Journal of Multivariate Analysis 191 (2022) 105007

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

CLT for linear spectral statistics of high-dimensional sample covariance matrices in elliptical distributions

Yangchun Zhang [a], Jiang Hu [b], Weiming Li [c,*]

[a] Department of Mathematics, Shanghai University, Shanghai 200444, China
[b] KLASMOE and School of Mathematics & Statistics, Northeast Normal University, Changchun 130024, China
[c] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

ARTICLE INFO

Article history:
Received 16 September 2021
Received in revised form 1 April 2022
Accepted 2 April 2022
Available online 20 May 2022

AMS 2020 subject classifications:
primary 62H12
secondary 62F12

Keywords:
Confidence interval
Covariance matrix
Elliptical distribution
Gaussian scale mixture
High-dimensional data

ABSTRACT

In this paper, we establish a new central limit theorem for the linear spectral statistics of high-dimensional sample covariance matrices. The underlying population belongs to the family of elliptical distributions, and the dimension of the population is allowed to grow to infinity, in proportion to the sample size. As an application, we construct confidence intervals for the model parameters of a Gaussian scale mixture.

1. Introduction

The spectral statistics of sample covariance matrices and their limiting behaviors play important roles in multivariate analysis [1]. When the number of observations p (referred to as the population dimension) is nonnegligible with respect to the sample size n, the classic central limit theorems (CLTs) are not applicable to such statistics since these theorems are justified only under a fixed-dimensional framework. That is, the population dimension p is fixed, while the sample size n tends to infinity. To address this discrepancy, many studies have been conducted to investigate the limiting behaviors of spectral statistics under high-dimensional asymptotic frameworks, which allow p and n to increase to infinity simultaneously (see, for instance, [3,4,10,13,19,21,24,25,27,28]). Among these works, [4] established a general CLT for the linear spectral statistics (LSSs) of sample covariance matrices in an asymptotic regime where

$$\min\{p, n\} \to \infty, \quad c_n \triangleq p/n \to c \in (0, \infty), \tag{1}$$

also referred to as the Marčenko–Pastur (MP) asymptotic regime [17], which is a commonly used setting in random matrix theory. Several extensions of this CLT can be found in [20,27]. These results assume that the population $\mathbf{x}$ follows an independent components (IC) model, i.e.,

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z}, \tag{2}$$

* Corresponding author.
E-mail addresses: zycstatis@shu.edu.cn (Y. Zhang), huj156@nenu.edu.cn (J. Hu), li.weiming@shufe.edu.cn (W. Li).

https://doi.org/10.1016/j.jmva.2022.105007
0047-259X/

where $\boldsymbol{\mu} \in \mathbb{R}^p$ denotes the population mean vector, $\mathbf{A} \in \mathbb{R}^{p\times p}$ is a deterministic matrix, and $\mathbf{z} = (z_1, \ldots, z_p)' \in \mathbb{R}^p$ consists of independent and identically distributed (i.i.d.) random variables with zero mean, unit variance, and finite fourth moment. Although this model is highly general, it can describe only the linear dependence among the population components and thus is unsuitable for situations where the observations are nonlinearly dependent, such as applications in biology and finance. To overcome the above limitation, [10] proposed a CLT for the LSSs from a population belonging to the family of elliptical distributions [12], where the population $\mathbf{x}$ admits the following stochastic representation:

$$\mathbf{x} = \boldsymbol{\mu} + \xi \mathbf{A}\mathbf{u}. \tag{3}$$

In this model, $\boldsymbol{\mu}$ and $\mathbf{A}$ are the same as in the IC model (2), $\mathbf{u} \in \mathbb{R}^p$ is a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$, and $\xi \geq 0$ is a scalar random variable that is independent of $\mathbf{u}$ and is normalized by $\mathbb{E}(\xi^2) = p$ for the identifiability of the triple product $\xi\mathbf{A}\mathbf{u}$. By choosing a suitable distribution for ξ, this model can fit various sample sets that exhibit light or heavy tails and are tail-dependent (see [8,9]). However, in the work of [10], the distribution of ξ is restricted to the following moment condition for some $\tau > 0$:

$$\frac{\mathbb{E}\xi^4}{(\mathbb{E}\xi^2)^2} = 1 + \frac{\tau}{p} + o\left(\frac{1}{p}\right). \tag{4}$$

This condition implies that $\xi/\sqrt{p}$ should converge in probability to the constant 1 as $p \to \infty$. Thus, their model cannot cover all elliptical distributions, resulting in the exclusion of some important distributions, such as the multivariate t-distribution and Gaussian scale mixtures. Taking the scale mixture as an example, the variable ξ in (3) can be written as the product of two independent random variables, i.e., $\xi = \eta\kappa$, where κ^2 follows a chi-squared distribution with p degrees of freedom and η follows a discrete distribution with a finite number of positive mass points. It almost surely holds that $\xi/\sqrt{p} \to \eta$ as $p \to \infty$. Therefore, it is worth noting that $\xi/\sqrt{p}$ does not degenerate asymptotically. Indeed, several studies have been published on this topic (see [16] for scale mixtures with spherical covariance matrices and [11] for complex-valued elliptical distributions). However, their results cannot cover the real-valued elliptical model.

In this paper, we investigate the fluctuations of the LSSs under the elliptical model (3) when $\xi/\sqrt{p}$ converges in distribution to a nondegenerate limit. A new CLT is established for the LSSs in the MP asymptotic regime, and we demonstrate that the convergence rate of these LSSs under our model is $O(1/\sqrt{p})$, which is much slower than the rate of $O(1/p)$ shown in [10]. Thus, as the fluctuation of $\xi/\sqrt{p}$ increases, the convergence rate of the LSSs exhibits a phase transition under the real-valued elliptical model. As an application, we consider a Gaussian scale mixture and construct confidence intervals for the model parameters by using our new CLT through the moment method.

The remainder of this paper is organized as follows. Section 2 details our model assumptions and presents our CLT for the LSSs under the real-valued elliptical model. Section 3 develops the confidence intervals for the model parameters of a Gaussian scale mixture. Furthermore, technical proofs are presented in Section 4. Some supporting lemmas are deferred to the Appendix.

2. CLT for the LSSs of sample covariance matrices in elliptical distributions

2.1. Preliminary

For any $p \times p$ real symmetric matrix $\mathbf{A}$ with eigenvalues $\{\tau_i\},\ i \in \{1, \ldots, p\}$, its empirical spectral distribution (ESD) is by definition the random measure

$$F^{\mathbf{A}} = \frac{1}{p}\sum_{i=1}^{p}\delta_{\tau_i},$$

where δ_a denotes the Dirac mass at point a. If ESD $F^{\mathbf{A}}$ has a limit as $p \to \infty$, this limit is referred to as the limiting spectral distribution (LSD). Given a real symmetric random matrix $\mathbf{A}$ and a function f supported on the real line, we call

$$\int f(x)dF^{\mathbf{A}}(x)$$

an LSS of the matrix $\mathbf{A}$ associated with the function f. For any probability measure G supported on the real line, its Stieltjes transform is defined as

$$m_G(z) = \int \frac{1}{x - z}dG(x), \quad z \in \mathbb{C}^+,$$

where $\mathbb{C}^+ \triangleq \{u + \mathrm{i}v : u, v \in \mathbb{R}, v > 0\}$ denotes the upper complex plane. This definition can be extended to the whole complex plane excluding the support of G.

2.2. Main results

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a sequence of i.i.d. random vectors from the elliptical distribution $\mathbf{x} \in \mathbb{R}^p$, defined in (3) with zero mean vector, i.e.,

$$\mathbf{x}_j = \xi_j \mathbf{A}\mathbf{u}_j,$$

for $j \in \{1, \ldots, n\}$. The population covariance matrix is denoted by $\boldsymbol{\Sigma} \triangleq \mathrm{Cov}(\mathbf{x}) = \mathbf{A}\mathbf{A}^\top$, and the sample covariance matrix is

$$\mathbf{B}_n = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^\top .$$

Our main results regarding the eigenvalues of $\mathbf{B}_n$ are based on the following assumptions:

A1 The population dimension p and the sample size n both tend to infinity such that $c_n = p/n \to c \in (0, \infty)$;
A2 The ESD $H_p \triangleq F^{\boldsymbol{\Sigma}}$ of the covariance matrix $\boldsymbol{\Sigma}$ weakly converges to a probability distribution H;
A3 The distribution M_p of ξ^2/p weakly converges to a nondegenerate probability distribution $\tilde{H}$, referred to as the limiting mixing distribution (LMD);
A4 The spectral norm of $\boldsymbol{\Sigma}$ is bounded in p;
A5 The support S_{M_p} of M_p is bounded in p, that is, $\cup_{p=1}^{\infty} S_{M_p} \subset [a, b]$ for some $a, b \in [0, \infty)$.

Assumptions A1, A2 and A4 are standard conditions for studying the spectral behaviors of a certain random matrix (see [4,17,22]). Assumption A3 is made for our elliptical model (3) to guarantee that $\xi/\sqrt{p}$ has a nondegenerate limit. Assumption A5 is an analog of Assumption A4, and both assumptions are required when establishing the CLT for the LSSs of $\mathbf{B}_n$.

Our first result concerns the convergence of the ESD $F^{\mathbf{B}_n}$. As this is a direct consequence of Theorem 1.2.1 in [26], its proof is omitted.

Theorem 1. *Suppose that Assumptions A1–A3 hold. Then, almost surely, the ESD $F^{\mathbf{B}_n}$ converges weakly to a probability distribution $F^{c,H,\tilde{H}}$. The Stieltjes transform $m(z)$ of $F^{c,H,\tilde{H}}$ is a solution to the following system of equations, which are defined on $\mathbb{C}^+$:*

$$zm(z) = -\int \frac{1}{1+s_2(z)t} dH(t), \quad z\underline{m}(z) = -\int \frac{1}{1+s_1(z)t} d\tilde{H}(t), \quad z\underline{m}(z) = -1 - zs_1(z)s_2(z), \tag{5}$$

where $\underline{m}(z) = -(1-c)/z + cm(z)$ is the Stieltjes transform of $(1-c)\delta_0 + cF^{c,H,\tilde{H}}$. Moreover, the three analytic functions $(m(z), s_1(z), s_2(z))$ are unique in the following set:

$$\{(m(z), s_1(z), s_2(z)) : \Im(m(z)) > 0, \Im(zs_1(z)) > 0, \Im(s_2(z)) > 0, z \in \mathbb{C}^+\}.$$

Theorem 1 describes the first-order convergence of the LSSs of $\mathbf{B}_n$. That is, for any function f analytic on an open interval containing

$$I_c \triangleq \left[\liminf_{p\to\infty} \lambda_{\min}^{\boldsymbol{\Sigma}} I_{(0,1)}(c) a(1-\sqrt{c})^2, \quad \limsup_{p\to\infty} \lambda_{\max}^{\boldsymbol{\Sigma}} b(1+\sqrt{c})^2\right], \tag{6}$$

we have almost surely, as $n, p \to \infty$,

$$\int f(x) dF^{\mathbf{B}_n}(x) \to \int f(x) dF^{c,H,\tilde{H}}(x).$$

To study the second-order convergence of these LSSs, we introduce two processes, namely,

$$G_{n1}(x) \triangleq F_n(x) - F^{c_n,H_p,\tilde{H}_n}(x), \quad G_{n2}(x) \triangleq F^{c_n,H_p,\tilde{H}_n}(x) - F^{c_n,H_p,M_p}(x), \tag{7}$$

where $F^{c_n,H_p,\tilde{H}_n}(x)$ and $F^{c_n,H_p,M_p}(x)$ are solutions to the system of equations in (5) with the parameters $(c, H, \tilde{H})$ replaced by $(c_n, H_p, \tilde{H}_n)$ and (c_n, H_p, M_p), respectively, and

$$\tilde{H}_n \triangleq \frac{1}{n}\sum_{j=1}^{n} \delta_{\xi_j^2/p}$$

denotes the empirical distribution generated by the sequence $\{\xi_j^2/p\}$. Note that $F^{c_n,H_p,\tilde{H}_n}$ is a random measure, while F^{c_n,H_p,M_p} is deterministic. With the two empirical processes $G_{n1}(x)$ and $G_{n2}(x)$, we can centralize the LSS $\int f(x)dF^{\mathbf{B}_n}(x)$ as

$$\int f(x) dG_n(x) \triangleq \int f(x) d[F^{\mathbf{B}_n}(x) - F^{c_n,H_p,M_p}(x)] = \int f(x) dG_{n1}(x) + \int f(x) dG_{n2}(x). \tag{8}$$

The asymptotic distribution of this centralized statistic is presented in the following theorem.

Theorem 2. *Suppose that Assumptions A1–A5 hold. Let $f_1, \ldots, f_k$ be k functions on $\mathbb{R}$ that are analytic on an open interval containing I_c defined in (6). Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two nonoverlapping, closed, positive-oriented contours in the complex plane, and let each contour enclose the support of $F^{c,H,\tilde{H}}$.*

(i) *The random vector*

$$p\left(\int f_1(x)dG_{n1}(x), \ldots, \int f_k(x)dG_{n1}(x)\right)$$

converges weakly to a Gaussian vector $(X_{f_1}, \ldots, X_{f_k})$. The mean function is

$$\mathbb{E}X_f = -\frac{1}{2\pi\mathrm{i}}\oint_{\mathcal{C}_1} f(z)\frac{P(z)-\underline{P}(z)R(z)}{1-R(z)}dz,$$

where

$$\begin{aligned}
P(z) =& -\frac{c}{z^3}\int\frac{t^2d\tilde{H}(t)}{(1+s_1(z)t)^2}\int\frac{t^2dH(t)}{(1+s_2(z)t)^3}\left[1-\int\frac{ct^2dH(t)}{z^2(1+s_2(z)t)^2}\int\frac{t^2d\tilde{H}(t)}{(1+s_1(z)t)^2}\right]^{-1}\\
&-\frac{c^2}{z^4}\int\frac{t^3d\tilde{H}(t)}{(1+s_1(z)t)^3}\int\frac{tdH(t)}{(1+s_2(z)t)^2}\int\frac{t^2dH(t)}{(1+s_2(z)t)^2}\left[1-\int\frac{ct^2dH(t)}{z^2(1+s_2(z)t)^2}\right.\\
&\times\left.\int\frac{t^2d\tilde{H}(t)}{(1+s_1(z)t)^2}\right]^{-1}-\frac{2s_1(z)}{z^2}\int\frac{t^2d\tilde{H}(t)}{(1+s_1(z)t)^3}\int\frac{tdH(t)}{(1+s_2(z)t)^2},\\
\underline{P}(z) =& -\frac{c}{z^3}\int\frac{t^2d\tilde{H}(t)}{(1+s_1(z)t)^3}\int\frac{t^2dH(t)}{(1+s_2(z)t)^2}\left[1-\int\frac{ct^2dH(t)}{z^2(1+s_2(z)t)^2}\int\frac{t^2d\tilde{H}(t)}{(1+s_1(z)t)^2}\right]^{-1}\\
&+\int\frac{2s_1^2(z)t^2d\tilde{H}(t)}{cz(1+s_1(z)t)^3},\\
R(z) =& \frac{c}{z}\int\frac{t^2d\tilde{H}(t)}{(1+s_1(z)t)^2}\int\frac{tdH(t)}{(1+s_2(z)t)^2}\left[\int\frac{td\tilde{H}(t)}{(1+s_1(z)t)^2}\right]^{-1}.
\end{aligned}$$

