mathrm{Var}(\tilde{x}_{11}^2) = O(p^{-4}), \quad \mathrm{Var}(\tilde{x}_{11}\tilde{x}_{12}) = \frac{1+o(1)}{p^4}, \tag{S2.128}$$

and similarly for the $\tilde{y}_{ij}$-terms:

$$\mathbb{E}\tilde{y}_{11}^2 = \mathrm{Var}(\tilde{y}_{11}) = \frac{p(1+\alpha^2) - (1+\alpha)^2}{(1+\alpha)^2 p^2(p+1)}, \quad \mathbb{E}\tilde{y}_{11}\tilde{y}_{12} = O(p^{-3}), \quad \mathbb{E}\tilde{y}_{11}^2\tilde{y}_{12}^2 = O(p^{-4}), \tag{S2.129}$$
$$\mathrm{Cov}(\tilde{y}_{11}, \tilde{y}_{12}) = \frac{p\alpha - (1+\alpha)^2}{(1+\alpha)^2 p^2(p+1)}, \quad \mathrm{Cov}(\tilde{y}_{11}, \tilde{y}_{13}) = \frac{-1}{p^2(p+1)}, \tag{S2.130}$$
$$\mathrm{Var}(\tilde{y}_{11}^2) = O(p^{-4}), \quad \mathrm{Var}(\tilde{y}_{11}\tilde{y}_{12}) = O(p^{-4}), \quad \mathrm{Var}(\tilde{y}_{11}\tilde{y}_{13}) = \frac{(1+\alpha^2)^2\{1+o(1)\}}{(1+\alpha)^4 p^4}. \tag{S2.131}$$

Proof Sketch of (S2.125): To simplify the proof, we first show that

$$\mathbb{E}\|\mathbf{B}_\ell - \mathbf{B}\|_F^2 = \mathbb{E}\|\widetilde{\mathbf{B}}_\ell - \mathbf{B}\|_F^2\{1+o(1)\}, \quad \ell = 0, 1, \tag{S2.132}$$

where

$$\widetilde{\mathbf{B}}_0 = \frac{p^2}{n}(\mathbf{X} - \mathbb{E}\mathbf{X})'(\mathbf{X} - \mathbb{E}\mathbf{X}), \qquad \widetilde{\mathbf{B}}_1 = \frac{p^2}{n}(\mathbf{Y} - \mathbb{E}\mathbf{Y})'(\mathbf{Y} - \mathbb{E}\mathbf{Y}).$$

Based on this result, we can prove (S2.125) with $\mathbf{B}_\ell$ replaced by $\widetilde{\mathbf{B}}_\ell$. Specifically, we will show that

$$\mathbb{E}\|\widetilde{\mathbf{B}}_0 - \mathbf{B}\|_F^2 = \frac{p^2}{n} + o(p), \tag{S2.133}$$
$$\mathbb{E}\|\widetilde{\mathbf{B}}_1 - \mathbf{B}\|_F^2 = \frac{(1+\alpha^2)^2}{(1+\alpha)^4}\frac{p^2}{n} + \frac{6\alpha^2}{(1+\alpha)^4}\frac{p^3}{(p+1)^2} + o(p). \tag{S2.134}$$

Now, we prove (S2.132), (S2.133), and (S2.134) as follows:

- **Proof of (S2.132)**: We first show the proof for $\ell = 0$. We begin with the following identity:

$$\mathbb{E}\|\mathbf{B}_0 - \mathbf{B}\|_F^2 = \mathbb{E}\|\mathbf{B}_0 - \widetilde{\mathbf{B}}_0\|_F^2 + \mathbb{E}\|\widetilde{\mathbf{B}}_0 - \mathbf{B}\|_F^2 + 2\mathbb{E}\mathrm{tr}\{(\mathbf{B}_0 - \widetilde{\mathbf{B}}_0)(\widetilde{\mathbf{B}}_0 - \mathbf{B})'\}.$$

