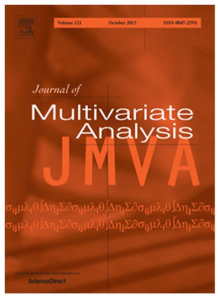
Accepted Manuscript



Estimation of two high-dimensional covariance matrices and the spectrum of their ratio

Jun Wen

PII: S0047-259X(17)30545-6
DOI: https://doi.org/10.1016/j.jmva.2018.06.008
Reference: YJMVA 4376

To appear in: *Journal of Multivariate Analysis*

Received date : 8 September 2017

Please cite this article as: J. Wen, Estimation of two high-dimensional covariance matrices and the spectrum of their ratio, *Journal of Multivariate Analysis* (2018), https://doi.org/10.1016/j.jmva.2018.06.008

Estimation of two high-dimensional covariance matrices and the spectrum of their ratio

Jun Wen[a]

[a] *Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546*

Abstract

Let $S_{p,1}$, $S_{p,2}$ be two independent $p \times p$ sample covariance matrices with degrees of freedom n_1 and n_2, respectively, whose corresponding population covariance matrices are $\Sigma_{p,1}$ and $\Sigma_{p,2}$, respectively. Knowing $S_{p,1}$, $S_{p,2}$, this article proposes a class of estimators for the spectrum (eigenvalues) of the matrix $\Sigma_{p,2}\Sigma_{p,1}^{-1}$ as well as the pair of the whole matrices $(\Sigma_{p,1}, \Sigma_{p,2})$. The estimators are created based on Random Matrix Theory. Under mild conditions, our estimator for the spectrum of $\Sigma_{p,2}\Sigma_{p,1}^{-1}$ is shown to be weakly consistent and the estimator for $(\Sigma_{p,1}, \Sigma_{p,2})$ is shown to be optimal in the sense of minimizing the asymptotic loss within the class of equivariant estimators as $n_1, n_2, p \to \infty$ with $p/n_1 \to c_1 \in (0, 1)$, $p/n_2 \to c_2 \in (0, 1) \cup (1, \infty)$. Also, our estimators are easy to implement. Even when p is 1000, our estimators can be computed in seconds using a personal laptop.

Keywords: Covariance matrix estimation, High-dimensional asymptotics, Marčenko–Pastur equation, Random matrix theory, Spectrum estimation, Two-sample problem.

1. Introduction

It is widely accepted that covariance matrices play a vital role in various statistical problems. However, in most real life applications, the true (population) covariance matrices are unknown. Therefore, a good estimate of it is much in demand. Traditionally, when the population covariance matrix is needed, what statisticians usually do is to use the sample covariance matrix instead. It is well known that when the dimension p of the covariance matrix is fixed and the sample size n tends to infinity, the sample covariance matrix is a consistent estimator of its population counterpart. However, when the dimension p of the covariance matrix is large, especially when the magnitude of p is comparable to the sample size n, the sample covariance matrix no longer performs as well as it does in the small p large n case; see, e.g., [19] for an illustration.

With the development of Random Matrix Theory (RMT), and especially spectral analysis of random matrices, quite a number of new statistical tools on covariance matrices related problems assuming p and n both large have been proposed, which largely stimulates the exploration of better covariance matrices estimators. One of the important techniques rooted in RMT is the so-called Marčenko–Pastur (or MP for short) equation technique. Initiated from the seminal paper [29], MP equation technique has been extensively studied in recent years, see [3–5, 32, 33, 40]. There has been quite a number of statistical applications resulting from it; see [2, 13, 20–22, 25, 26, 31, 39]. Sometimes not only are people interested in estimating one covariance matrix, but also in estimating two covariance matrices and the spectrum of their ratio; see [7, 10, 16, 17, 23, 24, 27, 28, 36]. In this article, we base our proposed covariance matrices estimators on this MP equation technique with the aim of getting substantial performance improvement over the traditional estimators.

In the following, for a symmetric positive semidefinite matrix A, its square root $A^{1/2}$ is defined as the unique symmetric positive semidefinite matrix such that $A^{1/2}A^{1/2} = A$. For a $p \times p$ matrix M with real eigenvalues $\lambda_1, \dots, \lambda_p$, we define the spectral distribution (SD) of M as $F^M = p^{-1} \sum_{i=1}^{p} \mathbf{1}(x \le \lambda_i)$, where $\mathbf{1}$ is the indicator function. Denoting the set of the eigenvalues of $\Sigma_{p,2}\Sigma_{p,1}^{-1}$ as $\mathbf{d}_p = \{d_{p,1}, \dots, d_{p,p}\}$ with $d_{p,1} \le \cdots \le d_{p,p}$, we impose the following three assumptions throughout this article.

Email address: `stawenj@nus.edu.sg` (Jun Wen)

Assumption 1. Let X_p, Y_p be two independent real $p \times n_1$ and $p \times n_2$ random matrices. Both X_p and Y_p consist of iid entries with mean 0, variance 1 and finite 12th moment. Denote $\Sigma_{p,1}$ and $\Sigma_{p,2}$ to be two $p \times p$ covariance matrices. The observed data are given as the sample covariance matrices $S_{p,1} = n_1^{-1}\Sigma_{p,1}^{1/2} X_p X_p^\top \Sigma_{p,1}^{1/2}$ and $S_{p,2} = n_2^{-1}\Sigma_{p,2}^{1/2} Y_p Y_p^\top \Sigma_{p,2}^{1/2}$.

Assumption 2. $n_1, n_2, p \to \infty$ with $\mathbf{c}_p = (c_{1,p}, c_{2,p}) = (p/n_1, p/n_2) \to (c_1, c_2)$ such that $c_1 \in (0, 1)$, $c_2 \in (0, 1) \cup (1, \infty)$.

Assumption 3. The SD of $\Sigma_{p,2}\Sigma_{p,1}^{-1}$, $D_p(t)$, converges weakly to a distribution $D(t)$ as $p \to \infty$ with $\mathrm{Supp}(D_p)$ and $\mathrm{Supp}(D)$, the supports of D_p and D, uniformly contained in a compact subinterval K of $(0, \infty)$ for all n_1, n_2, p large.

In Assumption 2, the case $c_2 = 1$ is ruled out for technical reasons. Nevertheless, our numerical computation can still address this case. Also, throughout this article, we regard n_1, n_2 as two functions of p. By saying $p \to \infty$, we mean $n_1, n_2, p \to \infty$ with $\mathbf{c}_p \to (c_1, c_2)$.

An estimator of the spectrum $\mathbf{d}_p$ is said to be weakly consistent if its empirical distribution converges in law to D as $p, n_1, n_2 \to \infty$ with $p/n_1 \to c_1$ and $p/n_2 \to c_2$. There are two goals to this article. One is to propose a weakly consistent estimator of the spectrum $\mathbf{d}_p$. The other is to construct a class of covariance matrices estimators $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$ having optimal risk performance. For the problem of estimating $\mathbf{d}_p$, El Karoui [13] and Ledoit and Wolf [21] have addressed the one-sample case. Specifically, in their problems, there is only one population covariance matrix Σ_p, and given the eigenvalues of the sample covariance matrix S_p, weakly consistent spectrum estimators of Σ_p have been constructed using the MP equation technique.

In this article, we propose two types of estimators, $\hat{\mathbf{d}}_p^{EK}$ and $\hat{\mathbf{d}}_p^{LW}$, of $\mathbf{d}_p$ which are the two-sample generalization of those of [13] and [21], respectively. It is shown that these estimators are weakly consistent given Assumptions 1–3. Moreover, if we regard them as p-variate vectors, such that

$$\hat{\mathbf{d}}_p^{EK} = (\hat{d}_{p,1}^{EK}, \ldots, \hat{d}_{p,p}^{EK}), \quad \hat{\mathbf{d}}_p^{LW} = (\hat{d}_{p,1}^{LW}, \ldots, \hat{d}_{p,p}^{LW}), \quad \mathbf{d}_p = (d_{p,1}, \ldots, d_{p,p})$$

then $p^{-1}\sum_{i=1}^p |\hat{d}_{p,i}^{LW} - d_{p,i}|$ and $p^{-1}\sum_{i=1}^p |\hat{d}_{p,i}^{EK} - d_{p,i}|$ both converge to 0 almost surely.

For a generic estimator $\hat{\mathbf{d}}_p = (\hat{d}_{p,1} \ldots, \hat{d}_{p,p})$, the absolute loss is also the L_1 distance between the vectors $\hat{\mathbf{d}}_p$ and $\mathbf{d}_p$ normalized by p. Some authors, e.g., [10, 21], use the L_2 distance $p^{-1}\sum_{i=1}^p (\hat{d}_{p,i} - d_{p,i})^2$ as the loss function. One of the advantages of using the L_1 distance instead of the L_2 distance is that the L_1 distance is more robust against large eigenvalue deviations. By making use of the weakly consistent estimators of $\mathbf{d}_p$, we further constructed our covariance matrices estimator $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$.

Stein [34, 35] proposed a class of rotation equivariant estimators for the one-sample problem using an unbiased estimate of risk (SURE). By generalizing Stein's SURE, Loh [28] proposed the two-sample Stein type estimator $(\hat{\Sigma}_{p,1}^{ST}, \hat{\Sigma}_{p,2}^{ST})$ which simultaneously estimates the pair of two covariance matrices $(\Sigma_{p,1}, \Sigma_{p,2})$. Seen from [28], the Stein type estimator $(\hat{\Sigma}_{p,1}^{ST}, \hat{\Sigma}_{p,2}^{ST})$ already has remarkable finite-sample performance. However, it still has some limitations.

The derivation of $(\hat{\Sigma}_{p,1}^{ST}, \hat{\Sigma}_{p,2}^{ST})$ requires normal distribution of the data while our estimators are distribution free. Furthermore, the estimator is only defined when $p \le \min(n_1, n_2)$ while our estimators also apply to the case when $p > n_2$. By comparing the limiting loss, we show that our estimators dominate the minimax estimator $(\hat{\Sigma}_{p,1}^{AU}, \hat{\Sigma}_{p,2}^{AU})$ (see [28]) in the limit and are thus also minimax asymptotically. Finally, we also compared the finite-sample risk performance of our estimators to the one-sample estimator of [22], which estimates $\Sigma_{p,1}$ and $\Sigma_{p,2}$ using data from each population individually. From the simulated results, we see that when $\Sigma_{p,1}$ is approximately proportional to $\Sigma_{p,2}$ but the eigenvalues of both $\Sigma_{p,1}$ and $\Sigma_{p,2}$ are far apart, our estimators largely outperform the one-sample estimator. The remainder of this article is organized as follows.

In Section 2, we present some preliminary results of RMT. In Section 3, we derive two types of the spectrum estimators using RMT. The main results are presented in Theorem 2 and 4. In Section 4, we focus on the problem of estimating the whole pair of covariance matrices $(\Sigma_{p,1}, \Sigma_{p,2})$. We show in Theorem 8 that our estimator is asymptotically optimal in minimizing the limiting loss function and present in Section 4.6 some other advantages of our estimator over the traditional decision-theoretic ones. In Section 5, we evaluate the finite-sample performance of our proposed estimators using extensive Monte Carlo simulations. The technical proofs are given in the Appendix.

Throughout this article, we use the following notation. The integers, real numbers, complex numbers are denoted by $\mathbb{Z}, \mathbb{R}, \mathbb{C}$ respectively. For any $z \in \mathbb{C}$, $\mathrm{Re}(z)$, $\mathrm{Im}(z)$ and $\bar{z}$ are the real part, imaginary part and conjugate of z, respectively. Furthermore, $\mathbb{C}^+ = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$, $\mathbb{C}^- = \{z \in \mathbb{C} : \mathrm{Im}(z) < 0\}$ and $\mathbb{Z}^+ = \{1, 2, \ldots\}$. We write

$\mathbf{i} = \sqrt{-1}$. For any matrix M, $M^\top$ represents the transpose of M, $\|M\|$ denote the spectral norm of M which equals to the largest singular value of M and tr$M$ stands for trace of M if M is a square matrix. The identity matrix with appropriate size according to the context is denoted as I. For a function $f(x)$, $f(x_0^+)$ is defined as the right limit of f at x_0 whenever the latter exists. For a finite signed measure G on the real line, Supp(G) denotes the support of G. Given G and a set $S \subset \mathbb{R}$, $G\{S\}$ denotes the mass that G puts on S. In particular, if S is a singleton, say $S = \{a\}$, then $G\{S\}$ is simply written as $G\{a\}$. Moreover if G is a probability distribution supported on $(0, \infty)$, we define G^{inv} as the probability distribution such that $G^{\text{inv}}\{[a, b]\} = G\{[1/b, 1/a]\}$ for all $b \geq a > 0$ and $G^{-1}(x)$ is defined as the quantile function of G for $x \in [0, 1]$. Specifically, $G^{-1}(x) = \sup\{u \in \mathbb{R} : G(u) < x\}$ for $x \in (0, 1]$ and $G^{-1}(0) = G^{-1}(0^+)$. Finally, $\rightsquigarrow$ is the symbol for weak convergence and $a.s.$ is short for almost surely.

2. Some preliminary results

2.1. The Marčenko–Pastur equation

The Marčenko–Pastur equation serves as a very powerful tool for high-dimensional multivariate data analysis. For overviews, we refer the readers to [1, 30].

Suppose Σ_p is a $p \times p$ real-valued symmetric positive definite matrix with its SD, $H_p(t)$ weakly converging to a distribution $H(t)$ as $p \to \infty$. Let X be a $p \times n$ real matrix consisting of iid entries with mean 0 and variance 1. Denote $S_p = n^{-1}\Sigma_p^{1/2}XX^\top\Sigma_p^{1/2}$. Then with probability 1, its SD F^{S_p} converges weakly to a nonrandom distribution $F(t)$ called the limiting spectral distribution (LSD) of S_p as $p, n \to \infty$ with $p/n \to c > 0$. To quantize the relationship between $H(t)$ and $F(t)$, we utilize an auxiliary quantity called the Stieltjes transform. For a distribution function $G(t)$ of a finite signed measure on the real line, its Stieltjes transform is defined for $z \in \mathbb{C}^+$ as

$$m_G(z) = \int_{-\infty}^{\infty} \frac{1}{t - z}\, dG(t).$$

Sometimes we also need to set $z \in \mathbb{C}^-$, whence it is easily seen that $m_G(z) = \bar{m}_G(\bar{z})$. For any continuity points a, b of G, the inversion formula of the Stieltjes transform (see Theorem B.8 of [5]) reads

$$G\{[a, b]\} = \lim_{\eta \to 0^+} \int_a^b \operatorname{Im}\{m_G(\xi + \mathbf{i}\eta)\} d\xi.$$

For a SD G of a symmetric matrix A that can be written as $A = PP^\top$, where the matrix P is of size, say $p \times n$, it is useful to consider the distribution $\underline{G}$ defined as the SD of $P^\top P$. In this definition, G and $\underline{G}$ are called the companion distributions of each other. It can be seen that $PP^\top$ and $P^\top P$ have the same set of eigenvalues up to $|p - n|$ 0s. Hence G and $\underline{G}$ are related by the equality $\underline{G}(t) = (1 - p/n)\mathbf{1}(0 \leq t) + (p/n)G(t)$, for all $t \in \mathbb{R}$. Correspondingly, the Stieltjes transforms of G and $\underline{G}$ are called the companion Stieltjes transforms of each other. If G and $\underline{G}$ converge weakly to some distributions as $p, n \to \infty$ with $p/n \to c > 0$, we also refer to the limiting distributions as the companion distributions of each other.

Now we come back to establish the quantitative relationship between the distributions H and F above. Silverstein [32] shows that for all $z \in \mathbb{C}^+$, $m_F(z)$ is the unique solution in $\{m \in \mathbb{C} : -(1 - c)/z + cm \in \mathbb{C}^+\}$ to the equation

$$m = \int \frac{dH(t)}{t(1 - c - czm) - z},$$

and equivalently, $m_{\underline{F}}(z)$ is the unique solution in $\mathbb{C}^+$ to the following Marčenko–Pastur (MP) equation (named after [29])

$$z = -\frac{1}{m} + c \int \frac{t}{1 + tm}\, dH(t), \tag{1}$$

where integrations without limits are meant to be integrating over the whole range here and after if no ambiguity occurs.

One may note that in statistics literature, the data are usually given as $\Sigma_p^{1/2}X$ whose sample and population covariance matrices are respectively $n^{-1}\Sigma_p^{1/2}XX^\top\Sigma_p^{1/2}$ and Σ_p. Thus the above result provides at least in theory a way to infer eigenvalues of population covariance matrices using sample covariance matrices.

Finally, it is important to know that for a Stieltjes transform $m_G(z)$ of some distribution G, Silverstein and Choi [33] have shown that if, for all z in $\mathbb{C}^+$, $m_G(z)$ is a solution to some Marčenko–Pastur equation

$$z = -\frac{1}{m} + c \int \frac{t}{1+tm}\, dH(t),$$

with respect to a constant $c > 0$ and a probability distribution $H(t)$, then $m_G(z)$ can be continuously extended to the whole real line except possibly at the origin. More specifically, it is shown that $\lim_{z \in \mathbb{C}^+ \to x} m_G(z)$ exists for all $x \in \mathbb{R} \backslash \{0\}$. In what follows, we will denote it $\breve{m}_G(x)$ as long as it exists. This quantity is useful in finding the density of G.

2.2. The limiting spectral distribution of $S_{p,2}S_{p,1}^{-1}$

Similar to the MP equation (1), which associates the LSD of the sample covariance matrix with that of the population covariance matrix, we can also establish a system of equations which associates the LSD of $S_{p,2}S_{p,1}^{-1}$ with the LSD of $\Sigma_{p,2}\Sigma_{p,1}^{-1}$. To facilitate the construction of the estimator $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$, we also consider the matrix $S_{p,2}(S_{p,1}+S_{p,2})^{-1}$.

Denote the sets of ascending eigenvalues of $S_{p,2}S_{p,1}^{-1}$ and $S_{p,2}(S_{p,1}+S_{p,2})^{-1}$ as $\{\ell_{p,1}, \ldots, \ell_{p,p}\}$ and $\{f_{p,1}, \ldots, f_{p,p}\}$, respectively. We see that it satisfies that $f_{p,i} = \ell_{p,i}/(1+\ell_{p,i})$. In the following, we summarize our notation of the matrices that we will work on frequently in the first two columns of Table 1 below. Under Assumptions 1–3, it is known from the result of [32] that the LSD of all theses matrices exist. The notation of SD, LSD together with their Stieltjes transforms (ST) of these matrices are presented in the last two columns of Table 1.

We see that T^{inv} and $\underline{T}^{\text{inv}}$ are associated by the equality $m_{\underline{T}^{\text{inv}}}(z) = -(1-c_1)/z + c_1 m_{T^{\text{inv}}}(z)$, and L and $\underline{L}$ are associated by the equality $m_{\underline{L}}(z) = -(1-c_2)/z + c_2 m_L(z)$. Using the results on MP equations [32], we immediately have the following results:

a) $m_{\underline{T}^{\text{inv}}}(z)$ is the unique value in $\mathbb{C}^-$ satisfying, for all $z \in \mathbb{C}^-$, the MP equation

$$z = -\frac{1}{m_{\underline{T}^{\text{inv}}}(z)} + c_1 \int \frac{t}{1+t m_{\underline{T}^{\text{inv}}}(z)}\, dD^{\text{inv}}(t).$$

b) $m_{\underline{L}}(z)$ is the unique value in $\mathbb{C}^+$ satisfying, for all $z \in \mathbb{C}^+$, the MP equation

$$z = -\frac{1}{m_{\underline{L}}(z)} + c_2 \int \frac{t}{1+t m_{\underline{L}}(z)}\, dT(t).$$

c) It is easy to verify that the above two equations are equivalent respectively to

$$\forall_{z \in \mathbb{C}^-} \quad z = -\frac{1}{m_{\underline{T}^{\text{inv}}}(z)} + c_1 \int \frac{dD(t)}{t + m_{\underline{T}^{\text{inv}}}(z)}, \tag{2}$$

Table 1: Notation of the matrices and their corresponding SD, LSD and Stieltjes transforms.

Notation	Definitions	SD/ST	LSD/ST
M_{L_p}	$S_{p,2}S_{p,1}^{-1}$,	$L_p/m_{L_p}(z)$	$L/m_L(z)$
$M_{\underline{L}_p}$	$n_2^{-1}Y_p^\top \Sigma_{p,2}^{1/2} S_{p,1}^{-1} \Sigma_{p,2}^{1/2} Y_p$,	$\underline{L}_p/m_{\underline{L}_p}(z)$	$\underline{L}/m_{\underline{L}}(z)$
M_{F_p}	$S_{p,2}(S_{p,1}+S_{p,2})^{-1}$	$F_p/m_{F_p}(z)$	$F/m_F(z)$
M_{T_p}	$S_{p,1}^{-1}\Sigma_{p,2}$,	$T_p/m_{T_p}(z)$	$T/m_T(z)$
$M_{T_p^{\text{inv}}}$	$S_{p,1}\Sigma_{p,2}^{-1}$,	$T_p^{\text{inv}}/m_{T_p^{\text{inv}}}(z)$	$T^{\text{inv}}/m_{T^{\text{inv}}}(z)$
$M_{\underline{T}_p^{\text{inv}}}$	$n_1^{-1}X_p^\top \Sigma_{p,1}^{1/2} \Sigma_{p,2}^{-1} \Sigma_{p,1}^{1/2} X_p$,	$\underline{T}_p^{\text{inv}}/m_{\underline{T}_p^{\text{inv}}}(z)$	$\underline{T}^{\text{inv}}/m_{\underline{T}^{\text{inv}}}(z)$
M_{R_p}	$S_{p,2}\Sigma_{p,1}^{-1}$	$R_p/m_{R_p}(z)$	$R/m_R(z)$
M_{D_p}	$\Sigma_{p,2}\Sigma_{p,1}^{-1}$,	$D_p/m_{D_p}(z)$	$D/m_D(z)$
$M_{D_p^{\text{inv}}}$	$\Sigma_{p,2}^{-1}\Sigma_{p,1}$.	$D_p^{\text{inv}}/m_{D_p^{\text{inv}}}(z)$	$D^{\text{inv}}/m_{D^{\text{inv}}}(z)$

and

$$\forall_{z\in\mathbb{C}^+} \quad z = -\frac{1}{m_{\underline{L}}(z)} + c_2 \int \frac{dT^{\mathrm{inv}}(t)}{t + m_{\underline{L}}(z)}. \tag{3}$$

The above results associate $L, \underline{L}, T, T^{\mathrm{inv}}, \underline{T}^{\mathrm{inv}}, D, D^{\mathrm{inv}}$ with each other. In practice, the SD of observed data matrix we have is L_p (or equivalently $\underline{L}_p$), and the population parameter we want to estimate is the set of eigenvalues of M_{D_p}. It is necessary that we associate L (or $\underline{L}$) directly with D.

In the following, we present a system of equations which associates the Stieltjes transform of $\underline{L}$ with the one of D. A recent publication by Zheng et al. [40] has also shown the same result on weaker conditions. Their paper is more concerned with the theoretical property of the spectrum of $S_{p,2}S_{p,1}^{-1}$ from the perspective of probability theory, whereas, our paper will utilize the result on estimating the spectrum of $\Sigma_{p,2}\Sigma_{p,1}^{-1}$.