The covariance function equals

$$\mathrm{Cov}\left(X_f, X_g\right) = \frac{1}{2\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)g(z_2)\frac{\partial^2}{\partial z_1\partial z_2}\left[\ln(1-a(z_1,z_2))+b(z_1,z_2)\right]dz_1dz_2$$

for $f, g \in \{f_1, \ldots, f_k\}$, where

$$a(z_1,z_2) = \frac{z_1s_2(z_1)-z_2s_2(z_2)}{s_1(z_1)-s_1(z_2)}\frac{z_1s_1(z_1)-z_2s_1(z_2)}{z_1z_2(s_2(z_1)-s_2(z_2))},\quad b(z_1,z_2) = \frac{(z_1s_2(z_1)-z_2s_2(z_2))s_1(z_1)s_1(z_2)}{c(s_1(z_1)-s_1(z_2))}.$$

(ii) *The random vector*

$$\sqrt{p}\left(\int f_1(x)dG_{n2}(x), \ldots, \int f_k(x)dG_{n2}(x)\right)$$

converges weakly to a zero-mean Gaussian vector $(Y_{f_1}, \ldots, Y_{f_k})$ with the covariance function

$$\mathrm{Cov}\left(X_f, X_g\right) = -\frac{1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)g(z_2)h(z_1,z_2)dz_1dz_2$$

for $f, g \in \{f_1, \ldots, f_k\}$, where

$$h(z_1,z_2) = c\int g(t,z_1)g(t,z_2)d\tilde{H}(t) - c\int g(t,z_1)d\tilde{H}(t)\int g(t,z_2)d\tilde{H}(t)$$

and $g(t,z) = (A(z)+tB(z))/(1+ts_1(z))$ with

$$A(z) = \int\frac{tdH(t)}{(1+s_2t)^2}\int\frac{t^2d\tilde{H}(t)}{(1+s_1t)^2}\left[\int\frac{z^2td\tilde{H}(t)}{(1+s_1t)^2}-\int\frac{cztdH(t)}{(1+s_2t)^2}\int\frac{t^2d\tilde{H}(t)}{(1+s_1t)^2}\right]^{-1},$$

$$B(z) = \int\frac{tdH(t)}{(1+s_2t)^2}\int\frac{td\tilde{H}(t)}{(1+s_1t)^2}\left[\int\frac{z^2td\tilde{H}(t)}{(1+s_1t)^2}-\int\frac{cztdH(t)}{(1+s_2t)^2}\int\frac{t^2d\tilde{H}(t)}{(1+s_1t)^2}\right]^{-1}.$$

Remark 1. Theorem 2 establishes a new CLT for the LSSs of $\mathbf{B}_n$ under the real-valued elliptical model. For the LSS defined in (8), the first part $\int f(x)dG_{n1}(x)$ converges in distribution to a Gaussian variable at the rate of $1/p$, while the second part

$\int f(x)dG_{n2}(x)$ converges in distribution to another Gaussian variable at the rate of $1/\sqrt{p}$. Therefore, the limiting distribution of the LSS is dominated by the second part. This is different from the case where the LMD $\tilde{H} = \delta_1$, as discussed in [10], under whose model settings ξ^2/p degenerates to the constant 1 according to (4), which accelerates the convergence rate of $\int f(x)dG_{n2}(x)$ to $1/p$. As a result, both G_{n1} and G_{n2} yield nontrivial contributions to the CLT of [10].

Remark 2. Our CLT parallels that of [11], who studied the LSSs of sample covariance matrices in complex-valued elliptical distributions. That is, the population has the form $\mathbf{x} = \xi\mathbf{A}\mathbf{v}$, with ξ and $\mathbf{A}$ defined as in our model (3), while the random direction $\mathbf{v}$ is uniformly distributed on the unit sphere in $\mathbb{C}^p$. Under this model, [11] established the asymptotic normality of $p\int f(x)dG_{n1}(x)$ and $\sqrt{p}\int f(x)dG_{n2}(x)$. Our results show that the limit of the second part involving G_{n2} is universal for both real-valued and complex-valued distributions. However, the limit of the first part is different, and this difference plays an important role in a high-order correction to the CLT. Specifically, if we obtain

$$p\int f(x)dG_{n1}(x) \xrightarrow{D} N\left(\mu_1, \sigma_1^2\right), \quad \sqrt{p}\int f(x)dG_{n2}(x) \xrightarrow{D} N\left(0, \sigma_2^2\right), \tag{9}$$

then a finite-dimensional correction to the distribution of $\sqrt{p}\int f(x)dG_n(x)$ is

$$\sqrt{p}\int f(x)dG_n(x) \dot{\sim} N\left(\mu_1/\sqrt{p}, \sigma_1^2/p + \sigma_2^2\right).$$

This finding follows from the fact that the two parts in (9) are asymptotically independent (following a conditioning argument as in the proof of Lemma 1). It is worth noting that for moderately large p and n, such a correction is often necessary; moreover, discounting the contribution of the first part may lead to significant bias in the CLT.

As an example, we consider the centralized statistic $\hat{\beta}_2 \triangleq \sqrt{p}\int x^2 dG_n(x)$ when the covariance matrix $\boldsymbol{\Sigma}$ is an identity covariance matrix. In this case, we can set $H = \delta_1$; thus, the three equations in (5) reduce to

$$z = -\frac{1}{m(z)} + \int \frac{t}{1 + ctm(z)} d\tilde{H}(t), \quad s_1(z) = cm(z), \quad s_2(z) = -1 - \frac{1}{zm(z)}.$$

Some algebra can reveal

$$\frac{P(z) - \underline{P}(z)R(z)}{1 - R(z)} = \int \frac{cm^3(z)t^2 d\tilde{H}(t)}{(1 + cm(z)t)^3}\left\{\left[1 - \int \frac{cm^2(z)t^2 d\tilde{H}(t)}{(1 + cm(z)t)^2}\right]^{-2} - 2\left[1 - \int \frac{cm^2(z)t^2 d\tilde{H}(t)}{(1 + cm(z)t)^2}\right]^{-1}\right\}$$

and

$$\begin{aligned}\frac{\partial^2\{\log[1 - a(z_1, z_2)] + b(z_1, z_2)\}}{\partial z_1 \partial z_2} =& \frac{1}{(z_1 - z_2)^2} - \frac{m'(z_1)m'(z_2)}{[m(z_1) - m(z_2)]^2} + \frac{1}{(m(z_1) - m(z_2))^3}\Big[(m^2(z_2)(m(z_2) - m(z_1))m'(z_1) \\ &+ m^2(z_1)(m(z_2) - m(z_1))m'(z_2) + 2(z_1 - z_2)m(z_1)m(z_2)m'(z_1)m'(z_2))\Big].\end{aligned}$$

From the above results and the calculation of some residuals (see [16]), one can obtain

$$p\int x^2 dG_{n1}(x) \xrightarrow{D} N\left(c\gamma_2, 4c^2\gamma_2^2\right), \quad \sqrt{p}\int x^2 dG_{n2}(x) \xrightarrow{D} N\left(0, c^3(\gamma_4 - \gamma_2^2) + 4c^2\gamma_1\gamma_3 + 4c(1-c)\gamma_1^2\gamma_2 - 4c\gamma_1^4\right),$$

where $\gamma_i = \int x^i d\tilde{H}(x)$ for $i \in \{1, \ldots, 4\}$.

Next, we numerically examine the fluctuation of $\hat{\beta}_2$ under a Gaussian scale mixture model, where the variable ξ has a discrete distribution with two mass points:

$$\Pr(\xi = 1.8\sqrt{p}) = 0.8, \quad \Pr(\xi = 1.5\sqrt{p}) = 0.2.$$

The dimensional settings are $(p, n) = (100, 150)$ and $(600, 900)$ and thus the ratio c is set to $2/3$. Then, the limiting distribution of $\hat{\beta}_2$ is $N(0, 9.93)$, while its finite-dimensional correction should be $N(0.628, 11.50)$ for $(p, n) = (100, 150)$ and $N(0.256, 10.19)$ for $(p, n) = (600, 900)$. Normal QQ plots for $\hat{\beta}_2$ from 5000 independent replications are displayed in Fig. 1. Here, we normalize $\hat{\beta}_2$ according to its corrected CLT. The results demonstrate that the empirical distribution of the normalized statistic can be well approximated by the standard Gaussian distribution under both the two dimensional settings.

3. Applications

3.1. Application to the interval estimation of the parameters in a Gaussian scale mixture

Recall that a p-dimensional vector $\mathbf{x} \in \mathbb{R}^p$ is a multivariate Gaussian mixture with k subpopulations if its density function has the form

$$f(\mathbf{x}) = \sum_{j=1}^{k} p_j \phi\left(\mathbf{x}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right), \tag{10}$$

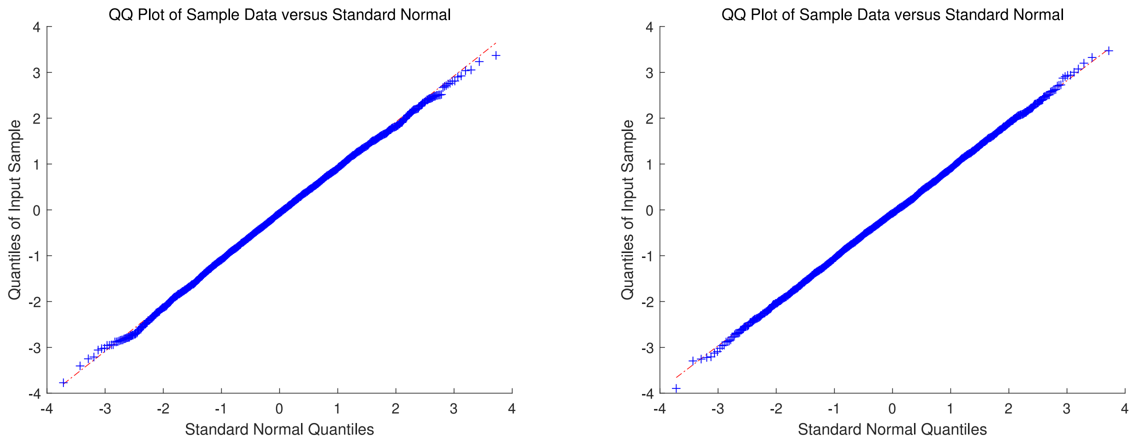

Fig. 1. QQ plots for the normalized $\hat{\beta}_2$ from 5000 independent replications with $(p, n) = (100, 150)$ (left panel) and $(p, n) = (600, 900)$ (right panel).

where (p_j) are the k mixing weights and $\phi(\cdot; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ denotes the density function of the jth subpopulation with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$. Assume that

$$\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k = \mathbf{0}, \quad \boldsymbol{\Sigma}_j = v_j \boldsymbol{\Sigma} \text{ for some } v_j > 0,\ j \in \{1, \ldots, k\},$$

where $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite matrix denoting the common shape matrix of the k subpopulations and is normalized by $\operatorname{tr}(\boldsymbol{\Sigma}) = p$. This mixture model is thus called a Gaussian scale mixture [8]. According to our elliptical model (3), such a mixture can be written as

$$\mathbf{x} = \xi \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{u} \triangleq \eta\kappa \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{u} \triangleq \eta \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{z}, \tag{11}$$

where η, κ and $\mathbf{u}$ are mutually independent, κ^2 follows a chi-squared distribution with p degrees of freedom such that $\mathbf{z} \triangleq \kappa \mathbf{u}$ is a standard Gaussian vector in $\mathbb{R}^p$, and η follows a discrete distribution $\tilde{H}$ with k positive mass points (v_i), i.e.,

$$\Pr(\eta = v_i) = p_i,\ i \in \{1, \ldots, k\}.$$

Note that we do not need any moment restriction on η^2 since the normalization condition $\operatorname{tr}(\boldsymbol{\Sigma}) = p$ is sufficient for identifying the product in (11). Our target in this section is to estimate the distribution of η; that is, we consider the interval estimation of $\boldsymbol{\theta} \triangleq \{v_1, \ldots, v_k, p_1, \ldots, p_{k-1}\} \in \Theta$ where

$$\Theta \triangleq \left\{ \boldsymbol{\theta} : 0 < v_1 < \cdots < v_k < \infty; \quad 0 < \min_i p_i, \quad \sum_{i=1}^{k} p_i = 1 \right\}$$

and the order k is assumed to be known. Our strategy is based on the moment method; that is, we first estimate the moments (w_j) of η and then solve a system of moment equations to obtain a consistent estimate of $\boldsymbol{\theta}$.

Specifically, let $j \in \{1, \ldots, 2k-1\}$, $\hat{\boldsymbol{\alpha}}_j = (\hat{\alpha}_1, \ldots, \hat{\alpha}_j)^\top$, $\boldsymbol{\alpha}_j = (\mathbb{E}\hat{\alpha}_1, \ldots, \mathbb{E}\hat{\alpha}_j)^\top$ and $\boldsymbol{w}_j = (w_1, \ldots, w_j)^\top$, where

$$\hat{\alpha}_j \triangleq c_n \int x^j dF^{\mathbf{B}_n}(x), \qquad w_j \triangleq \sum_{i=1}^{k} v_i^{2j} p_i.$$

In addition, denote

$$g_{1,2k-1} : \boldsymbol{w}_{2k-1} \longrightarrow \boldsymbol{\theta}, \quad g_{2,2k-1} : \boldsymbol{\alpha}_{2k-1} \longrightarrow \boldsymbol{w}_{2k-1}$$

as the mappings between the corresponding vectors. These two mappings are both one-to-one and the determinants of their Jacobian matrices are all nonzero (see [2,15]). Therefore, it holds true from Theorem 1 that

$$\hat{\boldsymbol{\alpha}}_{2k-1} - \boldsymbol{\alpha}_{2k-1} \xrightarrow{a.s.} \mathbf{0},$$

which is followed by

$$\hat{\boldsymbol{\theta}} \triangleq g_{1,2k-1} \circ g_{2,2k-1}(\hat{\boldsymbol{\alpha}}_{2k-1}) \xrightarrow{a.s.} \boldsymbol{\theta}.$$

Moreover, from Theorem 2 and a standard application of the delta method, one can obtain the asymptotic distribution of $\hat{\boldsymbol{\theta}}$.

Theorem 3. *Suppose that Assumptions A1–A5 hold and that the true value $\boldsymbol{\theta}$ is an inner point of Θ. Then, we have*

$$\sqrt{p}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N_{2k-1}(\mathbf{0}, \mathbf{J}_{1,2k-1}\mathbf{J}_{2,2k-1}\boldsymbol{\Psi}_{2k-1}\mathbf{J}_{2,2k-1}^\top \mathbf{J}_{1,2k-1}^\top), \tag{12}$$

where $\mathbf{J}_{\ell,2k-1}$ *represents the Jacobian matrix* $\partial g_{\ell,2k-1}/\partial \boldsymbol{w}_{2k-1}$ *for* $\ell \in \{1,2\}$ *and* $\boldsymbol{\Psi}_{2k-1}$ *denotes the limiting covariance matrix of* $\sqrt{p}\left(\hat{\boldsymbol{\alpha}}_{2k-1} - \boldsymbol{\alpha}_{2k-1}\right)$.

