High-dimensional inference on covariance structures via the extended cross-data-matrix methodology

Kazuyoshi Yata[a], Makoto Aoshima[a,1,]

[a] *Institute of Mathematics, University of Tsukuba, Ibaraki 305–8571, Japan*

Abstract

Tests of the correlation matrix between two subsets of a high-dimensional random vector are considered. The test statistic is based on the extended cross-data-matrix methodology (ECDM) and shown to be unbiased. The ECDM estimator is also proved to be consistent and asymptotically Normal in high-dimensional settings. The authors propose a test procedure based on the ECDM estimator and evaluate its size and power, both theoretically and numerically. They give several applications of the ECDM estimator and illustrate the performance of the test procedure using microarray data.

Keywords: Correlations test; Graphical modeling; Large p, small n; Partial correlation; Pathway analysis; RV-coefficient.

1. Introduction

Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be a random sample of size $n \geq 4$ from a p-variate distribution. We are interested here in situations where the data dimension, p, is very high compared to the sample size n.

For each $j \in \{1, \dots, n\}$, write $\boldsymbol{x}_j = (\boldsymbol{x}_{1j}^\top, \boldsymbol{x}_{2j}^\top)^\top$, where for $i \in \{1, 2\}$, $\boldsymbol{x}_{ij} \in \mathbb{R}^{p_i}$ with $p_1 \in \{1, \dots, p-1\}$ and $p_2 = p - p_1$. Assume that $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ have

Email address: aoshima@math.tsukuba.ac.jp (Makoto Aoshima)
[1] Phone & Fax: +81–298–53–6501

unknown mean vector, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$, and unknown covariance matrix,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_* \\ \boldsymbol{\Sigma}_*^\top & \boldsymbol{\Sigma}_2 \end{pmatrix} \geq \mathbf{0}.$$

In other words, for all $j \in \{1, \ldots, n\}$ and $i \in \{1, 2\}$,

$$\mathrm{E}(\boldsymbol{x}_{ij}) = \boldsymbol{\mu}_i, \quad \mathrm{var}(\boldsymbol{x}_{ij}) = \boldsymbol{\Sigma}_i, \quad \mathrm{cov}(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j}) = \mathrm{E}(\boldsymbol{x}_{1j}\boldsymbol{x}_{2j}^\top) - \boldsymbol{\mu}_1\boldsymbol{\mu}_2^\top = \boldsymbol{\Sigma}_*.$$

For all $i \in \{1, 2\}$ and $k \in \{1, \ldots, p_i\}$, the kth diagonal element σ_{ik} of $\boldsymbol{\Sigma}_i$ is assumed to be strictly positive. Then, for all $j \in \{1, \ldots, n\}$,

$$\mathrm{corr}(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j}) = \boldsymbol{P} = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{1p_1})^{-1/2}\boldsymbol{\Sigma}_*\mathrm{diag}(\sigma_{21}, \ldots, \sigma_{2p_2})^{-1/2}.$$

In this paper, we consider the problem of testing the hypotheses

$$\mathcal{H}_0 : \boldsymbol{P} = \mathbf{0} \quad \text{vs.} \quad \mathcal{H}_1 : \boldsymbol{P} \neq \mathbf{0} \tag{1}$$

in high-dimensional settings. When $(p_1, p_2) = (p-1, 1)$ or $(1, p-1)$, testing (1) amounts to testing correlation coefficients. Aoshima and Yata [1] proposed a statistic for the latter problem and Yata and Aoshima [19] improved this test statistic by using a method called the *extended cross-data-matrix methodology* (ECDM). However, tests on the correlation matrix are equally important, e.g., in pathway analysis or graphical modeling for high-dimensional data. One possible application pertains to the construction of gene networks, as portrayed in Figure 1.

Here, we consider testing partial correlation coefficients. When $\boldsymbol{\Sigma} > \mathbf{0}$, write

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_* \\ \boldsymbol{\Omega}_*^\top & \boldsymbol{\Omega}_2 \end{pmatrix} = (\omega_{ij}),$$

where, for $i \in \{1, 2\}$, $\boldsymbol{\Omega}_i$ is the corresponding $p_i \times p_i$ matrix. Here, (m_{ij}) denotes a matrix whose (i, j)th element is m_{ij}. When $i \neq j$, $-\omega_{ij}(\omega_{ii}\omega_{jj})^{-1/2}$ is the (i, j)th partial correlation coefficient; see, e.g., Drton and Perlman [5]. We denote the partial correlation coefficient matrix by

$$\boldsymbol{P}_\Omega = -\mathrm{diag}(\omega_{11}, \ldots, \omega_{p_1p_1})^{-1/2}\boldsymbol{\Omega}_*\mathrm{diag}(\omega_{p_1+1p_1+1}, \ldots, \omega_{pp})^{-1/2}$$

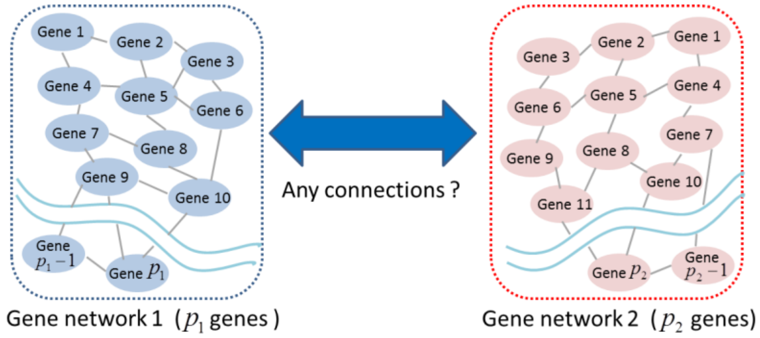


Figure 1: Relevance of hypotheses (1) illustrated in the context of gene networks.

and note that the test of the hypotheses

$$\mathcal{H}_0 : \boldsymbol{P}_\Omega = \boldsymbol{0} \quad \text{vs.} \quad \mathcal{H}_1 : \boldsymbol{P}_\Omega \neq \boldsymbol{0}$$

is equivalent to the test of hypotheses (1) since $\boldsymbol{\Omega}_* = \boldsymbol{0}$ is equivalent to $\boldsymbol{\Sigma}_* = \boldsymbol{0}$.

Drton and Perlman [5] and Wille et al. [16] considered pathway analysis or graphical modeling of microarray data by testing an individual partial correlation coefficient. For example, Wille et al. [16] analyzed gene networks of microarray data with $p = 834$ ($p_1 = 39$ and $p_2 = 795$) and $n = 118$. In contrast, Hero and Rajaratnam [8] considered correlation screening procedures for high-dimensional data by testing correlations. Lan et al. [10] and Zhong and Chen [20] considered tests of regression coefficient vectors in linear regression models. As for tests of independence, see, among others, Fujikoshi et al. [7], Hyodo et al. [9], Srivastava and Reid [13], and Yang and Pan [17]. Also, one may refer to Székely and Rizzo [14, 15] for distance correlation.

In Section 2, we set the notation and state several assumptions required for the construction of our high-dimensional correlation test of hypotheses (1). In Section 3, we produce a test statistic for this problem by using the ECDM methodology and show the unbiasedness of the ECDM estimator. We also show that the ECDM estimator is consistent and asymptotically Normal when $p \to \infty$ and $n \to \infty$. In Section 4, we propose a test procedure for (1) by the ECDM estimator and evaluate its asymptotic size and power when $p \to \infty$ and $n \to \infty$

theoretically and numerically. In Section 5, we give several applications of the ECDM estimator. Finally, we demonstrate how the test procedure performs in practice using microarray data.

2. Assumptions

In this section, we lay out the basic assumptions for the construction of our test of hypotheses (1). The eigenvalue decomposition of $\boldsymbol{\Sigma}$ is denoted by $\boldsymbol{\Sigma} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top$, where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_p)$ and $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$ are the eigenvalues of $\boldsymbol{\Sigma}$, while $\boldsymbol{H}$ is an orthogonal matrix of the corresponding eigenvectors.

For all $j \in \{1, \ldots, n\}$, let $\boldsymbol{x}_j = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}\boldsymbol{z}_j + \boldsymbol{\mu}$, where $\text{E}(\boldsymbol{z}_j) = \boldsymbol{0}$ and $\text{var}(\boldsymbol{z}_j) = \boldsymbol{I}_p$, the identity matrix of dimension p. Note that if $\boldsymbol{x}_j$ is Gaussian, the elements of $\boldsymbol{z}_j$ form a random sample from the standard Normal distribution, $\mathcal{N}(0,1)$. We assume that, for all $j \in \{1, \ldots, n\}$,

$$\boldsymbol{x}_j = \boldsymbol{\Gamma}\boldsymbol{w}_j + \boldsymbol{\mu}, \tag{2}$$

where $\boldsymbol{\Gamma}$ is a $p \times q$ matrix for some $q > 0$ such that $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top = \boldsymbol{\Sigma}$, and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ form a random sample, so that for every $j \in \{1, \ldots, n\}$, $\boldsymbol{w}_j = (w_{1j}, \ldots, w_{qj})^\top$, $\text{E}(\boldsymbol{w}_j) = \boldsymbol{0}$ and $\text{var}(\boldsymbol{w}_j) = \boldsymbol{I}_q$. Let $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1^\top, \boldsymbol{\Gamma}_2^\top)^\top$, where for $i \in \{1, 2\}$, $\boldsymbol{\Gamma}_i = (\boldsymbol{\gamma}_{i1}, \ldots, \boldsymbol{\gamma}_{iq})$ with $\boldsymbol{\gamma}_{ij} \in \mathbb{R}^{p_i}$, so that $\boldsymbol{x}_{ij} = \boldsymbol{\Gamma}_i\boldsymbol{w}_j + \boldsymbol{\mu}_i$. Note that

$$\boldsymbol{\Sigma}_* = \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^\top = \sum_{r=1}^{q} \boldsymbol{\gamma}_{1r}\boldsymbol{\gamma}_{2r}^\top.$$

Also note that Eq. (2) includes the case where $\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$ and $\boldsymbol{w}_j = \boldsymbol{z}_j$. For all $r \in \{1, \ldots, q\}$, let $\text{var}(w_{rj}^2) = M_r$ and assume that $\limsup_{p\to\infty} M_r < \infty$.

Following Aoshima and Yata [2] and Bai and Saranadasa [3], we assume that:

(A1) For all $r, s, t, u \in \{1, \ldots, q\}$ with $r \neq s, t, u$,

$$\text{E}(w_{rj}^2 w_{sj}^2) = \text{E}(w_{rj}^2)\text{E}(w_{sj}^2) = 1 \quad \text{and} \quad \text{E}(w_{rj}w_{sj}w_{tj}w_{uj}) = 0.$$

We also make the following assumption instead of (A1) whenever necessary:

(A2) For all $v \in \{2,\dots,8\}$, $r_1 \neq r_2 \neq \cdots \neq r_v \in [1,q]$ and $\alpha_1,\dots,\alpha_v \in [1,4]$ with $\alpha_1 + \cdots + \alpha_v \leq 8$,

$$\mathrm{E}(w_{r_1 j}^{\alpha_1} \cdots w_{r_v j}^{\alpha_v}) = \mathrm{E}(w_{r_1 j}^{\alpha_1}) \cdots \mathrm{E}(w_{r_v j}^{\alpha_v}).$$

See Chen and Qin [4] and Zhong and Chen [20] concerning (A2). Note that (A2) implies (A1). Further note that when $\boldsymbol{x}_j$ is Gaussian, $\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$ and $\boldsymbol{w}_j = \boldsymbol{z}_j$ in Eq. (2). In addition, (A2) is naturally satisfied when $\boldsymbol{x}_j$ is Gaussian because the elements of $\boldsymbol{z}_j$ are independent and $M_r = 2$ for all $r \in \{1,\dots,q\}$.

Furthermore, we impose the following assumption on $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ whenever required:

(A3) $\min\left\{\dfrac{\mathrm{tr}(\boldsymbol{\Sigma}_1^4)}{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)^2}, \dfrac{\mathrm{tr}(\boldsymbol{\Sigma}_2^4)}{\mathrm{tr}(\boldsymbol{\Sigma}_2^2)^2}\right\} \to 0$ as $p \to \infty$.

