应用概率统计　第 33 卷
第 4 期　2017 年 8 月

Chinese Journal of Applied Probability and Statistics
Aug., 2017, Vol. 33, No. 4, pp. 349-368
doi: 10.3969/j.issn.1001-4268.2017.04.003

Composite Hotelling's T-Square Test for High-Dimensional Data *

LI Ting

(*Department of Statistics, School of Management, Fudan University, Shanghai, 200433, China*)

Abstract: The paper is concerned with the two-sample mean testing problem in high-dimension settings. We propose a composite Hotelling's T-square test, establish its asymptotical normality and study its local power. The finite-sample priority of the proposed test over existing high-dimensional tests is shown by simulations and illustrated by a real data-set analysis.

Keywords: high dimension; Hotelling's T-square test; asymptotical normality; local power

2010 Mathematics Subject Classification: 62H15

Citation: Li T. Composite Hotelling's T-square test for high-dimensional data [J]. Chinese J. Appl. Probab. Statist., 2017, 33(4): 349–368.

§1. Introduction

The paper is concerned with the two-sample mean testing problem in high-dimension settings. An easy example of this problem arises from the case-control study if too many features are measured in both the case and control groups. Let $\{\boldsymbol{X}_{ij} : j = 1, 2, \ldots, n_i\}$ $(i = 1, 2)$ be two p-dimensional simple random samples with mean $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip})^\top$ and covariance matrix $\boldsymbol{\Sigma}_i$. The problem of interest is to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \longleftrightarrow H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \tag{1}$$

in the high-dimensional setting, where both the dimension $p = p_n$ and the total sample size $n = n_1 + n_2$ increase to infinity. Let $\overline{\boldsymbol{X}}_i$ and $\boldsymbol{S}_i$ be the sample mean and sample variance for the ith sample $(i = 1, 2)$ and define $\boldsymbol{S}_* = \{(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2\}/(n - 2)$. The standard approach to this problem in low dimensional setting is Hotelling's T-square test [1] $\mathrm{HT} = (n_1 n_2/n)(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)^\top \boldsymbol{S}_*^{-1}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)$. Although owning many nice properties, it does not work in the high-dimensional setting because the sample variance is not invertible when $p \geqslant n$.

*The research was supported in part by the National Natural Science Foundation of China (Grant No. 11371142).
Received September 10, 2015. Revised April 12, 2016.

In recent years, many remedies to Hotelling's T-square test have been proposed to adapt to the high dimensionality. Bai and Saranadasa [2], Srivastava and Du [3] and Srivastava [4] constructed test statistics by removing $\boldsymbol{S}_*^{-1}$ in HT or replacing it with the inverse of the diagonal of $\boldsymbol{S}_*$. Chen and Qin [5] proposed a new test by removing the cross-product terms $\sum\limits_{j=1}^{n_i} \boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^{\top}$ $(i=1,2)$ in [2]'s test. Alternative to the above sum-of-square type tests, Cai et al. [6] introduced a max type test statistic, which is defined to be the maximum squared length of certain projections of the mean difference $\overline{\boldsymbol{X}}_1-\overline{\boldsymbol{X}}_2$ on the p coordinates. An estimator of the covariance matrix is required in this test.

A natural property for a desirable high-dimensional test is component-wise scale invariance. That is, it should be invariant under the transformation $\boldsymbol{X}\mapsto\boldsymbol{B}\boldsymbol{X}$ with $\boldsymbol{B}$ being a invertible diagonal matrix. Unfortunately, except [3] and [4]'s tests, none of the existing tests own the component-wise scale-invariance. For these tests, a different conclusion would be achieved if the underlying data-set is re-scaled in some of its dimensions. Although [3] and [4]'s tests are component-wise invariant, they ignore the component dependency information in data. We may expect to have possible power gain if such information is taken into account appropriately.

We propose a composite Hotelling's T-square test (CHT) for (1) by integrating the Hotelling's T-square tests for all bivariate subvectors of the mean vectors. Its asymptotical normality is also established. The CHT not only is component-wise scale invariant, but also takes between-component correlation into account, therefore it is expected to have better testing power than existing high-dimensional tests especially when the between-component correlation in data is big. This point is confirmed by our local power analysis and finite-sample simulation study.

We end the introduction section by introducing necessary notation. Throughout the paper, we assume for convenience that there exists two constants $\eta_i\in(0,1)$ $(i=1,2)$ such that $n_i/n=\eta_i$ as $n\to\infty$, although it is more natural to assume that $n_i/n\to\eta_i$. For a vector $\boldsymbol{X}_{ij}$, let $X_{ij,k}$ be its kth component and let $\boldsymbol{X}_{ij(kl)}$ denote a bivariate vector consisting of the kth and lth components of $\boldsymbol{X}_{ij}$. Let $\boldsymbol{S}\equiv(s_{kl})_{p\times p}=\boldsymbol{S}_1/\eta_1+\boldsymbol{S}_2/\eta_2$ with $\boldsymbol{S}_i=(s_{i,kl})_{p\times p}$ $(i=1,2)$. We use $\boldsymbol{S}_{(kl)}$ to denote a 2×2 submatrix of $\boldsymbol{S}$ consisting of s_{kk}, s_{kl}, s_{lk} and s_{ll}. Let $\boldsymbol{\Sigma}\equiv(\sigma_{ij})=\boldsymbol{\Sigma}_1/\eta_1+\boldsymbol{\Sigma}_2/\eta_2$. The symbols $\boldsymbol{S}_{i(kl)}$, $\boldsymbol{\Sigma}_{(kl)}$, $\sigma_{i,kl}$ and $\boldsymbol{\Sigma}_{i(kl)}$ can be defined similarly. For clarity, we postpone all proofs to the Appendix.

§2. Composite Hotelling's T-Square Test

Testing hypothesis (1) is equivalent to simultaneously testing $H_{0(kl)}:\boldsymbol{\mu}_{1(kl)}=\boldsymbol{\mu}_{2(kl)}$ for all pairs (k,l) such that $1\leqslant k\neq l\leqslant p$. If for each (k,l), a desirable test can be

constructed for testing $H_{0(kl)}$, then pooling all these tests together gives a desirable test for hypothesis (1).

The commonly-used test for $H_{0(kl)}$ is the two-sample Hotelling's T-square test, which, up to a scale n_1n_2/n^2, is

$$\begin{aligned} & n(\overline{\boldsymbol{X}}_{1(kl)} - \overline{\boldsymbol{X}}_{2(kl)})^{\mathsf{T}}\{\boldsymbol{S}_{(kl)}\}^{-1}(\overline{\boldsymbol{X}}_{1(kl)} - \overline{\boldsymbol{X}}_{2(kl)}) \\ = & \frac{n}{n_1^2n_2^2}\sum_{r_1,r_2}\sum_{s_1,s_2}(\boldsymbol{X}_{1r_1(kl)} - \boldsymbol{X}_{2s_1(kl)})^{\mathsf{T}}\{\boldsymbol{S}_{(kl)}\}^{-1}(\boldsymbol{X}_{1r_2(kl)} - \boldsymbol{X}_{2s_2(kl)}). \end{aligned} \tag{2}$$

Here and in what follows, the subscripts r_1, r_2 run from 1 to n_1 and s_1, s_2 from 1 to n_2. Similar to the justification of [5] for their test, we find that the terms with $r_1 = r_2$ or $s_1 = s_2$ in (2) are not useful for testing hypothesis $H_{0(kl)}$. Naturally, we choose the following modified Hotelling's T-square test to serve this purpose,

$$T_{kl} = \sum_{r_1\neq r_2}\sum_{s_1\neq s_2}\frac{n(\boldsymbol{X}_{1r_1(kl)} - \boldsymbol{X}_{2s_1(kl)})^{\mathsf{T}}\{\boldsymbol{S}_{(kl)}\}^{-1}(\boldsymbol{X}_{1r_2(kl)} - \boldsymbol{X}_{2s_2(kl)})}{n_1(n_1-1)n_2(n_2-1)}.$$

Adding up T_{kl}'s for all different pairs (k,l) leads to the proposed composite Hotelling's T-square (CHT) test statistic, $T = \sum_{k\neq l} T_{kl}$.

The CHT test inherits at least two nice properties from the traditional Hotelling's T-square test. On one hand, it captures the information on correlation between any two dimensions of the data since each T_{kl} sufficiently incorporates this information. Intuitively, this may make the CHT test to have good power. On the other hand, it is invariant under a component-wise scale-transformation, i.e., the CHT test keeps unchanged when the data is recorded under different measurement units.

Roughly speaking, $\boldsymbol{S}_{(kl)} \approx \boldsymbol{\Sigma}_{(kl)}$ as n is large. Therefore $T_{kl} \approx T^*_{kl}$ where

$$T^*_{kl} = \sum_{r_1\neq r_2}\sum_{s_1\neq s_2}\frac{n(\boldsymbol{X}_{1r_1(kl)} - \boldsymbol{X}_{2s_1(kl)})^{\mathsf{T}}\{\boldsymbol{\Sigma}_{(kl)}\}^{-1}(\boldsymbol{X}_{1r_2(kl)} - \boldsymbol{X}_{2s_2(kl)})}{n_1(n_1-1)n_2(n_2-1)}.$$

Accordingly T can be approximated by $T^* = \sum_{k\neq l} T^*_{kl}$, whose limiting distribution is easier to derive than that of T itself. Since $\boldsymbol{X}_{1r_1}$'s and $\boldsymbol{X}_{2s_1}$'s are both independent and identically distributed random vector series and their means are $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, respectively. It can be shown that $\mathsf{E}(T^*) = 2n(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathsf{T}}\boldsymbol{A}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, where $\boldsymbol{A}$ is a positive definite matrix defined in Lemma 2. This implies that the null hypothesis H_0 in (1) is equivalent to $\mathsf{E}(T^*) = 0$. If H_0 does not holds, then $\mathsf{E}(T^*)$ tends to be large, so do T^* and T. Therefore our testing rule is to reject H_0 if T is too large. The critical values can be determined through the limiting distribution of T, presented in the next section.