3.2. An example and simulation results

We illustrate an example of a Gaussian scale mixture with two subpopulations, i.e., $k = 2$. In this case, the core statistic for the estimation is $\hat{\boldsymbol{\alpha}}_3 = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)^\top$. Its mean vector $\boldsymbol{\alpha}_3 = (\alpha_1, \alpha_2, \alpha_3)^\top$ is given by

$$\begin{aligned}
\alpha_1 &= n^{-1} w_1 \mathrm{tr}\boldsymbol{\Sigma}, \quad \alpha_2 = n^{-3}\left(n w_2((\mathrm{tr}\boldsymbol{\Sigma})^2 + 2\mathrm{tr}\boldsymbol{\Sigma}^2) + n_2 w_1^2 \mathrm{tr}\boldsymbol{\Sigma}^2\right), \\
\alpha_3 &= n^{-4}\left(n w_3((\mathrm{tr}\boldsymbol{\Sigma})^3 + 6\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 8\mathrm{tr}\boldsymbol{\Sigma}^3) + 3n_2 w_2 w_1(\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 2\mathrm{tr}\boldsymbol{\Sigma}^3) + n_3 w_1^3 \mathrm{tr}\boldsymbol{\Sigma}^3\right),
\end{aligned} \tag{13}$$

where $n_i = n!/(n-i)!$. In addition, its covariance matrix $\boldsymbol{\Psi}_3$ can be calculated directly, an approximation of which can be found in Section 5.

Next, we calculate $g_{1,3}$ and $g_{2,3}$. Solving the moment equations

$$w_\ell = \sum_{i=1}^{k} v_i^{2\ell} p_i, \;\; \ell \in \{0, 1, 2, 3\},$$

yields the mapping $g_{1,3}$,

$$v_1 = \sqrt{\frac{-a + \sqrt{a^2 - 4b}}{2}}, \quad v_2 = \sqrt{\frac{-a - \sqrt{a^2 - 4b}}{2}}, \quad p_1 = \frac{1}{2} + \frac{a + 2w_1}{2\sqrt{a^2 - 4b}},$$

where

$$a = \frac{w_1 w_2 - w_3}{w_2 - w_1^2}, \quad b = \frac{w_1 w_3 - w_2^2}{w_2 - w_1^2}.$$

The mapping $g_{2,3}$ can be derived directly from (13) and is given by

$$\begin{aligned}
w_1 &= \frac{n\alpha_1}{\mathrm{tr}\boldsymbol{\Sigma}}, \quad w_2 = \frac{n^2\alpha_2 - (n-1)w_1^2 \mathrm{tr}\boldsymbol{\Sigma}^2}{(\mathrm{tr}\boldsymbol{\Sigma})^2 + 2\mathrm{tr}\boldsymbol{\Sigma}^2}, \\
w_3 &= \frac{n^3\alpha_3 - 3(n-1)w_1 w_2(\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 2\mathrm{tr}\boldsymbol{\Sigma}^3) - (n-1)(n-2)w_1^3\mathrm{tr}\boldsymbol{\Sigma}^3}{(\mathrm{tr}\boldsymbol{\Sigma})^3 + 6\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 8\mathrm{tr}\boldsymbol{\Sigma}^3}.
\end{aligned}$$

Therefore, a confidence interval for $\boldsymbol{\theta}$ is obtained from Theorem 3 and the two mappings $g_{1,3}$ and $g_{2,3}$.

We note that such confidence interval involves some unknown quantities related to the common shape matrix $\boldsymbol{\Sigma}$, for instance, $\{\mathrm{tr}(\boldsymbol{\Sigma}^k)/p\}$. To obtain estimates of these quantities, we consider the spatial-sign covariance matrix of the sample $\{\mathbf{x}_j\}$, i.e.,

$$\mathbf{D}_n = \frac{1}{n}\sum_{j=1}^{n} \tilde{\mathbf{x}}_j \tilde{\mathbf{x}}_j^\top,$$

where

$$\tilde{\mathbf{x}}_j \triangleq \sqrt{p}\frac{\mathbf{x}_j}{\|\mathbf{x}_j\|} = \sqrt{p}\frac{\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{u}_j}{\|\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{u}_j\|}, \; j \in \{1, \ldots, n\},$$

are the spatial-sign transforms of $(\mathbf{x}_j)$. The first- and second-order convergence of the general LSSs of $\mathbf{D}_n$ have been established in [14]. Applying their results, we can obtain consistent estimates of $\mathrm{tr}(\boldsymbol{\Sigma}^k)/p$ for any $k \geq 1$. In particular, the second-order convergence rate of such estimates is $O(1/p)$, which is much faster than that of $\hat{\boldsymbol{\theta}}$; hence, substituting these estimates for $\{\mathrm{tr}(\boldsymbol{\Sigma}^k)/p\}$ in the confidence interval for $\boldsymbol{\theta}$ does not affect its asymptotic results.

Next, we numerically examine the finite-dimensional behaviors of our proposed confidence interval. Six models are considered in this experiment:

$$\begin{aligned}
&\text{Model 1.} \;\; v_1 = 1.5, \; v_2 = 0.5, \; p_1 = 0.3, \;\; \boldsymbol{\Sigma} = \mathbf{I}_p; \\
&\text{Model 2.} \;\; v_1 = 1.5, \; v_2 = 0.5, \; p_1 = 0.5, \;\; \boldsymbol{\Sigma} = \mathbf{I}_p; \\
&\text{Model 3.} \;\; v_1 = 1.5, \; v_2 = 0.5, \; p_1 = 0.7, \;\; \boldsymbol{\Sigma} = \mathbf{I}_p;
\end{aligned} \tag{14}$$

$$\begin{aligned}
&\text{Model 4.} \;\; v_1 = 1, \; v_2 = 0.5, \; p_1 = 0.3, \;\; \boldsymbol{\Sigma} = \mathrm{diag}\{\underbrace{1.2, \ldots, 1.2}_{p/2}, \underbrace{0.8, \ldots, 0.8}_{p/2}\}; \\
&\text{Model 5.} \;\; v_1 = 1, \; v_2 = 0.5, \; p_1 = 0.5, \;\; \boldsymbol{\Sigma} = \mathrm{diag}\{\underbrace{1.2, \ldots, 1.2}_{p/2}, \underbrace{0.8, \ldots, 0.8}_{p/2}\}; \\
&\text{Model 6.} \;\; v_1 = 1, \; v_2 = 0.5, \; p_1 = 0.7, \;\; \boldsymbol{\Sigma} = \mathrm{diag}\{\underbrace{1.2, \ldots, 1.2}_{p/2}, \underbrace{0.8, \ldots, 0.8}_{p/2}\}.
\end{aligned} \tag{15}$$

Table 1
Coverage probabilities for the three parameters v_1, v_2 and p_1 under Model 1 defined in (14).

p	$c_n = 1$			$c_n = 1/2$			$c_n = 1/3$		
	v_1	v_2	p_1	v_1	v_2	p_1	v_1	v_2	p_1
400	0.8922	0.7964	0.9358	0.9260	0.9092	0.9452	0.9344	0.9270	0.9462
600	0.9012	0.8164	0.9512	0.9442	0.9152	0.9480	0.9486	0.9302	0.9484
800	0.9290	0.8190	0.9492	0.9396	0.9180	0.9520	0.9528	0.9394	0.9466

Table 2
Coverage probabilities for the three parameters v_1, v_2 and p_1 under Model 2 defined in (14).

p	$c_n = 1$			$c_n = 1/2$			$c_n = 1/3$		
	v_1	v_2	p_1	v_1	v_2	p_1	v_1	v_2	p_1
400	0.9014	0.8784	0.9576	0.9286	0.9318	0.9564	0.9540	0.9386	0.9528
600	0.9116	0.8872	0.9512	0.9456	0.9358	0.9482	0.9482	0.9422	0.9496
800	0.9282	0.9012	0.9444	0.9494	0.9350	0.9442	0.9504	0.9478	0.9462

Table 3
Coverage probabilities for the three parameters v_1, v_2 and p_1 under Model 3 defined in (14).

p	$c_n = 1$			$c_n = 1/2$			$c_n = 1/3$		
	v_1	v_2	p_1	v_1	v_2	p_1	v_1	v_2	p_1
400	0.8652	0.9058	0.9370	0.9208	0.9248	0.9472	0.9370	0.9334	0.9516
600	0.8744	0.9176	0.9460	0.9322	0.9440	0.9566	0.9412	0.9478	0.9570
800	0.8882	0.9218	0.9480	0.9398	0.9456	0.9444	0.9384	0.9528	0.9442

Table 4
Coverage probabilities for the three parameters v_1, v_2 and p_1 under Model 4 defined in (15).

p	$c_n = 1$			$c_n = 1/2$			$c_n = 1/3$		
	v_1	v_2	p_1	v_1	v_2	p_1	v_1	v_2	p_1
400	0.9840	0.8790	0.9496	0.9708	0.9170	0.9680	0.9570	0.9334	0.9428
600	0.9642	0.8836	0.9522	0.9586	0.9142	0.9530	0.9510	0.9272	0.9550
800	0.9614	0.9010	0.9492	0.9584	0.9214	0.9482	0.9522	0.9324	0.9486

Table 5
Coverage probabilities for the three parameters v_1, v_2 and p_1 under Model 5 defined in (15).

p	$c_n = 1$			$c_n = 1/2$			$c_n = 1/3$		
	v_1	v_2	p_1	v_1	v_2	p_1	v_1	v_2	p_1
400	0.9818	0.9092	0.9644	0.9658	0.9254	0.9660	0.9544	0.9400	0.9560
600	0.9698	0.8982	0.9562	0.9612	0.9288	0.9440	0.9532	0.9424	0.9560
800	0.9602	0.9180	0.9502	0.9598	0.9312	0.9600	0.9550	0.9486	0.9548

Table 6
Coverage probabilities for the three parameters v_1, v_2 and p_1 under Model 6 defined in (15).

p	$c_n = 1$			$c_n = 1/2$			$c_n = 1/3$		
	v_1	v_2	p_1	v_1	v_2	p_1	v_1	v_2	p_1
400	0.9724	0.9000	0.9776	0.9684	0.9418	0.9672	0.9756	0.9468	0.9632
600	0.9682	0.9112	0.9664	0.9584	0.9330	0.9618	0.9632	0.9532	0.9626
800	0.9622	0.9390	0.9588	0.9608	0.9398	0.9550	0.9524	0.9466	0.9534

The dimensional settings are $p \in \{400, 600, 800\}$ and $n = p/c_n$ with $c_n = 1$, $1/2$ and $1/3$. The nominal confidence level is fixed at $1 - \alpha = 0.95$, and the number of independent replications is 5000.

Tables 1–6 summarize the empirical coverage probabilities for $\boldsymbol{\theta}$ in the corresponding models (Models 1–6). These results show that the coverage probabilities for the mass points v_1 and v_2 are slightly biased when the dimensions p and n are small. As the dimensions increase, the biases decrease, and the coverage probabilities approach the nominal level of 0.95.

4. Technical details

In this section, we prove Theorem 2 and the approximation of the covariance matrix $\boldsymbol{\Psi}_3$ used in Section 3.2. The proof of Theorem 2 follows the strategy in [11], which is a typical procedure for the proof of a CLT for LSSs in random matrix theory. Roughly speaking, via Cauchy's integral formula, we first transform LSSs into integrals of the Stieltjes transforms of G_{n1} and G_{n2} with respect to a contour enclosing the support of G. Then, the CLT is obtained by determining the weak convergences of the Stieltjes transforms to two-dimensional Gaussian processes. Thus, the main task of the proof of Theorem 2 is to obtain the following lemma.

Let $v_0 > 0$ be arbitrary, x_r be any number greater than the right end point of the interval I_c defined in (6), and x_l be any negative number if the left end point of the interval is zero; otherwise,

$$x_l \in \left(0, \liminf_{n,p\to\infty} \min_{1\le j\le n} \{\xi_j^2/p\}\lambda_{\min}^{\Sigma} I_{(0,1)}(c)(1-\sqrt{c})^2\right).$$

Define a contour $\mathcal{C}$ as

$$\mathcal{C} = \{x+iv : x \in \{x_r, x_l\}, v \in [-v_0, v_0]\} \cup \mathcal{C}_u, \quad \mathcal{C}_u = \{x \pm iv_0 : x \in [x_l, x_r]\}. \tag{16}$$

Let $(m_0(z), s_{10}(z), s_{20}(z))$ and $(m_1(z), s_{11}(z), s_{21}(z))$ be the solutions to Eqs. (5) with the triplet $(c, H, \tilde{H})$ replaced by $(c_n, H_p, \tilde{H}_n)$ and (c_n, H_p, M_p), respectively. Our aim is to study the fluctuations of the following two random processes:

$$M_n(z) = p\left[m_n(z) - m_0(z)\right], \quad \widetilde{M}_n(z) = \sqrt{p}\left[m_0(z) - m_1(z)\right], \quad z \in \mathcal{C}.$$

For the first process, we need to define its truncated version $\widehat{M}_n(\cdot)$. Choose a sequence $\{\varepsilon_n\}$ decreasing to zero satisfying $\varepsilon_n > n^{-a}$ for some $a \in (3/4, 1)$. For $z = x+iv \in \mathcal{C}$, denote

$$\widehat{M}_n(z) = \begin{cases} M_n(z), & z \in \mathcal{C}_n, \\ M_n(x + in^{-1}\varepsilon_n), & x \in \{x_l, x_r\}, \quad v \in [0, n^{-1}\varepsilon_n], \end{cases} \tag{17}$$

where $\mathcal{C}_n = \mathcal{C}_u \cup \{x \pm iv : x \in \{x_l, x_r\}, v \in [n^{-1}\varepsilon_n, v_0]\}$. Then, we have the following auxiliary lemma.

Lemma 1. *Suppose that Assumptions A1–A5 hold.*

(i) *The random process $\widehat{M}_n(\cdot)$ converges weakly to a two-dimensional Gaussian process $M(\cdot)$ on $\mathcal{C}$ with mean function*

$$\mathbb{E}M(z) = \frac{P(z) - \underline{P}(z)R(z)}{1 - R(z)}, \tag{18}$$

and the covariance function is

$$\operatorname{Cov}\left(M(z_1), M(z_2)\right) = -\frac{2\partial^2}{\partial z_1 \partial z_2}\left[\ln(1 - a(z_1, z_2)) + b(z_1, z_2)\right]. \tag{19}$$

(ii) *The random process $\widetilde{M}_n(\cdot)$ converges weakly to a two-dimensional Gaussian process $\widetilde{M}(\cdot)$ on $\mathcal{C}$ with zero-mean, and the covariance function equals*

$$\operatorname{Cov}\left(\widetilde{M}(z_1), \widetilde{M}(z_2)\right) = h(z_1, z_2).$$

Here, the functions $P(z)$, $\underline{P}(z)$, $R(z)$, $a(z_1, z_2)$, $b(z_1, z_2)$ and $h(z_1, z_2)$ are defined in Theorem 2.

