

# Statistical Inference for High-Dimensional Data

Runze Li

The Pennsylvania State University

June 2017

# Topics in this chapter

- a. One sample and two sample mean problems
- b. Tests on covariance structures

# 1 One sample and two sample mean problems

## One sample mean problem:

Suppose that  $\mathbf{x}_1, \dots, \mathbf{x}_n$  is a random sample IID from a  $p$ -dimensional population with mean  $\boldsymbol{\mu}$ . We want to test

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

## Two sample mean problem:

Suppose that  $\mathbf{x}_{kj}$ ,  $k = 1, 2$ , and  $j = 1, \dots, n_k$  is a random sample IID from a population with mean  $\boldsymbol{\mu}_k$ . To test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{versus} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

# Hotelling $T^2$ test for one sample problem

Given a random sample  $\mathbf{x}_1, \dots, \mathbf{x}_n$  from  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , define the sample mean  $\bar{\mathbf{x}}$  as

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i,$$

and the sample covariance matrix  $\mathbf{S}$  as

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T.$$

# Hotelling's $T^2$ test for one sample problem

Under  $p < n$ , the Hotelling's  $T^2$  test for

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \text{ vs } H_1 : \boldsymbol{\mu} \neq \mathbf{0}$$

has its test statistics defined as

$$T^2 = n\bar{\mathbf{x}}'\mathbf{S}^{-1}\bar{\mathbf{x}}.$$

Under  $H_0$ ,

$$T^2 \sim \frac{(n-1)p}{n-p} F_{p, n-p}.$$

For one sample normal mean problem,  $T^2$  is equivalent to the LRT.

# Hotelling's $T^2$ test for one sample problem

Under  $p < n$ , the Hotelling's  $T^2$  test for

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \text{ vs } H_1 : \boldsymbol{\mu} \neq \mathbf{0}$$

has its test statistics defined as

$$T^2 = n\bar{\mathbf{x}}' \mathbf{S}^{-1} \bar{\mathbf{x}}.$$

Under  $H_0$ ,

$$T^2 \sim \frac{(n-1)p}{n-p} F_{p, n-p}.$$

For one sample normal mean problem,  $T^2$  is equivalent to the LRT.

# Hotelling's $T^2$ test: two sample problem

Given two independent random samples,

- $\{\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1}\}$  from  $N_p(\boldsymbol{\mu}_1, \Sigma)$ ,
- $\{\mathbf{x}_{21}, \dots, \mathbf{x}_{2n_2}\}$  from  $N_p(\boldsymbol{\mu}_2, \Sigma)$ .

Sample means:  $\bar{\mathbf{x}}_1$  and  $\bar{\mathbf{x}}_2$ ,

Sample covariance matrix:  $\mathbf{S}_1$  and  $\mathbf{S}_2$ .

Define the pooled sample covariance matrix  $\mathbf{S}_p$  as

$$\begin{aligned}\mathbf{S}_p &= \frac{1}{n} \sum_{k=1}^2 \sum_{j=1}^{n_k} (\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^T \\ &= \frac{1}{n} ((n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2),\end{aligned}\tag{1}$$

where  $n = n_1 + n_2 - 2$ .

This is not equivalent to an LRT.

# Hotelling's $T^2$ test: two sample problem

Under  $p < n$ , the Hotelling's  $T^2$  test for

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ vs } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

has its test statistics defined as

$$T^2 = \left( \frac{1}{n_1} + \frac{1}{n_2} \right)^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_p^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2). \quad (2)$$

Under  $H_0$ ,

$$T^2 \sim \frac{np}{n - p + 1} F_{p, n-p+1}. \quad (3)$$



# Hotelling's $T^2$ test: two sample problem

Under  $p < n$ , the Hotelling's  $T^2$  test for

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ vs } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

has its test statistics defined as

$$T^2 = \left( \frac{1}{n_1} + \frac{1}{n_2} \right)^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_p^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2). \quad (2)$$

Under  $H_0$ ,

$$T^2 \sim \frac{np}{n - p + 1} F_{p, n-p+1}. \quad (3)$$

# Limitations of $T^2$ under HD setting

- When  $p > n$ , the Hotelling's  $T^2$  test is not well-defined due to the singularity of sample covariance matrix  $\mathbf{S}$  and  $\mathbf{S}_p$ .
- When  $p < n$ , the Hotelling's  $T^2$  test has poor power when  $p$  is close to  $n$ , where the sample covariance matrix is nearly singular (Bai and Saranadasa, 1996).

# Modification for one sample problem

## 1 Replace $\mathbf{S}^{-1}$

- Identity matrix:  $I$
- Diagonal matrix:  $D_S$ , the diagonal matrix of  $\mathbf{S}$
- Ridge-like estimator:  $(\mathbf{S} + \lambda I)^{-1}$

## 2 Projection

# Methods that replace $\mathbf{S}^{-1}$

Table 1: Overview of methods that replace  $\mathbf{S}^{-1}$

Method	Test Statistic	Null Distribution	Normality
Dempster (1958, 1960)	$T_D^2 = \bar{\mathbf{x}}^T \bar{\mathbf{x}} / (\frac{1}{n} \text{tr} \mathbf{S})$	$T_D^2 \sim F_{r, nr}$	Yes
Bai and Saranadasa (1996)	$T_{BS} = \bar{\mathbf{x}}^T \bar{\mathbf{x}} - \frac{1}{n} \text{tr} \mathbf{S}$	$T_{BS} / \sqrt{\text{Var}(T_{BS})} \rightarrow N(0, 1)$	No
Chen and Qin (2010)	$T_{CQ} = \sum_{i \neq j}^n \mathbf{x}_i^T \mathbf{x}_j$	$T_{CQ} / \sqrt{\text{Var}(T_{CQ})} \rightarrow N(0, 1)$	No
Srivastava and Du (2008)	$T_{SD(adj)} = \bar{\mathbf{x}}^T (\frac{1}{n} D_S)^{-1} \bar{\mathbf{x}} - \frac{(n-1)p}{n-3}$	$T_{SD(adj)} / \sqrt{c_{p,n} \text{Var}(T_{SD(adj)})} \rightarrow N(0, 1)$	Yes
Srivastava (2009)	$T_{SD(noadj)} = \bar{\mathbf{x}}^T (\frac{1}{n} D_S)^{-1} \bar{\mathbf{x}} - \frac{(n-1)p}{n-3}$	$T_{SD(noadj)} / \sqrt{\text{Var}(T_{SD(noadj)})} \rightarrow N(0, 1)$	No
Chen et al. (2011)	$RHT(\lambda) = n \bar{\mathbf{x}}^T (\mathbf{S} + \lambda \mathbf{I})^{-1} \bar{\mathbf{x}}$	$\frac{\sqrt{p(RHT(\lambda)/p - \Theta_1(\lambda, c))}}{(2\Theta_2(\lambda, c))^{1/2}} \rightarrow N(0, 1)$	Yes

# Methods that replace $\mathbf{S}^{-1}$

Table 1: Overview of methods that replace  $\mathbf{S}^{-1}$

Method	Test Statistic	Null Distribution	Normality
Dempster (1958, 1960)	$T_D^2 = \bar{\mathbf{x}}^T \bar{\mathbf{x}} / (\frac{1}{n} \text{tr} \mathbf{S})$	$T_D^2 \sim F_{r, nr}$	Yes
Bai and Saranadasa (1996)	$T_{BS} = \bar{\mathbf{x}}^T \bar{\mathbf{x}} - \frac{1}{n} \text{tr} \mathbf{S}$	$T_{BS} / \sqrt{\text{Var}(T_{BS})} \rightarrow N(0, 1)$	No
Chen and Qin (2010)	$T_{CQ} = \sum_{i \neq j}^n \mathbf{x}_i^T \mathbf{x}_j$	$T_{CQ} / \sqrt{\text{Var}(T_{CQ})} \rightarrow N(0, 1)$	No
Srivastava and Du (2008)	$T_{SD(adj)} = \bar{\mathbf{x}}^T (\frac{1}{n} D_S)^{-1} \bar{\mathbf{x}} - \frac{(n-1)p}{n-3}$	$T_{SD(adj)} / \sqrt{c_{p,n} \text{Var}(T_{SD(adj)})} \rightarrow N(0, 1)$	Yes
Srivastava (2009)	$T_{SD(noadj)} = \bar{\mathbf{x}}^T (\frac{1}{n} D_S)^{-1} \bar{\mathbf{x}} - \frac{(n-1)p}{n-3}$	$T_{SD(noadj)} / \sqrt{\text{Var}(T_{SD(noadj)})} \rightarrow N(0, 1)$	No
Chen et al. (2011)	$RHT(\lambda) = n \bar{\mathbf{x}}^T (\mathbf{S} + \lambda \mathbf{I})^{-1} \bar{\mathbf{x}}$	$\frac{\sqrt{p} (RHT(\lambda) / p - \Theta_1(\lambda, c))}{(2\Theta_2(\lambda, c))^{1/2}} \rightarrow N(0, 1)$	Yes

$$\widehat{\text{tr}(\Sigma^2)}_{BS} = \frac{(n-1)^2}{n(n+3)} \left[ \text{tr} \mathbf{S}^2 - (\text{tr} \mathbf{S})^2 / (n-1) \right] \quad (4)$$

$$\widehat{\text{tr}(\Sigma^2)}_{CQ} = \frac{1}{(n-1)n} \text{tr} \left( \sum_{j \neq k}^n (\mathbf{x}_j - \bar{\mathbf{x}}_{(j,k)}) \mathbf{x}_j^T (\mathbf{x}_k - \bar{\mathbf{x}}_{(j,k)}) \mathbf{x}_k^T \right) \quad (5)$$

# Methods that replace $\mathbf{S}^{-1}$

Table 1: Overview of methods that replace  $\mathbf{S}^{-1}$

Method	Test Statistic	Null Distribution	Normality
Dempster (1958, 1960)	$T_D^2 = \bar{\mathbf{x}}^T \bar{\mathbf{x}} / (\frac{1}{n} \text{tr} \mathbf{S})$	$T_D^2 \sim F_{r, nr}$	Yes
Bai and Saranadasa (1996)	$T_{BS} = \bar{\mathbf{x}}^T \bar{\mathbf{x}} - \frac{1}{n} \text{tr} \mathbf{S}$	$T_{BS} / \sqrt{\text{Var}(T_{BS})} \rightarrow N(0, 1)$	No
Chen and Qin (2010)	$T_{CQ} = \sum_{i \neq j}^n \mathbf{x}_i^T \mathbf{x}_j$	$T_{CQ} / \sqrt{\text{Var}(T_{CQ})} \rightarrow N(0, 1)$	No
Srivastava and Du (2008)	$T_{SD(adj)} = \bar{\mathbf{x}}^T (\frac{1}{n} D_S)^{-1} \bar{\mathbf{x}} - \frac{(n-1)p}{n-3}$	$T_{SD(adj)} / \sqrt{c_{p,n} \text{Var}(T_{SD(adj)})} \rightarrow N(0, 1)$	Yes
Srivastava (2009)	$T_{SD(noadj)} = \bar{\mathbf{x}}^T (\frac{1}{n} D_S)^{-1} \bar{\mathbf{x}} - \frac{(n-1)p}{n-3}$	$T_{SD(noadj)} / \sqrt{\text{Var}(T_{SD(noadj)})} \rightarrow N(0, 1)$	No
Chen et al. (2011)	$RHT(\lambda) = n \bar{\mathbf{x}}^T (\mathbf{S} + \lambda \mathbf{I})^{-1} \bar{\mathbf{x}}$	$\frac{\sqrt{p} (RHT(\lambda) / p - \Theta_1(\lambda, c))}{(2\Theta_2(\lambda, c))^{1/2}} \rightarrow N(0, 1)$	Yes

# Methods that replace $\mathbf{S}^{-1}$

Table 1: Overview of methods that replace  $\mathbf{S}^{-1}$

Method	Test Statistic	Null Distribution	Normality
Dempster (1958, 1960)	$T_D^2 = \bar{\mathbf{x}}' \bar{\mathbf{x}} / (\frac{1}{n} \text{tr} \mathbf{S})$	$T_D^2 \sim F_{r, nr}$	Yes
Bai and Saranadasa (1996)	$T_{BS} = \bar{\mathbf{x}}' \bar{\mathbf{x}} - \frac{1}{n} \text{tr} \mathbf{S}$	$T_{BS} / \sqrt{\text{Var}(\overline{T_{BS}})} \rightarrow N(0, 1)$	No
Chen and Qin (2010)	$T_{CQ} = \sum_{i \neq j}^n \mathbf{x}_i' \mathbf{x}_j$	$T_{CQ} / \sqrt{\text{Var}(\overline{T_{CQ}})} \rightarrow N(0, 1)$	No
Srivastava and Du (2008)	$T_{SD(adj)} = \bar{\mathbf{x}}' (\frac{1}{n} D_S)^{-1} \bar{\mathbf{x}} - \frac{(n-1)p}{n-3}$	$T_{SD(adj)} / \sqrt{c_{p,n} \text{Var}(\overline{T_{SD(adj)}})} \rightarrow N(0, 1)$	Yes
Srivastava (2009)	$T_{SD(noadj)} = \bar{\mathbf{x}}' (\frac{1}{n} D_S)^{-1} \bar{\mathbf{x}} - \frac{(n-1)p}{n-3}$	$T_{SD(noadj)} / \sqrt{\text{Var}(\overline{T_{SD(noadj)}})} \rightarrow N(0, 1)$	No
Chen et al. (2011)	$RHT(\lambda) = n \bar{\mathbf{x}}' (\mathbf{S} + \lambda \mathbf{I})^{-1} \bar{\mathbf{x}}$	$\frac{\sqrt{p} (RHT(\lambda) / p - \Theta_1(\lambda, c))}{(2\Theta_2(\lambda, c))^{1/2}} \rightarrow N(0, 1)$	Yes

# Extensions of multinormal distribution

## 1. Independent component model:

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Gamma} \mathbf{z},$$

where  $\boldsymbol{\Gamma}$  is a  $p \times m$  matrix for some  $m \geq p$  such that  $\boldsymbol{\Gamma} \boldsymbol{\Gamma}^T = \boldsymbol{\Sigma}$ , and  $\mathbf{z} = (z_1, \dots, z_m)^T$  consisting of  $m$ -variate independent and identically distributed random variables satisfying  $E(z_j) = 0$ ,  $\text{var}(z_j) = 1$ , and  $E(z_j^4) = \kappa < \infty$ .

Test-specific:

- BS:  $\mathbf{z}$  satisfies that  $E(\prod_{k=1}^m z_k^{v_k})$  equals 0 when there is at least one  $v_k = 1$  and equals 1 when there are two  $v_k$ 's equal to 2, whenever  $\sum_{k=1}^m v_k = 4$ .
- CQ:  $\mathbf{z}$  satisfies that  $E(\prod_{k=1}^q z_{l_k}^{\alpha_k}) = \prod_{k=1}^q E(z_{l_k}^{\alpha_k})$  for a positive integer  $q$  such that  $\sum_{l=1}^q \alpha_l \leq 8$  and  $l_1 \neq \dots \neq l_q$ .