§3. Asymptotics of the CHT Test

The limiting distribution of T is obtained through two steps. First, we shall derive the limiting distribution of T^*. Then we show that the difference between T and T^* is negligible compared with the standard deviation of T^*. These two assertions immediately implies that T and T^* have the same limiting distribution after standardization.

Following [2] and [5], we make the following assumption on data.

Assumption 1 The data come from $\boldsymbol{X}_{ij} = \boldsymbol{\Gamma}_i \boldsymbol{Z}_{ij} + \boldsymbol{\mu}_i$, where $\boldsymbol{\Gamma}_i = (\Gamma_{i,kl})$ is a $p \times m$ matrix, $\boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_i^{\mathsf{T}} = \boldsymbol{\Sigma}_i$, and $\boldsymbol{Z}_{ij} = (Z_{ij1}, Z_{ij2}, \ldots, Z_{ijm})^{\mathsf{T}}$'s are independent and identically distributed (i.i.d.) random vectors such that $\mathsf{E}(\boldsymbol{Z}_{ij}) = \boldsymbol{0}$ and $\mathsf{Var}(\boldsymbol{Z}_{ij}) = \boldsymbol{I}_m$, an $m \times m$ identity matrix. In addition, m is required to be larger than p and there exists an positive integer κ such that $\mathsf{E}(Z_{ijl}^{4\kappa}) \leqslant K_0 < \infty$, $\mathsf{E}(Z_{ijl_1}^{\alpha_1} Z_{ijl_2}^{\alpha_2} \cdots Z_{ijl_q}^{\alpha_q}) = \mathsf{E}(Z_{ijl_1}^{\alpha_1})\mathsf{E}(Z_{ijl_2}^{\alpha_2}) \cdots \mathsf{E}(Z_{ijl_q}^{\alpha_q})$, whenever $\sum\limits_{l=1}^{q} \alpha_l \leqslant 4\kappa$ and $l_1 \neq l_2 \neq \cdots \neq l_q$.

The i.i.d. assumption on $\boldsymbol{Z}_{ij}$'s implies that $\boldsymbol{X}_{ij}$ $(j = 1, 2, \ldots, n_i)$ are also i.i.d. observations for $i = 1$ and 2, respectively. The fact that m is arbitrary offers much flexibility in generating a rich collection of dependence structure [5]. If the observations come from two p-dimensional normal distributions, Assumption 1 is clearly satisfied. Therefore this assumption can be regarded as an extension of the normality assumption.

3.1 Limiting Distribution of T^*

We write T^* in a more compact form which will facilitate subsequent presentation.

Lemma 2 The T^* defined in Section 2 can be rewritten as

$$T^* = \frac{2n}{n_1(n_1-1)n_2(n_2-1)} \sum_{r_1 \neq r_2} \sum_{s_1 \neq s_2} (\boldsymbol{X}_{1r_1} - \boldsymbol{X}_{2s_1})^{\mathsf{T}} \boldsymbol{A} (\boldsymbol{X}_{1r_2} - \boldsymbol{X}_{2s_2}), \tag{3}$$

where $\boldsymbol{A} = (a_{kl})_{p \times p}$ with $a_{kk} = \sum\limits_{j \neq k} \sigma_{jj}/(\sigma_{kk}\sigma_{jj} - \sigma_{kj}^2)$ and $a_{kl} = -\sigma_{kl}/(\sigma_{kk}\sigma_{ll} - \sigma_{kl}^2)$.

The limiting distribution of T^* hinges on the expectation and variance of T^*, which are given in the following lemma.

Lemma 3 Under Assumption 1, we have $\mathsf{E}(T^*) = 2n(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathsf{T}} \boldsymbol{A} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ and $\mathsf{Var}(T^*) = 16n(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathsf{T}} \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + 8\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\}\{1 + o(1)\}$.

With the above preparations, we present the limiting distribution of T^*.

Theorem 4 Under Assumption 1, as n and p tend to infinity, if $n_i/n = \eta_i \in (0, 1)$ $(i = 1, 2)$ and $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathsf{T}} \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o(n^{-1}\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\})$, then

$$\{T^* - 2n(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathsf{T}} \boldsymbol{A} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\}/\sqrt{8\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\}} \xrightarrow{\mathrm{d}} \mathrm{N}(0, 1).$$

The assumption $n_i/n = \eta_i \in (0,1)$ is imposed only for convenience purpose. It is readily to be replaced by "$n_i/n \to \eta_i \in (0,1)$".

3.2 Magnitude of the Difference $T - T^*$

We need more conditions than to guarantee that the difference $T - T^*$ is negligible compared with the standard deviation of T^*. Let $\rho_{i,kl} = \sigma_{i,kl}/\sqrt{\sigma_{i,kk}\sigma_{i,ll}}$.

Lemma 5 Suppose Assumptions 1 holds and that there exists $\varepsilon_0 \in (0,1)$ such that $\max\limits_{i,k,l} |\rho_{i,kl}| \leqslant 1 - \varepsilon_0$ uniformly for all n. If $p^4 = o(n\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\})$ and $n\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2/p = O(1)$, then $(T - T^*)/\sqrt{\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\}} = o_p(1)$.

According to Lemma **??** in the Appendix, a sufficient and necessary condition for $p^4 = o(n\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\})$ is $p^2 = o(n\mathrm{tr}(\boldsymbol{R}^2))$, where $\boldsymbol{R} \equiv (\rho_{kl})_{p\times p} = \boldsymbol{D}_\sigma^{-1/2}\boldsymbol{\Sigma}\boldsymbol{D}_\sigma^{-1/2}$ with $\boldsymbol{D}_\sigma = \mathrm{diag}\{\sigma_{11}, \sigma_{22}, \ldots, \sigma_{pp}\}$. This condition characterises how the between-component correlation in data affects the data dimension that the proposed CHT test can handle. On one hand, if $\boldsymbol{\Sigma}$ is a diagonal matrix or there is not any correlation between any pair of dimensions of the data, then these two conditions are fulfilled if and only if $p = o(n)$. We do not recommend the CHT test in this situation because the correlation information it incorporates becomes noise, which downplays its testing capability. On the other hand, if the correlation coefficient of any two dimensions is a constant $\rho \in (0,1)$ then $\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\} = p^4\rho^2(1+\rho)^{-2}\{1 + o(1)\}$ (see the example in the next subsection), which means that $p^4 = o(n\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\})$ is always true for any p and any fixed $\rho \in (0,1)$ in this example.

3.3 Limiting Distribution of T

The asymptotical normality of T is a direct corollary of Theorem 4 and Lemma 5.

Theorem 6 If the conditions in Theorem 4 and Lemma 5 are all fulfilled, then

$$\{T - 2n(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathsf{T}}\boldsymbol{A}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\}/\sqrt{8\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1).$$

Although Theorem 6 shows in theory the limiting distribution of the proposed CHT test statistic, it is still not ready for practical use because the asymptotical variance $\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\}$ in Theorem 6 is still unknown and needs to be estimated appropriately.

Lemma 7 Assume the conditions in Theorem 6. If $p^2 = o(\sqrt{n}\,\mathrm{tr}(\boldsymbol{R}^2))$, then

$$\mathrm{tr}\{(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{A}})^2\}/\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\} = 1 + o_p(1),$$

where $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{A}}$ are the respective moment estimators of $\boldsymbol{\Sigma}$ and $\boldsymbol{A}$.

The condition $p^2 = o(\sqrt{n}\,\mathrm{tr}(\boldsymbol{R}^2))$ excludes the cases with no or too weak between-component correlation in data. According to the discussion after Lemma 5, when there is no or too weak between-component correlation, the correlation information incorporated in $\mathrm{tr}\{(\boldsymbol{\Sigma A})^2\}$ is more to be noise than to be signal. We exclude these cases so that the signal in $\mathrm{tr}\{(\boldsymbol{\Sigma A})^2\}$ is big enough.

Theorem 8 If the conditions in Theorem 6 and Lemma 7 are all fulfilled, then

$$\{T - 2n(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathsf{T}}\boldsymbol{A}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\}/\sqrt{8\mathrm{tr}\{(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{A}})^2\}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1).$$

Theorem 8 implies that if $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ is not true, then the proposed test statistic T tends to be infinity. Naturally, we propose to reject H_0 at α significance level if $T/\sqrt{8\mathrm{tr}\{(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{A}})^2\}} > \xi_{1-\alpha}$, the $1-\alpha$ quantile of the standard normal distribution.

The proposed CHT test also applies to the one-sample mean testing problem, which can be regarded as a degenerate two-sample mean problem. Suppose the data are $X_{11}, X_{12}, \ldots, X_{1n_1}$ with mean $\boldsymbol{\mu}_1$ and we wish to test $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ for a given value $\boldsymbol{\mu}_2$. In this case, $\boldsymbol{S} = \boldsymbol{S}_1$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1$, and the CHT test reduces to

$$T = \sum_{k\neq l} T_{kl} = \sum_{k\neq l}\sum_{r_1\neq r_2} \frac{n(\boldsymbol{X}_{1r_1(kl)} - \boldsymbol{\mu}_{2(kl)})^{\mathsf{T}}\{\boldsymbol{S}_{1(kl)}\}^{-1}(\boldsymbol{X}_{1r_2(kl)} - \boldsymbol{\mu}_{2(kl)})}{n_1(n_1-1)}, \tag{4}$$

and Theorem 8 still holds.

3.4 Local Power

Theorem 8 indicates that the proposed CHT test has an asymptotic normal distribution under the null hypotheses H_0. It also facilitates us to derive the asymptotic power of the CHT test under a local alternative,

$$\beta_{\mathrm{CHT}}(\delta) \approx \Phi\big(-\xi_\alpha + n\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\delta}/\sqrt{2\mathrm{tr}\{(\boldsymbol{\Sigma A})^2\}}\,\big),$$

where $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ and $\Phi(\cdot)$ is the standard normal distribution function.