With the aid of this lemma, we next prove Theorem 2, and we postpone the proof of Lemma 1 to Section 4.

Proof of Theorem 2. For any function f analytic on a simple connected domain D containing the support of a distribution function G, it holds that

$$\int f(x)dG(x) = -\frac{1}{2\pi\mathrm{i}}\oint_{\mathcal{C}} f(z)m_G(z)dz, \tag{20}$$

where $m_G(z)$ denotes the Stieltjes transform of G and $\mathcal{C} \subset D$ is a simple, closed, positively oriented contour enclosing the support of G. Similar to (16), we choose v_0, x_r, and x_l such that $f_1, \ldots, f_k$ are all analytic on and inside the contour $\mathcal{C}$. We denote by K a common upper bound of these functions on $\mathcal{C}$. Therefore, almost surely, for all large n, $\{f_1, \ldots, f_k\}$ satisfies the equation in (20) with $G = F^{\mathbf{B}_n}$; moreover,

$$\begin{aligned}\left|\int f_i(z)(M_n(z) - \widehat{M}_n(z))dz\right| \le & 4K\varepsilon_n \Big(\big|\max_j\{(\xi_j^2/p)\lambda_{\max}^{\Sigma}(1+\sqrt{c_n})^2, \lambda_{\max}^{\mathbf{B}_n}\} - x_r\big|^{-1} \\ & + \big|\min_j\{(\xi_j^2/p)\lambda_{\min}^{\Sigma} I_{(0,1)}(c_n)(1-\sqrt{c_n})^2, \lambda_{\min}^{\mathbf{B}_n}\} - x_l\big|^{-1} \Big),\end{aligned}$$

which converges to zero as $n \to \infty$. Since

$$\widehat{M}_n(\cdot) \longrightarrow \left(-\frac{1}{2\pi\mathrm{i}}\int f_1(z)\widehat{M}_n(z)dz, \ldots, -\frac{1}{2\pi\mathrm{i}}\int f_k(z)\widehat{M}_n(z)dz\right)$$

is a continuous mapping of $C(\mathcal{C},\mathbb{R}^2)$ into $\mathbb{R}^{2k}$, it follows that the above random vector converges in distribution to a multivariate Gaussian vector with mean and covariance functions being

$$\mathbb{E}(X_f) = -\frac{1}{2\pi \mathrm{i}}\oint_{\mathcal{C}_1} f(z)\mathbb{E}[M(z)]dz, \quad \mathrm{Cov}\left(X_f, X_g\right) = -\frac{1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)g(z_2)\mathrm{Cov}[M(z_1), M(z_2)]dz_1dz_2,$$

where $f, g \in \{f_1,\dots,f_k\}$ and $\{\mathcal{C}_i\}$ are two nonoverlapping analogs of the contour $\mathcal{C}$. Analogously, (ii) of Theorem 2 is straightforward to prove. □

Proof of (i) in Lemma 1. We prove this conclusion by assuming that the sequence $\{\xi_j\}$ is fixed. Then, the lemma automatically holds by noticing that the limiting distribution of $\widehat{M}_n(z)$ does not depend on a specific realization of $\{\xi_j\}$. Write for $z \in \mathcal{C}_n$, $\widehat{M}_n(z) = M_n^{(1)}(z) + M_n^{(2)}(z)$, where

$$M_n^{(1)}(z) = p[m_n(z) - \mathbb{E}m_n(z)], \quad M_n^{(2)}(z) = p[\mathbb{E}m_n(z) - m_0(z)].$$

Using the strategy in [4], we demonstrate the convergence of $\widehat{M}_n(z)$ by showing the following three facts:

Fact 1: Finite-dimensional convergence of $M_n^{(1)}(z)$ in distribution;
Fact 2: Tightness of $M_n^{(1)}(z)$ on $\mathcal{C}_n$;
Fact 3: Convergence of $M_n^{(2)}(z)$.

Finite-dimensional convergence of $M_n^{(1)}(z)$ in distribution: Here, we show for any positive integer r and any complex numbers $z_1,\dots,z_r \in \mathcal{C}_n$ that the random vector

$$[M_n^{(1)}(z_1),\dots,M_n^{(1)}(z_r)]$$

converges to a $2r$-dimensional Gaussian vector. Because of Assumption A4, without loss of generality, we may assume $\|\boldsymbol{\Sigma}\| \le 1$ for all p. Constants appearing in inequalities are denoted by K and may take on different values from one expression to the next.

We define some qualities that are frequently used in the following. Let $\mathbf{r}_j = (1/\sqrt{n})\mathbf{x}_j$,

$$\mathbf{D}(z) = \mathbf{B}_n - z\mathbf{I}, \quad \mathbf{D}_j(z) = \mathbf{D}(z) - \mathbf{r}_j\mathbf{r}_j^\top, \quad \mathbf{D}_{ij}(z) = \mathbf{D}(z) - \mathbf{r}_i\mathbf{r}_i^\top - \mathbf{r}_j\mathbf{r}_j^\top,$$
$$\varepsilon_j(z) = \mathbf{r}_j^\top\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{\xi_j^2}{np}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z), \quad \delta_j(z) = \mathbf{r}_j^\top\mathbf{D}_j^{-2}(z)\mathbf{r}_j - \frac{\xi_j^2}{np}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-2}(z),$$
$$\varepsilon_{j0}(z) = \mathbf{u}_j^\top\mathbf{A}^\top\mathbf{D}_j^{-1}(z)\mathbf{A}\mathbf{u}_j - \frac{1}{p}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z), \quad \delta_{j0}(z) = \mathbf{u}_j^\top\mathbf{A}^\top\mathbf{D}_j^{-2}(z)\mathbf{A}\mathbf{u}_j - \frac{1}{p}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-2}(z),$$
$$\beta_j(z) = \frac{1}{1+\mathbf{r}_j^\top\mathbf{D}_j^{-1}(z)\mathbf{r}_j}, \quad \bar{\beta}_j(z) = \frac{1}{1+(np)^{-1}\xi_j^2\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z)},$$
$$b_j(z) = \frac{1}{1+(np)^{-1}\xi_j^2\mathbb{E}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z)}.$$

Note that, for any $z = u + iv \in \mathbb{C}^+$, the last three quantities are bounded in absolute value by $|z|/v$. Moreover,

$$\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z) = -\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}(z)\beta_j(z), \tag{21}$$

and from Lemma 2.6 in [23], for any $p \times p$ matrix $\mathbf{B}$,

$$|\mathrm{tr}(\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z))\mathbf{B}| \le \frac{\|\mathbf{B}\|}{v}. \tag{22}$$

Let $\mathbb{E}_0(\cdot)$ denote the expectation and $\mathbb{E}_j(\cdot)$ denote the conditional expectation with respect to the σ-field generated by $\mathbf{r}_1,\dots,\mathbf{r}_j$. Following a martingale decomposition and according to the identity in (21), we have

$$\begin{aligned}M_n^{(1)}(z) =&\mathrm{tr}(\mathbf{D}^{-1}(z) - \mathbb{E}\mathbf{D}^{-1}(z)) = \sum_{j=1}^n \mathrm{tr}\mathbb{E}_j\mathbf{D}^{-1}(z) - \mathrm{tr}\mathbb{E}_{j-1}\mathbf{D}^{-1}(z)\\ =&\sum_{j=1}^n \mathrm{tr}\mathbb{E}_j[\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z)] - \mathrm{tr}\mathbb{E}_{j-1}[\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z)] = -\sum_{j=1}^n(\mathbb{E}_j - \mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j^\top\mathbf{D}_j^{-2}\mathbf{r}_j.\end{aligned}$$

Writing $\beta_j(z) = \bar{\beta}_j(z) - \bar{\beta}_j(z)\beta_j(z)\varepsilon_j(z) = \bar{\beta}_j(z) - \bar{\beta}_j^2\varepsilon_j(z) + \bar{\beta}_j^2(z)\beta_j(z)\varepsilon_j^2(z)$, we have

$$\begin{aligned}(\mathbb{E}_j - \mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j^\top\mathbf{D}_j^{-2}\mathbf{r}_j =&(\mathbb{E}_j - \mathbb{E}_{j-1})\left(\bar{\beta}_j(z)\delta_j(z) - \bar{\beta}_j^2(z)\varepsilon_j(z)\mathbf{r}_j^\top\mathbf{D}_j^{-2}(z)\mathbf{r}_j + \bar{\beta}_j^2(z)\beta_j(z)\varepsilon_j^2(z)\mathbf{r}_j^\top\mathbf{D}_j^{-2}(z)\mathbf{r}_j\right)\\ =&\frac{d}{dz}\mathbb{E}_j\bar{\beta}_j(z)\varepsilon_j(z) - (\mathbb{E}_j - \mathbb{E}_{j-1})\bar{\beta}_j^2(z)\left(\varepsilon_j(z)\delta_j(z) - \beta_j(z)\varepsilon_j^2(z)\mathbf{r}_j^\top\mathbf{D}_j^{-2}(z)\mathbf{r}_j\right).\end{aligned}$$

Notice that

$$\mathbb{E}\left|\sum_{j=1}^{n}(\mathbb{E}_j-\mathbb{E}_{j-1})\bar{\beta}_j^2(z)\varepsilon_j(z)\delta_j(z)\right|^2=\sum_{j=1}^{n}\mathbb{E}|(\mathbb{E}_j-\mathbb{E}_{j-1})\bar{\beta}_j^2(z)\varepsilon_j(z)\delta_j(z)|^2$$
$$\le 4\sum_{j=1}^{n}\mathbb{E}|\bar{\beta}_j^2(z)\varepsilon_j(z)\delta_j(z)|^2\le\frac{4|z|^4}{v^4}\sum_{j=1}^{n}\frac{\xi_j^8}{n^4}\mathbb{E}^{\frac{1}{2}}|\varepsilon_{j0}(z)|^4\mathbb{E}^{\frac{1}{2}}|\delta_{j0}(z)|^4,$$

which is $o(1)$ from Lemma 3. Similarly, $\mathbb{E}|\sum_{j=1}^{n}(\mathbb{E}_j-\mathbb{E}_{j-1})\bar{\beta}_j^2(z)\beta_j(z)\varepsilon_j^2(z)\mathbf{r}_j^\top\mathbf{D}_j^{-2}(z)\mathbf{r}_j|^2=o(1)$. Thus, we obtain

$$\sum_{j=1}^{n}(\mathbb{E}_j-\mathbb{E}_{j-1})\bar{\beta}_j^2(z)\left(\varepsilon_j(z)\delta_j(z)+\beta_j(z)\varepsilon_j^2(z)\mathbf{r}_j^\top\mathbf{D}_j^{-2}(z)\mathbf{r}_j\right)=o_p(1),$$

which implies that we need to consider only the limiting distribution of

$$-\frac{d}{dz}\sum_{j=1}^{n}\mathbb{E}_j\bar{\beta}_j(z)\varepsilon_j(z)=-\frac{d}{dz}\sum_{j=1}^{n}(\mathbb{E}_j-\mathbb{E}_{j-1})\bar{\beta}_j(z)\varepsilon_j(z)$$

in finite-dimensional situations. For any $\epsilon>0$,

$$\sum_{j=1}^{n}\mathbb{E}\left|\mathbb{E}_j\frac{d}{dz}\varepsilon_j(z)\bar{\beta}_j(z)\right|^2 I_{\left(\left|\mathbb{E}_j\frac{d}{dz}\varepsilon_j(z)\bar{\beta}_j(z)\right|\ge\epsilon\right)}\le\frac{1}{\epsilon^2}\sum_{j=1}^{n}\mathbb{E}\left|\mathbb{E}_j\frac{d}{dz}\varepsilon_j(z)\bar{\beta}_j(z)\right|^4$$
$$\le\frac{K}{\varepsilon^2}\sum_{j=1}^{n}\frac{\xi_j^8}{n^4}\left(\frac{|z|^4\mathbb{E}|\delta_{j0}(z)|^4}{v^4}+\frac{|z|^8p^4\mathbb{E}|\varepsilon_{j0}(z)|^4}{v^{16}n^4}\right),$$

which tends to zero according to Lemma 3 and thus verifies the Lyapunov condition. Therefore, from the martingale CLT (Lemma 7), the random vector $(M_n^{(1)}(z_j))$ tends to a $2r$-dimensional Gaussian vector $(M(z_j))$ with covariance function

$$\mathrm{Cov}(M(z_1),M(z_2))=\lim_{n\to\infty}\sum_{j=1}^{n}\frac{\partial^2}{\partial z_1\partial z_2}\mathbb{E}_{j-1}\left(\mathbb{E}_j\varepsilon_j(z_1)\bar{\beta}_j(z_1)\cdot\mathbb{E}_j\varepsilon_j(z_2)\bar{\beta}_j(z_2)\right)\tag{23}$$

if this limit exists. By the same arguments as on page 571 of [4], it is sufficient to show that

$$\sum_{j=1}^{n}\mathbb{E}_{j-1}\prod_{k=1}^{2}\mathbb{E}_j\bar{\beta}_j(z_k)\varepsilon_j(z_k)\tag{24}$$

converges in probability. Using (22), we have

$$\mathbb{E}|\bar{\beta}_j(z)-b_j(z)|^2=\frac{\xi_j^4}{p^2}|b_j(z)|^2n^{-2}\mathbb{E}|\bar{\beta}_j(z)(\mathrm{tr}\,\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z)-\mathbb{E}\mathrm{tr}\,\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z))|^2$$
$$\le\frac{K|z|^4}{n^2v^4}\mathbb{E}\left|\sum_{k\ne j}(\mathbb{E}_k-\mathbb{E}_{k-1})\mathrm{tr}(\mathbf{D}_j^{-1}(z)-\mathbf{D}_{jk}^{-1}(z))\right|^2\le\frac{K|z|^4}{n^2v^4}\mathbb{E}\sum_{k\ne j}\left|\mathrm{tr}(\mathbf{D}_j^{-1}(z)-\mathbf{D}_{jk}^{-1}(z))\right|^2\le\frac{K|z|^4}{v^6n},$$

which implies that we can replace $\bar{\beta}_j(z)$ with $b_j(z)$ in (24) and study the convergence of

$$\sum_{j=1}^{n}b_j(z_1)b_j(z_2)\mathbb{E}_{j-1}\left(\mathbb{E}_j\varepsilon_j(z_1)\mathbb{E}_j\varepsilon_j(z_2)\right),\tag{25}$$

whose second mixed partial derivative yields the limit of (23). From Lemma 4, we know that

$$(25)=\frac{2p}{p+2}(T_1-T_2),\tag{26}$$

where

$$T_1=\frac{1}{n^2}\sum_{j=1}^{n}b_j(z_1)b_j(z_2)\frac{\xi_j^4}{p^2}\mathrm{tr}\left[\mathbb{E}_j\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z_1)\mathbb{E}_j(\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z_2))\right],$$
$$T_2=\frac{1}{pn^2}\sum_{j=1}^{n}b_j(z_1)b_j(z_2)\frac{\xi_j^4}{p^2}\mathrm{tr}\left[\mathbb{E}_j\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z_1)\right]\mathrm{tr}\left[\mathbb{E}_j\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z_2)\right].$$