Theorem 1. *Under Assumptions 1–3 and following the notation in Table 1, let $m_0(z) = m_{\underline{T}^{\mathrm{inv}}}\{-m_{\underline{L}}(z)\}$. Then for all $z \in \mathbb{C}^+$, $m_0(z)$ is the unique solution in $\mathbb{C}^-$ to the equation*

$$z = \frac{(c_1 + c_2 - c_1c_2)}{c_1\left\{-\frac{1}{m_0} + c_1 \int \frac{dD(t)}{t+m_0}\right\}} + \frac{c_2}{c_1} m_0, \tag{4}$$

and

$$\forall_{z\in\mathbb{C}^+} \quad -m_{\underline{L}}(z) = -\frac{1}{m_0(z)} + c_1 \int \frac{dD(t)}{t + m_0(z)}. \tag{5}$$

According to the relationship between F_p and L_p, we observe that

$$m_{F_p}(z) = \int \frac{1}{\frac{y}{1+y} - z} dF_p\left(\frac{y}{1+y}\right) = \int \frac{1+y}{(1-z)y - z} dL_p(y) = \frac{1}{1-z} + \frac{m_{L_p}(\frac{z}{1-z})}{(1-z)^2}. \tag{6}$$

Taking limits on both sides, we obtain the relationship on the Stieltjes transforms of the limiting distributions, viz.

$$m_F(z) = (1-z)^{-1} + (1-z)^{-2} m_L\Big(\frac{z}{1-z}\Big).$$

Remark 1. Let G be a probability distribution such that $\mathrm{Supp}(G) \subset (0, \infty)$, and $k_1 \in (0, 1)$, $k_2 \in (0, 1) \cup (1, \infty)$ be two constants. Seen from Theorem 1, if we solve, for any $z \in \mathbb{C}^+$, the equation

$$z = \frac{(k_1 + k_2 - k_1k_2)}{k_1\left\{-\frac{1}{m_0} + k_1 \int \frac{dG(t)}{t+m_0}\right\}} + \frac{k_2}{k_1} m_0$$

with respect to m_0 in $\mathbb{C}^-$, we will obtain a value $m_0(z) \in \mathbb{C}^-$. Letting

$$m(z) = \frac{1}{m_0(z)} - k_1 \int \frac{dG(t)}{t + m_0(z)},$$

we get that the value $m(z)$ is the Stieltjes transform of a probability distribution.

In the sequel, we use the notation $\underline{L}_{G,k_1,k_2}$ to denote this obtained distribution. Its companion distribution is denoted as L_{G,k_1,k_2}. Recall that L_{G,k_1,k_2} satisfies that $L_{G,k_1,k_2}(x) = k_2^{-1}\underline{L}_{G,k_1,k_2}(x) - (1-k_2)k_2^{-1}\mathbf{1}(0 \le x)$ for all $x \in \mathbb{R}$. Besides, we define F_{G,k_1,k_2} as the distribution such that its Stieltjes transform satisfies

$$m_{F_{G,k_1,k_2}}(z) = (1-z)^{-1} + (1-z)^{-2} m_{L_{G,k_1,k_2}}\Big(\frac{z}{1-z}\Big). \tag{7}$$

The companion distribution of F_{G,k_1,k_2} is denoted as $\underline{F}_{G,k_1,k_2}$. We note that with this notation, $L = L_{D,c_1,c_2}$, $\underline{L} = \underline{L}_{D,c_1,c_2}$, $F = F_{D,c_1,c_2}$ and $\underline{F} = \underline{F}_{D,c_1,c_2}$.

3. Weakly consistent estimation of population eigenvalues

3.1. A weakly consistent estimator of El Karoui type

3.1.1. Formulation of the estimation problem

Let $h = c_1 + c_2 - c_1 c_2$. From (4) and (5), we see that for all $z \in \mathbb{C}^+$, $m_{\underline{L}}(z)$ satisfies

$$-m_{\underline{L}}(z) = -\frac{1}{\frac{c_1}{c_2}\left\{z + \frac{h}{c_1 m_{\underline{L}}(z)}\right\}} + c_1 \int \frac{dD(t)}{t + \frac{c_1}{c_2}\left\{z + \frac{h}{c_1 m_{\underline{L}}(z)}\right\}}. \tag{8}$$

Using (8), we construct the estimator of D_p based on the basis pursuit method proposed by El Karoui [13]. First, we see that D_p can be approximated by a weighted sum of point masses, viz.

$$D_p(x) \approx \sum_{k=1}^{N_p} w_k \mathbf{1}(t_k \leq x),$$

where N_p is a positive integer depending on p, $t_1, \ldots, t_{N_p}$ are pre-specified fixed points in $(0, \infty)$, and $w_1, \ldots, w_{N_p}$ are the unknown weights to be optimized over.

Then with $\underline{L}$ replaced by the observed distribution $\underline{L}_p$ and c_1, c_2, h replaced by their finite-sample counterparts $c_{1,p} = p/n_1$, $c_{2,p} = p/n_2$, $h_p = c_{1,p} + c_{2,p} - c_{1,p}c_{2,p}$, the finite-sample approximation of (8) reads

$$-m_{\underline{L}_p}(z) \approx -\frac{1}{\frac{c_{1,p}}{c_{2,p}}\left\{z + \frac{h_p}{c_{1,p} m_{\underline{L}_p}(z)}\right\}} + c_{1,p} \sum_{k=1}^{N_p} \frac{w_k}{t_k + \frac{c_{1,p}}{c_{2,p}}\left\{z + \frac{h_p}{c_{1,p} m_{\underline{L}_p}(z)}\right\}}. \tag{9}$$

Then the strategy is to find $w_1, \ldots, w_{N_p}$ such that (9) is "best" satisfied across a set of values of $\{z_1, \ldots, z_{J_p}\} \subset \mathbb{C}^+$, where J_p is an integer tending to infinity as $p \to \infty$.

3.1.2. Convex optimization

We observe that $w_1, \ldots, w_{N_p}$ are linear in (9), which allows us to reformulate the problem of finding $w_1, \ldots, w_{N_p}$ as a convex optimization problem. Denoting, for each $j \in \{1, \ldots, J_p\}$, the approximation error

$$e_j = m_{\underline{L}_p}(z_j) - \frac{1}{\frac{c_{1,p}}{c_{2,p}}\left\{z_j + \frac{h_p}{c_{1,p} m_{\underline{L}_p}(z_j)}\right\}} + c_{1,p} \sum_{k=1}^{N_p} \frac{w_k}{t_k + \frac{c_{1,p}}{c_{2,p}}\left\{z_j + \frac{h_p}{c_{1,p} m_{\underline{L}_p}(z_j)}\right\}},$$

we propose to

$$\text{Minimize} \quad \max_{j \in \{1,\ldots,J_p\}} \max\{|\mathrm{Re}(e_j)|, |\mathrm{Im}(e_j)|\}. \tag{10}$$

Observe that the only constraints that $w_1, \ldots, w_{N_p}$ should satisfy are $w_1, \ldots, w_{N_p} \in [0, 1]$ with $w_1 + \cdots + w_{N_p} = 1$. Therefore, (10) is essentially a linear program which can be equivalently expressed as:

$$\begin{cases} \underset{(w_1,\ldots,w_{N_p},u)}{\text{Minimize}}\ u, \\ \forall_{j\in\{1,\ldots,N_p\}} \quad -u \leq \mathrm{Re}(e_j) \leq u, \quad -u \leq \mathrm{Im}(e_j) \leq u, \\ \text{subject to } w_1, \ldots, w_{N_p} \in [0, 1] \text{ and } w_1 + \cdots + w_{N_p} = 1. \end{cases}$$

3.1.3. Consistency

In this section, we show that the estimator is consistent in the sense of weak convergence of probability distributions provided that the t_ks in (9) become dense in $\mathrm{Supp}(D)$ as $p \to \infty$.

Theorem 2. *Let J_p be an integer tending to ∞ as $p \to \infty$, C be a compact subset of $\mathbb{C}^+$ and $\{z_1, z_2, \ldots\} \subset C$ be a sequence of distinct complex numbers. Let $\tilde{D}_p^{EK}$ be the probability distribution defined as*

$$\tilde{D}_p^{EK} = \operatorname*{argmin}_{D} \max_{j \in \{1,\ldots,J_p\}} \left| m_{\underline{L}_p}(z_j) - \frac{1}{\frac{c_{1,p}}{c_{2,p}} \left\{ z_j + \frac{h_p}{c_{1,p} m_{\underline{L}_p}(z_j)} \right\}} + c_{1,p} \int \frac{dD(t)}{t + \frac{c_{1,p}}{c_{2,p}} \left\{ z_j + \frac{h_p}{c_{1,p} m_{\underline{L}_p}(z_j)} \right\}} \right|.$$

Then we have $\tilde{D}_p^{EK} \overset{a.s.}{\rightsquigarrow} D$ as $p \to \infty$.

In practice, it is infeasible to perform optimization over the set of all probability distributions. In the following, we propose an estimator optimizing over a properly chosen subset, but still keeping the consistent property. By saying a grid, we mean a set of points on $\mathbb{R}$. By saying that a grid covers a set, we mean the set is a subset of the interval between the smallest and largest points in the grid. The size of the grid is defined as the largest length of the gaps between adjacent points in the grid.

Corollary 1. *Let J_p be an integer tending to ∞ as $p \to \infty$, C be a compact subset of $\mathbb{C}^+$, $\{z_1, z_2, \ldots\} \subset C$ be a sequence of distinct complex numbers, G_p be a grid in a compact subinterval of $(0, \infty)$ such that G_p covers $\mathrm{Supp}(D_p)$ and $\mathrm{Supp}(D)$ for all large p. Let $\hat{D}_p^{EK}$ be the probability distribution defined as*

$$\hat{D}_p^{EK} = \operatorname*{argmin}_{D \in \mathcal{P}_{G_p}} \max_{j \in \{1,\ldots,J_p\}} \left| m_{\underline{L}_p}(z_j) - \frac{1}{\frac{c_{1,p}}{c_{2,p}} \left\{ z_j + \frac{h_p}{c_{1,p} m_{\underline{L}_p}(z_j)} \right\}} + c_{1,p} \int \frac{dD(t)}{t + \frac{c_{1,p}}{c_{2,p}} \left\{ z_j + \frac{h_p}{c_{1,p} m_{\underline{L}_p}(z_j)} \right\}} \right|,$$

where $\mathcal{P}_{G_p}$ is the set of discrete probability distributions supported on G_p. Suppose the size of G_p tends to 0 as $p \to \infty$. Then we have $\hat{D}_p^{EK} \overset{a.s.}{\rightsquigarrow} D$ as $p \to \infty$.

Usually the set of eigenvalues of $\Sigma_{p,2}\Sigma_{p,1}^{-1}$ of finite p is of more practical interest to us. Instead of regarding $\hat{D}_p^{EK}$ as a consistent estimator of the limiting population spectral distribution D, we can generate an estimator of the population spectrum $\mathbf{d}_p = \{d_{p,1}, \ldots, d_{p,p}\}$ from $\hat{D}_p^{EK}$, and examine its finite-sample performance. Specifically, we have the following result:

Corollary 2. *Under the assumptions of Corollary 1, we define $\{\hat{d}_{p,1}^{EK}, \ldots, \hat{d}_{p,p}^{EK}\}$ by*

$$\hat{d}_{p,i}^{EK} = p \int_{(i-1)/p}^{i/p} (\hat{D}_p^{EK})^{-1}(x) dx,$$

where $(\hat{D}_p^{EK})^{-1}$ is the quantile function of the distribution $\hat{D}_p^{EK}$. Then as $p \to \infty$,

$$\frac{1}{p} \sum_{i=1}^{p} \left| \hat{d}_{p,i}^{EK} - d_{p,i} \right| \overset{a.s.}{\to} 0.$$

Remark 2. We note that our algorithm (10) minimizes

$$\max_{j \le J_p} \max\{|\mathrm{Re}(e_j)|, |\mathrm{Im}(e_j)|\},$$

while the theory minimizes $\max(|e_1|, \ldots, |e_{J_p}|)$. To see the equivalence of the two, we just need to note that

$$\max\{|\mathrm{Re}(e_j)|, |\mathrm{Im}(e_j)|\} \le |e_j| \le \sqrt{2} \max\{|\mathrm{Re}(e_j)|, |\mathrm{Im}(e_j)|\}.$$

By reformulating the optimization problem as a linear programming, we can obtain our estimator fast and efficiently by using standard linear optimization algorithms. In our case, the problem can be solved in a few seconds by the `linprog` routine of MATLAB on a personal laptop.

3.1.4. Implementation details

Seen from Corollary 1, to implement the El Karoui type estimator, we need to pre-specify the sequence $\{z_1, \ldots, z_{J_p}\}$ and the grid of points G_p. Generally speaking, the criterion leading to the optimal choice of $\{z_1, \ldots, z_{J_p}\}$ and the grid G_p is still unknown and we leave this question to our future research. At this moment, we suggest the following choice of $\{z_1, \ldots, z_{J_p}\}$ and G_p, which according to our simulation study, shows a good result.

Choice of z_j In our simulation study, we set $J_p = \max\{2, \text{round}(p^{1/5})\}$ where round$(\cdot)$ is the rounding-off function. Then the values of z_j are set in the following recursive way: $z_1 = 1\mathbf{i}$, $z_2 = 2\mathbf{i}$ and $z_j = (z_{j-2} + z_{j-1})/2$ for $j \in \{3, \ldots, J_p\}$. For instance, if $p = 750$, then $J_p = 4$ and the elements in $\{z_1, \ldots, z_4\}$ are $z_1 = 1\,\mathbf{i}$, $z_2 = 2\,\mathbf{i}$, $z_3 = 1.5\,\mathbf{i}$ and $z_4 = 1.75\,\mathbf{i}$.

Choice of the grid For a finite set of real numbers S, define the set $M(S) = S \cup \{(s_j + s_{j+1})/2 : s_j \in S, 1 \le j \le |S| - 1\}$, where $|S|$ is the cardinality of S. Denote $\mathbf{h}_p = \{\ell_{p,1}, \ldots, \ell_{p,p}\}$. We choose the grid G_p in Corollary 1 to be $G_p = M\{M(\mathbf{h}_p)\}$.

3.2. A weakly consistent estimator of Ledoit and Wolf type

3.2.1. Quantized eigenvalues sampling transformation

In practice, the limiting quantities $D(t), c_1, c_2$ in (4) are not observable to us. Therefore, it is reasonable to assume for finite n_1, n_2 and p that $D(t) = D_p(t)$, $c_1 = c_{1,p}$, $c_2 = c_{2,p}$, namely, the LSD $D(t)$ equals the true finite-sample spectral distribution $D_p(t)$ of $\Sigma_{p,2}\Sigma_{p,1}^{-1}$ and the limiting ratios c_1, c_2 equal their finite-sample counterparts $c_{1,p}, c_{2,p}$. Then following Remark 1, we obtain a probability distribution $L_{D_p, c_{1,p}, c_{2,p}}$ which can be regarded as an approximation of the SD L_p of $S_{p,2}S_{p,1}^{-1}$. To construct the spectrum estimator, we define the function referred to as "Quantized Eigenvalues Sampling Transformation" or QuEST in short, whose name is inherited from [21]. In this article, QuEST is a generalization of the one in [21], from the one-sample problem to the two-sample problem.

Definition 1. Let $\mathbf{t} = (t_1, \ldots, t_r)$ be an r-dimensional vector whose components $t_1, \ldots, t_r$ have empirical distribution function G and $\mathbf{k} = (k_1, k_2)$ be a pair of numbers satisfying $k_1 \in (0, 1)$, $k_2 \in (0, 1) \cup (1, \infty)$. With the notation of Remark 1, QuEST is defined as the function $Q_{\mathbf{k},r}(\mathbf{t}) = (q^1_{\mathbf{k},r}(\mathbf{t}), \ldots, q^r_{\mathbf{k},r}(\mathbf{t}))^\top$ such that

$$\forall_{i \in \{1,\ldots,r\}} \quad q^i_{\mathbf{k},r}(\mathbf{t}) = r \int_{(i-1)/r}^{i/r} L^{-1}_{G,k_1,k_2}(u) du,$$

where L^{-1}_{G,k_1,k_2} is the quantile function of L_{G,k_1,k_2}.

Despite its complexity, in fact it is not difficult to compute the QuEST function along with its analytic Jacobian numerically; see Figure 1.

We now proceed to estimating the spectrum of $\Sigma_{p,2}\Sigma_{p,1}^{-1}$.

3.2.2. Constructing the weakly consistent estimator

Theorem 3. *Under Assumptions 1–3 and following the notation in Definition 1, let* $\tilde{\mathbf{d}}_p = (\tilde{d}_{p,1}, \ldots, \tilde{d}_{p,p})$ *be a vector such that the empirical distribution* $\tilde{D}_p$ *of* $\tilde{d}_{p,1}, \ldots, \tilde{d}_{p,p}$ *converges weakly to* D *as* $p \to \infty$ *and satisfies that* $\text{Supp}(\tilde{D}_p) \subset K$ *for all large* p. *Then we have as* $p \to \infty$,

$$\frac{1}{p} \sum_{i=1}^{p} \left\{ q^i_{\mathbf{c}_p,p}(\tilde{\mathbf{d}}_p) - \ell_{p,i} \right\}^2 \xrightarrow{a.s.} 0,$$

where $\ell_{p,i}$ *is the ith smallest eigenvalue of* $S_{p,2}S_{p,1}^{-1}$.

Theorem 4. *Under Assumptions 1–3 and following the notation in Definition 1 and Theorem 3, define*

$$\hat{\mathbf{d}}_p = \underset{\boldsymbol{\tau} \in K^p}{\text{argmin}} \frac{1}{p} \sum_{i=1}^{p} \left\{ q^i_{\mathbf{c}_p,p}(\boldsymbol{\tau}) - \ell_{p,i} \right\}^2,$$

where K^p *is the* p*-fold Cartesian product of* K. *Then as* $p \to \infty$,

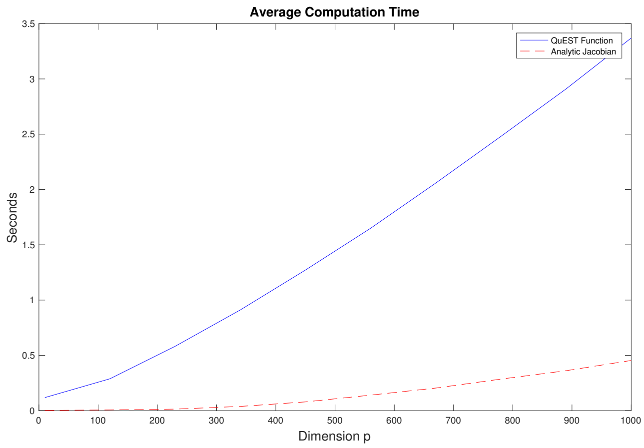


Figure 1: Average computation time of QuEST and its analytic Jacobian. The setting of the population parameters are p ranges from 20 to 1000 with $p/n_1, p/n_2$ fixed at $1/3$ and the ith smallest eigenvalue of $\Sigma_{p,2}\Sigma_{p,1}^{-1}$ is the $(i-0.5)/p$th theoretical quantile of the random variable $1+10\times\mathcal{B}(2,5)$, where $\mathcal{B}(2,5)$ is a random variable following the Beta distribution with parameters $(2,5)$. The QuEST function and the Jacobian are programmed in MATLAB. The computer is a 1.7 GHz dual-core laptop.

$$\frac{1}{p}\sum_{i=1}^{p}\left|\hat{d}_{p,i}-d_{p,i}\right| \xrightarrow{a.s.} 0,$$

where $\hat{d}_{p,i}$ is the ith smallest component of $\hat{\mathbf{d}}_p$.

Remark 3. The convergence result of Theorem 4 requires that we can obtain the global minimizer of the target function $p^{-1}\sum_{i=1}^{p}\{q^i_{\mathbf{c}_p,p}(\boldsymbol{\tau})-\ell_{p,i}\}^2$. While it is hardly possible to guarantee that the solution obtained from any prevailing optimization algorithms is the global minimizer due to the nonlinearity of the target function. Nonetheless, one way to still get a weakly consistent estimator in practice is to start from a good enough initial point. In particular, we have the following result.

Corollary 3. *Under Assumptions 1–3 and following the notation in Definition 1 and Theorem 3, suppose $\tilde{\mathbf{d}}_p$ is a p-variate vector whose components have empirical distribution function $\tilde{D}_p$ with* Supp$(\tilde{D}_p)$ *contained in K for all large p and $\tilde{D}_p$ converging weakly to D as $p\to\infty$. Define $\hat{\mathbf{d}}^{LW}_p=(\hat{d}^{LW}_{p,1},\ldots,\hat{d}^{LW}_{p,p})$ to be the vector such that $\hat{d}^{LW}_{p,1}\le\cdots\le\hat{d}^{LW}_{p,p}$, $\hat{d}^{LW}_{p,1},\ldots,\hat{d}^{LW}_{p,p}$ are uniformly contained in K and*

$$\frac{1}{p}\sum_{i=1}^{p}\left\{q^i_{\mathbf{c}_p,p}(\hat{\mathbf{d}}^{LW}_p)-\ell_{p,i}\right\}^2 \le \frac{1}{p}\sum_{i=1}^{p}\left\{q^i_{\mathbf{c}_p,p}(\tilde{\mathbf{d}}_p)-\ell_{p,i}\right\}^2,$$

for all large p. Then as $p\to\infty$,

$$\frac{1}{p}\sum_{i=1}^{p}\left|\hat{d}^{LW}_{p,i}-d_{p,i}\right| \xrightarrow{a.s.} 0.$$

3.2.3. Implementation details

Let $(\hat{\Sigma}^{ST}_{p,1},\hat{\Sigma}^{ST}_{p,2})$ be the Stein type estimator of the pair of population covariance matrices $(\Sigma_{p,1},\Sigma_{p,2})$ proposed in [28]. It is well known that $(\hat{\Sigma}^{ST}_{p,1},\hat{\Sigma}^{ST}_{p,2})$ already performs very well in estimating $(\Sigma_{p,1},\Sigma_{p,2})$. In our simulation study, we use the constrained nonlinear least square optimization package `lsqnonlin` in MATLAB to implement the optimization procedures. We start from two initial points, of which one is the spectrum of $\hat{\Sigma}^{ST}_{p,2}(\hat{\Sigma}^{ST}_{p,1})^{-1}$, the other is $(\hat{d}^{EK}_{p,1},\ldots,\hat{d}^{EK}_{p,p})$ presented in Corollary 2. Then we choose among the two obtained solutions from the optimization procedures the one with smaller value of the target function as our estimator denoted as $(\hat{d}^{LW}_{p,1}, \ldots, \hat{d}^{LW}_{p,p})$. Since it is proven that $(\hat{d}^{EK}_{p,1}, \ldots, \hat{d}^{EK}_{p,p})$ is weakly consistent, $(\hat{d}^{LW}_{p,1}, \ldots, \hat{d}^{LW}_{p,p})$ is thus also weakly consistent according to Corollary 3.

4. Estimation of $(\Sigma_{p,1}, \Sigma_{p,2})$

4.1. Introduction

Since Stein [34, 35] proposed the shrinkage estimator of covariance matrices, many efforts by various statisticians have been devoted to estimating covariance matrices in this direction. The literature includes [9, 15, 19, 20, 22, 28, 38]. It is well known that the eigenvalues of the sample covariance matrices are far more dispersed than those of the population eigenvalues. The idea of shrinkage estimation is to correct this distortion. In other words, it pulls down those large eigenvalues and pulls up the small ones. This shrinkage method gets the most risk savings when the population eigenvalues are close together. In this section, the problem of estimating two covariance matrices is examined. Namely, we consider the estimation of a pair of covariance matrices $(\Sigma_{p,1}, \Sigma_{p,2})$ with the aim of getting substantial savings in the risk when eigenvalues of $\Sigma_{p,2}\Sigma_{p,1}^{-1}$ are close together. This would be useful, for example, in estimating $(\Sigma_{p,1}, \Sigma_{p,2})$, when one has prior information that the eigenvalues of $\Sigma_{p,1}$, $\Sigma_{p,2}$ are likely to be far apart but the eigenvalues of $\Sigma_{p,1}$ are approximately proportional to those of $\Sigma_{p,2}$.

