

# A randomization test for mean vector in high dimension

Rui Wang<sup>a</sup>, Xingzhong Xu<sup>a,b,\*</sup>

<sup>a</sup>*School of Mathematics and Statistics, Beijing Institute of Technology, Beijing  
100081, China*

<sup>b</sup>*Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing  
100081, China*

---

## Abstract

The strength of randomization tests is that the size of the test is exact under certain symmetry assumption for distribution. Researchers have justified the robustness of the randomization tests when symmetry assumption breaks down and showed they have good power performance in fixed  $p$  situation. It's a theoretical and practical interest for the behavior of randomization tests in high dimensional setting. In this paper, we study the asymptotic behavior of a randomization test for testing high dimensional means. The robustness is justified and asymptotic power function is derived.

An important issue in high dimensional statistics is the computational efficiency. In classical statistics, randomization tests suffer from computational complexity, which limits their adoption. Surprisingly, in our problem the randomization test can be implemented even faster than asymptotic method. With good theoretical properties and fast algorithm, randomization test can be recommended in practice.

*Keywords:* `elsarticle.cls`, L<sup>A</sup>T<sub>E</sub>X, Elsevier, template

*2010 MSC:* 00-01, 99-00

---

---

<sup>\*</sup>Fully documented templates are available in the elsarticle package on CTAN.

<sup>\*</sup>Corresponding author

*Email address:* `xuxz@bit.edu.cn` (Xingzhong Xu)

## 1. Introduction

Consider i.i.d. random sample  $X_1, \dots, X_n \in \mathbb{R}^p$  which has means  $\mu = (\mu_1, \dots, \mu_p)^T$  and covariance matrix  $\Sigma$ . The one sample testing problem

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

has been extensively studied by many researchers. A classical test statistic is Hotelling's  $T^2$ ,  $T_1 = n\bar{X}^T S^{-1} \bar{X}$ , where  $\bar{X}$  and  $S = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$  are the sample mean vector and sample covariance matrix, respectively.

5 However, the high-dimensional case  $d > n$  invalidates the Hotelling's test. [1] is the first paper modifying Hotelling's  $T^2$  statistic for testing (1) in high dimensional setting. This seminal paper removed  $S^{-1}$  from  $T_1$  and proposed a test statistic based on  $T_2 = \bar{X}^T \bar{X}$ . Many subsequent papers relaxed the assumptions and generalized the idea of [1]. For example, [2] proposed a test based on  
10  $T_3 = \bar{X}^T [\text{diag}(S)]^{-1} \bar{X}$ , where  $\text{diag}(S)$  is a matrix with diagonal elements equal to that of  $S$  and off-diagonal elements equal to 0. [3] proposed a test based on  $T_4 = \sum_{i \neq j} X_i^T X_j$ . [4] proposed a test based on  $T_5 = \sum_{i \neq j} Y_i^T Y_j$ , where  $Y_i$  is defined as  $Y_i = X_i / \|X_i\|$  if  $X_i \neq 0$ ; and  $Y_i = 0$  if  $X_i = 0$ .

Note that the test statistics  $T_2$ – $T_5$  all can be written as a generalized quadratic  
15 form of data, see [5], and their theoretical proves mostly rely on an application of martingale central limit theorem (MCLT).

The critical value of existing high dimensional tests are often determined by asymptotical distribution. We will call it asymptotic method. Asymptotic method can guarantee the test level asymptotically. However, in many real  
20 world problems, e.g., gene testing (see [6]), sample size  $n$  may be very small. Hence the Type I error rates of Asymptotic method may be far away from nominal levels in this case.

The idea of randomization test dates back to [7], which is a tool to determine the critical value for a given test statistic. [8] described a general construction  
25 of the randomization test. In high dimension setting, randomization test is widely used in applied statistics, see [6]. It's strength is in that the resulting

test procedure has exact level under mild condition. Although there are many papers give theoretical analysis for fixed  $p$  case, to the best of our knowledge, there's no existing theoretical work concerning high dimension setting.

Suppose the distribution of  $X_i$  is symmetric about 0 under the null hypothesis.  $T$  is certain test statistic for testing (1). Let  $\epsilon_1, \dots, \epsilon_n$  be i.i.d. Rademacher variables ( $\Pr(\epsilon_i = 1) = \Pr(\epsilon_i = -1) = 1/2$ ) which are independent of the data. The conditional distribution

$$\mathcal{L}(T(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n) | X_1, \dots, X_n) \quad (2)$$

is the uniform distribution on  $2^n$  values. The critical value of the randomization test is defined as the  $1 - \alpha$  quantile of the above distribution. More specifically, the test function equals to 1 or 0 if  $T(X_1, \dots, X_n)$  is greater or not greater than the critical value. The resulting test is a level  $\alpha$  test if the distribution of  $X_i$  is symmetric under null. By some refinement of the test function when  $T(X_1, \dots, X_n)$  equals to the critical value, the test is exact, see [8]. Since such extreme case occurs with little probability  $2^{-n}$ , the refinement is often dropped in practice, which only losses a little power.

Equivalently, the randomization test can be implemented by  $p$ -value. Define

$$p(X_1, \dots, X_n) = \Pr(T(\epsilon_1 X_1, \dots, \epsilon_n X_n) \geq T(X_1, \dots, X_n) | X_1, \dots, X_n). \quad (3)$$

The randomization test rejects the null hypothesis if  $p(X_1, \dots, X_n) \leq \alpha$ .

It's easy to see that the randomization version of [1]'s test and [3]'s test are both equivalent to the randomization test based on

$$T(X_1, \dots, X_n) = \sum_{j < i} X_i^T X_j. \quad (4)$$

In this paper, we will study this special statistic for illustration. From now on,  $T(X_1, \dots, X_n)$  will refer to (4). Other quadratic based statistic can be studied by similar method.

Our results show that even if the null distribution is not symmetric, the randomization test is still asymptotically exact under mild assumptions. We also give the local asymptotic power.

45 In fixed  $p$  setting, it's well known that randomization test has much higher time complexity than asymptotic method, which historically hampered it's use. Surprisingly, in high dimension setting the randomization test can be implemented as efficiently as asymptotic method.

Maybe the most widely used randomization method is the two sample permutation test. As [8] pointed out, the asymptotic property of randomization tests depends heavily on the particular problem and, the two sample case is quite distinct from the one sample case. The method used in this paper can not be applied to permutation test.