From the definition of $(m_0(z), s_{10}(z), s_{20}(z))$, it is easy to verify the following relationship between $s_{10}(z)$ and $s_{20}(z)$:

$$zs_{10}(z) = -\int \frac{c_n t dH_p(t)}{1 + ts_{20}(z)}, \quad zs_{20}(z) = -\int \frac{t d\tilde{H}_n(t)}{1 + ts_{10}(z)}. \tag{27}$$

From the discussions in Chapter 4 in [26] and (22), we may conclude that for any $z \in \mathcal{C}_n$,

$$\frac{1}{n}\mathbb{E}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z) - s_{10}(z) \xrightarrow{a.s.} 0, \quad \frac{1}{np}\sum_{i \neq j} \xi_i^2 b_i(z) + zs_{20}(z) \xrightarrow{a.s.} 0. \tag{28}$$

Based on the strategy in [10], we obtain

$$T_1 = \frac{1}{n}\sum_{j=1}^{n} \frac{a_{n2}(z_1, z_2) b_j(z_1) b_j(z_2)(\xi_j^4/p^2)}{1 - n^{-1}\sum_{i<j} a_{n2}(z_1, z_2) b_i(z_1) b_i(z_2)(\xi_i^4/p^2)} + o_p(1) \xrightarrow{i.p.} \int_0^{a(z_1,z_2)} \frac{1}{1-z} dz = -\ln(1 - a(z_1, z_2)) \tag{29}$$

and

$$T_2 = c_n^{-1} a_{n1}(z_1, z_2) s_{10}(z_1) s_{10}(z_2) + o_p(1) \xrightarrow{i.p.} \frac{(z_1 s_2(z_1) - z_2 s_2(z_2)) s_1(z_1) s_1(z_2)}{c(s_1(z_1) - s_1(z_2))} = b(z_1, z_2). \tag{30}$$

We finally obtain

$$\mathrm{Cov}(M(z_1), M(z_2)) = -\frac{2\partial^2}{\partial z_1 \partial z_2}\left[\ln(1 - a(z_1, z_2)) + b(z_1, z_2)\right].$$

Tightness of $M_n^{(1)}(z)$: From [4], the tightness of $M_n^{(1)}(z)$ can be established by verifying the moment condition (12.51) of [5], i.e.,

$$\sup_{n, z_1, z_2 \in \mathcal{C}_n} \frac{\mathbb{E}|M_n^{(1)}(z_1) - M_n^{(1)}(z_2)|^2}{|z_1 - z_2|^2} < \infty. \tag{31}$$

The proof of (31) is analogous to pages 592–593 in [4] and Lemma 5.1 in [11]; thus, we omit the details of this proof here.

Convergence of $M_n^{(2)}(z)$: We first claim that the moments of $\mathbf{D}^{-1}(z)$, $\mathbf{D}_j^{-1}(z)$ and $\mathbf{D}_{ij}^{-1}(z)$ are all bounded in n and $z \in \mathcal{C}_n$. Taking $\mathbf{D}^{-1}(z)$ as an example, it is clear that $\mathbb{E}\|\mathbf{D}^{-1}(z)\|^q < 1/v_0^q$ for $z \in \mathcal{C}_u$. For $z \in \mathcal{C}_l \cup \mathcal{C}_r$, applying Lemma 8 with suitably large s yields

$$\mathbb{E}\|\mathbf{D}^{-1}(z)\|^q \le K_1 + \frac{1}{v^q}\Pr(\|\mathbf{B}_n\| > \eta_r \text{ or } \lambda_{\min}^{\mathbf{B}_n} < \eta_l) \le K_1 + K_2 n^q \varepsilon^{-q} n^{-s} \le K,$$

where the two constants η_r and η_l satisfy $\limsup_{n,p\to\infty} \max_{1\le j\le n}\{\xi_j^2/p\}\lambda_{\max}^{\boldsymbol{\Sigma}}(1+\sqrt{c})^2 < \eta_r < x_r$ and $x_l < \eta_l < \liminf_{n,p\to\infty} \min_{1\le j\le n}\{\xi_j^2/p\}\lambda_{\min}^{\boldsymbol{\Sigma}} I_{(0,1)}(c)(1-\sqrt{c})^2$. Therefore, for any positive q, we may assume that

$$\max\{\mathbb{E}\|\mathbf{D}^{-1}(z)\|^q, \mathbb{E}\|\mathbf{D}_j^{-1}(z)\|^q, \mathbb{E}\|\mathbf{D}_{ij}^{-1}(z)\|^q\} \le K_q. \tag{32}$$

Using the above argument, we can extend the inequality in Lemma 3 to

$$\left|\mathbb{E}\left[a(v)\prod_{l=1}^{q}\left(\mathbf{u}_1^\top \mathbf{A}^\top \mathbf{B}_l(v)\mathbf{A}\mathbf{u}_1 - \frac{1}{p}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{B}_l(v)\right)\right]\right| \le Kp^{-q/2}, \tag{33}$$

where the matrices $\mathbf{B}_l(v)$ are independent of $\mathbf{u}_1$ and

$$\max\{|a(v)|, \|\mathbf{B}_l(v)\|\} \le K\left[1 + p^s I\left(\|\mathbf{B}_l(v)\| \ge \eta_r \text{ or } \lambda_{\min}^{\tilde{\mathbf{B}}} \le \eta_l\right)\right] \tag{34}$$

with some positive s and $\tilde{\mathbf{B}}$ being $\mathbf{B}_n$ or $\mathbf{B}_n$ with the exclusion of some $\mathbf{r}_j$. In applications of (33), $a(v)$ can be a product of the factors $\beta_1(z)$ or $\mathbf{r}_1^\top \mathbf{D}_1^{-1}(z_1)\mathbf{D}_1^{-1}(z_2)\mathbf{r}_1$ or similar terms. It is easy to verify that these terms satisfy (34) (see pages 579 and 580 in [4] for further details).

Let

$$\gamma_j(z) = \mathbf{r}_j^\top \mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{\xi_j^2}{np}\mathbb{E}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z).$$

Now, we show the bounds of the moments of $\gamma_j(z)$. Using Lemma 6, (33) and the Hölder inequality, for even q, we have

$$\begin{aligned}\mathbb{E}|\gamma_j(z)-\varepsilon_j(z)|^q &= \left(\frac{\xi_j^2}{p}\right)^q \mathbb{E}\left|\frac{1}{n}\sum_{i\neq j}(\mathbb{E}_i-\mathbb{E}_{i-1})\mathrm{tr}\,\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z)\right|^q = \left(\frac{\xi_j^2}{p}\right)^q \mathbb{E}\left|\frac{1}{n}\sum_{i\neq j}(\mathbb{E}_i-\mathbb{E}_{i-1})\beta_{ij}(z)\mathbf{r}_i^\top\mathbf{D}_{ij}^{-1}(z)\boldsymbol{\Sigma}\mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i\right|^q \\ &\le \frac{K}{n^q}\mathbb{E}\left[\sum_{i\neq j}\left|(\mathbb{E}_i-\mathbb{E}_{i-1})\beta_{ij}(z)\mathbf{r}_i^\top\mathbf{D}_{ij}^{-1}(z)\boldsymbol{\Sigma}\mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i\right|^2\right]^{q/2} \\ &\le \frac{K}{n^{1+q/2}}\sum_{i\neq j}\mathbb{E}\left|(\mathbb{E}_i-\mathbb{E}_{i-1})\beta_{ij}(z)\mathbf{r}_i^\top\mathbf{D}_{ij}^{-1}(z)\boldsymbol{\Sigma}\mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i\right|^q \le \frac{K}{n^{q/2}}, \end{aligned} \tag{35}$$

where the last inequality uses the boundedness of $\mathbb{E}|\beta_{ij}(z)|^q$ and $\mathbb{E}|\mathbf{r}_i^\top\mathbf{D}_{ij}^{-1}(z)\boldsymbol{\Sigma}\mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i|^q$. From (33) and (35), we obtain

$$\mathbb{E}|\varepsilon_j(z)|^q \le Kn^{-q/2}, \quad \mathbb{E}|\gamma_j(z)|^q \le Kn^{-q/2} \tag{36}$$

for even q.

Next, we show that $b_j(z)$ is bounded for all n. By the equality $b_j(z)-\beta_j(z)=b_j(z)\beta_j(z)\gamma_j(z)$ and the boundedness of $\mathbb{E}|\beta_j(z)|^q$ and $\mathbb{E}|\gamma_j|^q$, we have

$$|b_j(z)| = |\mathbb{E}\beta_j(z)+\mathbb{E}\beta_j(z)b_j(z)\gamma_j(z)| \le K_1 + K_2|b_j(z)|n^{-1/2}.$$

Thus, for all large n,

$$|b_j(z)| \le \frac{K_1}{1-K_2n^{-1/2}} < K. \tag{37}$$

We now consider the convergence of $M_n^{(2)}(z)$. Let $\bar{b}_j(z) = 1/[1+\xi_j^2\mathbb{E}\mathrm{tr}\,\boldsymbol{\Sigma}\mathbf{D}^{-1}(z)/(np)]$, by which we define two matrices:

$$\mathbf{V}(z) = z\mathbf{I} - \frac{1}{np}\sum_{j=1}^n \xi_j^2\bar{b}_j(z)\boldsymbol{\Sigma}, \quad \underline{\mathbf{V}}(z) = \mathrm{diag}(z\bar{b}_1^{-1}(z),\ldots,z\bar{b}_n^{-1}(z)). \tag{38}$$

Similar to [4], we can verify that

$$\sup_{n,z\in\mathcal{C}_n}\|\mathbf{V}^{-1}(z)\| < \infty, \quad \sup_{n,z\in\mathcal{C}_n}\|\underline{\mathbf{V}}^{-1}(z)\| < \infty. \tag{39}$$

In addition, for any nonrandom $p\times p$ matrix $\mathbf{M}$,

$$\sup_{n,z\in\mathcal{C}_n}\mathbb{E}\left|\mathrm{tr}\mathbf{D}^{-1}(z)\mathbf{M}-\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}(z)\mathbf{M}\right|^2 \le K\|\mathbf{M}\|^2. \tag{40}$$

From (38), we can obtain two decompositions of $M_n^{(2)}(z)$, i.e.,

$$M_n^{(2)}(z) = [p\mathbb{E}m_n(z)+\mathrm{tr}\mathbf{V}^{-1}(z)] - [\mathrm{tr}\mathbf{V}^{-1}(z)+pm_0(z)] \triangleq P_n(z)-Q_n(z) \tag{41}$$

$$= [n\mathbb{E}\underline{m}_n(z)+\mathrm{tr}\underline{\mathbf{V}}^{-1}(z)] - [\mathrm{tr}\underline{\mathbf{V}}^{-1}(z)+n\underline{m}_0(z)] \triangleq \underline{P}_n(z)-\underline{Q}_n(z). \tag{42}$$

Notice that

$$\begin{aligned} Q_n(z) &= p\int\frac{dH_p(t)}{z-\frac{1}{np}\sum_{j=1}^n\xi_j^2\bar{b}_j(z)t} - p\int\frac{dH_p(t)}{z+zs_{20}(z)t} = \int\frac{\left[\frac{1}{n}\sum_{j=1}^n\xi_j^2\bar{b}_j(z)+pzs_{20}(z)\right]tdH_p(t)}{(z-\frac{1}{np}\sum_{j=1}^n\xi_j^2\bar{b}_j(z)t)(z+zs_{20}(z)t)}, \\ \underline{Q}_n(z) &= n\int\frac{d\tilde{H}_n(t)}{z+z\frac{1}{n}\mathbb{E}\mathrm{tr}\,\boldsymbol{\Sigma}\mathbf{D}^{-1}(z)t} - n\int\frac{d\tilde{H}_n(t)}{z+zs_{10}(z)t} = \int\frac{\left[nzs_{10}(z)-z\mathbb{E}\mathrm{tr}\,\boldsymbol{\Sigma}\mathbf{D}^{-1}(z)\right]td\tilde{H}_n(t)}{(z+z\frac{1}{n}\mathbb{E}\mathrm{tr}\,\boldsymbol{\Sigma}\mathbf{D}^{-1}(z)t)(z+zs_{10}(z)t)}. \end{aligned}$$

From these, together with the equality

$$\frac{1}{np}\sum_{j=1}^n\xi_j^2\bar{b}_j(z)+zs_{20}(z) = \left[s_{10}(z)-\frac{1}{n}\mathbb{E}\mathrm{tr}\,\boldsymbol{\Sigma}\mathbf{D}^{-1}(z)\right]\int\frac{t^2d\tilde{H}_n(t)}{(1+\frac{1}{n}\mathbb{E}\mathrm{tr}\,\boldsymbol{\Sigma}\mathbf{D}^{-1}(z)t)(1+s_{10}(z)t)},$$

we obtain

$$\frac{M_n^{(2)}(z)-P_n(z)}{M_n^{(2)}(z)-\underline{P}_n(z)}=\frac{Q_n(z)}{\underline{Q}_n(z)}=\frac{c_n}{z}\frac{\int\frac{t^2d\tilde{H}_n(t)}{(1+\frac{1}{n}\mathbb{E}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}^{-1}(z)t)(1+s_{10}(z)t)}\int\frac{tdH_p(t)}{(z-\frac{1}{np}\sum_{j=1}^n\xi_j^2\bar{b}_j(z)t)(z+zs_{20}(z)t)}}{\int\frac{td\tilde{H}_n(t)}{(z+z\frac{1}{n}\mathbb{E}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}^{-1}(z)t)(z+zs_{10}(z)t)}}$$
$$=\frac{c_n}{z}\frac{\int\frac{t^2d\tilde{H}_n(t)}{(1+s_{10}(z)t)^2}\int\frac{tdH_p(t)}{(1+s_{20}(z)t)^2}}{\int\frac{td\tilde{H}_n(t)}{(1+s_{10}(z)t)^2}}+o_{a.s.}(1)\triangleq R_n(z)+o_{a.s.}(1),$$

where the third equality uses analogous convergence in (28).