# Extensions of multinormal distribution

2. **Elliptical distribution:**  $\mathbf{x}$  has characteristic function

$$E \exp(i \mathbf{t}^T \mathbf{x}) = \exp(i \boldsymbol{\mu}^T \mathbf{t}) \phi(\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t})$$

If  $\mathbf{x}$  has a density, its density is of the form

$$f(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1/2} g\{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\},$$

and  $E(\mathbf{x}) = \boldsymbol{\mu}$  and  $\text{cov}(\mathbf{x}) \propto \boldsymbol{\Sigma}$  if they exist. Thus,  $\boldsymbol{\Sigma}$  is called the scatter matrix of  $\mathbf{x}$ .

Stochastic Representation:

$$\mathbf{x} = \boldsymbol{\mu} + R \boldsymbol{\Sigma}^{1/2} \mathbf{u},$$

where  $R$  is a nonnegative random variable and is independent of  $\mathbf{u}$ , which follows a uniform distribution over the unit sphere  $\{\mathbf{u} : \mathbf{u}^T \mathbf{u} = 1\}$ .

# Projection based test

- Lauter's test (Lauter, 1996; Lauter et al., 1998)
  - SS test
  - PC test
- Lopes' test: random projection (Lopes et al., 2012)

A  $n \times p$  matrix  $\mathbf{X}$  is left-spherically distributed if  $\mathbf{X}$  and  $C\mathbf{X}$  follow the same distribution for every fixed  $n \times n$  orthogonal matrix  $C$ .

Lauter's test is based on matrix spherical distribution theory.

- Two criteria in selecting weight vector  $\mathbf{d}$ :
  - (1)  $\mathbf{d}$  depends only on  $\mathbf{X}^T \mathbf{X}$ ; (2)  $P(\mathbf{d} \neq \mathbf{0}) = 1$
- Define  $\mathbf{z} = \mathbf{X}\mathbf{d}$ . Such  $\mathbf{z}$  follows a spherical distribution. Thus,

$$\sqrt{n}\bar{z}/s_z \sim t_{n-1}.$$

where  $s_z$  is the sample standard deviation of  $z_1, \dots, z_n$ .

# Lauter's test (II)

- SS test attains the highest power in the situation where all variables present nearly the same relative deviation and the same correlation to each other.
- PC test is especially appropriate with one-factor structure that  $\Sigma = K + \theta\theta'$  and  $\mu = \theta\delta$ , where  $K$  is a diagonal matrix,  $\theta$  is a vector and  $\delta$  is nonnegative scalar.
- SS test and PC test are power insufficient when the alternative mean vector  $\mu$  contains at least one zero element (Frick, 1996).

# Lopes's tests: single random projection

- Intuition

- Random projection can simultaneously reduce the dimension and preserve the length with high probability.

- Implementation

- Generate a random matrix  $P_k$  of dimension  $p \times k$  with entries from i.i.d  $N(0, 1)$
- Define  $\mathbf{Y} = \mathbf{X}P_k$  then perform Hotelling's  $T^2$  test with  $\mathbf{Y}$  such that

$$\begin{aligned}T_{RP_S}^2 &= n\bar{\mathbf{y}}^T(\mathbf{S}_y)^{-1}\bar{\mathbf{y}}, \\ &= n\bar{\mathbf{x}}^T P_k (P_k^T \mathbf{S}_x P_k)^{-1} P_k^T \bar{\mathbf{x}}\end{aligned}$$

- Advantages: Single projection results an exact test if  $\mathbf{x}$  is normal, and utilizes the off-dimensional information of sample covariance matrix.

# Goals

To address two interesting questions related to projection test

- what is the **optimal dimension** of projections?
- what are the **optimal projections**?

To construct an **exact** projection with **optimal** projection

# Projection test

Suppose that  $\mathbf{x}_i$ ,  $i = 1, \dots, n$  iid  $\sim N_p(\boldsymbol{\mu}, \Sigma)$ . Denote  $\zeta = \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}$ . Then

$$(n-p)T^2/(n-1)p \sim F_{p, n-p}(n\zeta),$$

the noncentral  $F$ -distribution with  $(p, n-p)$  degrees of freedom and noncentrality parameter  $n\zeta$ .

Let  $A$  be a  $p \times k$  full column-rank matrix with  $k \ll p$  and  $k < n$ .

Let  $\mathbf{y}_i = A^T \mathbf{x}_i$ ,  $i = 1, \dots, n$ , iid  $N_k(A^T \boldsymbol{\mu}, A^T \Sigma A)$ .

Under  $H_0$ ,  $E(\mathbf{y}_i) = \mathbf{0}$ . Define the **projection Hotelling's  $T^2$  test** to be

$$T_A^2 = n\bar{\mathbf{x}}^T A(A^T \mathbf{S} A)^{-1} A^T \bar{\mathbf{x}},$$

which is  $T^2$  test based on the  $\mathbf{y}_i$ 's.

# Optimal projection direction

Then  $T_A^2$  may be used to test  $H_{0A} : \mu_A = \mathbf{0}$ , where  $\mu_A = A^T \mu$ .

Of course, in general,  $H_{0A}$  does not imply  $H_0$ .

A critical issue here is how to determine  $k$  and how to choose  $A$  for a given  $k$  so that  $T_A^2$  maximizes the power under  $H_1$ .

## Theorem

*The projection test  $T_A^2$  reaches its best power for  $H_1$  at  $k = 1$  and  $A = \Sigma^{-1} \mu$ .*

This theorem provides us insights into how to construct a projection Hotelling's  $T^2$  test.



# Optimal projection direction

Let  $\mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}$  be the theoretical optimal direction for  $T_A^2$  and denote by  $T_{\mathbf{a}}^2$  the projection test  $T_A^2$  with  $A = \mathbf{a}$ .

Clearly, when  $A = \mathbf{a}$ ,  $H_{0A}$  becomes  $H_{0A} : \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} = 0$ , which is equivalent to  $H_0 : \boldsymbol{\mu} = \mathbf{0}$  under the assumption of  $\Sigma$  being full rank.

Consequently,  $T_{\mathbf{a}}^2$  serves as a feasible tool for the one sample problem.

Since  $\mathbf{a}$  depends on unknown population parameters, we need to estimate  $\mathbf{a}$  to carry out  $T_{\mathbf{a}}^2$  in practice.

Estimation of  $\mathbf{a}$  has been studied in the literature HD portfolio optimization and HD LDA.

# Estimation of optimal direction

To obtain an exact test, we randomly splitting the data into two parts:

- one part is used to estimate the optimal direction  $\mathbf{a}$ ;
- the other part is used to construct the projection test.

# Algorithm

Randomly partition the sample  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  into

$$\begin{cases} \text{estimating set: } \{\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1}\} \rightarrow (\bar{\mathbf{x}}_1, \mathbf{S}_1) \\ \text{testing set: } \{\mathbf{x}_{21}, \dots, \mathbf{x}_{2n_2}\} \end{cases},$$

where  $n_1 = \lfloor n\kappa \rfloor$  and  $n_2 = n - n_1$ .

(a) Estimate  $\mathbf{a}$  that

$$\hat{\mathbf{a}} = (\mathbf{S}_1 + \lambda D \mathbf{S}_1)^{-1} \bar{\mathbf{x}}_1,$$

(b) Construct linear score  $y_i = \hat{\mathbf{a}}^T \mathbf{x}_i$ ,  $i = 1, \dots, n_2$ . Then calculate the  $t$ -test statistic based on  $y_i$ 's. Under  $H_0$ , the  $t$ -test follows a  $t$ -distribution with  $\text{df} = n_2 - 1$ .

Set  $\kappa = 0.4$  and  $\lambda = n_1^{-0.5}$  based on simulation studies.

# Sample splitting strategy

## Advantages

Lead to an exact test.

## Disadvantages

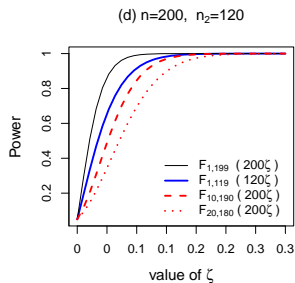
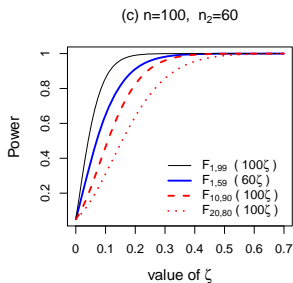
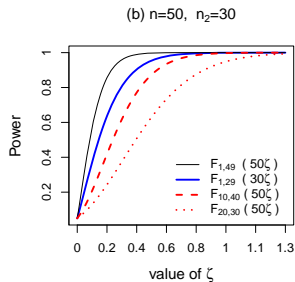
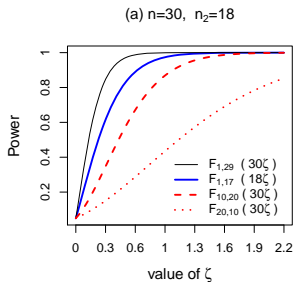
- Introduces two parameters:  $\kappa$  and  $\lambda$ .
- Reduces the sample size for constructing linear scores.

# Power loss due to sample splitting

- Goal: study the effect of sample size reduction on power.
- Method: compare power functions of the proposed test and several baseline curves under  $p < n$  setting.
  - Proposed projection test:  $F_{1,n_2-1}(n_2\zeta)$ .
  - Projection test with known optimal direction:  $F_{1,n-1}(n\zeta)$ .
  - Hotelling's  $T^2$  test:  $F_{p,n-p}(n\zeta)$ ,

where  $\zeta = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ .

- Setup:
  - $n = 30, 50, 100, 200$ .
  - $p = 10, 20$ .
  - Mimic 40% data used to estimate  $\mathbf{a}$  and 60% data used to test.



# Comparison of Power Function of Projection test vs Hotelling's $T^2$ test

# Plots of power functions implies:

If we can estimate  $\mathbf{a}$  well, the projection test can be more powerful than the classical Hotelling  $T^2$  test when  $p$  is close to, but less than  $n$ .

# Asymptotic power comparison

Under  $H_1 : \boldsymbol{\mu} = \boldsymbol{\delta}/\sqrt{n}$ , compare three tests:

- The proposed projection test
- $T^2 = n\bar{\mathbf{x}}^T \mathbf{S}^{-1} \bar{\mathbf{x}}$ , which shares asymptotic power function of Hotelling's  $T^2$ .
- $T_3^2 = n\|\bar{\mathbf{x}}\|^2$ , which shares asymptotic power function of BS1996: Bai & Saranadasa (1996), SD2008: Srivastava & Du (2008) and CQ2010: Chen & Qin (2010) under mild conditions.

Let  $\Phi(\cdot)$  and  $z_\alpha$  denote the cumulative distribution function and  $1 - \alpha$  quantile of  $N(0, 1)$ , respectively.



# Asymptotic function of new test and $T^2$

Assume that  $\hat{\mathbf{a}} \rightarrow \mathbf{a}_0 \propto \mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}$  in probability as  $n_1 \rightarrow \infty$ . Further assume that  $\sqrt{n_2/n} \rightarrow b > 0$  as  $n \rightarrow \infty$ , where  $n_2$  is the size of the second subsample. Then the asymptotic power function of  $T_{\mathbf{a}}^2$  at a given level  $\alpha$  is

$$\beta_{1p}(\eta) = \Phi(-z_{\alpha/2} + b\sqrt{\eta}),$$

where  $\eta = \boldsymbol{\delta}^T \Sigma^{-1} \boldsymbol{\delta}$ .

$T^2 \rightarrow \chi_p^2$  under  $H_0 : \boldsymbol{\mu} = 0$ , and  $T^2 \rightarrow \chi_p^2(\eta)$  under  $H_1 : \boldsymbol{\mu} = \boldsymbol{\delta}/\sqrt{n}$ . The asymptotic power function of  $T^2$  at level  $\alpha$  is

$$\beta_{2p}(\eta|\tau_p) = \Phi \left\{ -z_{\alpha}/\sqrt{1+2\tau_p} + \frac{1}{2}\sqrt{\tau_p/(0.5+\tau_p)} \cdot \sqrt{\eta} \right\},$$

where  $\tau_p = \eta/p$ .

# Comparison between $\beta_{1p}(\eta)$ and $\beta_{2p}(\eta|\tau_p)$

## Proposition

Assume that  $\sqrt{n_2/n} \rightarrow b$ . The following statements are valid.

- (a) If  $b > 0.5$  and  $\eta \rightarrow \infty$  as  $p \rightarrow \infty$ , then  $\beta_{1p}(\eta) - \beta_{2p}(\eta|\tau_p) > 0$  for large enough  $p$ .
- (b) If  $\sqrt{\eta}b \geq z_{\alpha/2} - z_{\alpha}$  and  $\tau_p \rightarrow 0$  as  $p \rightarrow \infty$ , then  $\beta_{1p}(\eta) - \beta_{2p}(\eta|\tau_p) > 0$  for large enough  $p$ .

Statement (a) implies that, if  $\eta \rightarrow \infty$  as  $p \rightarrow \infty$  and we always set  $S_2$  to include more than 25% samples so that  $b > 0.5$ , then the proposed projection test may be asymptotically more powerful than the traditional  $T^2$ .

In (b), the condition  $\tau_p \rightarrow 0$  implies that the signal in  $\mu$  is weak or sparse.

# Asymptotic function of $T_3^2$

Let  $\beta_{3p}(\boldsymbol{\delta}, \Sigma)$  be the asymptotic power function of  $T_3^2$  and

$$\beta_{3p}(\eta|\tau_p^*) = \Phi \left\{ -z_\alpha / \sqrt{1 + 2\tau_p^*} + \frac{1}{2} \sqrt{\tau_p^* / (0.5 + \tau_p^*)} \cdot \sqrt{\eta} \right\},$$

where  $\tau_p^* = \boldsymbol{\delta}^T \Sigma \boldsymbol{\delta} / \text{tr}(\Sigma^2)$ .

Under local alternative, it can be shown that  $\beta_{3p}(\boldsymbol{\delta}, \Sigma) \leq \beta_{3p}(\eta|\tau_p^*)$ .

# Comparison between $\beta_{1p}(\eta)$ and $\beta_{3p}(\eta|\tau_p^*)$

## Proposition

Assume that  $\sqrt{n_2/n} \rightarrow b$ . The following statements are valid.

- (I) If  $b > 0.5$  and  $\eta \rightarrow \infty$ , then  $\beta_{1p}(\eta) - \beta_{3p}(\eta|\tau_p^*) > 0$  for large enough  $p$ .
- (II) If  $\sqrt{\eta}b \geq z_{\alpha/2} - z_{\alpha}$  and  $\tau_p^* \rightarrow 0$  as  $p \rightarrow \infty$ , then  $\beta_{1p}(\eta) - \beta_{3p}(\eta|\tau_p^*) > 0$  for large enough  $p$ .

In (II), the condition  $\tau_p^* \rightarrow 0$  means that  $\delta^T \Sigma \delta = o\{\text{tr}(\Sigma^2)\}$ . Under this assumption,  $\beta_{3p} \approx \Phi \left\{ -z_{\alpha} + \frac{n\|\mu\|^2}{\sqrt{2\text{tr}(\Sigma^2)}} \right\}$ , which is exactly the asymptotic power function of BS1996, CQ2010, and SD2008 with all diagonal elements of  $\Sigma$  being 1.

# Examples

- Compound symmetry:  $\Sigma = (1 - r)I_p + r\mathbf{1}_p\mathbf{1}_p^T$  for  $r \in [0, 1)$ .
- AR correlation:  $\sigma_{ij} = r^{|i-j|}$  for  $r \in (0, 1)$

Under these correlation structure, we can get closed form for  $\eta$ ,  $\tau_p$  and  $\tau_p^*$ .  
We can verify conditions in (a), (b), (I) or (II).