4.2. Equivariant estimators

We will base our estimation on the equivariant estimators using the sample covariance matrices $S_{p,1}$ and $S_{p,2}$. Denote a generic estimator to be $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$. We consider the loss function

$$L(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2}; \Sigma_{p,1}, \Sigma_{p,2}) = \frac{1}{p}\sum_{i=1}^{2}\{\mathrm{tr}(\Sigma_{p,i}^{-1}\hat{\Sigma}_{p,i}) - \ln|\Sigma_{p,i}^{-1}\hat{\Sigma}_{p,i}| - p\}. \tag{11}$$

It follows from classic linear algebra (see p. 31 of [14]) that there exists $p \times p$ invertible matrix G_p and diagonal matrix $F_p = \mathrm{diag}(f_{p,1}, \ldots, f_{p,p})$ with $f_{p,1} \le \cdots \le f_{p,p}$, such that

$$G_p^{-1}(S_{p,1} + S_{p,2})G_p^{\top -1} = I, \quad G_p^{-1}S_{p,2}G_p^{\top -1} = F_p. \tag{12}$$

In particular, $f_{p,1}, \ldots, f_{p,p}$ are the eigenvalues of $S_{p,2}(S_{p,1} + S_{p,2})^{-1}$, and it satisfies that $f_{p,i} = \ell_{p,i}(1 + \ell_{p,i})^{-1}$ where $\ell_{p,i}$ is the ith smallest eigenvalue of $S_{p,2}S_{p,1}^{-1}$.

Then the equivariant estimator has a general from presented in the following theorem:

Theorem 5. *Under the transformation $S_{p,1} \to AS_{p,1}A^\top$, $S_{p,2} \to AS_{p,2}A^\top$, where A is an arbitrary invertible $p \times p$ matrix, and under the loss function* (11), *$(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$ is the equivariant estimator of $(\Sigma_{p,1}, \Sigma_{p,2})$ if and only if $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$ can be expressed as*

$$\hat{\Sigma}_{p,1}(S_{p,1}, S_{p,2}) = G_p\Psi_p(F_p)G_p^\top, \quad \hat{\Sigma}_{p,2}(S_{p,1}, S_{p,2}) = G_p\Phi_p(F_p)G_p^\top,$$

where ψ_p and ϕ_p are univariate functions referred to as the shrinkage functions, $\Psi_p(F_p) = \mathrm{diag}\{\psi_p(f_{p,1}), \ldots, \psi_p(f_{p,p})\}$, $\Phi_p(F_p) = \mathrm{diag}\{\phi_p(f_{p,1}), \ldots, \phi_p(f_{p,p})\}$ are two diagonal matrices and G_p is the matrix defined in (12). *The mappings ψ_p and ϕ_p and their arguments may both depend on $f_{p,1}, \ldots, f_{p,p}$.*

4.3. The limiting loss

Assumptions 1–3 together with Lemma A.5 in [37] imply that almost surely there exists a compact subinterval $\mathbb{K}$ of $(0, 1)$ containing $\mathrm{Supp}(F_p)\backslash\{0\}$ and $\mathrm{Supp}(F)\backslash\{0\}$ for all large p. In the following, we impose an additional assumption for the problem of estimating $(\Sigma_{p,1}, \Sigma_{p,2})$.

Assumption 4. *There are two functions ψ and ϕ defined on $\mathbb{K} \cup \{0\}$ and continuous and positive on* $\mathrm{Supp}(F)$*, such that as $p \to \infty$, $\psi_p(x) \stackrel{a.s.}{\to} \psi(x)$, and $\phi_p(x) \stackrel{a.s.}{\to} \phi(x)$ uniformly for all $x \in \mathrm{Supp}(F)$ (Recall the functions ψ_p and ϕ_p are defined in the statement of Theorem 5). On top of that, there exists a finite nonrandom positive number κ such that almost surely, $|\psi_p(x)| \le \kappa$ and $|\phi_p(x)| \le \kappa$ uniformly for all n_1, n_2, p large enough and $x \in \mathbb{K}$.*

Theorem 6. *Under Assumptions 1–4 and with the notation of Section 2, the loss function* (11) *of any equivariant estimator* $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$ *proposed in Theorem 5 has the following limit:*

$$L(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2}; \Sigma_{p,1}, \Sigma_{p,2}) \xrightarrow{a.s.} \int_{\mathrm{Supp}(F)\setminus\{0\}} \Big[\frac{1 - c_1 + 2c_1(1-x)\{1 + x\mathrm{Re}\,\check{m}_F(x)\}}{1-x}\psi(x) - \ln\left\{\frac{\psi(x)}{1-x}\right\}$$
$$+ \frac{1 - c_2 + 2c_2 x\{1 - (1-x)\mathrm{Re}\,\check{m}_F(x)\}}{x}\phi(x) - \ln\left\{\frac{\phi(x)}{1-x}\right\}\Big] dF(x)$$
$$- \int \ln(x) dT^{\mathrm{inv}}(x) - \int \ln(x) d\mathscr{M}_{c_1}(x) - 2 + A, \tag{13}$$

where $\underline{F}(x) = c_2 F(x) - (c_2 - 1)\mathbf{1}(0 \le x)$ *is the companion distribution of* F, $\mathscr{M}_{c_1}(x)$ *is the Marčenko–Pastur distribution with parameter* c_1 *(see 3.1.1 of [5]) and* $A = 0$ *if* $c_2 \in (0,1)$ *while if* $c_2 \in (1,\infty)$,

$$A = \frac{c_2 - 1}{c_2}\left[\left(1 + \frac{c_1}{c_2}\right)\psi(0) + \left\{1 - \check{m}_{\underline{F}}(0) + \frac{c_2 \int t^{-1} dD(t)}{c_2 - 1}\right\}\phi(0) - \ln\{\psi(0)\} - \ln\{\phi(0)\}\right].$$

4.4. The oracle shrinkage functions

By differentiating the limiting loss over ψ and ϕ in Theorem 6, we have the following result.

Theorem 7. *Under Assumptions 1–4 and with the notation of Theorem 6, a covariance matrices estimator* $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$ *minimizes, in the class of equivariant estimators proposed in Theorem 5, the asymptotic loss* (13) *if and only if its limiting shrinkage functions* (ψ, ϕ) *is such that, for all* $x \in \mathrm{Supp}(F)$, $(\psi, \phi) = (\psi^{\mathrm{or}}, \phi^{\mathrm{or}})$ *where as* $c_2 \in (0,1)$, ψ^{or} *and* ϕ^{or} *are given, for all* $x \in \mathrm{Supp}(F)$, *by*

$$\psi^{\mathrm{or}}(x) = \frac{1-x}{1 - c_1 + 2c_1(1-x)\{1 + x\mathrm{Re}\,\check{m}_F(x)\}}, \tag{14}$$
$$\phi^{\mathrm{or}}(x) = \frac{x}{1 - c_2 + 2c_2 x\{1 - (1-x)\mathrm{Re}\,\check{m}_F(x)\}}. \tag{15}$$

When $c_2 \in (1,\infty)$, (14) *and* (15) *are still valid except at* $x = 0$ *and*

$$\psi^{\mathrm{or}}(0) = \frac{c_2}{c_1 + c_2}, \quad \phi^{\mathrm{or}}(0) = \left\{1 - \check{m}_{\underline{F}}(0) + \frac{c_2}{c_2 - 1}\int t^{-1} dD(t)\right\}^{-1}.$$

Proposition 1. *Under Assumptions 1–4 and with the notation of Theorem 7,* $\psi^{\mathrm{or}}(x)$ *and* $\phi^{\mathrm{or}}(x)$ *are positive and uniformly bounded for all* $x \in \{0\} \cup \mathbb{K}$.

Remark 4. We note that the proof of Proposition 1 does not require the distribution F to satisfy any particular property except that it is obtained from solving the two-sample MP equation in the sense of Remark 1. In particular, if we replace D by any probability distribution compactly supported on $(0,\infty)$ and c_1, c_2 by any other positive constants k_1, k_2 such that $k_1 \in (0,1)$, $k_2 \in (0,1) \cup (1,\infty)$, the positiveness and boundedness of the resulting shrinkage functions should still hold. We will see that the shrinkage functions of our bona fide estimator presented later on has the same structure as ψ^{or} and ϕ^{or} except that the limiting population spectral distribution D is substituted for a weakly consistent estimate $\hat{D}_p$ and c_1, c_2 are substituted for their finite-sample counterparts p/n_1, p/n_2. This implies that Proposition 1 will still be true for the bona fide shrinkage functions.

4.5. The bona fide estimator

It can be seen that the oracle shrinkage functions ψ^{or} and ϕ^{or} depend on the knowledge of the LSD F. In practice, F is always unknown. Therefore, we need to further estimate it through the data. Now that Section 3 has provided estimates of D we can make use of them to get estimates of F. Let $\hat{D}_p$ be a probability distribution such that almost surely $\hat{D}_p$ weakly converges to D as $p \to \infty$ with $\mathrm{Supp}(\hat{D}_p)$ uniformly contained in a compact subinterval of $(0,\infty)$ for all large p.

Let $F_{\hat{D}_p,c_{1,p},c_{2,p}}$ and $\underline{F}_{\hat{D}_p,c_{1,p},c_{2,p}}$ be the distributions defined in Remark 1 with $\hat{D}_p, c_{1,p}, c_{2,p}$ in place of G, k_1, k_2. The Stieltjes transforms of $F_{\hat{D}_p,c_{1,p},c_{2,p}}$ and $\underline{F}_{\hat{D}_p,c_{1,p},c_{2,p}}$ are denoted respectively as $m_{F_{\hat{D}_p,c_{1,p},c_{2,p}}}(z)$ and $m_{\underline{F}_{\hat{D}_p,c_{1,p},c_{2,p}}}(z)$.

For each $i \in \{1,\ldots,p\}$, let $\hat{d}_{p,i}$ denote an estimate of the ith smallest population eigenvalue $d_{p,i}$ such that $p^{-1}\sum_{i=1}^{p}\hat{d}_{p,i}^{-1}$ converges almost surely to $\int t^{-1}dD(t)$ as $p \to \infty$. We note that in practice, we can use either $\hat{d}_{p,i}^{EK}$ obtained from Corollary 2 or $\hat{d}_{p,i}^{LW}$ from Corollary 3 as $\hat{d}_{p,i}$. Now we propose our bona fide estimator.

Theorem 8. *Under Assumptions 1–4 and following the notation above, we define the bona fide shrinkage functions, for any p, n_1, n_2 and for all $x \in \mathbb{K}$, by*

$$\psi_p^{BF}(x) = \frac{1-x}{1-\frac{p}{n_1}+2\frac{p}{n_1}(1-x)\left\{1+x\mathrm{Re}\,\check{m}_{F_{\hat{D}_p,c_{1,p},c_{2,p}}}(x)\right\}}, \tag{16}$$

$$\phi_p^{BF}(x) = \frac{x}{1-\frac{p}{n_2}+2\frac{p}{n_2}x\left\{1-(1-x)\,\mathrm{Re}\,\check{m}_{F_{\hat{D}_p,c_{1,p},c_{2,p}}}(x)\right\}}. \tag{17}$$

If $p > n_2$, in addition to (16) *and* (17)*, the shrinkage functions are also defined at $x = 0$, viz.*

$$\psi_p^{BF}(0) = \frac{p/n_2}{p/n_1 + p/n_2}, \quad \phi_p^{BF}(0) = \left\{1 - \check{m}_{\underline{F}_{\hat{D}_p,c_{1,p},c_{2,p}}}(0) + \frac{\sum_{i=1}^{p}\hat{d}_{p,i}^{-1}}{p-n_2}\right\}^{-1}.$$

Then the covariance matrices estimator $(\hat{\Sigma}_{p,1}^{BF}, \hat{\Sigma}_{p,2}^{BF}) = (G_p\Psi_p^{BF}G_p^\top, G_p\Phi_p^{BF}G_p^\top)$ minimizes in the class of equivariant estimators the asymptotic loss (13) *as $p \to \infty$, where*

$$\Psi_p^{BF} = \mathrm{diag}(\psi_{p,1}^{BF},\ldots,\psi_{p,p}^{BF}), \quad \Phi_p^{BF} = \mathrm{diag}(\phi_{p,1}^{BF},\ldots,\phi_{p,p}^{BF})$$

are diagonal matrices and for all $i \in \{1,\ldots,p\}$, $\psi_{p,i}^{BF} = \psi_p^{BF}(f_{p,i})$, $\phi_{p,i}^{BF} = \phi_p^{BF}(f_{p,i})$.

We note that $f_{p,1},\ldots,f_{p,p}$ could be outside $\mathrm{Supp}(F_{\hat{D}_p,c_{1,p},c_{2,p}})$. Although this does not affect the estimator's asymptotic property, to facilitate computation, we propose the modified random matrix (RM) estimator given by the following corollary.

Corollary 4. *Under Assumptions 1–4 and with the notation in Theorem 8, the covariance matrices estimator $(\hat{\Sigma}_{p,1}^{RM}, \hat{\Sigma}_{p,2}^{RM}) = (G_p\Psi_p^{RM}G_p^\top, G_p\Phi_p^{RM}G_p^\top)$ minimizes in the class of equivariant estimators the asymptotic loss* (13) *as $p \to \infty$, where $\Psi_p^{RM} = \mathrm{diag}(\psi_{p,1}^{RM},\ldots,\psi_{p,p}^{RM})$, $\Phi_p^{RM} = \mathrm{diag}(\phi_{p,1}^{RM},\ldots,\phi_{p,p}^{RM})$ are diagonal matrices and for all $i \in \{1,\ldots,p\}$,*

$$\psi_{p,i}^{RM} = p\int_{(i-1)/p}^{i/p}\psi_p^{BF}\{F_{\hat{D}_p,c_{1,p},c_{2,p}}^{-1}(x)\}dx, \quad \phi_{p,i}^{RM} = p\int_{(i-1)/p}^{i/p}\phi_p^{BF}\{F_{\hat{D}_p,c_{1,p},c_{2,p}}^{-1}(x)\}dx,$$

where $F_{\hat{D}_p,c_{1,p},c_{2,p}}^{-1}$ is the quantile function of the distribution $F_{\hat{D}_p,c_{1,p},c_{2,p}}$ and the shrinkage functions ψ_p^{BF}, ϕ_p^{BF} are defined in Theorem 8.

4.6. Comparison with other estimators

4.6.1. Compared to the Stein type estimator

By an approximate minimization of the unbiased estimate of risk, Stein [34] (see also [35]) constructed an estimator of the covariance matrix for the one-sample problem whose risk compares very favorably with the minimax risk. In particular, substantial savings in risk is obtained when eigenvalues of the population covariance matrix are close together. Extending Stein's one-sample estimator, Loh [28] proposed the two-sample analogue $(\hat{\Sigma}_{p,1}^{ST}, \hat{\Sigma}_{p,2}^{ST})$ for the case $p \le \min(n_1, n_2)$. Similar to Stein's one-sample estimator, since the shrunk eigenvalues may not follow the natural ascending or descending order or may even have negative values, the isotonic regression needs to be applied to get the positive ordered eigenvalues. Under the diagonalization (12), $(\hat{\Sigma}_{p,1}^{ST}, \hat{\Sigma}_{p,2}^{ST})$ has the form

$$(\hat{\Sigma}_{p,1}^{ST}, \hat{\Sigma}_{p,2}^{ST}) = (G_p\Psi_p^{ST}G_p^\top, G_p\Phi_p^{ST}G_p^\top),$$

where Ψ_p^{ST} and Φ_p^{ST} are two diagonal matrices. Prior to the application of isotonic regression, the unordered diagonal entries of Ψ_p^{ST} and Φ_p^{ST} can be expressed respectively, for each $i \in \{1, \dots, p\}$, as

$$\psi_{p,i}^{ST\text{raw}} = (1 - f_{p,i}) \Bigg/ \left\{1 - \frac{p-1}{n_1} - \frac{2}{n_1} \sum_{j \neq i} \frac{f_{p,j}(1 - f_{p,i})}{f_{p,i} - f_{p,j}}\right\},$$
$$\phi_{p,i}^{ST\text{raw}} = f_{p,i} \Bigg/ \left\{1 - \frac{p-1}{n_2} + \frac{2}{n_2} \sum_{j \neq i} \frac{f_{p,i}(1 - f_{p,j})}{f_{p,i} - f_{p,j}}\right\}.$$

Note that in [28], the simultaneous diagonalization finds the invertible matrix B_p such that

$$B_p(n_1 S_{p,1} + n_2 S_{p,2})B_p^\top = I, \quad B_p(n_2 S_{p,2})B_p^\top = F_p.$$

Here we have adjusted the expressions of $\psi_{p,i}^{ST\text{raw}}$ and $\phi_{p,i}^{ST\text{raw}}$ for each $i \in \{1, \dots, p\}$ according to the diagonalization (12).

For a function $g(t)$, its Cauchy principal value at a point $x \in \mathbb{R}$ is defined as the limit of the integral

$$\mathrm{PV} \int_{-\infty}^{\infty} \frac{g(t)}{t - x} dt = \lim_{\varepsilon \to 0^+} \left\{ \int_{-\infty}^{x-\varepsilon} \frac{g(t)}{t - x} dt + \int_{x+\varepsilon}^{\infty} \frac{g(t)}{t - x} dt \right\}.$$

Recall the definition of $F_p(x)$ from Table 1. For each $i \in \{1, \dots, p\}$, $\psi_{p,i}^{ST\text{raw}}$ and $\phi_{p,i}^{ST\text{raw}}$ can be rewritten as

$$\psi_{p,i}^{ST\text{raw}} =(1 - f_{p,i}) \Bigg/ \left[1 - \frac{p-1}{n_1} + \frac{2p}{n_1}(1 - f_{p,i}) \left\{\frac{p-1}{p} + f_{p,i} \mathrm{PV} \int_{-\infty}^{\infty} \frac{1}{x - f_{p,i}} dF_p(x)\right\}\right],$$
$$\phi_{p,i}^{ST\text{raw}} =f_{p,i} \Bigg/ \left[1 - \frac{p-1}{n_1} + \frac{2p}{n_1} f_{p,i} \left\{\frac{p-1}{p} - (1 - f_{p,i}) \mathrm{PV} \int_{-\infty}^{\infty} \frac{1}{x - f_{p,i}} dF_p(x)\right\}\right].$$

We also note that

$$\mathrm{Re}\, \check{m}_{F_{\hat{D}_p, c_{1,p}, c_{2,p}}}(x) = \mathrm{PV} \int_{-\infty}^{\infty} (t - x)^{-1} dF_{\hat{D}_p, c_{1,p}, c_{2,p}}(t).$$

So for each $i \in \{1, \dots, p\}$, the diagonal entries $\psi_{p,i}^{BF}$ and $\phi_{p,i}^{BF}$ of our bona fide estimator $(\hat{\Sigma}_{p,1}^{BF}, \hat{\Sigma}_{p,2}^{BF})$ can be rewritten as

$$\psi_{p,i}^{BF} =(1 - f_{p,i}) \Bigg/ \left[1 - \frac{p}{n_1} + \frac{2p}{n_1}(1 - f_{p,i}) \left\{1 + f_{p,i} \mathrm{PV} \int_{-\infty}^{\infty} \frac{1}{x - f_{p,i}} dF_{\hat{D}_p, c_{1,p}, c_{2,p}}(x)\right\}\right],$$
$$\phi_{p,i}^{BF} =f_{p,i} \Bigg/ \left[1 - \frac{p}{n_1} + \frac{2p}{n_1} f_{p,i} \left\{1 - (1 - f_{p,i}) \mathrm{PV} \int_{-\infty}^{\infty} \frac{1}{x - f_{p,i}} dF_{\hat{D}_p, c_{1,p}, c_{2,p}}(x)\right\}\right].$$

The only substantial difference is that the step function F_p is replaced by the smooth function $F_{\hat{D}_p, c_{1,p}, c_{2,p}}$.

Despite the similarity, there are some advantages of our proposed estimator. First, it has been shown that our shrinkage functions are always nonnegative, therefore the ad hoc correction of the negativeness and disorder of the shrunk eigenvalues via isotonic regression is no longer needed. Second, our estimator is able to deal with the more challenging singular case where $p > n_2$ while the Stein type estimator is only defined when $p \leq \min(n_1, n_2)$. Third, the optimality of our estimator does not require particular distribution assumption on the data matrix, while the Stein type estimator relies on the unbiased estimate of the risk which is derived only when the two sample covariance matrices follow Wishart distributions. Finally, simulation results show that the shrinkage functions $\psi_p^{BF}(x)$ and $\phi_p^{BF}(x)$ may follow neither ascending nor descending order, which suggests that the order correction via isotonic regression on the Stein type estimator may be misleading. The last point may seem counter-intuitive at the beginning, because it is expected that the shrinkage function should preserve the monotonicity of the sample eigenvalues $f_{p,1}, \dots, f_{p,p}$ just as it does in the one-sample case. In the following, we present a justification of it. We recall from Theorem 5 that the two-sample covariance matrices estimator is constructed to be of the form

$$(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2}) = (G_p \Psi_p(F_p) G_p', G_p \Phi_p(F_p) G_p'),$$

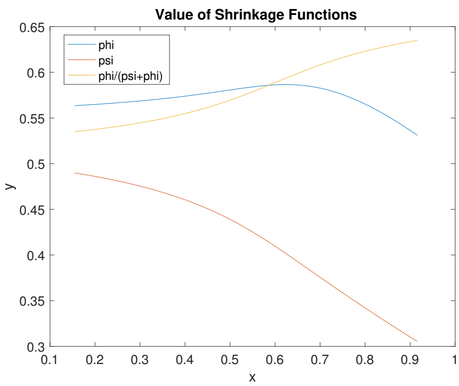


Figure 2: Value of the shrinkage functions $\psi_p^{BF}(x)$, $\phi_p^{BF}(x)$ and $\phi_p^{BF}(x)/\{\psi_p^{BF}(x)+\phi_p^{BF}(x)\}$ given $p = 200$, $n_1 = n_2 = 600$ and $\hat{D}_p = 2^{-1}\mathbf{1}(1 \le x) + 2^{-1}\mathbf{1}(2 \le x)$, where the blue, red and yellow curves are the graphs of $\phi_p^{BF}(x)$, $\psi_p^{BF}(x)$ and $\phi_p^{BF}(x)/\{\psi_p^{BF}(x)+\phi_p^{BF}(x)\}$ respectively. As displayed in the plot, $\phi_p^{BF}(x)$ is not monotone, $\psi_p^{BF}(x)$ is decreasing and $\phi_p^{BF}(x)/\{\psi_p^{BF}(x)+\phi_p^{BF}(x)\}$ is increasing.

where $\Psi_p(F_p) = \mathrm{diag}\{\psi_p(f_{p,1}), \ldots, \psi_p(f_{p,p})\}$ and $\Phi_p(F_p) = \mathrm{diag}\{\phi_p(f_{p,1}), \ldots, \phi_p(f_{p,p})\}$. A crucial difference between the two-sample and the one-sample shrinkage estimators is that in the two-sample case, since G_p is not an orthogonal matrix, $\psi_p(f_{p,i})$ and $\phi_p(f_{p,i})$ are not eigenvalues of the covariance matrices estimators $\hat{\Sigma}_{p,1}$ and $\hat{\Sigma}_{p,2}$. Consequently, the shrinkage functions may not follow the same order of $f_{p,i}$ as it does in the one-sample case.