## 2. Randomization test

We assume, like [3] and [1], the following multivariate model:

$$X_i = \mu + \Gamma Z_i \quad \text{for } i = 1, \dots, n, \quad (5)$$

where  $\Gamma$  is a  $p \times m$  matrix for some  $m \geq p$  such that  $\Gamma_i \Gamma_i^T = \Sigma_i$  and  $\{Z_i\}_{i=1}^n$  are  $m$ -variate i.i.d. random vectors satisfying  $E(Z_i) = 0$  and  $\text{Var}(Z_i) = I_m$ , the  $m \times m$  identity matrix. Write  $Z_i = (z_{i1}, \dots, z_{im})^T$ , we assume  $E(z_{ij}^4) = 3 + \Delta < \infty$  and

$$E(z_{il_1}^{\alpha_1} z_{il_2}^{\alpha_2} \dots z_{il_q}^{\alpha_q}) = E(z_{il_1}^{\alpha_1}) E(z_{il_2}^{\alpha_2}) \dots E(z_{il_q}^{\alpha_q}) \quad (6)$$

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

In the following,  $T(X_1, \dots, X_n)$  will be specialized to (4).

An important assumption in [3] is  $\text{tr}(\Sigma^4) = o\{\text{tr}^2(\Sigma^2)\}$ , which is equivalent to

$$\frac{\lambda_{\max}(\Sigma)}{\sqrt{\text{tr} \Sigma^2}} \rightarrow 0. \quad (7)$$

Although Chen and Qin's results is for two sample case, their results can be proved similarly for one sample case. We restate their theorems:

**Theorem 1.** *Under (5), (6), (7) and local alternatives*

$$\mu^T \Sigma \mu = o(n^{-1} \text{tr} \Sigma^2), \quad (8)$$

we have

$$\frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\frac{n(n-1)}{2} \text{tr} \Sigma^2}} \xrightarrow{\mathcal{L}} N(0, 1). \quad (9)$$

And

**Theorem 2.** Under (5), (6), (7) and

$$n^{-1} \text{tr}(\Sigma)^2 = o(\mu^T \Sigma \mu), \quad (10)$$

we have

$$\frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} \xrightarrow{\mathcal{L}} N(0, 1). \quad (11)$$

We will call the conditional distribution

$$\mathcal{L}\left(\frac{T(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \middle| X_1, \dots, X_n\right)$$

the randomization distribution. Let  $\xi_\alpha^*$  be the  $1-\alpha$  quantile of the randomization distribution. Then the test function  $\phi(X_1, \dots, X_n)$  of randomization test equals to 1 when

$$\frac{T(X_1, \dots, X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^*$$

and equals to 0 otherwise. Since  $\xi_\alpha^*$  relies on data, the rejection region is determined by not only  $T(X_1, \dots, X_n)$  but also randomization distribution.

To study the asymptotic property of  $\xi_\alpha^*$ , we need to derive the asymptotic behavior of randomization distribution. Since the randomization distribution itself is random, we need to define in what sense the convergence is. Let  $F$  and  $G$  be two distribution function on  $\mathbb{R}$  and the Levy metric  $\rho$  of  $F$  and  $G$  is defined as

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

It's well known that  $\rho(F_n, F) \rightarrow 0$  if and only if  $F_n \xrightarrow{\mathcal{L}} F$ . Our first result is:

**Theorem 3.** Under (5), (6), (7) and

$$\mu^T \mu = o(\sqrt{\text{tr} \Sigma^2}), \quad (12)$$

we have that

$$\rho\left(\mathcal{L}\left(\frac{T_2(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \middle| X_1, \dots, X_n\right), N(0, 1)\right) \xrightarrow{P} 0.$$

Under Theorem 3's conditions, the randomization distribution is asymptotically a normal distribution. It's also interesting to understand the behavior of randomization distribution when condition (12) is not valid. We have the following asymptotic result:

**Theorem 4.** Under (5), (6), (7) and

$$\sqrt{\text{tr}\Sigma^2} = o(\mu^T \mu), \quad (13)$$

we have

$$\rho\left(\mathcal{L}\left(\frac{T(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \middle| X_1, \dots, X_n\right), \frac{\sqrt{2}}{2}(\chi_1^2 - 1)\right) \xrightarrow{P} 0. \quad (14)$$

Once the asymptotic distribution of the randomization distribution is obtained, the asymptotic behavior of  $\xi_\alpha^*$  can be derived immediately. Let  $\Phi(\cdot)$  be the cumulative distribution function (CDF) of standard normal distribution, we have

**Corollary 1.** Under the conditions of Theorem 3,  $\xi_\alpha^* \xrightarrow{P} \Phi(1 - \alpha)$ .

**Corollary 2.** Under the conditions of Theorem 4,  $\xi_\alpha^* \xrightarrow{P} \frac{\sqrt{2}}{2}(\Phi^{-1}(1 - \frac{\alpha}{2}) - 1)$ .

Now we are ready to derive the asymptotic power of randomization test. Since the limit property of  $T$  is different under (8) and (10). The following two theorems give the power under (8) and (10), separately.

**Theorem 5.** Suppose conditions (5), (6), (7) and (8) holds.

If (12) holds,

$$\Pr\left(\frac{T(X_1, \dots, X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^*\right) = \Phi(-\Phi^{-1}(1 - \alpha) + \frac{\sqrt{n(n-1)}\mu^T \mu}{\sqrt{2\text{tr}\Sigma^2}}) + o(1).$$

If (13) holds,

$$\Pr \left( \frac{T(X_1, \dots, X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^* \right) \rightarrow 1.$$

**Theorem 6.** Under (5), (6), (7) (10) and either (12) or (13),

$$\Pr \left( \frac{T(X_1, \dots, X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^* \right) = \Phi \left( \frac{\sqrt{n} \mu^T \mu}{2 \sqrt{\mu^T \Sigma \mu}} \right) + o(1),$$

**Remark 1.** The theorem doesn't assume that the distribution of  $X_i$  is symmetric under null. Hence Theorem 5 shows that the level of randomization  
80 test is robust against asymmetry.

**Remark 2.** Neither (8) or (12) implies the other one. For example, suppose  $\Sigma = I_p$ , then (8) is equivalent to  $\mu^T \mu = o(p/n)$  and (12) is equivalent to  $\mu^T \mu = o(\sqrt{p})$ . In this case, if  $\sqrt{p}/n \rightarrow 0$ , then (8) implies (12); conversely, if  $\sqrt{p}/n \rightarrow \infty$ , then (12) implies (8).