# Simulation Studies

Generate  $\mathbf{x}_1, \dots, \mathbf{x}_n$  iid from  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  with

Parameter	Choices	More details
$\boldsymbol{\Sigma}$	$\boldsymbol{\Sigma}_1 = (\rho_{ij} = \rho, i \neq j)$ $\boldsymbol{\Sigma}_2 = (\rho_{ij} = \rho^{ i-j }, i \neq j)$ $\boldsymbol{\Sigma}_3 = 0.5\boldsymbol{\Sigma}_1 + 0.5\boldsymbol{\Sigma}_2$	$\rho = 0.25, 0.5, 0.75, 0.95$
$\boldsymbol{\mu}$	$(c\mathbf{1}'_{10}, \mathbf{0}'_{p-10})'$	$c = 0, 0.5, 1$
$n$	40, 160	
$p$	400, 1600	

Tests included in our comparison:

- D1958 (Dempster, 1958)
- BS1996 (Bai and Saranadasa, 1996)
- CQ2010 (Chen and Qin, 2010)
- SD2008w (Srivastava and Du, 2008)
- SD2008wo (Srivastava, 2009)
- L1996 (Lauter, 1996)
- LJW2011 (Lopes, Jacob, and Wainwright, 2011)

Table 1: Power comparison for  $N_p(\mu, \Sigma_1)$ 

Methods	$c = 0$				$c = 0.5$				$c = 1$			
	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
<b>n = 40, p = 400</b>												
New	5.16	5.04	4.86	5.05	48.60	70.34	94.10	99.98	98.81	99.54	99.96	100.00
D1958	6.77	6.22	5.71	5.49	12.63	8.13	6.98	6.46	80.44	22.71	13.06	10.23
BS1996	7.73	7.80	7.79	7.80	14.64	10.55	9.39	9.11	88.21	30.28	18.51	15.33
CQ2010	7.72	7.82	7.79	7.77	14.64	10.50	9.41	9.11	88.18	30.22	18.50	15.32
SD2008w	4.20	1.71	0.52	0.15	7.97	2.29	0.63	0.22	54.21	6.41	1.29	0.36
SD2008wo	8.48	8.21	7.87	7.71	16.34	11.15	9.53	8.96	90.25	32.69	18.93	15.06
L1996	5.18	5.18	5.17	5.15	5.66	5.21	5.17	5.11	6.25	5.59	5.31	5.22
LJW2011	5.01	4.99	4.86	5.03	13.80	20.65	40.58	98.34	54.05	74.46	95.94	100.00
<b>n = 40, p = 1600</b>												
New	4.91	4.82	4.85	4.69	13.92	23.54	53.93	98.13	70.54	80.97	95.99	99.99
D1958	6.93	6.19	5.73	5.48	7.81	6.71	5.96	5.73	12.45	8.12	6.96	6.45
BS1996	7.74	7.79	7.78	7.79	8.85	8.30	8.12	8.06	14.47	10.42	9.35	8.92
CQ2010	7.76	7.80	7.77	7.77	8.88	8.32	8.14	8.05	14.49	10.41	9.34	8.92
SD2008w	2.76	0.69	0.14	0.00	3.22	0.73	0.17	0.00	5.26	0.97	0.20	0.01
SD2008wo	8.37	8.12	7.86	7.71	9.67	8.70	8.24	7.94	15.79	11.14	9.52	8.78
L1996	5.15	5.15	5.15	5.15	5.30	5.22	5.19	5.18	5.50	5.29	5.23	5.18
LJW2011	4.65	5.08	4.99	4.95	6.77	7.68	11.52	52.16	14.49	20.55	42.17	98.29

Monte Carlo error =  $1.96\sqrt{0.05 \times 0.95/10000} = 0.43\%$ .

Well-controlled Type I errors are expected to vary roughly between 4.57% and 5.43%.

Table 2: Power comparison for  $N_p(\mu, \Sigma_2)$ 

Methods	$c = 0$				$c = 0.5$				$c = 1$			
	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
<b>n = 40, p = 400</b>												
New	5.01	5.04	5.02	5.02	46.81	34.83	20.93	13.46	99.99	99.52	91.13	68.01
D1958	5.06	4.97	4.75	5.30	89.47	77.24	51.45	17.29	100.00	100.00	99.96	84.57
BS1996	5.57	5.57	5.46	6.86	90.19	78.40	53.88	20.81	100.00	100.00	99.99	88.16
CQ2010	5.59	5.57	5.44	6.85	90.16	78.39	53.83	20.81	100.00	100.00	99.99	88.15
SD2008w	3.75	3.68	3.30	2.72	84.86	70.93	44.71	9.94	100.00	100.00	99.85	68.93
SD2008wo	7.25	7.28	7.61	8.52	90.57	80.54	57.97	23.86	100.00	100.00	99.96	87.61
L1996	4.69	4.67	4.93	4.96	36.78	28.64	16.49	6.63	92.38	77.29	40.81	9.26
LJW2011	5.52	5.11	5.00	4.97	12.71	12.15	11.51	15.28	44.17	43.04	42.40	60.42
<b>n = 40, p = 1600</b>												
New	4.93	5.07	4.83	5.16	17.36	13.68	9.80	6.53	94.73	84.77	58.18	22.55
D1958	4.91	5.14	4.88	4.74	48.45	37.63	23.47	9.96	99.99	99.91	94.58	42.28
BS1996	5.05	5.46	5.36	5.49	49.13	38.40	24.63	11.40	99.99	99.91	94.96	45.81
CQ2010	5.08	5.48	5.29	5.50	49.26	38.35	24.60	11.44	99.99	99.91	94.94	45.74
SD2008w	1.77	1.91	2.04	1.81	30.97	22.82	12.73	3.66	99.92	99.03	86.28	23.53
SD2008wo	7.04	7.13	7.19	7.55	53.38	43.79	29.11	14.45	99.98	99.79	95.07	50.80
L1996	4.92	5.11	5.08	4.99	15.69	13.57	9.77	6.17	45.99	34.47	19.55	7.99
LJW2011	4.61	4.99	4.87	4.89	6.04	6.47	6.17	6.68	11.71	12.12	11.46	13.14

Monte Carlo error =  $1.96\sqrt{0.05 \times 0.95/10000} = 0.43\%$ .

Well-controlled Type I errors are expected to vary roughly between 4.57% and 5.43%.



Table 2a: Power comparison for  $N_p(\mu, \Sigma_2)$ 

Methods	$c = 0$				$c = 0.5$				$c = 1$			
	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
$n = 160, p = 400$												
New	4.77	4.95	4.76	4.91	100.0	99.50	90.60	96.54	100.0	100.0	100.0	100.0
D1958	4.61	4.97	5.12	5.34	100.0	100.0	100.0	85.83	100.0	100.0	100.0	100.0
BS1996	5.03	5.50	5.83	6.61	100.0	100.0	100.0	89.10	100.0	100.0	100.0	100.0
CQ2010	5.03	5.49	5.83	6.62	100.0	100.0	100.0	89.10	100.0	100.0	100.0	100.0
SD2008w	4.20	4.42	4.17	2.73	100.0	100.0	100.0	72.60	100.0	100.0	100.0	100.0
SD2008wo	5.41	5.78	6.19	6.93	100.0	100.0	100.0	88.85	100.0	100.0	100.0	100.0
L1996	4.87	4.71	4.70	5.00	89.99	71.70	34.04	7.34	100.0	100.0	71.60	10.28
LJW2011	4.65	4.95	4.75	5.27	89.44	85.36	80.43	98.54	100.0	100.0	100.0	100.0
$n = 160, p = 1600$												
New	4.50	4.82	4.60	4.84	97.50	89.89	61.44	36.12	100.0	100.0	100.0	99.54
D1958	4.73	4.72	4.99	5.11	99.99	99.89	95.03	42.55	100.0	100.0	100.0	100.0
BS1996	4.86	5.00	5.30	5.98	100.0	99.90	95.35	45.67	100.0	100.0	100.0	100.0
CQ2010	4.86	4.98	5.29	5.99	100.0	99.90	95.35	45.66	100.0	100.0	100.0	100.0
SD2008w	3.47	3.46	3.57	2.70	100.0	99.83	93.02	29.91	100.0	100.0	100.0	99.99
SD2008wo	5.40	5.48	5.65	6.33	100.0	99.87	95.47	46.65	100.0	100.0	100.0	100.0
L1996	5.27	5.08	4.84	4.78	42.34	31.61	16.88	6.42	97.49	83.24	39.61	9.04
LJW2011	4.86	4.67	4.56	5.45	25.35	24.48	23.41	37.24	92.06	90.94	90.47	98.77

Monte Carlo error =  $1.96\sqrt{0.05 \times 0.95/10000} = 0.43\%$ .

Well-controlled Type I errors are expected to vary roughly between 4.57% and 5.43%.

Table 3: Power comparison for  $N_p(\mu, \Sigma_3)$ 

Methods	$c = 0$				$c = 0.5$				$c = 1$			
	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
<b>n = 40, p = 400</b>												
New	5.04	4.97	5.24	5.19	42.03	41.23	34.91	29.71	99.04	98.27	97.16	93.76
D1958	6.82	6.68	6.38	6.02	26.75	12.64	9.38	8.05	99.93	78.81	38.79	21.82
BS1996	7.37	7.73	7.75	7.86	29.27	14.57	11.73	10.39	99.95	86.13	48.62	29.47
CQ2010	7.40	7.71	7.71	7.89	29.30	14.54	11.68	10.35	99.96	86.09	48.67	29.42
SD2008w	5.39	4.27	2.71	1.32	20.68	8.06	4.18	1.88	99.61	52.80	15.70	5.47
SD2008wo	8.15	8.36	8.34	8.27	33.33	15.93	12.72	10.85	99.96	88.48	52.41	32.24
L1996	5.07	5.17	5.13	5.19	6.12	5.57	5.33	5.22	8.10	6.30	5.80	5.57
LJW2011	4.86	5.20	4.95	5.06	12.83	14.60	15.63	26.31	48.15	53.40	60.76	86.04
<b>n = 40, p = 1600</b>												
New	4.76	4.89	5.20	5.42	13.05	12.51	11.88	8.75	70.52	67.60	65.63	45.33
D1958	7.17	6.90	6.48	6.19	9.58	7.78	7.18	6.81	27.30	12.45	9.31	8.28
BS1996	7.74	7.80	7.80	7.78	10.34	9.01	8.46	8.30	30.00	14.44	11.59	10.42
CQ2010	7.76	7.81	7.78	7.72	10.28	8.95	8.46	8.29	30.05	14.45	11.54	10.40
SD2008w	4.19	2.70	1.44	0.73	5.79	3.18	1.61	0.77	15.53	5.26	2.19	1.04
SD2008wo	8.48	8.43	8.32	8.19	11.70	9.76	9.05	8.82	34.82	15.94	12.59	11.17
L1996	5.10	5.14	5.19	5.20	5.39	5.27	5.21	5.21	5.71	5.44	5.30	5.31
LJW2011	5.00	4.84	4.79	5.23	6.68	6.81	7.05	8.80	13.07	14.15	16.65	23.86

Monte Carlo error =  $1.96\sqrt{0.05 \times 0.95/10000} = 0.43\%$ .

Well-controlled Type I errors are expected to vary roughly between 4.57% and 5.43%.

# Real data example

- Dataset

- Comparative study for the bone density pattern of two adjacent bones in mouse's skull.
- Source: Data were collected at Center for Quantitative X-Ray Imaging at Pennsylvania State University.
- Data: Bone volume measured at many different bone density levels by high-resolution micro-computed tomography technique.

- Pre-process:

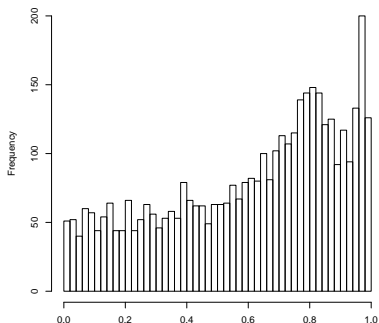
- Use volume difference at bone intensity levels from 10 to 100.
- Obtain normalized bone volume by divide the bone volume at each bone density level by the total bone volume
- Take difference of normalized bone volume at the corresponding density level for each subject.

The dimension and the sample size of this data set are  $p = 91$  and  $n = 22$ , respectively.

- the splitting percentage = 40% and  $\lambda = n_1^{-0.5}$ .
- All the tests have very small p-value, indicating strong evidence to reject the null.
- Set  $\bar{\mathbf{x}}$  to be the same mean, and  $\mathbf{r}_i = \mathbf{x}_i - \bar{\mathbf{x}}$ , the residual vector of the  $i$ -th subject, and then construct new data  $\mathbf{z}_i = \delta \bar{\mathbf{x}} + \mathbf{r}_i$  for  $i = 1, \dots, n$ .
- By the construction, smaller  $\delta$  results in weaker signals, and would make the test more challenging.

Table: P-values of tests  $(n, p) = (22, 91)$

$\delta$	New	D1958	BS1996	CQ2010	SD2008w	SD2008wo	L1996	LJW2011	LJW2012
1	0	0	0	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0	0.0005	0
0.6	0	0	0	0	0	0	0	0.0007	0
0.4	0.001	.0001	0	0	0	0	0.0008	0.0713	0.0002
0.2	0.003	0.098	0.0771	0.0782	0.2482	0.0583	0.1003	0.5337	0.6499



**Figure:** Histogram of absolute values of paired sample correlations among bone volumes at all different bone density levels.

Figure 1 indicates high correlations among variables. This may explain why the proposed test is more powerful than D1958, BS1996 and SD2010 because these tests ignore the correlation among variables.

Without normality assumption,  $T^2$  asymptotically follows  $\chi_p^2$  distribution.  
We can show that the optimal  $k = 1$  and the optimal direction  $\mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}$

For two-sample mean problem  $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  versus  $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ . We can show that the optimal  $k = 1$  and the optimal direction  $\mathbf{a} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ , the direction of Fisher's linear discriminant analysis.

