A feasible high dimensional randomization test for the mean vector

Rui Wang[a], Xingzhong Xu[a,b,*]

[a]*School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081,China*
[b]*Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China*

Abstract

The strength of randomization tests is that they are exact tests under certain symmetry assumption for distributions. In this paper, we propose a randomization test for the mean vector in high dimensional setting. We give an implementation of the proposed randomization test procedure, which is computationally feasible. So far, the asymptotic behaviors of randomization tests have only been studied in fixed dimension case. We investigate the asymptotic behavior of the proposed randomization test in high dimensional setting. It turns out that even if the symmetry assumption is violated, the proposed randomization test still has correct level asymptotically. The asymptotic power function is also given. Our theoretical and simulation results show that the proposed test has a wide application scope while still has good power behavior.

Keywords: Asymptotic power function, High dimension, Randomization test, Symmetry assumption

1. Introduction

Suppose $X_1, \ldots, X_n$ are p-variate independent and identically distributed (iid) random vectors with mean vector μ and covariance matrix Σ. In this paper, we consider the problem of testing the hypotheses

$$H_0 : \mu = 0_p \quad \text{versus} \quad H_1 : \mu \neq 0_p. \tag{1}$$

A classical test statistic for hypotheses (1) is Hotelling's T^2, defined as $n\bar{X}^T S^{-1}\bar{X}$, where $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ is the sample mean vector and $S = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$ is the sample covariance matrix. Under normal distribution, Hotelling's T^2 is the likelihood ratio test and enjoys desirable properties in fixed p setting. See, for example, Anderson (2003). However, Hotelling's test can not be defined when $p > n - 1$ due to the singularity of S. In a seminal paper, Bai and Saranadasa (1996) considered two sample testing problem and proposed a statistic by removing S^{-1} from Hotelling's T^2 statistic. They studied the asymptotic properties of their test statistic when p/n tends to a positive constant. Many subsequent papers generalized the idea of Bai and Saranadasa (1996) to more general models (Srivastava and Du, 2008; Chen

---

and Qin, 2010; Wang et al., 2015). The critical values of existing high dimensional tests mostly rely on the asymptotic normality of the test statistics. We call it asymptotic method. However, if Σ has spiked eigenvalues, the asymptotic normality of the test statistics is not be valid (Katayama et al., 2013). In this case, the asymptotic methods can not satisfactorily control the level.

The randomization test method is a tool to determine the critical value for a given test statistic. The idea of randomization tests dates back to Fisher (1935). See Romano (1990) for a general construction of randomization test. Its strength is in that the resulting test procedure has exact level under mild condition. There are many papers concerning the theoretical properties of randomization tests for fixed p case. See, for example, Romano (1990), Zhu (2000) and Chung and Romano (2016). In high dimensional setting, randomization tests are widely used in applied statistics (Subramanian et al., 2005; Efron and Tibshirani, 2007; Ko et al., 2016). However, little is known about the theoretical properties of the randomization test in high dimensional setting.

In this paper, we consider the following randomization method. Suppose $T(X_1, \ldots, X_n)$ is certain test statistic for hypotheses (1). Let $\epsilon_1, \ldots, \epsilon_n$ be iid Rademacher variables ($\Pr(\epsilon_i = 1) = \Pr(\epsilon_i = -1) = 1/2$) which are independent of data. The randomization test rejects the null hypothesis when $T(X_1, \ldots, X_n)$ is greater than the $1 - \alpha$ quantile of the conditional distribution of $T(\epsilon_1 X_1, \ldots, \epsilon_i X_i, \ldots, \epsilon_n X_n)$ given $X_1, \ldots, X_n$, and accepts the null hypothesis otherwise, where α is the significant level and the $1 - \alpha$ quantile of a distribution function $F(\cdot)$ is defined as $\inf\{y : F(y) \geq 1 - \alpha\}$. In fixed p setting, it's well known that randomization tests consume much more computing time than the asymptotic method, which historically hampered the use of randomization tests. The goal of this paper is to show that in high dimensional setting, randomization tests can be computationally feasible and have desirable statistic properties. Inspired by the work of Bai and Saranadasa (1996) and Chen and Qin (2010), we propose a randomization test for hypotheses (1). We give an implementation of the proposed randomization test, which is computationally feasible. We also investigate the asymptotic behavior of the test procedure. Our results show that even if the null distribution of X_1 is not symmetric, the randomization test is still asymptotically exact under fairly general assumptions. Hence the test procedure is robust. In particular, the proposed test can be applied to situations where the asymptotic method is not valid. We also derive the asymptotic power function of the proposed test. To the best of our knowledge, this is the first work which gives the asymptotic behavior of randomization tests in high dimensional setting. A simulation study is carried out to examine the numerical performance of the proposed test and compare with the asymptotic method and the bootstrap method. Compared with its competitors, the proposed test reduces the size distortion while still possesses reasonable test power.

The rest of the paper is organized in the following way. In Section 2, we propose a randomization test and give a fast implementation. In Section 3, we investigate the asymptotic behavior of the proposed test. The simulation results are reported in Section 4. The technical proofs are presented in Appendices.

2. Test Procedure

Consider testing the hypotheses (1) in high dimensional setting. It is known that Hotelling's T^2 can not be defined when $p > n - 1$. Bai and Saranadasa (1996) removed S^{-1} from Hotelling's T^2 statistic and proposed a statistic which has good power behavior in high dimensional setting. Their idea can also be used for testing hypotheses (1) and the statistic becomes $\bar{X}^T\bar{X}$. The asymptotic properties of the statistic requires p/n tends to a positive constant. Chen and Qin (2010) found that the restriction on p and n can be considerably relaxed by removing the diagonal elements in the statistic of Bai and Saranadasa (1996). For hypotheses (1), their statistic is $\sum_{i \neq j} X_i^T X_j$. Inspired by the statistics of Bai and Saranadasa (1996) and Chen and Qin (2010), we consider the statistic

$$T(X_1, \ldots, X_n) = \sum_{j < i} X_i^T X_j. \tag{2}$$

Bai and Saranadasa (1996) and Chen and Qin (2010) used asymptotic method to determine the critical value of $T(X_1, \ldots, X_n)$. However, as will be shown in Section 3, the critical value determined by the asymptotic method is not valid in some important cases. Hence it is desirable to find a better critical value.

The bootstrap and the randomization method are two popular methods to determine the critical value. However, the bootstrap method may not be a good choice for our problem. To see this, consider the following basic bootstrap procedure. Let $X_1^*, \ldots, X_n^*$ be a bootstrap sample which is randomly drawn from $\{X_1 - \bar{X}, \ldots, X_n - \bar{X}\}$ with replacement. Denote by

$$\mathcal{L}(T(X_1^*, \ldots, X_n^*)|X_1, \ldots, X_n), \tag{3}$$

the distribution of $T(X_1^*, \ldots, X_n^*)$ conditioning on $X_1, \ldots, X_n$. Then the bootstrap critical value for the statistic $T(X_1, \ldots, X_n)$ is defined to be the $1 - \alpha$ quantile of the bootstrap distribution (3). If this bootstrap method worked well, the bootstrap distribution (3) should mimic the null distribution of $T(X_1, \ldots, X_n)$. Then one may expect that the first two moments of (3) are close to the first two moments of $T(X_1, \ldots, X_n)$. Under the null hypothesis,

$$\mathrm{E}\,T(X_1, \ldots, X_n) = 0, \quad \mathrm{Var}(T(X_1, \ldots, X_n)) = \frac{n(n-1)}{2}\,\mathrm{tr}(\Sigma^2).$$

Also, it's straightforward to show that

$$\mathrm{E}(T(X_1^*, \ldots, X_n^*)|X_1, \ldots, X_n) = 0, \quad \mathrm{Var}(T(X_1^*, \ldots, X_n^*)|X_1, \ldots, X_n) = \frac{n(n-1)}{2}\Big(\frac{n-1}{n}\Big)^2\,\mathrm{tr}(S^2).$$

However, as pointed out by Bai and Saranadasa (1996), $\big((n-1)/n\big)^2\,\mathrm{tr}(S^2)$ is not even a ratio consistent estimator of $\mathrm{tr}(\Sigma^2)$ in high dimensional setting.

Now we consider the following randomization method. Let $\epsilon_1, \ldots, \epsilon_n$ be iid Rademacher variables which are independent of data. Denote by

$$\mathcal{L}\big(T(\epsilon_1 X_1, \ldots, \epsilon_i X_i, \ldots, \epsilon_n X_n)|X_1, \ldots, X_n\big) \tag{4}$$

the distribution of $T(\epsilon_1 X_1, \ldots, \epsilon_n X_n)$ conditioning on $X_1, \ldots, X_n$. We consider the test $\phi(X_1, \ldots, X_n)$ which equals to 1 if $T(X_1, \ldots, X_n)$ is greater than the $1 - \alpha$ quantile of the conditional distribution (4) and equals to 0 otherwise. It can be seen that

$$\mathrm{E}(T(\epsilon_1 X_1, \ldots, \epsilon_n X_n)|X_1, \ldots, X_n) = 0, \quad \mathrm{Var}(T(\epsilon_1 X_1, \ldots, \epsilon_n X_n)|X_1, \ldots, X_n) = \sum_{j<i}(X_i^T X_j)^2.$$

Under the null hypothesis, $\sum_{j<i}(X_i^T X_j)^2$ is a good estimator of $\mathrm{Var}(T(X_1, \ldots, X_n))$. In fact, it can be shown that it is unbiased and ratio consistent. Hence the randomization test $\phi(X_1, \ldots, X_n)$ may be a good choice for our problem.

Since $T(X_1, \ldots, X_n)$ equals to half of Chen and Qin (2010)'s statistic $\sum_{i\neq j} X_i^T X_j$, the test procedure $\phi(X_1, \ldots, X_n)$ is the randomization version of Chen and Qin (2010)'s test procedure. On the other hand, note that Bai and Saranadasa (1996)'s statistic $\bar{X}^T \bar{X}$ can be written as $n^{-2}\sum_{i=1}^n \sum_{j=1}^n X_i^T X_j$. Since $\sum_{i=1}^n X_i^T X_i$ is invariance under randomization, the test procedure $\phi(X_1, \ldots, X_n)$ is also the randomization version of Bai and Saranadasa (1996)'s test. Here we can see that the randomization test automatically removes the unwanted diagonal elements.

Under certain symmetry assumption, randomization tests are exact tests, which is a desirable property. See, for example, Lehmann and Romano (2005, Chapter 15). In our problem, the Type I error rate of $\phi(X_1, \ldots, X_n)$ does not exceed α provided that X_1 and $-X_1$ have the same distribution under null hypothesis. By refined definition of $\phi(X_1, \ldots, X_n)$ on the boundary of rejection region, one can obtain a test procedure with exact significant level. Such refinement only has minor effect on the test procedure and won't be considered in this paper. The test procedure $\phi(X_1, \ldots, X_n)$ can be equivalently implemented by p-value. Define

$$p(X_1, \ldots, X_n) = \mathrm{Pr}(T(\epsilon_1 X_1, \ldots, \epsilon_n X_n) \geq T(X_1, \ldots, X_n)|X_1, \ldots, X_n). \tag{5}$$

Then our test procedure rejects the null hypothesis when $p(X_1, \ldots, X_n) \leq \alpha$.

Given the data $X_1, \ldots, X_n$, the randomized statistic $T(\epsilon_1 X_1, \ldots, \epsilon_n X_n)$ is uniformly distributed on 2^n values. To compute the exact quantile of (4) or the p-value (5), one needs to calculate at least 2^n values, which is not feasible even when n is moderate. In practice, randomization test is often realized through an approximation of p-value (5). More specifically, we sample $\epsilon_1^*, \ldots, \epsilon_n^*$ and compute the randomized statistic $T^* = T(\epsilon_1^* X_1, \ldots, \epsilon_n^* X_n)$. Repeat B times for a large B and we obtain $T_1^*, \ldots, T_B^*$. Let $T_0 = T(X_1, \ldots, X_n)$ be the original statistic and define

$$\tilde{p}(X_1, \ldots, X_n) = \frac{1}{B+1}\Big(1 + \sum_{i=1}^B \mathbf{1}_{\{T_i^* \geq T_0\}}\Big).$$

The null hypothesis is rejected when $\tilde{p}(X_1, \ldots, X_n) \leq \alpha$. This test procedure can still control the significant level. In fact, we have $\mathrm{Pr}(\tilde{p}(X_1, \ldots, X_n) \leq u) \leq u$ for all $0 \leq u \leq 1$ and all positive integer B. See Lehmann

and Romano (2005, Page 636). By Bernoulli's law of large numbers, $\tilde{p}(X_1, \ldots, X_n)$ tends to $p(X_1, \ldots, X_n)$ in probability as $B \to \infty$. Here we emphasis that the convergence rate of $\tilde{p}(X_1, \ldots, X_n)$ to $p(X_1, \ldots, X_n)$ only relies on $p(X_1, \ldots, X_n)$. Hence the choice of B can be independent of the sample size n and the dimension of data p.