We recall from Theorem 8 that the two shrinkage functions ψ_p^{BF} and ϕ_p^{BF} are constructed such that $(\hat{\Sigma}_{p,1}^{BF}, \hat{\Sigma}_{p,2}^{BF})$ has optimal risk performance among the class of equivariant estimators. Except from being positive, continuous, bounded and uniformly convergent as required by Assumption 4, we do not impose any other restrictions on ψ_p^{BF} and ϕ_p^{BF}. In fact the two functions may even not be monotone; it is possible that $\phi_p^{BF}(f_{p,i})/\{\psi_p^{BF}(f_{p,i}) + \phi_p^{BF}(f_{p,i})\}$ has the same order of $f_{p,1}, \ldots, f_{p,p}$ because the former is the ith eigenvalue of $\hat{\Sigma}_{p,2}(\hat{\Sigma}_{p,1} + \hat{\Sigma}_{p,2})^{-1}$ which has the same eigenvectors of $S_{p,2}(S_{p,1} + S_{p,2})^{-1}$. Here we present a concrete example as an illustration.

Consider the case when $p = 200$, $n_1 = n_2 = 600$ and $\hat{D}_p = 2^{-1}\mathbf{1}(1 \le x) + 2^{-1}\mathbf{1}(2 \le x)$ where $\mathbf{1}$ is the indicator function. In other words, the estimated population spectral distribution $\hat{D}_p$ is a discrete distribution which places half its mass at 1 and the other half at 2. Drawing the result from numerical computation, we obtain the plot in Figure 2.

4.6.2. Minimaxity

Suppose the $p \times p$ matrix S follows the Wishart distribution $W_p(n^{-1}\Sigma, n)$. Under the transformation $\Sigma \to U\Sigma U^\top$, $S \to USU^\top$, where U is $p \times p$ orthogonal, Stein [34] and Dey and Srivinasan [11] independently proposed an estimator of Σ which is shown to be minimax. Its two-sample analogue $(\hat{\Sigma}_{p,1}^{AU}, \hat{\Sigma}_{p,2}^{AU})$ can be found in [27, 28] wherein this type of estimator is referred to as the adjusted usual estimator. With respect to the diagonalization (12), $(\hat{\Sigma}_{p,1}^{AU}, \hat{\Sigma}_{p,2}^{AU})$ has the form $(G_p\Psi_p^{AU}G_p^\top, G_p\Phi_p^{AU}G_p^\top)$ where Ψ_p^{AU} and Φ_p^{AU} are diagonal matrices whose ith diagonal entries can be expressed respectively, for each $i \in \{1, \ldots, p\}$, as

$$\psi_{p,i}^{AU} = (1 - f_{p,i}) \Big/ \left(1 + \frac{p+1}{n_1} - \frac{2i}{n_1}\right), \quad \phi_{p,i}^{AU} = f_{p,i} \Big/ \left(1 - \frac{p+1}{n_2} + \frac{2i}{n_2}\right).$$

Note that Loh [28] sorts the eigenvalues $f_{p,1}, \ldots, f_{p,p}$ in descending order whereas we sort them in ascending order.

With respect to the interval $\mathbb{K}$ in Assumption 4, suppose we define the shrinkage functions, for all $x \in \mathbb{K}$, by

$$\psi_p^{AU}(x) = \frac{1-x}{1 + \frac{p+1}{n_1} - \frac{2p}{n_1}F_p(x)}, \quad \phi_p^{AU}(x) = \frac{x}{1 - \frac{p+1}{n_2} + \frac{2p}{n_2}F_p(x)}.$$

Then we see that $\psi_{p,i}^{AU} = \psi_p^{AU}(f_{p,i})$, $\phi_{p,i}^{AU} = \phi_p^{AU}(f_{p,i})$. One can easily check that under Assumptions 1–3, $\psi_p^{AU}(x)$ and $\phi_p^{AU}(x)$ respectively converge uniformly almost surely for $x \in \mathbb{K}$ to

$$\frac{1-x}{1 + c_1 - 2c_1F(x)}, \quad \frac{x}{1 - c_2 + 2c_2F(x)},$$

which implies that the adjusted usual estimator belongs to the class of estimators satisfying Assumption 4.

We recall that the limiting shrinkage functions of the bona fide estimators equal the oracle shrinkage functions $(\psi^{\rm or}, \phi^{\rm or})$ which minimizes the asymptotic loss. This implies that given Wishart distributed sample covariance matrices $S_{p,1}$ and $S_{p,2}$, if we restrict the population covariance matrices in the class satisfying Assumption 3, our bona fide estimator dominates the adjusted usual estimator in the limit as $p, n_1, n_2 \to \infty$ with $p/n_1 \to c_1 \in (0, 1)$ and $p/n_2 \to c_2 \in (0, 1)$ and thus is minimax. Actually, our estimator minimizes the asymptotic loss for every limiting spectral distribution D of the ratio of the population covariance matrices. This is a much stronger notion of optimality than minimax which only minimizes the worst case risk.

4.6.3. Simultaneous estimation vs. estimating individually

Since the data are given as two independent sample covariance matrices $S_{p,1}$ and $S_{p,2}$, a natural question is how the performance of simultaneous estimation compares to that of estimating $\Sigma_{p,1}$ and $\Sigma_{p,2}$ individually. We call the estimator simultaneously estimating a pair of covariance matrices the two-sample estimator and the estimator estimating only one covariance matrix the one-sample estimator. Denoting the one-sample covariance matrix estimators proposed in [22] as $\hat{\Sigma}^{\rm one}_{p,1}$ and $\hat{\Sigma}^{\rm one}_{p,2}$ respectively for $\Sigma_{p,1}$ and $\Sigma_{p,2}$, we can construct an estimator of $(\Sigma_{p,1}, \Sigma_{p,2})$ as $(\hat{\Sigma}^{\rm one}_{p,1}, \hat{\Sigma}^{\rm one}_{p,2})$. Seen from the assumptions in [22], the one-sample covariance matrix estimator minimizes the limiting loss only when the eigenvalues of the population covariance matrices are bounded from both below and above for all large p, thus we may expect performance depreciation of $(\hat{\Sigma}^{\rm one}_{p,1}, \hat{\Sigma}^{\rm one}_{p,2})$ when the eigenvalues of $\Sigma_{p,1}$ and $\Sigma_{p,2}$ are highly dispersed. However, our two-sample estimator always performs well as long as the eigenvalues of $\Sigma_{p,2}\Sigma^{-1}_{p,1}$ are close together.

5. Simulations

In this section, we compare the performance of our estimators proposed in this article with several existing estimators for different settings of the population parameters.

For the problem of estimating the population spectrum $\mathbf{d}_p = \{d_{p,1}, \ldots, d_{p,p}\}$, let $\hat{\mathbf{d}}_p = \{\hat{d}_{p,1}, \ldots, \hat{d}_{p,p}\}$ such that $\hat{d}_{p,1} \leq \cdots \leq \hat{d}_{p,p}$ denote a generic estimator. The loss function we use is the normalized L_1 distance, viz.

$$\frac{1}{p}\sum_{i=1}^{p}\left|\hat{d}_{p,i} - d_{p,i}\right|.$$

The estimators to be compared are

$\hat{\mathbf{d}}^{Dey}_p$: The estimator proposed by Dey [10] which is shown to dominate the scaled eigenvalues of $S_{p,2}S^{-1}_{p,1}$ under the squared error loss $\sum_{i=1}^{p}(\hat{d}_{p,i} - d_{p,i})^2$.

$\hat{\mathbf{d}}^{ST}_p$: The spectrum of $\hat{\Sigma}^{ST}_{p,2}(\hat{\Sigma}^{ST}_{p,1})^{-1}$ where the pair $(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$ is the Stein type covariance matrices estimator proposed by Loh [28]. It is demonstrated by Monte Carlo simulations in [28] that $(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$ has good risk performance estimating $(\Sigma_{p,1}, \Sigma_{p,2})$. It can also be seen in our simulation study that the spectrum of $\hat{\Sigma}^{ST}_{p,2}(\hat{\Sigma}^{ST}_{p,1})^{-1}$ as an estimator of $\mathbf{d}_p$ also performs very well. Therefore, we consider this estimator as a benchmark for our proposed estimators to compare with.

$\hat{\mathbf{d}}^{EK}_p$: The El Karoui type estimator proposed in Corollary 2.

$\hat{\mathbf{d}}^{LW}_p$: The Ledoit and Wolf type estimator proposed in Corollary 3.

$\hat{\mathbf{d}}^{\rm one}_p$: The spectrum of the matrix $\hat{\Sigma}^{\rm one}_{p,2}(\hat{\Sigma}^{\rm one}_{p,1})^{-1}$ where $\hat{\Sigma}^{\rm one}_{p,1}$, $\hat{\Sigma}^{\rm one}_{p,2}$, are the one-sample covariance matrices estimators proposed by Ledoit and Wolf [22] based on RMT.

For the problem of estimating $(\Sigma_{p,1}, \Sigma_{p,2})$, let $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$ denote a generic estimator. The loss function we use is the Stein's loss, viz.

$$L(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2}; \Sigma_{p,1}, \Sigma_{p,2}) = \frac{1}{p}\sum_{i=1}^{2}\{\mathrm{tr}(\Sigma^{-1}_{p,i}\hat{\Sigma}_{p,i}) - \ln|\Sigma^{-1}_{p,i}\hat{\Sigma}_{p,i}| - p\}.$$

The estimators to be compared are

$(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$: The minimax adjusted usual estimator; see [28].

$(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$: The Stein type estimator proposed by Loh [28].

$(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$: The oracle estimator which is defined as the estimator obtained in Corollary 4 but with the true population spectral distribution D_p in place of the estimate $\hat{D}_p$. Since $(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$ incorporates the true population parameter D_p, theoretically the loss of this estimator is the best that a bona fide estimator can achieve. We use it as a benchmark for our proposed estimators to compare with.

$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$: Our random matrix estimator proposed in Corollary 4 with $\check{D}^{EK}_p$ acting as $\hat{D}_p$, where $\check{D}^{EK}_p$ is the empirical distribution of $\{\hat{d}^{EK}_{p,1}, \ldots, \hat{d}^{EK}_{p,p}\}$ defined in Corollary 2.

$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$: Our random matrix estimator proposed in Corollary 4 with $\hat{D}^{LW}_p$ acting as $\hat{D}_p$, where $\hat{D}^{LW}_p$ is the empirical distribution of $\{\hat{d}^{LW}_{p,1}, \ldots, \hat{d}^{LW}_{p,p}\}$ defined in Corollary 3.

$(\hat{\Sigma}^{\mathrm{one}}_{p,1}, \hat{\Sigma}^{\mathrm{one}}_{p,2})$: The pair of one-sample covariance matrices estimators proposed by Ledoit and Wolf [22] based on RMT.

We focus on three scenarios to demonstrate the performance of our proposed estimators. Since the covariance matrices estimators are equivariant, we may without loss of generality set $\Sigma_{p,1}$ and $\Sigma_{p,2}$ to be diagonal matrices. In the following, we let $\mathbf{d}_p = \{d_{p,1}, \ldots, d_{p,p}\}$ denote the spectrum of $\Sigma_{p,2}\Sigma_{p,1}^{-1}$. For $i \in \{1, 2\}$, let $\lambda_i = \{\lambda_{i,1}, \ldots, \lambda_{i,p}\}$ such that $\lambda_{i,1} \leq \cdots \leq \lambda_{i,p}$ denote the spectrum of $\Sigma_{p,i}$. The dimension p ranges in the set $\{50, 200, 500, 750, 1000\}$. Each time, we run 400 Monte Carlo repetitions to get the average loss and standard error.

Scenario 1. We consider three different settings of $(\Sigma_{p,1}, \Sigma_{p,2})$. In all the settings, we set the matrices X_p, Y_p to be independent, each consisting of iid standard normal $\mathcal{N}(0, 1)$ random variables. Then the observed sample covariance matrices are given as $n_1^{-1}\Sigma_{p,1}^{1/2}X_pX_p^\top\Sigma_{p,1}^{1/2}$ and $n_2^{-1}\Sigma_{p,2}^{1/2}Y_pY_p^\top\Sigma_{p,2}^{1/2}$. In the **id** setting, we set $\lambda_{1,j} = \lambda_{2,j} = d_{p,j} = 1$ for all $j \in \{1, \ldots, p\}$ so that $\Sigma_{p,2}\Sigma_{p,1}^{-1}$ is the identity matrix. In the **blk** setting, $\lambda_{1,j} = 1$ and $\lambda_{2,j} = d_{p,j} = \mathbf{1}(j \leq p/2) + 10\mathbf{1}(p/2 < j \leq p)$ for all $j \in \{1, \ldots, p\}$. In the **beta** setting, $\lambda_{1,j} = 1$ and $\lambda_{2,j}$ and $d_{p,j}$ equal the $(j - 0.5)/p$th theoretical quantile of the random variable following the distribution $1 + 10 \times \mathcal{B}(2, 5)$ for all $j \in \{1, \ldots, p\}$, where $\mathcal{B}(2, 5)$ is the beta distribution with parameters $(2, 5)$. For each dimension p, we consider three different pairs of the sample sizes (n_1, n_2): $(3p, 3p)$, $(3p, p)$ and $(3p, p/2)$.

For the last case, $S_{p,2}$ is singular and $\hat{\mathbf{d}}^{Dey}_p$, $\hat{\mathbf{d}}^{ST}_p$, $(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$ and $(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$ are not defined, so we compare only $\hat{\mathbf{d}}^{EK}_p$, $\hat{\mathbf{d}}^{LW}_p$, $(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$, $(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$ and $(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$. The results are summarized in Table 2 for $\hat{\mathbf{d}}_p$ and Table 5 for $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$. We see that our high-dimensional estimators both for estimating $\mathbf{d}_p$ and $(\Sigma_{p,1}, \Sigma_{p,2})$ outperform the traditional estimators for all the three settings of $(\Sigma_{p,1}, \Sigma_{p,2})$ especially when $p \geq 500$.

Scenario 2. We focus on the comparison between the two-sample estimators and the one-sample estimators. As explained in Section 4.6.3, when $\Sigma_{p,1}$, $\Sigma_{p,2}$ are approximately proportional but the eigenvalues of each of the population covariance matrices are far apart, the performance of the one-sample estimators may not be good, whereas our two-sample estimators still perform well. In this scenario, we still use normally distributed data and set $\mathbf{d}_p$ to be the same as those in **id** and **beta** settings of Scenario 1. But the individual spectra of $\Sigma_{p,1}$ and $\Sigma_{p,2}$ are more spread out. The ith smallest eigenvalues of $\Sigma_{p,1}$, $\Sigma_{p,2}$ are respectively set to be $\lambda_{1,j} = 10^{-3}\mathbf{1}(j \leq p/2) + 10^3\mathbf{1}(p/2 < j \leq p)$ and $\lambda_{2,j,} = d_{p,j} \times \lambda_{1,j}$ for all $j \in \{1, \ldots, p\}$. We set $(n_1, n_2) = (3p, 3p)$. The results are summarized in Table 3 for $\hat{\mathbf{d}}_p$ and Table 6 for $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$. We see that the performance of both $\hat{\mathbf{d}}^{\mathrm{one}}_p$ and $(\hat{\Sigma}^{\mathrm{one}}_{p,1}, \hat{\Sigma}^{\mathrm{one}}_{p,2})$ are not satisfying, while our two-sample estimators are still robust.

Scenario 3. We focus on the case where the data no longer follow normal distribution. Specifically, we set the matrices X_p, Y_p to be independent, each consisting of iid random variables following Student t-distribution with degrees of freedom 5. Moreover, these t-distributed random variables are multiplied by $\sqrt{3/5}$ so that they have unit variance. Then the observed sample covariance matrices are given as $n_1^{-1}\Sigma_{p,1}^{1/2}X_pX_p^\top\Sigma_{p,1}^{1/2}$ and $n_2^{-1}\Sigma_{p,2}^{1/2}Y_pY_p^\top\Sigma_{p,2}^{1/2}$. We set $(\Sigma_{p,1}, \Sigma_{p,2})$ to be the same as that in **beta** setting of Scenario 1. The sample sizes are (n_1, n_2): $(3p, 3p)$, $(3p, p)$ and

$(3p, p/2)$. The results are summarized in Table 4 for $\hat{\mathbf{d}}_p$ and Table 7 for $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$. It is seen that with non-normally distributed data, our proposed estimators still perform well.

L_1, Loss and SE in Tables 2 to 7 respectively stand for the normalized L_1 distance $p^{-1}\sum_{i=1}^{p}|\hat{d}_{p,i}-d_{p,i}|$ for estimating $\mathbf{d}_p$, the Stein's loss (11) and the estimated standard errors from the Monte Carlo repetitions.

Acknowledgments

This paper forms a part of my PhD thesis [37]. I would like to express my deep gratitude to my advisor, Professor Wei-Liem Loh, whose constant support, encouragement and guidance made this thesis possible. I would also like to thank Professors Olivier Ledoit and Michael Wolf for sending me their MATLAB program, Professors Wang Zhou, Shurong Zheng and Zhidong Bai for their technical help to my thesis and Professor Sanjay Chaudhuri for a discussion on the numerical implementation of the El Karoui type estimator. Finally I would like to thank the Editor-in-Chief and an anonymous referee for valuable comments, which have resulted in an improved exposition of this paper. I was supported in part by grant AcRF Tier 1 R-155-000-166-112 at National University of Singapore.

Appendix A. Proofs of mathematical results in Section 2

Appendix A.1. Proof of Theorem 1

Proof. Using the definition of Stieltjes transform, Eq. (3) can be rewritten, for all $z \in \mathbb{C}^+$, as

$$z = -\frac{1}{m_{\underline{L}}(z)} + c_2 m_{T^{\mathrm{inv}}}\{-m_{\underline{L}}(z)\}. \tag{A.1}$$

Using the equality $m_{\underline{T}^{\mathrm{inv}}}(z) = -(1-c_1)/z + c_1 m_{T^{\mathrm{inv}}}(z)$, (A.1) further gives rise to

$$z = -\frac{1}{m_{\underline{L}}(z)} + \frac{c_2}{c_1} m_{\underline{T}^{\mathrm{inv}}}\{-m_{\underline{L}}(z)\} - \frac{c_2 - c_1 c_2}{c_1 m_{\underline{L}}(z)} = -\frac{c_1 + c_2 - c_1 c_2}{c_1 m_{\underline{L}}(z)} + \frac{c_2}{c_1} m_0(z). \tag{A.2}$$

Replacing z by $-m_{\underline{L}}(z)$ in (2), we obtain, for all $z \in \mathbb{C}^+$,

$$-m_{\underline{L}}(z) = -\frac{1}{m_0(z)} + c_1 \int \frac{dD(t)}{t + m_0(z)}, \tag{A.3}$$

which is the desired Eq. (5).

Plugging (A.3) into (A.2), we obtain Eq. (4).

Next, let $z = z_1 + \mathbf{i}z_2$ and $m_0 = m_1 + \mathbf{i}m_2$, where z_1, z_2, m_1, m_2 are respectively the real part, the imaginary part of z, the real part and the imaginary part of m_0. We show the uniqueness of m_0 by arguing that for a fixed $z \in \mathbb{C}^+$, there is at most one $m_0 \in \mathbb{C}^-$ such that

$$z = \frac{c_1 + c_2 - c_1 c_2}{c_1\left\{-\frac{1}{m_0} + c_1 \int \frac{dD(t)}{t+m_0}\right\}} + \frac{c_2}{c_1} m_0. \tag{A.4}$$

Let $c_1 + c_2 - c_1 c_2 = h$ and rewrite Eq. (A.4) as

$$\frac{h}{zc_1 - c_2 m_0} = -\frac{1}{m_0} + c_1 \int \frac{dD(t)}{t + m_0}. \tag{A.5}$$

Taking imaginary parts and multiplying $1/m_2$ on both sides, we obtain

$$-\frac{h(z_2 c_1 - c_2 m_2)}{m_2 |zc_1 - c_2 m_0|^2} = \frac{1}{|m_0|^2} - c_1 \int \frac{dD(t)}{|t + m_0|^2}.$$

Since $z_2 > 0$, $m_2 < 0$, it follows that

$$0 < \frac{hc_2}{|zc_1 - c_2 m_0|^2} < -\frac{h(z_2 c_1 - c_2 m_2)}{m_2 |zc_1 - c_2 m_0|^2} = \frac{1}{|m_0|^2} - c_1 \int \frac{dD(t)}{|t + m_0|^2}. \tag{A.6}$$

Table 2: Comparison of average loss of the proposed spectrum estimators with other existing methods given normally distributed data.

(n_1, n_2)	$\hat{\mathbf{d}}_p$	$p = 50$		$p = 200$		$p = 500$		$p = 750$		$p = 1000$	
		L_1	SE	L_1	SE	L_1	SE	L_1	SE	L_1	SE
						id					
$(3p, 3p)$	$\hat{\mathbf{d}}_p^{Dey}$	0.534	(0.001)	0.534	(0.000)	0.533	(0.000)	0.533	(0.000)	0.533	(0.000)
	$\hat{\mathbf{d}}_p^{ST}$	0.091	(0.001)	0.033	(0.000)	0.016	(0.000)	0.011	(0.000)	0.009	(0.000)
	$\hat{\mathbf{d}}_p^{EK}$	0.070	(0.004)	0.023	(0.001)	0.011	(0.001)	0.008	(0.001)	0.007	(0.000)
	$\hat{\mathbf{d}}_p^{LW}$	0.097	(0.004)	0.036	(0.002)	0.017	(0.001)	0.013	(0.001)	0.012	(0.001)
$(3p, p)$	$\hat{\mathbf{d}}_p^{Dey}$	2.387	(0.006)	2.375	(0.001)	2.373	(0.001)	2.370	(0.000)	2.371	(0.000)
	$\hat{\mathbf{d}}_p^{ST}$	0.212	(0.002)	0.121	(0.001)	0.083	(0.000)	0.070	(0.000)	0.062	(0.000)
	$\hat{\mathbf{d}}_p^{EK}$	0.101	(0.005)	0.032	(0.002)	0.016	(0.001)	0.011	(0.001)	0.009	(0.001)
	$\hat{\mathbf{d}}_p^{LW}$	0.135	(0.005)	0.046	(0.002)	0.026	(0.001)	0.017	(0.001)	0.014	(0.001)
$(3p, p/2)$	$\hat{\mathbf{d}}_p^{EK}$	0.158	(0.007)	0.046	(0.002)	0.022	(0.001)	0.015	(0.001)	0.012	(0.001)
	$\hat{\mathbf{d}}_p^{LW}$	0.187	(0.008)	0.072	(0.003)	0.041	(0.002)	0.029	(0.002)	0.026	(0.002)
						blk					
$(3p, 3p)$	$\hat{\mathbf{d}}_p^{Dey}$	3.440	(0.007)	3.416	(0.002)	3.408	(0.001)	3.405	(0.000)	3.405	(0.000)
	$\hat{\mathbf{d}}_p^{ST}$	1.383	(0.009)	1.233	(0.002)	1.192	(0.001)	1.183	(0.001)	1.177	(0.001)
	$\hat{\mathbf{d}}_p^{EK}$	0.763	(0.028)	0.124	(0.006)	0.054	(0.003)	0.032	(0.002)	0.024	(0.001)
	$\hat{\mathbf{d}}_p^{LW}$	0.615	(0.017)	0.221	(0.010)	0.097	(0.005)	0.072	(0.004)	0.055	(0.003)
$(3p, p)$	$\hat{\mathbf{d}}_p^{Dey}$	11.671	(0.040)	11.594	(0.010)	11.576	(0.004)	11.565	(0.003)	11.565	(0.002)
	$\hat{\mathbf{d}}_p^{ST}$	2.310	(0.008)	2.224	(0.002)	2.196	(0.001)	2.189	(0.001)	2.186	(0.001)
	$\hat{\mathbf{d}}_p^{EK}$	1.311	(0.043)	0.281	(0.012)	0.138	(0.007)	0.097	(0.005)	0.070	(0.003)
	$\hat{\mathbf{d}}_p^{LW}$	1.122	(0.031)	0.343	(0.013)	0.175	(0.008)	0.118	(0.006)	0.092	(0.005)
$(3p, p/2)$	$\hat{\mathbf{d}}_p^{EK}$	3.132	(0.121)	1.117	(0.045)	0.509	(0.021)	0.352	(0.016)	0.288	(0.013)
	$\hat{\mathbf{d}}_p^{LW}$	2.554	(0.076)	1.151	(0.039)	0.651	(0.029)	0.518	(0.025)	0.411	(0.021)
						beta					
$(3p, 3p)$	$\hat{\mathbf{d}}_p^{Dey}$	1.492	(0.004)	1.484	(0.001)	1.481	(0.000)	1.480	(0.000)	1.480	(0.000)
	$\hat{\mathbf{d}}_p^{ST}$	0.363	(0.003)	0.319	(0.001)	0.335	(0.001)	0.340	(0.000)	0.342	(0.000)
	$\hat{\mathbf{d}}_p^{EK}$	0.545	(0.012)	0.420	(0.008)	0.278	(0.005)	0.251	(0.004)	0.230	(0.004)
	$\hat{\mathbf{d}}_p^{LW}$	0.716	(0.008)	0.450	(0.007)	0.288	(0.006)	0.235	(0.005)	0.191	(0.004)
$(3p, p)$	$\hat{\mathbf{d}}_p^{Dey}$	8.172	(0.024)	8.122	(0.006)	8.112	(0.002)	8.103	(0.002)	8.104	(0.001)
	$\hat{\mathbf{d}}_p^{ST}$	0.908	(0.005)	0.863	(0.001)	0.849	(0.001)	0.845	(0.000)	0.842	(0.000)
	$\hat{\mathbf{d}}_p^{EK}$	1.292	(0.032)	0.977	(0.013)	0.805	(0.008)	0.806	(0.006)	0.773	(0.007)
	$\hat{\mathbf{d}}_p^{LW}$	1.001	(0.012)	0.681	(0.011)	0.512	(0.012)	0.447	(0.012)	0.380	(0.011)
$(3p, p/2)$	$\hat{\mathbf{d}}_p^{EK}$	1.463	(0.026)	1.338	(0.018)	1.211	(0.017)	1.052	(0.014)	0.964	(0.012)
	$\hat{\mathbf{d}}_p^{LW}$	1.316	(0.018)	1.021	(0.010)	0.924	(0.009)	0.855	(0.009)	0.833	(0.008)

Table 3: Comparison of average loss of two-sample spectrum estimators with the one-sample estimators given normally distributed data.