Table 4: Power comparison for multivariate  $t_6$  with mean  $\mu$  and scatter matrix  $\Sigma_1$ 

Methods	$c = 0$				$c = 0.5$				$c = 1$			
	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
<b>n = 40, p = 400</b>												
New	5.09	4.94	4.88	4.71	37.02	57.48	89.13	99.94	96.51	98.59	99.86	100.00
D1958	5.07	5.44	5.16	5.05	8.52	7.12	6.27	5.84	38.09	14.37	9.86	8.29
BS1996	6.25	7.41	7.66	7.76	10.01	9.24	8.91	8.71	45.99	19.11	13.95	12.35
CQ2010	7.85	7.96	7.93	7.94	12.19	9.75	9.17	8.92	55.22	20.18	14.47	12.58
SD2008w	2.97	1.45	0.45	0.15	4.80	1.83	0.50	0.18	22.74	4.06	0.84	0.27
SD2008wo	7.14	7.86	7.74	7.57	11.36	9.75	9.06	8.52	54.58	21.00	14.34	12.11
L1996	4.66	4.80	4.81	4.82	5.17	4.95	4.85	4.79	5.89	5.34	5.00	4.90
LJW2011	4.33	4.15	4.36	4.23	11.65	16.81	34.72	96.43	45.75	65.28	91.90	100.00
<b>n = 40, p = 1600</b>												
New	4.72	4.62	4.69	4.71	11.88	18.94	44.98	96.04	61.34	74.40	92.50	99.97
D1958	5.44	5.81	5.57	5.37	6.04	6.09	5.76	5.54	8.52	7.16	6.36	5.99
BS1996	6.53	7.45	7.74	7.82	7.25	7.89	8.03	8.03	10.05	9.41	9.08	8.93
CQ2010	7.86	7.96	7.95	7.96	8.92	8.35	8.25	8.24	12.24	9.87	9.30	9.10
SD2008w	1.69	0.63	0.12	0.00	1.92	0.69	0.12	0.00	2.76	0.79	0.15	0.00
SD2008wo	7.31	7.97	7.85	7.64	8.16	8.42	8.12	7.88	11.55	9.90	9.14	8.68
L1996	5.12	5.18	5.20	5.21	5.18	5.14	5.16	5.19	5.28	5.17	5.14	5.15
LJW2011	4.35	4.36	4.39	4.29	5.93	7.02	9.91	45.00	12.36	17.23	36.15	96.25

Monte Carlo error =  $1.96\sqrt{0.05 \times 0.95/10000} = 0.43\%$ .

Well-controlled Type I errors are expected to vary roughly between 4.57% and 5.43%.

Table 5: Power comparison for multivariate  $t_6$  with mean  $\mu$  and scatter matrix  $\Sigma_2$

Methods	$c = 0$				$c = 0.5$				$c = 1$			
	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
<b><math>n = 40, p = 400</math></b>												
New	4.92	4.72	5.15	4.55	34.49	26.51	16.07	10.81	99.32	96.11	79.58	55.31
D1958	0.05	0.17	0.80	3.27	10.04	11.00	11.39	8.56	92.77	91.15	83.76	48.47
BS1996	0.08	0.25	1.02	4.65	12.42	13.43	14.05	11.29	94.45	92.97	86.68	55.35
CQ2010	5.49	5.71	5.83	6.77	68.47	55.52	36.29	15.44	100.00	99.97	97.97	64.15
SD2008w	0.04	0.05	0.37	1.34	7.13	8.03	8.00	3.84	91.59	89.34	79.00	33.06
SD2008wo	0.16	0.48	1.57	5.99	20.35	20.70	19.99	14.67	97.78	96.67	91.34	61.40
L1996	0.46	0.76	1.66	4.00	1.53	2.31	3.49	4.98	5.90	6.81	7.87	6.73
LJW2011	3.85	4.40	4.34	4.12	10.16	10.22	10.12	13.08	37.12	36.14	35.35	51.57
<b><math>n = 40, p = 1600</math></b>												
New	4.54	4.64	4.91	4.78	13.20	11.24	8.16	6.00	82.95	70.16	44.58	17.93
D1958	0.00	0.00	0.02	1.13	0.00	0.00	0.22	2.09	7.62	7.79	9.38	9.44
BS1996	0.00	0.00	0.06	1.58	0.00	0.00	0.30	2.72	9.59	9.90	11.96	11.86
CQ2010	5.07	5.16	5.23	5.93	30.83	24.57	16.62	9.58	98.44	94.23	75.89	29.10
SD2008w	0.00	0.00	0.00	0.11	0.00	0.00	0.02	0.30	1.74	2.05	2.91	2.41
SD2008wo	0.00	0.00	0.13	2.51	0.00	0.05	0.57	4.55	18.33	18.81	19.84	17.45
L1996	0.05	0.12	0.40	2.08	0.10	0.18	0.52	2.30	0.21	0.40	0.80	2.87
LJW2011	4.26	4.24	4.22	4.18	5.60	5.31	5.49	6.01	10.32	9.83	9.97	11.53

Monte Carlo error =  $1.96\sqrt{0.05 \times 0.95/10000} = 0.43\%$ .

Well-controlled Type I errors are expected to vary roughly between 4.57% and 5.43%.



Table 6: Power comparison for multivariate  $t_6$  with mean  $\mu$  and scatter matrix  $\Sigma_3$ 

Methods	$c = 0$				$c = 0.5$				$c = 1$			
	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
<b>n = 40, p = 400</b>												
New	5.01	5.19	5.24	4.88	31.76	31.21	27.25	22.56	96.29	95.28	92.42	85.93
D1958	3.39	5.02	5.51	5.40	8.88	8.38	7.51	6.95	78.22	37.80	20.45	13.58
BS1996	3.95	6.21	7.09	7.31	10.61	9.87	9.48	9.14	82.84	45.72	26.91	18.78
CQ2010	7.39	7.79	7.86	7.73	19.60	12.02	10.35	9.67	95.36	54.45	29.48	20.02
SD2008w	2.32	2.99	2.12	1.11	6.30	4.75	3.04	1.46	68.31	22.40	8.13	3.36
SD2008wo	4.89	7.07	7.71	7.74	13.44	11.53	10.47	9.61	90.14	54.13	30.77	20.85
L1996	4.07	4.66	4.86	4.83	4.88	5.13	5.05	4.92	6.48	5.94	5.58	5.30
LJW2011	4.11	4.41	3.92	4.70	10.21	11.67	12.90	22.49	40.10	45.59	52.01	78.51
<b>n = 40, p = 1600</b>												
New	4.97	4.68	5.08	5.06	10.21	10.33	9.94	7.48	59.66	57.14	54.29	36.00
D1958	3.55	5.45	5.70	5.73	4.40	6.01	6.18	6.11	8.72	8.47	7.65	7.19
BS1996	4.09	6.49	7.17	7.42	5.15	7.21	7.76	7.86	10.40	10.02	9.63	9.46
CQ2010	7.78	7.89	7.92	7.82	9.68	8.82	8.59	8.28	20.33	12.28	10.64	9.88
SD2008w	1.59	1.63	1.03	0.65	1.89	1.88	1.11	0.68	4.16	2.74	1.48	0.84
SD2008wo	5.04	7.22	7.82	7.85	6.30	8.12	8.46	8.37	13.12	11.46	10.54	9.99
L1996	4.45	5.11	5.15	5.18	4.57	5.11	5.18	5.14	4.76	5.25	5.24	5.20
LJW2011	4.27	4.09	4.21	4.36	5.68	5.62	6.36	7.28	11.38	11.96	14.38	19.23

Monte Carlo error =  $1.96\sqrt{0.05 \times 0.95/10000} = 0.43\%$ .

Well-controlled Type I errors are expected to vary roughly between 4.57% and 5.43%.

# Open problems

What is the asymptotic behavior of  $\hat{\mathbf{a}}$ ?

How to incorporate sparsity in the estimation of  $\mathbf{a}$ ?

Will multiple sample splitting improve the power while keep Type I error?

Can projection test be extended to other settings?

# Nonparametric/Robust test for HD data

Assume  $\mathbf{x}_i$  follows an elliptical distribution:

$$\mathbf{x}_i = \boldsymbol{\mu} + \epsilon_i, \quad \text{and} \quad \epsilon_i = \Gamma R_i \mathbf{u}_i,$$

where  $\Gamma$  is a  $p \times p$  matrix,  $\mathbf{u}_i$  is a random vector uniformly distributed on the unit sphere in  $\mathbb{R}^p$ , and  $R_i$  is a nonnegative random variable independent of  $\mathbf{u}_i$ .

The distribution of  $\mathbf{x}_i$  depends on  $\Gamma$  only through  $\Gamma\Gamma^T \triangleq \Omega$ .

In general,  $\mathbf{x}_i$ 's covariance matrix  $\Sigma$  is related to  $\Omega$  by  $\Sigma = p^{-1}E(R_i^2)\Omega$ .

An important special case of elliptical distribution is  $N(\boldsymbol{\mu}, \Sigma)$ , for which  $R_i^2 \sim \chi_p^2$  and  $\Omega = \Sigma$ .

# Chen and Qin (2010)'s test

For the independent component models, the test statistic of Bai and Saranadasa (1996) is based on  $\|\bar{\mathbf{x}}\|^2$ .

The test statistic of Chen and Qin (2010) is based on  $\sum_{i=1}^n \sum_{j=1, j \neq i}^n \mathbf{x}_i^T \mathbf{x}_j$ .

By removing the diagonal elements in  $\|\bar{\mathbf{x}}\|^2$ , Chen and Qin (2010) was able to considerably relax the conditions on  $p$  and  $n$ .

# New nonparametric test statistics

The **spatial sign function** of  $\mathbf{x}_i$  is defined as

$$\mathbf{z}_i = \begin{cases} \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}, & \mathbf{x}_i \neq 0, \\ 0, & \mathbf{x}_i = 0, \end{cases}$$

where  $\|\mathbf{x}_i\|^2 = \mathbf{x}_i^T \mathbf{x}_i$ . The spatial sign vector is simply the unit vector in the direction of  $\mathbf{x}_i$ . In the univariate case, it reduces to the familiar sign function. Define

$$T_n = \sum_{i=1}^n \sum_{j=1}^{i-1} \mathbf{z}_i^T \mathbf{z}_j,$$

which indeed is a  $U$ -statistic.

Under  $H_0$ ,  $E(\mathbf{z}_i) = 0$  which implies  $E(T_n) = 0$ .

The new test statistic dismiss the diagonal elements in defining  $T_n$ , and can be deemed as a nonparametric extension of Chen and Qin (2010).

# Connection to multivariate spatial sign test

The new test generalizes the multivariate spatial sign test (e.g., Brown, 1983; Chaudhuri, 1992; Möttönen and Oja, 1995) to the high-dimensional setting.

For fixed  $p$ , Möttönen, et al. (1997) derived the asymptotic relative efficiency (ARE) of the spatial sign test versus Hotelling's  $T^2$  test and established its theoretical advantage for heavy-tailed distributions:

For example, when  $\mathbf{x} \sim 10\text{-dim } t_\nu$ , the ARE of the spatial sign test versus  $T^2$  test is **2.42** when  $\nu = 3$ , and is **0.95** when  $\nu = \infty$  (multivariate normality).

However, similarly as Hotelling's  $T^2$  test, the multivariate spatial sign test is not defined when  $p > n$ .

Here we modify the sign test in a way such that its efficiency advantage can be preserved in the high-dimensional setting.

# Theoretical properties

Conditions:

$$(C1) \operatorname{Tr}(\Sigma^4) = o\{\operatorname{Tr}^2(\Sigma^2)\}.$$

$$(C2) \frac{\operatorname{Tr}^4(\Sigma)}{\operatorname{Tr}^2(\Sigma^2)} \exp\left\{-\frac{\operatorname{Tr}^2(\Sigma)}{128p\lambda_{\max}^2(\Sigma)}\right\} = o(1).$$

## Theorem

*Assume Conditions (C1) and (C2) hold. Then under  $H_0$ ,*

$$\frac{T_n}{\sqrt{\frac{n(n-1)}{2} \operatorname{Tr}(B^2)}} \rightarrow N(0, 1)$$

*in distribution, as  $n, p \rightarrow \infty$ , where  $B = E\left(\frac{\epsilon_i \epsilon_i^T}{\|\epsilon_i\|^2}\right)$ .*

# Comparison on dimensionality

Srivastava and Du (2008):  $n = O(p^\delta)$  for some  $1/2 < \delta \leq 1$ .

Srivastava (2009):  $n = O(p^\delta)$  for some  $0 < \delta \leq 1$ ,

Lee, Limb, Li, Vannuccid and Petkova (2012):  $p/n \rightarrow c > 0$

Srivastava, Katayama and Kano (2013):  $n = O(p^\delta)$ ,  $\delta > 1/2$ ,

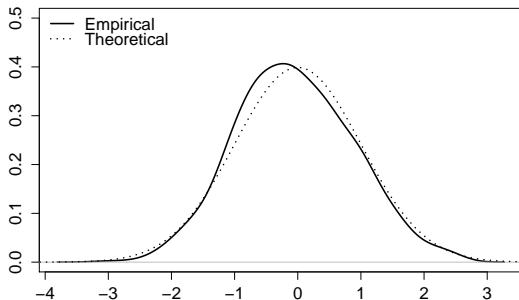
Chen and Qin (2010):  $\text{Trace}(\Sigma^4) = o(\text{Trace}^2(\Sigma^2))$ .



To apply  $T_n$  in practical data analysis, we need an estimator of  $\text{Tr}(B^2)$ . Following Chen and Qin (2010), we may estimate  $\text{Tr}(B^2)$  using the cross-validation approach as follows:

$$\widehat{\text{Tr}(B^2)} = \{n(n-1)\}^{-1} \text{Tr} \left\{ \sum_{1 \leq j \neq k \leq n} (\mathbf{z}_j - \bar{\mathbf{z}}_{(j,k)}) \mathbf{z}_j^T (\mathbf{z}_k - \bar{\mathbf{z}}_{(j,k)}) \mathbf{z}_k^T \right\},$$

where  $\bar{\mathbf{z}}_{(j,k)}$  is the sample mean after excluding  $\mathbf{z}_j$  and  $\mathbf{z}_k$ .



**Figure:** Density curves under the null hypothesis and  $\mathbf{x} \sim N(0, (0.8^{|i-j|}))$ .  
( $n = 50$ ,  $p = 1000$ )

# Local power analysis

For the local power analysis, we impose the following additional conditions.

$$(C3) \quad \exp\left(-\frac{\text{Tr}^2(\Sigma)}{256p\lambda_{\max}^2(\Sigma)}\right) = o\left(\min\left(\frac{\lambda_{\max}(\Sigma)}{\text{Tr}(\Sigma)}, \frac{\lambda_{\min}(\Sigma)}{\lambda_{\max}(\Sigma)}\right)\right).$$

$$(C4) \quad \lambda_{\max}(\Sigma) = o(\text{Tr}(\Sigma)).$$

$$(C5) \quad \|\mu\|^2 E(\|\epsilon\|^{-2}) = o\left(\min\left(n^{-1} \frac{\text{Tr}(\Sigma^2)}{\lambda_{\max}(\Sigma)\text{Tr}(\Sigma)}, n^{-1/2} \frac{\text{Tr}^{1/2}(\Sigma^2)}{\text{Tr}(\Sigma)}\right)\right).$$

$$(C6) \quad \text{For some } 0 < \delta < 1, \|\mu\|^{2\delta} E(\|\epsilon\|^{-2-2\delta}) = o(E^2(\|\epsilon\|^{-1})).$$

## Theorem

Assume conditions (C1)-(C6) hold. Letting  $A = E \left\{ \frac{1}{\|\epsilon_i\|} \left( I_p - \frac{\epsilon_i \epsilon_i^T}{\|\epsilon_i\|^2} \right) \right\}$ .  
Then as  $n, p \rightarrow \infty$ ,

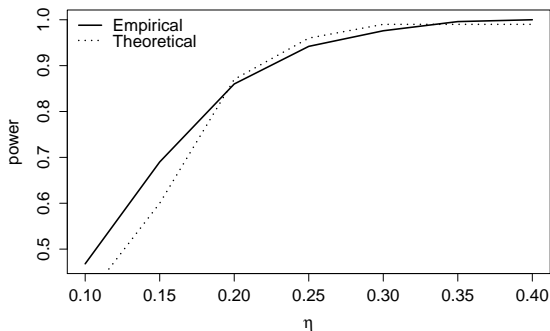
$$\frac{T_n - \frac{n(n-1)}{2} \boldsymbol{\mu}^T A^2 \boldsymbol{\mu} (1 + o(1))}{\sqrt{\frac{n(n-1)}{2} \text{Tr}(B^2)}} \rightarrow N(0, 1)$$

in distribution.