Our next task is to study the limits of $P_n(z)$ and $\underline{P}_n(z)$. For simplicity, we suppress the independent variable z from expressions in the following. All expressions and convergence statements hold uniformly for $z\in\mathcal{C}_n$. We first simplify the expression of P_n. Using the identity $\mathbf{r}_j^\top\mathbf{D}^{-1}=\mathbf{r}_j^\top\mathbf{D}_j^{-1}\beta_j$, we have

$$\begin{aligned}P_n=&\mathbb{E}\mathrm{tr}(\mathbf{D}^{-1}+\mathbf{V}^{-1})=\mathbb{E}\mathrm{tr}\left[\mathbf{V}^{-1}\left(\sum_{j=1}^n\mathbf{r}_j\mathbf{r}_j^\top-\frac{1}{np}\sum_{j=1}^n\xi_j^2\bar{b}_j\boldsymbol{\Sigma}\right)\mathbf{D}^{-1}\right]\\=&\sum_{j=1}^n\mathbb{E}\beta_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j-\frac{1}{np}\sum_{j=1}^n\xi_j^2\bar{b}_j\mathbb{E}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}^{-1}\mathbf{V}^{-1}.\end{aligned}\tag{43}$$

From (21) and $\beta_j=b_j-b_j\beta_j\gamma_j$,

$$\bar{b}_j-b_j=\bar{b}_jb_j\frac{\xi_j^2}{np}\mathbb{E}\mathrm{tr}\boldsymbol{\Sigma}(\mathbf{D}_j^{-1}-\mathbf{D}^{-1})=b_j\bar{b}_j\frac{\xi_j^2}{np}\mathbb{E}\beta_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\mathbf{r}_j=b_j^2\bar{b}_j\frac{\xi_j^2}{np}\mathbb{E}(1-\beta_j\gamma_j)\mathbf{r}_j^\top\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\mathbf{r}_j,\tag{44}$$
$$\mathbb{E}\mathrm{tr}\mathbf{V}^{-1}\boldsymbol{\Sigma}(\mathbf{D}_j^{-1}-\mathbf{D}^{-1})=\mathbb{E}\mathrm{tr}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\mathbf{r}_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\beta_j=b_j\mathbb{E}(1-\beta_j\gamma_j)\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\mathbf{r}_j.$$

From (33) and (36), for $\mathbf{M}=\boldsymbol{\Sigma}$ or $\mathbf{M}=\mathbf{V}^{-1}\boldsymbol{\Sigma}$,

$$|\mathbb{E}\beta_j\gamma_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{M}\mathbf{D}_j^{-1}\mathbf{r}_j|\le Kn^{-1/2},\quad \mathbb{E}\beta_j=b_j+O(n^{-1/2}).\tag{45}$$

Combining (43)–(45), we obtain

$$\begin{aligned}P_n=&\sum_{j=1}^n\mathbb{E}\beta_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j-\frac{1}{np}\sum_{j=1}^n\xi_j^2b_j\mathbb{E}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\\&-\frac{1}{(np)^3}\sum_{j=1}^n\xi_j^6b_j^3\mathbb{E}\mathrm{tr}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbb{E}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\mathbf{V}^{-1}+\frac{1}{(np)^2}\sum_{j=1}^n\xi_j^4b_j^2\mathbb{E}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}+o(1).\end{aligned}$$

Substituting $\beta_j=b_j-b_j^2\gamma_j+b_j^3\gamma_j^2-\beta_jb_j^3\gamma_j^3$ into the first term in the above equation, we obtain

$$\begin{aligned}\sum_{j=1}^n\mathbb{E}\beta_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j=&\frac{1}{np}\sum_{j=1}^n\xi_j^2b_j\mathbb{E}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}-\sum_{j=1}^nb_j^2\mathbb{E}\gamma_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j\\&+\sum_{j=1}^nb_j^3\mathbb{E}\gamma_j^2\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j-\sum_{j=1}^nb_j^3\mathbb{E}\gamma_j^3\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j.\end{aligned}$$

Note that, from (33), (36) and (40),

$$\begin{aligned}&\sum_{j=1}^n\mathbb{E}\gamma_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j=\sum_{j=1}^n\mathbb{E}\left[\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{r}_j-\frac{\xi_j^2}{np}\mathrm{tr}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\right]\left[\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j-\frac{\xi_j^2}{np}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\right]\\&+\frac{1}{n^2p^2}\sum_{j=1}^n\xi_j^4\mathrm{Cov}(\mathrm{tr}\mathbf{D}_j^{-1}\boldsymbol{\Sigma},\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma})=\sum_{j=1}^n\mathbb{E}\left[\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{r}_j-\frac{\xi_j^2}{np}\mathrm{tr}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\right]\left[\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j-\frac{\xi_j^2}{np}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\right]+o(1),\\&\sum_{j=1}^n\mathbb{E}\gamma_j^2\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j=\sum_{j=1}^n\mathbb{E}\gamma_j^2\left[\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j-\frac{\xi_j^2}{np}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\right]\\&+\frac{1}{np}\sum_{j=1}^n\xi_j^2\mathrm{Cov}(\gamma_j^2,\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma})+\frac{1}{np}\sum_{j=1}^n\xi_j^2\mathbb{E}\gamma_j^2\mathbb{E}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}=\frac{1}{np}\sum_{j=1}^n\xi_j^2\mathbb{E}\gamma_j^2\mathbb{E}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}+o(1),\end{aligned}$$

and

$$\sum_{j=1}^{n}\mathbb{E}\gamma_j^3\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j=o(1).$$

We thus arrive at

$$P_n=-\sum_{j=1}^{n}b_j^2\mathbb{E}\left[\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{r}_j-\frac{\xi_j^2}{np}\mathrm{tr}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\right]\left[\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{V}^{-1}\mathbf{r}_j-\frac{\xi_j^2}{np}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\right]+\frac{1}{np}\sum_{j=1}^{n}b_j^3\xi_j^2\mathbb{E}\gamma_j^2\mathbb{E}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}$$
$$-\frac{1}{(np)^3}\sum_{j=1}^{n}\xi_j^6b_j^3\mathbb{E}\mathrm{tr}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbb{E}\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\mathbf{V}^{-1}+\frac{1}{(np)^2}\sum_{j=1}^{n}\xi_j^4b_j^2\mathbb{E}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}+o(1).$$

On the other hand, according to the identity $\mathbf{r}_j^\top\mathbf{D}^{-1}=\mathbf{r}_j^\top\mathbf{D}_j^{-1}\beta_j$, we have

$$p+z\mathrm{tr}\mathbf{D}^{-1}=\mathrm{tr}(\mathbf{B}_n\mathbf{D}^{-1})=\sum_{j=1}^{n}\beta_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{r}_j=n-\sum_{j=1}^{n}\beta_j,$$

which implies $nz\underline{m}_n=-\sum_{j=1}^{n}\beta_j$. From this, together with $\beta_j=b_j-b_j^2\gamma_j+b_j^3\gamma_j^2-\beta_jb_j^3\gamma_j^3$, (33) and (44), we obtain

$$\underline{P}_n=-\frac{1}{z}\sum_{j=1}^{n}\mathbb{E}\left(\beta_j-\bar{b}_j\right)=-\frac{1}{z}\sum_{j=1}^{n}b_j^3\mathbb{E}\gamma_j^2+\frac{1}{z(np)^2}\sum_{j=1}^{n}b_j^3\xi_j^4\mathbb{E}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}+o(1).$$

Applying Lemma 4 to the simplified P_n and $\underline{P}_n$ yields

$$P_n=\frac{-1}{(np)^2}\sum_{j=1}^{n}\xi_j^4b_j^2\left[\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}-\frac{2}{p}\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathrm{tr}\mathbf{D}^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\right]$$
$$+\frac{1}{(np)^3}\sum_{j=1}^{n}\xi_j^6b_j^3\left[\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma}-\frac{2}{p}\mathbb{E}^2\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\right]\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}+o(1),\tag{46}$$
$$\underline{P}_n=\frac{-1}{z(np)^2}\sum_{j=1}^{n}\xi_j^4b_j^3\left[\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma}-\frac{2}{p}\mathbb{E}^2\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\right]+o(1).\tag{47}$$

We replace $p+2$ and $\mathbf{D}_j$ with p and $\mathbf{D}$, respectively, in the derived results.

To study the limits of P_n and $\underline{P}_n$ in (46) and (47), we need to reconsider the sum of $\mathbf{D}^{-1}$ and $\mathbf{V}^{-1}$. Similar to (43),

$$\begin{aligned}\mathbf{D}^{-1}+\mathbf{V}^{-1}&=\sum_{j=1}^{n}\beta_j\mathbf{V}^{-1}\mathbf{r}_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}-\frac{1}{np}\sum_{j=1}^{n}\xi_j^2b_j\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\\&=\sum_{j=1}^{n}b_j\mathbf{V}^{-1}\left(\mathbf{r}_j\mathbf{r}_j^\top-\frac{\xi_j^2}{np}\boldsymbol{\Sigma}\right)\mathbf{D}_j^{-1}+(b_j-\beta_j)\mathbf{V}^{-1}\mathbf{r}_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}+\frac{1}{np}\sum_{j=1}^{n}\xi_j^2b_j\mathbf{V}^{-1}\boldsymbol{\Sigma}(\mathbf{D}_j^{-1}-\mathbf{D}^{-1})\\&\triangleq\mathbf{A}_1+\mathbf{A}_2+\mathbf{A}_3.\end{aligned}\tag{48}$$

Let $\mathbf{M}$ be a $p\times p$ matrix. From (32), (37), (39), (40) and (33), we have

$$|\mathbb{E}\mathrm{tr}\mathbf{A}_2\mathbf{M}|\le n^{1/2}K(\mathbb{E}\|\mathbf{M}\|^4)^{1/4},\quad|\mathrm{tr}\mathbf{A}_3\mathbf{M}|\le K(\mathbb{E}\|\mathbf{M}\|^2)^{1/2}.\tag{49}$$

In the following, $\mathbf{M}$ is a nonrandom matrix with bounded spectral norm. Similar to the above inequalities, we have $|\mathbb{E}\mathrm{tr}\mathbf{A}_1\mathbf{M}|\le n^{1/2}K$. We write

$$\mathrm{tr}\mathbf{A}_1\boldsymbol{\Sigma}\mathbf{D}^{-1}\mathbf{M}=\sum_{j=1}^{n}b_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\boldsymbol{\Sigma}(\mathbf{D}^{-1}-\mathbf{D}_j^{-1})\mathbf{M}\mathbf{V}^{-1}\mathbf{r}_j+\sum_{j=1}^{n}b_j\left(\mathbf{r}_j^\top\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\mathbf{M}\mathbf{V}^{-1}\mathbf{r}_j-\frac{\xi_j^2}{np}\mathrm{tr}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\mathbf{M}\right)$$
$$+\frac{1}{np}\sum_{j=1}^{n}b_j\xi_j^2\mathrm{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}(\mathbf{D}_j^{-1}-\mathbf{D}^{-1})\mathbf{M}\mathbf{V}^{-1}\triangleq A_{11}+A_{12}+A_{13}.\tag{50}$$

It is clear that $\mathbb{E}A_{12}=0$ and that, from (40), $|\mathbb{E}A_{13}|\le K$. For the term A_{11}, using (21), (32), (33), (40) and (45), we obtain

$$n^{-1}\mathbb{E}A_{11}=-\frac{1}{n}\sum_{j=1}^{n}b_j\mathbb{E}\beta_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\mathbf{r}_j\mathbf{r}_j^\top\mathbf{D}_j^{-1}\mathbf{M}\mathbf{V}^{-1}\mathbf{r}_j=-\frac{1}{n^3p^2}\sum_{j=1}^{n}b_j^2\xi_j^4\mathbb{E}(\mathrm{tr}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\boldsymbol{\Sigma})(\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{M}\mathbf{V}^{-1}\boldsymbol{\Sigma})+o(1)$$

$$
\begin{aligned}
&= -\frac{1}{n^3p^2}\sum_{j=1}^{n} b_j^2\xi_j^4\mathbb{E}(\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma})(\mathrm{tr}\mathbf{D}^{-1}\mathbf{M}\mathbf{V}^{-1}\boldsymbol{\Sigma}) + o(1)\\
&= -\frac{1}{n^3p^2}\sum_{j=1}^{n} b_j^2\xi_j^4\mathbb{E}(\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma})\mathbb{E}(\mathrm{tr}\mathbf{D}^{-1}\mathbf{M}\mathbf{V}^{-1}\boldsymbol{\Sigma}) + o(1).
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma} &= -\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma} - \frac{1}{n^3p^2}\sum_{j=1}^{n} b_j^2\xi_j^4\mathbb{E}(\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma})\mathbb{E}(\mathrm{tr}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma}) + o(1)\\
&= -\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma}\left[1 + \frac{1}{n^2p^2}\sum_{j=1}^{n} b_j^2\xi_j^4\mathbb{E}(\mathrm{tr}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma})\right]^{-1} + o(1) \qquad (51)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma} &= -\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\left[1 + \frac{1}{n^2p^2}\sum_{j=1}^{n} b_j^2\xi_j^4\mathbb{E}(\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma})\right] + o(1)\\
&= -\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\left[1 + \frac{1}{n^2p^2}\sum_{j=1}^{n} b_j^2\xi_j^4\mathbb{E}(\mathrm{tr}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma})\right]^{-1} + o(1). \qquad (52)
\end{aligned}
$$

From (28) and (48)–(52), we obtain

$$
\begin{aligned}
&\frac{1}{np^2}\sum_{j=1}^{n} b_j^2\xi_j^4 = \int\frac{t^2d\tilde{H}_n(t)}{(1+s_{10}t)^2} + o(1), \quad \frac{1}{np^2}\sum_{j=1}^{n} b_j^3\xi_j^4 = \int\frac{t^2d\tilde{H}_n(t)}{(1+s_{10}t)^3} + o(1),\\
&\frac{1}{np^3}\sum_{j=1}^{n} b_j^3\xi_j^6 = \int\frac{t^3d\tilde{H}_n(t)}{(1+s_{10}t)^3} + o(1), \quad \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma} = -\int\frac{c_ntdH_p(t)}{z(1+s_{20}t)} + o(1) = s_{10} + o(1),\\
&\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma} = -\int\frac{c_ntdH_p(t)}{z^2(1+s_{20}t)^2} + o(1),\\
&\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma} = \int\frac{c_nt^2dH_p(t)}{z^2(1+s_{20}t)^2}\left[1 - \int\frac{c_nt^2dH_p(t)}{z^2(1+s_{20}t)^2}\int\frac{t^2d\tilde{H}_n(t)}{(1+s_{10}t)^2}\right]^{-1} + o(1),\\
&\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma} = \int\frac{c_nt^2dH_p(t)}{z^3(1+s_{20}t)^3}\left[1 - \int\frac{c_nt^2dH_p(t)}{z^2(1+s_{20}t)^2}\int\frac{t^2d\tilde{H}_n(t)}{(1+s_{10}t)^2}\right]^{-1} + o(1).
\end{aligned}
$$

Combining the above results with (46) and (47), we obtain

$$
\begin{aligned}
P_n = &-\int\frac{t^2d\tilde{H}_n(t)}{(1+s_{10}t)^2}\int\frac{c_nt^2dH_p(t)}{z^3(1+s_{20}t)^3}\left[1 - \int\frac{c_nt^2dH_p(t)}{z^2(1+s_{20}t)^2}\int\frac{t^2d\tilde{H}_n(t)}{(1+s_{10}t)^2}\right]^{-1}\\
&-\int\frac{t^3d\tilde{H}_n(t)}{(1+s_{10}t)^3}\int\frac{c_ntdH_p(t)}{z^2(1+s_{20}t)^2}\int\frac{c_nt^2dH_p(t)}{z^2(1+s_{20}t)^2}\left[1 - \int\frac{c_nt^2dH_p(t)}{z^2(1+s_{20}t)^2}\int\frac{t^2d\tilde{H}_n(t)}{(1+s_{10}t)^2}\right]^{-1}\\
&-\int\frac{2s_{10}t^2d\tilde{H}_n(t)}{(1+s_{10}t)^2}\int\frac{tdH_p(t)}{z^2(1+s_{20}t)^2} + \int\frac{2s_{10}^2t^3d\tilde{H}_n(t)}{(1+s_{10}t)^3}\int\frac{tdH_p(t)}{z^2(1+s_{20}t)^2} + o(1),\\
\underline{P}_n = &-\int\frac{t^2d\tilde{H}_n(t)}{z(1+s_{10}t)^3}\int\frac{c_nt^2dH_p(t)}{z^2(1+s_{20}t)^2}\left[1 - \int\frac{c_nt^2dH_p(t)}{z^2(1+s_{20}t)^2}\int\frac{t^2d\tilde{H}_n(t)}{(1+s_{10}t)^2}\right]^{-1} + \int\frac{2s_{10}^2t^2d\tilde{H}_n(t)}{zc_n(1+s_{10}t)^3} + o(1).
\end{aligned}
$$

Therefore, we obtain

$$
M_n^{(2)}(z) = \frac{P_n - \underline{P}_nR_n}{1-R_n} \to \frac{P(z) - \underline{P}(z)R(z)}{1-R(z)}
$$

as $n \to \infty$, where $P(z)$, $Q(z)$ and $R(z)$ are defined in the lemma. We complete the proof of (i) in Lemma 1. □

Proof of (ii) in Lemma 1. Similar to the proof of the convergence of $\widehat{M}_n(z)$, the convergence of $\widetilde{M}_n(z)$ can be demonstrated by verifying the following two facts:

Fact 1: Finite-dimensional convergence of $\widetilde{M}_n(z)$ in distribution. That is, for any positive integer r and the real constants $\alpha_1, \dots, \alpha_r$, the linear combination $\sum_{i=1}^r \alpha_i \widetilde{M}_n(z_i)$ converges in distribution to a Gaussian random variable.
Fact 2: Tightness of $\widetilde{M}_n(z)$ on $\mathcal{C}$. This can be confirmed by verifying a similar moment condition as in (31).