$\hat{\mathbf{d}}_p$	$p = 50$		$p = 200$		$p = 500$		$p = 750$		$p = 1000$	
	L_1	SE	L_1	SE	L_1	SE	L_1	SE	L_1	SE
id										
$\hat{\mathbf{d}}_p^{Dey}$	2.387	(0.006)	2.375	(0.001)	2.373	(0.001)	2.370	(0.000)	2.371	(0.000)
$\hat{\mathbf{d}}_p^{ST}$	0.212	(0.002)	0.121	(0.001)	0.083	(0.000)	0.070	(0.000)	0.062	(0.000)
$\hat{\mathbf{d}}_p^{EK}$	0.101	(0.005)	0.032	(0.002)	0.016	(0.001)	0.011	(0.001)	0.008	(0.001)
$\hat{\mathbf{d}}_p^{LW}$	0.133	(0.005)	0.046	(0.002)	0.026	(0.001)	0.017	(0.001)	0.014	(0.001)
$\hat{\mathbf{d}}_p^{\text{one}}$	200.770	(103.920)	54.310	(15.360)	19.287	(3.725)	9.752	(1.853)	6.961	(1.297)
beta										
$\hat{\mathbf{d}}_p^{Dey}$	8.172	(0.024)	8.122	(0.006)	8.112	(0.002)	8.103	(0.002)	8.104	(0.001)
$\hat{\mathbf{d}}_p^{ST}$	0.908	(0.005)	0.863	(0.001)	0.849	(0.001)	0.845	(0.000)	0.842	(0.000)
$\hat{\mathbf{d}}_p^{EK}$	1.292	(0.032)	0.976	(0.013)	0.805	(0.008)	0.808	(0.006)	0.774	(0.007)
$\hat{\mathbf{d}}_p^{LW}$	1.006	(0.012)	0.676	(0.011)	0.510	(0.011)	0.445	(0.012)	0.377	(0.011)
$\hat{\mathbf{d}}_p^{\text{one}}$	558.820	(219.860)	144.360	(31.900)	50.322	(9.619)	29.734	(5.476)	21.154	(2.434)

Table 4: Comparison of average loss of the proposed spectrum estimators with other existing methods given t_5 distributed data.

(n_1, n_2)	$\hat{\mathbf{d}}_p$	$p = 50$		$p = 200$		$p = 500$		$p = 750$		$p = 1000$	
		L_1	SE	L_1	SE	L_1	SE	L_1	SE	L_1	SE
$(3p, 3p)$	$\hat{\mathbf{d}}_p^{Dey}$	1.520	(0.004)	1.490	(0.001)	1.484	(0.000)	1.482	(0.000)	1.481	(0.000)
	$\hat{\mathbf{d}}_p^{ST}$	0.373	(0.004)	0.299	(0.002)	0.323	(0.001)	0.331	(0.001)	0.336	(0.000)
	$\hat{\mathbf{d}}_p^{EK}$	0.543	(0.011)	0.460	(0.009)	0.302	(0.006)	0.322	(0.007)	0.276	(0.005)
	$\hat{\mathbf{d}}_p^{LW}$	0.761	(0.008)	0.470	(0.008)	0.301	(0.006)	0.261	(0.006)	0.219	(0.005)
$(3p, p)$	$\hat{\mathbf{d}}_p^{Dey}$	8.457	(0.040)	8.225	(0.009)	8.166	(0.004)	8.141	(0.003)	8.132	(0.002)
	$\hat{\mathbf{d}}_p^{ST}$	0.952	(0.007)	0.861	(0.002)	0.845	(0.001)	0.843	(0.001)	0.840	(0.001)
	$\hat{\mathbf{d}}_p^{EK}$	1.490	(0.042)	0.976	(0.012)	0.812	(0.010)	0.801	(0.006)	0.782	(0.007)
	$\hat{\mathbf{d}}_p^{LW}$	1.065	(0.015)	0.692	(0.011)	0.494	(0.011)	0.452	(0.012)	0.383	(0.011)
$(3p, p/2)$	$\hat{\mathbf{d}}_p^{EK}$	1.644	(0.038)	1.430	(0.022)	1.198	(0.017)	1.084	(0.015)	0.975	(0.011)
	$\hat{\mathbf{d}}_p^{LW}$	1.437	(0.023)	1.052	(0.010)	0.931	(0.009)	0.880	(0.009)	0.845	(0.008)

Table 5: Comparison of average loss of the proposed covariance matrices estimators with other existing methods given normally distributed data.

(n_1, n_2)	$(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$	$p = 50$		$p = 200$		$p = 500$		$p = 750$		$p = 1000$	
		Loss	SE	Loss	SE	Loss	SE	Loss	SE	Loss	SE
						id					
$(3p, 3p)$	$(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$	0.1869	(0.0004)	0.1789	(0.0001)	0.1775	(0.0000)	0.1772	(0.0000)	0.1771	(0.0000)
	$(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$	0.2584	(0.0004)	0.2526	(0.0001)	0.2513	(0.0000)	0.2511	(0.0000)	0.2510	(0.0000)
	$(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$	0.1813	(0.0004)	0.1778	(0.0001)	0.1772	(0.0000)	0.1770	(0.0000)	0.1770	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.1822	(0.0004)	0.1780	(0.0001)	0.1772	(0.0000)	0.1771	(0.0000)	0.1770	(0.0000)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.1820	(0.0004)	0.1779	(0.0001)	0.1772	(0.0000)	0.1770	(0.0000)	0.1770	(0.0000)
$(3p, p)$	$(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$	0.4714	(0.0036)	0.3795	(0.0013)	0.3455	(0.0006)	0.3326	(0.0005)	0.3257	(0.0004)
	$(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$	0.6563	(0.0033)	0.6200	(0.0010)	0.6109	(0.0004)	0.6077	(0.0003)	0.6065	(0.0002)
	$(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$	0.2801	(0.0006)	0.2756	(0.0001)	0.2745	(0.0001)	0.2743	(0.0000)	0.2742	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.3303	(0.0054)	0.2938	(0.0023)	0.2837	(0.0013)	0.2822	(0.0010)	0.2805	(0.0008)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.3203	(0.0037)	0.2905	(0.0016)	0.2804	(0.0008)	0.2793	(0.0007)	0.2780	(0.0005)
$(3p, p/2)$	$(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$	0.3252	(0.0006)	0.3196	(0.0002)	0.3183	(0.0001)	0.3181	(0.0000)	0.3180	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.4997	(0.0133)	0.3824	(0.0054)	0.3417	(0.0021)	0.3312	(0.0014)	0.3271	(0.0009)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.4733	(0.0114)	0.3619	(0.0040)	0.3339	(0.0018)	0.3260	(0.0010)	0.3227	(0.0006)
						blk					
$(3p, 3p)$	$(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$	0.2722	(0.0005)	0.2641	(0.0001)	0.2625	(0.0000)	0.2622	(0.0000)	0.2621	(0.0000)
	$(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$	0.3006	(0.0005)	0.2946	(0.0001)	0.2933	(0.0000)	0.2931	(0.0000)	0.2930	(0.0000)
	$(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$	0.2603	(0.0005)	0.2559	(0.0001)	0.2551	(0.0000)	0.2549	(0.0000)	0.2549	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.2648	(0.0005)	0.2566	(0.0001)	0.2553	(0.0000)	0.2551	(0.0000)	0.2550	(0.0000)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.2617	(0.0005)	0.2562	(0.0001)	0.2552	(0.0000)	0.2550	(0.0000)	0.2549	(0.0000)
$(3p, p)$	$(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$	0.6238	(0.0035)	0.5473	(0.0011)	0.5195	(0.0006)	0.5089	(0.0004)	0.5033	(0.0003)
	$(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$	0.7230	(0.0033)	0.6874	(0.0009)	0.6785	(0.0004)	0.6753	(0.0003)	0.6741	(0.0002)
	$(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$	0.4643	(0.0008)	0.4582	(0.0002)	0.4567	(0.0001)	0.4563	(0.0001)	0.4562	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.4823	(0.0018)	0.4766	(0.0022)	0.4657	(0.0011)	0.4619	(0.0008)	0.4610	(0.0006)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.4941	(0.0027)	0.4710	(0.0014)	0.4627	(0.0008)	0.4602	(0.0006)	0.4590	(0.0005)
$(3p, p/2)$	$(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$	0.6287	(0.0009)	0.6212	(0.0002)	0.6197	(0.0001)	0.6194	(0.0001)	0.6193	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.8904	(0.0168)	0.8067	(0.0123)	0.7266	(0.0079)	0.6885	(0.0056)	0.6707	(0.0044)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.8579	(0.0150)	0.7381	(0.0093)	0.6826	(0.0060)	0.6533	(0.0039)	0.6447	(0.0033)
						beta					
$(3p, 3p)$	$(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$	0.2348	(0.0004)	0.2282	(0.0001)	0.2270	(0.0000)	0.2267	(0.0000)	0.2266	(0.0000)
	$(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$	0.2718	(0.0004)	0.2661	(0.0001)	0.2649	(0.0000)	0.2647	(0.0000)	0.2646	(0.0000)
	$(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$	0.2179	(0.0004)	0.2146	(0.0001)	0.2139	(0.0000)	0.2138	(0.0000)	0.2137	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.2199	(0.0004)	0.2152	(0.0001)	0.2143	(0.0000)	0.2141	(0.0000)	0.2141	(0.0000)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.2198	(0.0004)	0.2148	(0.0001)	0.2140	(0.0000)	0.2138	(0.0000)	0.2138	(0.0000)
$(3p, p)$	$(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$	0.5313	(0.0036)	0.4524	(0.0012)	0.4244	(0.0006)	0.4137	(0.0004)	0.4079	(0.0004)
	$(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$	0.6711	(0.0033)	0.6352	(0.0009)	0.6261	(0.0004)	0.6229	(0.0003)	0.6217	(0.0002)
	$(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$	0.3348	(0.0006)	0.3304	(0.0001)	0.3293	(0.0001)	0.3291	(0.0000)	0.3290	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.4012	(0.0058)	0.3618	(0.0030)	0.3468	(0.0017)	0.3417	(0.0012)	0.3399	(0.0011)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.3873	(0.0040)	0.3501	(0.0016)	0.3407	(0.0010)	0.3375	(0.0008)	0.3361	(0.0007)
$(3p, p/2)$	$(\hat{\Sigma}^{\mathrm{or}}_{p,1}, \hat{\Sigma}^{\mathrm{or}}_{p,2})$	0.3950	(0.0007)	0.3897	(0.0002)	0.3883	(0.0001)	0.3880	(0.0000)	0.3880	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.6604	(0.0168)	0.5727	(0.0109)	0.5298	(0.0081)	0.4911	(0.0063)	0.4771	(0.0056)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.5904	(0.0138)	0.5054	(0.0091)	0.4580	(0.0061)	0.4334	(0.0041)	0.4262	(0.0035)

Table 6: Comparison of average loss of two-sample covariance matrices estimators with the one-sample estimators given normally distributed data.

$(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$	$p = 50$		$p = 200$		$p = 500$		$p = 750$		$p = 1000$	
	Loss	SE	Loss	SE	Loss	SE	Loss	SE	Loss	SE
					id					
$(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$	0.1869	(0.0004)	0.1789	(0.0001)	0.1775	(0.0000)	0.1772	(0.0000)	0.1771	(0.0000)
$(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$	0.2584	(0.0004)	0.2526	(0.0001)	0.2513	(0.0000)	0.2511	(0.0000)	0.2510	(0.0000)
$(\hat{\Sigma}^{\text{or}}_{p,1}, \hat{\Sigma}^{\text{or}}_{p,2})$	0.1809	(0.0004)	0.1778	(0.0001)	0.1772	(0.0000)	0.1770	(0.0000)	0.1770	(0.0000)
$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.1822	(0.0004)	0.1780	(0.0001)	0.1772	(0.0000)	0.1770	(0.0000)	0.1770	(0.0000)
$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.1824	(0.0004)	0.1779	(0.0001)	0.1772	(0.0000)	0.1770	(0.0000)	0.1770	(0.0000)
$(\hat{\Sigma}^{\text{one}}_{p,1}, \hat{\Sigma}^{\text{one}}_{p,2})$	1649.1000	(61.3000)	348.9300	(15.2770)	133.7000	(5.9688)	85.1220	(3.6893)	63.3150	(2.5022)
					beta					
$(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$	0.2348	(0.0004)	0.2282	(0.0001)	0.2270	(0.0000)	0.2267	(0.0000)	0.2266	(0.0000)
$(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$	0.2718	(0.0004)	0.2661	(0.0001)	0.2649	(0.0000)	0.2647	(0.0000)	0.2646	(0.0000)
$(\hat{\Sigma}^{\text{or}}_{p,1}, \hat{\Sigma}^{\text{or}}_{p,2})$	0.2179	(0.0004)	0.2146	(0.0001)	0.2139	(0.0000)	0.2138	(0.0000)	0.2137	(0.0000)
$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.2199	(0.0004)	0.2150	(0.0001)	0.2141	(0.0000)	0.2140	(0.0000)	0.2139	(0.0000)
$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.2198	(0.0004)	0.2148	(0.0001)	0.2140	(0.0000)	0.2138	(0.0000)	0.2138	(0.0000)
$(\hat{\Sigma}^{\text{one}}_{p,1}, \hat{\Sigma}^{\text{one}}_{p,2})$	1579.5000	(50.5130)	326.5000	(13.7870)	179.4600	(7.2440)	152.9500	(5.0493)	143.6200	(3.3201)

Table 7: Comparison of average loss of the proposed covariance matrices estimators with other existing methods given t_5 distributed data.

(n_1, n_2)	$(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$	$p = 50$		$p = 200$		$p = 500$		$p = 750$		$p = 1000$	
		Loss	SE	Loss	SE	Loss	SE	Loss	SE	Loss	SE
$(3p, 3p)$	$(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$	0.2512	(0.0005)	0.2334	(0.0001)	0.2291	(0.0000)	0.2281	(0.0000)	0.2276	(0.0000)
	$(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$	0.2917	(0.0005)	0.2725	(0.0001)	0.2677	(0.0001)	0.2665	(0.0000)	0.2659	(0.0000)
	$(\hat{\Sigma}^{\text{or}}_{p,1}, \hat{\Sigma}^{\text{or}}_{p,2})$	0.2317	(0.0004)	0.2191	(0.0001)	0.2158	(0.0000)	0.2150	(0.0000)	0.2147	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.2344	(0.0005)	0.2195	(0.0001)	0.2159	(0.0000)	0.2152	(0.0000)	0.2148	(0.0000)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.2345	(0.0005)	0.2194	(0.0001)	0.2158	(0.0000)	0.2151	(0.0000)	0.2147	(0.0000)
$(3p, p)$	$(\hat{\Sigma}^{ST}_{p,1}, \hat{\Sigma}^{ST}_{p,2})$	0.5595	(0.0038)	0.4598	(0.0011)	0.4266	(0.0006)	0.4173	(0.0005)	0.4107	(0.0004)
	$(\hat{\Sigma}^{AU}_{p,1}, \hat{\Sigma}^{AU}_{p,2})$	0.7031	(0.0035)	0.6451	(0.0009)	0.6304	(0.0004)	0.6272	(0.0003)	0.6253	(0.0002)
	$(\hat{\Sigma}^{\text{or}}_{p,1}, \hat{\Sigma}^{\text{or}}_{p,2})$	0.3566	(0.0008)	0.3366	(0.0002)	0.3321	(0.0001)	0.3309	(0.0001)	0.3303	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.4368	(0.0062)	0.3685	(0.0028)	0.3507	(0.0016)	0.3438	(0.0013)	0.3417	(0.0011)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.4192	(0.0042)	0.3588	(0.0017)	0.3439	(0.0010)	0.3384	(0.0008)	0.3364	(0.0006)
$(3p, p/2)$	$(\hat{\Sigma}^{\text{or}}_{p,1}, \hat{\Sigma}^{\text{or}}_{p,2})$	0.4208	(0.0010)	0.3965	(0.0003)	0.3917	(0.0001)	0.3901	(0.0001)	0.3895	(0.0000)
	$(\hat{\Sigma}^{EK}_{p,1}, \hat{\Sigma}^{EK}_{p,2})$	0.7222	(0.0183)	0.5942	(0.0119)	0.5382	(0.0083)	0.5077	(0.0069)	0.4762	(0.0054)
	$(\hat{\Sigma}^{LW}_{p,1}, \hat{\Sigma}^{LW}_{p,2})$	0.6646	(0.0160)	0.5060	(0.0091)	0.4545	(0.0055)	0.4386	(0.0043)	0.4301	(0.0037)

Suppose $\mathbf{m}_0 = \mathbf{m}_1 + \mathbf{i}\mathbf{m}_2 \in \mathbb{C}^-$ is another solution different from m_0 to (A.5). Substituting m_0 and $\mathbf{m}_0$ in (A.5) and taking difference, we obtain

$$\frac{hc_2(m_0 - \mathbf{m}_0)}{(zc_1 - c_2m_0)(zc_1 - c_2\mathbf{m}_0)} = \frac{(m_0 - \mathbf{m}_0)}{m_0\mathbf{m}_0} - c_1 \int \frac{(m_0 - \mathbf{m}_0)dD(t)}{(t + m_0)(t + \mathbf{m}_0)}. \tag{A.7}$$

Dividing both sides of (A.7) by $m_0 - \mathbf{m}_0$, we get

$$\frac{hc_2}{(zc_1 - c_2m_0)(zc_1 - c_2\mathbf{m}_0)} = \frac{1}{m_0\mathbf{m}_0} - c_1 \int \frac{dD(t)}{(t + m_0)(t + \mathbf{m}_0)}. \tag{A.8}$$

Applying the Cauchy–Schwarz inequality to the right-hand side of (A.8) and using (A.6), we find

$$\begin{aligned}
\left|\frac{1}{m_0\mathbf{m}_0} - c_1 \int \frac{dD(t)}{(t + m_0)(t + \mathbf{m}_0)}\right| &\geq \left|\frac{1}{m_0\mathbf{m}_0}\right| - \left|c_1 \int \frac{dD(t)}{(t + m_0)(t + \mathbf{m}_0)}\right| \\
&\geq \frac{1}{|m_0||\mathbf{m}_0|} - \left\{c_1 \int \frac{dD(t)}{|t + m_0|^2}\right\}^{1/2} \left\{c_1 \int \frac{dD(t)}{|t + \mathbf{m}_0|^2}\right\}^{1/2} \\
&\geq \left\{\frac{1}{|m_0|^2} - c_1 \int \frac{dD(t)}{|t + m_0|^2}\right\}^{1/2} \left\{\frac{1}{|\mathbf{m}_0|^2} - c_1 \int \frac{dD(t)}{|t + \mathbf{m}_0|^2}\right\}^{1/2} \\
&> \left|\frac{hc_2}{(zc_1 - c_2m_0)(zc_1 - c_2\mathbf{m}_0)}\right|,
\end{aligned}$$

which contradicts (A.8).

To verify the second last inequality, we claim for any $x_1 \geq y_1 \geq 0,\ x_2 \geq y_2 \geq 0$,

$$x_1x_2 - y_1y_2 \geq (x_1^2 - y_1^2)^{1/2}(x_2^2 - y_2^2)^{1/2}.$$

If the claim is true, we take $x_1 = 1/|m_0|$, $x_2 = 1/|\mathbf{m_0}|$, $y_1 = \{c_1 \int |t + m_0|^{-2}dD(t)\}^{1/2}$, $y_2 = \{c_1 \int |t + \mathbf{m}_0|^{-2}dD(t)\}^{1/2}$, and the proof is done. To show the claim, we use the symbol $\iff$ for equivalence and see that

$$\begin{aligned}
x_1x_2 - y_1y_2 \geq (x_1^2 - y_1^2)^{1/2}(x_2^2 - y_2^2)^{1/2} &\iff (x_1x_2 - y_1y_2)^2 \geq (x_1^2 - y_1^2)(x_2^2 - y_2^2) \\
&\iff (x_1^2x_2^2 + y_1^2y_2^2 - 2x_1x_2y_1y_2) \geq (x_1^2x_2^2 + y_1^2y_2^2 - x_1^2y_2^2 - x_2^2y_1^2) \\
&\iff -2x_1y_2x_2y_1 \geq -x_1^2y_2^2 - x_2^2y_1^2 \\
&\iff 0 \geq -(x_1y_2 - x_2y_1)^2,
\end{aligned}$$

which is obviously true. □

Appendix B. Proofs of mathematical results in Section 3

Lemma 1. *Let* $\mathbf{x} = (x_1, \ldots x_p)$ *be a* p*-variate real vector with nonnegative components. Then*

$$p^{-1}\sum_{i=1}^{p} x_i \leq \sqrt{p^{-1}\sum_{i=1}^{p} x_i^2}.$$

Proof. Let $\mathbf{1} = (1, \ldots, 1)$ be the p-variate vector with all components 1. Then it follows from the Cauchy–Schwarz inequality that $|\mathbf{1} \cdot \mathbf{x}| \leq \sqrt{p \sum_{i=1}^{p} x_i^2}$. Dividing both sides by p, we get the result. □

Appendix B.1. Proofs of Theorem 2 and Corollary 1

We refer the readers to the proofs of Theorem 3.1 and Corollary 3.1 in Part I of [37] for the details.