Let $\overline{x}_i = \frac{1}{n}\sum_{k=1}^n x_{ki}$. From the identity

$$\frac{1}{n}\sum_{k=1}^n (x_{ki}-\overline{x}_i)(x_{kj}-\overline{x}_j) = \frac{1}{n}\sum_{k=1}^n \tilde{x}_{ki}\tilde{x}_{kj} - \frac{1}{n^2}\Big(\sum_{k=1}^n \tilde{x}_{ki}\Big)\Big(\sum_{k=1}^n \tilde{x}_{kj}\Big),$$

we have

$$\begin{aligned}
&\mathbb{E}\|\mathbf{B}_0 - \widetilde{\mathbf{B}}_0\|_F^2 \\
&= \frac{p^4}{n^4}\sum_{i,j}\mathbb{E}\Big\{\Big(\sum_{k=1}^n \tilde{x}_{ki}\Big)^2\Big(\sum_{k=1}^n \tilde{x}_{kj}\Big)^2\Big\} \\
&= \frac{p^5}{n^4}\mathbb{E}\Big(\sum_{k=1}^n \tilde{x}_{k1}\Big)^4 + \frac{p^4(p^2-p)}{n^4}\mathbb{E}\Big\{\Big(\sum_{k=1}^n \tilde{x}_{k1}\Big)^2\Big(\sum_{k=1}^n \tilde{x}_{k2}\Big)^2\Big\} \\
&= \frac{p^5}{n^4}\Big\{n\mathbb{E}\tilde{x}_{11}^4 + 3n(n-1)\{\mathbb{E}(\tilde{x}_{11}^2)\}^2\Big\} \\
&\qquad + \frac{p^4(p^2-p)}{n^4}\Big\{n\mathbb{E}\tilde{x}_{11}^2\tilde{x}_{12}^2 + n(n-1)\{\mathbb{E}\tilde{x}_{11}^2\}^2 + n(n-1)(\mathbb{E}\tilde{x}_{11}\tilde{x}_{12})^2\Big\} \\
&= O(p^2/n^2),
\end{aligned} \tag{S2.135}$$

where in the last equality we use (S2.127) and (S2.128). Moreover, we have

$$\begin{aligned}
&\mathbb{E}\mathrm{tr}\{(\mathbf{B}_0-\widetilde{\mathbf{B}}_0)(\widetilde{\mathbf{B}}_0-\mathbf{B})'\} = \sum_{i,j}\mathbb{E}\{(\mathbf{B}_1-\widetilde{\mathbf{B}}_0)_{ij}(\widetilde{\mathbf{B}}_0-\mathbf{B})_{ij}\} \\
&= \frac{p^4}{n^2}\sum_{i,j}\mathbb{E}\Big\{\Big(\sum_{k=1}^n \tilde{x}_{ki}\Big)\Big(\sum_{k=1}^n \tilde{x}_{kj}\Big)\frac{1}{n}\sum_{k=1}^n\Big(\tilde{x}_{ki}\tilde{x}_{kj} - p^{-2}B_{ij}\Big)\Big\} \\
&= \frac{p^4}{n^3}\sum_{i,j}\sum_{k=1}^n\Big(\mathbb{E}\tilde{x}_{ki}^2\tilde{x}_{kj}^2 - p^{-2}B_{ij}\mathbb{E}\tilde{x}_{ki}\tilde{x}_{kj}\Big) \\
&= O(p^2/n^2).
\end{aligned}$$

The above estimates and (S2.133) yield (S2.132) for $\ell = 0$.

For the case $\ell = 1$, we use the similar argument with the help of (S2.129) and (S2.130).

- **Proof of** (S2.133): Let $\widetilde{B}_{0,ij}$ be the (i,j)-th entry of $\widetilde{\mathbf{B}}_0$.

$$\begin{aligned}
\mathbb{E}\|\widetilde{\mathbf{B}}_0 - \mathbf{B}\|_F^2 &= \sum_{i,j}\mathrm{Var}(\widetilde{B}_{0,ij}) = \frac{p^4}{n^2}\sum_{i,j}\mathrm{Var}\Big(\sum_{k=1}^n \tilde{x}_{ki}\tilde{x}_{kj}\Big) \\
&= \frac{p^5}{n}\mathrm{Var}(\tilde{x}_{11}^2) + \frac{p^4(p^2-p)}{n}\mathrm{Var}(\tilde{x}_{11}\tilde{x}_{12}) \\
&\overset{\text{(S2.128)}}{=} \frac{p^2\{1+o(1)\}}{n}.
\end{aligned}$$

- **Proof of** (S2.134): Note that

$$\mathbb{E}\|\widetilde{\mathbf{B}}_1 - \mathbf{B}\|_F^2 = \mathbb{E}\|\widetilde{\mathbf{B}}_1 - \mathbb{E}\widetilde{\mathbf{B}}_1\|_F^2 + \|\mathbb{E}\widetilde{\mathbf{B}}_1 - \mathbf{B}\|_F^2.$$

Thus, we need to estimate two terms on the RHS.