We note that if $p_i \to \infty$ and $\mathrm{tr}(\boldsymbol{\Sigma}_i^4)/\mathrm{tr}(\boldsymbol{\Sigma}_i^2)^2 \to 0$ as $p \to \infty$, (A3) holds even when $p_{i'}$ is fixed for $i' \neq i$. Also note that "$\mathrm{tr}(\boldsymbol{\Sigma}_i^4)/\mathrm{tr}(\boldsymbol{\Sigma}_i^2)^2 \to 0$ as $p \to \infty$" is equivalent to "$\lambda_{\max}(\boldsymbol{\Sigma}_i)/\mathrm{tr}(\boldsymbol{\Sigma}_i^2)^{1/2} \to 0$ as $p \to \infty$," where $\lambda_{\max}(\boldsymbol{\Sigma}_i)$ denotes the largest eigenvalue of $\boldsymbol{\Sigma}_i$. Let $m = \min(p,n)$ and $\Delta = \mathrm{tr}(\boldsymbol{\Sigma}_*\boldsymbol{\Sigma}_*^\top)$ $(= ||\boldsymbol{\Sigma}_*||_F^2)$, where $||\cdot||_F$ is the Frobenius norm. We note that $\Delta = 0$ is equivalent to $\boldsymbol{P} = \boldsymbol{0}$.

Finally, we also make either one of the following assumptions whenever the need arises:

(A4) $\dfrac{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)\mathrm{tr}(\boldsymbol{\Sigma}_2^2)}{n^2\Delta^2} \to 0$ as $m \to \infty$;

(A5) $\limsup_{m\to\infty}\left\{\dfrac{n^2\Delta^2}{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)\mathrm{tr}(\boldsymbol{\Sigma}_2^2)}\right\} < \infty$.

Note that (A5) holds under the null hypothesis $\mathcal{H}_0$ in Eq. (1). Also, note that $\Delta^2\{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)\mathrm{tr}(\boldsymbol{\Sigma}_2^2)\}^{-1} \in [0,1]$ from Eq. (A.1) in the Appendix. If Δ is sufficiently large to ensure that $\Delta^{-2}\mathrm{tr}(\boldsymbol{\Sigma}_1^2)\mathrm{tr}(\boldsymbol{\Sigma}_2^2) = O(1)$, then (A4) holds. If Δ is small enough that $\Delta = O(1)$, (A5) holds when $\{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)\mathrm{tr}(\boldsymbol{\Sigma}_2^2)\}^{-1} = O\{(p_1p_2)^{-1}\}$ and $n = O\{(p_1p_2)^{1/2}\}$.

3. ECDM methodology

Yata and Aoshima [19] developed the ECDM methodology as an extension of the CDM methodology given by Yata and Aoshima [18]. One of the advantages of the ECDM methodology is to produce an unbiased estimator having small asymptotic variance at a low computational cost. See Section 2.5 of Yata and Aoshima [19] for details. In this section, we propose a statistic for testing the hypotheses (1) based on the ECDM methodology.

3.1. Unbiased estimator by ECDM

We consider an unbiased estimator of Δ by the ECDM methodology. Let $n_{(1)} = \lceil n/2 \rceil$ and $n_{(2)} = n - n_{(1)}$, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. For every $k \in \{3, \ldots, 2n-1\}$, let

$$\boldsymbol{V}_{n(1)(k)} = \begin{cases} \{\lfloor k/2 \rfloor - n_{(1)} + 1, \ldots, \lfloor k/2 \rfloor\} & \text{if } \lfloor k/2 \rfloor \geq n_{(1)}, \\ \{1, \ldots, \lfloor k/2 \rfloor\} \cup \{\lfloor k/2 \rfloor + n_{(2)} + 1, \ldots, n\} & \text{otherwise,} \end{cases}$$

$$\boldsymbol{V}_{n(2)(k)} = \begin{cases} \{\lfloor k/2 \rfloor + 1, \ldots, \lfloor k/2 \rfloor + n_{(2)}\} & \text{if } \lfloor k/2 \rfloor \leq n_{(1)}, \\ \{1, \ldots, \lfloor k/2 \rfloor - n_{(1)}\} \cup \{\lfloor k/2 \rfloor + 1, \ldots, n\} & \text{otherwise,} \end{cases}$$

where $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x. Let $\#\boldsymbol{A}$ denote the cardinality of the set $\boldsymbol{A}$. Note that for all $\ell \in \{1, 2\}$ and $k \in \{3, \ldots, 2n-1\}$,

$$\#\boldsymbol{V}_{n(\ell)(k)} = n_{(\ell)}, \quad \boldsymbol{V}_{n(1)(k)} \cap \boldsymbol{V}_{n(2)(k)} = \boldsymbol{\emptyset}, \quad \boldsymbol{V}_{n(1)(k)} \cup \boldsymbol{V}_{n(2)(k)} = \{1, \ldots, n\}.$$

Further note that

$$\forall_{1 \leq i < j \leq n} \quad i \in \boldsymbol{V}_{n(1)(i+j)} \quad \text{and} \quad j \in \boldsymbol{V}_{n(2)(i+j)}. \tag{3}$$

For all $\ell \in \{1, 2\}$ and $k \in \{3, \ldots, 2n-1\}$, let

$$\overline{\boldsymbol{x}}_{\ell(1)(k)} = \frac{1}{n_{(1)}} \sum_{j \in \boldsymbol{V}_{n(1)(k)}} \boldsymbol{x}_{\ell j} \quad \text{and} \quad \overline{\boldsymbol{x}}_{\ell(2)(k)} = \frac{1}{n_{(2)}} \sum_{j \in \boldsymbol{V}_{n(2)(k)}} \boldsymbol{x}_{\ell j}.$$

For every $1 \leq i < j \leq n$, further let

$$\widehat{\Delta}_{ij} = (\boldsymbol{x}_{1i} - \overline{\boldsymbol{x}}_{1(1)(i+j)})^\top (\boldsymbol{x}_{1j} - \overline{\boldsymbol{x}}_{1(2)(i+j)})(\boldsymbol{x}_{2i} - \overline{\boldsymbol{x}}_{2(1)(i+j)})^\top (\boldsymbol{x}_{2j} - \overline{\boldsymbol{x}}_{2(2)(i+j)}).$$

Then, in view of Eq. (3), for all $1 \leq i < j \leq n$, we have the following facts:

(i) For all $\ell \in \{1, 2\}$, $\boldsymbol{x}_{\ell i} - \overline{\boldsymbol{x}}_{\ell(1)(i+j)}$ and $\boldsymbol{x}_{\ell j} - \overline{\boldsymbol{x}}_{\ell(2)(i+j)}$ are independent.

(ii) $\mathrm{E}(\widehat{\Delta}_{ij}) = \Delta\{(n_{(1)}-1)(n_{(2)}-1)\}/(n_{(1)}n_{(2)})$.

Let

$$u_n = \frac{n_{(1)}n_{(2)}}{(n_{(1)}-1)(n_{(2)}-1)}. \tag{4}$$

We propose to estimate Δ by

$$\widehat{T}_n = \frac{2u_n}{n(n-1)}\sum_{i<j}^{n}\widehat{\Delta}_{ij}.$$

Then, we have that $\mathrm{E}(\widehat{T}_n) = \Delta$.

Remark 1. One can save the computational cost of $\widehat{T}_n$ by using previously calculated values of $\overline{\boldsymbol{x}}_{1(i)(k)}$ and $\overline{\boldsymbol{x}}_{2(i)(k)}$ for $i \in \{1,2\}$ and $k \in \{3,\ldots,2n-1\}$. Then, the computational cost of $\widehat{T}_n$ is of the order, $O(n^2p)$.

Set

$$\overline{\boldsymbol{x}}_{1n} = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_{1j}, \quad \overline{\boldsymbol{x}}_{2n} = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_{2j},$$

and

$$\boldsymbol{S}_* = \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{x}_{1j}-\overline{\boldsymbol{x}}_{1n})(\boldsymbol{x}_{2j}-\overline{\boldsymbol{x}}_{2n})^\top.$$

Then $\mathrm{tr}(\boldsymbol{S}_*\boldsymbol{S}_*^\top)$ is a naive estimator of Δ and under (A1),

$$\mathrm{E}\{\mathrm{tr}(\boldsymbol{S}_*\boldsymbol{S}_*^\top)\} = \Delta + O\{\mathrm{tr}(\boldsymbol{\Sigma}_1)\mathrm{tr}(\boldsymbol{\Sigma}_2)/n\}.$$

Note that the bias term of $\mathrm{tr}(\boldsymbol{S}_*\boldsymbol{S}_*^\top)$ becomes very large as p increases. Srivastava and Reid [13] suggested estimating Δ by

$$\widehat{\Delta}_{SR} = \frac{(n-1)^2}{(n-2)(n+1)}\Big\{\mathrm{tr}(\boldsymbol{S}_*\boldsymbol{S}_*^\top) - \frac{\mathrm{tr}(\boldsymbol{S}_1)\mathrm{tr}(\boldsymbol{S}_2)}{n-1}\Big\}$$

when the underlying distribution is Gaussian, where $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are the sample covariance matrices. They showed that $\mathrm{E}(\widehat{\Delta}_{SR}) = \Delta$. However, $\widehat{\Delta}_{SR}$ can be severely biased when the Gaussian assumption fails. In contrast, the proposed estimator, $\widehat{T}_n$, is always unbiased and one can claim that $\mathrm{E}(\widehat{T}_n) = \Delta$ without any assumptions.

Remark 2. We give the following `Mathematica` algorithm to calculate $\widehat{T}_n$:

Input: Sample size n and $n \times p_i$ data matrices $X[1], X[2]$ such as for $i \in \{1,2\}$, $X[i] = (\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{in})^\top$.

Mathematica code:

- $n1$ =Ceiling[$n/2$]; $n2 = n - n1$; $u = 2*n1*n2/((n1-1)*(n2-1)*n*(n-1))$
- V[$1, k\_, X\_$] :=If [Floor[$k/2$] $\geq n1$, Take[X, {Floor[$k/2$]$-n1+1$, Floor[$k/2$]}],
  Join[Take[X, {1, Floor[$k/2$]}], Take[X, {Floor[$k/2$] $+ n2 + 1, n$}]]]
- V[$2, k\_, X\_$] :=If [Floor[$k/2$] $\leq n1$, Take[X, {Floor[$k/2$] + 1, Floor[$k/2$] + $n2$}],
  Join[Take[X, {1, Floor[$k/2$] $- n1$}, Take[X, {Floor[$k/2$] $+ 1, n$}]]]
- Do[M[i, j, k] =Mean[$V[j, k, X[i]$], {$k, 3, 2*n-1$}, {$i, 1, 2$}, {$j, 1, 2$}]
- $T = u*$Sum[(Part[$X[1], i$]$-$M[$1, 1, i+j$]).(Part[$X[1], j$]$-$M[$1, 2, i+j$])
  $*$(Part[$X[2], i$]$-$M[$2, 1, i+j$]).(Part[$X[2], j$]$-$M[$2, 2, i+j$]), {$j, 2, n$}, {$i, 1, j-1$}]

Then $T = \widehat{T}_n$.

3.2. Asymptotic properties of $\widehat{T}_n$

We first consider the consistency of $\widehat{T}_n$ in the sense that $\widehat{T}_n/\Delta = 1 + o_P(1)$ as $m \to \infty$. Let $\delta = n^{-1}\{2\mathrm{tr}(\mathbf{\Sigma}_1^2)\mathrm{tr}(\mathbf{\Sigma}_2^2)\}^{1/2}$. Let $M_r' = M_r - 2$ for all $r \in \{1, \ldots, q\}$ and note that $M_r' = 0$ when the underlying distribution is Gaussian. We have the following result.

Lemma 3.1. *Assume (A1). Then, as $m \to \infty$,*

$$\mathrm{var}(\widehat{T}_n) = \Big\{4\frac{\mathrm{tr}(\mathbf{\Sigma}_1\mathbf{\Sigma}_*\mathbf{\Sigma}_2\mathbf{\Sigma}_*^\top) + \mathrm{tr}\{(\mathbf{\Sigma}_*\mathbf{\Sigma}_*^\top)^2\} + \sum_{j=1}^q M_j'(\boldsymbol{\gamma}_{1j}^\top\mathbf{\Sigma}_*\boldsymbol{\gamma}_{2j})^2}{n} + 2\,\frac{\Delta^2}{n^2} + \delta^2\Big\}\{1 + o(1)\} + O\Big[\frac{\{\mathrm{tr}(\mathbf{\Sigma}_1^4)\mathrm{tr}(\mathbf{\Sigma}_2^4)\}^{1/2}}{n^2}\Big].$$

Remark 3. When the underlying distribution is Gaussian and $\mathbf{\Sigma}_* = \mathbf{0}$, Srivastava and Reid [13] showed that, as $m \to \infty$,

$$\mathrm{var}(\widehat{\Delta}_{SR}) = \delta^2\{1 + o(1)\}$$

under a certain regularity condition which is stronger than (A3). Note that $\mathrm{var}(\widehat{T}_n)$ given in Lemma 3.1 is asymptotically equivalent to $\mathrm{var}(\widehat{\Delta}_{SR})$ under (A3) and $\mathbf{\Sigma}_* = \mathbf{0}$.