Appendix B.2. Proof of Corollary 2

It is shown in Corollary 3.2 in Part I of [37] that $p^{-1}\sum_{i=1}^{p}(\hat{d}_{p,i}^{EK}-d_{p,i})^2 \xrightarrow{a.s.} 0$. We note that it is assumed that $c_2 \in (0,1)$ in [37] whereas the arguments actually also apply to the case $c_2 \in (1,\infty)$. Therefore, the desired result follows from Lemma 1 and $p^{-1}\sum_{i=1}^{p}(\hat{d}_{p,i}^{EK}-d_{p,i})^2 \xrightarrow{a.s.} 0$.

Appendix B.3. Proof of Theorem 3

The proof is the same as the one of Theorem 3 in Part I of [37].

Appendix B.4. Proof of Theorem 4

It is shown in Theorem 3.3 in Part I in [37] that $p^{-1}\sum_{i=1}^{p}(\hat{d}_{p,i}-d_{p,i})^2 \xrightarrow{a.s.} 0$. We note that it is assumed that $c_2 \in (0,1)$ in [37] whereas the arguments actually also apply to the case $c_2 \in (1,\infty)$. Therefore, the desired result follows from Lemma 1 and $p^{-1}\sum_{i=1}^{p}(\hat{d}_{p,i}-d_{p,i})^2 \xrightarrow{a.s.} 0$.

Appendix B.5. Proof of Corollary 3

It is shown in Theorem 3.4 in Part I [37] that $p^{-1}\sum_{i=1}^{p}(\hat{d}_{p,i}^{LW}-d_{p,i})^2 \xrightarrow{a.s.} 0$. We note that it is assumed that $c_2 \in (0,1)$ in [37] whereas the arguments also apply to the case $c_2 \in (1,\infty)$. Therefore, the desired result follows from Lemma 1 and $p^{-1}\sum_{i=1}^{p}(\hat{d}_{p,i}^{LW}-d_{p,i})^2 \xrightarrow{a.s.} 0$.

Appendix C. Proofs of mathematical results in Section 4

With the notation of Assumptions 1–4, let $q_k = n_2^{-1/2}\Sigma_{p,2}^{1/2}Y_{\cdot k}$ for each $k \in \{1,\dots,n_2\}$, where $Y_{\cdot k}$ is the kth column of Y_p, $S_{p,2}^{(k)} = S_{p,2} - q_k q_k^\top$, $W_p = S_{p,1} + S_{p,2}$ and $W_p^{(k)} = W_p - q_k q_k^\top$. We note that since $p/n_1 \to c_1 \in (0,1)$, $S_{p,1}$ is invertible for all large p, n_1. Therefore W_p and $W_p^{(k)}$ are invertible for all large p, n_1, n_2.

Denote $(W_p^{(k)})^{-1/2}$ as the square root of $(W_p^{(k)})^{-1}$, $\tilde{S}_{p,2}^{(k)} = (W_p^{(k)})^{-1/2}S_{p,2}^{(k)}(W_p^{(k)})^{-1/2}$ and $\tilde{\Sigma}_{p,2}^{(k)} = (W_p^{(k)})^{-1/2}\Sigma_{p,2}(W_p^{(k)})^{-1/2}$. It follows from Assumptions 3 that the spectral norms of $\tilde{S}_{p,2}^{(k)}$ and $\tilde{\Sigma}_{p,2}^{(k)}$ satisfy that $\|\tilde{S}_{p,2}^{(k)}\|$ and $\|\tilde{\Sigma}_{p,2}^{(k)}\|$ are uniformly bounded for all large p and $k \in \{1,\dots,n_2\}$. Then we have the following results.

Lemma 2. *Let $\Theta_p(z) = p^{-1}\mathrm{tr}\{(S_{p,2}-zW_p)^{-1}\Sigma_{p,2}\}$, $z \in \mathbb{C}^+$. Then under Assumptions 1–3, $\Theta_p(z)$ converges almost surely to a nonrandom limit, viz.*

$$\Theta(z) = \frac{1+zm_F(z)}{1-c_2(1-z)-c_2z(1-z)m_F(z)},$$

where m_F is defined in Table 1.

Proof. Observe that $S_{p,2}W_p^{-1} - zI + zI = S_{p,2}W_p^{-1} = \sum_{k=1}^{n_2} q_k q_k^\top W_p^{-1}$. Multiplying $p^{-1}(S_{p,2}W_p^{-1}-zI)^{-1}$ on both sides of $S_{p,2}W_p^{-1} - zI + zI = \sum_{k=1}^{n_2} q_k q_k^\top W_p^{-1}$ and taking trace, we get

$$1 + z\frac{1}{p}\mathrm{tr}\Big(S_{p,2}W_p^{-1}-zI\Big)^{-1} = \frac{1}{p}\sum_{k=1}^{n_2} q_k^\top W_p^{-1}\Big(S_{p,2}W_p^{-1}-zI\Big)^{-1}q_k = \frac{1}{p}\sum_{k=1}^{n_2} q_k^\top\Big(S_{p,2}-zW_p\Big)^{-1}q_k.$$

It follows from Eq. (2.2) of [3] that

$$\begin{aligned}
\frac{1}{p}\sum_{k=1}^{n_2} q_k^\top\Big(S_{p,2}-zW_p\Big)^{-1}q_k &= \frac{1}{p}\sum_{k=1}^{n_2}\frac{(1-z)^{-1}q_k^\top(S_{p,2}^{(k)}-zW_p^{(k)})q_k}{(1-z)^{-1}+q_k^\top(S_{p,2}^{(k)}-zW_p^{(k)})q_k}\\
&= \frac{n_2(1-z)^{-1}}{p} - \frac{1}{p}\sum_{k=1}^{n_2}\frac{(1-z)^{-2}}{(1-z)^{-1}+q_k^\top(S_{p,2}^{(k)}-zW_p^{(k)})q_k}.\\
&= \frac{n_2(1-z)^{-1}}{p} - \frac{n_2p^{-1}(1-z)^{-2}}{(1-z)^{-1}+n_2^{-1}\mathrm{tr}\{(S_{p,2}-zW_p)^{-1}\Sigma_{p,2}\}} + \delta_p, \qquad \text{(C.1)}
\end{aligned}$$

where

$$\delta_p = \frac{1}{p}\sum_{k=1}^{n_2} \frac{(1-z)^{-2}\Big[q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})q_k - n_2^{-1}\mathrm{tr}\{(S_{p,2} - zW_p)^{-1}\Sigma_{p,2}\}\Big]}{\Big\{(1-z)^{-1} + q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})q_k\Big\}\Big[(1-z)^{-1} + n_2^{-1}\mathrm{tr}\{(S_{p,2} - zW_p)^{-1}\Sigma_{p,2}\}\Big]}.$$

It follows from Lemma B.26 of [5] that there exists a constant α_p depending only on p such that

$$\begin{aligned}\mathrm{E}\Big|q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})q_k - n_2^{-1}\mathrm{tr}\{(S_{p,2}^{(k)} - zW_p^{(k)})^{-1}\Sigma_{p,2}\}\Big|^6 &\le \alpha_p p^3 n_2^{-6}\|\tilde\Sigma_{p,2}^{(k)}\|^6 \|(\tilde S_{p,2}^{(k)} - zI)^{-1}\|^6 \\ &\le \alpha_p p^3 (\mathrm{Im}\, z)^{-6} n_2^{-6}\|\tilde\Sigma_{p,2}^{(k)}\|^6\end{aligned}$$

for all $k \in \{1,\ldots,n_2\}$. From the uniform boundedness of $\|\tilde\Sigma_{p,2}^{(k)}\|$ and Borel–Cantelli Lemma, it follows that

$$\max_{k\in\{1,\ldots,n_2\}} \Big|q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})q_k - n_2^{-1}\mathrm{tr}\{(S_{p,2}^{(k)} - zW_p^{(k)})^{-1}\Sigma_{p,2}\}\Big| \xrightarrow{a.s.} 0. \tag{C.2}$$

We observe that $(S_{p,2}^{(k)} - zW_p^{(k)})^{-1} - (S_{p,2} - zW_p)^{-1} = (1-z)(S_{p,2} - zW_p)^{-1} q_k q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})^{-1}$. Using Eq. (2.2) of [3], we have

$$\begin{aligned} n_2^{-1}\Big|\mathrm{tr}\Big[\Big\{(S_{p,2}^{(k)} - zW_p^{(k)})^{-1} - (S_{p,2} - zW_p)^{-1}\Big\}\Sigma_{p,2}\Big]\Big| &= n_2^{-1}|1-z|^{-1}\Big|\mathrm{tr}\Big\{(S_{p,2} - zW_p)^{-1} q_k q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})^{-1}\Sigma_{p,2}\Big\}\Big| \\ &= n_2^{-1}|1-z|^{-1}\left|\frac{q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})^{-1}\Sigma_{p,2}(S_{p,2}^{(k)} - zW_p^{(k)})^{-1}q_k}{(1-z)^{-1} + q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})^{-1}q_k}\right| \\ &= n_2^{-1}|1-z|^{-1}\left|\frac{q_k^\top (W_p^{(k)})^{-1/2}(\tilde S_{p,2}^{(k)} - zI)^{-1}\tilde\Sigma_{p,2}^{(k)}(\tilde S_{p,2}^{(k)} - zI)^{-1}(W_p^{(k)})^{-1/2}q_k}{(1-z)^{-1} + q_k^\top (W_p^{(k)})^{-1/2}(\tilde S_{p,2}^{(k)} - zI)^{-1}(W_p^{(k)})^{-1/2}q_k}\right| \\ &\le n_2^{-1}\frac{|1-z|^{-1}\,\|\tilde\Sigma_{p,2}^{(k)}\|\,\|(\tilde S_{p,2}^{(k)} - zI)^{-1}(W_p^{(k)})^{-1/2}q_k\|_2^2}{\mathrm{Im}\Big\{(1-z)^{-1} + q_k^\top (W_p^{(k)})^{-1/2}(\tilde S_{p,2}^{(k)} - zI)^{-1}(W_p^{(k)})^{-1/2}q_k\Big\}},\end{aligned}$$

where $\|\cdot\|_2$ is the Euclidean norm of vectors.

Writing $\tilde S_{p,2}^{(k)} = \sum_{i=1}^p \theta_i e_i e_i^\top$, where the e_is are the orthonormal eigenvectors of $\tilde S_{p,2}^{(k)}$ and the θ_is are the eigenvalues, we have

$$\|(\tilde S_{p,2}^{(k)} - zI)^{-1}(W_p^{(k)})^{-1/2}q_k\|_2^2 = \sum_{i=1}^p \frac{|e_i^\top (W_p^{(k)})^{-1/2}q_k|^2}{|\theta_i - z|^2},$$

and

$$\begin{aligned}\mathrm{Im}\Big\{(1-z)^{-1} + q_k^\top (W_p^{(k)})^{-1/2}(\tilde S_{p,2}^{(k)} - zI)^{-1}(W_p^{(k)})^{-1/2}q_k\Big\} &= \frac{\mathrm{Im}\, z}{|1-z|^2} + \mathrm{Im}\, z\sum_{i=1}^p \frac{|e_i^\top (W_p^{(k)})^{-1/2}q_k|^2}{|\theta_i - z|^2} \\ &\ge \mathrm{Im}\, z\sum_{i=1}^p \frac{|e_i^\top (W_p^{(k)})^{-1/2}q_k|^2}{|\theta_i - z|^2}.\end{aligned}$$

Then it follows that

$$n_2^{-1}\Big|\mathrm{tr}\Big[\Big\{(S_{p,2}^{(k)} - zW_p^{(k)})^{-1} - (S_{p,2} - zW_p)^{-1}\Big\}\Sigma_{p,2}\Big]\Big| \le n_2^{-1}(\mathrm{Im}\, z)^{-1}|1-z|^{-1}\|\tilde\Sigma_{p,2}^{(k)}\|. \tag{C.3}$$

From (C.2), (C.3) and the uniform boundedness of $\|\tilde\Sigma_{p,2}^{(k)}\|$, we get that, for $z \in \mathbb{C}^+$, and as $p \to \infty$,

$$\max_{k\in\{1,\ldots,n_2\}} \Big|q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})q_k - n_2^{-1}\mathrm{tr}\{(S_{p,2} - zW_p)^{-1}\Sigma_{p,2}\}\Big| \xrightarrow{a.s.} 0.$$

Hence, taking limits on both sides of (C.1) we get

$$1 + zm_F(z) = c_2^{-1}(1-z)^{-1} - \frac{c_2^{-1}(1-z)^{-2}}{(1-z)^{-1} + c_2\Theta(z)}. \tag{C.4}$$

The desired result follows from algebraic transformation of (C.4). □

Lemma 3. *Let G_p be the matrix in (12). Write $G_p = (g_1, \ldots, g_p)$ with the column vectors $g_1, \ldots, g_p$. Define the function $\Delta_p(x) = p^{-1} \sum_{i=1}^{p} \{g_i^\top \Sigma_{p,2}^{-1} g_i \mathbf{1}(f_{p,i} \le x)\}$. Let $\Omega_p(z)$ for $z \in \mathbb{C}^+$ be the Stieltjes transform of Δ_p. Then under Assumptions 1–3, $\Omega_p(z)$ converges almost surely to*

$$\Omega(z) = z^{-1} \left[m_F(z) \{1 - c_2 + 2c_2 z - c_2 z(1-z) m_F(z)\} - 1 + c_2 - \int t^{-1} dD(t) \right],$$

where m_F and D are defined in Table 1.

Proof. We observe that

$$\begin{aligned}
\Omega_p(z) &= \frac{1}{p} \sum_{i=1}^{p} g_i^\top \Sigma_{p,2}^{-1} g_i (f_{p,i} - z)^{-1} = \frac{1}{p} \mathrm{tr}\Big\{\Sigma_{p,2}^{-1} G_p (F_p - zI)^{-1} G_p^\top\Big\} \\
&= \frac{1}{p} \mathrm{tr}\Big\{\Sigma_{p,2}^{-1} W_p G_p^{\top -1} (F_p - zI)^{-1} G_p^{-1} W_p\Big\} = \frac{1}{p} \mathrm{tr}\Big[\Sigma_{p,2}^{-1} W_p \Big\{S_{p,2} - z(S_{p,1} + S_{p,2})\Big\}^{-1} W_p\Big] \\
&= \frac{1}{p} \mathrm{tr}\Big\{W_p \Sigma_{p,2}^{-1} \Big(S_{p,2} W_p^{-1} - zI\Big)^{-1}\Big\}.
\end{aligned}$$

Multiplying $p^{-1} W_p \Sigma_{p,2}^{-1} (S_{p,2} W_p^{-1} - zI)^{-1}$ on both sides of $S_{p,2} W_p^{-1} - zI + zI = \sum_{k=1}^{n_2} q_k q_k^\top W_p^{-1}$, taking trace and using Eq. (2.2) of [3] , we get

$$\begin{aligned}
\frac{1}{p} \mathrm{tr}\Big(W_p \Sigma_{p,2}^{-1}\Big) + z \frac{1}{p} \mathrm{tr}\Big\{W_p \Sigma_{p,2}^{-1} \Big(S_{p,2} W_p^{-1} - zI\Big)^{-1}\Big\} &= \frac{1}{p} \sum_{k=1}^{n_2} q_k^\top \Sigma_{p,2}^{-1} \Big(S_{p,2} W_p^{-1} - zI\Big)^{-1} q_k = \frac{1}{p} \sum_{k=1}^{n_2} q_k^\top \Sigma_{p,2}^{-1} W_p \Big(S_{p,2} - zW_p\Big)^{-1} q_k \\
&= \frac{1}{p} \sum_{k=1}^{n_2} q_k^\top \Sigma_{p,2}^{-1} (W_p^{(k)} + q_k q_k^\top) \Big\{S_{p,2}^{(k)} - zW_p^{(k)} + (1-z) q_k q_k^\top\Big\}^{-1} q_k \\
&= \frac{1}{p} \sum_{k=1}^{n_2} q_k^\top \Sigma_{p,2}^{-1} (W_p^{(k)} + q_k q_k^\top) \left\{ \frac{(S_{p,2}^{(k)} - zW_p^{(k)})^{-1} q_k}{1 + (1-z) q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})^{-1} q_k} \right\}
\end{aligned}$$

and hence

$$\begin{aligned}
&\frac{1}{p} \mathrm{tr}\Big(W_p \Sigma_{p,2}^{-1}\Big) + z \frac{1}{p} \mathrm{tr}\Big\{W_p \Sigma_{p,2}^{-1} \Big(S_{p,2} W_p^{-1} - zI\Big)^{-1}\Big\} \\
&\qquad = \frac{1}{p} \sum_{k=1}^{n_2} \frac{q_k^\top \Sigma_{p,2}^{-1} W_p^{(k)} (S_{p,2}^{(k)} - zW_p^{(k)})^{-1} q_k + q_k^\top \Sigma_{p,2}^{-1} q_k q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})^{-1} q_k}{1 + (1-z) q_k^\top (S_{p,2}^{(k)} - zW_p^{(k)})^{-1} q_k}. \qquad \text{(C.5)}
\end{aligned}$$

Let $\Theta(z)$ be the Stieltjes transform defined in Lemma 2. Analogous to the arguments in the proof of Lemma 2, taking limits on both sides of (C.5), we get

$$\int t dT^{\mathrm{inv}}(t) + 1 + z\Omega(z) = \frac{m_F(z) + c_2 \Theta(z)}{1 + (1-z) c_2 \Theta(z)},$$

where T^{inv} is defined in Table 1.

We note that since $p^{-1} \mathrm{tr} S_{p,1} \Sigma_{p,2}^{-1}$ converges to $\int t dT^{\mathrm{inv}}(t)$ whereas $\mathrm{E} p^{-1} \mathrm{tr} S_{p,1} \Sigma_{p,2}^{-1}$ converges to $\int t^{-1} dD(t)$, which implies $\int t^{-1} dD(t) = \int t dT^{\mathrm{inv}}(t)$. Therefore, plugging in the expression of $\Theta(z)$ from Lemma 2, we get the final result. □

Lemma 4. *Under Assumptions 1–3 and with the notation in Section 2 and Lemma 3, Δ_p converges almost surely vaguely (see p. 85 of [8] for the definition of vague convergence) to a finite measure denoted as Δ. Furthermore,* *defining* $\underline{F}(x) = c_2 F(x) - (c_2 - 1)\mathbf{1}(0 \le x)$ *as the companion distribution of* F*, we have that the distribution function* $\Delta(x)$ *satisfies* $\Delta(x) = \int_{-\infty}^{x} \delta(y) dF(y)$ *for all* $x \in \mathbb{R}$*, where*

$$\delta(y) = \begin{cases} \left\{1 - \check{m}_{\underline{F}}(0) + \frac{c_2 \int t^{-1} dD(t)}{c_2 - 1}\right\} \mathbf{1}(y = 0, c_2 > 1) & \text{if } y \le 0, \\ \dfrac{1 - c_2 + 2c_2 y \{1 - (1 - y)\mathrm{Re}\check{m}_F(y)\}}{y} & \text{if } y > 0. \end{cases}$$

Proof. The first assertion follows from Lemma 3 and Theorem B.9 of [5]. Using Theorem B.10 of [5], the density of the continuous part of Δ can be calculated as follows:

$$\begin{aligned} d\Delta(x) &= \frac{1}{\pi} \mathrm{Im} \lim_{z \in \mathbb{C}^+ \to x} \Omega(z) dx \\ &= \frac{1}{\pi} \mathrm{Im} \lim_{z \in \mathbb{C}^+ \to x} z^{-1} \left[m_F(z) \{1 - c_2 + 2c_2 z - c_2 z(1 - z) m_F(z)\} - 1 + c_2 - \int t^{-1} dD(t) \right] dx \\ &= \frac{1 - c_2 + 2c_2 x \{1 - (1 - x)\mathrm{Re}\check{m}_F(x)\}}{x} \times \frac{1}{\pi} \mathrm{Im}\, \check{m}_F(x) dx \\ &= \frac{1 - c_2 + 2c_2 x \{1 - (1 - x)\mathrm{Re}\check{m}_F(x)\}}{x} dF(x). \end{aligned}$$

We note that the Stieltjes transform of $\underline{F}$ satisfies $m_F(z) = c_2^{-1}(1 - c_2) z^{-1} + c_2^{-1} m_{\underline{F}}(z)$. Therefore, $\Omega(z)$ can be rewritten as

$$\Omega(z) = z^{-1} \left[\left\{c_2^{-1}(1 - c_2) + c_2^{-1} z m_{\underline{F}}(z)\right\} \left\{1 + c_2 - (1 - z) m_{\underline{F}}(z)\right\} - 1 + c_2 - \int t^{-1} dD(t) \right].$$

When $c_2 > 1$, F has $1 - c_2^{-1}$ point mass at 0 and meanwhile $\underline{F}$ is a continuous probability distribution. Applying the Inversion formula of Stieltjes transform (see Theorem B.8 of [5]) and Lemma 9 of [18], we get that

$$\Delta\{0\} = (1 - c_2^{-1})\{1 - \check{m}_{\underline{F}}(0)\} + \int t^{-1} dD(t).$$

From the fact that F has $1 - c_2^{-1}$ point mass at 0, it follows that for x in a neighborhood of 0,

$$\Delta(x) = \int_{-\infty}^{x} \left\{1 - \check{m}_{\underline{F}}(0) + \frac{c_2 \int t^{-1} dD(t)}{c_2 - 1}\right\} dF(y).$$

Thus the proof is complete. □

Proposition 2. *Under Assumptions 1–4 and with the notation of Lemma 4, for any equivariant estimator* $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$ *proposed in Theorem 5, we have as* $p \to \infty$,

$$\begin{aligned} \frac{1}{p} \mathrm{tr}\left(\Sigma_{p,2}^{-1} \hat{\Sigma}_{p,2}\right) - \frac{1}{p} \ln\det\left(\Sigma_{p,2}^{-1} \hat{\Sigma}_{p,2}\right) - 1 \xrightarrow{a.s.} & \int_{\mathrm{Supp}(F) \backslash \{0\}} \left[\frac{1 - c_2 + 2c_2 x \{1 - (1 - x)\mathrm{Re}\, \check{m}_F(x)\}}{x} \phi(x) - \ln\left\{\frac{\phi(x)}{1 - x}\right\} \right] dF(x) \\ & - \int \ln(x) dT^{\mathrm{inv}}(x) - 1 + A, \end{aligned}$$

where $A = 0$ *if* $c_2 \in (0, 1)$,

$$A = \frac{c_2 - 1}{c_2} \left[\left\{1 - \check{m}_{\underline{F}}(0) + \frac{c_2 \int t^{-1} dD(t)}{c_2 - 1}\right\} \phi(0) - \ln\{\phi(0)\} \right]$$

if $c_2 \in (1, \infty)$*, and* T^{inv} *is the distribution defined in Table 1.*

Proof. Recall that $\mathbb{K}$ denotes the compact subinterval of (0, 1) containing $\mathrm{Supp}(F_p)\backslash\{0\}$ and $\mathrm{Supp}(F)\backslash\{0\}$ for all large p. With the notation of Lemmas 3–4, we have

$$\frac{1}{p}\,\mathrm{tr}\left(\Sigma_{p,2}^{-1}\hat{\Sigma}_{p,2}\right) = \frac{1}{p}\,\mathrm{tr}G_p^\top\Sigma_{p,2}^{-1}G_p\Phi_p = \frac{1}{p}\sum_{i=1}^{p} g_i^\top\Sigma_{p,2}^{-1}g_i\phi_p(f_{p,i}) = \int_{\mathbb{K}\cup\{0\}}\phi_p(x)d\Delta_p(x).$$

Since by Assumption 4, $\phi = \lim_{p\to\infty}\phi_p$ is continuous on the compact set $\mathrm{Supp}(F)$, using the result from Lemma 4 that Δ_p converges vaguely to Δ almost surely, we have as $p \to \infty$,

$$\int_{\mathrm{Supp}(F)}\phi(x)d\Delta_p(x) \xrightarrow{a.s.} \int_{\mathrm{Supp}(F)}\phi(x)d\Delta(x). \tag{C.6}$$

We observe that

$$\Delta_p\{\mathbb{R}\} = \frac{1}{p}\,\mathrm{tr}(S_{p,1}\Sigma_2^{-1}) \le \lambda_{\max}(S_{p,1}\Sigma_2^{-1}),$$

where $\lambda_{\max}(S_{p,1}\Sigma_2^{-1})$ stands for the largest eigenvalue of $S_{p,1}\Sigma_2^{-1}$. Assumption 3 and Theorem 1.1 of [4] imply that with probability 1, $\Delta_p\{\mathbb{R}\}$ is uniformly bounded for all large p, namely, there exists $N_1 \in \mathbb{Z}^+$ and $\kappa_1 > 0$ such that, for all $p \ge N_1$, $\Delta_p\{\mathbb{R}\} \le \kappa_1$.