Let $\widetilde{B}_{1,ij}$ be the (i,j)-th entry of $\widetilde{\mathbf{B}}_1$. Then, we have

$$\begin{aligned}\mathbb{E}\|\widetilde{\mathbf{B}}_1 - \mathbb{E}\widetilde{\mathbf{B}}_1\|_F^2 &= \sum_{i,j} \mathrm{Var}(\widetilde{B}_{1,ij}) = \frac{p^4}{n^2}\sum_{i,j}\mathrm{Var}\Big(\sum_{k=1}^{n} \tilde{y}_{ki}\tilde{y}_{kj}\Big) \\ &= \frac{p^5}{n}\mathrm{Var}(\tilde{y}_{11}^2) + \frac{2p^5}{n}\mathrm{Var}(\tilde{y}_{11}\tilde{y}_{12}) + \frac{p^4(p^2-3p)}{n}\mathrm{Var}(\tilde{y}_{11}\tilde{y}_{13}) \\ &\overset{(S2.131)}{=} \frac{(1+\alpha^2)^2}{(1+\alpha)^4}\frac{p^2\{1+o(1)\}}{n}.\end{aligned}$$

By using (S2.127), (S2.129), and (S2.130), we obtain

$$\mathbf{B} = \mathbb{E}\widetilde{\mathbf{B}}_0 = \begin{pmatrix} d_0 & b & b & \cdots & b & b \\ b & d_0 & b & \ddots & & b \\ b & b & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & b \\ b & & \ddots & \ddots & \ddots & b \\ b & b & \cdots & b & b & d_0 \end{pmatrix}, \quad \mathbb{E}\widetilde{\mathbf{B}}_1 = \begin{pmatrix} d_1 & a & b & \cdots & b & a \\ a & d_1 & a & \ddots & & b \\ b & a & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & b \\ b & & \ddots & \ddots & \ddots & a \\ a & b & \cdots & b & a & d_1 \end{pmatrix},$$

where

$$d_0 = \frac{p-1}{p+1}, \quad b = \frac{-1}{p+1}, \quad d_1 = \frac{p(1+\alpha^2)-(1+\alpha)^2}{(1+\alpha)^2(p+1)}, \quad a = \frac{p\alpha-(1+\alpha)^2}{(1+\alpha)^2(p+1)}.$$

Then, we have

$$\|\mathbb{E}\widetilde{\mathbf{B}}_1 - \mathbf{B}\|_F^2 = p(d_1-d_0)^2 + 2p(a-b)^2 = \frac{6\alpha^2}{(1+\alpha)^4}\frac{p^3}{(p+1)^2}.$$

Combining these results, we obtain (S2.134).

S2.19 Proof of power analysis in Section 3.2

The goal is to show that

$$\mathbb{E}\|\mathbf{B}_0 - \mathbf{B}\|_F^2 = \frac{p^2\lambda^2}{n} + o(p), \qquad \mathbb{E}\|\mathbf{B}_1 - \mathbf{B}\|_F^2 = \frac{\lambda^2(\sum_{k=1}^{p} d_k^2)^2}{n} + O(p). \tag{S2.136}$$

We first list some moments (will be proven later) of $\tilde{y}_{ij}$ as follows:

$$\mathbb{E}\tilde{y}_{11}\tilde{y}_{12} = O(p^{-2}), \quad \mathbb{E}\tilde{y}_{11}^2 = O(p^{-2}), \quad \mathbb{E}\tilde{y}_{11}^2\tilde{y}_{12}^2 = O(p^{-4}), \quad \mathbb{E}\tilde{y}_{11}^4 = O(p^{-4}), \tag{S2.137}$$

$$\mathrm{Var}(\tilde{y}_{1i}^2) = \frac{d_i^4(\kappa_4+2\sigma^4)}{p^4\mu^4}\Big\{1+O_P(p^{-1/2})\Big\}, \quad \mathrm{Var}(\tilde{y}_{1i}\tilde{y}_{1j}) = \frac{d_i^2 d_j^2\sigma^4}{p^4\mu^4}\Big\{1+O_P(p^{-1/2})\Big\}, \tag{S2.138}$$

where $\kappa_4 := \mathbb{E}(w_{11}-\mu)^4 - 3\sigma^4$ denotes the fourth cumulant of w_{11}.