For comparison, we review the asymptotical powers of several existing competitors of the CHT test. Under the same heterogeneity assumption on variance, Chen and Qin [5] (CQ for short) derived the local power of their s test, i.e.,

$$\beta_{\mathrm{CQ}}(\boldsymbol{\delta}) \approx \Phi\big(-\xi_\alpha + n\|\boldsymbol{\delta}\|^2/\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}\,\big).$$

When variance homogeneity is assumed, say $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_c = (\sigma_{c,ij})$, Bai and Saranadasa [2] (BS) and Srivastava and Du [3] (SD) found that the asymptotical powers of their tests are

$$\beta_{\mathrm{BS}}(\boldsymbol{\delta}) \approx \Phi\big(-\xi_\alpha + n\eta_1\eta_2\|\boldsymbol{\delta}\|^2/\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}_c^2)}\,\big),$$

$$\beta_{\mathrm{SD}}(\boldsymbol{\delta}) \approx \Phi\big(-\xi_\alpha + n\eta_1\eta_2\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{D}_\sigma^{-1}\boldsymbol{\delta}/\sqrt{2\mathrm{tr}(\boldsymbol{R}_c^2)}\,\big),$$

where $\boldsymbol{D}_\sigma = \mathrm{diag}(\sigma_{c,11}, \sigma_{c,22}, \ldots, \sigma_{c,pp})$ and $\boldsymbol{R}_c = (\rho_{c,kl})$ with $\rho_{c,kl} = \sigma_{c,kl}/\sqrt{\sigma_{c,kk}\sigma_{c,ll}}$.

Given that direct comparison on the above powers is intractable, we consider a special case: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_c = (\sigma_{c,kl})$ with $\sigma_{c,kk} = 1$, $\sigma_{c,kl} = \rho \in (0,1)$ for $k \neq l$. Because the two variances are equal and the diagonal elements in Σ_c are all one, it follows that $\beta_{\mathrm{CQ}}(\boldsymbol{\delta}) \approx \beta_{\mathrm{SD}}(\boldsymbol{\delta}) \approx \beta_{\mathrm{BS}}(\boldsymbol{\delta})$. Therefore the asymptotical relative efficiency (ARE) of the CHT test over the CQ, SD and BS tests are all

$$\mathrm{ARE} = \frac{\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\delta}}{\eta_1\eta_2\|\boldsymbol{\delta}\|^2} \cdot \sqrt{\frac{\mathrm{tr}(\boldsymbol{\Sigma}_c^2)}{\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\}}}.$$

Since $\mathrm{tr}(\boldsymbol{\Sigma}_c^2) = p^2\rho^2\{1+o(1)\}$, $\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\} = p^4\rho^2(1+\rho)^{-2}\{1+o(1)\}$ and

$$\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\delta} = \eta_1\eta_2\{(p-1+\rho)\|\boldsymbol{\delta}\|^2 - \rho(\boldsymbol{\delta}^{\mathsf{T}}\mathbf{1}_p)^2\}/(1-\rho^2),$$

where $\mathbf{1}_p = (1,1,\ldots,1)^{\mathsf{T}}$ is a p-variate vector, we have for large p,

$$\mathrm{ARE} = \frac{(p-1+\rho)\|\boldsymbol{\delta}\|^2 - \rho(\boldsymbol{\delta}^{\mathsf{T}}\mathbf{1}_p)^2}{(1-\rho^2)\|\boldsymbol{\delta}\|^2}\frac{1+\rho}{p}\{1+o(1)\} = \frac{1-\rho t}{1-\rho}\{1+o(1)\},$$

where $t = (\boldsymbol{\delta}^{\mathsf{T}}\mathbf{1}_p)^2/(p\|\boldsymbol{\delta}\|^2)$. By Cauchy-Schwarz inequality, we find $(\boldsymbol{\delta}^{\mathsf{T}}\mathbf{1}_p)^2 \leqslant p\|\boldsymbol{\delta}\|^2$ for any $\boldsymbol{\delta} \neq \mathbf{0}$, which implies $t \in [0,1]$. Because $\rho \geqslant 0$ in this example, we conclude that (i) ARE $\geqslant 1$, that is, the CHT test is asymptotically more powerful than the rest three tests excluding the degenerate and trivial cases $\rho = 0$ or $\boldsymbol{\delta}^{\mathsf{T}}\mathbf{1}_p = 0$; and (ii) the advantage of the CHT test increases as the correlation coefficient ρ increases.

§4. Simulation and a Real Application

We report a limited simulation study and a real application to investigate the finite-sample performance of the proposed CHT test. We compare it with the CQ, BS and SD tests as mentioned before. When $p < n-2$, Hotelling's T-square test is also considered. Throughout the simulation study, all numbers reported are obtained based on 2 000 random samples and the significance level is 5%.

4.1 A Simulation Study

Let $\boldsymbol{X}_{ij} = (X_{ij,1}, X_{ij,2}, \ldots, X_{ij,p})^{\mathsf{T}}$ denote the jth observation in the ith sample ($i = 1,2$; $j = 1,2,\ldots,n_i$). We generate $X_{ij,k}$'s from the following moving average model [5]:

$$X_{ij,k} = Z_{ij,k} + Z_{ij,k+1} + \cdots + Z_{ij,k+p-1} + \mu_{ij}, \qquad k = 1,2,\ldots,p,$$

where $Z_{ij,k}$'s are i.i.d. from the centralized Gamma$(4,1)$ distribution in the case of equal variance. In the case of unequal variance, $Z_{2j,k}$'s are replaced with i.i.d. observations from the centralized Gamma$(3,1)$ distribution. We fix $\boldsymbol{\mu}_1 = 0$ and choose $\boldsymbol{\mu}_2$ in the same way as [7]. The percentage of true null hypotheses is set to be 0%, 25%, 50%, 75%, 95% and 100%. For each percentage level of true nulls, three patterns of allocations as specified in [7] are considered: (i) the equal allocation where all the nonzero $\mu_{2,l}$ are equal; (ii) linearly decreasing allocations and; (iii) linearly increasing allocations. In addition, $\boldsymbol{\mu}_2$ is subject to $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|/\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)} = 0.1$ throughout this simulation. We choose $p = 500$ and $n_1 = n_2 = 158$.

Table 1 Powers and sizes in percentage under the moving average model with $p = 500$ and $n = 158$

Allocation	% of true null	Equal variance				Unequal variance			
		CHT	CQ	BS	SD	CHT	CQ	BS	SD
Equal	0	37.16	34.32	34.64	4.32	41.42	37.35	37.38	5.78
	25	42.82	33.64	33.96	3.84	48.03	37.47	37.45	4.45
	50	52.40	33.30	33.62	2.98	58.97	39.33	39.36	3.38
	75	73.88	34.08	34.32	1.48	82.74	40.09	40.01	2.89
	95	100.00	30.82	31.24	0.88	100.00	42.13	42.17	1.03
	100	7.06	6.58	6.68	0.28	7.11	7.02	7.03	0.42
Increasing	0	34.36	34.34	34.56	5.42	40.03	38.72	38.74	6.02
	25	42.10	34.88	35.12	4.42	44.68	36.34	36.32	4.94
	50	50.14	33.58	33.96	2.90	57.45	39.38	39.32	3.10
	75	72.04	34.40	34.76	1.98	79.14	37.68	37.60	2.08
	95	99.86	27.44	28.00	0.94	100.00	34.38	34.28	0.82
	100	7.64	7.19	7.29	0.18	8.01	7.88	07.85	0.38
Decreasing	0	38.74	34.54	34.84	4.08	41.13	36.56	36.54	6.13
	25	44.66	32.72	33.00	3.70	52.36	37.80	37.82	3.88
	50	55.28	34.52	34.82	3.00	61.13	40.14	40.12	3.75
	75	73.32	30.50	31.04	1.78	81.10	36.14	36.00	2.24
	95	99.94	24.88	25.28	0.70	100.00	30.46	30.38	0.94
	100	7.30	6.75	6.87	0.28	6.65	6.74	06.75	0.31

Table 1 tabulates our simulated sizes and powers under the moving average model. The results in the cases of equal variance and unequal variance are very similar to each other. When the percentage of true nulls is 100%, the reported numbers are type I errors. We observe that all type I errors are acceptable and close to 5% although those of the CHT test are generally slightly larger. In view of power comparison, the CHT test is always

more powerful than the rest three tests. When none of the hypotheses is true, the CHT test has comparable and slightly larger powers than the CQ and BS tests. As the percentage of true null increases, the power of the CHT test becomes larger and larger, however those of the CQ and BS tests remain nearly unchanged. This observation coincides with the local power of CQ and BS tests, which are increasing functions of $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|/\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}$. When 95% of the hypotheses is true, the powers of the CHT test reaches 100%. Recall that we fixed $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|/\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)} = 0.1$. This implies that the CHT test has clear advantages over the CQ and BS tests if the signal is sparse but large.

The above simulation results show the nice testing performance of the CHT test when the dimension p is much larger than the sample size n, in which Hotelling's T-square test is not applicable. Naturally we would ask, how does it perform when the the dimension p is less than but close to the sample size n? In this case, Hotelling's T-square test (T^2 for short) is applicable and can be taken as a benchmark for comparison. To this end, we generate data from the same moving average model under two settings: (i) $n_1 = n_2 = 200$, $p = 100$ and (ii) $n_1 = n_2 = 100$, $p = 190$. Throughout this simulation, we choose $\boldsymbol{\mu}_2$ such that $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|/\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)} = 0.01$.

The simulation results are reported in Table 2. We have the same findings as in Table 1 about the type I error and power comparisons of the CHT test and both the CQ and BS tests. We turn our attention to the comparison of Hotelling's T-square test and the above three tests. The CHT, CQ and BS tests have similar and slightly inflated type I errors, while that of Hotelling's T-square test is almost equal to the nominal, indicating that it has a better control on type I error. Under setting (i) where p is much smaller than $n_1 + n_2$, Hotelling's T-square test works well and is clearly the most powerful among all tests. In comparison, the CHT test has very close performance and it loses less power than the CQ and BS tests, whose powers keep almost unchanged. As p approaches $n_1 + n_2$ as in setting (ii), sample covariance matrix $\boldsymbol{S}$ becomes ill conditioned. Hotelling's T-square test loses power and superiority, and is uniformly outperformed by the CHT test.

