Accepted Manuscript

Simultaneous testing of mean vector and covariance matrix for high-dimensional data

Zhongying Liu, Baisen Liu, Shurong Zheng, Ning-Zhong Shi

PII: S0378-3758(17)30055-1
DOI: http://dx.doi.org/10.1016/j.jspi.2017.03.009
Reference: JSPI 5556

To appear in: *Journal of Statistical Planning and Inference*

Received date: 2 March 2016
Revised date: 22 November 2016
Accepted date: 22 March 2017

Simultaneous Testing of Mean Vector and Covariance Matrix for High-dimensional Data

Zhongying Liu[a], Baisen Liu[b], Shurong Zheng[a,*], Ning-Zhong Shi[a]

[a] *KLAS and School of Mathematics and Statistics, Northeast Normal University, Changchun, 130024, China*

[b] *School of Statistics, Dongbei University of Finance and Economics, Dalian, 116025, China*

Abstract

A new method is proposed to simultaneously test mean vector and covariance matrix for high-dimensional data. It allows for the case of large dimension p and small sample size n, and it is also robust against non-Gaussian data. Besides, the asymptotic null distribution is derived and the asymptotic theoretical power function is explicitly achieved. The local power of the new method is studied and the proposed test is proved to be asymptotically unbiased. Finally, the efficiency of the new method is assessed by numerical simulations.

Keywords: High dimensional covariance matrix, High dimensional mean vector, Martingale difference sequence, Simultaneous test, Theoretical power function

*Corresponding email: zhengsr@nenu.edu.cn (S.R.Zheng)

1. Introduction

Mean vector and covariance matrix are fundamental in multivariate statistical analysis. High dimensional mean vector and high dimensional covariance matrix are widely used in many fields, for example, in gene expression data and fMRI studies etc. The statistical inference for mean vector and covariance matrix is still an important task in high dimensional data analysis.

Let $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ be the i.i.d. samples from a p-dimensional population with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Define the corresponding sample mean and sample covariance matrix as

$$\bar{\mathbf{x}} = n^{-1}\sum_{i=1}^{n}\mathbf{x}_i, \quad \mathbf{S}_n = n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T.$$

Simultaneous testing of mean vector and covariance matrix is an interesting problem in multivariate statistical analysis, that is, to test the null hypothesis

$$H_0 : \boldsymbol{\mu} = \mathbf{0}_p, \quad \boldsymbol{\Sigma} = \mathbf{I}_p, \tag{1.1}$$

where $\mathbf{0}_p$ is the $p-$dimensional zero vector and $\mathbf{I}_p$ is the $p \times p$ dimensional identity matrix. In the literature, some researches have been done on this issue for Gaussian population. For example, in the classical setting which assumes that the dimension p is fixed and the sample size n tends to infinity, Anderson (2003) suggested a classical likelihood ratio (LR) test

$$n^{-1}\mathrm{tr}\mathbf{B}_n - \log|\mathbf{B}_n| + \bar{\mathbf{x}}^T\bar{\mathbf{x}}, \tag{1.2}$$

and provided its asymptotic null distribution where $\mathbf{B}_n = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$ and $\mathrm{tr}\mathbf{B}_n$ denotes the trace of the matrix $\mathbf{B}_n$. Sugiura and Nagao (1968) and Das Gupta (1969) studied the unbiasedness of the LR test. Nagarsenker and Pillai (1974) and Pham-Gia and Turkkan (2009) concerned an exact null distribution of the LR test. Chou et al. (2003) discussed an approximation approach for the power function of the LR test.