Now we consider the implementation of the randomization test procedure. The computation of T_0 costs $O(n^2 p)$ operations. To obtain T_i^*, $i = 1, \ldots, B$, we need to generate $\epsilon_1, \ldots, \epsilon_n$ and compute

$$T(\epsilon_1 X_1, \ldots, \epsilon_n X_n) = \sum_{1 \leq j < i \leq n} X_i^T X_j \epsilon_i \epsilon_j.$$

Note that $X_i^T X_j$ ($1 \leq j < i \leq n$) can be computed beforehand. Once we obtain $X_i^T X_j$, the computation of T_i^* costs $O(n^2)$ operations. Thus, the randomization test costs $O(n^2 p + n^2 B)$ operations in total. As mentioned earlier, the choice of B can be independent of n and p. Hence when p is large, the computational complexity of the entire procedure is $O(n^2 p)$. The randomization test doesn't need a variance estimator which is a necessary for the asymptotic method. Bai and Saranadasa (1996) used the following estimator for $\mathrm{tr}(\Sigma^2)$:

$$\frac{(n-1)^2}{(n+1)(n-2)} \Big(\mathrm{tr}(S_n^2) - \frac{1}{n-1} (\mathrm{tr}\, S_n)^2 \Big).$$

Even with a good implementation, the computation of this estimator still costs $O(np \min(n, p))$ operations. Chen and Qin (2010) used a more complicated estimator of $\mathrm{tr}(\Sigma^2)$. Consequently, if p is large compared with n, the randomization test is competitive compared with the asymptotic method in terms of computational complexity. This is different from the low dimensional setting where randomization tests usually consume much more computing time than the asymptotic methods.

If we only care about the decision (reject or accept) and the p-value is not needed, the computing time of the randomization test can be further reduced. In fact, the rejection region $\tilde{p}(X_1, \ldots, X_n) \leq \alpha$ can be written as

$$\sum_{i=1}^{B} (1 - \mathbf{1}_{\{T_i^* \geq T_0\}}) \geq B + 1 - (B+1)\alpha.$$

Since the left hand side is a sum of non-negative values, we can reject the null hypothesis once $\sum_{i=1}^{B_0} (1 - \mathbf{1}_{\{T_i^* \geq T_0\}}) \geq B + 1 - (B+1)\alpha$ for some B_0. Similarly, the acceptance region can be written as

$$\sum_{i=1}^{B} \mathbf{1}_{\{T_i^* \geq T_0\}} > (B+1)\alpha - 1.$$

we can accept the null hypothesis once $\sum_{i=1}^{B_0} \mathbf{1}_{\{T_i^* \geq T_0\}} > (B+1)\alpha - 1$ for some B_0. Algorithm 1 summarizes our computing method.

3. Asymptotic properties

As we have noted, the randomization test is exact provided that X_1 and $-X_1$ have the same distribution under null hypothesis. However, under this condition, all p variables are symmetric about 0, which is

Algorithm 1: Randomization Algorithm

Data: Data $X_1, \ldots, X_n$

Result: Reject or accept the null hypothesis

1 for $i \leftarrow 2$ **to** n **do**

2   **for** $j \leftarrow 1$ **to** $i - 1$ **do**

3 $D_{ij} \leftarrow X_i^T X_j$;

4   **end**

5 end

6 Compute $T_0 \leftarrow \sum_{1 \leq j < i \leq n} D_{ij}$;

7 Set $A \leftarrow 0$;

8 for $i = 1$ *to* B **do**

9   Generate $\epsilon_1, \ldots, \epsilon_n$ according to $\Pr(\epsilon_i = 1) = \Pr(\epsilon_i = -1) = \frac{1}{2}$;

10   **if** $\sum_{1 \leq j < i \leq n} D_{ij} \epsilon_i \epsilon_j \geq T_0$ **then**

11 $A \leftarrow A + 1$;

12   **if** $A > (B + 1)\alpha - 1$ **then return** *Accept*;

13   **else**

14   **if** $i - A \geq B + 1 - (B + 1)\alpha$ **then return** *Reject*;

15   **end**

16 end

---

unrealistic for many applications. In this section, we investigate the asymptotic properties of the test procedure $\phi(X_1, \ldots, X_n)$ without symmetry assumption.

We assume, like Chen and Qin (2010) and Bai and Saranadasa (1996), the following multivariate model:

$$X_i = \mu + \Gamma Z_i \text{ for } i = 1, \ldots, n, \tag{6}$$

where Γ is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma\Gamma^T = \Sigma$ and $Z_1, \ldots, Z_n$ are m-variate iid random vectors satisfying $\mathrm{E}(Z_i) = 0$ and $\mathrm{Var}(Z_i) = I_m$, the $m \times m$ identity matrix. Write $Z_i = (z_{i1}, \ldots, z_{im})^T$. We assume $\mathrm{E}(z_{ij}^4) = 3 + \Delta < \infty$ and

$$\mathrm{E}(z_{il_1}^{\alpha_1} z_{il_2}^{\alpha_2} \cdots z_{il_q}^{\alpha_q}) = \mathrm{E}(z_{il_1}^{\alpha_1})\,\mathrm{E}(z_{il_2}^{\alpha_2}) \cdots \mathrm{E}(z_{il_q}^{\alpha_q}) \tag{7}$$

for a positive integer q such that $\sum_{l=1}^{q} \alpha_l \leq 8$ and $l_1 \neq l_2 \neq \cdots \neq l_q$.

Let $\lambda_i(\Sigma)$ be the ith largest eigenvalue of Σ. In Bai and Saranadasa (1996), a key assumption for the eigenvalues of Σ is

$$\frac{\lambda_1(\Sigma)}{\sqrt{\mathrm{tr}(\Sigma^2)}} \to 0. \tag{8}$$

Correspondingly, Chen and Qin (2010) imposed the condition $\mathrm{tr}(\Sigma^4) = o\big(\mathrm{tr}^2(\Sigma^2)\big)$. From the inequality

$$\frac{\lambda_1(\Sigma)^4}{(\sum_{i=1}^{p} \lambda_i(\Sigma)^2)^2} \leq \frac{\sum_{i=1}^{p} \lambda_i(\Sigma)^4}{(\sum_{i=1}^{p} \lambda_i(\Sigma)^2)^2} \leq \frac{\lambda_1(\Sigma)^2 \sum_{i=1}^{p} \lambda_i(\Sigma)^2}{(\sum_{i=1}^{p} \lambda_i(\Sigma)^2)^2}$$

we can see that these two conditions are equivalent. As pointed out by Katayama et al. (2013), however, the condition (8) excludes a typical situation where the population covariance matrix has spiked eigenvalues. See also Ma et al. (2015). To include such important cases, we consider a condition that is weaker than (8).

Suppose that there is a fixed integer $r \geq 0$ and positive constants $\kappa_1, \ldots, \kappa_r$ such that

$$\frac{\lambda_i(\Sigma)}{\sqrt{\mathrm{tr}(\Sigma^2)}} \to \kappa_i \text{ for } i = 1, \ldots, r \quad \text{and} \quad \frac{\lambda_i(\Sigma)}{\sqrt{\mathrm{tr}(\Sigma^2)}} \to 0 \text{ for } i > r. \tag{9}$$

The following theorem gives the asymptotic distribution of $T(X_1, \ldots, X_n)$ under (6), (7) and (9).

Theorem 1. *Under* (6), (7) *and* (9), *we have*

(i) *if* $\mu^T \Sigma \mu = o(n^{-1} \mathrm{tr}(\Sigma^2))$,

$$\frac{T(X_1, \ldots, X_n) - \frac{n(n-1)}{2}\|\mu\|^2}{\sqrt{\frac{n(n-1)}{2} \mathrm{tr}(\Sigma^2)}} \xrightarrow{w} \frac{\sqrt{2}}{2} \sum_{i=1}^{r} \kappa_i(\xi_i^2 - 1) + (1 - \sum_{i=1}^{r} \kappa_i^2)^{1/2} \xi_{r+1},$$

where $\xi_1, \ldots \xi_{r+1}$ *are iid standard normal random variables;*

(ii) *if* $n^{-1} \mathrm{tr}(\Sigma^2) = o(\mu^T \Sigma \mu)$,

$$\frac{T(X_1, \ldots, X_n) - \frac{n(n-1)}{2}\|\mu\|^2}{(n-1)\sqrt{n\mu^T \Sigma \mu}} \xrightarrow{\mathcal{L}} N(0,1).$$

7

Remark 1. If $r = 0$, Theorem 1 implies that $T(X_1, \ldots, X_n)$ is asymptotically normal distributed under the null hypothesis. This is a known result (Chen and Qin, 2010, Theorem 1). However, if $r > 0$, i.e. Σ has a few spiked eigenvalues, $T(X_1, \ldots, X_n)$ is not asymptotically normal distributed. Note that a similar result has been provided by Katayama et al. (2013, Theorem 3.1). Compared with their result, Theorem 1 doesn't need the normal assumption.

The asymptotic method of Chen and Qin (2010) rejects the null hypothesis when

$$\frac{\sum_{j<i} X_i^T X_j}{\sqrt{\frac{n(n-1)}{2} \widehat{\mathrm{tr}(\Sigma^2)}}} > \Phi^{-1}(1 - \alpha),$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution,

$$\widehat{\mathrm{tr}(\Sigma^2)} = \frac{1}{n(n-1)} \mathrm{tr} \left(\sum_{i \neq j} (X_i - \bar{X}_{(i,j)}) X_i^T (X_j - \bar{X}_{(i,j)}) X_j^T \right)$$

is a ratio consistent estimator of $\mathrm{tr}(\Sigma^2)$ and $\bar{X}_{(i,j)}$ is the sample mean after excluding X_i and X_j. Since the critical value is determined by normal distribution, (i) of Theorem 1 implies that the critical value of the asymptotic method is not valid if Σ has a few spiked eigenvalues.

Now we study the asymptotic properties of the randomization test. The conditional distribution

$$\mathcal{L}\left(\frac{T(\epsilon_1 X_1, \ldots, \epsilon_n X_n)}{\sqrt{\frac{n(n-1)}{2} \mathrm{tr}(\Sigma^2)}} \Bigg| X_1, \ldots, X_n \right) \tag{10}$$

plays an important role in our analysis. Let q_α^* be the $1 - \alpha$ quantile of the distribution (10). Then the test function $\phi(X_1, \ldots, X_n)$ equals to 1 if

$$\frac{T(X_1, \ldots, X_n)}{\sqrt{\frac{n(n-1)}{2} \mathrm{tr}(\Sigma^2)}} > q_\alpha^* \tag{11}$$

and equals to 0 otherwise. The asymptotic behavior of the left hand side of (11) is given by Theorem 1. We need to investigate the asymptotic property of q_α^*. Denote by $F^*(\cdot)$ the distribution function of

$$\frac{\sqrt{2}}{2} \sum_{i=1}^r \kappa_i (\xi_i^2 - 1) + (1 - \sum_{i=1}^r \kappa_i^2)^{1/2} \xi_{r+1},$$

where $\xi_1, \ldots \xi_{r+1}$ are iid standard normal random variables.

Theorem 2. *Under (6), (7) and (9), we have*

(i) *if $\mu^T \mu = o\left(\sqrt{\mathrm{tr}(\Sigma^2)}\right)$, then $q_\alpha^* \xrightarrow{P} F^{*-1}(1 - \alpha)$;*

(ii) *if $\sqrt{\mathrm{tr}(\Sigma^2)} = o(\mu^T \mu)$, $\left(\sqrt{\mathrm{tr}(\Sigma^2)}/(\mu^T \mu)\right) q_\alpha^*$ converges in probability to the $1 - \alpha$ quantile of $(\chi_1^2 - 1)/\sqrt{2}$, the centralized chi-squared distribution with degree of freedom 1.*

From (i) of Theorem 2 we can see that under the null hypothesis, the randomization can produce the correct level asymptotically. Note that this result doesn't assume that the distribution of X_1 is symmetric

under null hypothesis. This implies that the proposed test is robust when the symmetry assumption breaks down. This property is not held by all randomization tests. See, for example, Romano (1990). Compared with the asymptotic method, the validity of the proposed test does not rely on the asymptotic normality of $T(X_1, \ldots, X_n)$ and thus has a wider application scope.

The asymptotic power function of the randomization test can be derived using Theorem 1 and Theorem 2.