The randomization test rejects the null when

$$T(X_1, \dots, X_n) > \sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2} \xi_\alpha^*.$$

The asymptotic method rejects the null when

$$T(X_1, \dots, X_n) > \sqrt{\frac{n(n-1)}{2} \text{tr} \Sigma^2 \Phi^{-1}(1-\alpha)}.$$

Note that the larger the reject region, the more powerful the test is. Thus we compare  $\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2} \xi_\alpha^*$  and  $\sqrt{\frac{n(n-1)}{2} \text{tr} \hat{\Sigma}^2 \Phi^{-1}(1-\alpha)}$ . Suppose  $\text{tr} \hat{\Sigma}^2$  is a ratio consistent estimator of  $\text{tr} \Sigma^2$ . By Lemma 5 in appendix,

$$\frac{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2} \xi_\alpha^*}{\sqrt{\frac{n(n-1)}{2} \text{tr} \hat{\Sigma}^2 \Phi^{-1}(1-\alpha)}} = (1 + o_P(1)) \frac{\sqrt{\text{tr}(\Sigma + \mu \mu^T)^2} \xi_\alpha^*}{\sqrt{\text{tr} \Sigma^2 \Phi^{-1}(1-\alpha)}},$$

85 which tends to 1 when (12) holds, and tends to  $\infty$  when (13) holds. Hence randomization may loss some power.

### 3. Algorithm

Because the randomized test statistic is (conditionally) uniformly distributed on  $2^n$  different values, the quantile  $\xi_\alpha^*$  is not computationally feasible even when  $n$  is moderate. In practice, randomization test is often realized through an approximation of  $p$ -value (3). More specifically, we sample  $\epsilon_1^*, \dots, \epsilon_n^*$  and compute  $T^* = T(\epsilon_1^* X_1, \dots, \epsilon_n^* X_n)$ . Repeat  $B$  times for a large  $B$  and we obtain  $T_i^*$ ,  $i = 1, \dots, B$ . Let

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

It can be shown

$$\Pr(\tilde{p}(X_1, \dots, X_n) \leq u) \leq u \quad \text{for all } 0 \leq u \leq 1,$$

and as  $B \rightarrow \infty$ ,  $\tilde{p}(X_1, \dots, X_n) - p(X_1, \dots, X_n) \xrightarrow{P} 0$  where  $p(X_1, \dots, X_n)$  is defined by (3) (see [9] page 636).

Note that

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

90 We need to compute  $T^*$  for  $B$  times, which is seemingly time consuming. Nevertheless,  $X_i^T X_j$  ( $1 \leq j < i \leq n$ ) can be computed in advance, which has complexity  $O(n^2 p)$  and may be the most time consuming part in large  $p$  small  $n$  case. Once we have obtained  $X_i^T X_j$ , the computation of  $T_i^*$  has only complexity  $O(n^2)$ . The total complexity is thus  $O(n^2 p + n^2 B)$ . The randomiza-  
95 tion procedure only add minor complexity when  $p$  is large. The randomization method doesn't need an estimator of variance, which is a must need in asymptotic method. A good variance estimator is complicated in form, see [3], which add much time complexity to asymptotic method. Hence the randomization method is very competitive compared to asymptotic method even in the sense  
100 of computation efficiency. This is different from low dimension setting where randomization is a lot slower than asymptotic method.



If we only care about the decision (reject or accept) and the  $p$ -value is not needed, the computation efficiency of randomization method can be further improved. In fact, the rejection region is  $\tilde{p}(X_1, \dots, X_n) \leq \alpha$  or

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

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

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

we can accept the null once the sum of left hand side exceeds the right hand side. The algorithm 1 summarizes our previous argument.

#### 4. Simulation studies

105 In this section, we report the simulation performance of the randomization test in various settings. We compare randomization method with asymptotic method. Asymptotic method needs a ratio consistent estimator of  $\text{tr}\Sigma^2$ . However, a good estimator is often complicated and thus time-consuming. For simplicity, throughout the simulation the asymptotic method is implemented by  
 110 using true  $\text{tr}\Sigma^2$ , which will in principle perform better. The empirical power and size are computed based on 1000 simulations.

Let  $c = \sqrt{n(n-1)}\mu^T\mu/\sqrt{2\text{tr}\Sigma^2}$  be the signal to noise ratio (SNR). The theoretic asymptotic power is an increasing function of SNR. Throughout the simulation, we will scale  $\mu$  to reach different level of SNR. Our simulation  
 115 consider two mean structure: dense mean and sparse mean. In the dense mean setting, each coordinate of  $\mu$  is independently generated from  $U(2, 3)$  and then  $\mu$  is scaled to reach a given SNR. In the sparse mean setting, we randomly select 5% of  $\mu$  coordinates to be non-zero. Each non-zero coordinate is again independently generated from  $U(2, 3)$  and then scaled to reach a given SNR.

120 We consider two innovation structure: the moving average model and the factor model.

---

**Algorithm 1** Randomization Algorithm

---

**Require:**  $\alpha, B$

**for**  $1 \leq j < i \leq n$  **do**

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

**end for**

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

Set  $A \leftarrow 0$ .

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

Generate  $\epsilon_1, \dots, \epsilon_n$  according to  $\Pr(\epsilon_i = 1) = \Pr(\epsilon_i = -1) = \frac{1}{2}$ .

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

$$A \leftarrow A + 1$$

**if**  $A > (B + 1)\alpha - 1$  **then**

**return** Accept

**end if**

**else**

**if**  $i - A \geq B - (B + 1)\alpha + 1$  **then**

**return** Reject

**end if**

**end if**

**end for**

---

We first conduct the moving average model. In model (5), we set  $m = p + k$  and  $(\Gamma)_{ij} = \rho_{j-i}$  for  $j = i, \dots, i + k$  and  $(\Gamma)_{ij} = 0$  otherwise. More precisely,

$$X_{ij} = \sum_{l=0}^k \rho_l Z_{i,j+l} \quad (15)$$

for  $i = 1, \dots, n$  and  $j = 1, \dots, p$ . Here  $Z_{ij}$ 's are i.i.d. random variables with distribution  $F$  for  $i = 1, \dots, n$  and  $j = 1, \dots, p + k$ . Like [3], we consider two different  $F$ . One is  $N(0, 1)$ , and the other is  $(\text{Gamma}(4, 1) - 4)/2$ . We set  $k = 5$  and  $k = p$ . The  $\rho_i$ 's are generated independently from  $U(2, 3)$  and kept fixed throughout the simulation. Table 1 and Table 2 list the empirical power and size for the moving average model. It's not surprising that randomization test can control level well in normal case since it can be proved in theory. The results also show that the randomization method can control level well even in Gamma case, which is not symmetric under null. It justify the robustness of randomization method. On the other hand, asymptotic method has small size when dependence is weak and has inflated size when dependence is strong. Randomization method also shares similar power with asymptotic method.