This theorem implies that under the local alternatives, the proposed test at level  $\alpha$  has the local power

$$\beta_n = \Phi \left( -z_\alpha + \sqrt{\frac{n(n-1)}{2}} \frac{\boldsymbol{\mu}^T A^2 \boldsymbol{\mu} (1 + o(1))}{\sqrt{\text{Tr}(B^2)}} \right),$$

where  $\Phi(\cdot)$  and  $z_\alpha$  denote the cumulative distribution function and the upper  $\alpha$  quantile of the  $N(0, 1)$  distribution, respectively.



**Figure:** Comparing the empirical distribution of the new test with the theoretical distribution ( $n = 50$ ,  $p = 1000$ ) when  $X \sim t_3(\Omega)$  with  $\Omega = (0.8^{|i-j|})$ . Here  $\eta = \mu^T A^2 \mu / \sqrt{\text{Tr}(B^2)}$ .

# Asymptotic Relative Efficiency (ARE)

For multivariate  $t$ -distribution  $t_\nu(\Omega)$  with  $\Omega = I_p$ , when  $p$  is large, the ARE of  $T_n$  versus the Chen and Qin's test is approximately

$$\text{ARE}_{T_n, \text{CQ}} \approx \frac{2}{\nu - 2} \left( \frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu/2)} \right)^2.$$

$\text{ARE} \approx 2.54, 1.76, 1.51, 1.38$  for  $\nu = 3, 4, 5, 6$ , respectively, and  $\text{ARE} \approx 1$  for  $\nu = \infty$  (i.e.,  $N(0, I_p)$ ).

Recall that for  $\nu = 3$ , the ARE of the classical spatial sign test versus Hotelling's  $T^2$  is 2.02 for  $p = 3$  and 2.09 for  $p = 10$  in the fixed dimensional case.

This suggests that nonparametric test may have more substantial power gain in the high-dimensional case.

# Beyond Elliptical Distribution

*Symmetric independent component models* (e.g., Ilmonen and Paindaveine, 2011):

$$\mathbf{x}_i = \mu + \Gamma \mathbf{z}_i,$$

where  $\Gamma$  is a full rank  $p \times p$  positive definite matrix;  $\mathbf{z}_i = (Z_{i1}, \dots, Z_{ip})'$  has independent components  $Z_{ij}$  and  $Z_{ij}$  is symmetric about zero. Assume that  $Z_{ij}$  are standardized such that  $\text{Var}(Z_{ij}) = 1$ .

## Theorem

*Assume (C1) and (C2) hold for symmetric independent component models,  $E(Z_{ij}^4)$  is finite, and  $\text{Tr}(\Sigma^2) = o(\text{Tr}^2(\Sigma))$ . Then*

$$\frac{T_n}{\sqrt{\frac{n(n-1)}{2} \text{Tr}(B^2)}} \rightarrow N(0, 1) \text{ in distribution under } H_0, \text{ as } n, p \rightarrow \infty.$$

- Goal: to compare the performance of the new test with four alternatives:
- (1) the test of Chen and Qin (CQ test, 2010),
  - (2) the test of Srivastava, Katayama and Kano (SKK test, 2013),
  - (3) the test based on multiple comparison with Bonferroni correction (BF test),
  - (4) the test based on multiple comparison with FDR control (FDR test, Benjamini and Hochberg, 1995).

The BF test controls Type I error rate at 0.05 and the FDR test controls the false discovery rate at 0.05. Both the BF test and FDR test are computed using the p-values from the  $t$  tests for the marginal hypotheses and reject  $H_0$  if at least one marginal test is significant.



# Simulation Studies

Sample size:  $n = 20, 50$

Dimension:  $p = 200, 1000$  and  $2000$ . To save space, we report the results for  $p = 1000$  and  $2000$  here.

The performance of the five tests are evaluated on 1000 simulation runs.

## Example 1. $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$

The three choices for  $\boldsymbol{\mu}$  are:

- (1) the null hypothesis  $\boldsymbol{\mu}_0 = (0, \dots, 0)^T$ ;
- (2) the alternative  $\boldsymbol{\mu}_1 = (0.25, 0.25, \dots, 0.25)^T$ ;
- (3) the alternative  $\boldsymbol{\mu}_2 = (\boldsymbol{\mu}_{21}, \dots, \boldsymbol{\mu}_{2p})^T$  with  $\boldsymbol{\mu}_{21} = \dots = \boldsymbol{\mu}_{2\frac{p}{3}} = 0$ ,  $\boldsymbol{\mu}_{2(\frac{p}{3}+1)} = \dots = \boldsymbol{\mu}_{2(\frac{2p}{3})} = 0.2$  and  $\boldsymbol{\mu}_{2(\frac{2p}{3}+1)} = \dots = \boldsymbol{\mu}_{2p} = -0.2$ .

The three choices for  $\Sigma$  are:

- (1)  $\sigma_{ii} = 1$  and  $\sigma_{ij} = 0.2$  ( $i \neq j$ ); (denoted by  $\Sigma_1$  in tables)
- (2)  $\sigma_{ij} = 0.8^{|i-j|}$ ; (denoted by  $\Sigma_2$ )
- (3)  $\Sigma = DRD$ , where  $D = \text{diag}(d_1, \dots, d_p)$  with  $d_i = 2 + (p - i + 1)/p$ ,  $R = (r_{ij})$  with  $r_{ii} = 1$  and  $r_{ij} = (-1)^{i+j}(0.2)^{|i-j|^{0.1}}$  for  $i \neq j$ . Denoted by  $\Sigma_3$ . This was considered in Srivastava, Katayama and Kano (2013).

Table 1a: Example 1 - multivariate normal distribution

$\Sigma$	$\mu$	$n$	$p$	New	CQ	SKK	BF	FDR
$\Sigma_1$	$\mu_0$	20	1000	0.066	0.069	0.061	0.046	0.046
		20	2000	0.073	0.070	0.046	0.052	0.053
		50	1000	0.059	0.060	0.043	0.035	0.038
		50	2000	0.058	0.061	0.029	0.043	0.047
	$\mu_1$	20	1000	0.723	0.723	0.692	0.405	0.471
		20	2000	0.720	0.729	0.638	0.385	0.447
		50	1000	0.975	0.976	0.962	0.842	0.890
		50	2000	0.970	0.976	0.945	0.850	0.901
	$\mu_2$	20	1000	0.951	0.826	0.650	0.382	0.443
		20	2000	0.954	0.821	0.567	0.404	0.464
		50	1000	1.000	1.000	1.000	0.964	0.997
		50	2000	1.000	1.000	1.000	0.973	0.998

Table 1b: Example 1 - multivariate normal distribution

$\Sigma$	$\mu$	$n$	$p$	New	CQ	SKK	BF	FDR
$\Sigma_2$	$\mu_0$	20	1000	0.055	0.055	0.067	0.052	0.052
		20	2000	0.059	0.061	0.048	0.042	0.044
		50	1000	0.052	0.052	0.061	0.042	0.044
		50	2000	0.045	0.045	0.052	0.048	0.048
	$\mu_1$	20	1000	0.490	0.438	0.514	0.127	0.132
		20	2000	0.646	0.594	0.556	0.113	0.119
		50	1000	1.000	0.998	1.000	0.296	0.331
		50	2000	1.000	1.000	1.000	0.331	0.369
	$\mu_2$	20	1000	0.242	0.225	0.335	0.100	0.104
		20	2000	0.342	0.310	0.318	0.084	0.092
		50	1000	0.932	0.862	0.982	0.239	0.260
		50	2000	0.991	0.987	1.000	0.269	0.296

Table 1c: Example 1 - multivariate normal distribution

$\Sigma$	$\mu$	$n$	$p$	New	CQ	SKK	BF	FDR
$\Sigma_3$	$\mu_0$	20	1000	0.052	0.051	0.047	0.038	0.041
		20	2000	0.058	0.059	0.023	0.047	0.047
		50	1000	0.048	0.050	0.051	0.043	0.048
		50	2000	0.060	0.060	0.052	0.054	0.061
	$\mu_1$	20	1000	0.795	0.797	0.815	0.122	0.138
		20	2000	0.969	0.968	0.930	0.134	0.145
		50	1000	0.999	0.999	0.999	0.357	0.430
		50	2000	1.000	1.000	1.000	0.416	0.479
	$\mu_2$	20	1000	0.540	0.549	0.579	0.092	0.102
		20	2000	0.790	0.788	0.695	0.102	0.112
		50	1000	0.992	0.991	0.994	0.265	0.289
		50	2000	1.000	1.000	1.000	0.343	0.381

## Example 2: $\mathbf{x} \sim t_3(\Sigma)$

We simulate  $\mathbf{x}_i$  from a  $p$ -variate  $t$  distribution with mean vector  $\boldsymbol{\mu}$ , covariance matrix  $\Sigma$  and 3 degrees of freedom.

The choices of  $\boldsymbol{\mu}$  and  $\Sigma$  are set to be the same as those in Example 1. The distribution is heavy-tailed in this example.

Table 2a: Example 2 - multivariate  $t$ -distribution

$\Sigma$	$\mu$	$n$	$p$	New	CQ	SKK	BF	FDR
$\Sigma_1$	$\mu_0$	20	1000	0.083	0.088	0.012	0.011	0.013
		20	2000	0.064	0.072	0.007	0.010	0.010
		50	1000	0.053	0.063	0.015	0.011	0.012
		50	2000	0.069	0.076	0.008	0.010	0.010
	$\mu_1$	20	1000	0.633	0.472	0.222	0.153	0.183
		20	2000	0.631	0.468	0.171	0.117	0.138
		50	1000	0.941	0.721	0.493	0.424	0.491
		50	2000	0.921	0.736	0.448	0.438	0.492
	$\mu_2$	20	1000	0.815	0.371	0.076	0.129	0.150
		20	2000	0.830	0.363	0.040	0.107	0.122
		50	1000	1.000	0.803	0.333	0.427	0.485
		50	2000	1.000	0.825	0.288	0.493	0.571

Table 2b: Example 2 - multivariate  $t$ -distribution

$\Sigma$	$\mu$	$n$	$p$	New	CQ	SKK	BF	FDR
$\Sigma_2$	$\mu_0$	20	1000	0.054	0.058	0.001	0.023	0.023
		20	2000	0.061	0.057	0.000	0.011	0.011
		50	1000	0.056	0.057	0.002	0.015	0.015
		50	2000	0.048	0.042	0.001	0.015	0.015
	$\mu_1$	20	1000	0.355	0.174	0.001	0.033	0.034
		20	2000	0.488	0.224	0.003	0.022	0.022
		50	1000	0.979	0.465	0.021	0.080	0.084
		50	2000	0.998	0.624	0.015	0.095	0.103
	$\mu_2$	20	1000	0.198	0.113	0.001	0.034	0.034
		20	2000	0.251	0.141	0.000	0.018	0.019
		50	1000	0.766	0.249	0.010	0.059	0.064
		50	2000	0.922	0.349	0.005	0.070	0.073



Table 2c: Example 2 - multivariate  $t$ -distribution

$\Sigma$	$\mu$	$n$	$p$	New	CQ	SKK	BF	FDR
$\Sigma_3$	$\mu_0$	20	1000	0.052	0.053	0.000	0.011	0.012
		20	2000	0.058	0.060	0.000	0.013	0.013
		50	1000	0.052	0.053	0.000	0.015	0.015
		50	2000	0.059	0.060	0.000	0.020	0.021
	$\mu_1$	20	1000	0.682	0.349	0.013	0.029	0.033
		20	2000	0.883	0.512	0.002	0.038	0.040
		50	1000	0.996	0.780	0.120	0.094	0.102
		50	2000	1.000	0.933	0.091	0.118	0.128
	$\mu_2$	20	1000	0.441	0.228	0.001	0.027	0.029
		20	2000	0.654	0.339	0.000	0.027	0.028
		50	1000	0.942	0.570	0.037	0.067	0.071
		50	2000	0.998	0.785	0.018	0.077	0.086

## Example 3: scale mixture of two normals

We simulate  $\mathbf{x}_i$  from a scale mixture of two multivariate normal distributions  $0.9 * N_p(\boldsymbol{\mu}, \Sigma) + 0.1 * N_p(\boldsymbol{\mu}, 9\Sigma)$ .

We consider the same choices of  $\boldsymbol{\mu}$  and  $\Sigma$  as in Example 1.

Table 3a: Example 3 - mixture of multivariate normal distributions

$\Sigma$	$\mu$	$n$	$p$	New	CQ	SKK	BF	FDR
$\Sigma_1$	$\mu_0$	20	1000	0.063	0.070	0.007	0.014	0.015
		20	2000	0.045	0.049	0.007	0.015	0.018
		50	1000	0.063	0.066	0.014	0.020	0.021
		50	2000	0.042	0.040	0.007	0.015	0.016
	$\mu_1$	20	1000	0.649	0.548	0.277	0.209	0.259
		20	2000	0.627	0.542	0.229	0.193	0.231
		50	1000	0.941	0.859	0.730	0.600	0.682
		50	2000	0.964	0.867	0.701	0.619	0.687
	$\mu_2$	20	1000	0.870	0.449	0.109	0.175	0.201
		20	2000	0.882	0.492	0.089	0.224	0.243
		50	1000	1.000	0.966	0.700	0.663	0.759
		50	2000	1.000	0.968	0.577	0.698	0.800

Table 3b: Example 3 - mixture of multivariate normal distributions

$\Sigma$	$\mu$	$n$	$p$	New	CQ	SKK	BF	FDR
$\Sigma_2$	$\mu_0$	20	1000	0.054	0.053	0.008	0.020	0.020
		20	2000	0.039	0.045	0.002	0.022	0.022
		50	1000	0.050	0.050	0.000	0.026	0.026
		50	2000	0.035	0.034	0.000	0.021	0.022
	$\mu_1$	20	1000	0.342	0.207	0.045	0.058	0.061
		20	2000	0.493	0.307	0.072	0.046	0.049
		50	1000	0.995	0.730	0.050	0.124	0.129
		50	2000	1.000	0.879	0.035	0.111	0.119
	$\mu_2$	20	1000	0.178	0.130	0.030	0.046	0.050
		20	2000	0.249	0.173	0.043	0.047	0.048
		50	1000	0.797	0.421	0.012	0.100	0.110
		50	2000	0.947	0.565	0.011	0.092	0.093

Table 3c: Example 3 - mixture of multivariate normal distributions

$\Sigma$	$\mu$	$n$	$p$	New	CQ	SKK	BF	FDR
$\Sigma_3$	$\mu_0$	20	1000	0.046	0.063	0.006	0.019	0.019
		20	2000	0.050	0.047	0.004	0.017	0.018
		50	1000	0.039	0.049	0.000	0.021	0.023
		50	2000	0.041	0.039	0.001	0.020	0.021
	$\mu_1$	20	1000	0.678	0.485	0.093	0.059	0.067
		20	2000	0.914	0.700	0.126	0.069	0.070
		50	1000	0.998	0.943	0.230	0.157	0.177
		50	2000	1.000	0.995	0.167	0.190	0.211
	$\mu_2$	20	1000	0.437	0.285	0.059	0.050	0.053
		20	2000	0.687	0.478	0.088	0.047	0.050
		50	1000	0.953	0.759	0.045	0.111	0.118
		50	2000	0.998	0.940	0.040	0.150	0.169

# An application

Type 2 diabetes is one of the most common chronic diseases.