4.2 Real Data Analysis

We illustrate the usefulness of the CHT test by analyzing an amino acid data set, which is available upon request. The data set consists of hydrophobicity measurements of 114 family GH11 xylanase amino acid sequences ($n_1 = 23$ basophilic and $n_2 = 91$ acidophilous), with 202 charged residues ($p = 202$) in each sequence. The problem of interest is to test whether the two groups of amino acid sequences have different characteristics at these 202 sites. The results are shown in Table 3. All the four tests lead to the conclu-

Table 2 Powers and sizes in percentage under moving average model with p close to n

Allocation	% of true null	$p=100$, $n_1=n_2=200$					$p=190$, $n_1=n_2=100$				
		CHT	CQ	BS	SD	T^2	CHT	CQ	BS	SD	T^2
Equal	0	42.12	39.36	39.50	13.22	59.00	67.12	62.62	62.60	24.52	21.80
	25	48.76	37.52	37.82	10.84	90.42	79.08	66.14	66.14	23.24	35.78
	50	54.70	35.42	35.64	8.82	99.08	89.94	70.36	70.36	22.20	54.74
	75	85.84	43.12	43.42	8.20	100.00	99.64	76.68	76.66	16.96	88.46
	95	100.00	76.10	76.80	9.34	100.00	100.00	100.00	100.00	15.72	100.00
	100	7.37	7.32	7.43	0.86	4.90	6.53	6.19	6.17	0.45	4.94
Increasing	0	39.44	38.84	38.94	14.04	42.16	65.48	64.16	64.18	27.22	15.62
	25	46.14	38.38	38.56	12.24	75.18	75.00	64.38	64.36	23.90	25.84
	50	58.14	38.28	38.64	10.34	94.60	88.88	69.54	69.54	22.56	40.98
	75	75.38	34.78	35.20	6.58	100.00	99.30	79.10	79.12	19.70	77.38
	95	99.30	23.40	23.64	2.92	100.00	100.00	82.66	82.60	7.84	100.00
	100	7.20	7.01	7.10	0.90	5.09	7.71	7.34	7.35	0.71	4.87
Decreasing	0	43.74	38.28	38.46	12.06	84.48	72.92	64.10	64.10	25.36	29.72
	25	49.30	36.46	36.62	10.38	99.06	83.40	66.88	66.86	24.50	55.18
	50	63.18	39.32	39.66	10.36	100.00	92.38	69.16	69.12	20.52	75.52
	75	75.20	32.90	33.60	5.74	100.00	99.88	81.28	81.30	17.62	98.44
	95	86.98	16.74	17.06	2.20	100.00	100.00	59.10	59.14	5.00	100.00
	100	6.89	6.43	6.49	0.79	4.90	7.27	7.06	7.06	0.72	5.23

sion that there is difference in hydrophobicity between the two groups. And clearly the proposed CHT test provides the strongest evidence for the difference of the two groups.

Table 3 Testing results for the amino acid data set

	CHT	CQ	BS	SD
Statistic	7.3482	4.5109	6.1661	4.7529
p-value	1.0048e-13	3.2269e-06	3.4998e-10	1.0027e-06

Acknowledgments The author would like to thank the support and guide of Dr. Yukun Liu at East China Normal University. The author also would like to thank the editor, associate editor and referees for their valuable comments.

Appendix A: Proofs

Let $\boldsymbol{D} = \mathrm{diag}\{\sigma_{11}^{-1/2}, \sigma_{22}^{-1/2}, \ldots, \sigma_{pp}^{-1/2}\}$ where σ_{jj} are the diagonal elements of Σ. It is clear that the CHT based on $\boldsymbol{X}_{ij}$'s is equal to that based on $\boldsymbol{Y}_{ij} = \boldsymbol{D}\boldsymbol{X}_{ij}$'s. A nice

property of $\boldsymbol{Y}_{ij}$'s is that their variances and covariances are uniformly bounded, because

$$|\mathsf{Cov}\,(Y_{ij,k}, Y_{ij,l})| = \frac{|\sigma_{i,kl}|}{\sqrt{\sigma_{kk}\sigma_{ll}}} \leqslant \frac{\eta_i|\sigma_{i,kl}|}{\sqrt{\sigma_{i,kk}\sigma_{i,ll}}} = \eta_i|\rho_{i,kl}| \leqslant \eta_i \leqslant 1, \tag{5}$$

where $\rho_{i,kl} = \sigma_{i,kl}/\sqrt{\sigma_{i,kk}\sigma_{i,ll}}$ are the correlation coefficients of the ith population. Henceforth we shall take $\boldsymbol{X}_{ij}$'s as $\boldsymbol{Y}_{ij}$'s for convenience. The quantities T_{kl}, T, T^*_{kl} and T^* all keep unchanged.

Under Assumption 1, it follows from $\boldsymbol{X}_{ij} = \boldsymbol{\Gamma}_i\boldsymbol{Z}_{ij}$ with $\boldsymbol{\Gamma}_i = (\Gamma_{i,ju})_{p\times m}$ that $X_{ij,r} = \sum\limits_{u=1}^{m} \Gamma_{i,ju}Z_{iu,r}$. This together with fact $\mathsf{Var}\,(X_{ij,r}) \leqslant \eta_i$ implies $\sum\limits_{u} \Gamma^2_{i,ju} \leqslant \eta_i$, and

$$\mathsf{E}(X^4_{ij,r}) = \sum_{u=1}^{m} \Gamma^4_{i,ju}\mathsf{E}(Z^4_{iu,r}) + 3\sum_{1\leqslant u_1\neq u_2\leqslant m} \Gamma^2_{i,ju_1}\Gamma^2_{i,ju_2} \leqslant K_0\eta_i + 3\eta_i^2 \leqslant K_0 + 3.$$

We further conclude that s_{kl} converges in probability to σ_{kl} uniformly. See Lemma 9. Thus under Assumption 1, all $\boldsymbol{S}_{(kl)}$ are close enough to $\boldsymbol{\Sigma}_{(kl)}$ as n is large. Keep in mind that the $\boldsymbol{\Sigma}_{(kl)}$ and $\boldsymbol{S}_{(kl)}$ are calculated based on $\boldsymbol{Y}_{ij}$'s.

Proof of Lemma 2 It can be found that

$$T^* = \sum_{r_1\neq r_2}\sum_{s_1\neq s_2} \frac{n}{n_1(n_1-1)n_2(n_2-1)} U(\boldsymbol{X}_{1r_1} - \boldsymbol{X}_{2s_1}, \boldsymbol{X}_{1r_2}\boldsymbol{X}_{2s_2}), \tag{6}$$

where $U(\boldsymbol{X}, \boldsymbol{Y}) = \sum\limits_{k\neq l} \boldsymbol{X}^{\mathsf{T}}_{(kl)}\{\boldsymbol{\Sigma}_{(kl)}\}^{-1}\boldsymbol{Y}_{(kl)}$. If $\boldsymbol{\Sigma} = (\sigma_{kl})_{p\times p}$, then

$$(\boldsymbol{\Sigma}_{(kl)})^{-1} = \begin{pmatrix} \sigma_{kk} & \sigma_{kl} \\ \sigma_{kl} & \sigma_{ll} \end{pmatrix}^{-1} = \frac{1}{\sigma_{kk}\sigma_{ll} - \sigma^2_{kl}} \begin{pmatrix} \sigma_{ll} & -\sigma_{kl} \\ -\sigma_{kl} & \sigma_{kk} \end{pmatrix}.$$

Let $X_{(k)}$ denote the kth component of $\boldsymbol{X}$. Since $\sigma_{kl} = \sigma_{lk}$, it follows that

$$U(\boldsymbol{X}, \boldsymbol{Y}) = 2\sum_{k\neq l} \frac{X_{(k)}Y_{(k)}\sigma_{ll} - X_{(k)}Y_{(l)}\sigma_{kl}}{\sigma_{kk}\sigma_{ll} - \sigma^2_{kl}} = 2\sum_{k,l} X_{(k)}Y_{(l)}a_{kl} = 2\boldsymbol{X}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{Y}.$$

This together with (6) implies (3). □

Proof of Lemma 3 Since $\boldsymbol{X}_{1i}$'s are independent and identically distributed (i.i.d.) with mean $\boldsymbol{\mu}_1$, $\boldsymbol{X}_{2i}$'s i.i.d. with mean $\boldsymbol{\mu}_2$ and they are independent of each other, it can be verified that $\mathsf{E}(T^*) = 2n(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathsf{T}}\boldsymbol{A}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$.

We now compute the variance of T^*. Let $\boldsymbol{Y}_{ij} = \boldsymbol{X}_{ij} - \boldsymbol{\mu}_i$. Then

$$\begin{aligned} T^* - \mathsf{E}(T^*) = {} & \frac{2n}{n_1(n_1-1)} \sum_{r_1\neq r_2} \boldsymbol{Y}^{\mathsf{T}}_{1r_1}\boldsymbol{A}\boldsymbol{Y}_{1r_2} + \frac{4n}{n_1}\sum_{r_1} \boldsymbol{Y}^{\mathsf{T}}_{1r_1}\boldsymbol{A}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\ & + \frac{2n}{n_2(n_2-1)} \sum_{s_1\neq s_2} \boldsymbol{Y}^{\mathsf{T}}_{2s_1}\boldsymbol{A}\boldsymbol{Y}_{2s_2} - \frac{4n}{n_2}\sum_{s_1} \boldsymbol{Y}^{\mathsf{T}}_{2s_1}\boldsymbol{A}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \end{aligned}$$

$$-\frac{4n}{n_1n_2}\sum_{r_1}\sum_{s_1}\boldsymbol{Y}_{1r_1}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{Y}_{2s_1}\equiv\sum_{k=1}^{5}\Delta_k.$$

Since the expectations of Δ_k's are all zero and that of their cross product are also zero, it follows that

$$\mathsf{Var}\,(T^*)=\mathsf{E}\{T^*-\mathsf{E}(T^*)\}^2=\sum_{k=1}^{5}\mathsf{E}(\Delta_k^2).$$