Theorem 3. *Under* (6), (7) *and* (9), *we have*

(i) *if* $\mu^T \Sigma \mu = o\big(n^{-1} \operatorname{tr}(\Sigma^2)\big)$,

$$\mathrm{E}\,\phi(X_1, \ldots, X_n) = 1 - F^*\left(F^{*-1}(1 - \alpha) - \sqrt{\frac{n(n-1)}{2}} \frac{\mu^T \mu}{\sqrt{\operatorname{tr}(\Sigma^2)}}\right) + o(1);$$

(ii) *if* $n^{-1} \operatorname{tr}(\Sigma^2) = o(\mu^T \Sigma \mu)$,

$$\mathrm{E}\,\phi(X_1, \ldots, X_n) = \Phi\left(\frac{\sqrt{n}\mu^T \mu}{2\sqrt{\mu^T \Sigma \mu}}\right) + o(1).$$

The condition $\mu^T \Sigma \mu = o\big(n^{-1} \operatorname{tr}(\Sigma^2)\big)$ can be viewed as a high-dimensional version of the local alternative hypothesis, and its opposite, $n^{-1} \operatorname{tr}(\Sigma^2) = o\big(\mu^T \Sigma \mu\big)$, can be viewed as a version of fixed alternative hypothesis. See Chen and Qin (2010). Hence (i) and (ii) of Theorem 3 give the asymptotic power under the local and fixed alternatives, respectively. Compared Theorem 3 with the results given by Chen and Qin (2010), it can be seen that under the condition (8), the proposed randomization test has the same asymptotic power as the asymptotic method. In summary, our theoretical results show that the proposed test has a wider application scope while still has good power behavior compared with the asymptotic method.

4. Simulation Studies

In this section, we report the simulation performance of the proposed randomization method (RM) in various settings. For comparison purposes, we also carried out simulations for the bootstrap method (BM) and the asymptotic method (AM) of Chen and Qin (2010). Throughout the simulation, the randomization method and the bootstrap method are calculated by 1000 resamplings.

We consider three innovation structures: the moving average model, the factor model and the compound symmetry structure. The moving average model has the following structure:

$$X_{ij} = \sum_{l=0}^{k} \rho_l Z_{i,j+l},$$

$i = 1, \ldots, n$, $j = 1, \ldots, p$, where Z_{ij}'s are iid random variables with distribution F for $i = 1, \ldots, n$ and $j = 1, \ldots, p + k$. Like Chen and Qin (2010), we consider two different F. One is $N(0, 1)$, and the other is $(\text{Gamma}(4, 1) - 4)/2$. We also consider different k. The ρ_i's are generated independently from $U(2, 3)$ and are kept fixed throughout the simulation. The second model we consider is the factor model in Fan

et al. (2007). In the simulation study of Fan et al. (2007), the factor model is used to reflect aspects of gene expression data. The model involves three group factor and one common factor among all p variables. We denote by $\{\xi_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq p}$ a sequence of independent $N(0,1)$ random variables and by $\{\chi_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq 4}$ a sequence of independent random variables with distribution $(\chi_6^2 - 6)/\sqrt{12}$. Note that χ_{ij} has mean 0, variance 1 and skewness $\sqrt{12}/3$. The data is generated by model

$$X_{ij} = \frac{a_{j1}\chi_{i1} + a_{j2}\chi_{i2} + a_{j3}\chi_{i3} + b_j\chi_{i4} + \xi_{ij}}{(1 + a_{j1}^2 + a_{j2}^2 + a_{j3}^2 + b_j^2)^{1/2}},$$

$i = 1, \ldots, n$, $j = 1, \ldots, p$, where $a_{jk} = 0$ except that $a_{j1} = a_j$ for $j = 1, \ldots, \frac{1}{3}p$, $a_{j2} = a_j$ for $\frac{1}{3}p + 1, \ldots, \frac{2}{3}p$ and $a_{j3} = a_j$ for $\frac{2}{3}p + 1, \ldots, p$. As in Fan et al. (2007), we consider two configurations of factor loadings. In case I we set $a_j = 0.25$ and $b_j = 0.1$ for $j = 1, \ldots, p$. In case II, a_i and b_i are generated independently from $U(0, 0.4)$ and $U(0, 0.2)$. We also consider the compound symmetry structure:

$$X_i \sim N(0, (1-c)I_p + c1_p 1_p^T),$$

where $0 < c < 1$ and 1_p is the p dimensional vector with all entries one. As shown by Katayama et al. (2013), this structure violates (8).

To control the significant level, the null distribution of a p-value should be close to $U(0, 1)$, the uniform distribution on $(0, 1)$. We simulate the p-value of the randomization method $\tilde{p}(X_1, \ldots, X_n)$, the p-value of the asymptotic method

$$p_{CQ}(X_1, \ldots, X_n) = 1 - \Phi\left(\frac{\sum_{j<i} X_i^T X_j}{\sqrt{\frac{n(n-1)}{2} \widehat{\mathrm{tr}(\Sigma^2)}}}\right),$$

and the bootstrap p-value. Figures 1 and 2 show the empirical distribution function (ECDF) of three p-values. It can be seen that the p-values of RM are uniformly distributed in all cases. In contrast, the uniformity of p-values of AM and BM depends on the innovation structure. The p-value of AM is not uniformly distributed when variables are strongly correlated. This is not surprising since the strongly correlated variables often produce a strongly spiked covariance matrix which exceeds the application scope of AM. In comparison, the p-value of BM is not uniformly distributed when variables are weakly correlated.

Now we simulate the empirical size and power of AM, BM and RM. It can be seen from (i) of Theorem 3 that the local asymptotic power of RM is an increasing function of the following signal to noise ratio (SNR):

$$\mathrm{SNR} = \sqrt{\frac{n(n-1)}{2}} \frac{\mu^T \mu}{\sqrt{\mathrm{tr}(\Sigma^2)}}.$$

We scale μ to reach different levels of SNR. Our simulations consider two mean structures: dense mean and sparse mean. In the dense mean setting, we independently generate each coordinate of μ from $U(2, 3)$ and then scale μ to reach a given SNR. In the sparse mean setting, we randomly select 5% of μ's p coordinates to be non-zero. Each non-zero coordinate is again independently generated from $U(2, 3)$ and then scaled to
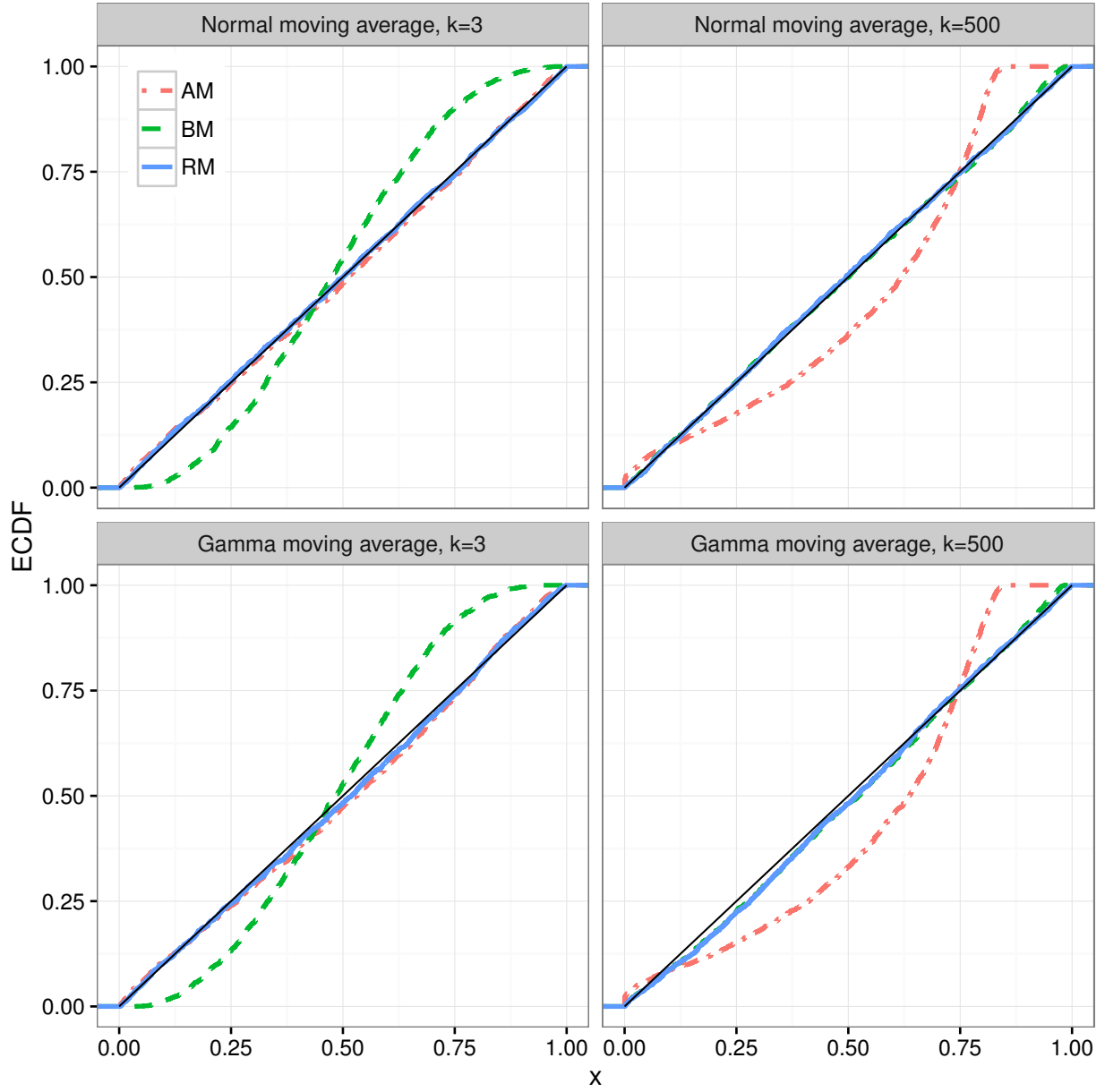
10

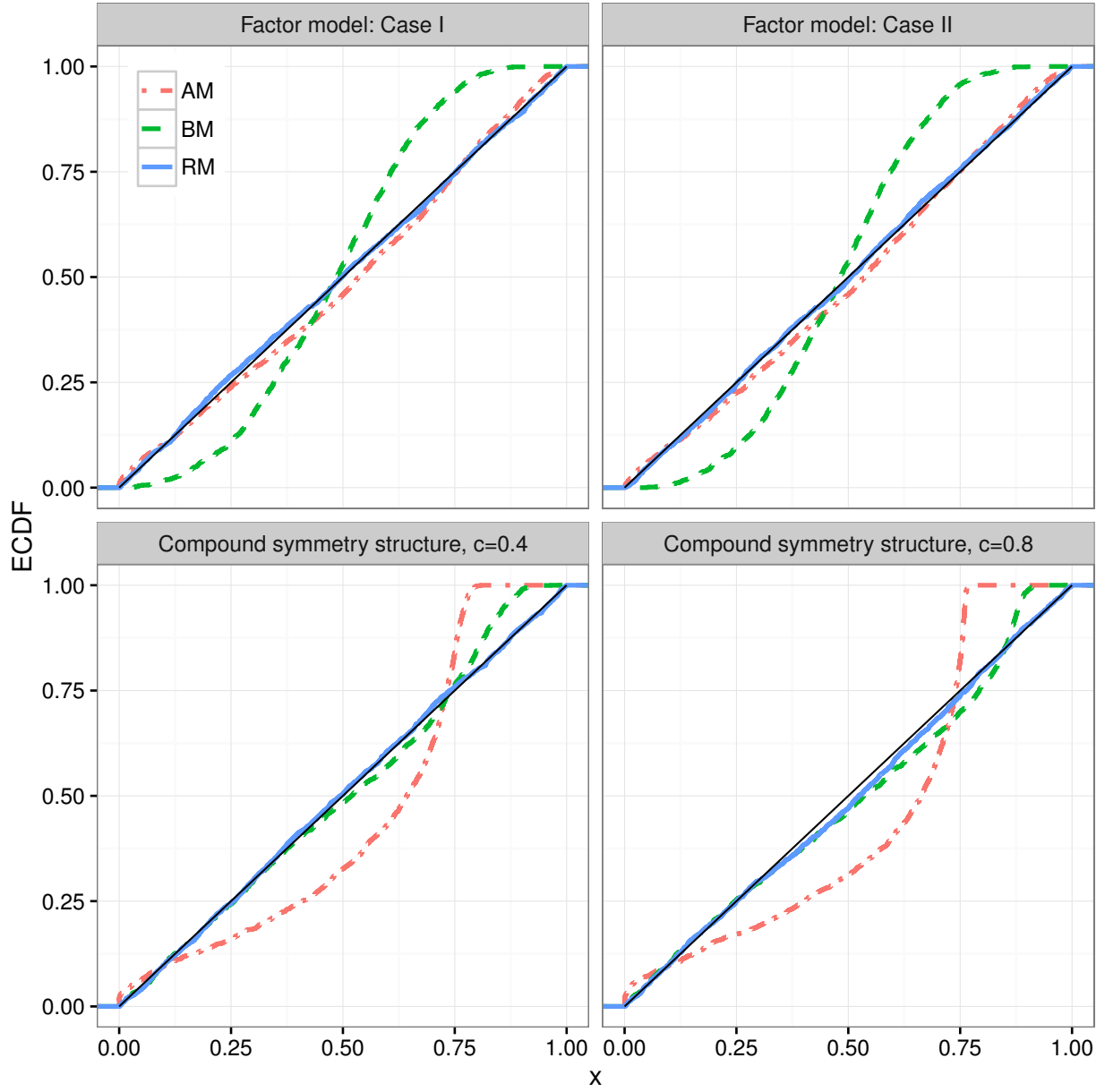Figure 1: The ECDF of 2000 independent p-values of AM, BM and RM. $n = 100$, $p = 600$.