Insulin resistance in skeletal muscle is a prominent feature of Type 2 diabetes.

To study insulins ability to regulate gene expression, an experiment performed microarray analysis using the Affymetrix Hu95A chip of human skeletal muscle biopsies from 15 diabetic patients both before and after insulin treatment (Wu et al., 2007).

The gene expression alterations are promising to provide insights on new therapeutic targets for the treatment of this common disease.

Hence, we are interested in testing the hypothesis whether there is a change of the gene expression level due to the treatment.

The underlying genetics of Type 2 diabetes were recognized to be very complex.

It is believed that Type 2 diabetes is resulted from interactions between many genetic factors and the environment. The data were normalized by quantile normalization.

When multiple probes are associated with the same gene, their expression values are consolidated by taking the average.

In our analysis, we considered 2519 curated gene sets. The gene sets we used are from the C2 collection of the GSEA online pathway databases ([http://www.broadinstitute.org/gsea/msigdb/collection\\_details.jsp#C2](http://www.broadinstitute.org/gsea/msigdb/collection_details.jsp#C2)). The largest gene set contains 1607 genes, which makes the hypothesis testing problem a high-dimensional one.

We applied both the new test and the CQ test at 5% significance level with the Bonferroni correction to control the family-wise error rate at 0.05 level.

For the CQ method, 520 gene sets (20.64% of all candidates) are identified as significant

For the new method, 954 gene sets (37.87% of all candidates) are selected as significant.

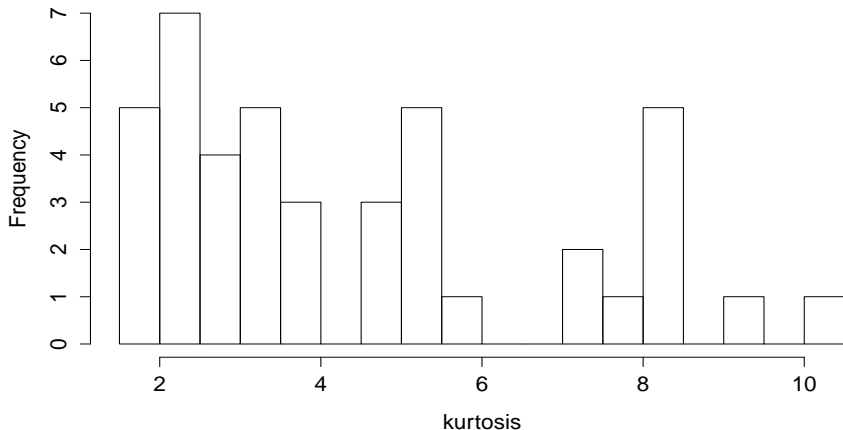
We observe that the significant gene sets selected by the new test include those identified by the CQ test with only one exception (HASLINGER\_B\_CLL\_WITH\_CHROMOSOME\_12\_TRISOMY).



**Table:** The top 10 significant gene sets selected by the new test and the CQ test

Gene set	New test	CQ test
ZWANG_CLASS_2_TRANSIENTLY_INDUCED_BY_EGF	24.34	20.11
NAGASHIMA_EGF_SIGNALING_UP	22.44	17.01
SHIPP_DLBCL_CURED_VS_FATAL_DN	22.34	18.24
WILLERT_WNT_SIGNALING	19.66	
UZONYI_RESPONSE_TO_LEUKOTRIENE_AND_THROMBIN	19.63	18.46
PID_HIF2PATHWAY	19.46	15.65
PHONG_TNF_TARGETS_UP	19.21	18.90
AMIT_EGF_RESPONSE_60_HELA	18.64	16.38
MCCLUNG_CREB1_TARGETS_DN	18.43	
SEMENZA_HIF1_TARGETS	18.34	
AMIT_SERUM_RESPONSE_40_MCF10A		15.98
AMIT_SERUM_RESPONSE_60_MCF10A		15.43
PLASARI_TGFB1_TARGETS_1HR_UP		15.00

## histogram of kurtosis



**Figure:** The histogram of marginal kurtosises for all genes in MCCLUNG\_CREB1\_TARGETS\_DN gene set.

## 2. Test of covariance structures

Test of covariance structure is of great interest in classical multivariate data analysis:

- Chapters 8 and 11 of Muirhead (1982)
- Chapters 9 and 10 of Anderson (2003)

## Earlier work on this topic

Ledoit and Wolf (2002, AoS) studied two tests on covariance

$$H_{10} : \Sigma = \sigma^2 I \quad \text{versus} \quad H_{11} : \Sigma \neq \sigma^2 I$$

and

$$H_{10a} : \Sigma = I \quad \text{versus} \quad H_{11a} : \Sigma \neq I$$

when we have an IID sample from  $N_p(\mu, \Sigma)$ , where  $I$  is the  $p \times p$  identity matrix.

The idea can be directly applicable for  $H_{10} : \Sigma = \sigma^2 \Sigma_0$  with known  $\Sigma_0$ .

$\bar{\mathbf{x}}$ : the sample mean,  $\mathbf{S}$ : the sample covariance matrix

The LRTs for  $H_{10}$  and  $H_{10a}$  are

$$U = \frac{1}{p} \text{tr} \left[ \left\{ \frac{\mathbf{S}}{p^{-1} \text{tr}(\mathbf{S})} - I \right\}^2 \right] \quad \text{and} \quad V = \frac{1}{p} \text{tr}\{(\mathbf{S} - I)^2\}$$

# A quick take-home message

Dimensionality:  $p/n \rightarrow y \in (0, \infty)$ .  $p$  and  $n$  have the same order.

$U$  is robust against high dimensionality, but  $V$  is not.

Denote by  $\lambda_1, \dots, \lambda_p$  the eigenvalues of  $\Sigma$ .

Let  $\bar{\lambda} = p^{-1} \sum_{j=1}^p \lambda_j = p^{-1} \text{tr}(\Sigma)$  and  $s_{\lambda}^2 = p^{-1} \sum_{j=1}^p (\lambda_j - \bar{\lambda})^2$ .

This paper shows that under some conditions (the sample 3rd and 4th moments of  $\lambda_j$ s have finite limites

$$p^{-1} \text{tr}(S) \rightarrow \bar{\lambda} \quad \text{and} \quad p^{-1} \text{tr}(S^2) \rightarrow (1 + y) \bar{\lambda}^2 + s_{\lambda}^2.$$

in probability.

Impact of dimensionality?

$s_{\lambda}^2 = 0$  if  $\Sigma = I$ .

$$U = \frac{p^{-1}\text{tr}(\mathbf{S}^2)}{[p^{-1}\text{tr}(\mathbf{S})]^2} - 1 \rightarrow y + \frac{s_\lambda^2}{\bar{\lambda}^2}$$

Corrected LRT is to take  $y$  into consideration by approximating  $y$  by  $p/n$ . This constitutes an  $(n, p)$ -consistent test.

The authors proposed to modify  $V$  by

$$W = \frac{1}{p}\text{tr}\{(\mathbf{S} - I)^2\} - \frac{p}{n}\{p^{-1}\text{tr}(\mathbf{S})\}^2 + p/n$$

Adding  $p/n$  makes  $W$  have a limiting distribution of  $\chi^2_{p(p+1)/2}$ .

This paper assumes that (1)  $n = O(p^\delta)$  with  $0 < \delta \leq 1$ , and (2) normality. Let  $a_j = p^{-1} \text{tr}(\Sigma^j)$ . The author proved that

$$\hat{a}_1 = p^{-1} \text{tr} \mathbf{S}$$

and

$$\hat{a}_2 = \frac{n^2}{(n-1)(n+2)} \frac{1}{p} \{ \text{tr} \mathbf{S}^2 - \frac{1}{n} (\text{tr} \mathbf{S})^2 \}$$

are unbiased and consistent estimate for  $a_1$  and  $a_2$ .

Apply these results for testing  $H_{10}$ ,  $H_{10a}$  and test of  $\Sigma$  being diagonal (test of independence under multinormality).

## Some more works

Birke and Detter (2005, Statist & Probab. Letter) considered  $H_{10}$  and  $H_{10a}$  under normality assumption with  $p/n \rightarrow y$ , and  $y = 0$  or  $\infty$ .

Schott (2007, CSDA) consider  $H_0 : \Sigma_1 = \cdots = \Sigma_K$  under normality assumption.

LRT for this null hypothesis is

$$M = n \log |\mathbf{S}| - \sum_{k=1}^K n_k \log |\mathbf{S}_k|$$

where  $\mathbf{S}$  is the pooled sample covariance matrix, and  $\mathbf{S}_k$  is the sample covariance matrix from the  $k$ -th population.

**Problem:**  $\mathbf{S}_k$  becomes singular if  $n_k < p$ .



He proposed using Wald statistic:

$$W = \frac{n}{2} \left\{ \sum_{k=1}^K \frac{n_k}{n} \text{tr}(\mathbf{S}_k \mathbf{S}^{-1} \mathbf{S}_k \mathbf{S}^{-1}) - \sum_{k=1}^K \sum_{l=1}^K \frac{n_k n_l}{n^2} \text{tr}(\mathbf{S}_k \mathbf{S}^{-1} \mathbf{S}_l \mathbf{S}^{-1}) \right\}$$

if  $\sum_{k=1}^K n_k \geq p$ .

# LRT under normality assumption

Under multinormality assumption, Jiang and Yang (2013) obtained the limiting distribution of LRT tests studied in Chapter 8 of Muirhead (1982) for  $p/n \rightarrow y \in (0, 1)$ . and Jiang and Qi (2015) further obtained the limiting distribution when  $y = 1$ .

Cai and Ma (2013) proposed a further corrected LRT that allows  $p/n \rightarrow \infty$ .

# Test of covariance for independent component models

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{w},$$

where  $\mathbf{w}$  has iid component with  $E(w_j) = 0$ ,  $\text{var}(w_j) = 1$ , and  $E(w_j^4) = \kappa < \infty$ .

Major probability tools: random matrix theory.

**For  $H_{10a} : \Sigma = I$ :**

Bai, Jiang, Yao and Zheng (2009, AoS) proposed correcting LRT when  $p/n \rightarrow y \in (0, 1)$ .

Wang, Yang, Miao and Cao (2013) also studied LRT

**For  $H_{10}$ , test of sphericity:**

Chen, Zhang and Zhong (2010, JASA) demonstrated the classical LRT may become invalid for HD data and proposed a test based on U-statistics.

Wang and Yao (2013) proposed the corrected LRT with  $y \in (0, 1)$  and corrected John's test with  $y \in (0, \infty)$ .

# Test of linear structures

Develop two tests on linear covariance structure for HD data.

For a set of pre-specified symmetric  $p \times p$  matrices  $(\mathbf{A}_1, \dots, \mathbf{A}_K)$  with fixed and finite  $K$ ,

$$H_0 : \Sigma = \theta_1 \mathbf{A}_1 + \theta_2 \mathbf{A}_2 + \dots + \theta_K \mathbf{A}_K,$$

where  $\{\theta_j, j = 1, \dots, K\}$  are unknown parameters

- *Assumption A. (ICM)* Assume that the population  $\mathbf{X}$  can be represented as  $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{w}$ , where  $\mathbf{w} = (w_1, \dots, w_p)^T$ , and  $w_1, \dots, w_p$  being IID and  $E(w_j) = 0$ ,  $E(w_j^2) = 1$  and  $E(w_j^4) = \kappa < \infty$ .
- *Assumption B.* Denote  $y_n = p/n$ . Assume that  $y_n \rightarrow y \in (0, \infty)$ .

# Parameter estimation

Without a likelihood, we propose estimating  $\theta$  by minimizing the following squared loss function

$$\min_{\theta} \text{tr}(\mathbf{S}_n - \theta_1 \mathbf{A}_1 - \dots - \theta_K \mathbf{A}_K)^2, \quad (1)$$

where  $\mathbf{S}_n$  is the sample covariance matrix.

Let  $\mathbf{C}$  be a  $K \times K$  matrix with  $(i, j)$ -element being  $\text{tr} \mathbf{A}_i \mathbf{A}_j$  and  $\mathbf{a}$  be a  $K \times 1$  vector with  $j$ -th element being  $\text{tr} \mathbf{S}_n \mathbf{A}_j$ . Minimizing (1) yields a least squares type estimate for  $\theta$ :

$$\hat{\theta} = \mathbf{C}^{-1} \mathbf{a}. \quad (2)$$

We can show that under Assumptions A and B,  $\hat{\theta}_k = \theta_k + O_p(n^{-1})$ ,  $k = 1, \dots, K$ .

Denote  $\Sigma_0 = \theta_1 \mathbf{A}_1 + \theta_2 \mathbf{A}_2 + \dots + \theta_K \mathbf{A}_K$ .

$$H_0 : \Sigma = \Sigma_0.$$

Under  $H_0$ , we estimate  $\boldsymbol{\theta}$  by  $\hat{\boldsymbol{\theta}}$  given in (2), and then an estimator of  $\Sigma$  is  $\hat{\Sigma}_0 = \hat{\theta}_1 \mathbf{A}_1 + \dots + \hat{\theta}_K \mathbf{A}_K$ .

Without the linear structure assumption, a natural estimator for  $\Sigma$  is the sample covariance matrix  $\mathbf{S}_n$ .

# Entropy loss based test

Motivated by the entropy loss (EL) used for covariance matrix estimation (James and Stein, 1961; Muirhead, 1982), we propose our first test for  $H_0$ . For  $p < n - 1$ ,

$$T_{n1} = \text{tr} \mathbf{S}_n \hat{\Sigma}_0^{-1} - \log(|\mathbf{S}_n \hat{\Sigma}_0^{-1}|).$$

Denote  $\{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p\}$  to be the eigenvalues of  $\hat{\Sigma}_0^{-1/2} \mathbf{S}_n \hat{\Sigma}_0^{-1/2}$ . Then we can write  $T_{n1}$  as

$$T_{n1} = p \left( p^{-1} \sum_{j=1}^p \lambda_j - p^{-1} \sum_{j=1}^p \log \lambda_j \right).$$

This motivates us to further extend the test to the situation with  $p \geq n - 1$  by defining

$$T_{n1} = (n - 1) \left( p^{-1} \sum_{j=1}^{n-1} \lambda_j - (n - 1)^{-1} \sum_{j=1}^{n-1} \log \lambda_j \right).$$



# Entropy loss based test

Define  $q = \min\{p, n - 1\}$ .  $T_{n1}$  can be written in a unified form for  $p < n - 1$  and  $p \geq n - 1$ :

$$T_{n1} = q \left( p^{-1} \sum_{j=1}^q \lambda_j - q^{-1} \sum_{j=1}^q \log \lambda_j \right). \quad (3)$$

Since this test is motivated by the entropy loss, we refer this test to be as **EL-test**.

# Quadratic loss based test

Motivated by the quadratic loss (QL), another popular loss function in covariance matrix estimation (Olkin and Selliah, 1977; Haff, 1980; Muirhead, 1982), we propose the second test

$$T_{n2} = \text{tr}(\mathbf{S}_n \hat{\Sigma}_0^{-1} - \mathbf{I}_p)^2, \quad (4)$$

and refer this test to be as **QL-test**.