The idea of the proof of (S2.136) is similar to that of (S2.125) in the previous section. It suffices to show the following three results:

$$\mathbb{E}\|\mathbf{B}_\ell - \mathbf{B}\|_F^2 = \mathbb{E}\|\widetilde{\mathbf{B}}_\ell - \mathbf{B}\|_F^2\{1+o(1)\}, \quad \ell = 0, 1, \tag{S2.139}$$

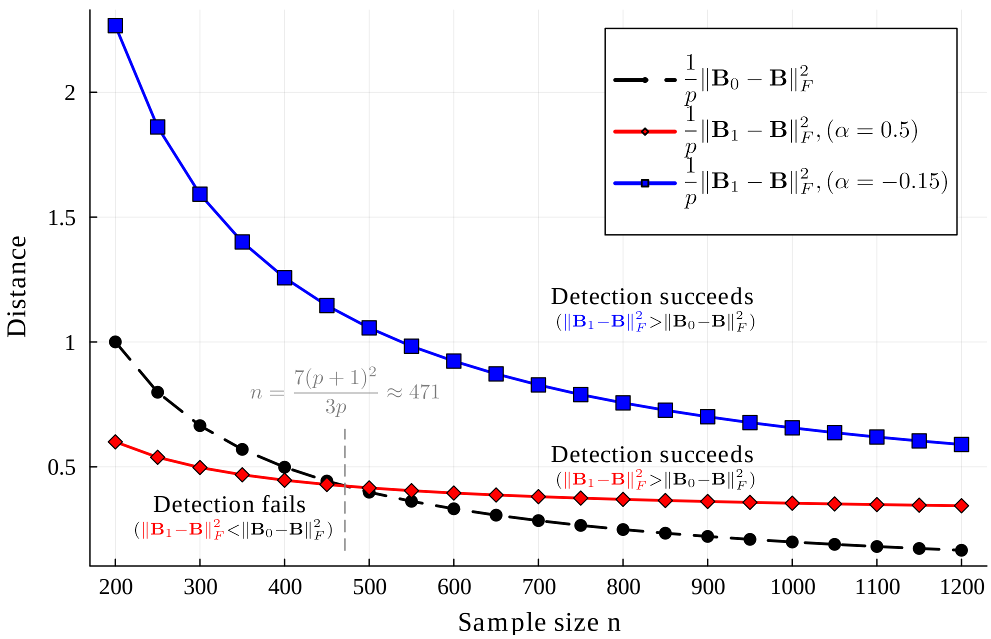


Figure S.5: Numerical simulation for non-diagonal alternative. Comparison of normalized Frobenius norm distances $\frac{1}{p}\|\mathbf{B}_0 - \mathbf{B}\|_F^2$ and $\frac{1}{p}\|\mathbf{B}_1 - \mathbf{B}\|_F^2$ against sample size n under the setting (S2.126), with fixed $p = 200$. Results are averaged over 1000 independent replications. For $\alpha = 0.5$, the vertical dashed line marks the critical threshold $n = \frac{7(p+1)^2}{3p}$, below which the proposed test cannot effectively detect such non-diagonal covariance structure. For $\alpha = -0.15$, the proposed test can detect such structure for any sample size n.