With tedious algebra, we have (only the leading terms are kept)

$$\begin{aligned}
&\mathsf{E}(\Delta_1^2)=\frac{8n^2}{n_1(n_1-1)}\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{A}\boldsymbol{\Sigma}_1\boldsymbol{A}), &&\mathsf{E}(\Delta_2^2)=\frac{16n^2}{n_1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\Sigma}_1\boldsymbol{A}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2),\\
&\mathsf{E}(\Delta_3^2)=\frac{8n^2}{n_2(n_2-1)}\mathrm{tr}(\boldsymbol{\Sigma}_2\boldsymbol{A}\boldsymbol{\Sigma}_2\boldsymbol{A}), &&\mathsf{E}(\Delta_4^2)=\frac{16n^2}{n_2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathsf{T}}\boldsymbol{A}\boldsymbol{\Sigma}_2\boldsymbol{A}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2),\\
&\mathsf{E}(\Delta_5^2)=\frac{16n^2}{n_1n_2}\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{A}\boldsymbol{\Sigma}_2\boldsymbol{A}).
\end{aligned}$$

The lemma is proved by summing up the above terms and noting that $\boldsymbol{\Sigma}=(n/n_1)\boldsymbol{\Sigma}_1+(n/n_2)\boldsymbol{\Sigma}_2$. □

Proof of Theorem 4 This theorem follows from the proof of Theorem 1 of [5] with our $\boldsymbol{A}^{1/2}\boldsymbol{X}_{ij}$ in place of their $\boldsymbol{X}_{ij}$. □

Lemma 9 If there exists an absolute constant $K>0$ and $\kappa\geqslant 1$ such that $\mathsf{E}(X_{ij,k}^{\kappa})\leqslant K$ uniformly for all i,j,k, then $\max\limits_{k,l}|s_{kl}-\sigma_{kl}|=o_p(1)$.

Proof To prove this lemma, it suffices to show $\max\limits_{k,l}|s_{i,kl}-\sigma_{i,kl}|=o_p(1)$ for $i=1,2$, because

$$\max_{k,l}|s_{kl}-\sigma_{kl}|\leqslant\max_{k,l}|s_{1,kl}-\sigma_{1,kl}|/\eta_1+\max_{k,l}|s_{2,kl}-\sigma_{2,kl}|/\eta_2.$$

We only prove $\max\limits_{k,l}|s_{1,kl}-\sigma_{1,kl}|=o_p(1)$ since the proof for $i=2$ is the same. The first subscript 1 will be dropped for convenience if no confusion is caused. Throughout this proof we assume $\boldsymbol{\mu}_1=\mathbf{0}$. It is easy to see that

$$s_{kl}-\sigma_{kl}=\frac{1}{n_1}\sum_{r=1}^{n_1}(X_{r,l}X_{r,k}-\sigma_{kl})-\overline{X}_{(l)}\overline{X}_{(k)},$$

where $\overline{X}_{(l)}$ is the lth component of $\overline{X}$. For any positive ε, we have

$$\begin{aligned}
&\mathsf{P}\Big(\max_{1\leqslant l\leqslant k\leqslant p}|s_{lk}-\sigma_{lk}|>2\varepsilon\Big)\leqslant\max_{1\leqslant l\leqslant k\leqslant p}\mathsf{P}(|s_{lk}-\sigma_{lk}|>2\varepsilon)\\
&\leqslant\max_{1\leqslant l\leqslant k\leqslant p}\mathsf{P}\Big(\Big|\frac{1}{n_1}\sum_{r=1}^{n_1}(X_{r,l}X_{r,k}-\sigma_{lk})\Big|>\varepsilon\Big)+\max_{1\leqslant l\leqslant k\leqslant p}\mathsf{P}(|\overline{X}_{(l)}\overline{X}_{(k)}|>\varepsilon)\\
&\leqslant\max_{1\leqslant l\leqslant k\leqslant p}\frac{1}{\varepsilon^{2\kappa}}\mathsf{E}\Big\{\frac{1}{n_1}\sum_{r=1}^{n_1}(X_{r,l}X_{r,k}-\sigma_{lk})\Big\}^{2\kappa}+\max_{1\leqslant l\leqslant k\leqslant p}\frac{1}{\varepsilon^{2\kappa}}\mathsf{E}\{\overline{X}_{(l)}\overline{X}_{(k)}\}^{2\kappa}.
\end{aligned}$$

We now study the magnitude of the two moments. Note that for each fixed pair (l,k), $X_{r,l}X_{r,k}-\sigma_{lk}$ $(r=1,2,\ldots,n)$ are independent and identically distributed random variables with mean zero. The condition $\mathsf{E}(X_{j,r}^{4\kappa})<\infty$ implies that they also have finite 2κ moments. Consequently it can be verified after straightforward but lengthy algebra that there exists a constant $C>0$ not depending on n such that

$$\mathsf{E}\Big\{\frac{1}{n_1}\sum_{r=1}^{n_1}(X_{r,l}X_{r,k}-\sigma_{lk})\Big\}^{2\kappa}\leqslant Cn^{-\kappa}$$

and $\mathsf{E}\{\overline{X}_{(l)}\overline{X}_{(k)}\}^{2\kappa}\leqslant Cn^{-2\kappa}$.

Thus, it follows that

$$\mathsf{P}\Big(\max_{l,k}|s_{lk}-\sigma_{lk}|>2\varepsilon\Big)\leqslant\frac{p(p+1)}{2}\cdot\frac{Cn^{-\kappa}+Cn^{-2\kappa}}{\varepsilon^{2\kappa}},$$

which under the assumption $p=o(n^{\kappa/2})$ means that $\max\limits_{l,k}|s_{lk}-\sigma_{lk}|=o_p(1)$. □

Lemma 10 Under the conditions of Lemma 5, $|\sigma_{lk}|\leqslant 1-\varepsilon_0$ for any $1\leqslant l\neq k\leqslant p$.

Proof Since $\sigma_{lk}=\sigma_{1,lk}/\eta_1+\sigma_{2,lk}/\eta_2$ and $|\rho_{i,lk}|=|\sigma_{i,lk}|/\sqrt{\sigma_{i,ll}\sigma_{i,kk}}\leqslant 1-\varepsilon_0$, we have

$$|\sigma_{lk}|\leqslant(1-\varepsilon_0)\Big(\frac{\sqrt{\sigma_{1,ll}\sigma_{1,kk}}}{\eta_1}+\frac{\sqrt{\sigma_{2,ll}\sigma_{2,kk}}}{\eta_2}\Big).\tag{7}$$

Meanwhile it can be seen that

$$\begin{aligned}\Big(\frac{\sqrt{\sigma_{1,ll}\sigma_{1,kk}}}{\eta_1}+\frac{\sqrt{\sigma_{2,ll}\sigma_{2,kk}}}{\eta_2}\Big)^2&\leqslant\frac{\sigma_{1,ll}\sigma_{1,kk}}{\eta_1^2}+\frac{\sigma_{2,ll}\sigma_{2,kk}}{\eta_2^2}+\frac{\sigma_{1,ll}\sigma_{2,kk}+\sigma_{1,kk}\sigma_{2,ll}}{\eta_1\eta_2}\\&=\Big(\frac{\sigma_{1,ll}}{\eta_1}+\frac{\sigma_{2,ll}}{\eta_2}\Big)\Big(\frac{\sigma_{1,kk}}{\eta_1}+\frac{\sigma_{2,kk}}{\eta_2}\Big).\end{aligned}$$

Since $1=\sigma_{ll}=\sigma_{1,ll}/\eta_1+\sigma_{2,ll}/\eta_2$ for any $1\leqslant l\leqslant p$, we have $\eta_1^{-1}\sqrt{\sigma_{1,ll}\sigma_{1,kk}}+\eta_1^{-2}\sqrt{\sigma_{2,ll}\sigma_{2,kk}}\leqslant 1$. It then follows from (7) that $|\sigma_{lk}|\leqslant 1-\varepsilon_0$. □

Proof of Lemma 5 Let $\boldsymbol{U}=(U_{kl})_{p\times p}$ with

$$U_{kl}=\sum_{r_1\neq r_2}\sum_{s_1\neq s_2}\frac{n(\boldsymbol{X}_{1r_1,k}-\boldsymbol{X}_{2s_1,k})(\boldsymbol{X}_{1r_2,l}-\boldsymbol{X}_{2s_2,l})}{n_1(n_1-1)n_2(n_2-1)}.$$

Then it can be verified that $T_{kl}-T_{kl}^*=\mathrm{tr}[\{(\boldsymbol{S}_{(kl)})^{-1}-(\boldsymbol{\Sigma}_{(kl)})^{-1}\}\boldsymbol{U}_{(kl)}]$.

We have shown in Lemma 9 that $\max\limits_{k,l}|s_{kl}-\sigma_{kl}|=o_p(1)$. Thus given the ε_0 in Lemma 5, as n is large, $\mathsf{P}\big(\max\limits_{k,l}|s_{kl}-\sigma_{kl}|\leqslant\varepsilon_0/3\big)$ can be as close to one as possible. In the rest part of this proof, we study $T_{kl}-T_{kl}^*$ conditionally the event $E_n=\big\{\max\limits_{k,l}|s_{kl}-\sigma_{kl}|\leqslant\varepsilon_0/3\big\}$.

When applying the first order Taylor expansion to study the elements of $(\boldsymbol{S}_{(kl)})^{-1}-(\boldsymbol{\Sigma}_{(kl)})^{-1}$, we find that they all are bounded by

$$C_1|s_{kk}-\sigma_{kk}|+C_2|s_{kl}-\sigma_{kl}|+C_3|s_{kl}-\sigma_{kl}|.\tag{8}$$

Take for example the $(1,1)$ element, which is $s_{ll}/(s_{kk}s_{ll}-s_{kl}^2)-\sigma_{ll}/(\sigma_{kk}\sigma_{ll}-\sigma_{kl}^2)$. Conditionally on E_n, $|s_{kl}|<1-(2/3)\varepsilon_0$ for $k\neq l$, and $|s_{kl}|<|\sigma_{kl}|+\varepsilon_0/3\leqslant 1-(2/3)\varepsilon_0$. The partial derivative of $s_{ll}/(s_{kk}s_{ll}-s_{kl}^2)$ with respect to s_{kk} is bounded by

$$\Big|\frac{s_{ll}^2}{(s_{kk}s_{ll}-s_{kl}^2)^2}\Big|\leqslant\Big|\frac{(1+\varepsilon_0/3)^2}{\{(1-\varepsilon_0/3)^2-(1-(2/3)\varepsilon_0)^2\}^2}\Big|.$$

Similarly its partial derivatives with respect to s_{kl} and s_{ll} are also bounded.