## Example 1. Test of sphericity: $H_{10} : \Sigma = \sigma^2 I_p$

Chen, Zhang & Zhong (2010) demonstrated that the classical LRT may become invalid for HD data and proposed a test based on U-statistics with  $p, n \rightarrow \infty$ .

Jiang & Yang (2013) derived the asymptotic distribution of the LRT with  $p/n \rightarrow (0, 1]$  under normality assumption on  $\mathbf{w}$ .

Wang & Yao (2013) proposed the corrected LRT with  $p/n \rightarrow (0, 1)$  and the corrected John's (CJ) test with  $p/n \rightarrow y \in (0, \infty)$

The CJ test behaves similarly with the proposal by Chen, Zhang & Zhong (2010) on powers as  $p/n \rightarrow y \in (0, \infty)$  and the corrected LRT had greater powers than the CJ test and Chen, Zhang & Zhong (2010)'s test when the dimension  $p$  is not large relative to the sample size  $n$ . But when  $p$  is large relative to  $n$  ( $p < n$ ), the corrected LRT had less powers than the CJ test and the test proposed in Chen, Zhang & Zhong (2010).

## Example 1. Test of sphericity: $H_{10} : \Sigma = \sigma^2 I_p$

Under  $H_{10}$ ,  $\sigma^2$  can be estimated by  $\hat{\sigma}^2 = p^{-1} \text{tr} \mathbf{S}_n$ . It is easy to see that  $\hat{\sigma}^2 = \sigma^2 + O_p(n^{-1})$  under the condition that  $p/n \rightarrow y \in (0, \infty)$ . Let  $\{\lambda_1 \geq \lambda_2 \dots \geq \lambda_p\}$  be the eigenvalues of  $\mathbf{S}_n / (p^{-1} \text{tr} \mathbf{S}_n)$ . The EL and QL tests have the following expressive forms:

$$T_{n1} = q \left( p^{-1} \sum_{j=1}^q \lambda_j - q^{-1} \sum_{j=1}^q \log \lambda_j \right), \quad (5)$$

where  $q = \min\{p, n-1\}$ , and

$$T_{n2} = \text{tr}[\mathbf{S}_n / (p^{-1} \text{tr} \mathbf{S}_n) - \mathbf{I}_p]^2. \quad (6)$$

$T_{n1}$  is equivalent to the LRT under normality assumption and  $p < n-1$ .  
 $T_{n2}$  in (6) coincides with the CJ test proposed by Wang & Yao (2013).

## Example 2: Compound symmetric structure

Let  $\mathbf{A}_1 = \mathbf{I}_p$  and  $\mathbf{A}_2 = \mathbf{1}_p \mathbf{1}_p^T$ , where  $\mathbf{1}_p$  stands for a  $p$ -dimensional column vectors with all elements being 1. Testing compound symmetric structure is to test

$$H_{20} : \Sigma = \theta_1 \mathbf{A}_1 + \theta_2 \mathbf{A}_2,$$

where  $\theta_1 > 0$  and  $-1/(p-1) < \theta_2/(\theta_1 + \theta_2) < 1$ .

Under normality assumption, Kato, Yamada and Fujikoshi (2010) studied the asymptotic behavior of the corresponding likelihood ratio test when  $p < n$ , and Srivastava and Reid (2012) proposed a new test statistic for  $H_{20}$  when  $p \geq n$ .

## Example 2: Compound symmetric structure

Under  $H_{20}$ ,

$$\hat{\theta}_1 = p^{-1}(p-1)^{-1}(p \operatorname{tr} \mathbf{S}_n - \mathbf{1}_p^T \mathbf{S}_n \mathbf{1}_p), \quad \hat{\theta}_2 = p^{-1}(p-1)^{-1}(\mathbf{1}_p^T \mathbf{S}_n \mathbf{1}_p - \operatorname{tr} \mathbf{S}_n).$$

respectively. It can be shown that  $\hat{\theta}_k = \theta_k + O_p(n^{-1})$ ,  $k = 1, 2$

$T_{n1}$  is equivalent to the LRT in Kato, Yamada and Fujikoshi (2010) under normality assumption when  $p < n - 1$ .

Srivastava and Reid (2012) recast this test problem to test of independence under normality. So it is different from the LRT,  $T_{n1}$  and  $T_{n2}$ .

## Example 3: Testing bandedness structure

Let  $\mathbf{A}_1 = \mathbf{I}_p$  and for  $2 \leq k \leq K$ ,  $\mathbf{A}_k$  to be a  $p \times p$  matrix with  $(i, j)$ -element being 1 if  $|i - j| = k - 1$  and 0 otherwise. Testing the  $(K - 1)$ -banded covariance matrix is to test

$$H_{30} : \Sigma = \theta_1 \mathbf{A}_1 + \cdots + \theta_K \mathbf{A}_K,$$

where  $\theta_k$ 's are unknown. By (2), we have

$$\begin{aligned}\hat{\theta}_1 &= p^{-1} \text{tr} \mathbf{S}_n \\ \hat{\theta}_k &= \frac{1}{2} (p - k + 1)^{-1} \text{tr} \mathbf{S}_n \mathbf{A}_k, \quad \text{for } 2 \leq k \leq K.\end{aligned}$$

Under Assumptions A and C,  $\hat{\theta}_k = \theta_k + O_p(n^{-1})$  if  $p/n \rightarrow y \in (0, \infty)$  when  $K$  is a finite.

Qiu and Chen (2012) proposed a U-statistic test for the banded structure, and is different from  $T_{n1}$  and  $T_{n2}$ .

## Example 4: Factor model

A factor model assumes  $\mathbf{x} = v_1 \mathbf{U}_1 + \cdots + v_{K-1} \mathbf{U}_{K-1} + \epsilon$ , where  $v_1, \cdots, v_{K-1}$  are random variables and  $\mathbf{U}_k, k = 1, \cdots, K-1$  are random vectors.

Suppose that  $v_1, \cdots, v_{K-1}, \mathbf{U}_1, \dots, \mathbf{U}_{K-1}$  and  $\epsilon$  are mutually independent. Suppose that  $\text{Cov}(\epsilon) = \theta_1 \mathbf{I}_p$  and conditioning on  $\mathbf{U}_k, k = 1, \cdots, K-1$ , the factor model has a covariance structure  $\Sigma = \theta_1 \mathbf{I}_p + \theta_2 \mathbf{U}_1 \mathbf{U}_1^T + \cdots + \theta_K \mathbf{U}_{K-1} \mathbf{U}_{K-1}^T$ , where  $\theta_{k+1} = \text{Var}(v_k)$ .

Let  $\mathbf{A}_1 = \mathbf{I}_p$ , and  $\mathbf{A}_{k+1} = \mathbf{U}_k \mathbf{U}_k^T$  for  $k = 1, \cdots, K-1$ . Thus, it is of interest to test

$$H_{40} : \Sigma = \theta_1 \mathbf{A}_1 + \cdots + \theta_K \mathbf{A}_K,$$

where  $\theta_k$ 's are unknown parameters.



## Example 4: Factor model

In practice one typically sets  $\mathbf{U}_k$  to be orthogonal to each other and have been standardized so that  $\mathbf{U}_s^T \mathbf{U}_t = p$  for  $s = t$  and 0 for  $s \neq t$ . The parameters  $\theta_k$  can be estimated by

$$\hat{\theta}_1 = (p - K + 1)^{-1} \left( \text{tr} \mathbf{S}_n - p^{-1} \sum_{k=1}^{K-1} \mathbf{U}_k^T \mathbf{S}_n \mathbf{U}_k \right)$$

$$\hat{\theta}_{k+1} = p^{-2} (\mathbf{U}_k^T \mathbf{S}_n \mathbf{U}_k - p \hat{\theta}_1)$$

for  $k = 1, \dots, K - 1$ . Thus, when  $K$  is finite and  $p/n$  has a finite positive limit, then  $\hat{\theta}_k = \theta_k + O_p(n^{-1})$  under  $H_{40}$ . Then testing  $H_{40}$  can be carried out by using the EL and QL tests.

## Example 5: Test a particular pattern (McDonald, 1974)

For even  $p$  which is finite and fixed, McDonald (1974) considered

$$H_{50} : \Sigma = \begin{pmatrix} \theta_1 \mathbf{I}_{p/2} + \theta_2 \mathbf{1}_{p/2} \mathbf{1}_{p/2}^T & \theta_3 \mathbf{I}_{p/2} \\ \theta_3 \mathbf{I}_{p/2} & \theta_1 \mathbf{I}_{p/2} + \theta_2 \mathbf{1}_{p/2} \mathbf{1}_{p/2}^T \end{pmatrix}.$$

Let  $\mathbf{A}_1 = \mathbf{I}_p$ ,

$$\mathbf{A}_2 = \begin{pmatrix} \mathbf{1}_{p/2} \mathbf{1}_{p/2}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{p/2} \mathbf{1}_{p/2}^T \end{pmatrix} \quad \text{and} \quad \mathbf{A}_3 = \begin{pmatrix} \mathbf{0} & \mathbf{I}_{p/2} \\ \mathbf{I}_{p/2} & \mathbf{0} \end{pmatrix}.$$

Then  $H_{50}$  can be written as  $H_{50} : \Sigma = \theta_1 \mathbf{A}_1 + \theta_2 \mathbf{A}_2 + \theta_3 \mathbf{A}_3$ . Thus, the proposed EL and QL test procedure can be used to test  $H_{50}$  with high-dimensional data.

# Some results on random matrix

To derive the limiting distribution of  $T_{n1}$  and  $T_{n2}$ , we need the limiting distributions of  $\log(|\mathbf{S}_n \hat{\Sigma}_0^{-1}|)$ ,  $\text{tr}(\mathbf{S}_n \hat{\Sigma}_0^{-1})$  and  $\text{tr}\{(\mathbf{S}_n \hat{\Sigma}_0^{-1})^2\}$ .

All these quantities are functional of eigenvalues of  $\hat{\Sigma}_0^{-1/2} \mathbf{S}_n \hat{\Sigma}_0^{-1/2}$ . This motivates us to study the limiting distribution of  $\Sigma^{-1/2} \mathbf{S}_n \Sigma^{-1/2} \triangleq \mathbf{F}$ , the sample covariance matrix of  $\mathbf{w}_1, \dots, \mathbf{w}_n$ .

Zheng, Chen, Cui and Li (2017) derived the limiting distribution of  $(\text{tr} \mathbf{F}, \sum_{j=1}^q \log\{\lambda_j(\mathbf{F})\})$ .

## Theorem 1.

Suppose that  $\mathbf{C}_0$  is a  $p \times p$  symmetric constant matrix, and  $p^{-1}\text{tr}\mathbf{C}_0$  has finite limit. Under Assumptions A and B, it follows that

$$p^{-1}\text{tr}\mathbf{F}\mathbf{C}_0 - p^{-1}\text{tr}\mathbf{C}_0 = o_p(1), \quad (7)$$

$$p^{-1}\text{tr}\mathbf{F}^2\mathbf{C}_0 - (1 + y_{n-1})p^{-1}\text{tr}\mathbf{C}_0 = o_p(1). \quad (8)$$

## Theorem 2.

Suppose that  $\mathbf{C}_1$  and  $\mathbf{C}_2$  are symmetric constant matrices, and their ESDs weakly converge as  $p \rightarrow \infty$ . Assume that  $\Sigma_n^{(1)} \rightarrow \Sigma^{(1)}$  exists. Under Assumptions A and B, it follows that

$$\begin{pmatrix} \text{tr} \mathbf{F} \mathbf{C}_1 \mathbf{F} \mathbf{C}_1 \\ \text{tr} \mathbf{F} \mathbf{C}_2 \end{pmatrix} - \boldsymbol{\mu}_n^{(1)} \rightarrow N(\mathbf{0}, \Sigma^{(1)})$$

in distribution as  $n \rightarrow \infty$ , where  $\boldsymbol{\mu}_n^{(1)}$ ,  $\Sigma_n^{(1)}$  and  $\Sigma^{(1)}$  are given in Zheng, Chen, Cui and Li (2017).

## Theorem 3.

Under  $H_0$  and Assumptions A and B, it follows that

(a) For  $p < n - 1$ ,

$$\frac{T_{n1} + (p - n + 1) \log(1 - y_{n-1}) - 2p + \alpha_1(y_{n-1})}{\sqrt{\alpha_2(y_{n-1})}} \rightarrow N(0, 1), \quad (9)$$

where  $\alpha_1(y) = 0.5 \log(1 - y) - 0.5(\kappa - 3)y$  and  
 $\alpha_2(y) = -2y - 2 \log(1 - y)$ ;

# Theorem 3

(b) For  $p \geq (n-1)$ ,

$$\frac{T_{n1} - p[y_{n-1}^{-1} - \alpha_3(y_{n-1})] - y_{n-1}^{-1}m_1(y_{n-1}) + m_2(y_{n-1})}{\sqrt{y^{-2}v_{11}(y_{n-1}) + v_{22}(y_{n-1}) - 2y_{n-1}^{-1}v_{12}(y_{n-1})}} \rightarrow N(0, 1) \quad (10)$$

where  $\alpha_3(y)$ ,  $m_1(y)$ ,  $m_2(y)$ ,  $v_{11}(y)$ ,  $v_{22}(y)$ ,  $v_{12}(y)$  are given in the paper.

(c)  $\frac{1}{2}[T_{n2} - py_{n-1} - (\kappa - 2)y]/\sigma \rightarrow N(0, 1)$ , where

$$\sigma^2 = y_{n-1}^2 + 2y_{n-1}^3 p^{-1} \text{tr}(\Sigma_0 \mathbf{B})^2 + (\kappa - 3)y_{n-1}^3 p^{-1} \sum_{i=1}^p (\mathbf{e}_i^T \Gamma^T \mathbf{B} \Gamma \mathbf{e}_i)^2 - (\kappa - 1)y_{n-1}^3$$

with  $\mathbf{B} = \sum_{k=1}^K d_k \mathbf{A}_k$ ,  $\mathbf{d} = (d_1, \dots, d_K)^T = \mathbf{C}^{-1} \mathbf{c}$  and  $\mathbf{c}$  being a  $K$ -dimensional column vector with  $k$ -th element being  $\text{tr} \mathbf{A}_k \Sigma_0^{-1}$ .

The limiting null distribution can be used to construct the rejection regions for  $T_{n1}$  and  $T_{n2}$ .

We next study the asymptotic power of these two tests. Define

$$\begin{pmatrix} \theta_1^* \\ \theta_2^* \\ \dots \\ \theta_K^* \end{pmatrix} = \begin{pmatrix} \text{tr} \mathbf{A}_1^2 & \text{tr} \mathbf{A}_2 \mathbf{A}_1 & \dots & \text{tr} \mathbf{A}_K \mathbf{A}_1 \\ \text{tr} \mathbf{A}_1 \mathbf{A}_2 & \text{tr} \mathbf{A}_2^2 & \dots & \text{tr} \mathbf{A}_K \mathbf{A}_2 \\ \dots & \dots & \dots & \dots \\ \text{tr} \mathbf{A}_1 \mathbf{A}_K & \text{tr} \mathbf{A}_2 \mathbf{A}_K & \dots & \text{tr} \mathbf{A}_K^2 \end{pmatrix}^{-1} \begin{pmatrix} \text{tr} \Sigma \mathbf{A}_1 \\ \text{tr} \Sigma \mathbf{A}_2 \\ \dots \\ \text{tr} \Sigma \mathbf{A}_K \end{pmatrix}$$

and  $\Sigma_{0*} = \theta_1^* \mathbf{A}_1 + \dots + \theta_K^* \mathbf{A}_K$ . Under  $H_0 : \Sigma = \Sigma_0$ , it follows that  $\widehat{\Sigma}_{0*} = \Sigma_0$ . Under Assumptions A and B, it can be shown that  $\widehat{\theta}_k = \theta_k^* + O_p(n^{-1})$ ,  $k = 1, \dots, K$ .