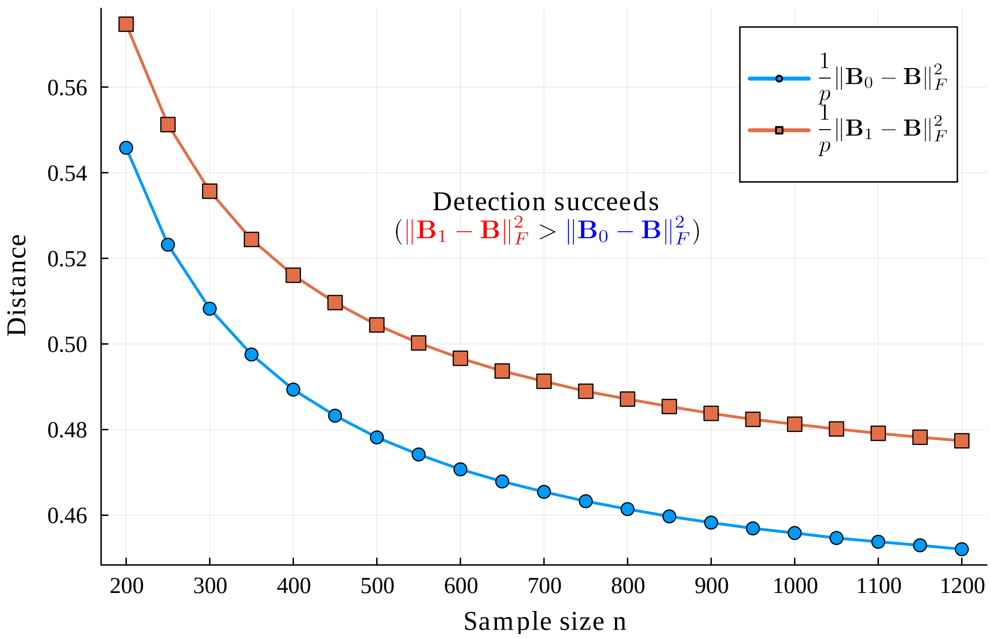


Figure S.6: Numerical simulation for diagonal alternative. Comparison of normalized Frobenius norm distances $\frac{1}{p}\|\mathbf{B}_0 - \mathbf{B}\|_F^2$ and $\frac{1}{p}\|\mathbf{B}_1 - \mathbf{B}\|_F^2$ against sample size n, with fixed $p = 200$. The diagonal matrix $\Sigma = \mathrm{diag}(3, 1, \ldots, 1)$ represents a single spike alternative. The basis data $\{w_{ij}\} \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(1)$ follow an exponential distribution. Results are averaged over 1000 independent replications. As expected, $\|\mathbf{B}_1 - \mathbf{B}\|_F^2 > \|\mathbf{B}_0 - \mathbf{B}\|_F^2$ across all sample sizes, confirming the test's ability to detect diagonal alternatives.

$$\mathbb{E}\|\widetilde{\mathbf{B}}_0 - \mathbf{B}\|_F^2 = \frac{\lambda^2 p^2}{n} + o(p), \tag{S2.140}$$

$$\mathbb{E}\|\widetilde{\mathbf{B}}_1 - \mathbf{B}\|_F^2 = \frac{\lambda^2 (\sum_{k=1}^p d_k^2)^2}{n} + O(p), \tag{S2.141}$$

where $\mathbf{B} = \frac{p\nu_2}{p-1}\mathbf{G}_p$.

Now, we prove (S2.137), (S2.138), (S2.139), (S2.140), and (S2.141) as follows:

- **Proof of** (S2.137) **and** (S2.138): Let

$$S = \sum_{i=1}^p d_i w_{1i}, \qquad \delta_i = w_{1i} - \mu.$$

Because $d_k's$ are of constant order, and $\sum_{i=1}^p d_i = p$, we have

$$\varepsilon := \frac{S - p\mu}{p\mu} = \frac{1}{p\mu}\sum_{i=1}^p d_i \delta_i = O_P(p^{-1/2}).$$

Hence, for each i,

$$y_{1i} = \frac{d_i w_{1i}}{S} = \frac{d_i w_{1i}}{p\mu}(1+\varepsilon)^{-1} = \frac{d_i w_{1i}}{p\mu}\Big(1 - \varepsilon + \varepsilon^2 - \cdots\Big), \qquad \tilde{y}_{1i} = \frac{d_i \delta_i}{p\mu} + O_P(p^{-3/2}).$$