Therefore it follows that

$$|T_{kl}-T_{kl}^*|\leqslant(C_1|s_{kk}-\sigma_{kk}|+C_2|s_{kl}-\sigma_{kl}|+C_3|s_{kl}-\sigma_{kl}|)\cdot(|U_{kk}|+2|U_{kl}|+|U_{ll}|).\quad(9)$$

We study the magnitude of $\mathsf{E}(|U_{kl}|)$, which is controlled by

$$\sqrt{\mathsf{E}(U_{kl}^2)}=\sqrt{\mathsf{Var}\,(U_{kl})+\{\mathsf{E}(U_{kl})\}^2}.$$

We need to calculate both $\mathsf{E}(U_{kl})$ and $\mathsf{Var}\,(U_{kl})$. It is easy to see $\mathsf{E}(U_{kl})=n(\mu_{1,k}-\mu_{2,k})\cdot(\mu_{1,l}-\mu_{2,l})$. Let $\boldsymbol{M}_{(kl)}$ be a 2×2 matrix with its $(2,1)$ element being 1 and the rest elements 0. Then $U_{kl}=Q_{kl}$ which is defined in (11). Lemma 12 implies

$$\begin{aligned}\mathsf{Var}\,(U_{kl})&=4\{1+O(n^{-1})\}+4n\{(\sigma_{1,ll}/\eta_1)(\mu_{1,k}-\mu_{2,k})^2+(\sigma_{2,kk}/\eta_2)(\mu_{1,l}-\mu_{2,l})^2\}\\&\leqslant 4\{1+O(n^{-1})\}+4n\{(\mu_{1,k}-\mu_{2,k})^2+(\mu_{1,l}-\mu_{2,l})^2\},\end{aligned}$$

where we have used the facts $\sigma_{kk}=1$ and $\sigma_{i,kk}\leqslant\eta_i$. Therefore

$$\mathsf{E}(U_{kl}^2)\leqslant\{2+n(\mu_{1,k}-\mu_{2,k})^2\}\{2+n(\mu_{1,l}-\mu_{2,l})^2\}+o(1),$$

which means $\sqrt{\mathsf{E}(U_{kl}^2)}\leqslant 2+n(\mu_{1,k}-\mu_{2,k})^2+n(\mu_{1,l}-\mu_{2,l})^2+o(1)$.

Under Assumption 1, there exists a constant $K_1>0$ such that $\mathsf{E}X_{ij,k}^4<K_1$ and therefore $\mathsf{E}|s_{kk}-\sigma_{kk}|^2\leqslant K_2/n$ for some $K_2>0$. Thus summing up both sides of (9) and employing the Schwarz inequality, we have for some constant $K>0$

$$\begin{aligned}\mathsf{E}|T-T^*|&\leqslant\mathsf{E}\sum_{k,l}|T_{kl}-T_{kl}^*|\\&\leqslant Kn^{-1/2}(p-1)\Big\{p+o(p)+2n\sum_{k=1}^{p}(\mu_{1,k}-\mu_{2,k})^2\Big\}=O(n^{-1/2}p^2).\end{aligned}$$

Since $p^4=o(n\mathrm{tr}\{(\boldsymbol{\Sigma A})^2\})$, we conclude $T-T^*=o(\sqrt{\mathrm{tr}\{(\boldsymbol{\Sigma A})^2\}}\,)$. □

Lemma 11 Let the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{A}$ be those defined in Theorem 4 and $\boldsymbol{R}\equiv(\rho_{ij})_{p\times p}=\boldsymbol{D}_\sigma^{-1/2}\boldsymbol{\Sigma}\boldsymbol{D}_\sigma^{-1/2}$ with $\boldsymbol{D}_\sigma=\mathsf{diag}\{\sigma_{11},\sigma_{22},\ldots,\sigma_{pp}\}$.

(a) We have $\mathrm{tr}\{(\boldsymbol{\Sigma A})^2\}\geqslant(p-1)^2\mathrm{tr}(\boldsymbol{R}^2)$.

(b) If there exists $\varepsilon_0 \in (0,1)$ such that $\max_{i,j}|\rho_{ij}| \leqslant 1-\varepsilon_0$ holds uniformly for all n, then $\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\} \leqslant 4(p-1)^2\mathrm{tr}(\boldsymbol{R}^2)/\varepsilon_0^2$.

Proof Clearly, $-1 \leqslant \rho_{rs} \leqslant 1$ and $\rho_{rr}=1$. Let $\boldsymbol{R}_{(rs,kl)}$ be a 2×2 matrix consisting of the elements at the crosses of the rth, sth rows and the kth, lth columns of $\boldsymbol{R}$. It is not hard to verify that

$$\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\} = \mathrm{tr}\{\boldsymbol{R}(\boldsymbol{D}_\sigma^{1/2}\boldsymbol{A}\boldsymbol{D}_\sigma^{1/2})\boldsymbol{R}(\boldsymbol{D}_\sigma^{1/2}\boldsymbol{A}\boldsymbol{D}_\sigma^{1/2})\} = \sum_{k\neq l}\sum_{r\neq s} x_{rs,kl},$$

where $x_{rs,kl} = \mathrm{tr}\{\boldsymbol{R}_{(rs,kl)}(\boldsymbol{R}_{(kl)})^{-1}\boldsymbol{R}_{(kl,rs)}(\boldsymbol{R}_{(rs)})^{-1}\}$.

Using the fact that

$$(\boldsymbol{R}_{(kl)})^{-1} = \frac{1}{2(1-\rho_{kl}^2)}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1-\rho_{kl} & 0 \\ 0 & 1+\rho_{kl} \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

we find

$$\begin{aligned} x_{rs,kl} = {} & \frac{1}{2(1-\rho_{kl}^2)(1-\rho_{rs}^2)} \\ & \cdot [(1-\rho_{kl})\{(\rho_{rk}+\rho_{rl})^2 - 2\rho_{rs}(\rho_{rk}+\rho_{rl})(\rho_{sk}+\rho_{sl}) + (\rho_{sk}+\rho_{sl})^2\} \\ & + (1+\rho_{kl})\{(\rho_{rk}-\rho_{rl})^2 - 2\rho_{rs}(\rho_{sk}-\rho_{sl})(\rho_{rk}-\rho_{rl}) + (\rho_{sk}-\rho_{sl})^2\}], \end{aligned} \tag{10}$$

indicating that $x_{rs,kl}$ is always nonnegative for all $r\neq s$ and $k\neq l$.

We first prove part (a). After splitting $(\rho_{sk}+\rho_{sl})^2$ into $\rho_{rs}^2(\rho_{sk}+\rho_{sl})^2+(1-\rho_{rs}^2)(\rho_{sk}+\rho_{sl})^2$, we have

$$\begin{aligned} x_{rs,kl} = {} & \frac{(1-\rho_{kl})\{(\rho_{rk}+\rho_{rl})-\rho_{rs}(\rho_{sk}+\rho_{sl})\}^2 + (1+\rho_{kl})\{(\rho_{rk}-\rho_{rl})-\rho_{rs}(\rho_{sk}-\rho_{sl})\}^2}{2(1-\rho_{kl}^2)(1-\rho_{rs}^2)} \\ & + \frac{(1-\rho_{kl})(\rho_{sk}+\rho_{sl})^2 + (1+\rho_{kl})(\rho_{sk}-\rho_{sl})^2}{2(1-\rho_{kl}^2)}. \end{aligned}$$

Since the first term is nonnegative and both $(1-\rho_{rs})$ and $(1+\rho_{rs})$ are no less than $(1-|\rho_{rs}|)$, it then follows that

$$x_{rs,kl} \geqslant \frac{(\rho_{sk}+\rho_{sl})^2+(\rho_{sk}-\rho_{sl})^2}{2(1+|\rho_{kl}|)} \geqslant \frac{1}{2}(\rho_{sk}^2+\rho_{sl}^2).$$

This immediately implies

$$\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\} = \sum_{k\neq l}\sum_{r\neq s} x_{rs,kl} \geqslant \frac{1}{2}\sum_{k\neq l}\sum_{r\neq s}(\rho_{sk}^2+\rho_{sl}^2) = (p-1)^2\mathrm{tr}(\boldsymbol{R}^2).$$

We now prove part (b). Using the fact that $|\rho_{rs}|\leqslant 1$ and Cauchy-Schwarz inequality, we have

$$-2\rho_{rs}(\rho_{rk}+\rho_{rl})(\rho_{sk}+\rho_{sl}) \leqslant (\rho_{rk}+\rho_{rl})^2 + (\rho_{sk}+\rho_{sl})^2.$$

Then it follows that

$$\begin{aligned} x_{rs,kl} &\leqslant \frac{(1-\rho_{kl})\{(\rho_{rk}+\rho_{rl})^2+(\rho_{sk}+\rho_{sl})^2\}+(1+\rho_{kl})\{(\rho_{rk}-\rho_{rl})^2+(\rho_{sk}-\rho_{sl})^2\}}{(1-\rho_{kl}^2)(1-\rho_{rs}^2)} \\ &\leqslant \frac{2(\rho_{rk}^2+\rho_{rl}^2+\rho_{sk}^2+\rho_{sl}^2)}{(1-|\rho_{kl}|)(1-\rho_{rs}^2)}. \end{aligned}$$