From Lemma 3.1, we can also deduce that $\widehat{T}_n$ is consistent, as stated next.

Theorem 3.1. *Assume (A1) and (A4). Then, as* $m \to \infty$, $\widehat{T}_n/\Delta = 1 + o_P(1)$.

While consistency holds whenever (A4) is satisfied, we can show that $\widehat{T}_n$ is asymptotically Normal under a different set of assumptions, as detailed below.

Lemma 3.2. *Assume (A1), (A3) and (A5). Then, as* $m \to \infty$, $\mathrm{var}(\widehat{T}_n) = \delta^2\{1 + o(1)\}$.

From Lemma 3.2, we have the asymptotic normality of $\widehat{T}_n$ as follows.

Theorem 3.2. *Assume (A2), (A3) and (A5). Then, as* $m \to \infty$

$$\frac{\widehat{T}_n - \Delta}{\sqrt{\mathrm{var}(\widehat{T}_n)}} = \frac{\widehat{T}_n - \Delta}{\delta} + o_P(1) \rightsquigarrow \mathcal{N}(0,1),$$

where $\rightsquigarrow$ *denotes convergence in distribution and* $\mathcal{N}(0,1)$ *denotes a random variable distributed as the standard Normal distribution.*

3.3. Estimation of $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)$

Given that $\mathrm{tr}(\boldsymbol{\Sigma}_1^2)$ and $\mathrm{tr}(\boldsymbol{\Sigma}_2^2)$ are unknown in δ, it is necessary to estimate them to construct a test of the hypotheses (1). Following Yata and Aoshima [19], an estimator of $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)$ is given, for $i \in \{1,2\}$, by

$$W_{in} = \frac{2u_n}{n(n-1)} \sum_{r<s}^{n} \big\{(\boldsymbol{x}_{ir} - \overline{\boldsymbol{x}}_{i(1)(r+s)})^\top (\boldsymbol{x}_{is} - \overline{\boldsymbol{x}}_{i(2)(r+s)})\big\}^2.$$

Note that $\mathrm{E}(W_{in}) = \mathrm{tr}(\boldsymbol{\Sigma}_i^2)$. From Lemma 3.1, we have the following result.

Lemma 3.3. *Assume (A1). Then, for* $i \in \{1,2\}$, *as* $m \to \infty$,

$$\mathrm{var}\Big\{\frac{W_{in}}{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}\Big\} = \Big[\frac{4}{n\mathrm{tr}(\boldsymbol{\Sigma}_i^2)^2}\Big\{2\mathrm{tr}(\boldsymbol{\Sigma}_i^4) + \sum_{j=1}^{q} M_j'(\boldsymbol{\gamma}_{ij}^\top \boldsymbol{\Sigma}_i \boldsymbol{\gamma}_{ij})^2\Big\} + \frac{4}{n^2}\Big]\{1 + o(1)\} \to 0.$$

Remark 4. In Section 2.5 of Yata and Aoshima [19], they compared W_{in} with other estimators of $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)$ theoretically and computationally. They showed that W_{in} has small asymptotic variance at a low computational cost.

Let $\widehat{\delta} = n^{-1}(2W_{1n}W_{2n})^{1/2}$. Then, by combining Theorem 3.2 with Lemma 3.3, we have the following result.

Corollary 3.1. *Assume (A2), (A3) and (A5). Then, as* $m \to \infty$, $(\widehat{T}_n - \Delta)/\widehat{\delta} \rightsquigarrow \mathcal{N}(0,1)$.

As an illustration, we consider a simple example in which

$$p_1 = p_2, \quad \boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\Sigma}_1 = (0.3^{|i-j|^{1/3}}), \quad \boldsymbol{\Sigma}_2 = (0.4^{|i-j|^{1/3}}), \quad \boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}.$$

For $i \in \{1,2\}$, let $\boldsymbol{\Sigma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i \boldsymbol{H}_i^\top$, where $\boldsymbol{\Lambda}_i = \mathrm{diag}(\lambda_{i1}, \ldots, \lambda_{ip_i})$ with eigenvalues $\lambda_{i1} \geq \cdots \geq \lambda_{ip_i} \geq 0$, and $\boldsymbol{H}_i$ is an orthogonal matrix with the corresponding eigenvectors. We consider two scenarios:

(a) $\Delta = 0$, in which case

$$\boldsymbol{x}_{1j} = \boldsymbol{H}_1 \boldsymbol{\Lambda}_1^{1/2} (w_{1j}, \ldots, w_{p_1 j})^\top, \quad \boldsymbol{x}_{2j} = \boldsymbol{H}_2 \boldsymbol{\Lambda}_2^{1/2} (w_{p_1+1j}, \ldots, w_{pj})^\top.$$

(b) $\Delta = \lambda_{13}\lambda_{23}$, in which case

$$\begin{aligned} \boldsymbol{x}_{1j} &= \boldsymbol{H}_1 \boldsymbol{\Lambda}_1^{1/2} (w_{1j}, \ldots, w_{p_1 j})^\top \\ \boldsymbol{x}_{2j} &= \boldsymbol{H}_2 \boldsymbol{\Lambda}_2^{1/2} (w_{p_1+1j}, w_{p_1+2j}, w_{3j}, w_{p+4j}, \ldots, w_{pj})^\top. \end{aligned}$$

For each choice of $(p,n) \in \{(10,25), (200,50), (4000,150)\}$, vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ were generated independently from a pseudo-random Normal distribution with mean vector zero and covariance matrix $\boldsymbol{\Sigma}$. Note that (A2), (A3) and (A5) hold from the fact that $\Delta = O(1)$.

Displayed in Figure 2 are two histograms of 2000 independent outcomes of $\widehat{T}_n/\widehat{\delta}$ in scenarios (a), (b), and (p,n), together with probability densities of $\mathcal{N}(0,1)$ and $\mathcal{N}(\Delta/\delta, 1)$. From Corollary 3.1, we expect that $\widehat{T}_n/\widehat{\delta}$ is close to $\mathcal{N}(0,1)$ when $\Delta = 0$ and $\mathcal{N}(\Delta/\delta, 1)$ when $\Delta \neq 0$. When $(p,n) = (10,25)$, the histograms detract considerably from the asymptotic densities. When $(p,n) = (200,50)$, the histogram for (a) approaches the $\mathcal{N}(0,1)$ fairly well. However, the histogram for (b) is still far from the $\mathcal{N}(\Delta/\delta, 1)$. This is because the convergence in Lemma 3.2 is slow for $\Delta \neq 0$ compared to $\Delta = 0$. As expected, both the histograms match the limiting distributions very closely when $(p,n) = (4000,150)$. For other simulation settings such as $p_1 = p - 1$ and $p_2 = 1$, see Section 2 of Yata and Aoshima [19].

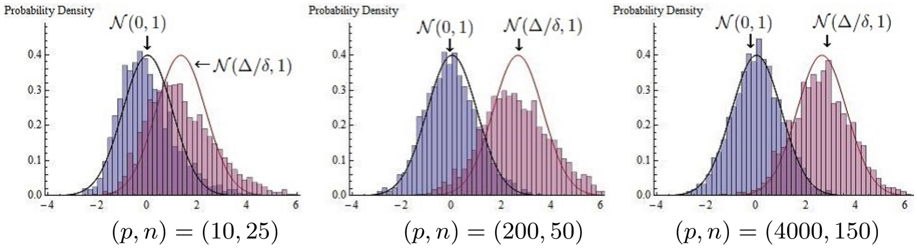


Figure 2: The solid lines are probability densities of $\mathcal{N}(0,1)$ and $\mathcal{N}(\Delta/\delta,1)$. The histograms of $\widehat{T}_n/\widehat{\delta}$ for cases of (a) $\Delta=0$ and (b) $\Delta\neq 0$ fit the solid lines with increasing dimension and sample size: $(p,n)=(10,25)$, $(200,50)$ and $(4000,150)$.

4. Test of high-dimensional correlations

In this section, we propose a test of the hypotheses (1) in high-dimensional settings.

4.1. Test procedure for hypotheses (1)

Let $\alpha\in(0,1/2)$ be a prespecified constant. Let z_α be a constant such that $\Pr\{\mathcal{N}(0,1)>z_\alpha\}=\alpha$. From Corollary 3.1, we propose to

$$\text{rejecting } \mathcal{H}_0 \quad \Leftrightarrow \quad \widehat{T}_n/\widehat{\delta}>z_\alpha. \tag{5}$$

Then, we have the following result.

Theorem 4.1. *Under (A2) and (A3), the size* Π_0 *and power* Π *of testing procedure* (5) *are such that*

$$\Pi_0=\alpha+o(1) \quad \textit{and} \quad \Pi(\Delta_\star)-\Phi\Big(\frac{\Delta_\star}{\delta}-z_\alpha\Big)=o(1),$$

where Φ *denotes the cumulative distribution function of* $\mathcal{N}(0,1)$ *and* $\Pi(\Delta_\star)$ *denotes the power when* $\Delta=\Delta_\star$ *for given* $\Delta_\star>0$.

When (A4) is met, we have the following result from Theorem 3.1.

Corollary 4.1. *Assume (A1) and (A4) under* $\mathcal{H}_1$. *Then the test* (5) *is such that, for any* $\Delta>0$, *as* $m\to\infty$, $\Pi(\Delta)=1+o(1)$.

Remark 5. Let

$$K = \Big\{ 4 \frac{\mathrm{tr}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_* \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_*^\top) + \mathrm{tr}\{(\boldsymbol{\Sigma}_* \boldsymbol{\Sigma}_*^\top)^2\} + \sum_{j=1}^q M_j' (\boldsymbol{\gamma}_{1j}^\top \boldsymbol{\Sigma}_* \boldsymbol{\gamma}_{2j})^2}{n} + 2\frac{\Delta^2}{n^2} + \delta^2 \Big\}^{1/2}.$$

Then, from Lemma 3.1, one has $\mathrm{var}(\widehat{T}_n)\, K^{-2} \to 1$ as $m \to \infty$ under (A1) and (A3). Hence, from Theorem 3.2, one may write the power in Theorem 4.1 as

$$\Pi(\Delta_\star) - \Phi\Big(\frac{\Delta_\star}{K} - \frac{z_\alpha \delta}{K}\Big) = o(1).$$

4.2. Simulation

In order to study the performance of the test (5), we used computer simulations. We set $\alpha = 0.05$, $p_1 = p_2$, $\boldsymbol{\mu} = \boldsymbol{0}$,

$$\boldsymbol{\Sigma}_1 = \boldsymbol{B}(0.3^{|i-j|^{1/3}})\boldsymbol{B}, \quad \boldsymbol{\Sigma}_2 = \boldsymbol{B}(0.4^{|i-j|^{1/3}})\boldsymbol{B}, \quad \boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2},$$

where

$$\boldsymbol{B} = \mathrm{diag}[\{0.5 + 1/(p_1+1)\}^{1/2}, \ldots, \{0.5 + p_1/(p_1+1)\}^{1/2}].$$

Note that for $i \in \{1,2\}$, $\mathrm{tr}(\boldsymbol{\Sigma}_i) = p_i$. We set (a) $\Delta = 0$ and (b) $\Delta = \lambda_{13}\lambda_{23}$, which are the same settings as in Figure 2. We considered three distributions for $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, namely

(I) $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$;

(II) for all $r \in \{1, \ldots, q\}$, $w_{rj} = 2^{-1/2}(v_{rj} - 1)$, where v_{rj}s are i.i.d. as $\chi^2_{(1)}$, the chi-squared distribution with 1 degree of freedom;

(III) $\boldsymbol{w}_j$s are i.i.d. as p-variate t-distribution, $t_p(\nu)$, with mean zero, covariance matrix $\boldsymbol{I}_p$ and degrees of freedom $\nu = 10$.