Figure 2: The ECDF of 2000 independent p-values of AM, BM and RM. $n = 100$, $p = 600$.

reach a given SNR. The empirical size and power are computed based on 2000 simulations. The nominal level α is 0.05. Tables 1-4 list the empirical size and power for AM, BM and RM. The results show that RM can control the Type I error rate well even when the distribution of X_1 is not symmetric under the null hypothesis. This justifies the robustness of the randomization method. Also, RM has reasonable power behavior. In contrast, the size of AM is a little inflated when the correlations between variables are strong and the size of BM is too conservative when the correlations between variables are weak.

One may argue that even in the worst case, the asymptotic method exhibits only a slight size distortion (about 2%). For a standard testing problem, such a size distortion may be tolerable. For some advanced applications, however, the test is the basic component of a multiple comparison procedure. See, for example, Section 4 of Chen and Qin (2010). It is shown by Fan et al. (2007) that small size distortion for component p-values may have a great impact on the overall size. In such applications, the proposed test may establish a clear superiority.

Table 1: Empirical size and power of moving average model with normal innovation.

| | | Dense means | | | | | | Sparse means | | | | | |
| | | $k = 3$ | | | $k = 500$ | | | $k = 3$ | | | $k = 500$ | | |
(n, p)	SNR	RM	AM	BM	RM	AM	BM	RM	AM	BM	RM	AM	BM
$(100, 600)$	0 (size)	0.040	0.048	0.004	0.040	0.063	0.043	0.047	0.051	0.002	0.054	0.072	0.056
	0.5	0.187	0.207	0.020	0.141	0.180	0.145	0.171	0.190	0.011	0.079	0.110	0.080
	1	0.405	0.424	0.086	0.244	0.301	0.248	0.376	0.411	0.052	0.148	0.200	0.152
	1.5	0.626	0.652	0.220	0.314	0.377	0.328	0.650	0.672	0.173	0.231	0.324	0.238
	2	0.800	0.819	0.398	0.378	0.438	0.385	0.863	0.880	0.400	0.378	0.542	0.376
	2.5	0.906	0.915	0.612	0.470	0.530	0.476	0.961	0.966	0.642	0.644	0.826	0.640
	3	0.966	0.970	0.792	0.554	0.622	0.562	0.992	0.994	0.846	0.867	0.975	0.858
$(200, 1002)$	0 (size)	0.046	0.052	0.003	0.054	0.076	0.056	0.042	0.048	0.002	0.044	0.064	0.044
	0.5	0.184	0.200	0.028	0.126	0.154	0.129	0.174	0.186	0.024	0.090	0.120	0.088
	1	0.386	0.404	0.102	0.226	0.268	0.230	0.385	0.400	0.074	0.151	0.204	0.152
	1.5	0.656	0.672	0.271	0.308	0.366	0.302	0.669	0.688	0.240	0.240	0.316	0.236
	2	0.824	0.836	0.484	0.406	0.472	0.406	0.860	0.874	0.482	0.395	0.536	0.393
	2.5	0.934	0.939	0.693	0.492	0.552	0.494	0.965	0.971	0.746	0.642	0.810	0.642
	3	0.972	0.974	0.846	0.574	0.634	0.580	0.996	0.996	0.895	0.882	0.984	0.872

5. Concluding Remark

In this paper, we considered a randomization test for mean vector in high dimensional setting. A fast implementation was provided. We also derived some asymptotic properties of the test procedure. Our theoretical results and simulations showed that the proposed test has better performance than the asymptotic

Table 2: Empirical size and power of moving average model with Gamma innovation.

| | | Dense means | | | | | | Sparse means | | | | | |
| | | k = 3 | | | k = 500 | | | k = 3 | | | k = 500 | | |
(n, p)	SNR	RM	AM	BM	RM	AM	BM	RM	AM	BM	RM	AM	BM
(100, 600)	0 (size)	0.050	0.056	0.002	0.048	0.070	0.050	0.046	0.049	0.002	0.048	0.068	0.052
	0.5	0.181	0.194	0.016	0.136	0.183	0.141	0.172	0.186	0.013	0.080	0.112	0.082
	1	0.382	0.408	0.091	0.218	0.268	0.222	0.397	0.420	0.066	0.148	0.210	0.149
	1.5	0.628	0.648	0.198	0.294	0.352	0.300	0.662	0.680	0.194	0.247	0.338	0.248
	2	0.808	0.821	0.418	0.406	0.463	0.411	0.858	0.872	0.397	0.368	0.512	0.372
	2.5	0.910	0.921	0.607	0.479	0.552	0.490	0.966	0.970	0.668	0.624	0.823	0.616
	3	0.970	0.976	0.776	0.548	0.604	0.551	0.994	0.996	0.849	0.882	0.974	0.878
(200, 1002)	0 (size)	0.042	0.046	0.002	0.049	0.074	0.050	0.054	0.058	0.004	0.048	0.068	0.050
	0.5	0.181	0.194	0.024	0.127	0.162	0.128	0.166	0.183	0.018	0.089	0.128	0.086
	1	0.388	0.407	0.098	0.222	0.270	0.222	0.384	0.402	0.082	0.158	0.217	0.158
	1.5	0.634	0.649	0.258	0.327	0.378	0.329	0.668	0.692	0.240	0.227	0.318	0.234
	2	0.828	0.840	0.485	0.412	0.460	0.411	0.851	0.864	0.468	0.392	0.519	0.392
	2.5	0.938	0.942	0.691	0.512	0.564	0.515	0.960	0.967	0.714	0.620	0.792	0.616
	3	0.983	0.984	0.852	0.591	0.658	0.598	0.992	0.994	0.890	0.874	0.980	0.874

method and the bootstrap method. In fact, the algorithm and the proof method can also be applied to other quadratic form statistics.

In classical statistics, randomization test procedures are time consuming. Nevertheless, the computational complexity of our randomization test procedure (which is $O(n^2 B)$) is not affected by the data dimension p. Hence we have reason to believe that randomization tests may be generally suitable for high dimensional problems.

Our randomization test can be immediately generalized to the two sample problem for paired data (Konietschke and Pauly, 2014). However, maybe the most widely used randomization test for two sample problem is the permutation test. As pointed by Romano (1990), the asymptotic property of randomization tests depends heavily on the particular problem and the randomization method. The proof method used in this paper can not be applied to two sample permutation test. It is interesting to understand the behavior of permutation tests in high dimensional setting. We leave it for further work.

Acknowledgments

Table 3: Empirical size and power of factor model innovation.

| | | Dense means | | | | | | | | | Sparse means | | | | | | | |
| | | Case I | | | Case II | | | Case I | | | Case II | | |
(n, p)	SNR	RM	AM	BM	RM	AM	BM	RM	AM	BM	RM	AM	BM
$(100, 600)$	0 (size)	0.055	0.066	0.003	0.054	0.063	0.004	0.048	0.060	0.004	0.046	0.060	0.002
	0.5	0.146	0.166	0.030	0.142	0.160	0.020	0.102	0.128	0.014	0.109	0.128	0.004
	1	0.258	0.288	0.068	0.254	0.282	0.053	0.200	0.235	0.022	0.208	0.237	0.014
	1.5	0.366	0.410	0.138	0.374	0.402	0.090	0.367	0.422	0.050	0.364	0.399	0.039
	2	0.488	0.522	0.186	0.500	0.536	0.162	0.548	0.601	0.086	0.562	0.606	0.060
	2.5	0.571	0.608	0.266	0.605	0.636	0.234	0.723	0.776	0.162	0.744	0.788	0.150
	3	0.642	0.671	0.328	0.664	0.689	0.312	0.857	0.889	0.276	0.866	0.889	0.242
$(200, 1002)$	0 (size)	0.053	0.066	0.012	0.044	0.057	0.004	0.046	0.061	0.012	0.046	0.058	0.008
	0.5	0.134	0.156	0.052	0.142	0.160	0.032	0.096	0.124	0.020	0.107	0.130	0.014
	1	0.264	0.295	0.122	0.248	0.280	0.082	0.182	0.238	0.048	0.202	0.228	0.036
	1.5	0.362	0.406	0.186	0.372	0.410	0.167	0.328	0.392	0.097	0.353	0.412	0.056
	2	0.446	0.482	0.254	0.453	0.484	0.226	0.525	0.600	0.169	0.542	0.604	0.120
	2.5	0.568	0.600	0.356	0.580	0.620	0.329	0.716	0.775	0.286	0.727	0.770	0.236
	3	0.656	0.686	0.438	0.673	0.710	0.396	0.868	0.910	0.452	0.876	0.904	0.392

Appendices

The following notations are used throughout the proofs. Denote by $\Gamma^T \Gamma = Q \Lambda Q^T$ the spectral decomposition of $\Gamma^T \Gamma$, where Q is a $m \times p$ column orthogonal matrix and $\Lambda = \operatorname{diag}(\lambda_1(\Sigma), \ldots, \lambda_p(\Sigma))$. Let $\Lambda_1 = \operatorname{diag}(\lambda_1(\Sigma), \ldots, \lambda_r(\Sigma))$, $\Lambda_2 = \operatorname{diag}(\lambda_{r+1}(\Sigma), \ldots, \lambda_p(\Sigma))$. Write $Q = (Q_1, Q_2)$ where Q_1 and Q_2 are the first r columns and last $p - r$ columns of Q respectively. Then $\Gamma^T \Gamma = Q_1 \Lambda_1 Q_1^T + Q_2 \Lambda_2 Q_2^T$. We denote

$$Y_{ij} = \left(\frac{n(n-1)}{2} \operatorname{tr}(\Sigma^2) \right)^{-1/2} Z_i^T Q_2 \Lambda_2 Q_2^T Z_j, \quad Y_i^* = n^{-1/2} Q_1^T Z_i.$$

Appendix A. Proof of Theorem 1

Lemma 1. *Under* (7), *we have*

(i) *for any $m \times m$ positive semidefinite matrix B,*

$$\mathrm{E}(Z_1^T B Z_1)^2 \le (3 + \Delta) \operatorname{tr}^2(B), \quad \operatorname{Var}(Z_1^T B Z_1) \le (2 + \Delta) \operatorname{tr}(B^2);$$

(ii) *for any $m \times m$ positive semidefinite matrix B,*

$$\mathrm{E}(Z_1^T B Z_2)^4 \le (3 + \Delta)^2 \operatorname{tr}^2(B^2).$$

15

Table 4: Empirical size and power of compound symmetry structure innovation.

		Dense means						Sparse means					
		$c=0.4$			$c=0.8$			$c=0.4$			$c=0.8$		
(n,p)	SNR	RM	AM	BM	RM	AM	BM	RM	AM	BM	RM	AM	BM
$(100,600)$	0 (size)	0.056	0.081	0.058	0.046	0.068	0.047	0.058	0.079	0.057	0.046	0.070	0.053
	0.5	0.136	0.172	0.136	0.130	0.168	0.136	0.075	0.108	0.081	0.084	0.118	0.090
	1	0.222	0.274	0.224	0.235	0.288	0.237	0.118	0.160	0.127	0.118	0.174	0.120
	1.5	0.318	0.375	0.326	0.308	0.370	0.314	0.209	0.304	0.216	0.207	0.300	0.212
	2	0.385	0.449	0.388	0.376	0.438	0.386	0.338	0.505	0.338	0.350	0.526	0.356
	2.5	0.454	0.526	0.459	0.454	0.520	0.462	0.621	0.840	0.618	0.600	0.832	0.592
	3	0.522	0.590	0.525	0.525	0.590	0.528	0.856	0.980	0.849	0.872	0.978	0.862
$(200,1002)$	0 (size)	0.050	0.071	0.050	0.064	0.082	0.064	0.056	0.074	0.058	0.051	0.069	0.050
	0.5	0.126	0.166	0.130	0.130	0.162	0.130	0.080	0.106	0.082	0.074	0.112	0.077
	1	0.214	0.257	0.214	0.221	0.268	0.220	0.136	0.194	0.136	0.128	0.175	0.133
	1.5	0.292	0.344	0.294	0.298	0.359	0.299	0.216	0.300	0.214	0.204	0.275	0.204
	2	0.384	0.450	0.392	0.371	0.436	0.377	0.304	0.466	0.310	0.330	0.502	0.333
	2.5	0.478	0.545	0.484	0.455	0.516	0.464	0.582	0.862	0.586	0.576	0.853	0.574
	3	0.538	0.591	0.533	0.511	0.572	0.509	0.870	0.992	0.876	0.887	0.994	0.894

Proof of Lemma 1. The inequality (i) follows from (i) of Chen et al. (2010, Proposition A.1). The inequality (ii) is obtained by using (i) twice:

$$\mathrm{E}(Z_1^T B Z_2)^4 = \mathrm{E}\,\mathrm{E}\left[(Z_1^T B Z_2 Z_2^T B Z_1)^2 | Z_2\right]$$
$$\leq (3+\Delta)\,\mathrm{E}\,\mathrm{tr}^2(B Z_2 Z_2^T B) = (3+\Delta)\,\mathrm{E}(Z_2^T B^2 Z_2)^2 \leq (3+\Delta)^2\,\mathrm{tr}^2(B^2).$$

$\square$

The following proposition gives the asymptotic distribution of $T(X_1,\ldots,X_n)$ under the null hypothesis.