In the simulation study of [10], data are generated from a factor model to reflect aspects of gene expression data. The model involves three group factor and one common factor among all  $p$  variables. Their data generation mechanism is adopted in our next simulation study. We denote by  $\{\epsilon_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq p}$  a sequence of independent  $N(0, 1)$  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  $\chi_{ij}$  has mean 0, variance 1 and skewness  $\sqrt{12}/3$ . The data is generated by model

$$X_{i,j} = \frac{a_{j1}\chi_{i1} + a_{j2}\chi_{i2} + a_{j3}\chi_{i3} + b_j\chi_{i4} + \epsilon_{ij}}{(1 + a_{j1}^2 + a_{j2}^2 + a_{j3}^2 + b_j^2)^{1/2}} \quad i = 1, \dots, n \text{ and } j = 1, \dots, p,$$

where  $a_{jk} = 0$  except that  $a_{j1} = a_j$  for  $j = 1, \dots, \frac{1}{3}p$ ,  $a_{j2} = a_j$  for  $\frac{1}{3}p + 1, \dots, \frac{2}{3}p$  and  $a_{j3} = a_j$  for  $\frac{2}{3}p + 1, \dots, p$ . As in [10], we consider two configurations of factor loadings. In case I we set  $a_j = 0.25$  and  $b_j = 0.1$  for  $j = 1, \dots, p$  and in case II  $a_i$  and  $b_i$  are generated independently from  $U(0, 0.4)$  and  $U(0, 0.2)$ . Table 3 list the simulation results. Although the distribution is not symmetric, the

Table 1: Empirical power and size of moving average model with normal innovation.  $p = 600$ ,  $n = 100$ . TP is the theoretical asymptotic power, RM represents the randomization method, AM represents the asymptotic method.

SNR	TP	Dense means				Sparse means			
		$k = 3$		$k = 500$		$k = 3$		$k = 500$	
		RM	AM	RM	AM	RM	AM	RM	AM
0.0	0.050	0.040	0.006	0.042	0.065	0.048	0.015	0.047	0.059
0.5	0.126	0.190	0.069	0.125	0.169	0.146	0.063	0.081	0.111
1.0	0.260	0.429	0.235	0.218	0.262	0.361	0.175	0.145	0.193
1.5	0.442	0.629	0.425	0.314	0.384	0.633	0.413	0.229	0.313
2.0	0.639	0.795	0.632	0.399	0.458	0.817	0.652	0.415	0.524
2.5	0.804	0.923	0.828	0.469	0.529	0.960	0.861	0.645	0.839
3.0	0.912	0.966	0.919	0.582	0.647	0.998	0.975	0.819	0.999

results show that the level of randomization method is robust while asymptotic  
140 method suffers from level inflation.

## 5. Conclusion remark

## 6. Appendix

*CLT for quadratic form of Rademacher variables.* The proof of the Theorem 3  
is based on a CLT of the quadratic form of Rademacher variables. Such a CLT  
145 can be also used to study the asymptotic behavior of many other randomization  
test. Let  $\epsilon_1, \dots, \epsilon_n$  be indepent Rademacher variables. Consider quadratic form  
 $W_n = \sum_{1 \leq j < i \leq n} a_{ij} \epsilon_i \epsilon_j$ , where  $\{a_{ij}\}$  are nonrandom numbers. Here  $\{\epsilon_i\}$  and  
 $\{a_{ij}\}$  may depend on  $n$ , a parameter we suppress. Obviously,  $E(W_n) = 0$  and  
 $\text{Var}(W_n) = \sum_{1 \leq j < i \leq n} a_{ij}^2$ .

**Proposition 1.** *A sufficient condition for*

$$\frac{W_n}{\sqrt{\sum_{1 \leq j < i \leq n} a_{ij}^2}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (16)$$

Table 2: Empirical power and size of moving average model with Gamma innovation.  $p = 600$ ,  $n = 100$ . TP is the theoretical asymptotic power, RM represents the randomization method, AM represents the asymptotic method.

SNR	TP	Dense means				Sparse means			
		$k = 3$		$k = 500$		$k = 3$		$k = 500$	
		RM	AM	RM	AM	RM	AM	RM	AM
0.0	0.050	0.050	0.016	0.043	0.060	0.033	0.005	0.050	0.069
0.5	0.126	0.211	0.077	0.128	0.160	0.159	0.062	0.065	0.098
1.0	0.260	0.410	0.228	0.255	0.308	0.386	0.159	0.151	0.188
1.5	0.442	0.609	0.418	0.318	0.375	0.643	0.410	0.231	0.321
2.0	0.639	0.799	0.634	0.392	0.449	0.875	0.685	0.373	0.501
2.5	0.804	0.912	0.809	0.496	0.550	0.961	0.858	0.589	0.828
3.0	0.912	0.963	0.908	0.543	0.596	0.985	0.967	0.843	1.000

is that

$$\sum_{j < k} \left( \sum_{i: i > k} a_{ij} a_{ik} \right)^2 + \sum_{j < i} a_{ij}^4 + \sum_{j < k < i} a_{ij}^2 a_{ik}^2 = o\left(\left(\sum_{j < i} a_{ij}^2\right)^2\right). \quad (17)$$

*Proof.* Define  $U_{in} = \epsilon_i \sum_{j=1}^{i-1} a_{ij} \epsilon_j$ ,  $i = 2, \dots, n$ , and  $\mathcal{F}_{in} = \sigma\{\epsilon_1, \dots, \epsilon_i\}$ ,  $i = 1, \dots, n$ . Now  $W_n = \sum_{i=2}^n U_{in}$ ,  $\{U_{in}\}$  is a martingale difference array with respect to  $\{\mathcal{F}_{in}\}$ . To prove the proposition, we shall verify two conditions (See [11]):

$$\frac{\sum_{i=2}^n \mathbb{E}(U_{in}^2 | \mathcal{F}_{i-1,n})}{\sum_{1 \leq j < i \leq n} a_{ij}^2} \xrightarrow{P} 1, \quad (18)$$

and

$$\frac{\sum_{i=2}^n \mathbb{E}(U_{in}^2 \{U_{in}^2 > \epsilon \sum_{1 \leq j < i \leq n} a_{ij}^2\} | \mathcal{F}_{i-1,n})}{\sum_{1 \leq j < i \leq n} a_{ij}^2} \xrightarrow{P} 0, \quad (19)$$

150 for every  $\epsilon > 0$ .