Sample mean vector and sample covariance matrix are the natural estimators of the population mean vector and covariance matrix; consequently, a natural test statistic will be a function based on the sample mean vector and the sample covariance matrix. In high dimensional setting that the dimension of data and the sample size increase proportionally, Marčenko and Pastur (1967) found that the empirical spectral distribution (*empirical distribution of eigenvalues*) of the sample covariance matrix does not converge to the true distribution of eigenvalues of the population covariance matrix. Yin, Bai and Krishnaiah (1988) proved that the maximum sample eigenvalue of high dimensional sample covariance matrix does not converge to the maximum population eigenvalue. Even for testing $H_{01} : \mathbf{\Sigma} = \mathbf{I}_p$, Bai et al. (2009) showed that the classical LR test became invalid for high-dimensional data by some numerical studies. As a remedy, a corrected LR test is developed in Bai et al. (2009). Chen, Zhang and Zhong (2010) further proposed a new test for H_{01} based on U-statistics. Bai and Hewa (1996) showed that the classical Hotelling test is invalid when studying the equality of two high dimensional mean vectors. Therefore, assuming that $(n, p) \to \infty$ and $p < n$, Jiang and Yang (2013) and Jiang and Qi (2015) studied the simultaneous testing of mean vector and covariance matrix and give some central limit theorem (CLT) results of the LR test. However, their works still require that the data are sampled from a multivariate Gaussian population.

This paper will target at testing the mean vector and covariance matrix simultaneously for high dimensional data but relaxing the assumption of distributions and only requiring the population fourth moment exists. Moreover, the proposed test is also applicable when $p > n$. Three-fold contributions of this paper are: (1) proposing a test statistic for $H_0 : \boldsymbol{\mu} = \mathbf{0}_p, \boldsymbol{\Sigma} = \mathbf{I}_p$; (2) establishing the central limit theorem of the proposed test statistic under the asymptotic regime that the dimension of data and the sample size tend to infinity proportionally; (3) providing an explicit form of the theoretical power function and proving that this theoretical power function is unbiased.

The rest of this paper is arranged as follows: Section 2 develops a statistic for testing the high dimensional mean vector and high dimensional covariance matrix simultaneously. Section 3 performs some simulation studies to evaluate the efficiency of test proposed. Some conclusions are provided in Section 4 and the proofs needed are given in the Appendix.

2. Main results

Before making statistical inference, some often-used assumptions in random matrix theories are imposed on the population.

Assumption [A]. The random vectors $\mathbf{x}_i$'s satisfy the independent component structure, $\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{w}_i$, where $\mathbf{w}_i = (w_{1i}, \ldots, w_{pi})^T$ with w_{ji}'s being i.i.d. and satisfying

$$\mathrm{E}w_{ji} = 0, \quad \mathrm{E}w_{ji}^2 = 1, \quad \text{and} \quad \beta_w = \mathrm{E}w_{ji}^4 - 3.$$

Assumption [B]. The spectral norm of $\boldsymbol{\Sigma}$ is bounded and the empirical spectral distribution (*ESD*) $F_n(x)$ of $\boldsymbol{\Sigma}$ converges in distribution to a limit distribution $\widetilde{F}(\cdot)$, where $F_n(x) = p^{-1}\sum_{j=1}^{p}\delta_{\{\lambda_j \le x\}}$ with the eigenvalues $\{\lambda_j\}_{j=1}^{p}$ of $\boldsymbol{\Sigma}$. The asymptotic regime is $p/n \to y \in (0, \infty)$.

Assumption [A] requires that the population only satisfies the independent component structure and has the finite fourth moment. Assumption [B] then gives the asymptotic regime that the dimension of data tends to infinity with the sample size proportionally. Moreover, $\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}$ and the ESD of $\boldsymbol{\Sigma}$ are required to converge.

For testing the null hypothesis (1.1), Jiang and Yang (2013) established the asymptotic distribution of the LR test (1.2) when $p/n \to y \in (0, 1]$ in the normal population. However, the LR test (1.2) is valid only for $p < n$. But in practice, we often encounter the cases of $p \ge n$. In the case of $p > n$, the LR test (1.2) does not exist since that a rank-deficient matrix $\mathbf{A}$ has a zero determinant. In fact, the statistic $n^{-1}\mathrm{tr}\mathbf{B}_n - \log|\mathbf{B}_n| + \bar{\mathbf{x}}^T\bar{\mathbf{x}}$ has two parts: one part is the term $\bar{\mathbf{x}}^T\bar{\mathbf{x}}$ which is the mean loss for mean estimation; another part is the term $n^{-1}\mathrm{tr}\mathbf{B}_n - \log|\mathbf{B}_n|$ which is the entropy loss (EL) used for covariance matrix estimation. Motivated by the LR test, we propose the following statistic