Note that (A2) is met in (I) and (II). However, (A1) — or (A2) — is not met in scenario (III). We set $p = 2^s$ ($s \in \{4, \ldots, 11\}$) and $n = 4\lceil p_1^{1/2} \rceil$. We note that (A3) and (A5) hold for (a) and (b).

We compared the performance of $\widehat{T}_n$ with $\widehat{\Delta}_{SR}/\widehat{\delta}_{SR}$ of Srivastava and Reid [13], where, for $i \in \{1,2\}$,

$$\begin{aligned} \widehat{\delta}_{SR} &= \{2W_{1(SR)}W_{2(SR)}\}^{1/2}/n, \\ W_{i(SR)} &= (n-1)^2\{(n-2)(n+1)\}^{-1}\{\mathrm{tr}(\boldsymbol{S}_i^2) - (n-1)^{-1}\mathrm{tr}(\boldsymbol{S}_i)^2\}. \end{aligned}$$

Srivastava and Reid [13] showed that $\widehat{\Delta}_{SR}/\widehat{\delta}_{SR}$ is asymptotically Normal when the underlying distribution is Gaussian and $\Delta = 0$. Also, note that $\mathrm{E}(\widehat{\Delta}_{SR}) = \Delta$ only under the Gaussian assumption. In contrast, from Corollary 3.1, $\widehat{T}_n/\widehat{\delta}$ is asymptotically Normal even for non-Gaussian situations and $\Delta \neq 0$. Also, one can claim that $\mathrm{E}(\widehat{T}_n) = \Delta$ without any assumptions such as (A1).

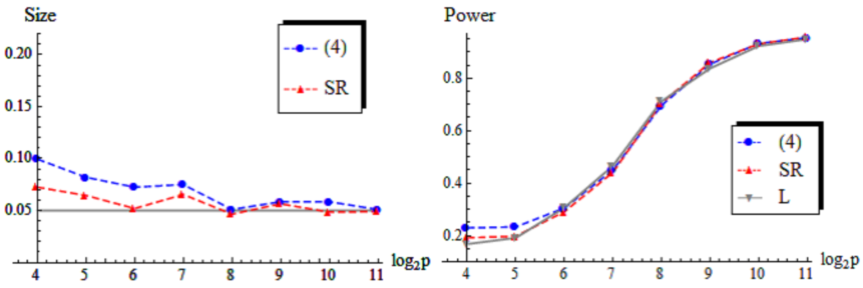
Figure 3 summarizes the findings obtained by averaging the outcomes from 4000 ($= 2R$, say) replications for scenarios (I)–(III). Here, the first 2000 replications were generated for (a) when $\Delta = 0$ and the last 2000 replications were generated for (b) when $\Delta \neq 0$. We defined $P_r = 1$ (or 0) when $\mathcal{H}_0$ was falsely rejected (or not) for all $r \in \{1, \ldots, 2000\}$, and when $\mathcal{H}_1$ was falsely rejected (or not) for all $r \in \{2001, \ldots, 4000\}$. We used

$$\overline{\alpha} = \frac{1}{R}\sum_{r=1}^{R} P_r \quad \text{and} \quad 1 - \overline{\beta} = 1 - \frac{1}{R}\sum_{r=R+1}^{2R} P_r$$

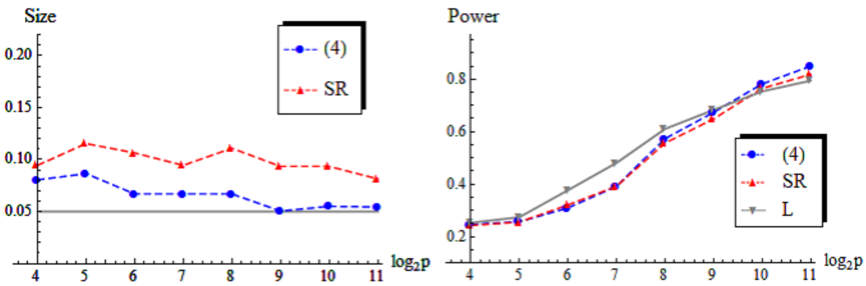
to estimate the size in the left panels, and the power in the right panels, respectively. Their standard deviations are less than 0.011.

Let $L = \Phi(\Delta K^{-1} - z_\alpha \delta K^{-1})$. From Theorem 4.1, and in view of Remark 5, we expected that $\overline{\alpha}$ and $1 - \overline{\beta}$ for (5) would be close to 0.05 and L, respectively. Figure 4 exhibits the averages (in the left panels) and the sample variances (in the right panels) of $\widehat{T}_n/\Delta$ and $\widehat{\Delta}_{SR}/\Delta$ for the outcomes of (b) when $\Delta \neq 0$ in scenarios (I)–(III). From Remark 5, the asymptotic variance for $\widehat{T}_n/\Delta$ was given by K^2/Δ^2.

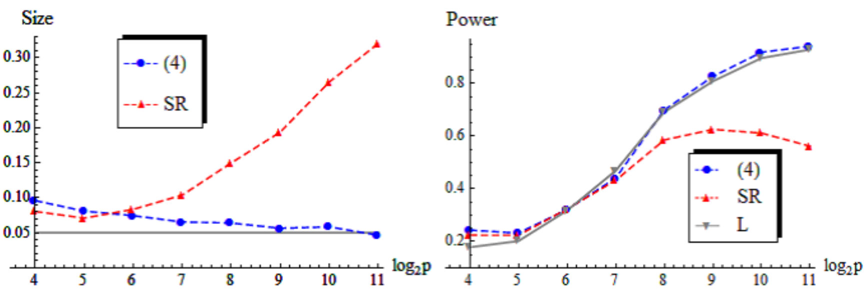
From Figures 3 and 4, we observe that $\widehat{\Delta}_{SR}$ performs well in the Gaussian case. However, for non-Gaussian cases such as (II) and (III), $\widehat{\Delta}_{SR}$ does not do so well and was particularly bad under scenario (III). This is probably because (A1) — or (A2) — is not met in scenario (III). In contrast, the behavior of $\widehat{T}_n$ was adequate for high-dimensional cases, even in the non-Gaussian situations. We further note that $\widehat{T}_n$ is quite robust against other non-Gaussian situations. Therefore, we can recommend to use $\widehat{T}_n$ for testing hypotheses (1) and for the estimation of Δ.



(I) $\mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$.

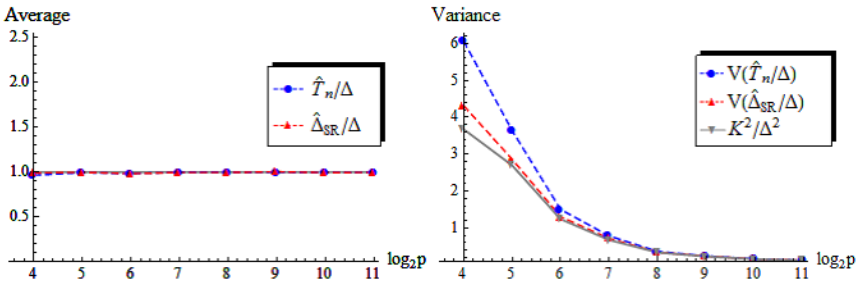


(II) The chi-squared distribution with 1 degree of freedom.

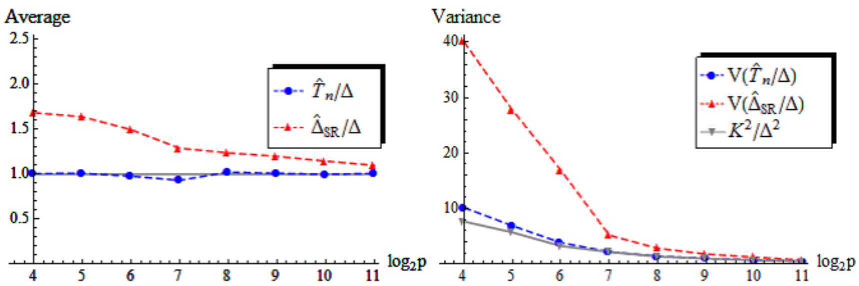


(III) $t_p(10)$.

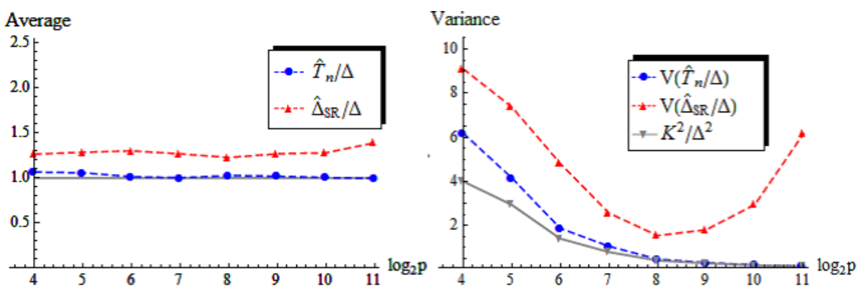
Figure 3: The values of $\overline{\alpha}$ are denoted by the dashed lines in the left panels and the values of $1 - \overline{\beta}$ are denoted by the dashed lines in the right panels for the tests by (5) and $\widehat{\Delta}_{SR}/\widehat{\delta}_{SR}$ (SR) in scenarios (I)–(III). The asymptotic powers were given by $L = \Phi(\Delta K^{-1} - z_\alpha \delta K^{-1})$, which was denoted by the solid lines in the right panels.



(I) $\mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$.



(II) The chi-squared distribution with 1 degree of freedom.



(III) $t_p(10)$.

Figure 4: The averages of $\widehat{T}_n/\Delta$ and $\widehat{\Delta}_{SR}/\Delta$ are denoted by the dashed lines in the left panels and their sample variances, $V(\widehat{T}_n/\Delta)$ and $V(\widehat{\Delta}_{SR}/\Delta)$, are denoted by the dashed lines in the right panels for (b) in scenarios (I)–(III). The asymptotic variance of $\widehat{T}_n/\Delta$ was given by K^2/Δ^2, which was denoted by the solid lines in the right panels.

5. Applications

In this section, we give several applications of the results stated in Section 3.

5.1. Confidence interval for Δ

A confidence interval of asymptotic level $\alpha \in (0,1)$ for Δ is given by

$$I = [\max\{\widehat{T}_n - z_{\alpha/2}\widehat{\delta},\ 0\},\ \widehat{T}_n + z_{\alpha/2}\widehat{\delta}].$$

Indeed, from Corollary 3.1, one has, as $m \to \infty$,

$$\Pr(\Delta \in I) = 1 - \alpha + o(1)$$

under (A2), (A3) and (A5). Hence, one can estimate Δ by I. If one considers $\boldsymbol{\Sigma}_0$ as a candidate of $\boldsymbol{\Sigma}_*$, one can check whether $\boldsymbol{\Sigma}_0$ is a valid candidate or not according as $||\boldsymbol{\Sigma}_0||_F^2 \in I$ or not.

5.2. Checking whether (A4) holds or not

As discussed in Section 3, $\widehat{T}_n$ is consistent when (A4) is met, and $\widehat{T}_n$ is asymptotically Normal when (A5) is met. Here, we propose a method to check whether (A4) holds or not.

Let $\widehat{\kappa} = W_{1n}W_{2n}(n\widehat{T}_n)^{-2}$. We have the following result.

Proposition 5.1. *Assume (A1). Then, as* $m \to \infty$*,*

$$\widehat{\kappa} = o_P(1) \quad \textit{under (A4)}; \quad \widehat{\kappa}^{-1} = O_P(1) \quad \textit{under (A5).}$$

From Proposition 5.1, one can distinguish (A4) and (A5). If $\widehat{\kappa}$ is sufficiently small, one may call on (A4); otherwise one can invoke (A5).

5.3. Estimation of the RV-coefficient

Let $\rho = \Delta\{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)\mathrm{tr}(\boldsymbol{\Sigma}_2^2)\}^{-1/2}$. Here, ρ is the (population) RV-coefficient, which is a multivariate generalization of the squared Pearson correlation coefficient. Note that $\rho \in [0,1]$; see Robert and Escoufier [11] for details. Smilde et al. [12] considered the RV-coefficient for high-dimensional data.

Let $\widehat{\rho} = \widehat{T}_n(W_{1n}W_{2n})^{-1/2}$. We then have the following result.

Proposition 5.2. *Assume (A1). Then, as $m \to \infty$,*

$$\widehat{\rho} = \rho + O_P(1/n + \rho/n^{1/2}) + O_P\Big[\Big\{\frac{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_*\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_*^\top)}{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)\mathrm{tr}(\boldsymbol{\Sigma}_2^2)n}\Big\}^{1/2}\Big] = \rho + O_P(n^{-1/2}).$$

Thus, one can estimate the RV coefficient by $\widehat{\rho}$ for high-dimensional data.