Table 3: Empirical power and size of factor model innovation.  $p = 600$ ,  $n = 100$ . TP is the theoretical asymptotic power, RM represents the randomization method, AM represents the asymptotic method.

SNR	TP	Dense means				Sparse means			
		Case I		Case II		Case I		Case II	
		RM	AM	RM	AM	RM	AM	RM	AM
0.0	0.050	0.040	0.050	0.046	0.057	0.052	0.061	0.047	0.060
0.5	0.126	0.141	0.166	0.136	0.154	0.104	0.126	0.105	0.129
1.0	0.260	0.222	0.256	0.233	0.266	0.207	0.239	0.215	0.240
1.5	0.442	0.363	0.411	0.369	0.392	0.367	0.411	0.373	0.415
2.0	0.639	0.460	0.501	0.470	0.509	0.517	0.581	0.532	0.589
2.5	0.804	0.582	0.613	0.565	0.607	0.718	0.757	0.728	0.780
3.0	0.912	0.646	0.678	0.673	0.701	0.834	0.889	0.870	0.900

*Proof of (18).* Since  $E(U_{in}^2 | \mathcal{F}_{i-1,n}) = (\sum_{j=1}^{i-1} a_{ij} \epsilon_j)^2$ , we have

$$\begin{aligned}
\sum_{i=2}^n E(U_{in}^2 | \mathcal{F}_{i-1,n}) &= \sum_{i=2}^n \left( \sum_{j=1}^{i-1} a_{ij} \epsilon_j \right)^2 = \sum_{i=2}^n \left( \sum_{j=1}^{i-1} a_{ij}^2 + 2 \sum_{j,k:j < k < i} a_{ij} a_{ik} \epsilon_j \epsilon_k \right) \\
&= \sum_{i=2}^n \sum_{j=1}^{i-1} a_{ij}^2 + 2 \sum_{j < k < i} a_{ij} a_{ik} \epsilon_j \epsilon_k.
\end{aligned}$$

But

$$\begin{aligned}
E\left(\sum_{j < k < i} a_{ij} a_{ik} \epsilon_j \epsilon_k\right)^2 &= E\left(\sum_{j < k} \left(\sum_{i:i > k} a_{ij} a_{ik}\right) \epsilon_j \epsilon_k\right)^2 \\
&= \sum_{j < k} \left(\sum_{i:i > k} a_{ij} a_{ik}\right)^2 = o\left(\left(\sum_{j < i} a_{ij}^2\right)^2\right),
\end{aligned}$$

where the last equality holds by assumption. Hence (18) holds.

*Proof of (19).* By Markov's inequality, we only need to prove

$$\frac{\sum_{i=2}^n E(U_{in}^4 | \mathcal{F}_{i-1,n})}{\left(\sum_{1 \leq j < i \leq n} a_{ij}^2\right)^2} \xrightarrow{P} 0. \quad (20)$$

Since the relevant random variables are all positive, we only need to prove (20) converges to 0 in mean. But

$$\begin{aligned}
\sum_{i=2}^n \mathbb{E} U_{in}^4 &= \sum_{i=2}^n \mathbb{E} \left( \sum_{j:j < i} a_{ij} \epsilon_j \right)^4 = \sum_{i=2}^n \mathbb{E} \left( \sum_{j:j < i} a_{ij}^2 + 2 \sum_{j,k:j < k < i} a_{ij} a_{ik} \epsilon_j \epsilon_k \right)^2 \\
&= \sum_{i=2}^n \left( \left( \sum_{j:j < i} a_{ij}^2 \right)^2 + 4 \mathbb{E} \left( \sum_{j,k:j < k < i} a_{ij} a_{ik} \epsilon_j \epsilon_k \right)^2 \right) \\
&= \sum_{i=2}^n \left( \sum_{j:j < i} a_{ij}^4 + 6 \sum_{j,k:j < k < i} a_{ij}^2 a_{ik}^2 \right) \\
&= \sum_{j < i} a_{ij}^4 + 6 \sum_{j < k < i} a_{ij}^2 a_{ik}^2 = o \left( \left( \sum_{j < i} a_{ij}^2 \right)^2 \right),
\end{aligned}$$

where the last equality holds by assumption. Hence (19) holds.  $\square$

## 7. Asymptotic normality

**Lemma 1.** *Suppose  $\{\eta_n\}$  is a sequence of 1-dimensional random variables, weakly converges to  $\eta$ , a random variable with continuous distribution function. Then we have*

$$\sup_x |\Pr(\eta_n \leq x) - \Pr(\eta \leq x)| \rightarrow 0.$$

**Lemma 2.** *Suppose  $A = (a_{ij})$  is an  $m \times m$  positive semi-definite matrix. Under (6), we have*

$$\mathbb{E}(Z_i^T A Z_i)^2 \asymp (\text{tr} A)^2 \quad (21)$$

*Proof.* Notice that

$$\begin{aligned}
(Z_i^T A Z_i)^2 &= \left( \sum_{j=1}^m a_{jj} z_{ij}^2 + 2 \sum_{k < j} a_{jk} z_{ij} z_{ik} \right)^2 \\
&= \left( \sum_{j=1}^m a_{jj} z_{ij}^2 \right)^2 + 4 \left( \sum_{j=1}^m a_{jj} z_{ij}^2 \right) \left( \sum_{k < j} a_{jk} z_{ij} z_{ik} \right) + 4 \left( \sum_{k < j} a_{jk} z_{ij} z_{ik} \right)^2 \\
&= \sum_{j=1}^m a_{jj}^2 z_{ij}^4 + 2 \sum_{k < j} a_{jj} a_{kk} z_{ij}^2 z_{ik}^2 + 4 \left( \sum_{j=1}^m a_{jj} z_{ij}^2 \right) \left( \sum_{k < j} a_{jk} z_{ij} z_{ik} \right) \\
&\quad + 4 \left( \sum_{k < j} a_{jk}^2 z_{ij}^2 z_{ik}^2 + \sum_{k < j, l < \alpha: \text{card}(\{k,j\} \cap \{l,\alpha\}) < 2} a_{jk} a_{\alpha l} z_{ij} z_{ik} z_{i\alpha} z_{il} \right)
\end{aligned} \quad (22)$$