Proposition 1. *Under* (7) *and* (9), *we have*

$$\frac{\sum_{j<i} Z_i^T \Gamma^T \Gamma Z_j}{\sqrt{\frac{n(n-1)}{2}\,\mathrm{tr}(\Sigma^2)}} \xrightarrow{w} \frac{\sqrt{2}}{2}\sum_{i=1}^r \kappa_i(\xi_i^2 - 1) + (1 - \sum_{i=1}^r \kappa_i^2)^{1/2}\xi_{r+1},$$

where $\xi_1,\ldots\xi_{r+1}$ *are iid standard normal random variables.*

Proof of Proposition 1. Note that

$$\frac{\sum_{j<i} Z_i^T \Gamma^T \Gamma Z_j}{\sqrt{\frac{n(n-1)}{2}\,\mathrm{tr}(\Sigma^2)}} = \frac{\sum_{j<i} Z_i^T Q_1 \Lambda_1 Q_1^T Z_j}{\sqrt{\frac{n(n-1)}{2}\,\mathrm{tr}(\Sigma^2)}} + \frac{\sum_{j<i} Z_i^T Q_2 \Lambda_2 Q_2^T Z_j}{\sqrt{\frac{n(n-1)}{2}\,\mathrm{tr}(\Sigma^2)}}$$
$$= \left(2\frac{n-1}{n}\,\mathrm{tr}(\Sigma^2)\right)^{-1/2}(\sum_{i=1}^n Y_i^*)^T \Lambda_1 \sum_{i=1}^n Y_i^* + \sum_{j<i} Y_{ij} - \left(2\frac{n-1}{n}\,\mathrm{tr}(\Sigma^2)\right)^{-1/2}\sum_{i=1}^n Y_i^{*T}\Lambda_1 Y_i^*.$$

16

To deal with the asymptotic distribution of the first two terms, we would like to show that

$$\Big(\big(\sum_{i=1}^{n} Y_i^*\big)^T, \sum_{j<i} Y_{ij}\Big)^T \xrightarrow{\mathcal{L}} N_{r+1}\Big(0_{r+1}, \operatorname{diag}\big(I_r, 1 - \sum_{i=1}^{r} \kappa_i^2\big)\Big). \tag{A.1}$$

To prove this, it suffices to show that for any nonzero constants $t \in \mathbb{R}^r$ and $c \in \mathbb{R}$,

$$t^T \sum_{i=1}^{n} Y_i^* + c \sum_{j<i} Y_{ij} \xrightarrow{\mathcal{L}} N\big(0, \|t\|^2 + c^2(1 - \sum_{i=1}^{r} \kappa_i^2)\big). \tag{A.2}$$

Write $t^T \sum_{i=1}^{n} Y_i^* + c \sum_{j<i} Y_{ij} = \sum_{i=1}^{n} U_{in}$, where $U_{in} = t^T Y_i^* + c \sum_{j:j<i} Y_{ij}$. Let $\mathcal{F}_{in}$ be the σ-field generated by $X_1, \ldots, X_i$, $i = 1, \ldots, n$. Then $\{U_{in}\}_{i=1}^{n}$ is a martingale difference array with respect to $\{\mathcal{F}_{in}\}_{i=1}^{n}$. By the martingale central limit theorem (Pollard, 1984, Chapter VIII, Theorem 1), (A.2) holds provided

$$\sum_{i=1}^{n} \mathrm{E}(U_{in}^2 | \mathcal{F}_{i-1,n}) \xrightarrow{P} \|t\|^2 + c^2(1 - \sum_{i=1}^{r} \kappa_i^2), \tag{A.3}$$

and for every $\epsilon > 0$,

$$\sum_{i=1}^{n} \mathrm{E}\big(U_{in}^2 \{U_{in}^2 > \epsilon\} | \mathcal{F}_{i-1,n}\big) \xrightarrow{P} 0. \tag{A.4}$$

We first verify (A.3). By direct calculation, we have

$$\sum_{i=1}^{n} \mathrm{E}(U_{in}^2 | \mathcal{F}_{i-1,n}) = \|t\|^2 + c^2 \big(\frac{n(n-1)}{2} \operatorname{tr}(\Sigma^2)\big)^{-1} \sum_{i=1}^{n} \Big(\big(\sum_{j:j<i} Z_j\big)^T Q_2 \Lambda_2^2 Q_2^T \sum_{j:j<i} Z_j\Big).$$

Note that

$$\sum_{i=1}^{n} \Big(\big(\sum_{j:j<i} Z_j\big)^T Q_2 \Lambda_2^2 Q_2^T \sum_{j:j<i} Z_j\Big) = \sum_{i=1}^{n} \Big(\sum_{j:j<i} Z_j^T Q_2 \Lambda_2^2 Q_2^T Z_j + 2 \sum_{j,k:k<j<i} Z_j^T Q_2 \Lambda_2^2 Q_2^T Z_k\Big)$$

$$= \sum_{j=1}^{n} (n-j) Z_j^T Q_2 \Lambda_2^2 Q_2^T Z_j + 2 \sum_{k<j} (n-j) Z_j^T Q_2 \Lambda_2^2 Q_2^T Z_k. \tag{A.5}$$

The first term of (A.5) satisfies

$$\mathrm{E} \sum_{j=1}^{n} (n-j) Z_j^T Q_2 \Lambda_2^2 Q_2^T Z_j = \frac{n(n-1)}{2} \operatorname{tr}(\Lambda_2^2)$$

and

$$\operatorname{Var}\Big(\sum_{j=1}^{n} (n-j) Z_j^T Q_2 \Lambda_2^2 Q_2^T Z_j\Big) = \frac{(n-1)n(2n-1)}{6} \operatorname{Var}\big(Z_1^T Q_2 \Lambda_2^2 Q_2^T Z_1\big) \le (2+\Delta) n^3 \operatorname{tr}(\Lambda_2^4),$$

where the last inequality follows from (i) of Lemma 1. Hence

$$\sum_{j=1}^{n} (n-j) Z_j^T Q_2 \Lambda_2^2 Q_2^T Z_j = \frac{n(n-1)}{2} \operatorname{tr}(\Lambda_2^2) + O_P\big(n^{3/2} \operatorname{tr}^{1/2}(\Lambda_2^4)\big)$$

$$= \big(\frac{n(n-1)}{2} \operatorname{tr}(\Sigma^2)\big) \Big(\frac{\operatorname{tr}(\Lambda_2^2)}{\operatorname{tr}(\Sigma^2)} + o_P(1)\Big) = \big(\frac{n(n-1)}{2} \operatorname{tr}(\Sigma^2)\big) \Big(1 - \sum_{i=1}^{r} \kappa_i^2 + o_P(1)\Big). \tag{A.6}$$

As for the second term of (A.5), we have

$$\mathrm{E}\,\big(\sum_{k<j}(n-j)Z_j^T Q_2\Lambda_2^2 Q_2^T Z_k\big)^2 = \sum_{k<j}(n-j)^2\,\mathrm{E}\,\big(Z_j^T Q_2\Lambda_2^2 Q_2^T Z_k\big)^2 = \sum_{k<j}(n-j)^2\,\mathrm{tr}(\Lambda_2^4)$$

$$\leq n^4\,\mathrm{tr}(\Lambda_2^4)\leq n^4\lambda_{r+1}^2(\Sigma)\,\mathrm{tr}(\Lambda_2^2) = o\Big(\big(\frac{n(n-1)}{2}\,\mathrm{tr}(\Sigma^2)\big)^2\Big),$$

where the last equality follows from the assumption $\lambda_{r+1}(\Sigma)/\sqrt{\mathrm{tr}(\Sigma^2)}\to 0$. Hence

$$\sum_{k<j}(n-j)Z_j^T Q_2\Lambda_2^2 Q_2^T Z_k = o_P\big(\frac{n(n-1)}{2}\,\mathrm{tr}(\Sigma^2)\big). \tag{A.7}$$

Combining (A.5), (A.6) and (A.7) yields

$$\big(\frac{n(n-1)}{2}\,\mathrm{tr}(\Sigma^2)\big)^{-1}\sum_{i=1}^{n}\Big(\big(\sum_{j:j<i}Z_j\big)^T Q_2\Lambda_2^2 Q_2^T\sum_{j:j<i}Z_j\Big) = 1 - \sum_{i=1}^{r}\kappa_i^2 + o_P(1).$$

This proves (A.3).

Now we verify (A.4). By Markov's inequality, it suffices to show $\sum_{i=1}^{n}\mathrm{E}\,\big(U_{in}^4\big)\to 0$. Note that

$$\mathrm{E}\,\big(U_{in}^4\big)\leq 8n^{-2}\,\mathrm{E}\,\big(t^T Q_1^T Z_i\big)^4 + 8c^4\big(\frac{n(n-1)}{2}\,\mathrm{tr}(\Sigma^2)\big)^{-2}\,\mathrm{E}\,\big(Z_i^T Q_2\Lambda_2 Q_2^T\sum_{j:j<i}Z_j\big)^4.$$

Apply (i) of Lemma 1, we have

$$\mathrm{E}\,\big(t^T Q_1^T Z_i\big)^4\leq \mathrm{E}\,\big(Z_i^T Q_1 t t^T Q_1^T Z_i\big)^2\leq (3+\Delta)\,\mathrm{tr}^2(Q_1 t t^T Q_1^T) = (3+\Delta)\|t\|^4.$$

Also by (i) of Lemma 1,

$$\mathrm{E}\,\big(Z_i^T Q_2\Lambda_2 Q_2^T\sum_{j:j<i}Z_j\big)^4 = \mathrm{E}\,\big(Z_i^T Q_2\Lambda_2 Q_2^T(\sum_{j:j<i}Z_j)(\sum_{j:j<i}Z_j)^T Q_2\Lambda_2 Q_2^T Z_i\big)^2$$

$$\leq (3+\Delta)\,\mathrm{E}\,\mathrm{tr}^2\big(Q_2\Lambda_2 Q_2^T(\sum_{j:j<i}Z_j)(\sum_{j:j<i}Z_j)^T Q_2\Lambda_2 Q_2^T\big)$$

$$= (3+\Delta)\,\mathrm{E}\,\Big(\big(\sum_{j:j<i}Z_j\big)^T Q_2\Lambda_2^2 Q_2^T\big(\sum_{j:j<i}Z_j\big)\Big)^2$$

$$= (3+\Delta)\,\mathrm{E}\,\Big(\sum_{j:j<i}Z_j^T Q_2\Lambda_2^2 Q_2^T Z_j + 2\sum_{j,k:k<j<i}Z_j^T Q_2\Lambda_2^2 Q_2^T Z_k\Big)^2$$

$$= (3+\Delta)\,\mathrm{E}\,\Big(\sum_{j:j<i}Z_j^T Q_2\Lambda_2^2 Q_2^T Z_j\Big)^2 + 4(3+\Delta)\,\mathrm{E}\,\Big(\sum_{j,k:k<j<i}Z_j^T Q_2\Lambda_2^2 Q_2^T Z_k\Big)^2$$

$$\leq (3+\Delta)(i-1)^2\,\mathrm{E}(Z_1^T Q_2\Lambda_2^2 Q_2^T Z_1)^2 + 4(3+\Delta)\frac{(i-1)(i-2)}{2}\,\mathrm{E}(Z_2^T Q_2\Lambda_2^2 Q_2^T Z_1)^2$$

$$\leq (3+\Delta)^2(i-1)^2\,\mathrm{tr}^2(\Lambda_2^2) + 2(3+\Delta)(i-1)(i-2)\,\mathrm{tr}(\Lambda_2^4)$$

$$\leq 2(3+\Delta)^2 i^2\,\mathrm{tr}^2(\Lambda_2^2).$$

Thus,

$$\mathrm{E}\,\big(U_{in}^4\big)\leq 8(3+\Delta)\|t\|^4 n^{-2} + 16c^4(3+\Delta)^2 i^2\big(\frac{n(n-1)}{2}\big)^{-2}.$$

18

It follows that $\sum_{i=1}^{n} \mathrm{E}\left(U_{in}^4\right) \to 0$. This completes the proof of (A.1).