The condition $|\rho_{rs}| \leqslant 1-\varepsilon_0$ implies that $1-|\rho_{kl}| \geqslant \varepsilon_0$ and $(1-\rho_{rs}^2) \geqslant 2\varepsilon_0$. Therefore we have $x_{rs,kl} \leqslant (\rho_{rk}^2+\rho_{rl}^2+\rho_{sk}^2+\rho_{sl}^2)/\varepsilon_0^2$, which means

$$\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\} = \sum_{k\neq l}\sum_{r\neq s} x_{rs,kl} \leqslant \sum_{k\neq l}\sum_{r\neq s}(\rho_{rk}^2+\rho_{rl}^2+\rho_{sk}^2+\rho_{sl}^2)/\varepsilon_0^2 = 4(p-1)^2\mathrm{tr}(\boldsymbol{R}^2)/\varepsilon_0^2. \qquad \square$$

Lemma 12 For any $1 \leqslant k, l \leqslant p$, let $\boldsymbol{M}_{(kl)}$ be a generic 2×2 matrix and let

$$Q_{kl} = \sum_{r_1\neq r_2}\sum_{s_1\neq s_2} \frac{n(\boldsymbol{X}_{1r_1(kl)}-\boldsymbol{X}_{2s_1(kl)})^{\mathsf{T}}\boldsymbol{M}_{(kl)}(\boldsymbol{X}_{1r_2(kl)}-\boldsymbol{X}_{2s_2(kl)})}{n_1(n_1-1)n_2(n_2-1)}. \tag{11}$$

Under Assumption 1, we have

$$\begin{aligned} \mathsf{Cov}\,(Q_{kl}, Q_{rs}) &= 2\mathrm{tr}\{\boldsymbol{\Sigma}_{(rs,kl)}\boldsymbol{M}_{(kl)}\boldsymbol{\Sigma}_{(kl,rs)}\boldsymbol{M}_{(rs)}^{\mathsf{T}}\}\times\{1+O(n^{-1})\} \\ &\quad + \frac{4n^2}{n_1}(\boldsymbol{\mu}_{1(kl)}-\boldsymbol{\mu}_{2(kl)})^{\mathsf{T}}\boldsymbol{M}_{(kl)}\boldsymbol{\Sigma}_{1(kl,rs)}\boldsymbol{M}_{(rs)}^{\mathsf{T}}(\boldsymbol{\mu}_{1(rs)}-\boldsymbol{\mu}_{2(rs)}) \\ &\quad + \frac{4n^2}{n_2}(\boldsymbol{\mu}_{1(kl)}-\boldsymbol{\mu}_{2(kl)})^{\mathsf{T}}\boldsymbol{M}_{(kl)}^{\mathsf{T}}\boldsymbol{\Sigma}_{2(kl,rs)}\boldsymbol{M}_{(rs)}(\boldsymbol{\mu}_{1(rs)}-\boldsymbol{\mu}_{2(rs)}). \end{aligned}$$

Proof We rewrite Q_{kl} as

$$\begin{aligned} Q_{kl} &= \sum_{r_1\neq r_2}\frac{n\boldsymbol{X}_{1r_1(kl)}^{\mathsf{T}}\boldsymbol{M}_{(kl)}\boldsymbol{X}_{1r_2(kl)}}{n_1(n_1-1)} + \sum_{s_1\neq s_2}\frac{nX_{2s_1(kl)}^{\mathsf{T}}\boldsymbol{M}_{(kl)}X_{2s_2(kl)}}{n_2(n_2-1)} \\ &\quad - \sum_{r_2,s_1}\frac{2n\boldsymbol{X}_{2s_1(kl)}^{\mathsf{T}}\boldsymbol{M}_{(kl)}\boldsymbol{X}_{1r_2(kl)}}{n_1n_2}. \end{aligned}$$

Let $\boldsymbol{Y}_1 = \boldsymbol{X}_1 - \boldsymbol{\mu}_1$ and $\boldsymbol{Y}_2 = \boldsymbol{X}_2 - \boldsymbol{\mu}_2$. It then follows that

$$\begin{aligned} Q_{kl} - \mathsf{E}\{Q_{kl}\} &= \sum_{r_1\neq r_2}\frac{n\boldsymbol{Y}_{1r_1(kl)}^{\mathsf{T}}\boldsymbol{M}_{(kl)}\boldsymbol{Y}_{1r_2(kl)}}{n_1(n_1-1)} + \sum_{r_2}\frac{2n(\boldsymbol{\mu}_{1(kl)}-\boldsymbol{\mu}_{2(kl)})^{\mathsf{T}}\boldsymbol{M}_{(kl)}\boldsymbol{Y}_{1r_2(kl)}}{n_1} \\ &\quad + \sum_{s_1\neq s_2}\frac{n\boldsymbol{Y}_{2s_1(kl)}^{\mathsf{T}}\boldsymbol{M}_{(kl)}\boldsymbol{Y}_{2s_2(kl)}}{n_2(n_2-1)} - \sum_{s_1}\frac{2n\boldsymbol{Y}_{2s_1(kl)}^{\mathsf{T}}\boldsymbol{M}_{(kl)}(\boldsymbol{\mu}_{1(kl)}-\boldsymbol{\mu}_{2(kl)})}{n_2} \\ &\quad - \sum_{r_2}\sum_{s_1}\frac{2n\boldsymbol{Y}_{2s_1(kl)}^{\mathsf{T}}\boldsymbol{M}_{(kl)}\boldsymbol{Y}_{1r_2(kl)}}{n_1n_2} \equiv \sum_{h=1}^{5} J_{h(kl)} \end{aligned}$$

and $\mathsf{Cov}\,(Q_{kl}) = \mathsf{E}(Q_{kl}-\mathsf{E}\{Q_{kl}\})^2 = \mathsf{E}\{\sum_{h=1}^{5} J_{h(kl)}\}^2$.

It can be verified that the expectation of $J_{r(kl)}J_{s(rs)}$ for $1 \leqslant r \neq s \leqslant 5$ are all equal to zero. Let $\boldsymbol{\Sigma}_{1(rs,kl)}$ be a 2×2 submatrix of $\boldsymbol{\Sigma}_1$ consisting of the elements at the crosses of the rth, sth rows and kth, lth columns. With tedious algebra, we obtain

$$\begin{aligned}
\mathsf{E}(J_{1(kl)}J_{1(rs)}) &= \frac{2n^2}{n_1(n_1-1)}\mathrm{tr}\{\boldsymbol{\Sigma}_{1(rs,kl)}\boldsymbol{M}_{(kl)}\boldsymbol{\Sigma}_{1(kl,rs)}\boldsymbol{M}_{(rs)}^{\mathsf{T}}\},\\
\mathsf{E}(J_{3(kl)}J_{3(rs)}) &= \frac{2n^2}{n_2(n_2-1)}\mathrm{tr}(\boldsymbol{\Sigma}_{2(rs,kl)}\boldsymbol{M}_{(kl)}\boldsymbol{\Sigma}_{2(kl,rs)}\boldsymbol{M}_{(rs)}^{\mathsf{T}}),\\
\mathsf{E}(J_{2(kl)}J_{2(rs)}) &= \frac{4n^2}{n_1}(\boldsymbol{\mu}_{1(kl)}-\boldsymbol{\mu}_{2(kl)})^{\mathsf{T}}\boldsymbol{M}_{(kl)}\boldsymbol{\Sigma}_{1(kl,rs)}\boldsymbol{M}_{(rs)}^{\mathsf{T}}(\boldsymbol{\mu}_{1(rs)}-\boldsymbol{\mu}_{2(rs)}),\\
\mathsf{E}(J_{4(kl)}J_{4(rs)}) &= \frac{4n^2}{n_2}(\boldsymbol{\mu}_{1(kl)}-\boldsymbol{\mu}_{2(kl)})^{\mathsf{T}}\boldsymbol{M}_{(kl)}^{\mathsf{T}}\boldsymbol{\Sigma}_{2(kl,rs)}\boldsymbol{M}_{(rs)}(\boldsymbol{\mu}_{1(rs)}-\boldsymbol{\mu}_{2(rs)}),\\
\mathsf{E}(J_{5(kl)}J_{5(rs)}) &= \frac{4n^2}{n_1n_2}\mathrm{tr}(\boldsymbol{\Sigma}_{2(rs,kl)}^{\mathsf{T}}\boldsymbol{M}_{(kl)}\boldsymbol{\Sigma}_{1(kl,rs)}\boldsymbol{M}_{(rs)}^{\mathsf{T}}).
\end{aligned}$$

The lemma is proved by summing up the above covariances and simplifying the summation using the fact $\boldsymbol{\Sigma}_{(kl,rs)} = (n/n_1)\boldsymbol{\Sigma}_{1(kl,rs)} + (n/n_2)\boldsymbol{\Sigma}_{2(kl,rs)}$. $\square$

Proof of Lemma 7 Without loss of generality, we assume $\sigma_{kk} = 1$ for all $k = 1, 2, \ldots, p$. Let E_n be the event defined in the proof of Lemma 7 and $x_{rs,kl}$ be the quantity defined in the proof of Lemma 12. Lemma (9) implies that the probability of E_n converges to 1 as n is large. Therefore if Lemma 7 holds conditionally on E_n, then Lemma 7 is proved.

Define $\widehat{x}_{rs,kl}$ be $x_{rs,kl}$ with all ρ_{rs}'s replaced by $\widehat{\rho}_{rs} = \widehat{\sigma}_{rs}/\sqrt{\widehat{\sigma}_{rr}\widehat{\sigma}_{ss}}$. Thus we have

$$|\mathrm{tr}\{(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{A}})^2\} - \mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\}| = |\sum_{r,s,k,l}\widehat{x}_{rs,kl} - x_{rs,kl}| \leqslant \sum_{r\neq s,k\neq l}|\widehat{x}_{rs,kl} - x_{rs,kl}|.$$

Conditionally on E_n, there exists a constant C_1 (dependent on ε_0 not on n, i, j, k, l) such that

$$|\widehat{x}_{rs,kl} - x_{rs,kl}| \leqslant C_1 \sum_{r,s\in\{r,s,k,l\}}|\widehat{\sigma}_{rs} - \sigma_{rs}|.$$

The above two inequalities imply

$$|\mathrm{tr}\{(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{A}})^2\} - \mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\}| \leqslant C_1 \sum_{i\neq s,k\neq l}\sum_{r,s\in\{r,s,k,l\}}|\widehat{\sigma}_{rs} - \sigma_{rs}| \leqslant 16C_1p^2\sum_{r,s}|\widehat{\sigma}_{rs} - \sigma_{rs}|.$$

According to Lemma 11, $\mathrm{tr}\{(\boldsymbol{\Sigma}\boldsymbol{A})^2\} \geqslant (p-1)^2\mathrm{tr}(\boldsymbol{R}^2)$. Thus Lemma 7 will be proved if we can show $\sum_{r,s}|\widehat{\sigma}_{rs} - \sigma_{rs}| = o_p(\mathrm{tr}(\boldsymbol{R}^2))$.