For $\varepsilon > 0$, from the uniform convergence of ϕ_p to ϕ on $\mathrm{Supp}(F)$ as assumed in Assumption 4, it follows that there exists $N_2 \in \mathbb{Z}^+$ such that

$$\forall_{p\ge N_2}\quad \forall_{x\in\mathrm{Supp}(F)}\quad |\phi_p(x) - \phi(x)| \le \frac{\varepsilon}{2\kappa_1}.$$

Then, as $p \ge \max(N_1, N_2)$

$$\left|\int_{\mathrm{Supp}(F)}\phi_p(x)d\Delta_p(x) - \int_{\mathrm{Supp}(F)}\phi(x)d\Delta_p(x)\right| \le \int_{\mathrm{Supp}(F)}\left|\phi_p(x) - \phi(x)\right| d\Delta_p(x) \le \frac{\varepsilon}{2}. \tag{C.7}$$

The boundedness condition of ϕ_p assumed in Assumption 4 implies that almost surely there exists $N_3 \in \mathbb{Z}^+$ such that as $p \ge N_3$, $|\phi_p(x)| \le \kappa$ for all $x \in \mathbb{K}$.

Applying Portmanteau Theorem (see Theorem 2.1 of [6]), we conclude that there exists $N_4 \in \mathbb{Z}^+$ such that

$$\forall_{p\ge N_4}\quad \Delta_p\{\mathbb{K}\cup\{0\}\backslash\mathrm{Supp}(F)\} \le \frac{\varepsilon}{2\kappa}.$$

Hence as $p \ge \max(N_3, N_4)$, we get

$$\left|\int_{\mathbb{K}\cup\{0\}}\phi_p(x)d\Delta_p(x) - \int_{\mathrm{Supp}(F)}\phi_p(x)d\Delta_p(x)\right| = \left|\int_{\mathbb{K}\cup\{0\}\backslash\mathrm{Supp}(F)}\phi_p(x)d\Delta_p(x)\right| \le \frac{\varepsilon}{2}. \tag{C.8}$$

Then, it follows from (C.6), (C.7) and (C.8) that as $p \to \infty$,

$$\int\phi_p(x)d\Delta_p(x) = \int_{\mathbb{K}\cup\{0\}}\phi_p(x)d\Delta_p(x) \xrightarrow{a.s.} \int_{\mathrm{Supp}(F)}\phi(x)d\Delta(x). \tag{C.9}$$

For the second term of the loss function, as $p \to \infty$ we have

$$\begin{aligned}\frac{1}{p}\ln\det\left(\Sigma_{p,2}^{-1}\hat{\Sigma}_{p,2}\right) &= \frac{1}{p}\ln\det\left(\Sigma_{p,2}^{-1}G_p\Phi_pG_p^\top\right) = \frac{1}{p}\ln\det\left\{\Sigma_{p,2}^{-1}G_p(I-F_p)G_p^\top\Phi_p(I-F_p)^{-1}\right\}\\ &= \frac{1}{p}\sum_{i=1}^{p}\ln\left\{\frac{\phi_p(f_{p,i})}{1-f_p}\right\} + \frac{1}{p}\ln\det\left(\Sigma_{p,2}^{-1}S_{p,1}\right) = \int\ln\left\{\frac{\phi_p(x)}{1-x}\right\}dF_p(x) + \int\ln(x)dT_p^{\mathrm{inv}}(x),\end{aligned}$$

where T_p^{inv} is define in Table 1.

Arguments analogous to (C.6)–(C.9) yield that as $p \to \infty$,

$$\int\ln\left\{\frac{\phi_p(x)}{1-x}\right\}dF_p(x) + \int\ln(x)dT_p^{\mathrm{inv}}(x) \xrightarrow{a.s.} \int_{\mathrm{Supp}(F)}\ln\left\{\frac{\phi(x)}{1-x}\right\}dF(x) + \int\ln(x)dT^{\mathrm{inv}}(x).$$

Putting all the limits together, using Lemma 4, we get the desired result. □

Proposition 3. *Under Assumptions 1–4, for any equivariant estimator $(\hat{\Sigma}_{p,1}, \hat{\Sigma}_{p,2})$ proposed in Theorem 5, we have as $p \to \infty$,*

$$\frac{1}{p}\mathrm{tr}\left(\Sigma_{p,1}^{-1}\hat{\Sigma}_{p,1}\right) - \frac{1}{p}\ln\det\left(\Sigma_{p,1}^{-1}\hat{\Sigma}_{p,1}\right) - 1 \overset{a.s.}{\to} \int_{\mathrm{Supp}(F)\backslash\{0\}} \left[\frac{1 - c_1 + 2c_1(1-x)\{1 + x\mathrm{Re}\,\check{m}_F(x)\}}{1-x}\psi(x) - \ln\left\{\frac{\psi(x)}{1-x}\right\}\right] dF(x)$$
$$- \int \ln(x) d\mathscr{M}_{c_1}(x) - 1 + A,$$

where $A = 0$ if $c_2 \in (0,1)$,

$$A = \frac{c_2 - 1}{c_2}\left[\left(1 + \frac{c_1}{c_2}\right)\psi(0) + \ln\{\psi(0)\}\right]$$

if $c_2 \in (1,\infty)$, and $\mathscr{M}_{c_1}(x)$ is the Marčenko–Pastur distribution with parameter c_1 (see 3.1.1 of [5]).

Proof. With the notation of Proposition 2, we define the measure $\bar{\Delta}_p(x) = p^{-1}\sum_{i=1}^{p} g_i^\top \Sigma_{p,1}^{-1} g_p \mathbf{1}(1 - f_{p,i} \le x)$ and the probability distribution functions $\bar{F}(x) = 1 - \lim_{y \to x^+} F(1-x)$, $\underline{\bar{F}}(x) = 1 - \lim_{y\to x^+} \underline{F}(1-x)$. From the same arguments of the proof of Lemmas 2–4 with the roles of $S_{p,1}$ and $S_{p,2}$ interchanged, it follows that as $p \to \infty$, $\bar{\Delta}_p(x)$ converges vaguely to a nonrandom measure $\bar{\Delta}(x)$ whose Stieltjes transform denoted as $\bar{\Omega}(z)$ is given by

$$\bar{\Omega}(z) = z^{-1}\left[m_{\bar{F}}(z)\{1 - c_1 + 2c_1 z - c_1 z(1-z)m_{\bar{F}}(z)\} - 1 + c_1 - \int t dD(t)\right].$$

Analogous to Proposition 2, one can show that $\bar{\Delta}\{1\} = (c_2 - 1)c_2^{-1}(1 + c_1 c_2^{-1}) > 0$. We note that for $a, b \in [0,1]$, $\bar{F}\{[a,b]\} = F\{[1-b, 1-a]\}$ and thus

$$m_{\bar{F}}(z) = \int \frac{1}{x - z} d\bar{F}(x) = \int \frac{1}{1 - x - z} dF(x) = -m_F(1-z).$$

Applying the same arguments of the proof of Proposition 2, it follows that

$$\frac{1}{p}\mathrm{tr}\left(\Sigma_{p,1}^{-1}\hat{\Sigma}_{p,1}\right) - \frac{1}{p}\ln\det\left(\Sigma_{p,1}^{-1}\hat{\Sigma}_{p,1}\right) - 1 \overset{a.s.}{\to}$$
$$\int_{\mathrm{Supp}(\bar{F})\backslash\{1\}} \left[\frac{1 - c_1 + 2c_1 x\{1 - (1-x)\mathrm{Re}\,\check{m}_{\bar{F}}(x)\}}{x}\psi(1-x) - \ln\left\{\frac{\psi(1-x)}{x}\right\}\right] d\bar{F}(x)$$
$$+ \mathbf{1}(c_2 > 1)\frac{c_2 - 1}{c_2}\left[\left(1 + \frac{c_1}{c_2}\right)\psi(0) - \ln\{\psi(0)\}\right] - \int \ln(x) d\mathscr{M}_{c_1}(x) - 1.$$

Changing measures from $\bar{F}$ to F, the final result follows. □

Appendix C.1. Proof of Theorem 5

Suppose $(\hat{\Sigma}_{p,1}(S_{p,1}, S_{p,2}), \hat{\Sigma}_{p,2}(S_{p,1}, S_{p,2}))$ is an equivariant estimator of $(\Sigma_{p,1}, \Sigma_{p,2})$. Then, for each $i \in \{1,2\}$,

$$\hat{\Sigma}_{p,i}(S_{p,1}, S_{p,2}) = A\hat{\Sigma}_{p,i}(A^{-1}S_{p,1}A^{\top -1}, A^{-1}S_{p,2}A^{\top -1})A^\top$$

for all invertible $p \times p$ matrix A. Plugging in the matrix G_p, we get

$$\hat{\Sigma}_{p,i}(S_{p,1}, S_{p,2}) = G_p\hat{\Sigma}_{p,i}(G_p^{-1}S_{p,1}G_p^{\top -1}, G_p^{-1}S_{p,2}G_p^{\top -1})G_p^\top = G_p\hat{\Sigma}_{p,i}(I - F_p, F_p)G^\top$$

for each $i \in \{1,2\}$. Let H be a $p \times p$ diagonal matrix such that its diagonal entries are either 1 or -1. By equivariance,

$$\hat{\Sigma}_{p,i}(I - F_p, F_p) = H\hat{\Sigma}_{p,i}(I - F_p, F_p)H^\top \tag{C.10}$$

for each $i \in \{1,2\}$.

Since (C.10) holds for all such matrix H, this implies $\hat{\Sigma}_{p,i}(I - F_p, F_p)$ is diagonal for each $i \in \{1,2\}$. Writing

$$\Psi_p(F_p) = \hat{\Sigma}_{p,1}(I - F_p, F_p), \quad \Phi_p(F_p) = \hat{\Sigma}_{p,2}(I - F_p, F_p),$$

proves the necessity part of the equivariance. For the sufficiency part, the proof is straightforward and is omitted. Furthermore, since the diagonal matrices $\Psi_p(F_p)$ and $\Phi_p(F_p)$ only depend on F_p, we can just express them using some generic functions ψ_p and ϕ_p depending on F_p such that $\Psi_p(F_p) = \mathrm{diag}\{\psi_p(f_{p,1}), \ldots, \psi_p(f_{p,p})\}$ and $\Phi_p(F_p) = \mathrm{diag}\{\phi_p(f_{p,1}), \ldots, \phi_p(f_{p,p})\}$. □

Appendix C.2. Proof of Theorem 6

Summing up the limiting loss obtained from Propositions 2 and 3, the result follows.

Appendix C.3. Proof of Proposition 1

Letting z converge to $y \in (0, \infty)$, we get from (3) that

$$\forall_{y \in (0,\infty)} \quad y = -\frac{1}{\check{m}_{\underline{L}}(y)} + c_2 \int \frac{dT^{\text{inv}}(t)}{t + \check{m}_{\underline{L}}(y)}. \tag{C.11}$$

Denoting $\check{m}_{\underline{L}}(y) = \underline{m}_1(y) + \mathbf{i}\underline{m}_2(y)$ and taking imaginary parts on both sides of (C.11), we get, for all $y \in (0, \infty)$,

$$0 = \frac{\underline{m}_2(y)}{|\check{m}_{\underline{L}}(y)|^2} - c_2 \int \frac{\underline{m}_2(y) dT^{\text{inv}}(t)}{|t + \check{m}_{\underline{L}}(y)|^2}. \tag{C.12}$$

Then it follows from (C.12) that, for all $y \in \{w : \underline{m}_2(w) > 0\}$,

$$1 = c_2 \int \frac{|\check{m}_{\underline{L}}(y)|^2 dT^{\text{inv}}(t)}{|t + \check{m}_{\underline{L}}(y)|^2}. \tag{C.13}$$

Since $\check{m}_{\underline{L}}(y)$ is continuous on $(0, \infty)$ and $\{w : \underline{m}_2(w) > 0\}$ is dense in $\text{Supp}(L)\backslash\{0\}$ (see [33]), it follows that (C.13) holds for all $y \in \text{Supp}(L)\backslash\{0\}$. Applying the equation $\check{m}_{\underline{L}}(y) = -(1 - c_2)/y + c_2\check{m}_L(y)$, we have from (C.11) that

$$1 - c_2 - 2c_2 y\check{m}_L(y) = 1 - c_2 + 2c_2 \int \frac{t dT^{\text{inv}}(t)}{t + \check{m}_{\underline{L}}(y)} \tag{C.14}$$

for all $y \in (0, \infty)$. Sending $p \to \infty$ and $z \to x \in (0, \infty)$, we get from (6)

$$\check{m}_F(x) = (1 - x)^{-1} + (1 - x)^{-2}\check{m}_L\Big(\frac{x}{1 - x}\Big). \tag{C.15}$$

Denoting $x/(1 - x) = y$, it follows from (C.13), (C.14) and (C.15) that, for all $y \in \{w : \underline{m}_2(w) > 0\}$,

$$1 - c_2 + 2c_2 x\{1 - (1 - x)\text{Re}\,\check{m}_F(x)\} = 1 - c_2 - 2c_2 y\text{Re}\,\check{m}_L(y) = c_2 \int \frac{t^2 dT^{\text{inv}}(t)}{|t + \check{m}_{\underline{L}}(y)|^2} > 0. \tag{C.16}$$

Suppose $y \in (0, \infty)\backslash\text{Supp}(L)$. It it shown in [33] that $\text{Re}\,\check{m}_{\underline{L}}(y) = \check{m}_{\underline{L}}(y)$, $-\check{m}_{\underline{L}}(y) \notin \text{Supp}(T^{\text{inv}})$ and $\partial y/\partial\check{m}_{\underline{L}}(y) > 0$. Calculating the derivative using (C.11), we get

$$\int \Big[\check{m}_{\underline{L}}(y)/\{t + \check{m}_{\underline{L}}(y)\}\Big]^2 dT^{\text{inv}}(t) < c_2^{-1},$$

for $y \in (0, \infty)\backslash\text{Supp}(L)$. By Jensen's inequality,

$$\Bigg\{\int \frac{\check{m}_{\underline{L}}(y)}{t + \check{m}_{\underline{L}}(y)} dT^{\text{inv}}(t)\Bigg\}^2 \leq \int \Bigg\{\frac{\check{m}_{\underline{L}}(y)}{t + \check{m}_{\underline{L}}(y)}\Bigg\}^2 dT^{\text{inv}}(t) < c_2^{-1},$$

which implies that, for all $y \in (0, \infty)\backslash\text{Supp}(L)$,

$$-\sqrt{c_2^{-1}} \leq \int \frac{\check{m}_{\underline{L}}(y)}{t + \check{m}_{\underline{L}}(y)} dT^{\text{inv}}(t) \leq \sqrt{c_2^{-1}}.$$

It then follows from (C.14) that, for all $y \in (0, \infty)\backslash\text{Supp}(L)$,

$$1 - c_2 - 2c_2 y\check{m}_L(y) = 1 + c_2 - 2c_2 \int \frac{\check{m}_{\underline{L}}(y) dT^{\text{inv}}(t)}{t + \check{m}_{\underline{L}}(y)} \geq 1 + c_2 - 2\sqrt{c_2} = (1 - \sqrt{c_2})^2 > 0 \tag{C.17}$$

This shows that $\phi^{\mathrm{or}}(x) > 0$ when $x \in (0,1)$ and $c_2 \in (0,1) \cup (1,\infty)$. The boundedness of $\phi^{\mathrm{or}}(x)$ on $\mathbb{K}$ follows from the continuity of $\phi^{\mathrm{or}}(x)$.

When $c_2 > 1$, one can check that $1 - \check{m}_{\underline{F}}(0) = -\check{m}_{\underline{L}}(0)$ and $\check{m}_{\underline{L}}(0) > 0$. Also we note that $\int t^{-1} dD(t) = \int t T^{\mathrm{inv}}(t)$. It follows from (C.11) that

$$1 - c_2^{-1} = \int \frac{t dT^{\mathrm{inv}}(t)}{t + \check{m}_{\underline{L}}(0)}.$$

Then we see that

$$\phi^{\mathrm{or}}(0) = -\check{m}_{\underline{L}}(0) + \left\{ \int t dT^{\mathrm{inv}}(t) \right\} \Big/ \int \frac{t dT^{\mathrm{inv}}(t)}{t + \check{m}_{\underline{L}}(0)}.$$

Therefore, $\phi^{\mathrm{or}}(0) > 0$ follows from

$$\int \frac{t dT^{\mathrm{inv}}(t)}{t + \check{m}_{\underline{L}}(0)} < \frac{\int t dT^{\mathrm{inv}}(t)}{\check{m}_{\underline{L}}(0)}.$$

Next we show the result for ψ^{or}. First we observe that the denominator of $\psi^{\mathrm{or}}(x)$ equals $1 + c_1 + 2c_1 x(1-x)^{-1} \mathrm{Re}\, \check{m}_L\{x/(1-x)\}$. Denote $y = x/(1-x)$.

If $c_2 < 1$, by changing the distribution from L to L^{inv}, the denominator of ψ^{or} can be rewritten as $1 - c_1 - 2c_1 y^{-1} \mathrm{Re}\, \check{m}_{L^{\mathrm{inv}}}(y^{-1})$, where L^{inv} is the LSD of $S_{p,1} S_{p,2}^{-1}$ and it satisfies that $L^{\mathrm{inv}}\{[a,b]\} = L\{[1/b, 1/a]\}$ for all $b \geq a > 0$. By symmetry, the positiveness and boundedness of ψ^{or} can be shown using arguments similar to those for ϕ^{or}.

When $c_2 > 1$, more caution is needed as L has point mass at 0 and thus L^{inv} is not well defined. In this case, we proceed by perturbing $S_{p,2}$ by a small amount. Define for a small number $\varepsilon > 0$, $L_{p,\varepsilon}$ and $R_{p,\varepsilon}$ as the spectral distributions of $(S_{p,2} + \varepsilon\Sigma_{p,1}) S_{p,1}^{-1}$ and $\Sigma_{p,1}^{-1/2}(S_{p,2} + \varepsilon\Sigma_{p,1})\Sigma_{p,1}^{-1/2}$ with Stieltjes transforms $m_{L_{p,\varepsilon}}(z)$ and $m_{R_{p,\varepsilon}}(z)$, respectively.

We observe that under Assumptions 1–3, the weak limit of $R_{p,\varepsilon}$ exists as $p \to \infty$. Denote the weak limit of $R_{p,\varepsilon}$ as R_ε with Stieltjes transform $m_{R_\varepsilon}(z)$. With the notation in Table 1, it satisfies with probability 1 that, for all $z \in \mathbb{C}^+$,

$$m_{R_\varepsilon}(z) = \lim_{p\to\infty} \frac{1}{p} \mathrm{tr} \left\{ S_{p,2}\Sigma_{p,1}^{-1} - (z-\varepsilon) I \right\}^{-1} = m_R(z - \varepsilon).$$

From the result of [32], the SD of the matrix $S_{p,1}(S_{p,2} + \varepsilon\Sigma_{p,1})^{-1}$ also converges weakly to some nonrandom distribution almost surely as $p \to \infty$. Since each eigenvalue of $(S_{p,2} + \varepsilon\Sigma_{p,1})S_{p,1}^{-1}$ is the reciprocal of that of $S_{p,1}(S_{p,2} + \varepsilon\Sigma_{p,1})^{-1}$, it follows that $L_{p,\varepsilon}$ which is the SD of $(S_{p,2} + \varepsilon\Sigma_{p,1})S_{p,1}^{-1}$ converges weakly almost surely.

Moreover, denoting $\check{S}_{p,2} = S_{p,1}^{-1/2} S_{p,2} S_{p,1}^{-1/2}$ and $\check{\Sigma}_{p,1} = S_{p,1}^{-1/2}\Sigma_{p,1} S_{p,1}^{-1/2}$, we obtain from the inequality that $\mathrm{tr}(AB) \leq p \times \|A\| \times \|B\|$ for any $p \times p$ matrices A, B and the inequality $\|(A - zI)^{-1}\| \leq 1/\mathrm{Im}\, z$ for any real symmetric matrix A and $z \in \mathbb{C}^+$ that, for all $p \in \mathbb{Z}^+$ and $z \in \mathbb{C}^+$,

$$\begin{aligned}
\left| m_{L_{p,\varepsilon}}(z) - m_{L_p}(z) \right| &= \frac{1}{p} \left| \mathrm{tr}\{ (S_{p,2}S_{p,1}^{-1} + \varepsilon\Sigma_{p,1}S_{p,1}^{-1} - zI)^{-1} - (S_{p,2}S_{p,1}^{-1} - zI)^{-1} \} \right| \\
&= \frac{\varepsilon}{p} \left| \mathrm{tr}\{ (S_{p,2}S_{p,1}^{-1} - zI)^{-1} \Sigma_{p,1} S_{p,1}^{-1} (S_{p,2}S_{p,1}^{-1} + \varepsilon\Sigma_{p,1}S_{p,1}^{-1} - zI)^{-1} \} \right| \\
&= \frac{\varepsilon}{p} \left| \mathrm{tr}\{ (\check{S}_{p,2} - zI)^{-1} \check{\Sigma}_{p,1} (\check{S}_{p,2} + \varepsilon\check{\Sigma}_{p,1} - zI)^{-1} \} \right| \\
&\leq \varepsilon\, (\mathrm{Im}\, z)^{-2} \|\check{\Sigma}_{p,1}\|.
\end{aligned} \tag{C.18}$$

Let L_ε denote the weak limit of $L_{p,\varepsilon}$ as $p \to \infty$ whose Stieltjes transform is denoted as $m_{L_\varepsilon}(z)$. Eq. (C.18) implies that, for all $z \in \mathbb{C}^+$,

$$\lim_{\varepsilon \to 0^+} m_{L_\varepsilon}(z) = m_L(z).$$

We define $L_\varepsilon^{\mathrm{inv}}$ to be the probability distribution satisfying $L_\varepsilon^{\mathrm{inv}}\{[a,b]\} = L_\varepsilon\{[1/b, 1/a]\}$ for any $b \geq a > 0$. We note that $L_\varepsilon^{\mathrm{inv}}$ is also the LSD of $S_{p,1}(S_{p,2} + \varepsilon\Sigma_{p,1})^{-1}$ as $p \to \infty$.