Let  $G_p(t) = p^{-1} \sum_{j=1}^p I(\lambda_j \leq t)$  be the ESD of  $\Sigma_{0*}^{-1/2} \Sigma \Sigma_{0*}^{-1/2}$ .

If  $G_p(t) \rightarrow G(t)$ , in which  $G(t)$  is assumed not to be degenerated to a single point distribution. That is,  $\Sigma \neq \tau \Sigma_{0*}$  for some constant  $\tau > 0$ . Under this condition, we have the limiting distributions of  $T_{n1}$  and  $T_{n2}$ .



## Theorem 4

Under Assumptions A and B, and under  $H_1 : G_p(t) \rightarrow G(t)$ , a non-degenerated distribution, it follows that

$$\frac{T_{n1} - pF_2^{y_{n-1}, G} - \mu_2^{(1)}}{\sigma_{2n}^{(1)}} \rightarrow N(0, 1), \quad \text{for } p < n - 1$$

$$\frac{T_{n1} - pF_3^{y_{n-1}, G} - \mu_3^{(1)}}{\sigma_{3n}^{(1)}} \rightarrow N(0, 1), \quad \text{for } p \geq n - 1$$

$$\frac{T_{n2} - \mu_1^{(1)}}{\sigma_{1n}^{(1)}} \rightarrow N(0, 1),$$

where  $\mu_j^{(1)}, j = 1, 2, 3$ ,  $F_j^{y_{n-1}, G}, j = 2, 3$  and  $\sigma_{jn}^{(1)}, j = 1, 2, 3$  are given in the proof of Theorem .

# Power function of $T_{n2}$ and unbiased test

For a level  $\alpha$ , the power function of  $T_{n2}$  is

$$1 - \Phi((\mu_0 - \mu_1^{(1)})/\sigma_{1n}^{(1)} - 2q_{\alpha/2}\sigma/\sigma_{1n}^{(1)}) + \Phi((\mu_0 - \mu_1^{(1)})/\sigma_{1n}^{(1)} + 2q_{\alpha/2}\sigma/\sigma_{1n}^{(1)}),$$

where  $q_{\alpha/2}$  is the  $\alpha/2$  quantile of  $N(0, 1)$  and  $\mu_0 = py_{n-1} + (\kappa - 2)y$ .

## Theorem 5

Suppose that Assumptions  $A$  and  $B$  are satisfied and the limit of  $\sigma_{1n}^{(1)}$  exists. If  $\Gamma^T \Sigma_{0*}^{-1} \Gamma = \mathbf{I}_p + \mathbf{A}$  whose ESD weakly converges and  $\text{tr} \mathbf{A}^2 > \delta > 0$ , then we have

$$\beta_{T_{n2}} > \alpha$$

when  $n$  is large enough and  $\delta$  is any given small constant.

Thus,  $T_{n2}$  is an asymptotically unbiased test. Furthermore, if  $p^{-1} \text{tr} \mathbf{A} \rightarrow c_1 \neq 0$ , then  $\beta_{T_{n2}} \rightarrow 1$  when  $n \rightarrow \infty$ .

**Estimation of  $\kappa$ :** Under Assumption A, we construct an consistent estimate for  $\kappa$  by using

$$E\{\mathbf{w}^T \Sigma \mathbf{w} - \text{tr}(\Sigma)\}^2 = 2\text{tr}(\Sigma^2) + (\kappa - 3) \sum_{j=1}^p \sigma_{jj}^2.$$

Our simulation shows that we may get more stable estimate for  $\kappa$  by using the linear structure under  $H_0$ .

# Numerical Examples

In our simulation,  $\mathbf{x} = \Sigma^{1/2}\mathbf{w}$ , where  $\Sigma$  will be set according to the hypothesis to be tested.

In order to examine the performances of the proposed tests under different distributions, we consider the elements of  $\mathbf{w}$  being independent and identically distributed as (a)  $N(0, 1)$  or (b)  $\text{Gamma}(4, 2) - 2$ . The two distributions are both normalized so that their means 0 and variances 1.

For each setting, we conduct 1000 Monte Carlo simulations. The Monte Carlo simulation error rate is  $1.96\sqrt{0.05 \times 0.95/1000} \approx 0.0135$  at level 0.05.

## Test of compound symmetry $H_{20} : \Sigma = \theta_1 \mathbf{I}_p + \theta_2 \mathbf{1}_p \mathbf{1}_p^T$

We set  $\Sigma = \theta_1 \mathbf{I}_p + \theta_2 \mathbf{1}_p \mathbf{1}_p^T + \theta_3 \mathbf{u}_p \mathbf{u}_p^T$ , where  $\mathbf{u}_p$  is a  $p$ -dimensional random vector following uniform distribution over  $[-1, 1]$ .

The third term is to examine the empirical powers when  $\theta_3 \neq 0$ . In our simulation, we set  $(\theta_1, \theta_2) = (6, 1)$  and  $\theta_3 = 0, 0.5, 1$ , respectively.

We set  $\theta_3 = 0$  to examine Type I error rate and  $\theta_3 = 0.5, 1$  to study the empirical power of the proposed tests. The sample size is set as  $n = 100, 200$  and the dimension is taken to be  $p = 50, 100, 500, 1000$ .

Compare the performances of proposed testing procedures and the test proposed in Srivastava and Reid (2012)

Table: Empirical power for  $H_{20}$  (in percentage) with  $n = 100$

$\theta_3$	Test	$w_j \sim N(0, 1)$				$w_j \sim \text{Gamma}(4, 2) - 2$			
		$p = 50$	100	500	1000	50	100	500	1000
0	QL	5.23	5.40	5.12	5.19	6.48	5.99	5.64	5.41
	EL	5.32	6.51	5.12	5.18	5.77	6.35	5.46	5.54
	SR	4.90	5.01	4.91	4.98	9.60	8.96	8.15	8.04
0.5	QL	40.25	80.58	100.00	100.00	41.01	80.70	100.00	100.00
	EL	13.42	11.38	99.78	100.00	13.74	11.42	99.74	100.00
	SR	24.46	59.04	99.99	100.00	41.22	73.86	100.00	100.00
1.0	QL	95.88	99.97	100.00	100.00	95.98	99.97	100.00	100.00
	EL	53.53	29.97	100.00	100.00	53.71	30.18	100.00	100.00
	SR	87.90	99.67	100.00	100.00	93.61	99.87	100.00	100.00

Table: Empirical power for  $H_{20}$  (in percentage) with  $n = 200$

$\theta_3$	Test	$w_j \sim N(0, 1)$				$w_j \sim \text{Gamma}(4, 2) - 2$			
		$p = 50$	100	500	1000	50	100	500	1000
0	QL	5.22	5.14	5.12	5.19	6.32	5.78	5.31	5.34
	EL	5.18	5.12	5.05	5.13	5.94	5.42	5.23	5.31
	SR	4.98	4.93	4.93	5.03	9.95	9.23	8.44	8.43
0.5	QL	79.86	99.32	100.00	100.00	78.56	99.28	100.00	100.00
	EL	42.00	58.62	100.00	100.00	41.22	58.62	100.00	100.00
	SR	61.74	95.79	100.00	100.00	75.78	98.24	100.00	100.00
1.0	QL	99.98	100.00	100.00	100.00	99.96	100.00	100.00	100.00
	EL	97.23	99.53	100.00	100.00	96.86	99.55	100.00	100.00
	SR	99.81	100.00	100.00	100.00	99.91	100.00	100.00	100.00

# Test covariance matrix structure in $H_{30}$

We construct a banded matrix defined in Example 3 with width of band  $K = 3$ . Therefore, the null hypothesis  $H_{30}$  has the linear decomposition  $\Sigma = \theta_1 \mathbf{I}_p + \theta_2 \mathbf{A}_2 + \theta_3 \mathbf{A}_3 + \theta_4 \mathbf{u}_p \mathbf{u}_p^T$ , where  $\mathbf{A}_2$  and  $\mathbf{A}_3$  are defined in Example 3 and  $\mathbf{u}_p$  is generated by the same way in the last example.

We take  $(\theta_1, \theta_2, \theta_3) = (6, 1, 0.5)$  and  $\theta_4 = 0$  to examine Type I error rates and take  $\theta_4 = 0.5, 1$  to examine powers. In the simulation studies, we still set sample size  $n = 100, 200$  and dimension  $p = 50, 100, 500, 1000$ .

In this example, we compare the test proposed in Qiu and Chen (2012) for the banded covariance matrix with our proposed tests, and referred their test as “QC” test hereinafter.



Table: Simulation Results for  $H_{30}$  (in percentage)

$\theta_4$	$n$	Test	$w_j \sim N(0, 1)$				$w_j \sim \text{Gamma}(4, 2) - 2$			
			$p = 50$	100	500	1000	50	100	500	1000
0	100	QL	5.30	5.19	5.29	5.34	6.61	6.19	5.83	5.89
		EL	5.31	6.34	5.20	5.36	5.83	6.32	5.53	5.81
		QC	5.00	5.50	5.50	5.90	5.00	5.80	6.10	5.20
0.5	100	QL	46.01	84.91	100.00	100.00	45.13	83.62	100.00	100.00
		EL	15.20	12.23	99.90	100.00	15.35	12.13	99.90	100.00
		QC	17.00	70.00	100.00	100.00	24.00	85.30	100.00	100.00
1.0	100	QL	97.49	99.98	100.00	100.00	96.96	99.98	100.00	100.00
		EL	60.45	33.68	100.00	100.00	59.80	33.71	100.00	100.00
		QC	83.00	100.00	100.00	100.00	78.00	99.90	100.00	100.00

Table: Simulation Results for  $H_{30}$  (in percentage)

			$w_j \sim N(0, 1)$				$w_j \sim \text{Gamma}(4, 2) - 2$			
$\theta_4$	$n$	Test	$p = 50$	100	500	1000	50	100	500	1000
0	200	QL	5.25	5.19	5.16	5.10	6.44	5.84	5.56	5.55
		EL	5.24	5.15	5.04	4.94	6.01	5.46	5.25	5.26
		QC	6.00	5.40	5.50	4.50	4.00	4.50	5.80	5.10
0.5	200	QL	85.30	99.66	100.00	100.00	84.09	99.64	100.00	100.00
		EL	48.46	65.39	100.00	100.00	47.62	65.20	100.00	100.00
		QC	49.00	100.00	100.00	100.00	49.00	99.70	100.00	100.00
1.0	200	QL	99.99	100.00	100.00	100.00	99.98	100.00	100.00	100.00
		EL	98.27	99.80	100.00	100.00	98.00	99.76	100.00	100.00
		QC	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

# Test of factor model $H_{40}$

In this example, we examine Type I error rate and empirical power of the proposed tests for factor model.

We first generate several mutually orthogonal factors. Suppose that  $\mathbf{u}_k^*$ ,  $k = 1, \dots, K$  are IID random vectors following  $N_p(\mathbf{0}, \mathbf{I}_p)$ . Let  $\mathbf{u}_1 = \mathbf{u}_1^*$  and  $\mathbf{u}_k = (\mathbf{I}_p - \mathbf{P}_k)\mathbf{u}_k^*$ , where  $\mathbf{P}_k$  is the projection matrix on  $\mathbf{u}_1, \dots, \mathbf{u}_{k-1}$  for  $k = 2, \dots, K$ .

Given  $\mathbf{u}_k$ 's, we have the covariance matrix structure

$$\Sigma = \theta_0 \mathbf{I}_p + \sum_{k=1}^K \theta_k \mathbf{u}_k \mathbf{u}_k^T$$

for the factor model.

We set  $K = 4$  and the coefficient vector  $(\theta_0, \theta_1, \theta_2, \theta_3)^T = (4, 3, 2, 1)^T$ .

Similarly,  $\theta_4 = 0$  is for Type I error rates and  $\theta_4 = 0.5, 1$  is for powers.

Table: Empirical powers for  $H_{40}$  (in percentage)

$\theta_5$	$n$	Test	$w_j \sim N(0, 1)$				$w_j \sim \text{Gamma}(4, 2) - 2$			
			$p = 50$	100	500	1000	50	100	500	1000
0	100	QL	5.46	5.46	6.02	6.27	6.89	6.53	6.58	6.94
		EL	5.40	6.40	5.79	6.20	6.03	6.42	6.30	6.52
0.5	100	QL	99.99	100.00	100.00	100.00	99.96	100.00	100.00	100.00
		EL	97.58	86.91	100.00	100.00	97.38	87.25	100.00	100.00
1.0	100	QL	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		EL	100.00	99.93	100.00	100.00	99.99	99.93	100.00	100.00
0	200	QL	5.32	5.28	5.42	5.52	6.57	6.05	5.89	6.05
		EL	5.30	5.22	5.33	5.65	6.15	5.61	5.60	5.65
0.5	200	QL	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		EL	99.99	100.00	100.00	100.00	100.00	100.00	100.00	100.00
1.0	200	QL	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		EL	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

# Test for special pattern $H_{50}$

To investigate the performance of QL and EL tests for  $H_{50}$ , represent  $\Sigma$  as a linear combination

$$\Sigma = \sum_{k=1}^4 \theta_k \mathbf{A}_k,$$

where  $\mathbf{A}_1$ ,  $\mathbf{A}_2$  and  $\mathbf{A}_3$  are defined in Example 5 and  $\mathbf{A}_4 = \mathbf{u}_p \mathbf{u}_p^T$  with  $\mathbf{u}_p \sim N_p(\mathbf{0}, \mathbf{I}_p)$ . We set the first three coefficients  $(\theta_1, \theta_2, \theta_3) = (6, 0.5, 0.1)$  and  $\theta_4 = 0, 0.5$  and  $1$  for examining Type I error rates and powers, respectively.

Table: Empirical power for  $H_{50}$  (in percentage)

			$w_j \sim N(0, 1)$				$w_j \sim \text{Gamma}(4, 2) - 2$			
$\theta_5$	$n$	Test	$p = 50$	100	500	1000	50	100	500	1000
0	100	QL	5.28	5.19	5.23	5.41	6.59	6.15	5.84	6.16
		EL	5.27	6.33	5.17	5.48	5.85	6.35	5.61	5.73
0.5	100	QL	99.64	100.00	100.00	100.00	99.61	100.00	100.00	100.00
		EL	85.90	57.86	100.00	100.00	85.89	59.17	100.00	100.00
1.0	100	QL	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		EL	99.87	97.67	100.00	100.00	99.81	97.86	100.00	100.00
0	200	QL	5.25	5.11	5.08	5.18	6.40	5.84	5.58	5.62
		EL	5.25	5.15	5.06	5.05	6.01	5.44	5.29	5.30
0.5	200	QL	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		EL	99.75	99.99	100.00	100.00	99.72	99.99	100.00	100.00
1.0	200	QL	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		EL	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00