$$T_n = \bar{\mathbf{x}}^T \bar{\mathbf{x}} + \mathrm{tr}(\mathbf{S}_n - \mathbf{I}_p)^2, \tag{2.1}$$

where $\mathrm{tr}(\mathbf{S}_n - \mathbf{I}_p)^2$ is the quadratic loss for covariance matrix estimation. From now on, in this article, the notations $\xrightarrow{D}$ and $\xrightarrow{P}$ mean that "converge in distribution" and "converge in probability", respectively. The following theorem will give the asymptotic null distribution of T_n under H_0.

Theorem 2.1. *Suppose that Assumption [A]-[B] hold. Under the null hypothesis H_0, when both n and p go to infinity with $p/n \to y \in (0, \infty)$, we have*

$$\sigma_0^{-1}(T_n - \mu_0) \xrightarrow{D} N(0, 1),$$

where

$$\mu_0 = p^2/n + p(\beta_w + 2)/n \quad \text{and} \quad \sigma_0^2 = 4y^3(2 + \beta_w) + 4y^2.$$

Under H_0, we have $\mathbf{x}_i = \mathbf{w}_i$. Thus the parameter β_w can be estimated by the method of moments estimation, that is,

$$\hat{\beta}_w = (np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} x_{ji}^4 - 3.$$

Let α be the true test size and $q_{1-\alpha}$ be the $100(1-\alpha)\%$ quantile of $N(0, 1)$. Then the rejection region for H_0 is given by

$$\{\mathbf{x}_1, \ldots, \mathbf{x}_n : T_n > \mu_0 + \sigma_0 q_{1-\alpha/2} \text{ or } T_n < \mu_0 + \sigma_0 q_{\alpha/2}\}.$$

The following theorem will give the asymptotic distribution of T_n under the alternative hypothesis. Essentially, Theorem 2.1 is the special case of the following Theorem 2.2 when $\boldsymbol{\mu} = \mathbf{0}_p$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$.

Theorem 2.2. *Suppose that the Assumption [A]-[B] hold. When both n and p go to infinity with $p/n \to y \in (0, \infty)$, we have*

$$\sigma_A^{-1}(T_n - \mu_A) \xrightarrow{D} N(0,1),$$

with

$$\begin{aligned}
\mu_A &= n^{-1}\mathrm{tr}\boldsymbol{\Sigma} + \mathrm{tr}(\boldsymbol{\Sigma} - \mathbf{I}_p)^2 - \boldsymbol{\mu}^T\boldsymbol{\mu} \\
&\quad + n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2 + \beta_w n^{-1}\sum_{k=1}^{p}(\mathbf{e}_k^T\boldsymbol{\Sigma}\mathbf{e}_k)^2 + n^{-1}(\mathrm{tr}\boldsymbol{\Sigma})^2 \\
&\quad + 2(n+1)n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu} + 4n^{-1}\mathrm{E}(\mathbf{x}_1 - \boldsymbol{\mu})^T(\mathbf{x}_1 - \boldsymbol{\mu})(\mathbf{x}_1 - \boldsymbol{\mu})^T\boldsymbol{\mu} \\
&\quad + 2\boldsymbol{\mu}^T\boldsymbol{\mu}(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) + (\boldsymbol{\mu}^T\boldsymbol{\mu})^2,
\end{aligned}$$

and $\sigma_A^2 \le C_1 + C_2(\boldsymbol{\mu}^T\boldsymbol{\mu})^2 + C_3\boldsymbol{\mu}^T\boldsymbol{\mu}$ where C_1, C_2, C_3 are positive constants and $\mathbf{e}_k$ is the kth column of the identity matrix. Moreover, the theoretical power function is given by

$$\beta_{T_n} = 1 - \Phi\left(\frac{\mu_0 - \mu_A + \sigma_0 q_{1-\alpha/2}}{\sigma_A}\right) + \Phi\left(\frac{\mu_0 - \mu_A + \sigma_0 q_{\alpha/2}}{\sigma_A}\right) \qquad (2.2)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of $N(0,1)$ and $q_{\alpha/2}$ is the $\alpha/2$ quantile of $N(0,1)$.