5.4. Test of high-dimensional covariance structures

We consider testing

$$\mathcal{H}_0 : \boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}_0 \quad \text{vs.} \quad \mathcal{H}_1 : \boldsymbol{\Sigma}_* \neq \boldsymbol{\Sigma}_0, \tag{6}$$

where $\boldsymbol{\Sigma}_0$ is a candidate covariance structure. Let $\Delta_0 = ||\boldsymbol{\Sigma}_* - \boldsymbol{\Sigma}_0||_F^2$ and

$$\begin{aligned} \widehat{\Delta}_{ij,0} \;=\;& u_n\widehat{\Delta}_{ij} - n_{(1)}(\boldsymbol{x}_{1i} - \overline{\boldsymbol{x}}_{1(1)(i+j)})^\top\boldsymbol{\Sigma}_0(\boldsymbol{x}_{2i} - \overline{\boldsymbol{x}}_{2(1)(i+j)})/(n_{(1)} - 1) \\ & - n_{(2)}(\boldsymbol{x}_{1j} - \overline{\boldsymbol{x}}_{1(2)(i+j)})^\top\boldsymbol{\Sigma}_0(\boldsymbol{x}_{2j} - \overline{\boldsymbol{x}}_{2(2)(i+j)})/(n_{(2)} - 1), \end{aligned}$$

where u_n is defined as in Eq. (4). Note that

$$\mathrm{E}(\widehat{\Delta}_{ij,0}) = ||\boldsymbol{\Sigma}_*||_F^2 - 2\mathrm{tr}(\boldsymbol{\Sigma}_*^\top\boldsymbol{\Sigma}_0) = \Delta_0 - ||\boldsymbol{\Sigma}_0||_F^2.$$

Then, we can test hypotheses (6) using the statistic

$$\widehat{T}_{n,0} = \frac{2}{n(n-1)}\sum_{i<j}^{n}\widehat{\Delta}_{ij,0} + ||\boldsymbol{\Sigma}_0||_F^2.$$

Note that $\mathrm{E}(\widehat{T}_{n,0}) = \Delta_0$.

Let $\boldsymbol{\Sigma}_{*0} = \boldsymbol{\Sigma}_* - \boldsymbol{\Sigma}_0$. Then, we have the following result.

Lemma 5.1. *Assume (A1). Then, as $m \to \infty$,*

$$\begin{aligned} \mathrm{var}(\widehat{T}_{n,0}) = \Big[4\frac{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_{*0}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_{*0}^\top) + \mathrm{tr}\{(\boldsymbol{\Sigma}_*\boldsymbol{\Sigma}_{*0}^\top)^2\} + \sum_{j=1}^q M_j'(\boldsymbol{\gamma}_{1j}^\top\boldsymbol{\Sigma}_{*0}\boldsymbol{\gamma}_{2j})^2}{n} \\ + 2\,\frac{\Delta^2}{n^2} + \delta^2\Big]\{1 + o(1)\} + O\Big[\frac{\{\mathrm{tr}(\boldsymbol{\Sigma}_1^4)\mathrm{tr}(\boldsymbol{\Sigma}_2^4)\}^{1/2}}{n^2}\Big]. \end{aligned}$$

From Lemma 5.1, Theorems 3.1 and 3.2, we have the following results.

Corollary 5.1. *Assume (A1). Assume also (A4) with $\Delta = \Delta_0$. Then, as $m \to \infty$, $\widehat{T}_{n,0}/\Delta_0 = 1 + o_P(1)$.*

Corollary 5.2. *Assume (A2) and (A3). Assume also (A5) with* $\Delta = \Delta_0$*. Then, as* $m \to \infty$*,*

$$\frac{\widehat{T}_{n,0} - \Delta_0}{(\delta^2 + 2\Delta^2/n^2)^{1/2}} \rightsquigarrow \mathcal{N}(0,1).$$

Hence, from Lemma 3.3 and Corollary 5.2, we can test (6) by

$$\text{rejecting } \mathcal{H}_0 \quad \Leftrightarrow \quad \widehat{T}_{n,0}/(\widehat{\delta}^2 + 2||\mathbf{\Sigma}_0||_F^4/n^2)^{1/2} > z_\alpha.$$

Under (A2) and (A3), the size of this test is $\alpha + o(1)$, as $m \to \infty$.

6. Application

In this section, we demonstrate how the proposed test procedures perform in practice using microarray data. We analyzed gene expression data of *Arabidopsis thaliana* given by Wille et al. [16] in which the data set consists of 118 samples having 834 ($= p$) genes: 39 ($= p_1$) isoprenoid genes and 795 ($= p_2$) additional genes. All the data were logarithmic transformed. Wille et al. [16] considered a genetic network between the two sets of genes. By using graphical Gaussian modeling, they constructed the isoprenoid gene network given in Figure 2 of [16]. In Figure 5, we illustrate the isoprenoid gene network and the additional genes.

We first consider testing (1) using (5). See Figure 1 for illustration. Let $\alpha = 0.05$. We found $\widehat{T}_n = 352.5$ and $\widehat{\delta} = 7.296$, so that $\widehat{T}_n/\widehat{\delta} = 48.3$. From (5) and $z_\alpha = 1.645$, we are led to reject $\mathcal{H}_0$ and to conclude that the two networks are connected. In addition, we found $\widehat{\kappa} = 0.000214$. Thus, with the help of Proposition 5.1 one may conclude that (A4) is met, so that the power of the test is 1 asymptotically and $\widehat{T}_n/\Delta = 1 + o_P(1)$ from Theorem 3.1 and Corollary 4.1. Also, with the help of Proposition 5.2 we found $\widehat{\rho} = 0.579$ as an estimate of the RV-coefficient.

Next, we considered testing hypotheses (1) between some part of the isoprenoid genes and the additional genes. The isoprenoid genes consisted of three types as the MEP pathway (19 genes), the MVA pathway (15 genes) and mitochondrion (5 genes). See [16] for details. From Figure 5 we expect that (i) the

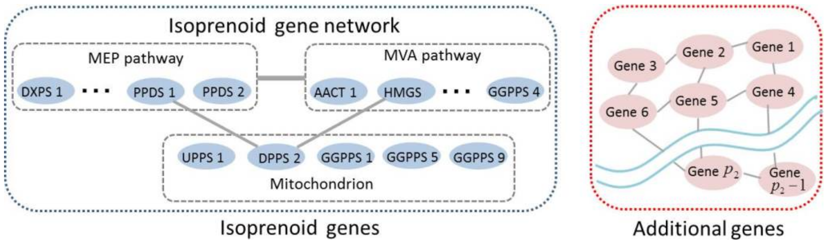


Figure 5: Illustration of the isoprenoid gene network given by Figure 2 in Wille et al. [16] and the additional genes, where DXPS1, PPDS1 and so on, are names of genes. DPPS2 is connected with both the MEP pathway and the MVA pathway. Other genes of mitochondrion are not connected with either the MEP pathway or the MVA pathway.

correlation between DPPS2 and the additional genes is high, and (ii) the correlation between the genes of mitochondrion (except DPPS2) and the additional genes is low. We set $\boldsymbol{x}_{2j}$ as the additional genes ($p_2 = 795$). We considered three tests for $\boldsymbol{x}_{1j}$: (a) the genes of mitochondrion ($p_1 = 5$); (b) DPPS2 ($p_1 = 1$); and (c) UPPS1, GGPPS1, 5, 9 ($p_1 = 4$). By using the first 50 samples ($n = 50$) of the 118 samples, we constructed (5). At level $\alpha = 0.05$, we can then reject $\mathcal{H}_0$ for (a) since $\widehat{T}_n/\widehat{\delta} = 12.27$ and for (b) since $\widehat{T}_n/\widehat{\delta} = 13.23$. On the other hand, we fail to reject $\mathcal{H}_0$ for (c) since $\widehat{T}_n/\widehat{\delta} = 1.417$. Hence, we could conclude (i) and (ii).

We also considered the correlation test for the genes of mitochondrion by the multiple testing procedure with FWER ≤ 0.05 given by (22) in Yata and Aoshima [19]. This led to the conclusion that UPPS1 and DPPS2 have correlations with the additional genes, that is (ii) fails the multiple test.

Proceeding as in Section 5 in [19], we also considered a high-dimensional linear regression model, viz.

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\Theta} + \boldsymbol{E},$$

where $\boldsymbol{Y}$ is an $n \times p_2$ response matrix, $\boldsymbol{X}$ is an $n \times k$ fixed design matrix, $\boldsymbol{E}$ is an $n \times p_2$ error matrix with mean zero, and $\boldsymbol{\Theta}$ is a $k \times p_2$ parameter matrix.

Let $\boldsymbol{x}_{1j}$ be the jth sample of the 35 isoprenoid genes (except UPPS1, and GGPPS1, 5, 9). For each $j \in \{1, \dots, 118\}$, let $\boldsymbol{x}_{1(j)} = (1, \boldsymbol{x}_{1j}^\top)^\top$. We set $\boldsymbol{Y} =$

$(\boldsymbol{x}_{21}, \dots, \boldsymbol{x}_{2n})^\top$ and $\boldsymbol{X} = (\boldsymbol{x}_{1(1)}, \dots, \boldsymbol{x}_{1(n)})^\top$ with $k = 36$. We note that the standard elements of $\boldsymbol{\Theta}$ are path coefficients from the isoprenoid genes to the additional genes. By using the first 50 observations as a training data set, we obtained the least squared estimator of $\boldsymbol{\Theta}$ by $\widehat{\boldsymbol{\Theta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}$.

We investigated the prediction accuracy of the regression with $\widehat{\boldsymbol{\Theta}}$ by using the remaining 68 observations ($68 = 118 - 50$) as a test data set. To this end, we used the prediction mean squared error (PMSE), viz.

$$\mathrm{E}(||\boldsymbol{x}_{2j} - \widehat{\boldsymbol{\Theta}}^\top \boldsymbol{x}_{1(j)}||^2 | \widehat{\boldsymbol{\Theta}}).$$

By using the test samples $\boldsymbol{x}_{1(j)}$ and $\boldsymbol{x}_{2j}$ for $j \in \{51, \dots, 118\}$, we applied the bias-corrected and accelerated (BCa) bootstrap by Efron [6]. Then, we constructed a 95% confidence interval (CI) of the PMSE by $[837.6, 1189.5]$ from 10,000 replications. On the other hand, we considered the PMSE for the full isoprenoid (39 genes). Then, similar to above, we constructed 95% CI of the PMSE by $[1088.7, 1581.3]$. The PMSE by the 35 isoprenoid genes is probably smaller than that of the full isoprenoid genes. Thus we conclude that the test (5) effectively works for this data set.