Moreover, we have

$$\tilde{y}_{1i}^2 = \frac{d_i^2 \delta_i^2}{p^2\mu^2} + O_P(p^{-5/2}), \quad \tilde{y}_{1i}^4 = \frac{d_i^4 \delta_i^4}{p^4\mu^4} + O_P(p^{-9/2}),$$

and thus

$$\mathrm{Var}(\tilde{y}_{1i}^2) = \frac{d_i^4(\kappa_4 + 2\sigma^4)}{p^4\mu^4}\Big\{1 + O_P(p^{-1/2})\Big\},$$

where $\kappa_4 := \mathbb{E}(w_{11} - \mu)^4 - 3\sigma^4$ denotes the fourth cumulant of w_{11}. Similarly, for $i \neq j$, we have

$$\mathrm{Var}(\tilde{y}_{1i}\tilde{y}_{1j}) = \frac{d_i^2 d_j^2 \sigma^4}{p^4\mu^4}\Big\{1 + O_P(p^{-1/2})\Big\}.$$

- **Proof of** (S2.139): We only show the proof for $\ell = 1$, since the case $\ell = 0$ is a special case of $\ell = 1$ by taking $d_1 = \cdots = d_p = 1$.

We begin with the following identity:

$$\mathbb{E}\|\mathbf{B}_1 - \mathbf{B}\|_F^2 = \mathbb{E}\|\mathbf{B}_1 - \widetilde{\mathbf{B}}_1\|_F^2 + \mathbb{E}\|\widetilde{\mathbf{B}}_1 - \mathbf{B}\|_F^2 + 2\mathbb{E}\mathrm{tr}\{(\mathbf{B}_1 - \widetilde{\mathbf{B}}_1)(\widetilde{\mathbf{B}}_1 - \mathbf{B})'\}.$$

Let $\overline{y}_i = \frac{1}{n}\sum_{k=1}^n y_{ki}$. From the identity

$$\frac{1}{n}\sum_{k=1}^n (y_{ki} - \overline{y}_i)(y_{kj} - \overline{y}_j) = \frac{1}{n}\sum_{k=1}^n \tilde{y}_{ki}\tilde{y}_{kj} - \frac{1}{n^2}\Big(\sum_{k=1}^n \tilde{y}_{ki}\Big)\Big(\sum_{k=1}^n \tilde{y}_{kj}\Big),$$

we have

$$\mathbb{E}\|\mathbf{B}_1 - \widetilde{\mathbf{B}}_1\|_F^2$$

$$
\begin{aligned}
&= \frac{p^4}{n^4}\sum_{i,j}\mathbb{E}\bigg\{\Big(\sum_{k=1}^n \tilde{y}_{ki}\Big)^2\Big(\sum_{k=1}^n \tilde{y}_{kj}\Big)^2\bigg\},\\
&= \frac{p^4}{n^4}\sum_{i=1}^p\mathbb{E}\Big(\sum_{k=1}^n \tilde{y}_{ki}\Big)^4 + \frac{p^4}{n^4}\sum_{i\neq j}\mathbb{E}\bigg\{\Big(\sum_{k=1}^n \tilde{y}_{ki}\Big)^2\Big(\sum_{k=1}^n \tilde{y}_{kj}\Big)^2\bigg\}\\
&= \frac{p^4}{n^4}\sum_{i=1}^p\Big\{n\mathbb{E}\tilde{y}_{1i}^4 + 3n(n-1)\mathbb{E}\tilde{y}_{1i}^2\tilde{y}_{2i}^2\Big\}\\
&\qquad + \frac{p^4}{n^4}\sum_{i\neq j}\Big\{n\mathbb{E}\tilde{y}_{1i}^2\tilde{y}_{1j}^2 + (n^2-n)\mathbb{E}\tilde{y}_{1i}^2\mathbb{E}\tilde{y}_{1j}^2 + (n^2-n)(\mathbb{E}\tilde{y}_{1i}\tilde{y}_1)^2\Big\}\\
&= O(p^2/n^2),
\end{aligned}
$$

where in the last equality we use (S2.127) and (S2.128). Moreover,

$$
\mathbb{E}\mathrm{tr}\{(\mathbf{B}_1-\widetilde{\mathbf{B}}_1)(\widetilde{\mathbf{B}}_1-\mathbf{B})'\} = \sum_{i,j}\mathbb{E}\{(\mathbf{B}_1-\widetilde{\mathbf{B}}_1)_{ij}(\widetilde{\mathbf{B}}_1-\mathbf{B})_{ij}\}
$$