Remark 1. *In computing the theoretical power function of (2.2), $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known. Let $\mathbf{w}_i = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu})$. Then the parameter β_w can be estimated by the method of moments estimation, that is,*

$$\hat{\beta}_w = (np)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{p} w_{ji}^4 - 3.$$

Theorem 2.3. *Under Assumptions [A]-[B], if $\boldsymbol{\mu} = \epsilon\mathbf{1}_p, \boldsymbol{\Sigma} = \mathbf{I}_p$ or $\boldsymbol{\mu} = \mathbf{0}_p, \boldsymbol{\Sigma} - \mathbf{I}_p = \epsilon\mathbf{I}_p$ where ϵ is a non-negative constant and $\mathbf{1}_p$ is a p-dimensional column vector with all elements equalling to 1, we have*

$$\beta_{T_n} \to 1.$$

Remark 2. *In fact, only if ϵ is a very small positive constant, we have $\beta_{T_n} > \alpha$ when n and p are large enough where α is the true test size . That is, the test T_n is an asymptotic unbiased test whose asymptotic power function is greater than the test size even although the alternative hypothesis is very close to the null hypothesis.*

Remark 3. *If we want to test*

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0, \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0,$$

where $\boldsymbol{\mu}_0$ *and* $\boldsymbol{\Sigma}_0$ *are specified known vector and positive definite matrix, respectively, then we transform the samples* $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ *as* $\{\boldsymbol{\Sigma}_0^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu}_0), i = 1, \ldots, n\}$. *Then under* H_0, *we have* $\mathrm{E}\boldsymbol{\Sigma}_0^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu}_0) = \mathbf{0}_p$ *and* $\mathrm{Cov}[\boldsymbol{\Sigma}_0^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu}_0)] = \mathbf{I}_p$. *Consequently, testing*

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0, \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0,$$

is equivalent to testing $H_0 : \tilde{\boldsymbol{\mu}} = \mathbf{0}_p, \widetilde{\boldsymbol{\Sigma}} = \mathbf{I}_p$ *where* $\tilde{\boldsymbol{\mu}}$ *and* $\widetilde{\boldsymbol{\Sigma}}$ *are the population mean vector and covariance matrix of the samples* $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$, *respectively, where* $\mathbf{y}_i = \boldsymbol{\Sigma}_0^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu}_0)$.

3. Simulations

The random samples are $\{\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}(\mathbf{w}_i - m(F))/\sigma(F), i = 1, \ldots, n\}$ where the elements $\{w_{1i}, \ldots, w_{pi}\}$ of $\mathbf{w}_i$ are i.i.d. generated from the distribution $F(\cdot)$ with mean $m(F)$ and standard deviation $\sigma(F)$. We consider the following scenarios for the distribution $F(\cdot)$: (1) Gaussian distribution $N(0, 1)$; (2) Gamma distribution Gamma(4,2) with mean 2; and (3) χ^2 distribution with degrees of freedom 3. The sample size is taken as $n = 55, 105, 205, 305$ and the dimension is taken as $p = 50, 100, 200, 300, 500$ and $1,000$. The true test size is set as 5%; while the underlying population mean and covariance matrix of the random samples $\{\mathbf{x}_i\}_{i=1}^n$ are given as:

- Model I: $\boldsymbol{\mu} = \mathbf{0}_p$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$.
- Model II: $\boldsymbol{\mu} = (\epsilon \mathbf{1}_{p_1}, \mathbf{0}_{p-p_1})^T$ with $\epsilon = 0.3, 0.5$ and $p_1 = [p/5]$ where $[x]$ denotes the integer truncation of x, and $\boldsymbol{\Sigma} = \mathbf{I}_p$.