It follows from (A.1) and Slutsky's theorem that

$$\left(2\frac{n-1}{n}\operatorname{tr}(\Sigma^2)\right)^{-1/2}\left(\sum_{i=1}^{n} Y_i^*\right)^T \Lambda_1 \sum_{i=1}^{n} Y_i^* + \sum_{j<i} Y_{ij} \xrightarrow{w} \frac{\sqrt{2}}{2}\sum_{i=1}^{r}\kappa_i\xi_i^2 + \left(1-\sum_{i=1}^{r}\kappa_i^2\right)^{1/2}\xi_{r+1}.$$

It remains to show that

$$\left(2\frac{n-1}{n}\operatorname{tr}(\Sigma^2)\right)^{-1/2}\sum_{i=1}^{n} Y_i^{*T}\Lambda_1 Y_i^* \xrightarrow{P} \frac{\sqrt{2}}{2}\sum_{i=1}^{r}\kappa_i.$$

This follows from

$$\mathrm{E}\left(2\frac{n-1}{n}\operatorname{tr}(\Sigma^2)\right)^{-1/2}\sum_{i=1}^{n} Y_i^{*T}\Lambda_1 Y_i^* = \left(2\frac{n-1}{n}\operatorname{tr}(\Sigma^2)\right)^{-1/2}\operatorname{tr}(\Lambda_1) = \frac{\sqrt{2}}{2}\sum_{i=1}^{r}\kappa_i + o(1)$$

and

$$\operatorname{Var}\left(\left(2\frac{n-1}{n}\operatorname{tr}(\Sigma^2)\right)^{-1/2}\sum_{i=1}^{n} Y_i^{*T}\Lambda_1 Y_i^*\right) \le \left(2\frac{n-1}{n}\operatorname{tr}(\Sigma^2)\right)^{-1}\frac{(2+\Delta)\operatorname{tr}(\Lambda_1^2)}{n} = o(1).$$

This completes the proof.

$\square$

Proof of Theorem 1. Note that

$$T(X_1,\ldots,X_n) - \frac{n(n-1)}{2}\|\mu\|^2 = \sum_{j<i} Z_i^T\Gamma^T\Gamma Z_j + (n-1)\sum_{i=1}^{n}\mu^T\Gamma Z_i. \tag{A.8}$$

By central limit theorem,

$$\frac{\sum_{i=1}^{n}\mu^T\Gamma Z_i}{\sqrt{n\mu^T\Sigma\mu}} \xrightarrow{\mathcal{L}} N(0,1). \tag{A.9}$$

If $\mu^T\Sigma\mu = o(n^{-1}\operatorname{tr}(\Sigma^2))$, the first term of (A.8) dominates the second term of (A.8), and the theorem follows from Proposition 1. On the other hand, if $n^{-1}\operatorname{tr}(\Sigma^2) = o(\mu^T\Sigma\mu)$, the second term of (A.8) dominates the first term of (A.8), and the theorem follows from (A.9).

$\square$

Appendix B. Proof of Theorem 2

To obtain the asymptotic property of q_α^*, we need to investigate the asymptotic behavior of the distribution (10). Note that the distribution (10) itself is random. To study its asymptotic behavior, we need to define in what sense the convergence is. Let F and G be two distribution functions on $\mathbb{R}$, Levy metric ρ of F and G is defined as

$$\rho(F,G) = \inf\{\epsilon : F(x-\epsilon) - \epsilon \le G(x) \le F(x+\epsilon) + \epsilon \text{ for all } x\}.$$

It is well known that $\rho(F_n, F) \to 0$ if and only if $F_n \xrightarrow{\mathcal{L}} F$.

A central limit theorem for quadratic form of Rademacher variables. The proof of Theorem 2 is based on a central limit theorem of the quadratic form of Rademacher variables. Such a central limit theorem can be also used to study the asymptotic behavior of many other randomization tests. Let $\epsilon_1, \ldots, \epsilon_n$ be independent Rademacher variables. Consider the quadratic form $W_n = \sum_{1 \leq j < i \leq n} a_{ij} \epsilon_i \epsilon_j + \sum_{i=1}^{n} b_i \epsilon_i$, where $\{a_{ij}\}_{1 \leq j < i \leq n}$ and $\{b_i\}_{1 \leq i \leq n}$ are nonrandom numbers. Here $\{\epsilon_i\}$, $\{a_{ij}\}$ and $\{b_i\}$ may depend on n, a parameter we suppress. By direct calculation, we have $\mathrm{E}(W_n) = 0$ and $\mathrm{Var}(W_n) = \sum_{1 \leq j < i \leq n} a_{ij}^2 + \sum_{i=1}^{n} b_i^2$.

Proposition 2. *A sufficient condition for*

$$\frac{W_n}{\sqrt{\sum_{1 \leq j < i \leq n} a_{ij}^2 + \sum_{i=1}^{n} b_i^2}} \xrightarrow{\mathcal{L}} N(0,1)$$

is that

$$\sum_{k<j}\Big(\sum_{i:i>j} a_{ij} a_{ik}\Big)^2 + \sum_{j<i} a_{ij}^4 + \sum_{k<j<i} a_{ij}^2 a_{ik}^2 + \sum_{j=1}^{n}\Big(\sum_{i:i>j} a_{ij} b_i\Big)^2 + \sum_{i=1}^{n} b_i^4 = o\Big(\Big(\sum_{j<i} a_{ij}^2 + \sum_{i=1}^{n} b_i^2\Big)^2\Big). \qquad \text{(B.1)}$$

Proof of Proposition 2. Note that we have the decomposition $W_n = \sum_{i=1}^{n} U_{in}$, where $U_{in} = \epsilon_i(\sum_{j:j<i} a_{ij} \epsilon_j + b_i)$, $i = 1, \ldots, n$. Let $\mathcal{F}_{in}$ be the σ-field generated by $\epsilon_1, \ldots, \epsilon_i$, $i = 1, \ldots, n$. Then $\{U_{in}\}_{i=1}^{n}$ is a martingale difference array with respect to $\{\mathcal{F}_{in}\}_{i=1}^{n}$. By the martingale central limit theorem (Pollard, 1984, Theorem 1 of Chapter VIII), the conclusion holds if the following two conditions are satisfied:

$$\frac{\sum_{i=1}^{n} \mathrm{E}(U_{in}^2 | \mathcal{F}_{i-1,n})}{\sum_{j<i} a_{ij}^2 + \sum_{i=1}^{n} b_i^2} \xrightarrow{P} 1, \qquad \text{(B.2)}$$

and

$$\frac{\sum_{i=1}^{n} \mathrm{E}\left(U_{in}^2 \{U_{in}^2 > \epsilon(\sum_{j<i} a_{ij}^2 + \sum_{i=1}^{n} b_i^2)\} \big| \mathcal{F}_{i-1,n}\right)}{\sum_{j<i} a_{ij}^2 + \sum_{i=1}^{n} b_i^2} \xrightarrow{P} 0, \qquad \text{(B.3)}$$

for every $\epsilon > 0$.

We first verify (B.2). By direct calculation, we have

$$\sum_{i=1}^{n} \mathrm{E}(U_{in}^2 | \mathcal{F}_{i-1,n}) = \sum_{i=1}^{n} \Big(\sum_{j:j<i} a_{ij} \epsilon_j + b_i\Big)^2$$

$$= \sum_{j<i} a_{ij}^2 + \sum_{i=1}^{n} b_i^2 + 2 \sum_{k<j} \sum_{i:i>j} (a_{ij} a_{ik}) \epsilon_j \epsilon_k + 2 \sum_{j=1}^{n} \Big(\sum_{i:i>j} a_{ij} b_i\Big) \epsilon_j.$$

By assumption, we have

$$\mathrm{E}\Big(\sum_{k<j}\big(\sum_{i:i>j} a_{ij} a_{ik}\big) \epsilon_j \epsilon_k\Big)^2 = \sum_{k<j}\Big(\sum_{i:i>j} a_{ij} a_{ik}\Big)^2 = o\Big(\big(\sum_{j<i} a_{ij}^2 + \sum_{i=1}^{n} b_i^2\big)^2\Big)$$

and

$$\mathrm{E}\Big(\sum_{j=1}^{n}\big(\sum_{i:i>j} a_{ij} b_i\big) \epsilon_j\Big)^2 = \sum_{j=1}^{n}\Big(\sum_{i:i>j} a_{ij} b_i\Big)^2 = o\Big(\big(\sum_{j<i} a_{ij}^2 + \sum_{i=1}^{n} b_i^2\big)^2\Big).$$

Hence (B.2) holds.

20

Now we verify (B.3). By Markov inequality, it's sufficient to prove

$$\frac{\sum_{i=1}^n \mathrm{E}\left(U_{in}^4\big|\mathcal{F}_{i-1,n}\right)}{\left(\sum_{j<i} a_{ij}^2 + \sum_{i=1}^n b_i^2\right)^2} \xrightarrow{P} 0. \tag{B.4}$$

Since the relevant random variables are all positive, we only need to prove (B.4) converges to 0 in mean. But

$$\sum_{i=1}^n \mathrm{E}\, U_{in}^4 = \sum_{i=1}^n \mathrm{E}(\sum_{j:j<i} a_{ij}\epsilon_j + b_i)^4 \leq 8 \sum_{i=1}^n \mathrm{E}(\sum_{j:j<i} a_{ij}\epsilon_j)^4 + 8 \sum_{i=1}^n b_i^4.$$

By assumption, we have

$$\sum_{i=1}^n b_i^4 = o\Big(\big(\sum_{j<i} a_{ij}^2 + \sum_{i=1}^n b_i^2\big)^2\Big)$$

and

$$\sum_{i=1}^n \mathrm{E}(\sum_{j:j<i} a_{ij}\epsilon_j)^4 = \sum_{i=1}^n \mathrm{E}\,\big(\sum_{j:j<i} a_{ij}^2 + 2 \sum_{j,k:k<j<i} a_{ij}a_{ik}\epsilon_j\epsilon_k\big)^2$$

$$= \sum_{i=1}^n \big((\sum_{j:j<i} a_{ij}^2)^2 + 4\,\mathrm{E}(\sum_{j,k:k<j<i} a_{ij}a_{ik}\epsilon_j\epsilon_k)^2\big) = \sum_{i=1}^n(\sum_{j:j<i} a_{ij}^4 + 6 \sum_{j,k:k<j<i} a_{ij}^2 a_{ik}^2)$$

$$= \sum_{j<i} a_{ij}^4 + 6 \sum_{k<j<i} a_{ij}^2 a_{ik}^2 = o\Big(\big(\sum_{j<i} a_{ij}^2 + \sum_{i=1}^n b_i^2\big)^2\Big).$$

Hence (B.3) holds. This completes the proof. $\qquad\square$

Proposition 3. *Under (7) and (9), for nonzero constants $c \in \mathbb{R}$ and $t \in \mathbb{R}^r$, we have*

$$\sum_{k<j}(\sum_{i:i>j} Y_{ij}Y_{ik})^2 + \sum_{j<i} Y_{ij}^4 + \sum_{k<j<i} Y_{ij}^2 Y_{ik}^2 + \sum_{j=1}^n \big(\sum_{i:i>j} Y_{ij}t^T Y_i^*\big)^2 + \sum_{i=1}^n (t^T Y_i^*)^4 \xrightarrow{P} 0, \tag{B.5}$$

$$c^2 \sum_{j<i} Y_{ij}^2 + \sum_{i=1}^n (t^T Y_i^*)^2 = c^2(1 - \sum_{i=1}^r \kappa_i^2) + \|t\|^2 + o_P(1). \tag{B.6}$$