Appendix: Proofs

Throughout, we assume that $\boldsymbol{\mu}_1 = \boldsymbol{0}$ and $\boldsymbol{\mu}_2 = \boldsymbol{0}$ without loss of generality. Let $\Upsilon = \mathrm{tr}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_* \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_*^\top)$, $\Psi = \mathrm{tr}(\boldsymbol{\Sigma}_1^2)\mathrm{tr}(\boldsymbol{\Sigma}_2^2)$ and $\Omega = \mathrm{tr}(\boldsymbol{\Sigma}_1^4)\mathrm{tr}(\boldsymbol{\Sigma}_2^4)$. Note that

$$\sum_{i=1}^q (\boldsymbol{\gamma}_{1i}^\top \boldsymbol{\Sigma}_* \boldsymbol{\gamma}_{2i})^2 \le \sum_{i,j}^q (\boldsymbol{\gamma}_{1i}^\top \boldsymbol{\Sigma}_* \boldsymbol{\gamma}_{2j})^2 = \Upsilon;$$

$$\mathrm{tr}\{(\boldsymbol{\Sigma}_* \boldsymbol{\Sigma}_*^\top)^2\} = \sum_{i,j}^q (\boldsymbol{\gamma}_{1i}^\top \boldsymbol{\Sigma}_* \boldsymbol{\gamma}_{2j})(\boldsymbol{\gamma}_{1j}^\top \boldsymbol{\Sigma}_* \boldsymbol{\gamma}_{2i}) \le \sum_{i,j}^q (\boldsymbol{\gamma}_{1i}^\top \boldsymbol{\Sigma}_* \boldsymbol{\gamma}_{2j})^2 = \Upsilon;$$

$$\Delta = \sum_{i,j}^q (\boldsymbol{\gamma}_{1i}^\top \boldsymbol{\gamma}_{1j} \boldsymbol{\gamma}_{2i}^\top \boldsymbol{\gamma}_{2j}) \le \prod_{\ell=1}^2 \Big\{ \sum_{i,j}^q (\boldsymbol{\gamma}_{\ell i}^\top \boldsymbol{\gamma}_{\ell j})^2 \Big\}^{1/2} = \Psi^{1/2}; \quad \text{and}$$

$$\Upsilon = \sum_{i,j}^q (\boldsymbol{\gamma}_{1i}^\top \boldsymbol{\Sigma}_1 \boldsymbol{\gamma}_{1j})(\boldsymbol{\gamma}_{2i}^\top \boldsymbol{\Sigma}_2 \boldsymbol{\gamma}_{2j}) \le \prod_{\ell=1}^2 \Big\{ \sum_{i,j}^q (\boldsymbol{\gamma}_{\ell i}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\gamma}_{\ell j})^2 \Big\}^{1/2} = \Omega^{1/2} \le \Psi \quad \text{(A.1)}$$

from the fact that for $i \in \{1,2\}$, $\mathrm{tr}(\boldsymbol{\Sigma}_i^4) \le \mathrm{tr}(\boldsymbol{\Sigma}_i^2)^2$. Then, we note that $K^2 = O(\Psi n^{-2} + \Upsilon n^{-1})$, where K is given in Remark 5. Let $y_{ij} = u_n \widehat{\Delta}_{ij} - \Delta$ and $\varepsilon_{ij} = \boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1j} \boldsymbol{x}_{2i}^\top \boldsymbol{x}_{2j} - \Delta$ for all $i < j$. Note that

$$\widehat{T}_n - \Delta = \frac{2}{n(n-1)} \sum_{i<j}^{n} y_{ij}.$$

For all $i \neq j$, let

$$\eta_{ij} = \sum_{r\neq s}^{q} \sum_{t=1}^{q} \boldsymbol{\gamma}_{1r}^\top \boldsymbol{\gamma}_{1t} \boldsymbol{\gamma}_{2s}^\top \boldsymbol{\gamma}_{2t} w_{ri} w_{si} (w_{tj}^2 - 1),$$

$$\psi_{ij} = \sum_{r,t}^{q} \boldsymbol{\gamma}_{1r}^\top \boldsymbol{\gamma}_{1t} \boldsymbol{\gamma}_{2r}^\top \boldsymbol{\gamma}_{2t} (w_{ri}^2 - 1)(w_{tj}^2 - 1),$$

and

$$\phi_{ij} = \sum_{r\neq s}^{q} \sum_{t\neq u}^{q} \boldsymbol{\gamma}_{1r}^\top \boldsymbol{\gamma}_{1t} \boldsymbol{\gamma}_{2s}^\top \boldsymbol{\gamma}_{2u} w_{ri} w_{si} w_{tj} w_{uj}.$$

Note that $\mathrm{E}(\phi_{ij}) = 0$ for all $i \neq j$ and $\mathrm{E}(\phi_{ij}\phi_{i'j}) = 0$ for all $i \neq i' \neq j$. Let

$$U_n = \frac{2}{n(n-1)} \sum_{i<j}^{n} \varepsilon_{ij}, \quad V_n = \frac{2}{n(n-1)} \sum_{i<j}^{n} \phi_{ij}$$

and $B = \mathrm{E}(\phi_{ij}^2)$ for any $i \neq j$. Furthermore, for all $i < j$, set

$$\widehat{\boldsymbol{\Sigma}}_{*,ij(1)} = n_{(1)}(n_{(1)} - 1)^{-1} (\boldsymbol{x}_{1i} - \overline{\boldsymbol{x}}_{1(1)(i+j)})(\boldsymbol{x}_{2i} - \overline{\boldsymbol{x}}_{2(1)(i+j)})^\top$$

and

$$\widehat{\boldsymbol{\Sigma}}_{*,ij(2)} = n_{(2)}(n_{(2)} - 1)^{-1} (\boldsymbol{x}_{1j} - \overline{\boldsymbol{x}}_{1(2)(i+j)})(\boldsymbol{x}_{2j} - \overline{\boldsymbol{x}}_{2(2)(i+j)})^\top.$$

Proof of Lemma 3.1. For all $i < j$, write

$$y_{ij} = \mathrm{tr}\{(\widehat{\boldsymbol{\Sigma}}_{*,ij(1)} - \boldsymbol{\Sigma}_*)(\widehat{\boldsymbol{\Sigma}}_{*,ij(2)} - \boldsymbol{\Sigma}_*)^\top\} + \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_{*,ij(1)} \boldsymbol{\Sigma}_*^\top) + \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_{*,ij(2)} \boldsymbol{\Sigma}_*^\top) - 2\Delta$$

and

$$\varepsilon_{ij} = \phi_{ij} + \eta_{ij} + \eta_{ji} + \psi_{ij} + \mathrm{tr}(\boldsymbol{x}_{1i} \boldsymbol{x}_{2i}^\top \boldsymbol{\Sigma}_*^\top) + \mathrm{tr}(\boldsymbol{x}_{1j} \boldsymbol{x}_{2j}^\top \boldsymbol{\Sigma}_*^\top) - 2\Delta. \tag{A.2}$$

For all $i < j$, we note that ϕ_{ij}, η_{ij}, η_{ji} and ψ_{ij} are uncorrelated under (A1). Also note that

$$\sum_{i=1}^{q} (\boldsymbol{\gamma}_{1i}^\top \boldsymbol{\Sigma}_1 \boldsymbol{\gamma}_{1i})(\boldsymbol{\gamma}_{2i}^\top \boldsymbol{\Sigma}_2 \boldsymbol{\gamma}_{2i}) \le \sum_{i,j}^{q} |(\boldsymbol{\gamma}_{1i}^\top \boldsymbol{\Sigma}_1 \boldsymbol{\gamma}_{1j})(\boldsymbol{\gamma}_{2i}^\top \boldsymbol{\Sigma}_2 \boldsymbol{\gamma}_{2j})| \le \Omega^{1/2}.$$

Under (A1), one then has, for all $i \neq j$,

$$\mathrm{E}(\psi_{ij}^2) = O\Big\{\sum_{r,t}^{q}(\boldsymbol{\gamma}_{1r}^\top\boldsymbol{\gamma}_{1t}\boldsymbol{\gamma}_{2r}^\top\boldsymbol{\gamma}_{2t})^2\Big\} = O\Big(\sum_{r=1}^{q}\boldsymbol{\gamma}_{1r}^\top\boldsymbol{\Sigma}_1\boldsymbol{\gamma}_{1r}\boldsymbol{\gamma}_{2r}^\top\boldsymbol{\Sigma}_2\boldsymbol{\gamma}_{2r}\Big) = O(\Omega^{1/2}).$$

Similarly, under (A1), one also has $\mathrm{E}(\eta_{ij}^2) = O(\Omega^{1/2})$ for all $i \neq j$. Then, we have that under (A1), for all $i < j$,

$$\mathrm{E}(\varepsilon_{ij}^2) = \Psi + \Delta^2 + O(\Upsilon + \Omega^{1/2});$$

and for all $i < j < k$,

$$\begin{aligned}\mathrm{E}(\varepsilon_{ij}\varepsilon_{ik}) &= \mathrm{E}(\varepsilon_{ik}\varepsilon_{jk}) = \mathrm{var}\{\mathrm{tr}(\boldsymbol{x}_{1i}\boldsymbol{x}_{2i}^\top\boldsymbol{\Sigma}_*^\top)\} \\ &= \Upsilon + \mathrm{tr}\{(\boldsymbol{\Sigma}_*\boldsymbol{\Sigma}_*^\top)^2\} + \sum_{r=1}^{q}(M_r - 2)(\boldsymbol{\gamma}_{1r}^\top\boldsymbol{\Sigma}_*\boldsymbol{\gamma}_{2r})^2.\end{aligned}$$

We also have that under (A1), for all $i < j$ and $k < \ell$; $i \neq j \neq k \neq \ell$, $\mathrm{E}(\varepsilon_{ij}\varepsilon_{k\ell}) = 0$. Then, under (A1), we have, as $m \to \infty$,

$$\mathrm{var}(U_n) = \mathrm{E}(U_n^2) = K^2\{1 + o(1)\} + O(\Omega^{1/2}/n^2) = O(K^2) \tag{A.3}$$

On the other hand, we have that under (A1), for all $i < j$, $\mathrm{E}\{(y_{ij} - \varepsilon_{ij})^2\} = O(\Psi/n)$; and for all $i < j < k$,

$$\begin{aligned}\mathrm{E}\{(y_{ij} - \varepsilon_{ij})(y_{ik} - \varepsilon_{ik})\} &= O(\Psi/n^2 + \Upsilon/n), \\ \mathrm{E}\{(y_{ik} - \varepsilon_{ik})(y_{jk} - \varepsilon_{jk})\} &= O(\Psi/n^2 + \Upsilon/n).\end{aligned}$$

We also have that under (A1), for all $i < j$ and $k < \ell$; $i \neq j \neq k \neq \ell$,

$$\mathrm{E}\{(y_{ij} - \varepsilon_{ij})(y_{k\ell} - \varepsilon_{k\ell})\} = O(\Psi/n^3 + \Upsilon/n^2).$$

Then, under (A1), we have that, as $m \to \infty$,

$$\mathrm{var}(U_n - \widehat{T}_n) = \mathrm{E}[\{U_n - (\widehat{T}_n - \Delta)\}^2] = o(K^2). \tag{A.4}$$

Hence, by combining Eq. (A.3) with Eq. (A.4), we have that under (A1), as $m \to \infty$,

$$\begin{aligned}\mathrm{var}(\widehat{T}_n) &= \mathrm{var}(U_n) + \mathrm{var}(U_n - \widehat{T}_n) - 2\mathrm{E}[\{U_n - (\widehat{T}_n - \Delta)\}U_n] \\ &= K^2\{1 + o(1)\} + O(\Omega^{1/2}/n^2)\end{aligned}$$

from the facts that $\mathrm{var}(\widehat{T}_n) = \mathrm{E}[\{(\widehat{T}_n - \Delta) - U_n + U_n\}^2]$ and

$$|\mathrm{E}[\{U_n - (\widehat{T}_n - \Delta)\}U_n]| \le \{\mathrm{var}(U_n - \widehat{T}_n)\mathrm{var}(U_n)\}^{1/2}$$

by the Cauchy–Schwarz inequality. This concludes the argument. □

Proof of Lemma 3.2. Let $r_* = \mathrm{rank}(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_*)$. When we consider the singular value decomposition of $\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_*$, it follows that $\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_* = \sum_{j=1}^{r_*}\lambda_{*j}\boldsymbol{h}_{*j(1)}\boldsymbol{h}_{*j(2)}^\top$, where $\lambda_{*1} \ge \cdots \ge \lambda_{*r_*} > 0$ denote the singular values of $\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_*$ and for each $j \in \{1, \ldots, r_*\}$, $\boldsymbol{h}_{*j(1)}$ (or $\boldsymbol{h}_{*j(2)}$) denotes a unit left- (or right-) singular vector corresponding to λ_{*j}. Note that $\Upsilon = \mathrm{tr}(\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_*\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_*^\top\boldsymbol{\Sigma}_1^{1/2})$. Then

$$\begin{aligned}\Upsilon &= \mathrm{tr}\Big\{\Big(\sum_{j=1}^{r_*}\lambda_{*j}\boldsymbol{h}_{*j(1)}\boldsymbol{h}_{*j(2)}^\top\Big)\boldsymbol{\Sigma}_2\Big(\sum_{j=1}^{r_*}\lambda_{*j}\boldsymbol{h}_{*j(2)}\boldsymbol{h}_{*j(1)}^\top\Big)\Big\}\\ &= \sum_{j=1}^{r_*}\lambda_{*j}^2\boldsymbol{h}_{*j(2)}^\top\boldsymbol{\Sigma}_2\boldsymbol{h}_{*j(2)} \le \lambda_{\max}(\boldsymbol{\Sigma}_2)\sum_{j=1}^{r_*}\lambda_{*j}^2 = \lambda_{\max}(\boldsymbol{\Sigma}_2)\mathrm{tr}(\boldsymbol{\Sigma}_*^\top\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_*).\end{aligned}$$