- Model III: $\boldsymbol{\mu} = (\epsilon \mathbf{1}_{p_1}, \mathbf{0}_{p-p_1})^T$ where $\epsilon = 0.1$, $p_1 = 0$ or $[p/2]$ and $\boldsymbol{\Sigma} = (\sigma_{ij})_{p\times p}$ where $\sigma_{ij} = 1$ for $i = j$, $\sigma_{ij} = 0.1$ for $0 < |i - j| \leq 3$, and $\sigma_{ij} = 0$ for $|i - j| > 3$.

- Model IV: $\boldsymbol{\mu} = (\epsilon \mathbf{1}_{p_1}, \mathbf{0}_{p-p_1})^T$ where $\epsilon = 0.1$, $p_1 = 0$ or $[p/2]$ and $\boldsymbol{\Sigma} = (1-\rho)\mathbf{I}_p + \rho \mathbf{J}_p$ with $\rho = 0.05$, where $\mathbf{J}_p$ denotes the $p \times p$ matrix with all elements being 1.

For each model, we run $10,000$ simulations to obtain the empirical test sizes and empirical powers of T_n. As comparisons, we also evaluate the CLRT results of Jiang and Yang (2013). Table 1 displays the empirical test sizes of T_n and the CLRT of Jiang and Yang (2013) under Model I for Gaussian data. The simulation result shows that, as n and p increase, the empirical test sizes of these two tests are similar and both are so close to the true test size $\alpha = 0.05$. Table 2-4 display the empirical powers of T_n and CLRT under Model II-IV for Gaussian data, from which we can see that T_n has much larger powers than the CLRT of Jiang and Yang (2013). Moreover, T_n performs very well for the case of $p > n$ while the CLRT is valid only for the case of $n > p+1$. Furthermore, Table 5 displays the empirical sizes of T_n for non-Gaussian data. Obviously, the empirical test sizes are also so close to the true test size $\alpha = 0.05$, which indicates that our proposed method is robust against non-normal data. Table 6-8 display the empirical powers of T_n for Gamma population and χ^2 population under the alternative hypothesis. We found there is an interesting phenomenon shown in Table 2-4 (which is also shown in Table 6-8): when fixed sample size n and increasing dimension p, the powers in Table 2 and Table 4 increase, while the powers in Table 3 decrease. We think that this discrepancy is because the covariance matrix $\mathbf{\Sigma}$ in Model III becomes more sparse with increasing p. On the contrary, $\mathbf{\Sigma}$ in Model II and Model IV is not sparse which results in the powers increasing when p increasing. We also compare the empirical powers and theoretical powers of T_n. From Figure 1-2, it can be seen that the empirical powers and the theoretical powers are very close.

Table 1: Empirical test sizes (in percentage) of T_n and CLRT (Jiang and Yang , 2013) under Model I with Gaussian data based on 10,000 replications.

n	Method	Dimension p					
		50	100	200	300	500	1,000
55	CLRT	5.28	-	-	-	-	-
	T_n	4.60	4.77	4.41	5.02	5.03	5.12
105	CLRT	5.44	5.41	-	-	-	-
	T_n	4.81	4.87	4.63	4.75	5.04	4.87
205	CLRT	5.62	5.25	5.35	-	-	-
	T_n	4.97	4.80	5.08	4.89	4.85	4.93
305	CLRT	5.70	5.36	5.40	5.34	-	-
	T_n	4.93	4.77	4.58	4.98	5.00	5.00

Table 2: Empirical test powers (in percentage) of T_n and CLRT (Jiang and Yang , 2013) under Model II with Gaussian data based on 10,000 replications.