Proof of Proposition 3. To prove (B.5), we only need to show that the left hand side of (B.5) converges to 0 in mean. For the first three terms of (B.5), we have

$$\mathrm{E}\sum_{k<j}\big(\sum_{i:i>j} Y_{ij}Y_{ik}\big)^2 + \mathrm{E}\sum_{j<i} Y_{ij}^4 + \mathrm{E}\sum_{k<j<i} Y_{ij}^2 Y_{ik}^2$$

$$= \mathrm{E}\sum_{k<j}\Big(\sum_{i:i>j} Y_{ij}^2 Y_{ik}^2 + 2\sum_{i_1,i_2:i_1>i_2>j} Y_{i_1 j}Y_{i_1 k}Y_{i_2 j}Y_{i_2 k}\Big) + \mathrm{E}\sum_{j<i} Y_{ij}^4 + \mathrm{E}\sum_{k<j<i} Y_{ij}^2 Y_{ik}^2$$

$$= 2\,\mathrm{E}\sum_{k<j<i_2<i_1} Y_{i_1 j}Y_{i_1 k}Y_{i_2 j}Y_{i_2 k} + 2\,\mathrm{E}\sum_{k<j<i} Y_{ij}^2 Y_{ik}^2 + \mathrm{E}\sum_{j<i} Y_{ij}^4$$

$$= \frac{(n-2)(n-3)\,\mathrm{tr}(\Lambda_2^4)}{3n(n-1)\,\mathrm{tr}^2(\Sigma^2)} + \frac{n(n-1)(n-2)}{3}\,\mathrm{E}(Y_{31}^2 Y_{32}^2) + \frac{n(n-1)}{2}\,\mathrm{E}\, Y_{21}^4$$

$$\leq \frac{(n-2)(n-3)\lambda_{r+1}(\Sigma)\,\mathrm{tr}(\Lambda_2^2)}{3n(n-1)\,\mathrm{tr}^2(\Sigma^2)} + \Big(\frac{n(n-1)(n-2)}{3} + \frac{n(n-1)}{2}\Big)\,\mathrm{E}\, Y_{21}^4$$

$$\leq \frac{(n-2)(n-3)\lambda_{r+1}(\Sigma)\,\mathrm{tr}(\Lambda_2^2)}{3n(n-1)\,\mathrm{tr}^2(\Sigma^2)} + \Big(\frac{n(n-1)(n-2)}{3} + \frac{n(n-1)}{2}\Big)\frac{4(3+\Delta)^2\,\mathrm{tr}^2(\Lambda_2^2)}{n^2(n-1)^2\,\mathrm{tr}^2(\Sigma^2)} \to 0,$$

21

where the last inequality follows from (ii) of Lemma 1 and the assumption $\lambda_{r+1}(\Sigma)/\sqrt{\operatorname{tr}(\Sigma^2)} \to 0$. As for the fourth term of (B.5), we have

$$
\begin{aligned}
\mathrm{E}\sum_{j=1}^{n}\Big(\sum_{i:i>j}Y_{ij}t^TY_i^*\Big)^2 &= \mathrm{E}\sum_{j<i}Y_{ij}^2(t^TY_i^*)^2 = \frac{n(n-1)}{2}\,\mathrm{E}\,Y_{21}^2(t^TY_2^*)^2\\
&=\frac{n(n-1)}{2}\,\mathrm{E}\,\mathrm{E}\,\big(Y_{21}^2(t^TY_2^*)^2\big|Z_2\big) = \big(n\operatorname{tr}(\Sigma^2)\big)^{-1}\,\mathrm{E}\,\big(Z_2^TQ_2\Lambda_2^2Q_2^TZ_2Z_2^TQ_1tt^TQ_1^TZ_2\big)\\
&\le\big(n\operatorname{tr}(\Sigma^2)\big)^{-1}\big(\mathrm{E}(Z_2^TQ_2\Lambda_2^2Q_2^TZ_2)^2\big)^{1/2}\big(\mathrm{E}(Z_2^TQ_1tt^TQ_1^TZ_2)^2\big)^{1/2}\\
&\le\frac{(3+\Delta)\|t\|^2}{n}\frac{\operatorname{tr}(\Lambda_2^2)}{\operatorname{tr}(\Sigma^2)}\to0,
\end{aligned}
$$

where the last inequality follows from (i) of Lemma 1. Also by (i) of Lemma 1, the last term of (B.5) satisfies

$$
\mathrm{E}\sum_{i=1}^{n}(t^TY_i^*)^4 = n\,\mathrm{E}(t^TY_1^*)^4 = \frac{1}{n}\,\mathrm{E}(Z_1^TQ_1tt^TQ_1^TZ_1)^2 \le \frac{(3+\Delta)\|t\|^4}{n}\to0.
$$

This completes the proof of (B.5).

To prove (B.6), we show that

$$
\sum_{j<i}Y_{ij}^2 = (1-\sum_{i=1}^{r}\kappa_i^2)+o_P(1), \tag{B.7}
$$

$$
\sum_{i=1}^{n}(t^TY_i^*)^2 = \|t\|^2+o_P(1). \tag{B.8}
$$

To prove (B.7), note that

$$
\mathrm{E}\sum_{j<i}Y_{ij}^2 = \frac{\operatorname{tr}(\Lambda_2^2)}{\operatorname{tr}(\Sigma^2)} = (1-\sum_{i=1}^{r}\kappa_i^2)+o(1).
$$

Hence we only need to show $\operatorname{Var}\big(\sum_{j<i}Y_{ij}^2\big)=o(1)$. Write $\big(\sum_{j<i}Y_{ij}^2\big)^2 = \big(\sum_{j<i}Y_{ij}^2\big)\big(\sum_{l<k}Y_{kl}^2\big)$. Collect terms according to $\operatorname{card}(\{i,j\}\cap\{k,l\})=0,1,2$, we have

$$
\Big(\sum_{j<i}Y_{ij}^2\Big)^2 = \sum_{j<i,l<k:\{i,j\}\cap\{k,l\}=\phi}Y_{ij}^2Y_{kl}^2 + 2\sum_{j<i<k}\Big(Y_{ij}^2Y_{ki}^2+Y_{ij}^2Y_{kj}^2+Y_{kj}^2Y_{ki}^2\Big) + \sum_{j<i}Y_{ij}^4.
$$

There are $n(n-1)(n-2)(n-3)/4$, $n(n-1)(n-2)/6$ and $n(n-1)/2$ terms in each sum, respectively. For distinct i,j,k,l,

$$
\mathrm{E}\,Y_{ij}^2Y_{kl}^2 = (\mathrm{E}\,Y_{21}^2)^2 = \frac{4\operatorname{tr}^2(\Lambda_2^2)}{n^2(n-1)^2\operatorname{tr}^2(\Sigma^2)}.
$$

Otherwise, if $\operatorname{card}(\{i,j\},\{k,l\})\ge1$, by (ii) of Lemma 1,

$$
\mathrm{E}\,Y_{ij}^2Y_{kl}^2 \le \mathrm{E}\,Y_{21}^4 \le \frac{4(3+\Delta)^2\operatorname{tr}^2(\Lambda_2^2)}{n^2(n-1)^2\operatorname{tr}^2(\Sigma^2)}.
$$

Thus,

$$
\begin{aligned}
\mathrm{E}\Big(\sum_{j<i}Y_{ij}^2\Big)^2 &= \frac{(n-2)(n-3)}{n(n-1)}\frac{\operatorname{tr}^2(\Lambda_2^2)}{\operatorname{tr}^2(\Sigma^2)} + O(1)\big(n(n-1)(n-2)+\frac{n(n-1)}{2}\big)\frac{4(3+\Delta)^2\operatorname{tr}^2(\Lambda_2^2)}{n^2(n-1)^2\operatorname{tr}^2(\Sigma^2)}\\
&=(1-\sum_{i=1}^{r}\kappa_i^2)^2+o(1).
\end{aligned}
$$

22

It follows that $\text{Var}\left(\sum_{j<i} Y_{ij}^2\right) = \text{E}\left(\sum_{j<i} Y_{ij}^2\right)^2 - \left(\text{E}\sum_{j<i} Y_{ij}^2\right)^2 = o(1)$, which completes the proof of (B.7).

Finally, (B.8) follows from $\text{E}\sum_{i=1}^n (t^T Y_i^*)^2 = \|t\|^2$ and

$$\text{Var}\left(\sum_{i=1}^n (t^T Y_i^*)^2\right) = \frac{1}{n}\text{Var}(Z_1^T Q_1 t t^T Q_1^T Z_1) \le \frac{1}{n}(2+\Delta)\|t\|^4 \to 0,$$

where the last inequality follows from (i) of Lemma 1. This completes the proof.

$\square$

The following proposition gives the asymptotic behavior of (10) under the null hypothesis.

Proposition 4. *Under* (6), (7), (9), *we have*

$$\rho\left(\mathcal{L}\left(\frac{\sum_{j<i} Z_i^T \Gamma^T \Gamma Z_j \epsilon_i \epsilon_j}{\sqrt{\frac{n(n-1)}{2}\text{tr}(\Sigma^2)}}\Big| Z_1,\ldots,Z_n\right), \mathcal{L}\left(\frac{\sqrt{2}}{2}\sum_{i=1}^r \kappa_i(\xi_i^2-1) + (1-\sum_{i=1}^r \kappa_i^2)^{1/2}\xi_{r+1}\right)\right) \xrightarrow{P} 0,$$

where $\xi_1,\ldots,\xi_{r+1}$ *are iid standard normal random variables.*

Proof of Proposition 4. Note that

$$\frac{\sum_{j<i} Z_i^T \Gamma^T \Gamma Z_j \epsilon_i \epsilon_j}{\sqrt{\frac{n(n-1)}{2}\text{tr}(\Sigma^2)}} = \frac{\sum_{j<i} Z_i^T Q_1 \Lambda_1 Q_1^T Z_j \epsilon_i \epsilon_j}{\sqrt{\frac{n(n-1)}{2}\text{tr}(\Sigma^2)}} + \frac{\sum_{j<i} Z_i^T Q_2 \Lambda_2 Q_2^T Z_j \epsilon_i \epsilon_j}{\sqrt{\frac{n(n-1)}{2}\text{tr}(\Sigma^2)}}$$

$$= \left(2\frac{n-1}{n}\text{tr}(\Sigma^2)\right)^{-1/2}\left(\left(\sum_{i=1}^n Y_i^* \epsilon_i\right)^T \Lambda_1 \sum_{i=1}^n Y_i^* \epsilon_i + \sum_{j<i} Y_{ij}\epsilon_i\epsilon_j\right) - \left(2\frac{n-1}{n}\text{tr}(\Sigma^2)\right)^{-1/2}\sum_{i=1}^n Y_i^{*T}\Lambda_1 Y_i^*.$$

To deal with the asymptotic distribution of the first two terms, we would like to show that

$$\rho\left(\mathcal{L}\left(\left(\left(\sum_{i=1}^n Y_i^* \epsilon_i\right)^T, \sum_{j<i} Y_{ij}\epsilon_i\epsilon_j\right)^T \Big| Z_1,\ldots,Z_n\right), N_{r+1}\left(0_{r+1}, \text{diag}(I_r, 1-\sum_{i=1}^r \kappa_i^2)\right)\right) \xrightarrow{P} 0. \tag{B.9}$$

To prove this, it suffices to show that for every nonzero constants $t \in \mathbb{R}^r$ and $c \in \mathbb{R}$,

$$\rho\left(\mathcal{L}\left(\sum_{i=1}^n t^T Y_i^* \epsilon_i + c\sum_{j<i} Y_{ij}\epsilon_i\epsilon_j \Big| Z_1,\ldots,Z_n\right), N\left(0, \|t\|^2 + c^2(1-\sum_{i=1}^r \kappa_i^2)\right)\right) \xrightarrow{P} 0.$$

We apply Proposition 2 to prove this. Let $a_{ij} = cY_{ij}$, $b_i = t^T Y_i^*$. For every subsequence of $\{n\}$, Proposition 3 implies that there is a further subsequence along which

$$\sum_{j<i} a_{ij}^2 + \sum_{i=1}^n b_i^2 \xrightarrow{a.s.} \|t\|^2 + c^2(1-\sum_{i=1}^r \kappa_i^2) \tag{B.10}$$

and (B.1) holds almost surely. Then it follows from Proposition 2, (B.10) and Slutsky's theorem that

$$\rho\left(\mathcal{L}\left(\frac{\sum_{i=1}^n t^T Y_i^* \epsilon_i + c\sum_{j<i} Y_{ij}\epsilon_i\epsilon_j}{\sqrt{\|t\|^2 + c^2(1-\sum_{i=1}^r \kappa_i^r)}}\Big| Z_1,\ldots,Z_n\right), N(0,1)\right) \xrightarrow{a.s.} 0$$

along this further subsequence. That is to say,

$$\rho\left(\mathcal{L}\left(\frac{\sum_{i=1}^n t^T Y_i^* \epsilon_i + c\sum_{j<i} Y_{ij}\epsilon_i\epsilon_j}{\sqrt{\|t\|^2 + c^2(1-\sum_{i=1}^r \kappa_i^r)}}\Big| Z_1,\ldots,Z_n\right), N(0,1)\right) \xrightarrow{P} 0.$$

23

Hence (B.9) holds. Then Slutsky's theorem implies that

$$\rho\Big(\mathcal{L}\big((2\frac{n-1}{n}\operatorname{tr}(\Sigma^2))^{-1/2}(\sum_{i=1}^{n}Y_i^*\epsilon_i)^T\Lambda_1\sum_{i=1}^{n}Y_i^*\epsilon_i+\sum_{j<i}Y_{ij}\epsilon_i\epsilon_j|Z_1,\ldots,Z_n),$$

$$\mathcal{L}\big(\frac{\sqrt{2}}{2}\sum_{i=1}^{r}\kappa_i\xi_i^2+(1-\sum_{i=1}^{r}\kappa_i^2)^{1/2}\xi_{r+1}\big)\Big)\xrightarrow{P}0. \tag{B.11}$$

We have shown in the proof of Proposition 1 that

$$\big(2\frac{n-1}{n}\operatorname{tr}(\Sigma^2)\big)^{-1/2}\sum_{i=1}^{n}Y_i^{*T}\Lambda_1 Y_i^*\xrightarrow{P}\frac{\sqrt{2}}{2}\sum_{i=1}^{r}\kappa_i. \tag{B.12}$$

Then the conclusion follows from (B.12), (B.11) and Slutsky's theorem.