Hence

$$\begin{aligned} \mathbb{E}(Z_i^T A Z_i)^2 &= \sum_{j=1}^n a_{jj}^2 \mathbb{E} z_{ij}^4 + 2 \sum_{k < j} a_{jj} a_{kk} \mathbb{E}(z_{ij}^2 z_{ik}^2) + 4 \sum_{k < j} a_{jk}^2 \mathbb{E}(z_{ij}^2 z_{ik}^2) \\ &\asymp \sum_{j=1}^n \sum_{k=1}^n a_{jj} a_{kk} + \sum_{j=1}^n \sum_{k=1}^n a_{jk}^2 = (\text{tr}(A))^2 + \text{tr} A^2. \end{aligned} \quad (23)$$

By Cauchy inequality,  $0 \leq \text{tr} A^2 \leq (\text{tr} A)^2$ . The conclusion holds.  $\square$

155

**Lemma 3.** Under (5), (6), for  $i \neq j$  we have

$$\mathbb{E}(X_i^T X_j)^4 = O(1) \left( \text{tr}(\Sigma + \mu \mu^T)^2 \right)^2. \quad (24)$$

*Proof.*

$$\begin{aligned} (X_i^T X_j)^4 &= (Z_i^T \Gamma^T \Gamma Z_j + \mu^T \Gamma Z_i + \mu^T \Gamma Z_j + \mu^T \mu)^4 \\ &\leq 64 ((Z_i^T \Gamma^T \Gamma Z_j)^4 + (\mu^T \Gamma Z_i)^4 + (\mu^T \Gamma Z_j)^4 + (\mu^T \mu)^4) \end{aligned} \quad (25)$$

We deal with the first term by applying Lemma 2 twice.

$$\begin{aligned} \mathbb{E}(Z_i^T \Gamma^T \Gamma Z_j)^4 &= \mathbb{E}(Z_i^T \Gamma^T \Gamma Z_j Z_j^T \Gamma^T \Gamma Z_i)^2 = \mathbb{E} \mathbb{E}((Z_i^T \Gamma^T \Gamma Z_j Z_j^T \Gamma^T \Gamma Z_i)^2 | Z_j) \\ &\asymp \mathbb{E}(Z_j^T \Gamma^T \Sigma \Gamma Z_j)^2 \asymp (\text{tr} \Sigma^2)^2, \end{aligned} \quad (26)$$

Similarly, we have

$$\begin{aligned} \mathbb{E}(\mu^T \Gamma Z_i)^4 &= \mathbb{E}(Z_i^T \Gamma^T \mu \mu^T \Gamma Z_i)^2 \asymp (\mu^T \Sigma \mu)^2 \\ &\leq \lambda_{\max}^2(\Sigma) (\mu^T \mu)^2 \leq \text{tr}(\Sigma^2) (\mu^T \mu)^2 \leq (\text{tr}(\Sigma^2))^2 + (\mu^T \mu)^4. \end{aligned} \quad (27)$$

Thus, the conclusion holds.  $\square$

**Lemma 4.** Under (5), (6), suppose  $i \neq j$ ,  $i \neq k$ ,  $j \neq k$ , we have

$$\mathbb{E}(X_i^T X_j)^2 (X_k^T X_i)^2 = O(1) \left( \text{tr}(\Sigma + \mu \mu^T)^2 \right)^2. \quad (28)$$

*Proof.* Note that

$$\begin{aligned} \mathbb{E}(X_i^T X_j)^2 (X_k^T X_i)^2 &= \mathbb{E} \mathbb{E}((X_i^T X_j)^2 (X_k^T X_i)^2 | X_i) = \mathbb{E}(X_i^T (\Sigma + \mu \mu^T) X_i)^2 \\ &= \mathbb{E}(Z_i^T \Gamma^T (\Sigma + \mu \mu^T) \Gamma Z_i + 2\mu^T (\Sigma + \mu \mu^T) \Gamma Z_i + \mu \Sigma \mu + (\mu^T \mu)^2)^2 \\ &\leq 4\mathbb{E}(Z_i^T \Gamma^T (\Sigma + \mu \mu^T) \Gamma Z_i)^2 + 16\mathbb{E}(\mu^T (\Sigma + \mu \mu^T) \Gamma Z_i)^2 + 4(\mu \Sigma \mu)^2 + 4(\mu^T \mu)^4. \end{aligned}$$



By Lemma (2),

$$\begin{aligned} \mathbb{E}(Z_i^T \Gamma^T (\Sigma + \mu \mu^T) \Gamma Z_i)^2 &\asymp (\text{tr}(\Gamma^T (\Sigma + \mu \mu^T) \Gamma))^2 \\ &= (\text{tr} \Sigma^2 + \mu^T \Sigma \mu)^2 \leq 2(\text{tr} \Sigma^2)^2 + 2(\mu^T \Sigma \mu)^2. \end{aligned}$$

And

$$\begin{aligned} \mathbb{E}(\mu^T (\Sigma + \mu \mu^T) \Gamma Z_i)^2 &= \mu^T (\Sigma + \mu \mu^T) \Sigma (\Sigma + \mu \mu^T) \mu \\ &= \mu^T \Sigma^3 \mu + 2(\mu^T \mu)(\mu^T \Sigma^2 \mu) + (\mu^T \mu)^2 (\mu^T \Sigma \mu). \end{aligned}$$

But for  $i = 1, 2, \dots$ , we have

$$\mu^T \Sigma^i \mu \leq \lambda_{\max}^i(\Sigma) \mu^T \mu \leq (\text{tr}(\Sigma^2))^{i/2} \mu^T \mu.$$

Thus,

$$\begin{aligned} &\mathbb{E}(X_i^T X_j)^2 (X_k^T X_i)^2 \\ &= O(1) \left( (\text{tr} \Sigma^2)^2 + (\text{tr} \Sigma^2)^{3/2} (\mu^T \mu) + (\text{tr} \Sigma^2) (\mu^T \mu)^2 + (\text{tr} \Sigma^2)^{1/2} (\mu^T \mu)^3 + (\mu^T \mu)^4 \right) \\ &= O(1) \left( \text{tr}(\Sigma + \mu \mu^T)^2 \right)^2. \end{aligned} \tag{29}$$

□

**Lemma 5.** Under (5) and (6), we have

$$\frac{\sum_{j < i} (X_i^T X_j)^2}{\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu \mu^T)^2} \xrightarrow{P} 1. \tag{30}$$