$$
\begin{aligned}
&= \frac{p^4}{n^2}\sum_{i,j}\mathbb{E}\bigg\{\Big(\sum_{k=1}^n \tilde{y}_{ki}\Big)\Big(\sum_{k=1}^n \tilde{y}_{kj}\Big)\frac{1}{n}\sum_{k=1}^n\Big(\tilde{y}_{ki}\tilde{y}_{kj} - p^{-2}B_{ij}\Big)\bigg\}\\
&= \frac{p^4}{n^3}\sum_{i,j}\bigg[\sum_{k=1}^n\Big(\mathbb{E}\tilde{y}_{ki}^2\tilde{y}_{kj}^2 - p^{-2}B_{ij}\mathbb{E}\tilde{y}_{ki}\tilde{y}_{kj}\Big) + \sum_{k\neq\ell}\mathbb{E}\Big\{\tilde{y}_{ki}\tilde{y}_{kj}\Big(\tilde{y}_{\ell i}\tilde{y}_{\ell j} - p^{-2}B_{ij}\Big)\Big\}\bigg]\\
&= O(p^2/n^2).
\end{aligned}
$$

The above estimates and imply (S2.139) for $\ell = 1$.

- **Proof of** (S2.140) **and** (S2.141): By using (S2.138), we have

$$
\begin{aligned}
\mathbb{E}\|\widetilde{\mathbf{B}}_1 - \mathbb{E}\widetilde{\mathbf{B}}_1\|_F^2 &= \sum_{i,j}\mathrm{Var}(B_{1,ij}) = \frac{p^4}{n^2}\sum_{i,j}\mathrm{Var}\Big(\sum_{k=1}^n \tilde{y}_{ki}\tilde{y}_{kj}\Big)\\
&= \frac{p^4}{n}\bigg\{\sum_{i=1}^p\mathrm{Var}(\tilde{y}_{1i}^2) + \sum_{i\neq j}\mathrm{Var}(\tilde{y}_{1i}\tilde{y}_{1j})\bigg\}\\
&= \frac{1}{n}\bigg\{\frac{\kappa_4+2\sigma^4}{\mu^4}\sum_{i=1}^p d_i^4 + \frac{\sigma^4}{\mu^4}\sum_{i\neq j}d_i^2d_j^2\bigg\}\Big\{1+O_P(p^{-1/2})\Big\}\\
&= \frac{1}{n}\bigg\{\lambda^2\Big(\sum_{i=1}^p d_i^2\Big)^2 + \frac{\kappa_4+\sigma^4}{\mu^4}\sum_{i=1}^p d_i^4\bigg\}\Big\{1+O_P(p^{-1/2})\Big\}\\
&= \frac{\lambda^2(\sum_{i=1}^p d_i^2)^2\{1+o(1)\}}{n}.
\end{aligned}
$$

Hence, we have

$$
\mathbb{E}\|\widetilde{\mathbf{B}}_1 - \mathbf{B}\|_F^2 = \mathbb{E}\|\widetilde{\mathbf{B}}_1 - \mathbb{E}\widetilde{\mathbf{B}}_1\|_F^2 + \|\mathbb{E}\widetilde{\mathbf{B}}_1 - \mathbf{B}\|_F^2 = \frac{\lambda^2(\sum_{i=1}^p d_i^2)^2}{n} + O(p).
$$

Under the null case, we have $d_1 = \cdots = d_p = 1$, and thus

$$
\mathbb{E}\|\widetilde{\mathbf{B}}_0 - \mathbb{E}\widetilde{\mathbf{B}}_0\|_F^2 = \frac{\lambda^2p^2\{1+o(1)\}}{n}.
$$

The above estimates complete the proof of (S2.140) and (S2.141).

S3 Simulation of CLT for $M_p(z)$

In this section, we compare the empirical mean and covariance of $M_p(z) = \operatorname{tr}(\mathbf{B}_p^0 - z\mathbf{I}_p)^{-1} - pm_{F^{c_n}}(z)$ with their theoretical limits as stated in Proposition 5.1. This proposition is a key step for the proof of our main result, Theorem 2.5. Readers are referred to Section 5 for more details of $M_p(z)$. We consider two types of data distribution of w_{ij} as follows:

1. w_{ij} follows the exponential distribution with rate parameter 5;
2. w_{ij} follows the Chi-square distribution with degree of freedom 1.