Denote, for $x > 0$, $y = x/(1-x)$. Let $\{z_1, z_2, \dots\} \subset \mathbb{C}^+$ and $\{\varepsilon_1, \varepsilon_2, \dots\} \subset (0, \infty)$ be sequences converging in appropriate rates to y^{-1} and 0 respectively such that

$$\lim_{k\to\infty} m_{L_{\varepsilon_k}}(z_k^{-1}) = \overline{\check{m}}_L(y),$$

where we get the complex conjugate $\overline{\check{m}}_L(y)$ instead of $\check{m}_L(y)$ because z_k^{-1} converges from the lower half complex plane to y.

Then it follows that

$$1 + c_1 + 2c_1 y \mathrm{Re}\, \check{m}_L(y) = \mathrm{Re} \lim_{k\to\infty}\{1 + c_1 + 2c_1 z_k^{-1} m_{L_{\varepsilon_k}}(z_k^{-1})\} = \mathrm{Re} \lim_{k\to\infty}\{1 - c_1 - 2c_1 z_k m_{L^{\mathrm{inv}}_{\varepsilon_k}}(z_k)\}.$$

Define the companion Stieltjes transform $m_{\underline{L}^{\mathrm{inv}}_{\varepsilon}}(z) = -(1 - c_1)/z + c_1 m_{L^{\mathrm{inv}}_{\varepsilon}}(z)$. Similar to (3), it holds that for all $k \in \{1, 2, \dots\}$,

$$z_k = -\frac{1}{m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)} + c_1 \int \frac{dR_{\varepsilon_k}(t)}{t + m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)}. \tag{C.19}$$

Taking imaginary parts on both sides of (C.19), we get that

$$\frac{\mathrm{Im}\, z_k}{\mathrm{Im}\, m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)} |m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)|^2 + c_1 \int \frac{|m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)|^2 dR_{\varepsilon_k}(t)}{|t + m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)|^2} = 1. \tag{C.20}$$

Using the equation $m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k) = -(1 - c_1)/z_k + c_1 m_{L^{\mathrm{inv}}_{\varepsilon_k}}(z_k)$, we get again from (C.19) that

$$1 - c_1 - 2c_1 z_k m_{L^{\mathrm{inv}}_{\varepsilon_k}}(z_k) = 1 - c_1 + 2c_1 \int \frac{t dR_{\varepsilon_k}(t)}{t + m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)}. \tag{C.21}$$

It then follows from (C.20) and (C.21) that

$$\mathrm{Re}\{1 - c_1 - 2c_1 z_k m_{L^{\mathrm{inv}}_{\varepsilon_k}}(z_k)\} = \frac{\mathrm{Im}\, z_k}{\mathrm{Im}\, m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)} |m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)|^2 + c_1 \int \frac{t^2 dR_{\varepsilon_k}(t)}{|t + m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)|^2} \geq c_1 \int \frac{t^2 dR_{\varepsilon_k}(t)}{|t + m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)|^2}.$$

Therefore, the denominator of $\psi^{\mathrm{or}}(x)$ satisfies that

$$1 - c_1 + 2c_1(1 - x)\{1 + x\mathrm{Re}\, \check{m}_F(x)\} = 1 + c_1 + 2c_1 y \mathrm{Re}\, \check{m}_L(y) \geq \liminf_{k\to\infty} c_1 \int \frac{t^2 dR_{\varepsilon_k}(t)}{|t + m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k)|^2} > 0,$$

where the last inequality follows from the fact that the mass of R_{ε_k} does not all escape to 0 as $k \to \infty$ and $m_{\underline{L}^{\mathrm{inv}}_{\varepsilon_k}}(z_k) = -(1 - c_1)/z_k + c_1 m_{L^{\mathrm{inv}}_{\varepsilon_k}}(z_k) = -z_k^{-1} - c_1 z_k^{-2} m_{L_{\varepsilon_k}}(z_k^{-1})$ is uniformly bounded for all k.

The proof is complete. □

Appendix C.4. Proof of Theorem 8

Recall $\mathbb{K}$ is the compact interval containing $\mathrm{Supp}(F_p)\backslash\{0\}$ and $\mathrm{Supp}(F)\backslash\{0\}$ for all large p. According to Assumption 4 and Theorem 6, we only need to show that almost surely the two functions $\psi_p^{BF}(x)$ and $\phi_p^{BF}(x)$ are uniformly bounded for all large p on $\mathbb{K}$ and uniformly converge to $\psi^{\mathrm{or}}(x)$ and $\phi^{\mathrm{or}}(x)$ (see (14) and (15)) on $\mathrm{Supp}(F)$ as $p \to \infty$. We show in the following a stronger result that the uniform convergence is true on any compact subinterval of $(0, 1)$. Thus the conditions for $\psi_p^{BF}(x)$ and $\phi_p^{BF}(x)$ are naturally satisfied provided that $\psi_p^{BF}(0)$ and $\phi_p^{BF}(0)$ also converge to $\psi^{\mathrm{or}}(0)$ and $\phi^{\mathrm{or}}(0)$ respectively when $c_2 > 1$. Observe from (16) and (17) that on any compact subsets of $(0, 1)$, to show the uniform convergence of $\psi_p^{BF}(x)$ and $\phi_p^{BF}(x)$ to $\psi^{\mathrm{or}}(x)$ and $\phi^{\mathrm{or}}(x)$, it suffices to show the uniform convergence of $\check{m}_{F_{\hat{D}_p, c_{1,p}, c_{2,p}}}(x)$ to $\check{m}_F(x)$.

Let $\underline{\hat{T}}_p^{\mathrm{inv}}$ be the probability distribution having Stieltjes transform $m_{\underline{\hat{T}}_p^{\mathrm{inv}}}(z)$ such that for $z \in \mathbb{C}^+$, $m_{\underline{\hat{T}}_p^{\mathrm{inv}}}(z)$ is the unique value in $\mathbb{C}^+$ satisfying

$$z = -\{m_{\underline{\hat{T}}_p^{\mathrm{inv}}}(z)\}^{-1} + c_{1,p} \int_0^\infty \{t + m_{\underline{\hat{T}}_p^{\mathrm{inv}}}(z)\}^{-1} d\hat{D}_p(t).$$

Let $\hat{T}_p^{\mathrm{inv}}$ be the probability distribution function such that $\underline{\hat{T}}_p^{\mathrm{inv}}(t) = (1 - c_{1,p})\mathbf{1}(0 \leq t) + c_{1,p}\hat{T}_p^{\mathrm{inv}}(t)$ for $t \in \mathbb{R}$ and $\hat{T}_p$ be the probability distribution such that $\hat{T}_p\{[a, b]\} = \hat{T}_p^{\mathrm{inv}}\{[1/b, 1/a]\}$ for any $b \geq a > 0$.

With the notation in Table 1, it follows from Lemma A.4 in [37] that $\hat{\underline{T}}_p^{\text{inv}}$ almost surely converges weakly to $\underline{T}^{\text{inv}}$, and thus $\hat{T}_p^{\text{inv}}$ and $\hat{T}_p$ converge weakly respectively to T^{inv} and T almost surely. By Lemma A.6 in [37], we get that almost surely $\text{Supp}(\hat{T}_p)$ are uniformly bounded away from 0 and ∞ for all large p.

Let $\underline{L}_{\hat{D}_p,c_{1,p},c_{2,p}}$ be the distribution defined in Remark 1 with $\hat{D}_p, c_{1,p}, c_{2,p}$ in place of G, k_1, k_2. It follows from [32] that $m_{\underline{L}_{\hat{D}_p,c_{1,p},c_{2,p}}}(z)$ is the unique value in $\mathbb{C}^+$ satisfying, for all $z \in \mathbb{C}^+$,

$$z = -\frac{1}{m_{\underline{L}_{\hat{D}_p,c_{1,p},c_{2,p}}}(z)} + c_{2,p}\int_0^\infty \frac{td\hat{T}_p(t)}{1 + tm_{\underline{L}_{\hat{D}_p,c_{1,p},c_{2,p}}}(z)}.$$

Let $\check{m}_{\underline{L}_{\hat{D}_p,c_{1,p},c_{2,p}}}(x) = \lim_{z\in\mathbb{C}^+\to x} m_{\underline{L}_{\hat{D}_p,c_{1,p},c_{2,p}}}(z)$ which exists for all $x \in \mathbb{R}\backslash\{0\}$. Applying Theorem A.2 in [37], we have, as $p \to \infty$,

$$\check{m}_{\underline{L}_{\hat{D}_p,c_{1,p},c_{2,p}}}(x) \to \check{m}_{\underline{L}_D,c_1,c_2}(x)$$

uniformly on any compact subset of $(0, \infty)$ almost surely. Letting $z \in \mathbb{C}^+ \to x$, it follows from (7) that

$$\check{m}_{F_{\hat{D}_p,c_{1,p},c_{2,p}}}(x) = \frac{c_{2,p}x + 1 - c_{2,p}}{c_{2,p}x(1-x)} + \frac{\check{m}_{\underline{L}_{\hat{D}_p,c_{1,p},c_{2,p}}}\{x/(1-x)\}}{c_{2,p}(1-x)^2}.$$

Therefore, the uniform convergence of $\check{m}_{F_{\hat{D}_p,c_{1,p},c_{2,p}}}(x)$ on compact subsets of $(0, 1)$ follows from the uniform convergence of $\check{m}_{\underline{L}_{\hat{D}_p,c_{1,p},c_{2,p}}}(x)$ on compact subsets of $(0, \infty)$.

When $c_2 > 1$, the convergence of $\psi_p^{BF}(x)$ and $\phi_p^{BF}(x)$ at $x = 0$ follows from the weak convergence of $\underline{F}_{\hat{D}_p,c_{1,p},c_{2,p}}$ to $\underline{F}$ and the almost sure convergence of $p^{-1}\sum_{i=1}^p \hat{d}_{p,i}^{-1}$ to $\int t^{-1}dD(t)$. □

Appendix C.5. Proof of Corollary 4

We only have to show that as $p \to \infty$, the difference of the loss function vanishes asymptotically, i.e.,

$$L(\hat{\Sigma}_{p,1}^{RM}, \hat{\Sigma}_{p,2}^{RM}; \Sigma_{p,1}, \Sigma_{p,2}) - L(\hat{\Sigma}_{p,1}^{BF}, \hat{\Sigma}_{p,2}^{BF}; \Sigma_{p,1}, \Sigma_{p,2}) \overset{a.s.}{\to} 0.$$

Observe that

$$\begin{aligned}
&L(\hat{\Sigma}_{p,1}^{RM}, \hat{\Sigma}_{p,2}^{RM}; \Sigma_{p,1}, \Sigma_{p,2}) - L(\hat{\Sigma}_{p,1}^{BF}, \hat{\Sigma}_{p,2}^{BF}; \Sigma_{p,1}, \Sigma_{p,2}) \\
&\quad = \frac{1}{p}\sum_{i=1}^{2}\left[\left\{\text{tr}(\Sigma_{p,i}^{-1}\hat{\Sigma}_{p,i}^{RM}) - \ln|\Sigma_{p,i}^{-1}\hat{\Sigma}_{p,i}^{RM}| - p\right\} - \left\{\text{tr}(\Sigma_{p,i}^{-1}\hat{\Sigma}_{p,i}^{BF}) - \ln|\Sigma_{p,i}^{-1}\hat{\Sigma}_{p,i}^{BF}| - p\right\}\right] \\
&\quad = \frac{1}{p}\sum_{i=1}^{2}\left[\text{tr}\left\{\Sigma_{p,i}^{-1}(\hat{\Sigma}_{p,i}^{RM} - \hat{\Sigma}_{p,i}^{BF})\right\} - \ln|\hat{\Sigma}_{p,i}^{RM}| + \ln|\hat{\Sigma}_{p,i}^{BF}|\right].
\end{aligned}$$

We show that $p^{-1}\text{tr}\{\Sigma_{p,i}^{-1}(\hat{\Sigma}_{p,i}^{RM} - \hat{\Sigma}_{p,i}^{BF})\} \overset{a.s.}{\to} 0$ and $p^{-1}(\ln|\hat{\Sigma}_{p,i}^{RM}| - \ln|\hat{\Sigma}_{p,i}^{BF}|) \overset{a.s.}{\to} 0$ for $i = 2$. When $i = 1$, the result follows similarly.

With the notation of Lemma 3, we obtain from the diagonalization (12) that

$$W_p^{-1/2}S_{p,2}W_p^{-1/2} = W_p^{-1/2}G_pF_pG_p^\top W_p^{1/2},$$

which implies there exists an orthogonal matrix E_p such that $W_p^{-1/2}G_p = E_p$. Writing $E_p = (e_1, \ldots, e_p)$ such that $e_1, \ldots, e_p$ are the orthonormal columns of E_p, we get that the ith column of G_p, $g_i = W_p^{1/2}e_i$. Thus under Assumptions 1–3, it follows that with probability 1, there exists a constant $C_1 > 0$ such that for all large p

$$\begin{aligned}
\frac{1}{p}\text{tr}\left\{\Sigma_{p,2}^{-1}(\hat{\Sigma}_{p,2}^{RM} - \hat{\Sigma}_{p,2}^{BF})\right\} &= \frac{1}{p}\text{tr}\left\{G_p^\top\Sigma_{p,2}^{-1}G_p(\Phi_p^{RM} - \Phi_p^{BF})\right\} = \frac{1}{p}\sum_{i=1}^{p} g_i^\top\Sigma_{p,2}^{-1}g_i(\phi_{p,i}^{RM} - \phi_{p,i}^{BF}) \\
&\le \frac{1}{p}\sum_{i=1}^{p}\left|e_i^\top W_p^{1/2}\Sigma_{p,2}^{-1}W_p^{1/2}e_i\right|\left|\phi_{p,i}^{RM} - \phi_{p,i}^{BF}\right| \le C_1\frac{1}{p}\sum_{i=1}^{p}\left|\phi_{p,i}^{RM} - \phi_{p,i}^{BF}\right|.
\end{aligned}$$

We also note that $\phi_{p,i}^{BF} = p\int_{(i-1)/p}^{i/p} \phi_p^{BF}\{F_p^{-1}(x)\}dx$, for all $i \in \{1,\ldots,p\}$. Therefore,

$$\frac{1}{p}\sum_{i=1}^{p}\left|\phi_{p,i}^{RM}-\phi_{p,i}^{BF}\right| = \frac{1}{p}\sum_{i=1}^{p}\left|p\int_{(i-1)/p}^{i/p}\left[\phi_p^{BF}\{F_{\hat{D}_p,c_{1,p},c_{2,p}}^{-1}(x)\}-\phi_p^{BF}\{F_p^{-1}(x)\}\right]dx\right|$$
$$\leq \int_0^1\left|\phi_p^{BF}\{F_{\hat{D}_p,c_{1,p},c_{2,p}}^{-1}(x)\}-\phi_p^{BF}\{F_p^{-1}(x)\}\right|dx. \tag{C.22}$$

By the proof of Theorem 3.2.2 in [12], $F_{\hat{D}_p,c_{1,p},c_{2,p}}^{-1}(x) - F_p^{-1}(x)$ converges almost surely to 0 for all but a countable number of x in $[0,1]$ as $p \to \infty$.

According to the almost sure uniform convergence of the function $\phi_p^{BF}(x)$ to $\phi^{\text{or}}(x)$ on any compact subset of $(0,1)$ (see the proof of Theorem 8), the almost sure uniform boundedness of $F_{\hat{D}_p,c_{1,p},c_{2,p}}^{-1}(x)$ and $F_p^{-1}(x)$ for all $x \in [0,1]$ and all large p (from Assumption 3, Lemma A.6 in [37] and Lemma 2.1 in Part I of [37]) and Lebesgue's Dominated Convergence Theorem, we get (C.22) converges almost surely to 0, which implies that, as $p \to \infty$,

$$\frac{1}{p}\text{tr}\left\{\Sigma_{p,2}^{-1}(\hat{\Sigma}_{p,2}^{RM}-\hat{\Sigma}_{p,2}^{BF})\right\} \xrightarrow{a.s.} 0.$$

For $p^{-1}(\ln|\hat{\Sigma}_{p,2}^{RM}| - \ln|\hat{\Sigma}_{p,2}^{BF}|) \xrightarrow{a.s.} 0$, using the inequality $\ln(x) \leq x-1$ for $x>0$, we get

$$\frac{1}{p}\left(\ln|\hat{\Sigma}_{p,2}^{RM}|-\ln|\hat{\Sigma}_{p,2}^{BF}|\right) = \frac{1}{p}\sum_{i=1}^{p}\ln\left(\phi_{p,i}^{RM}/\phi_{p,i}^{BF}\right) \leq \frac{1}{p}\sum_{i=1}^{p}\left|\left(\phi_{p,i}^{RM}-\phi_{p,i}^{BF}\right)/\phi_{p,i}^{BF}\right|. \tag{C.23}$$

Again using almost sure uniform convergence of the function $\phi_p^{BF}(x)$ to $\phi^{\text{or}}(x)$ on any compact subset of $(0,1)$ and Proposition. 1, we can conclude that $|\phi_{p,i}^{BF}|$ are almost surely uniformly bounded away from 0 and ∞ for all large p and all $i \in \{1,\ldots,p\}$. Therefore, continuing with (C.23) and using the convergence result of (C.22), we have almost surely as $p \to \infty$,

$$\frac{1}{p}\left(\ln|\hat{\Sigma}_{p,2}^{RM}|-\ln|\hat{\Sigma}_{p,2}^{BF}|\right) \leq C_2\frac{1}{p}\sum_{i=1}^{p}\left|\phi_{p,i}^{RM}-\phi_{p,i}^{BF}\right| \xrightarrow{a.s.} 0,$$

where $C_2>0$ is a constant independent with p, n_1, n_2.

Reference

[1] Z.D. Bai, Methodologies in spectral analysis of large dimensional random matrices: A review, Statist. Sinica 9 (1999) 611–677.
[2] Z.D. Bai, J.Q. Chen, J.F. Yao, On estimation of the population spectral distribution from a high-dimensional sample covariance matrix, Aust. New Zealand J. Statist. 52 (2010) 423–437.
[3] Z.D. Bai, J.W. Silverstein, On the empirical distribution of eigenvalues of a class of large dimensional random matrices, J. Multivariate Anal. 54 (1995) 175–192.
[4] Z.D. Bai, J.W. Silverstein, No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices, Ann. Probab. 26 (1998) 316–345.
[5] Z.D. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, 2nd ed., Springer, New York, 2010.
[6] P. Billingsley, Convergence of Probability Measures, 2nd ed., Wiley, New York, 1999.
[7] M. Bilodeau, M.S. Srivastava, Estimation of the eigenvalues of $\Sigma_1\Sigma_2^{-1}$, J. Multivariate Anal. 41 (1992) 1–13.
[8] K.L. Chung, A Course in Probability Theory, Academic Press, 3rd ed., 2001.
[9] M.J. Daniels, R.E. Kass, Shrinkage estimators for covariance matrices, Biometrics 57 (2001) 1173–1184.
[10] D.K. Dey, Simultaneous estimation of eigenvalues, Ann. Inst. Statist. Math. 40 (1988) 137–147.
[11] D.K. Dey, C. Srivinasan, Estimation of a covariance matrix under Stein's loss, Ann. Statist. 13 (1985) 1581–1591.
[12] R. Durrett, Probability: Theory and Examples, 4th ed., Cambridge Univ. Press, Cambridge, 2010.
[13] N. El Karoui, Spectrum estimation for large dimensional covariance matrices using random matrix theory, Ann. Statist. 36 (2008) 2757–2790.
[14] K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd ed., Academic Press, New York, 1990.
[15] L.R. Haff, The variational form of certain Bayesian estimators, Ann. Statist. 19 (1991) 1163–1190.
[16] Y. Konno, A note on estimating eigenvalues of scale matrix of the multivariate F-distribution, Ann. Inst. Statist. Math. 43 (1991) 157–165.
[17] Y. Konno, On estimating eigenvalues of the scale matrix of the multivariate F-distribution, Indian Statist. Inst. 54 (1992) 241–251.
[18] O. Ledoit, S. Péché, Eigenvectors of some large sample covariance matrix ensembles, Probab. Theory Relat. Fields 151 (2011) 233–264.
[19] O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivariate Anal. 88 (2004) 365–411.

[20] O. Ledoit, M. Wolf, Nonlinear shrinkage estimation of large-dimensional covariance matrices, Ann. Statist. 40 (2012) 1024–1060.
[21] O. Ledoit, M. Wolf, Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions, J. Multivariate Anal. 139 (2015) 360–384.
[22] O. Ledoit, M. Wolf, Optimal estimation of a large-dimensional covariance matrix under Stein's loss, Bernoulli 24 (2018) 3791–3821.
[23] P.L. Leung, Estimation of eigenvalues of the scale matrix of the multivariate F-distribution, Commun. Statist. Theory Methods 21 (1992) 1845–1856.
[24] P.L. Leung, W.Y. Chan, Estimation of the scale matrix and its eigenvalues in the Wishart and the multivariate F-Distributions, Ann. Inst. Statist. Math. 50 (1998) 523–530.
[25] W.M. Li, J.Q. Chen, Y.L. Qin, J.F. Yao, Z.D. Bai, Estimation of the population spectral distribution from a large dimensional sample covariance matrix, J. Statist. Plann. Inf. 143 (2013) 1887–1897.
[26] W.M. Li, J.F. Yao, On generalized expectation-based estimation of a population spectral distribution from high-dimensional data, Ann. Inst. Statist. Math. 67 (2015) 359–373.
[27] W.L. Loh, Estimating Covariance Matrices, Ph.D. thesis, Stanford University, Stanford, CA, 1988.
[28] W.L. Loh, Estimating covariance matrices, Ann. Statist. 19 (1991) 283–296.
[29] V.A. Marcĕnko, L.A. Pastur, Distribution of eigenvalues for some sets of random matrices, Sbornik: Math. 1 (1967) 457–483.
[30] D. Paul, A. Aue, Random matrix theory in statistics: A review, J. Statist. Plann. Inf. 150 (2014) 1–29.
[31] N.R. Rao, J.A. Mingo, R. Speicher, A. Edelman, Statistical eigen-inference from large Wishart matrices, Ann. Statist. 36 (2008) 2850–2885.
[32] J.W. Silverstein, Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices, J. Multivariate Anal. 55 (1995) 331–339.
[33] J.W. Silverstein, S.I. Choi, Analysis of the limiting spectral distribution of large dimensional random matrices, J. Multivariate Anal. 54 (1995) 295–309.
[34] C. Stein, Rietz lecture, 39th annual meeting IMS, Lectures, IMS, Atlanta, GA, 1975.
[35] C. Stein, Lectures on the theory of estimation of many parameters, J. Math. Sci. 34 (1986) 1373–1403.
[36] H. Tsukuma, Y. Konno, Simultanuous estimation of scale matrices in two-sample problem under elliptically contoured distributions, J. Japanese Soc. Comput. Statist. 16 (2003) 1–22.
[37] J. Wen, On High-Dimensional Covariance Matrices Estimation, Ph.D. thesis, Dept. Statist. & Appl. Probab., National University of Singapore, `http://scholarbank.nus.edu.sg/handle/10635/135450`, 2017.
[38] R. Yang, J. Berger, Estimation of a covariance matrix using the reference prior, Ann. Statist. 22 (1994) 1195–1211.
[39] J.F. Yao, Z.D. Bai, S.R. Zheng, Large Sample Covariance Matrices and High-Dimensional Data Analysis, Cambridge Univ. Press, Cambridge, 2015.
[40] S.R. Zheng, Z.D. Bai, J.F. Yao, CLT for large dimensional general Fisher matrices and its applications in high-dimensional data analysis, Bernoulli 23 (2017) 1130–1178.