Similarly, we can see that $\mathrm{tr}(\boldsymbol{\Sigma}_*^\top\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_*) \le \lambda_{\max}(\boldsymbol{\Sigma}_1)\mathrm{tr}(\boldsymbol{\Sigma}_*^\top\boldsymbol{\Sigma}_*) = \lambda_{\max}(\boldsymbol{\Sigma}_1)\Delta$, so that

$$\Upsilon \le \lambda_{\max}(\boldsymbol{\Sigma}_1)\lambda_{\max}(\boldsymbol{\Sigma}_2)\Delta. \tag{A.5}$$

Thus under (A3), one has $\Upsilon = o(\Delta\Psi^{1/2})$ as $p \to \infty$. It follows that $n\Upsilon\Psi^{-1} = o(n\Delta\Psi^{-1/2})$, so that under (A3) and (A5), as $m \to \infty$,

$$n\Upsilon\Psi^{-1} = o(1). \tag{A.6}$$

By noting that

$$\sum_{i=1}^{q}(\boldsymbol{\gamma}_{1i}^\top\boldsymbol{\Sigma}_*\boldsymbol{\gamma}_{2i})^2 \le \Upsilon, \quad \mathrm{tr}\{(\boldsymbol{\Sigma}_*\boldsymbol{\Sigma}_*^\top)^2\} \le \Upsilon,$$

from Lemma 3.1 and Eq. (A.6), we see that, as $m \to \infty$, $\mathrm{var}(\widehat{T}_n)/\delta^2 = 1 + o(1)$ under (A1), (A3) and (A5) from the fact that $\Delta^2\Psi^{-1} = o(1)$ under (A5). □

Proof of Theorem 3.1. It follows from Eq. (A.5) that $\Upsilon \le \Psi^{1/2}\Delta$, and hence $K^2 = O(\Psi n^{-2} + \Psi^{1/2}\Delta n^{-1})$. From Lemma 3.1 and the fact that $\Omega^{1/2} \le \Psi$, it follows that, as $m \to \infty$, $\mathrm{var}(T_n\Delta^{-1}) = O\{(n^2\Delta^2)^{-1}\Psi + (n\Delta)^{-1}\Psi^{1/2}\}$ under (A1). Thus, under (A4), Chebyshev's inequality's allows us to conclude. □

The following lemmas are instrumental in the proof of Theorem 3.2.

Lemma A.1. *Under (A2), one has*

(i) $\mathrm{E}(\phi_{ij}^2\phi_{i'j}^2) = O(\Psi^2)$ *for all* $i, i' \neq j$;

(ii) $\mathrm{E}(\phi_{ij}\phi_{i'j}\phi_{ij'}\phi_{i'j'}) = O(\Omega)$ *for all* $i \neq i' \neq j \neq j'$.

Proof. To prove (i), let $\zeta_{rstu} = \boldsymbol{\gamma}_{1r}^\top\boldsymbol{\gamma}_{1t}\boldsymbol{\gamma}_{2s}^\top\boldsymbol{\gamma}_{2u}$ for all r, s, t, u. Let also

$$A_1 = \sum_{r\neq s}^{q}\sum_{t\neq u}^{q}\zeta_{rstu}(\zeta_{rstu} + \zeta_{srtu} + \zeta_{rsut} + \zeta_{srut})w_{ri}^2w_{si}^2w_{tj}^2w_{uj}^2$$

and $A_2 = \phi_{ij}^2 - A_1$ for $i \neq j$. Note that $\mathrm{E}(A_1) = B$ and $\mathrm{E}(A_2) = 0$ under (A2). We can see that

$$\sum_{r\neq s}^{q}\sum_{t\neq u}^{q}(\zeta_{rstu}^2 + \zeta_{srtu}^2 + \zeta_{rsut}^2 + \zeta_{srut}^2) = O(\Psi),$$

and hence

$$\sum_{r\neq s}^{q}\sum_{t\neq u}^{q}(|\zeta_{rstu}| + |\zeta_{srtu}| + |\zeta_{rsut}| + |\zeta_{srut}|)^2 = O(\Psi).$$

Then, under (A2), we have

$$\begin{aligned}\mathrm{E}(A_1^2) &\leq \mathrm{E}\Big\{\Big(\sum_{r\neq s}^{q}\sum_{t\neq u}^{q}(|\zeta_{rstu}| + |\zeta_{srtu}| + |\zeta_{rsut}| + |\zeta_{srut}|)^2w_{ri}^2w_{si}^2w_{tj}^2w_{uj}^2\Big)^2\Big\}\\ &= O(\Psi^2). \qquad \text{(A.7)}\end{aligned}$$

For $\mathrm{E}(A_2^2)$, it is necessary to consider the terms of $w_{ri}^3w_{r'i}^3w_{r''i}^2$ $(r \neq r' \neq r'')$ because it does not hold that $\mathrm{E}(w_{ri}^3w_{r'i}^3w_{r''i}^2) = 0$ $(r \neq r' \neq r'')$ unless $\mathrm{E}(w_{ri}^3) = 0$ or $\mathrm{E}(w_{r'i}^3) = 0$. Here, under (A2), we can assert that for sufficiently large

$C > 0$,

$$\Big|\mathrm{E}\Big(\sum_{r\neq r'\neq r''}^{q}\sum_{t\neq u}^{q}\zeta_{rr'tu}\zeta_{rr''tu}\sum_{t'\neq u'}^{q}\zeta_{rr't'u'}\zeta_{r'r''t'u'}w_{ri}^3w_{r'i}^3w_{r''i}^2w_{tj}^2w_{uj}^2w_{t'j}^2w_{u'j}^2\Big)\Big|$$

$$\leq C\sum_{r\neq r'\neq r''}^{q}\sum_{t\neq u}^{q}|\zeta_{rr'tu}\zeta_{rr''tu}|\sum_{t'\neq u'}^{q}|\zeta_{rr't'u'}\zeta_{r'r''t'u'}|$$

$$\leq C\sum_{r,r',r''}^{q}\Big\{\Big(\sum_{t,u}^{q}\zeta_{rr'tu}^2\Big)\Big(\sum_{t,u}^{q}\zeta_{rr''tu}^2\Big)\Big\}^{1/2}\Big\{\Big(\sum_{t,u}^{q}\zeta_{rr'tu}^2\Big)\Big(\sum_{t,u}^{q}\zeta_{r'r''tu}^2\Big)\Big\}^{1/2}$$

$$\leq C\Big\{\sum_{r,r',r''}^{q}\Big(\sum_{t,u}^{q}\zeta_{rr'tu}^2\Big)\Big(\sum_{t,u}^{q}\zeta_{rr''tu}^2\Big)\Big\}^{1/2}\Big\{\sum_{r,r',r''}^{q}\Big(\sum_{t,u}^{q}\zeta_{rr'tu}^2\Big)\Big(\sum_{t,u}^{q}\zeta_{r'r''tu}^2\Big)\Big\}^{1/2}$$

$$\leq C\Big(\sum_{r,r'}^{q}\sum_{t,u}^{q}\zeta_{rr'tu}^2\Big)\Big(\sum_{r,r''}^{q}\sum_{t,u}^{q}\zeta_{rr''tu}^2\Big)^{1/2}\Big(\sum_{r',r''}^{q}\sum_{t,u}^{q}\zeta_{r'r''tu}^2\Big)^{1/2}=O(\Psi^2),$$

where the fact that $|\mathrm{E}(w_{ri}^3)| \leq \{\mathrm{E}(w_{ri}^4)\mathrm{E}(w_{ri}^2)\}^{1/2} \leq M_r^{1/2}$ for all $r \in \{1,\dots,q\}$ has been used. Similarly, for other terms, we can evaluate the order to be $O(\Psi^2)$. Hence, we can claim that $\mathrm{E}(A_2^2) = O(\Psi^2)$ under (A2), so that $\mathrm{E}(\phi_{ij}^4) = O\{\mathrm{E}(A_1^2) + \mathrm{E}(A_2^2)\} = O(\Psi^2)$ from Eq. (A.7). Finally, noting that

$$\mathrm{E}(\phi_{ij}^2\phi_{i'j}^2) \leq \{\mathrm{E}(\phi_{ij}^4)\mathrm{E}(\phi_{i'j}^4)\}^{1/2},$$

we conclude that (i) holds.

To show (ii), note that from Eq. (A.1), we have that, under (A2),

$$\mathrm{E}(\phi_{ij}\phi_{i'j}\phi_{ij'}\phi_{i'j'}) = O(\Omega) + O(\Upsilon^2) = O(\Omega)$$

for all $i \neq i' \neq j \neq j'$. This is enough to conclude. □

Lemma A.2. *Under (A1), (A3) and (A5), one has, as $m \to \infty$,* $\mathrm{var}(\widehat{T}_n - V_n) = o(\delta^2)$.

Proof. From Eq. (A.2), we have that under (A1),

$$\mathrm{E}\{(\phi_{ij}-\varepsilon_{ij})^2\} = O(\Upsilon + \Omega^{1/2}) \quad \text{for all } i \neq j;$$
$$\mathrm{E}\{(\phi_{ij}-\varepsilon_{ij})(\phi_{ik}-\varepsilon_{ik})\} = O(\Upsilon) \quad \text{for all } i \neq j \neq k;$$
$$\text{and}\quad \mathrm{E}\{(\phi_{ij}-\varepsilon_{ij})(\phi_{k\ell}-\varepsilon_{k\ell})\} = 0 \quad \text{for all } i \neq j \neq k \neq \ell.$$

Then, from Eq. (A.6), we have, under (A1), (A3) and (A5), that as $m \to \infty$,

$$\mathrm{var}(U_n - V_n) = O(\Upsilon/n + \Omega^{1/2}/n^2) = o(\delta^2). \tag{A.8}$$

By combining Eq. (A.8) with Eq. (A.4), and using the fact that $\mathrm{var}(\widehat{T}_n - V_n) = O\{\mathrm{var}(\widehat{T}_n - U_n) + \mathrm{var}(U_n - V_n)\}$, we can conclude. $\square$

Proof of Theorem 3.2. For each $j \in \{2, \dots, n\}$, let

$$v_j = \frac{2}{n(n-1)} \sum_{i=1}^{j-1} \phi_{ij}.$$

Note that

$$\sum_{j=2}^{n} v_j = \frac{2}{n(n-1)} \sum_{i<j}^{n} \phi_{ij} = V_n$$

and that, for all $j \in \{3, \dots, n\}$, $\mathrm{E}(v_j | v_{j-1}, \dots, v_2) = 0$.

Now for each $j \in \{2, \dots, n\}$, let $\xi_j = v_j[2B\{n(n-1)\}^{-1}]^{-1/2}$. Note that

$$\sum_{j=2}^{n} \mathrm{E}(\xi_j^2) = 1, \quad \mathrm{var}\Big(\sum_{j=2}^{n} \xi_j\Big) = 1$$

from the fact that $\mathrm{var}(\sum_{j=2}^{n} v_j) = 2B/\{n(n-1)\}$. Let $I(\cdot)$ denote the indicator function. By noting that

$$\sum_{i=1}^{q} (\boldsymbol{\gamma}_{1i}^\top \boldsymbol{\Sigma}_1 \boldsymbol{\gamma}_{1i})(\boldsymbol{\gamma}_{2i}^\top \boldsymbol{\Sigma}_2 \boldsymbol{\gamma}_{2i}) \le \Omega^{1/2}$$

from Eq. (A.1), we can deduce, under (A2) and (A3), that as $p \to \infty$,

$$B = \Psi + \Delta^2 + O(\Omega^{1/2}) = \Psi\{1 + o(1)\} + \Delta^2. \tag{A.9}$$

Then, by using Chebyshev's inequality and the Cauchy–Schwarz inequality, from Lemma A.1, under (A2) and (A3), the Lindeberg condition holds that, as $m \to \infty$,

$$\sum_{j=2}^{n} \mathrm{E}\{\xi_j^2 I(\xi_j^2 \ge \tau)\} \le \sum_{j=2}^{n} \frac{\mathrm{E}(\xi_j^4)}{\tau} = O\Big(\frac{\Psi^2}{B^2 n}\Big) \to 0 \tag{A.10}$$

for any $\tau > 0$. Hence, from Lemma A.1, Eq. (A.9) and Eq. (A.10), we deduce that under (A2) and (A3), as $m \to \infty$,

$$\sum_{2 \le i < j \le n} \mathrm{E}[\{\xi_i^2 - \mathrm{E}(\xi_i^2)\}\{\xi_j^2 - \mathrm{E}(\xi_j^2)\}] = O\Big(\frac{\Psi^2}{B^2 n} + \frac{\Omega}{B^2}\Big) \to 0$$

and

$$\sum_{j=2}^{n} \mathrm{E}[\{\xi_j^2 - \mathrm{E}(\xi^2)\}^2] \leq \sum_{j=2}^{n} \mathrm{E}(\xi_j^4) \to 0,$$

so that

$$\mathrm{var}\Big(\sum_{j=2}^{n} \xi_j^2\Big) = \mathrm{E}\Big[\Big\{\sum_{j=2}^{n} \{\xi_j^2 - \mathrm{E}(\xi_j^2)\}\Big\}^2\Big] \to 0. \tag{A.11}$$