Under Assumption 1 and the assumption $\sigma_{kk} = 1$, X_{ij}'s have uniformly bounded fourth moments, i.e., there exists $C_2 > 0$ such that $\max_{i,j}\mathsf{E}|X_{ij}^4| < C_2$. At the same time, there exists a constant $C_3 > 0$ such that $\max_{k,l}\mathsf{E}\{(\widehat{\sigma}_{kl} - \sigma_{kl})^2\} < C_3/n$. Thus

$$\mathsf{E}\Big(\sum_{k,l}|\widehat{\sigma}_{kl} - \sigma_{kl}|\Big)^2 \leqslant p^2\sum_{k,l}\mathsf{E}(\widehat{\sigma}_{kl} - \sigma_{kl})^2 \leqslant p^4C_3/n = o(\{\mathrm{tr}(\boldsymbol{R}^2)\}^2),$$

where the last equality follows from the assumption $p^2 = o(\sqrt{n}\mathrm{tr}(\boldsymbol{R}^2))$. Consequently we have $\sum_{k,l} |\widehat{\sigma}_{kl} - \sigma_{kl}| = o_p(\mathrm{tr}(\boldsymbol{R}^2))$. $\square$

Appendix B: Simulation for One-Sample Case

Denote $\boldsymbol{R} = (\rho_{kl})$ with $\rho_{kk} = 1$ and $\rho_{kl} = \rho \in (0, 1)$. Let $\boldsymbol{Z}$ be a random vector from $\mathrm{N}(\boldsymbol{0}, \boldsymbol{R})$ and W be a random variable from χ^2_v, independent of $\boldsymbol{Z}$. The degree of freedom v may be 4 or 8. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^\intercal$. We generate data from three models: (1) multi-normal distribution $\boldsymbol{X} = \boldsymbol{Z} + \boldsymbol{\mu}$, (2) multi-t distribution $\boldsymbol{X} = \boldsymbol{Z}/W + \boldsymbol{\mu}$, and (3) multi-$\chi^2$ distribution $X_i = \sum_{j=1}^{v} Z^2_{j,i} + \mu_i$, where $Z_{j,i}$ and μ_i denote the ith components of $\boldsymbol{Z}_j$ and $\boldsymbol{\mu}$, and $\boldsymbol{Z}_j$'s are i.i.d. copies of $\boldsymbol{Z}$. Suppose the hypothesis to test is $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$.

We fix $n = 100$ and $p = 300$. Four choices of ρ are considered: 0, 0.2, 0.5 and 0.8. The simulated type I errors are tabulated in the first panel of Table 4. We find that when $\rho > 0$, the type I errors of the CHT test are comparable with those of the CQ and BS tests. When $\rho = 0$, the type I errors of the CHT test is much smaller than the significance level 5%, indicating that the CHT test is very conservative in this situation compared with the rest three tests. This is probably because the estimation of the zero ρ values enlarges the estimated variance of the CHT test.

In power comparison, we generated data-sets with $\mu_1 = \mu_2 = \cdots = \mu_6 = 1$ and $\mu_7 = \mu_8 = \cdots = \mu_p = 0$. The simulated powers are presented in the second panel of Table 4. We also transform the generated data-sets by a component-wise transformation, $\boldsymbol{D} = \mathrm{diag}(d_{11}, d_{22}, \ldots, d_{pp})$ with d_{ii}'s i.i.d. from the uniform distribution on $(0, 10)$; The $\boldsymbol{D}$ keeps unchanged throughout this simulation. The corresponding simulated powers of the CQ and BS tests are given in the last panel of Table 4.

When $\rho > 0$, the CHT test have uniformly larger powers than the rest three tests. In particular, under models (2) with $v = 4$ and $\rho = 0.5$ or 0.8, the power gain of the CHT test over the CQ and BS tests can be as large as 60%. Meanwhile under component-wise transformation, the CHT test keeps unchanged while the CQ and BS tests both suffer from component-wise scale changes in data. For example, the powers of the CQ and BS tests can vary from 41% to 77% under model (1) with $\rho = 0.8$, indicating that they show some sensitivity to scale transformations. When $\rho = 0$, the CHT test seems inferior to the rest three tests, which is probably due to the too conservative variance estimation of T. There is room for improvement of the CHT test when the data have little between-component correlation.

Table 4 Simulated rejection rates (%) in the one-sample case

model	v	ρ	Type I error				Power				Power (scaled)	
			CHT	CQ	BS	SD	CHT	CQ	BS	SD	CQ	BS
(1)		0	0.44	5.09	4.88	5.17	100.00	100.00	100.00	100.00	100.00	100.00
		0.2	6.31	6.77	6.76	4.62	100.00	100.00	100.00	100.00	100.00	100.00
		0.5	7.10	7.05	7.05	1.44	100.00	94.74	94.74	19.20	84.86	84.80
		0.8	7.45	7.37	7.37	0.30	100.00	41.28	41.30	1.38	77.08	77.12
(2)	4	0	0.01	5.41	0.03	0.03	99.56	99.99	96.75	96.47	99.94	96.36
		0.2	5.09	7.55	5.51	3.13	97.78	96.16	89.05	75.35	98.52	94.75
		0.5	6.77	7.21	6.79	1.33	94.65	33.46	31.38	5.62	16.00	14.89
		0.8	6.95	7.40	7.15	0.30	99.96	17.89	17.25	0.72	14.40	13.93
	8	0	0.12	5.84	1.09	0.95	100.00	100.00	100.00	100.00	100.00	100.00
		0.2	5.93	7.16	6.47	3.95	100.00	99.99	99.97	99.25	100.00	99.98
		0.5	7.25	7.51	7.39	1.53	100.00	64.67	63.35	10.68	63.84	62.73
		0.8	7.01	7.07	6.99	0.28	100.00	25.68	25.45	1.34	19.52	19.22
(3)	4	0	2.36	5.83	5.42	14.38	64.78	87.33	86.35	91.37	90.06	89.42
		0.2	9.56	7.51	7.04	7.18	20.22	16.36	15.23	11.96	25.83	24.12
		0.5	8.91	7.75	7.50	2.36	14.57	10.22	9.92	3.11	9.21	8.96
		0.8	8.33	7.77	7.59	0.78	19.66	8.61	8.45	0.84	8.06	7.95
	8	0	1.03	5.38	5.00	9.12	14.97	43.8	42.65	47.73	46.37	45.61
		0.2	8.14	7.45	7.25	6.09	11.76	11.2	10.77	8.01	9.46	9.08
		0.5	8.01	7.27	7.17	1.91	10.17	8.37	8.23	2.27	8.47	8.32
		0.8	7.84	7.46	7.43	0.52	11.97	7.97	7.91	0.42	8.04	7.95

References

[1] Hotelling H. The generalization of Student's ratio [J]. *Ann. Math. Statist.*, 1931, **2(3)**: 360–378.

[2] Bai Z D, Saranadasa H. Effect of high dimension: by an example of a two sample problem [J]. *Statist. Sinica*, 1996, **6(2)**: 311–329.

[3] Srivastava M S, Du M. A test for the mean vector with fewer observations than the dimension [J]. *J. Multivariate Anal.*, 2008, **99(3)**: 386–402.

[4] Srivastava M S. A test for the mean vector with fewer observations than the dimension under non-normality [J]. *J. Multivariate Anal.*, 2009, **100(3)**: 518–532.

[5] Chen S X, Qin Y L. A two-sample test for high-dimensional data with applications to gene-set testing [J]. *Ann. Statist.*, 2010, **38(2)**: 808–835.

[6] Cai T T, Liu W D, Xia Y. Two-sample test of high dimensional means under dependence [J]. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 2014, **76(2)**: 349–372.

[7] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing [J]. *J. Roy. Statist. Soc. Ser. B*, 1995, **57(1)**: 289–300.

[8] Bullmore E, Fadili J, Breakspear M, et al. Wavelets and statistical analysis of functional magnetic resonance images of the human brain [J]. *Stat. Methods Med. Res.*, 2003, **12(5)**: 375–399.

[9] Crooks S A, Gray L J. Characterization of the 11-Year solar signal using a multiple regression analysis of the ERA-40 dataset [J]. *J. Climate*, 2005, **18(7)**: 996–1015.

[10] Fan J Q. A selective overview of nonparametric methods in financial econometrics (with discussion) [J]. *Statist. Sci.*, 2005, **20(4)**: 317–357.

[11] Goeman J J, Bühlmann P. Analyzing gene expression data in terms of gene sets: methodological issues [J]. *Bioinformatics*, 2007, **23(8)**: 980–987.

[12] Lu Y, Liu P Y, Xiao P, et al. Hotelling's T^2 multivariate profiling for detecting differential expression in microarrays [J]. *Bioinformatics*, 2005, **21(14)**: 3105–3113.

用于高维数据的复合Hotelling's T^2检验

李 婷

(复旦大学管理学院统计系, 上海, 200433)

摘 要: 本文研究高维数据下两样本均值的检验问题. 基于Hotelling's T^2检验, 我们提出了适用于高维数据均值检验的复合Hotelling's T^2检验统计量, 证明了其渐近正态性并研究了其渐近功效. 我们通过模拟和实例分析展示了该检验在有限样本下相比现有高维检验方法的优良性.

关键词: 高维均值检验; Hotelling's T^2检验; 渐近正态; 局部功效

中图分类号: O212.4