These two facts can be demonstrated by similar arguments as in the proof of Lemma 5.2 in [11] (see pages 7178–7179 in this reference). We thus omit the details here. □

5. An approximation of the covariance matrix Ψ_3

In this section, we present an approximation of $\boldsymbol{\Psi}_3 = (\psi_{ij})_{i,j\in\{1,2,3\}}$, where

$$\begin{aligned}
&\psi_{11} = pn^{-2}\mathbb{E}(\mathrm{tr}\mathbf{B}_n)^2 - p\alpha_1^2, \quad \psi_{12} = \psi_{21} = pn^{-2}\mathbb{E}\mathrm{tr}\mathbf{B}_n^2\mathrm{tr}\mathbf{B}_n - p\alpha_1\alpha_2,\\
&\psi_{22} = pn^{-2}\mathbb{E}(\mathrm{tr}\mathbf{B}_n^2)^2 - p\alpha_2^2, \quad \psi_{23} = \psi_{32} = pn^{-2}\mathbb{E}\mathrm{tr}\mathbf{B}_n^3\mathrm{tr}\mathbf{B}_n^2 - p\alpha_2\alpha_3,\\
&\psi_{33} = pn^{-2}\mathbb{E}(\mathrm{tr}\mathbf{B}_n^3)^2 - p\alpha_3^2, \quad \psi_{13} = \psi_{31} = pn^{-2}\mathbb{E}\mathrm{tr}\mathbf{B}_n^3\mathrm{tr}\mathbf{B}_n - p\alpha_1\alpha_3.
\end{aligned}$$

Thus, the main task is to approximate $\mathbb{E}(\mathrm{tr}\mathbf{B}_n)^2$, $\mathbb{E}(\mathrm{tr}\mathbf{B}_n^2)^2$, $\mathbb{E}(\mathrm{tr}\mathbf{B}_n^3)^2$, $\mathbb{E}\mathrm{tr}\mathbf{B}_n^2\mathrm{tr}\mathbf{B}_n$, $\mathbb{E}\mathrm{tr}\mathbf{B}_n^3\mathrm{tr}\mathbf{B}_n$ and $\mathbb{E}\mathrm{tr}\mathbf{B}_n^3\mathrm{tr}\mathbf{B}_n^2$. Direct calculations yield

$$\mathbb{E}(\mathrm{tr}\mathbf{B}_n)^2 = n^{-2}\left(nw_2((\mathrm{tr}\boldsymbol{\Sigma})^2 + 2\mathrm{tr}\boldsymbol{\Sigma}^2) + n_2w_1^2(\mathrm{tr}\boldsymbol{\Sigma})^2\right)$$

and

$$\mathbb{E}(\mathrm{tr}\mathbf{B}_n^2)^2 = n^{-4}(V_1+V_2+V_3+V_4+V_5),$$

where

$$\begin{aligned}
&V_1 = nw_4\left((\mathrm{tr}\boldsymbol{\Sigma})^4 + 12(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^2\right) + o(n^4), \quad V_2 = n_2w_3w_1\left((\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^2 + 4\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma} + 2(\mathrm{tr}\boldsymbol{\Sigma}^2)^2\right) + o(n^4),\\
&V_3 = n_2w_2^2\left((\mathrm{tr}\boldsymbol{\Sigma})^4 + 4(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^2 + 10(\mathrm{tr}\boldsymbol{\Sigma}^2)^2\right) + o(n^4), \quad V_4 = n_3w_2w_1^2\left(2(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^2 + 8(\mathrm{tr}\boldsymbol{\Sigma}^2)^2 + 8\mathrm{tr}\boldsymbol{\Sigma}^4\right),\\
&V_5 = n_4w_1^4(\mathrm{tr}\boldsymbol{\Sigma}^2)^2.
\end{aligned}$$

Analogously, we have

$$\mathbb{E}(\mathrm{tr}\mathbf{B}_n^3)^2 = n^{-6}(W_1+W_2+W_3+W_4+W_5+W_6+W_7+W_8+W_9+W_{10}+W_{11}),$$

where

$$\begin{aligned}
W_1 &= nw_6\left((\mathrm{tr}\boldsymbol{\Sigma})^6 + 30(\mathrm{tr}\boldsymbol{\Sigma})^4\mathrm{tr}\boldsymbol{\Sigma}^2\right) + o(n^6), \quad W_2 = 6n_2w_5w_1\left((\mathrm{tr}\boldsymbol{\Sigma})^4\mathrm{tr}\boldsymbol{\Sigma}^2 + 8(\mathrm{tr}\boldsymbol{\Sigma})^3\mathrm{tr}\boldsymbol{\Sigma}^3 + 12(\mathrm{tr}\boldsymbol{\Sigma}^2)^2(\mathrm{tr}\boldsymbol{\Sigma})^2\right)\\
&\quad + o(n^6),\\
W_3 &= n_3w_4w_1^2\left((\mathrm{tr}\boldsymbol{\Sigma})^3\mathrm{tr}\boldsymbol{\Sigma}^3 + 9(\mathrm{tr}\boldsymbol{\Sigma}^2)^2(\mathrm{tr}\boldsymbol{\Sigma})^2 + 5(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^4 + 108\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 18(\mathrm{tr}\boldsymbol{\Sigma}^2)^3\right) + o(n^6),\\
W_4 &= n_2w_4w_2\left(6(\mathrm{tr}\boldsymbol{\Sigma})^4\mathrm{tr}\boldsymbol{\Sigma}^2 + 63(\mathrm{tr}\boldsymbol{\Sigma}^2)^2(\mathrm{tr}\boldsymbol{\Sigma})^2 + 48(\mathrm{tr}\boldsymbol{\Sigma})^3\mathrm{tr}\boldsymbol{\Sigma}^3\right) + o(n^6),\\
W_5 &= n_4w_3w_1^3\left(2(\mathrm{tr}\boldsymbol{\Sigma})^3\mathrm{tr}\boldsymbol{\Sigma}^3 + 30\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 52(\mathrm{tr}\boldsymbol{\Sigma}^3)^2 + 36\mathrm{tr}\boldsymbol{\Sigma}^4\mathrm{tr}\boldsymbol{\Sigma}^2 + 36\mathrm{tr}\boldsymbol{\Sigma}^5\mathrm{tr}\boldsymbol{\Sigma}\right) + o(n^6),\\
W_6 &= n_3w_3w_2w_1\left(6(\mathrm{tr}\boldsymbol{\Sigma})^4\mathrm{tr}\boldsymbol{\Sigma}^2 + 12(\mathrm{tr}\boldsymbol{\Sigma})^3\mathrm{tr}\boldsymbol{\Sigma}^3 + 54(\mathrm{tr}\boldsymbol{\Sigma}^2)^2(\mathrm{tr}\boldsymbol{\Sigma})^2 + 336\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 36(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^4\right)\\
&\quad + o(n^6),\\
W_7 &= n_2w_3^2\left((\mathrm{tr}\boldsymbol{\Sigma})^6 + 12(\mathrm{tr}\boldsymbol{\Sigma})^4\mathrm{tr}\boldsymbol{\Sigma}^2 + 16(\mathrm{tr}\boldsymbol{\Sigma})^3\mathrm{tr}\boldsymbol{\Sigma}^3 + 63(\mathrm{tr}\boldsymbol{\Sigma}^2)^2(\mathrm{tr}\boldsymbol{\Sigma})^2\right) + o(n^6),\\
W_8 &= n_5w_2w_1^4\left(6\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 21(\mathrm{tr}\boldsymbol{\Sigma}^3)^2 + 18\mathrm{tr}\boldsymbol{\Sigma}^6\right),\\
W_9 &= n_4w_2^2w_1^2\left(9(\mathrm{tr}\boldsymbol{\Sigma}^2)^2(\mathrm{tr}\boldsymbol{\Sigma})^2 + 54\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 108(\mathrm{tr}\boldsymbol{\Sigma}^3)^2 + 36\mathrm{tr}\boldsymbol{\Sigma}^5\mathrm{tr}\boldsymbol{\Sigma} + 18\mathrm{tr}\boldsymbol{\Sigma}^4\mathrm{tr}\boldsymbol{\Sigma}^2\right) + o(n^6),\\
W_{10} &= n_3w_2^3\left(9(\mathrm{tr}\boldsymbol{\Sigma}^2)^2(\mathrm{tr}\boldsymbol{\Sigma})^2 + 36\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 6(\mathrm{tr}\boldsymbol{\Sigma}^2)^3 + 18\mathrm{tr}\boldsymbol{\Sigma}^4(\mathrm{tr}\boldsymbol{\Sigma})^2\right) + o(n^6), \quad W_{11} = n_6w_1^6(\mathrm{tr}\boldsymbol{\Sigma}^3)^2.
\end{aligned}$$

We show the following algebraically:

$$\mathbb{E}\mathrm{tr}\mathbf{B}_n^2\mathrm{tr}\mathbf{B}_n = n^{-3}\left(nw_3((\mathrm{tr}\boldsymbol{\Sigma})^3 + 6\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 8\mathrm{tr}\boldsymbol{\Sigma}^3) + n_2w_2w_1((\mathrm{tr}\boldsymbol{\Sigma})^3 + 4\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma} + 4\mathrm{tr}\boldsymbol{\Sigma}^3) + n_3w_1^3\mathrm{tr}\boldsymbol{\Sigma}^2\mathrm{tr}\boldsymbol{\Sigma}\right)$$

and

$$\mathbb{E}\mathrm{tr}\mathbf{B}_n^3\mathrm{tr}\mathbf{B}_n = n^{-4}(U_1+U_2+U_3+U_4+U_5),$$

where

$$\begin{aligned}
&U_1 = nw_4\left((\mathrm{tr}\boldsymbol{\Sigma})^4 + 12(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^2\right) + o(n^4), \quad U_2 = n_2w_3w_1\left((\mathrm{tr}\boldsymbol{\Sigma})^4 + 9(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^2 + 20\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma} + 6(\mathrm{tr}\boldsymbol{\Sigma}^2)^2\right),\\
&U_3 = n_2w_2^2\left(3(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^2 + 6\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma} + 6(\mathrm{tr}\boldsymbol{\Sigma}^2)^2\right) + o(n^4), \quad U_4 = n_3w_2w_1^2\left(3(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^2 + 9\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma} + 6\mathrm{tr}\boldsymbol{\Sigma}^4\right),\\
&U_5 = n_4w_1^4\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma}.
\end{aligned}$$

Analogously, we have

$$\mathbb{E}\mathrm{tr}\mathbf{B}_n^3\mathrm{tr}\mathbf{B}_n^2 = n^{-5}(Z_1 + Z_2 + Z_3 + Z_4 + Z_5 + Z_6 + Z_7),$$

where

$$Z_1 = nw_5\left((\mathrm{tr}\boldsymbol{\Sigma})^5 + 20(\mathrm{tr}\boldsymbol{\Sigma})^3\mathrm{tr}\boldsymbol{\Sigma}^2\right) + o(n^5), \quad Z_2 = 5n_2w_4w_1\left((\mathrm{tr}\boldsymbol{\Sigma})^3\mathrm{tr}\boldsymbol{\Sigma}^2 + 6(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^3 + 6(\mathrm{tr}\boldsymbol{\Sigma}^2)^2\mathrm{tr}\boldsymbol{\Sigma}\right) + o(n^5),$$
$$Z_3 = n_3w_3w_1^2\left((\mathrm{tr}\boldsymbol{\Sigma})^3\mathrm{tr}\boldsymbol{\Sigma}^2 + 12(\mathrm{tr}\boldsymbol{\Sigma}^2)^2\mathrm{tr}\boldsymbol{\Sigma} + 3(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^3 + 38\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma}^2 + 24\mathrm{tr}\boldsymbol{\Sigma}^4\mathrm{tr}\boldsymbol{\Sigma}\right) + o(n^5),$$
$$Z_4 = n_2w_3w_2\left((\mathrm{tr}\boldsymbol{\Sigma})^5 + 11(\mathrm{tr}\boldsymbol{\Sigma})^3\mathrm{tr}\boldsymbol{\Sigma}^2 + 36(\mathrm{tr}\boldsymbol{\Sigma}^2)^2\mathrm{tr}\boldsymbol{\Sigma} + 26(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^3\right) + o(n^5),$$
$$Z_5 = n_4w_2w_1^3\left((\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^3 + 3(\mathrm{tr}\boldsymbol{\Sigma}^2)^2\mathrm{tr}\boldsymbol{\Sigma} + 14\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma}^2 + 12\mathrm{tr}\boldsymbol{\Sigma}^5\right),$$
$$Z_6 = n_3w_2^2w_1\left(3(\mathrm{tr}\boldsymbol{\Sigma})^3\mathrm{tr}\boldsymbol{\Sigma}^2 + 6(\mathrm{tr}\boldsymbol{\Sigma})^2\mathrm{tr}\boldsymbol{\Sigma}^3 + 12(\mathrm{tr}\boldsymbol{\Sigma}^2)^2\mathrm{tr}\boldsymbol{\Sigma} + 42\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma}^2 + 12\mathrm{tr}\boldsymbol{\Sigma}^4\mathrm{tr}\boldsymbol{\Sigma}\right) + o(n^5),$$
$$Z_7 = n_5w_1^5\mathrm{tr}\boldsymbol{\Sigma}^3\mathrm{tr}\boldsymbol{\Sigma}^2.$$

Acknowledgments

We thank the Editor, Associate Editor and Referees, as well as our financial sponsors. Yangchun Zhang's research was supported by the Shanghai Sailing Program, China (21YF1413500). Jiang Hu's research was supported by the National Natural Science Foundation of China (NSFC) (Nos. 12171078 and 11971097) and the National Key R&D Program of China (No. 2020YFA0714102). Weiming Li's research was supported by the NSFC, China (Nos. 11971293 and 12141107).