Note that

$$\mathrm{var}(\widehat{T}_n)^{1/2}[2B\{n(n-1)\}^{-1}]^{-1/2} = \delta[2B\{n(n-1)\}^{-1}]^{-1/2} + o(1) \to 1$$

and $\widehat{T}_n - \Delta = V_n + o_P(\delta)$ as $m \to \infty$ under (A2), (A3) and (A5) from Lemmas 3.2 and A.2. Then, proceeding as in the proof of Theorem 2.1 in [19], from Eq. (A.10) and Eq. (A.11), under (A2), (A3) and (A5), we obtain that, as $m \to \infty$,

$$\frac{\widehat{T}_n - \Delta}{\sqrt{\mathrm{var}(\widehat{T}_n)}} = \frac{\widehat{T}_n - \Delta}{\delta} + o_P(1) = \sum_{j=2}^{n} \xi_j + o_P(1) \Rightarrow \mathcal{N}(0,1). \tag{A.12}$$

This concludes the argument. □

Proof of Lemma 3.3. Upon replacing $(\boldsymbol{\Sigma}_2, \boldsymbol{\gamma}_{2j}, \boldsymbol{\Sigma}_*, \Delta)$ with $(\boldsymbol{\Sigma}_1, \boldsymbol{\gamma}_{1j}, \boldsymbol{\Sigma}_1, \mathrm{tr}(\boldsymbol{\Sigma}_1^2))$ in Lemma 3.1, we can get the result when $i = 1$. The result for $i = 2$ follows in a similar way. □

Proof of Corollary 3.1. It suffices to combine Theorem 3.2 with Lemma 3.3 to conclude. □

Proofs of Theorem 4.1 and Corollary 4.1. To prove Corollary 4.1, under (A1) and (A4), from Theorem 3.1 and Lemma 3.3, we obtain that, as $m \to \infty$,

$$\Pr(\widehat{T}_n/\widehat{\delta} > z_\alpha) = \Pr(\widehat{T}_n/\Delta > z_\alpha \widehat{\delta}/\Delta) = \Pr\{1 + o_P(1) > o_P(1)\} \to 1.$$

We conclude the result of Corollary 4.1.

Next, to prove Theorem 4.1, one can proceed as in the proof of Theorem 2.2 in [19]. The results concerning the size and power when (A5) is met can be deduced from Corollary 3.1. We note that $\Phi(\Delta\delta^{-1} - z_\alpha) \to 1$ as $m \to \infty$ under

(A4), so that we obtain the result of power when (A4) is met from Corollary 4.1. Hence, by considering the convergent subsequence of Δ/δ, we can obtain the power result stated in Theorem 4.1. This concludes the proof. □

Proof of Proposition 5.1. We first consider the case when (A4) is met. From Theorem 3.1 and Lemma 3.3, one has $\widehat{\kappa} = (n\Delta)^{-2}\Psi\{1 + o_P(1)\} = o_P(1)$ as $m \to \infty$ under (A1) and (A4). Thus the result holds when (A4) is met.

Next, we consider the case when (A5) is met. From Eq. (A.5), one has

$$\Upsilon \le \lambda_{\max}(\mathbf{\Sigma}_1)\lambda_{\max}(\mathbf{\Sigma}_2)\Delta \le \Psi^{1/2}\Delta,$$

so that $n\Upsilon\Psi^{-1} = O(1)$ under (A5). Then, from Lemma 3.1 and (A.1), under (A1) and (A5), we deduce that $\mathrm{var}(\widehat{T}_n) = O(\Psi n^{-2})$ as $m \to \infty$. Note that $\Delta = O(\Psi^{1/2}n^{-1})$ under (A5). Thus under (A1) and (A5), one has $\widehat{T}_n = \Delta + O_P(\Psi^{1/2}n^{-1}) = O_P(\Psi^{1/2}n^{-1})$ as $m \to \infty$. Then, from Lemma 3.3, under (A1) and (A5), we have that, as $m \to \infty$,

$$\widehat{\kappa}^{-1} = \Psi^{-1}n^2\widehat{T}_n^2\{1 + o_P(1)\} = O_P(1).$$

Therefore, the result holds when (A5) is met and the proof is complete. □

Proof of Proposition 5.2. The conclusion derives easily by combining Lemmas 3.1 and 3.3. □

Proof of Lemma 5.1. For all $i < j$, let $y_{ij,0} = \widehat{\Delta}_{ij,0} + ||\mathbf{\Sigma}_0||_F^2 - \Delta_0$ and

$$\varepsilon_{ij,0} = \varepsilon_{ij} - \boldsymbol{x}_{1i}^\top\mathbf{\Sigma}_0\boldsymbol{x}_{2i} - \boldsymbol{x}_{1j}^\top\mathbf{\Sigma}_0\boldsymbol{x}_{2j} + 2\mathrm{tr}(\mathbf{\Sigma}_*\mathbf{\Sigma}_0^\top).$$

From Eq. (A.2), we can write

$$y_{ij,0} = \mathrm{tr}\{(\widehat{\mathbf{\Sigma}}_{*,ij(1)} - \mathbf{\Sigma}_*)(\widehat{\mathbf{\Sigma}}_{*,ij(2)} - \mathbf{\Sigma}_*)^\top\} + \mathrm{tr}(\widehat{\mathbf{\Sigma}}_{*,ij(1)}\mathbf{\Sigma}_{*0}^\top) + \mathrm{tr}(\widehat{\mathbf{\Sigma}}_{*,ij(2)}\mathbf{\Sigma}_{*0}^\top) - 2\mathrm{tr}(\mathbf{\Sigma}_*\mathbf{\Sigma}_{*0}^\top)$$

and

$$\varepsilon_{ij,0} = \phi_{ij} + \eta_{ij} + \eta_{ji} + \psi_{ij} + \mathrm{tr}(\boldsymbol{x}_{1i}\boldsymbol{x}_{2i}^\top\mathbf{\Sigma}_{*0}^\top) + \mathrm{tr}(\boldsymbol{x}_{1j}\boldsymbol{x}_{2j}^\top\mathbf{\Sigma}_{*0}^\top) - 2\mathrm{tr}(\mathbf{\Sigma}_*\mathbf{\Sigma}_{*0}^\top).$$

The rest of the argument is similar to the proof of Lemma 3.1. □

Proofs of Corollaries 5.1 and 5.2. We first consider the proof of Corollary 5.1. Let $\Upsilon_0 = \mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_{*0}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_{*0}^\top)$. Similar to Eq. (A.5), it holds that $\Upsilon_0 \le \Psi^{1/2}\Delta_0$. Then, by noting that

$$\sum_{i=1}^{q}(\boldsymbol{\gamma}_{1i}^\top\boldsymbol{\Sigma}_{*0}\boldsymbol{\gamma}_{2i})^2 \le \Upsilon_0$$

and $\mathrm{tr}\{(\boldsymbol{\Sigma}_{*0}\boldsymbol{\Sigma}_*^\top)^2\} \le \Upsilon_0$, from Lemma 5.1, we have that as $m \to \infty$

$$\mathrm{var}(\widehat{T}_{n,0}/\Delta_0) = O\{\Psi/(n^2\Delta_0^2) + \Psi^{1/2}/(n\Delta_0)\}$$

under (A1). Thus, under (A4) with $\Delta = \Delta_0$, from Chebyshev's inequality, we can claim the result of Corollary 5.1.

Next, we consider the proof of Corollary 5.2. Similar to the proof of Lemma A.2, under (A1), (A3) and (A5) with $\Delta = \Delta_0$, we can claim that $\mathrm{var}(\widehat{T}_{n,0} - V_n) = o(\delta^2)$ as $m \to \infty$. From Eq. (A.9), we also note that, as $m \to \infty$,

$$(\delta^2 + 2\Delta^2 n^{-2})^{1/2}[2B\{n(n-1)\}^{-1}]^{-1/2} \to 1$$

under (A2) and (A3). Thus, by analogy with Eq. (A.12), we can conclude. □

Acknowledgment

The authors would like to thank two anonymous referees for their constructive comments. The authors are also grateful to the Editor-in-Chief, Christian Genest, for his substantial editorial assistance. The research of the first author was partially supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 26800078. The research of the second author was partially supported by Grants-in-Aid for Scientific Research (B) and Challenging Exploratory Research, JSPS, under Contract Numbers 22300094 and 26540010.

References

[1] M. Aoshima, K. Yata, Two-stage procedures for high-dimensional data, Sequential Anal. 30 (2011) 356–399.

[2] M. Aoshima, K. Yata, Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions, Methodol. Comput. Appl. Probab. 17 (2015) 419–439.

[3] Z. Bai, H. Saranadasa, Effect of high dimension: By an example of a two sample problem, Statist. Sinica 6 (1996) 311–329.

[4] S.X. Chen, Y.-L. Qin, A two-sample test for high-dimensional data with applications to gene-set testing, Ann. Statist. 38 (2010) 808-835.

[5] M. Drton, M.D. Perlman, Multiple testing and error control in Gaussian graphical model selection, Statist. Sci. 22 (2007) 430–449.

[6] B. Efron, Better bootstrap confidence intervals, J. Amer. Statist. Assoc. 82 (1987) 171–185.

[7] Y. Fujikoshi, V. Ulyanov, R. Shimizu, Multivariate Statistics: High-Dimensional and Large-Sample Approximations, Wiley, Hoboken, New Jersey, 2010.

[8] A. Hero, B. Rajaratnam, Large-scale correlation screening, J. Amer. Statist. Assoc. 106 (2011) 1540–1552.

[9] M. Hyodo, N. Shutoh, T. Nishiyama, T. Pavlenko, Testing block-diagonal covariance structure for high-dimensional data, Statist. Neerlandica 69 (2015) 460–482.

[10] W. Lan, H. Wang, C.-L. Tsai, Testing covariates in high-dimensional regression, Ann. Inst. Statist. Math. 66 (2014) 279–301.

[11] P. Robert, Y. Escoufier, A unifying tool for linear multivariate statistical methods: the RV-coefficient, J. R. Statist. Soc. Ser. C 25 (1976) 257–265.

[12] A.K. Smilde, H.A. Kiers, S. Bijlsma, C.M. Rubingh, M.J. van Erk, Matrix correlations for high-dimensional data: the modified RV-coefficient, Bioinformatics 25 (2009) 401–405.

[13] M.S. Srivastava, N. Reid, Testing the structure of the covariance matrix with fewer observations than the dimension, J. Multivariate Anal. 112 (2012) 156–171.

[14] G.J. Székely, M.L. Rizzo, Brownian distance covariance, Ann. Appl. Statist. 3 (2009) 1236–1265.

[15] G.J. Székely, M.L. Rizzo, The distance correlation t-test of independence in high dimension, J. Multivariate Anal. 117 (2013) 193–213.

[16] A. Wille, P. Zimmermann, E. Vranova, A. Fürholz, O. Laule, S. Bleuler, L. Hennig, A. Prelic, P. von Rohr, L. Thiele, E. Zitzler, W. Gruissem, P. Bühlmann, Sparse graphical Gaussian modeling of the isoprenoid gene network in Arabidopsis thaliana, Genome Biol. 5 (2004) R92.

[17] Y. Yang, G. Pan, Independence test for high dimensional data based on regularized canonical correlation coefficients, Ann. Statist. 43 (2015) 467–500.

[18] K. Yata, M. Aoshima, Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix, J. Multivariate Anal. 101 (2010) 2060–2077.

[19] K. Yata, M. Aoshima, Correlation tests for high-dimensional data using extended cross-data-matrix methodology, J. Multivariate Anal. 117 (2013) 313–331.

[20] P.S. Zhong, S.X. Chen, Tests for high-dimensional regression coefficients with factorial designs, J. Amer. Statist. Assoc. 106 (2011) 260–274.