*Proof.* Since

$$\begin{aligned} \mathbb{E}(X_i^T X_j)^2 &= \mathbb{E}(X_i^T X_j X_j^T X_i) = \mathbb{E}(X_i^T (\Sigma + \mu \mu^T) X_i) \\ &= \mathbb{E} \text{tr}((\Sigma + \mu \mu^T) X_i X_i^T) = \text{tr}(\Sigma + \mu \mu^T)^2, \end{aligned} \tag{31}$$

we have

$$\mathbb{E} \sum_{j < i} (X_i^T X_j)^2 = \frac{n(n-1)}{2} \text{tr}(\Sigma + \mu \mu^T)^2 \tag{32}$$

So we only need to consider the variance. According to  $\text{card}(\{i, j\} \cap \{k, l\}) = 0, 1, 2$ , we have

$$\begin{aligned} \left( \sum_{j < i} (X_i^T X_j)^2 \right)^2 &= \sum_{j < i} (X_i^T X_j)^4 + \sum_{j < i, k < l: \{i, j\} \cap \{k, l\} = \emptyset} (X_i^T X_j)^2 (X_k^T X_l)^2 \\ &\quad + 2 \sum_{j < i < k} ((X_i^T X_j)^2 (X_k^T X_i)^2 + (X_i^T X_j)^2 (X_k^T X_j)^2 + (X_k^T X_j)^2 (X_k^T X_i)^2). \end{aligned} \tag{33}$$

In (33), there are  $n(n-1)/2$ ,  $n(n-1)(n-2)(n-3)/4$  and  $n(n-1)(n-2)/6$  terms in each summation. By Lemma 3 and Lemma 4, we have

$$\begin{aligned} \mathbb{E}(\sum_{j<i} (X_i^T X_j)^2)^2 &= \frac{n(n-1)(n-2)(n-3)}{4} (\text{tr}(\Sigma + \mu\mu^T))^2 \\ &\quad + O(1)(\frac{n(n-1)}{2} + n(n-1)(n-2))(\text{tr}(\Sigma + \mu\mu^T))^2. \end{aligned} \quad (34)$$

Hence

$$\frac{\text{Var}(\sum_{j<i} (X_i^T X_j)^2)}{(\mathbb{E} \sum_{j<i} (X_i^T X_j)^2)^2} = \frac{\mathbb{E}(\sum_{j<i} (X_i^T X_j)^2)^2 - (\mathbb{E} \sum_{j<i} (X_i^T X_j)^2)^2}{(\mathbb{E} \sum_{j<i} (X_i^T X_j)^2)^2} = O(\frac{1}{n}).$$

Thus the conclusion holds.  $\square$

**Lemma 6.** Under (5), (6), (7) and

$$\mu^T \mu = o(\sqrt{\text{tr} \Sigma^2}), \quad (35)$$

we have

$$\sum_{j<k} (\sum_{i:i>k} X_i^T X_j X_i^T X_k)^2 = o_P\left(\left(\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu\mu^T)^2\right)^2\right). \quad (36)$$

$$\sum_{j<k} (X_i^T X_j)^4 = o_P\left(\left(\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu\mu^T)^2\right)^2\right) \quad (37)$$

$$\sum_{j<k<i} (X_i^T X_j)^2 (X_i^T X_k)^2 = o_P\left(\left(\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu\mu^T)^2\right)^2\right) \quad (38)$$

*Proof.*

$$\begin{aligned} &\mathbb{E} \sum_{j<k} (\sum_{i:i>k} X_i^T X_j X_i^T X_k)^2 \\ &= \mathbb{E} \sum_{j<k} \left( \sum_{i:i>k} (X_i^T X_j)^2 (X_i^T X_k)^2 + 2 \sum_{i_1, i_2: i_1>i_2>k} X_{i_1}^T X_j X_{i_1}^T X_k X_{i_2}^T X_j X_{i_2}^T X_k \right). \end{aligned}$$

By Lemma 3, we have

$$\mathbb{E} \sum_{j<k<i} (X_i^T X_j)^2 (X_i^T X_k)^2 = O(n^3) (\text{tr}(\Sigma + \mu\mu^T))^2.$$

And

$$\begin{aligned}
& \mathbb{E} \sum_{j < k < i_2 < i_1} X_{i_1}^T X_j X_{i_1}^T X_k X_{i_2}^T X_j X_{i_2}^T X_k \\
&= \frac{n(n-1)(n-2)(n-3)}{6} \text{tr}(\Sigma + \mu\mu^T)^4 \\
&\leq \frac{n(n-1)(n-2)(n-3)}{6} 8(\text{tr}(\Sigma)^4 + (\mu^T \mu)^4) \\
&\leq O(n^4)(\lambda_{\max}^2(\Sigma) \text{tr}(\Sigma)^2 + (\mu^T \mu)^4) \\
&= o\left(n^4(\text{tr}\Sigma^2)^2\right),
\end{aligned}$$

where the last line follows by assumption (7) and (35). Thus (36) holds. (37)

160 and (38) follow by Lemma 3 and Lemma 4. □

*Proof of Theorem 3.* By a standard subsequence argument, we only need to prove

$$\rho\left(\mathcal{L}\left(\frac{T_2(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \middle| X_1, \dots, X_n\right), N(0, 1)\right) \xrightarrow{a.s.} 0 \quad (39)$$

along a subsequence. But there is a subsequence  $\{n(k)\}$  along which (36), (37) and (38) holds almost surely. By Proposition 1, we have

$$\mathcal{L}\left(\frac{T_2(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \middle| X_1, \dots, X_n\right) \xrightarrow{\mathcal{L}} N(0, 1)$$

almost surely along  $\{n(k)\}$ , which means (39) holds along  $\{n(k)\}$ . □

*Proof of Theorem 5.* Note that

$$\begin{aligned}
& \Pr\left(\frac{T(X_1, \dots, X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^*\right) \\
&= \Pr\left(\frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^* - \frac{\frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}}\right) \quad (40)
\end{aligned}$$

If (12) holds, by Lemma 5, we have

$$\frac{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}{\frac{n(n-1)}{2} \text{tr}\Sigma^2} \xrightarrow{P} 1.$$

By Corollary 1, we have  $\xi_\alpha^* \xrightarrow{P} \Phi(1 - \alpha)$ . Thus,

$$\begin{aligned}
(40) &= \Pr \left( \frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\frac{n(n-1)}{2} \text{tr} \Sigma^2}} - \frac{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}}{\sqrt{\frac{n(n-1)}{2} \text{tr} \Sigma^2}} \xi_\alpha^* > -\frac{\sqrt{n(n-1)} \mu^T \mu}{\sqrt{2 \text{tr} \Sigma^2}} \right) \\
&= \Pr \left( N(0, 1) - \Phi(1 - \alpha) > -\frac{\sqrt{n(n-1)} \mu^T \mu}{\sqrt{2 \text{tr} \Sigma^2}} \right) + o(1) \\
&= \Phi(-\Phi(1 - \alpha) + \frac{\sqrt{n(n-1)} \mu^T \mu}{\sqrt{2 \text{tr} \Sigma^2}}) + o(1),
\end{aligned}$$

where the last two equality holds by Theorem 1, Slutsky's theorem and Lemma 1.