$\square$

Proof of Theorem 2. Note that

$$\sum_{j<i}X_i^T X_j\epsilon_i\epsilon_j=\sum_{j<i}Z_i^T\Gamma^T\Gamma Z_j\epsilon_i\epsilon_j+\sum_{j<i}\mu^T\Gamma Z_i\epsilon_i\epsilon_j+\sum_{j<i}\mu^T\Gamma Z_j\epsilon_i\epsilon_j+\mu^T\mu\sum_{j<i}\epsilon_i\epsilon_j$$

$$\overset{def}{=}C_1+C_2+C_3+C_4.$$

The asymptotic behavior of C_1 is given by Proposition 4. Note that C_4 is independent of $X_1,\ldots,X_n$. By central limit theorem and Slutsky's theorem,

$$\frac{C_4}{\sqrt{\frac{n(n-1)}{2}(\mu^T\mu)^2}}=\frac{\sqrt{2}}{2}\sqrt{\frac{n}{n-1}}\Big((\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i)^2-1\Big)\xrightarrow{\mathcal{L}}\frac{\sqrt{2}}{2}(\chi_1^2-1).$$

Suppose $\mu^T\mu=o(\sqrt{\operatorname{tr}(\Sigma^2)})$. It can be seen that

$$\frac{C_4}{\sqrt{\frac{n(n-1)}{2}\operatorname{tr}(\Sigma^2)}}\xrightarrow{P}0.$$

The cross terms satisfy

$$\mathrm{E}(C_2^2)=\mathrm{E}(C_3^2)=\frac{n(n-1)}{2}\mu^T\Sigma\mu\leq\frac{n(n-1)}{2}\sqrt{\operatorname{tr}(\Sigma^2)}\mu^T\mu=o\Big(\frac{n(n-1)}{2}\operatorname{tr}(\Sigma^2)\Big).$$

By Markov's inequality, we have

$$\mathrm{E}\left(\left(\frac{C_i}{\sqrt{\frac{n(n-1)}{2}\operatorname{tr}(\Sigma^2)}}\right)^2\Big|X_1,\ldots,X_n\right)\xrightarrow{P}0,\quad i=2,3.$$

Then Proposition 4 and a standard subsequence argument yields

$$\rho\left(\mathcal{L}\Big(\frac{T(\epsilon_1 X_1,\ldots,\epsilon_n X_n)}{\sqrt{\frac{n(n-1)}{2}\operatorname{tr}(\Sigma^2)}}\Big|X_1,\ldots,X_n\Big),\mathcal{L}\big(\frac{\sqrt{2}}{2}\sum_{i=1}^{r}\kappa_i(\xi_i^2-1)+(1-\sum_{i=1}^{r}\kappa_i^2)^{1/2}\xi_{r+1}\big)\right)\xrightarrow{P}0.$$

And (i) of Theorem 2 follows.

If $\sqrt{\operatorname{tr}(\Sigma^2)}=o(\mu^T\mu)$, it can be similarly shown that C_4 is the dominating term and

$$\rho\left(\mathcal{L}\Big(\frac{T(\epsilon_1 X_1,\ldots,\epsilon_n X_n)}{\sqrt{\frac{n(n-1)}{2}(\mu^T\mu)^2}}\Big|X_1,\ldots,X_n\Big),\frac{\sqrt{2}}{2}(\chi_1^2-1)\right)\xrightarrow{P}0.$$

This proves (ii) of Theorem 2.

24

Appendix C. Proof of Theorem 3

Lemma 2. *Suppose $\{\eta_n\}_{n=1}^{\infty}$ is a sequence of random variables, weakly converges to η, a random variable with continuous distribution function. Then we have*

$$\sup_{x \in \mathbb{R}} |\Pr(\eta_n \leq x) - \Pr(\eta \leq x)| \to 0.$$

Proof of Theorem 3. By a standard subsequence argument, we can without loss of generality and assume

$$\frac{\mu^T \mu}{\sqrt{\operatorname{tr}(\Sigma^2)}} \to c \in [0, +\infty].$$

First we prove (i) of Theorem 3. Suppose $\mu^T \Sigma \mu = o\big(n^{-1} \operatorname{tr}(\Sigma^2)\big)$. Note that

$$\operatorname{E} \phi(X_1, \ldots, X_n) = \Pr\left(\frac{T(X_1, \ldots, X_n)}{\sqrt{\frac{n(n-1)}{2} \operatorname{tr}(\Sigma^2)}} > q_\alpha^*\right)$$

$$= \Pr\left(\frac{T(X_1, \ldots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\frac{n(n-1)}{2} \operatorname{tr}(\Sigma^2)}} + F^{*-1}(1-\alpha) - q_\alpha^* > F^{*-1}(1-\alpha) - \sqrt{\frac{n(n-1)}{2}} \frac{\mu^T \mu}{\sqrt{\operatorname{tr}(\Sigma^2)}}\right). \quad \text{(C.1)}$$

If $c = 0$, then the conclusion follows from (i) of Theorem 1, (i) of Theorem 2, Slutsky's theorem and Lemma 2. If $c > 0$, (i) of Theorem 2 can not be applied. Hence we use Markov's inequality to bound q_α^*. For any $t > 0$,

$$\Pr\left(\frac{T(\epsilon_1 X_1, \ldots, \epsilon_n X_n)}{\sqrt{\frac{n(n-1)}{2} \operatorname{tr}(\Sigma^2)}} > t \,\Big|\, X_1, \ldots, X_n\right)$$

$$\leq \frac{2}{n(n-1) \operatorname{tr}(\Sigma^2) t^2} \operatorname{E}\big(T^2(\epsilon_1 X_1, \ldots, \epsilon_n X_n) \big| X_1, \ldots, X_n\big) = \frac{2 \sum_{j<i} (X_i^T X_j)^2}{n(n-1) \operatorname{tr}(\Sigma^2) t^2}.$$

We take

$$t = \sqrt{\frac{2 \sum_{j<i} (X_i^T X_j)^2}{n(n-1) \operatorname{tr}(\Sigma^2) \alpha}},$$

then

$$\Pr\left(\frac{T(\epsilon_1 X_1, \ldots, \epsilon_n X_n)}{\sqrt{\frac{n(n-1)}{2} \operatorname{tr}(\Sigma^2)}} > \sqrt{\frac{2 \sum_{j<i} (X_i^T X_j)^2}{n(n-1) \operatorname{tr}(\Sigma^2) \alpha}} \,\Big|\, X_1, \ldots, X_n\right) \leq \alpha.$$

This implies

$$q_\alpha^* \leq \sqrt{\frac{2 \sum_{j<i} (X_i^T X_j)^2}{n(n-1) \operatorname{tr}(\Sigma^2) \alpha}}.$$

Then

$$\operatorname{E}(q_\alpha^{*2}) \leq \frac{\operatorname{tr}(\Sigma + \mu \mu^T)^2}{\operatorname{tr}(\Sigma^2) \alpha} \leq 2 \alpha^{-1}\left(1 + \frac{(\mu^T \mu)^2}{\operatorname{tr}(\Sigma^2)}\right).$$

Hence

$$|q_\alpha^*| = O_P\left(1 + \frac{\mu^T \mu}{\sqrt{\operatorname{tr}(\Sigma^2)}}\right). \quad \text{(C.2)}$$

Therefore, the probability (C.1) equals

$$\Pr\left(O_P(1) + O_P\left(1 + \frac{\mu^T \mu}{\sqrt{\operatorname{tr}(\Sigma^2)}}\right) > O_P(1) - \sqrt{\frac{n(n-1)}{2}} \frac{\mu^T \mu}{\sqrt{\operatorname{tr}(\Sigma^2)}}\right) \to 1.$$

This completes the proof of (i) of Theorem 3.

Now we prove (ii) of Theorem 3. Note that

$$\Pr\left(\frac{T(X_1,\ldots,X_n)}{\sqrt{\frac{n(n-1)}{2}\operatorname{tr}(\Sigma^2)}} > q_\alpha^*\right)$$

$$= \Pr\left(\frac{T(X_1,\ldots,X_n) - \frac{n(n-1)}{2}\mu^T\mu}{(n-1)\sqrt{n\mu^T\Sigma\mu}} > \sqrt{\frac{\operatorname{tr}(\Sigma^2)}{2(n-1)\mu^T\Sigma\mu}}q_\alpha^* - \frac{\sqrt{n}\mu^T\mu}{2\sqrt{\mu^T\Sigma\mu}}\right).$$

If $c = 0$, then it follows from $n^{-1}\operatorname{tr}(\Sigma^2) = o(\mu^T\Sigma\mu)$ and (C.2) that

$$\sqrt{\frac{\operatorname{tr}(\Sigma^2)}{2(n-1)\mu^T\Sigma\mu}}q_\alpha^* = o_P(1).$$

If $c > 0$, then (C.2) implies

$$\sqrt{\frac{\operatorname{tr}(\Sigma^2)}{2(n-1)\mu^T\Sigma\mu}}q_\alpha^* \Big/ \frac{\sqrt{n}\mu^T\mu}{2\sqrt{\mu^T\Sigma\mu}} = \sqrt{\frac{2}{n(n-1)}}\frac{\sqrt{\operatorname{tr}(\Sigma^2)}}{\mu^T\mu}q_\alpha^* = O_P\Big(\frac{1}{n}\Big(\frac{\sqrt{\operatorname{tr}(\Sigma^2)}}{\mu^T\mu}+1\Big)\Big) = o_P(1).$$

Then the conclusion follows from (ii) of Theorem 1, Slutsky's theorem and Lemma 2. $\qquad\square$

References

Anderson, T.W., 2003. An Introduction to Multivariate Statistical Analysis. 3rd ed., Wiley, New York.

Bai, Z., Saranadasa, H., 1996. Effect of high dimension: by an example of a two sample problem. Statistica Sinica 6, 311–329.

Chen, S.X., Qin, Y.L., 2010. A two-sample test for high-dimensional data with applications to gene-set testing. The Annals of Statistics 38, 808–835.

Chen, S.X., Zhang, L.X., Zhong, P.S., 2010. Tests for high-dimensional covariance matrices. Journal of the American Statistical Association 105, 810–819.

Chung, E., Romano, J.P., 2016. Multivariate and multiple permutation tests. Journal of Econometrics 193, 76–91.

Efron, B., Tibshirani, R., 2007. On testing the significance of sets of genes. The Annals of Applied Statistics 1, 107–129.

Fan, J., Hall, P., Yao, Q., 2007. To how many simultaneous hypothesis tests can normal, student's t or bootstrap calibration be applied? Journal of the American Statistical Association 102, 1282–1288.

Fisher, R.A., 1935. The design of experiments. Oliver and Boyd, Edinburgh.

Katayama, S., Kano, Y., Srivastava, M.S., 2013. Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis 116, 410 – 421.

Ko, H., Kim, K., Sun, H., 2016. Multiple group testing procedures for analysis of high-dimensional genomic data. Genomics & Informatics 14, 187.

Konietschke, F., Pauly, M., 2014. Bootstrapping and permuting paired t-test type statistics. Statistics and Computing 24, 283–296.

Lehmann, E.L., Romano, J.P., 2005. Testing Statistical Hypotheses. 3rd ed., Springer New York.

Ma, Y., Lan, W., Wang, H., 2015. A high dimensional two-sample test under a low dimensional factor structure. Journal of Multivariate Analysis 140, 162 – 170.

Pollard, D., 1984. Convergence of Stochastic Processes. 1st ed., Springer New York, New York.

Romano, J.P., 1990. On the behavior of randomization tests without a group invariance assumption. Journal of the American Statistical Association 85, 686–692.

Srivastava, M.S., Du, M., 2008. A test for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis 99, 386–402.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences 102, 15545–15550.

Wang, L., Peng, B., Li, R., 2015. A high-dimensional nonparametric multivariate test for mean vector. Journal of the American Statistical Association 110, 1658–1669.

Zhu, L.X., 2000. Nonparametric monte carlo tests for multivariate distributions. Biometrika 87, 919–928.