Appendix

This section includes some of the supporting lemmas used in Section 4.

Lemma 2. *Let $\mathbf{z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ be a standard p-dimensional Gaussian random vector. For any $p \times p$ nonrandom matrix $\mathbf{C}$ and $\alpha \in \mathbb{R}$, we have*

$$\mathbb{E}\left(\frac{\mathbf{z}^\top\mathbf{C}\mathbf{z}}{\|\mathbf{z}\|^2}\right)^\alpha = \frac{\mathbb{E}(\mathbf{z}^\top\mathbf{C}\mathbf{z})^\alpha}{\mathbb{E}(\|\mathbf{z}\|^{2\alpha})}. \tag{53}$$

Proof. From the fact that $\mathbf{z}^\top\mathbf{C}\mathbf{z}/\|\mathbf{z}\|^2$ is independent of $\|\mathbf{z}\|^2$, we obtain

$$\mathbb{E}(\mathbf{z}^\top\mathbf{C}\mathbf{z})^\alpha = \mathbb{E}\left(\frac{\mathbf{z}^\top\mathbf{C}\mathbf{z}}{\|\mathbf{z}\|^2}\|\mathbf{z}\|^2\right)^\alpha = \mathbb{E}\left(\frac{\mathbf{z}^\top\mathbf{C}\mathbf{z}}{\|\mathbf{z}\|^2}\right)^\alpha\mathbb{E}(\|\mathbf{z}\|^{2\alpha}),$$

which implies the conclusion of the lemma. □

Lemma 3. *For any $p \times p$ complex matrix $\mathbf{C}$ and $\mathbf{u} = \mathbf{z}/\|\mathbf{z}\|$ with $\mathbf{z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$, we have*

$$\mathbb{E}\left|\mathbf{u}^\top\mathbf{C}\mathbf{u} - \frac{1}{p}\mathrm{tr}\mathbf{C}\right|^q \le K_q\|\mathbf{C}\|^q p^{-q/2}, \ q \ge 2, \tag{54}$$

where K_q is a positive constant depending only on q.

Proof. This lemma is a direct consequence of Lemma 1 in [18]. □

Lemma 4. *For any $p \times p$ complex matrices $\mathbf{C}$ and $\tilde{\mathbf{C}}$,*

$$\mathbb{E}\left(\mathbf{u}^\top\mathbf{C}\mathbf{u} - \frac{1}{p}\mathrm{tr}\mathbf{C}\right)\left(\mathbf{u}^\top\tilde{\mathbf{C}}\mathbf{u} - \frac{1}{p}\mathrm{tr}\tilde{\mathbf{C}}\right) = \frac{\mathrm{tr}\mathbf{C}\tilde{\mathbf{C}}^\top + \mathrm{tr}\mathbf{C}\tilde{\mathbf{C}}}{p(p+2)} - \frac{2\mathrm{tr}\mathbf{C}\mathrm{tr}\tilde{\mathbf{C}}}{p^2(p+2)}.$$

Proof. Let $\mathbf{u} = (u_1, \ldots, u_p)^\top = \mathbf{z}/\|\mathbf{z}\|$ with $\mathbf{z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$. Because $\|\mathbf{z}\|$ is independent of $\mathbf{z}/\|\mathbf{z}\|$, for $i_j \in \mathbb{R}$, $j \in \{1, 2, 3, 4\}$, we have

$$\mathbb{E}\left(\prod_{j=1}^4 u_j^{i_j}\right) = \mathbb{E}\left(\prod_{j=1}^4 z_j^{i_j}\right)\Big/\mathbb{E}\|\mathbf{z}\|^{i_1+i_2+i_3+i_4}.$$

We thus obtain

$$\mathbb{E}(u_1^4) = \frac{3}{p(p+2)}, \quad \mathbb{E}(u_1^2u_2^2) = \frac{1}{p(p+2)}, \quad \mathbb{E}(u_1^2u_2u_3) = \mathbb{E}(u_1^3u_2) = \mathbb{E}(u_1u_2u_3u_4) = 0.$$

Using these results, we obtain

$$\mathbb{E}(\mathbf{u}^\top\mathbf{C}\mathbf{u}\mathbf{u}^\top\tilde{\mathbf{C}}\mathbf{u}) = \left(\mathbb{E}u_1^4 - 3\mathbb{E}u_1^2u_2^2\right)\sum_{i=1}^{p} c_{ii}\tilde{c}_{ii} + \mathbb{E}u_1^2u_2^2\left(\mathrm{tr}\mathbf{C}\mathrm{tr}\tilde{\mathbf{C}} + \mathrm{tr}\mathbf{C}\tilde{\mathbf{C}}^\top + \mathrm{tr}\mathbf{C}\tilde{\mathbf{C}}\right)$$
$$= \frac{1}{p(p+2)}\left(\mathrm{tr}\mathbf{C}\mathrm{tr}\tilde{\mathbf{C}} + \mathrm{tr}\mathbf{C}\tilde{\mathbf{C}}^\top + \mathrm{tr}\mathbf{C}\tilde{\mathbf{C}}\right),$$

which yields the result of the lemma. □

Lemma 5 ([7]). *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then, for $q \geq 2$,*

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

Lemma 6 ([7]). *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then, for $q > 1$,*

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

Lemma 7 (*Theorem 35.12 of* [6]). *Suppose that for each n $Y_{n1},\ldots,Y_{nr_n}$ is a real martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_{nj}\}$ having second moments. If for each $\varepsilon > 0$, as $n \to \infty$,*

$$\sum_{j=1}^{r_n}\mathbb{E}(Y_{nj}^2 I_{(|Y_{nj}|\geq\varepsilon)}) \to 0, \quad \sum_{j=1}^{r_n}\mathbb{E}(Y_{nj}^2|\mathcal{F}_{n,j-1}) \xrightarrow{i.p.} \sigma^2,$$

where σ^2 is a positive constant, then

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

Lemma 8. *Suppose that Assumptions A1–A5 hold. Then, for any positive s,*

$$\Pr(\|\mathbf{B}_n\| > \eta_r) = o(n^{-s}), \quad \eta_r > \limsup_{n,p\to\infty}\max_{1\leq j\leq n}\{\xi_j^2/p\}\lambda_{\max}^{\boldsymbol{\Sigma}}(1+\sqrt{c})^2,$$
$$\Pr(\lambda_{\min}^{\mathbf{B}_n} < \eta_l) = o(n^{-s}), \quad 0 < \eta_l < \lim\inf_{p\to\infty}\lambda_{\min}^{\boldsymbol{\Sigma}} I_{(0,1)}(c)(1-\sqrt{c})^2.$$

Proof. Let $\mathbf{x}_j = \mathbf{A}\mathbf{z}_j$, where $\mathbf{A}\mathbf{A}^\top = \boldsymbol{\Sigma}$ and $\mathbf{z}_j \sim N(\mathbf{0},\mathbf{I}_p), j \in \{1,\ldots,n\}$. In addition, let $\mathbf{B}_n^{(0)} = (1/n)\sum_{j=1}^n \mathbf{A}\mathbf{z}_j\mathbf{z}_j^\top\mathbf{A}^\top$. From [4], the conclusions of this lemma hold for $\mathbf{B}_n^{(0)}$. Choose $\eta_r^{(0)}$ and $\eta_l^{(0)}$ such that they satisfy

$$\eta_l < \eta_l^{(0)} < \lim\inf_{p\to\infty}\lambda_{\min}^{\boldsymbol{\Sigma}} I_{(0,1)}(c)(1-\sqrt{c})^2, \quad \limsup_{n,p\to\infty}\max_{1\leq j\leq n}\{\xi_j^2/p\}\lambda_{\max}^{\boldsymbol{\Sigma}}(1+\sqrt{c})^2 < \eta_r^{(0)} < \eta_r.$$

From the inequalities

$$\min_{1\leq j\leq n}\frac{\xi_j^2}{\|\mathbf{z}_j\|^2}\lambda_{\min}^{\mathbf{B}_n^{(0)}} \leq \lambda_{\min}^{\mathbf{B}_n} \leq \|\mathbf{B}_n\| \leq \max_{1\leq j\leq n}\frac{\xi_j^2}{\|\mathbf{z}_j\|^2}\|\mathbf{B}_n^{(0)}\|,$$

we have

$$\Pr(\|\mathbf{B}_n\| > \eta_r) \leq \Pr\left(\|\mathbf{B}_n^{(0)}\| > \eta_r^{(0)}\right) + \Pr\left(\max_{1\leq j\leq n}\frac{\xi_j^2}{\|\mathbf{z}_j\|^2}\|\mathbf{B}_n^{(0)}\| > \eta_r, \|\mathbf{B}_n^{(0)}\| \leq \eta_r^{(0)}\right)$$
$$\leq \Pr\left(\max_{1\leq j\leq n}\frac{\xi_j^2}{\|\mathbf{z}_j\|^2} > \frac{\eta_r}{\eta_r^{(0)}}\right) + o(n^{-s}) \leq n\Pr\left(\frac{\xi_1^2}{\|\mathbf{z}_1\|^2} > \frac{\eta_r}{\eta_r^{(0)}}\right) + o(n^{-s}),$$

and similarly,

$$\Pr(\lambda_{\min}^{\mathbf{B}_n} < \eta_l) \leq n\Pr\left(\frac{\xi_1^2}{\|\mathbf{z}_1\|^2} < \frac{\eta_l}{\eta_l^{(0)}}\right) + o(n^{-s}).$$

To complete our proof, it is sufficient to show that, for any fixed $\varepsilon > 0$ and $s > 0$,

$$\Pr\left(\left|\frac{\xi_1^2}{\|\mathbf{z}_1\|^2} - 1\right| > \varepsilon\right) = o(n^{-s}).$$

In fact,

$$\begin{aligned}\Pr\left(\left|\frac{\xi_1^2}{\|\mathbf{z}_1\|^2} - 1\right| > \varepsilon\right) &\le \Pr\left(\left|\frac{p}{\|\mathbf{z}_1\|^2} - 1\right|\frac{\xi_1^2}{p} > \frac{\varepsilon}{2}\right) + \Pr\left(\left|\frac{\xi_1^2}{p} - 1\right| > \frac{\varepsilon}{2}\right) \\ &\le \Pr\left(\left|\frac{p}{\|\mathbf{z}_1\|^2} - 1\right|\frac{\xi_1^2}{p} > \frac{\varepsilon}{2}, \frac{\xi_1^2}{p} < 2\right) + o(n^{-s}) \le \Pr\left(\left|\frac{p}{\|\mathbf{z}_1\|^2} - 1\right| > \varepsilon\right) + o(n^{-s}) = o(n^{-s}),\end{aligned}$$

where the last equality can be obtained by the Chernoff bound of the chi-squared distribution. □

References

[1] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley Series in Probability and Statistics, third ed., Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2003.
[2] Z. Bai, J. Chen, J. Yao, On estimation of the population spectral distribution from a high-dimensional sample covariance matrix, Australian New Zealand J. Stat. 52 (2010) 423–437.
[3] Z. Bai, D. Jiang, J. Yao, S. Zheng, Corrections to LRT on large-dimensional covariance matrix by RMT, Ann. Statist. 37 (2009) 3822–3840.
[4] Z. Bai, J.W. Silverstein, Central limit theorem for linear spectral statistics of large dimensional sample covariance matrices, Ann. Probab. 32 (2004) 553–605.
[5] P. Billingsley, Convergence of Probability Measures, John Wiley & Sons, Inc., New York-London-Sydney, 1968.
[6] P. Billingsley, Probability and measure, third ed., in: Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1995.
[7] D.L. Burkhölder, Distribution function inequalities for martingales, Ann. Probab. 1 (1973) 19–42.
[8] K. Fang, Y. Zhang, Generalized Multivariate Analysis, Springer-Verlag, Science Press, Berlin, Beijing, 1990.
[9] A.K. Gupta, T. Varga, T. Bodnar, Elliptically Contoured Models in Statistics and Portfolio Theory, second ed., Springer, New York, 2013.
[10] J. Hu, W. Li, Z. Liu, W. Zhou, High-dimensional covariance matrices in elliptical distributions with application to spherical test, Ann. Statist. 47 (2019) 527–555.
[11] J. Hu, W. Li, W. Zhou, Central limit theorem for mutual information of large mimo systems with elliptically correlated channels, IEEE Trans. Inform. Theory 65 (2019) 7168–7180.
[12] D. Kelker, Distribution theory of spherical distributions and a location-scale parameter generalization, Sankhyā. Indian J. Stat. Ser. A 32 (1970) 419–438.
[13] O. Ledoit, M. Wolf, Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size, Ann. Statist. 30 (2002) 1081–1102.
[14] W. Li, Q. Wang, J. Yao, W. Zhou, On eigenvalues of a high-dimensional spatial-sign covariance matrix, Bernoulli 28 (2022) 606–637.
[15] W. Li, J. Yao, A local moment estimator of the spectrum of a large dimensional covariance matrix, Statist. Sinica 24 (2014) 919–936.
[16] W. Li, J. Yao, On structure testing for component covariance matrices of a high dimensional mixture, J. R. Stat. Soc. Ser. B Stat. Methodol. 80 (2018) 293–318.
[17] V.A. Marčenko, L.A. Pastur, Distribution of eigenvalues in certain sets of random matrices, Mat. Sb. (N.S.) 72 (1967) 507–536.
[18] D. Morales-Jimenez, I. Johnstone, M. McKay, J. Yang, Asymptotics of eigenstructure of sample correlation matrices for high-dimensional spiked models, Statist. Sinica 31 (2021) 571–601.
[19] D. Paindaveine, T. Verdebout, On high-dimensional sign tests, Bernoulli 22 (2016) 1745–1769.
[20] G. Pan, W. Zhou, Central limit theorem for signal-to-interference ratio of reduced rank linear receiver, Ann. Appl. Probab. 18 (2008) 1232–1270.
[21] J.R. Schott, A test for the equality of covariance matrices when the dimension is large relative to the sample sizes, Comput. Statist. Data Anal. 51 (2007) 6535–6542.
[22] J.W. Silverstein, Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices, J. Multivariate Anal. 55 (1995) 331–339.
[23] J.W. Silverstein, Z. Bai, On the empirical distribution of eigenvalues of a class of large dimensional random matrices, J. Multivariate Anal. 54 (1995) 75–192.
[24] M.S. Srivastava, Some tests concerning the covariance matrix in high dimensional data, J. Japan Statist. Soc. 35 (2005) 251–272.
[25] Q. Wang, J. Yao, On the sphericity test with large-dimensional observations, Electron. J. Stat. 7 (2013) 2164–2192.
[26] L. Zhang, Large Dimensional General Sample Covariance Matrices (Ph. D. thesis), National University of Singapore, 2006.
[27] S. Zheng, Z. Bai, J. Yao, Substitution principle for clt of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing, Ann. Statist. 43 (2015) 546–591.
[28] C. Zou, L. Peng, L. Feng, Z. Wang, Multivariate sign-based high-dimensional tests for sphericity, Biometrika 101 (2014) 229–236.