If (13) holds, by Lemma 5, we have

$$\frac{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}{\frac{n(n-1)}{2} (\mu^T \mu)^2} \xrightarrow{P} 1.$$

165 Thus

$$\begin{aligned}
&\frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \\
&= \frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\frac{n(n-1)}{2} \text{tr} \Sigma^2}} \frac{\sqrt{\frac{n(n-1)}{2} \text{tr} \Sigma^2}}{\sqrt{\frac{n(n-1)}{2} (\mu^T \mu)^2}} \frac{\sqrt{\frac{n(n-1)}{2} (\mu^T \mu)^2}}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \xrightarrow{P} 0.
\end{aligned}$$

Since  $\xi_\alpha^* \xrightarrow{P} \frac{\sqrt{2}}{2} (\Phi^{-1}(1 - \frac{\alpha}{2}) - 1)$  by Corollary 2, we have that (40)  $\rightarrow 1$ .

□

*Proof of Theorem 6.* Note that

$$\begin{aligned}
&\Pr \left( \frac{T(X_1, \dots, X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^* \right) \\
&= \Pr \left( \frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} > \frac{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} \xi_\alpha^* - \frac{\frac{n(n-1)}{2} \mu^T \mu}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} \right).
\end{aligned} \tag{41}$$

If (12) holds, the theorem follows by Theorem 2 and the fact that if (12) holds, the coefficient of  $\xi_\alpha^*$  in (41) tends to 0.

If (13) holds, the theorem follows by noting that

$$(41) = \Pr\left(\frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} > -(1 + o_P(1)) \frac{\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}}\right).$$

170

□

*Proof of Theorem 4.*

$$\begin{aligned} \sum_{j < i} X_i^T X_j \epsilon_i \epsilon_j &= \sum_{j < i} Z_i^T \Gamma^T \Gamma Z_j \epsilon_i \epsilon_j \\ &\quad + \sum_{j < i} \mu^T \Gamma Z_i \epsilon_i \epsilon_j + \sum_{j < i} \mu^T \Gamma Z_j \epsilon_i \epsilon_j + \mu^T \mu \sum_{j < i} \epsilon_i \epsilon_j \\ &\stackrel{\text{def}}{=} C_1 + C_2 + C_3 + C_4. \end{aligned}$$

Term  $C_4$  plays a major role. Note that

$$C_4 = \frac{n}{2} \mu^T \mu \left( \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i \right)^2 - 1 \right).$$

By central limit theorem, we have

$$\rho\left(\mathcal{L}\left(\frac{C_4}{\frac{n}{2}\mu^T\mu} \middle| X_1, \dots, X_n\right), \chi_1^2 - 1\right) \xrightarrow{a.s.} 0.$$

Next we show that  $C_1$ ,  $C_2$  and  $C_3$  are negligible under the assumptions of the theorem. By a standard subsequence argument and Slutsky's theorem, we can obtain

$$\rho\left(\mathcal{L}\left(\frac{T(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\frac{n}{2}\mu^T\mu} \middle| X_1, \dots, X_n\right), \chi_1^2 - 1\right) \xrightarrow{P} 0 \quad (42)$$

by showing that

$$\mathbb{E}\left(\left(\frac{C_i}{\frac{n}{2}\mu^T\mu}\right)^2 \middle| X_1, \dots, X_n\right) \xrightarrow{P} 0, \quad i = 1, 2, 3.$$

It in turn suffices to show

$$\mathbb{E}\left(\frac{C_i}{\frac{n}{2}\mu^T\mu}\right)^2 \rightarrow 0, \quad i = 1, 2, 3. \quad (43)$$

By direct calculation, we have

$$\mathbb{E}(C_1^2) = \mathbb{E}\mathbb{E}(C_1^2 | X_1, \dots, X_n) = \sum_{j < i} \mathbb{E}(Z_i^T \Gamma^T \Gamma Z_j)^2 = \frac{n(n-1)}{2} \text{tr} \Sigma^2,$$

and

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

Thus (43) follows by Assumption (13). Having (42) holds, the theorem follows by Slutsky's theorem, Lemma 5 and Assumption (13).  $\square$

## References

- 175 [1] Z. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem, *Statistica Sinica* 6 (2) (1996) 311–329.
- [2] M. S. Srivastava, A test for the mean vector with fewer observations than the dimension under non-normality, *Journal of Multivariate Analysis* 100 (3) (2009) 518–532. As the access to this document is restricted, you may want to look for a different version under "Related research" (further below) or for a different version of it.
- 180 [3] S. X. Chen, Y. L. Qin, A two-sample test for high-dimensional data with applications to gene-set testing, *Annals of Statistics* 38 (2) (2010) 808–835.
- [4] L. Wang, B. Peng, R. Li, A high-dimensional nonparametric multivariate test for mean vector., *Journal of the American Statistical Association* 110 (512) (2015) 00–00.
- 185 [5] P. D. Jong, A central limit theorem for generalized quadratic forms, *Probability Theory & Related Fields* 75 (2) (1987) 261–277.
- [6] R. J. Bradley Efron, On testing the significance of sets of genes, *The Annals of Applied Statistics* 1 (1) (2007) 107–129.
- 190 [7] R. A. Fisher, The design of experiments.
- [8] J. P. Romano, On the behavior of randomization tests without a group invariance assumption, *Journal of the American Statistical Association* 85 (411) (1990) 686–692.

- 195 [9] J. P. R. E. L. Lehmann, Testing Statistical Hypotheses, Springer New York,  
2005. doi:10.1007/0-387-27605-X.
- [10] J. Fan, P. Hall, Q. Yao, To how many simultaneous hypothesis tests can  
normal, student's t or bootstrap calibration be applied?, Journal of the  
American Statistical Association 102 (480) (2007) 1282–1288.
- 200 [11] D. Pollard, Convergence of stochastic processes, 1984.