Empirical values of $\mathbb{E}M_p(z)$ and $\operatorname{Cov}(M_p(z_1), M_p(z_2))$ are calculated for various combinations of (p,n) with $p/n = 3/4$ or $p/n = 5/4$. For each pair of (p,n), 2000 independent replications are used to obtain the empirical values. Table S.5 reports the empirical mean of $M_p(z)$ with $z = \pm 3 + 2i$ for both Exp(5) population and $\chi^2(1)$ population. The empirical results of $\operatorname{Cov}(M_p(z_1), M_p(z_2))$ are reported in Table S.6. As shown in Tables S.5 – S.6, the empirical values of $\mathbb{E}M_p(z)$ and $\operatorname{Cov}(M_p(z_1), M_p(z_2))$ closely match their respective theoretical limits under all scenarios.

Table S.5: Empirical mean of $M_p(z)$ with $z = \mp 3 + 2i$.

			Exp(5)		$\chi^2(1)$	
	p/n	n	-3+2i	3+2i	-3+2i	3+2i
Emp	3/4	100	0.0586+0.0857i	-0.0373-0.249i	0.1405+0.1628i	-0.55-0.2732i
		200	0.0582+0.0858i	-0.0311-0.2526i	0.1459+0.1697i	-0.5761-0.3089i
		300	0.0567+0.0844i	-0.0336-0.2566i	0.1465+0.1712i	-0.5705-0.3212i
		400	0.0596+0.0878i	-0.0352-0.2528i	0.1463+0.172i	-0.5631-0.3465i
Theo			**0.0587+0.0872i**	**-0.029-0.2529i**	**0.15+0.1768i**	**-0.5792-0.3764i**
Emp	5/4	100	0.0547+0.0766i	-0.1069-0.2671i	0.1366+0.1473i	-0.5458-0.1545i
		200	0.0572+0.0793i	-0.1109-0.2757i	0.1395+0.1518i	-0.5847-0.1787i
		300	0.0587+0.0808i	-0.1074-0.2752i	0.1382+0.1511i	-0.5747-0.1934i
		400	0.0559+0.0778i	-0.0949-0.2733i	0.1434+0.1553i	-0.5751-0.1933i
Theo			**0.0578+0.0804i**	**-0.0919-0.2764i**	**0.1432+0.1569i**	**-0.6025-0.2149i**

Table S.6: Empirical covariance between $M_p(z_1)$ and $M_p(z_2)$.

	p/n	n	Exp(5)		$\chi^2(1)$	
			(-3+2i,-1+1i)†	(3+2i,5+1i)	(-3+2i,-1+1i)	(3+2i,5+1i)
Emp	3/4	100	-0.0038+0.0147i	-0.04+0.0035i	0+0.0304i	0.089+0.014i
		200	-0.0041+0.0163i	-0.0418+0.0022i	0.0004+0.0326i	0.117+0.0284i
		300	-0.0043+0.0171i	-0.0446+0.0011i	0+0.0335i	0.1372+0.0294i
		400	-0.0043+0.0168i	-0.0465-0.0003i	0.0002+0.0356i	0.1273+0.036i
Theo			**-0.0044+0.0172**i	**-0.0448-0.0002**i	**0.0006+0.0363**i	**0.1491+0.0524**i
Emp	5/4	100	-0.0032+0.0197i	-0.0483+0.0765i	0.0025+0.0349i	0.0931-0.0373i
		200	-0.0032+0.0196i	-0.0545+0.0763i	0.0032+0.035i	0.0991-0.0406i
		300	-0.0036+0.0212i	-0.0566+0.0708i	0.0026+0.0336i	0.0955-0.0209i
		400	-0.0032+0.02i	-0.0594+0.0742i	0.0038+0.0374i	0.1138-0.0297i
Theo			**-0.0034+0.0206**i	**-0.0624+0.0743**i	**0.0035+0.0388**i	**0.1099-0.0323**i

†This row denotes different combinations of (z_1, z_2).