ϵ	n	Method	Dimension p					
			50	100	200	300	500	1,000
0.3	55	CLRT	8.68	-	-	-	-	-
		T_n	19.2	30.5	50.5	66.8	86.2	98.8
	105	CLRT	33.8	13.9	-	-	-	-
		T_n	46.4	69.3	92.2	98.2	99.9	100
	205	CLRT	89.1	83.5	30.9	-	-	-
		T_n	89.3	98.9	99.9	100	100	100
	305	CLRT	99.8	99.6	97.1	54.3	-	-
		T_n	99.3	100	100	100	100	100
0.5	55	CLRT	28.0	-	-	-	-	-
		T_n	93.3	99.7	100	100	100	100
	105	CLRT	97.5	63.4	-	-	-	-
		T_n	100	100	100	100	100	100
	205	CLRT	100	100	98.2	-	-	-
		T_n	100	100	100	100	100	100
	305	CLRT	100	100	100	99.9	-	-
		T_n	100	100	100	100	100	100

Table 3: Empirical test powers (in percentage) of T_n and CLRT (Jiang and Yang , 2013) under Model III with Gaussian data based on 10,000 replications.

p_1	n	Method	Dimension p					
			50	100	200	300	500	1,000
0	55	CLRT	32.9	-	-	-	-	-
		T_n	54.4	41.3	27.8	21.4	15.4	10.9
	105	CLRT	98.9	72.6	-	-	-	-
		T_n	97.7	95.0	85.5	75.0	57.5	36.9
	205	CLRT	100	100	99.5	-	-	-
		T_n	100	100	100	99.9	99.9	98.0
	305	CLRT	100	100	100	100	-	-
		T_n	100	100	100	100	100	100
$[p/2]$	55	CLRT	37.6	-	-	-	-	-
		T_n	65.1	54.7	44.2	39.7	35.6	36.1
	105	CLRT	99.6	79.7	-	-	-	-
		T_n	99.0	98.7	96.2	93.2	88.9	85.4
	205	CLRT	100	100	99.9	-	-	-
		T_n	100	100	100	100	100	100
	305	CLRT	100	100	100	100	-	-
		T_n	100	100	100	100	100	100

Table 4: Empirical test powers (in percentage) of T_n and CLRT (Jiang and Yang , 2013) under Model IV with Gaussian data based on 10,000 replications.

p_1	n	Method	Dimension p					
			50	100	200	300	500	1,000
0	55	CLRT	11.8	-	-	-	-	-
		T_n	47.9	69.3	85.5	90.8	94.6	97.1
	105	CLRT	55.5	34.0	-	-	-	-
		T_n	88.9	98.8	99.9	100	99.9	100
	205	CLRT	98.2	99.7	86.9	-	-	-
		T_n	99.9	100	100	100	100	100
	305	CLRT	100	100	100	99.6	-	-
		T_n	100	100	100	100	100	100
$[p/2]$	55	CLRT	14.6	-	-	-	-	-
		T_n	57.4	77.9	91.4	95.0	97.4	98.7
	105	CLRT	68.7	41.9	-	-	-	-
		T_n	94.3	99.5	100	100	100	100
	205	CLRT	99.6	99.9	93.4	-	-	-
		T_n	100	100	100	100	100	100
	305	CLRT	100	100	100	99.94	-	-
		T_n	100	100	100	100	100	100

Table 5: Empirical test sizes (in percentage) of T_n under Model I with non-Gaussian data based on 10,000 replications.

n	Dimension p											
	Gamma(4,2)						$\chi^2(3)$					
	50	100	200	300	500	1,000	50	100	200	300	500	1,000
55	5.22	5.05	5.01	5.04	4.91	4.73	6.05	5.24	5.23	5.47	5.57	4.82
105	5.00	4.92	4.73	5.02	4.91	5.15	5.97	5.15	5.32	5.14	5.44	5.21
205	5.50	5.20	4.88	4.68	4.67	4.79	6.48	5.90	5.41	5.08	4.98	5.40
305	5.48	5.39	5.14	4.99	5.11	5.12	7.14	5.63	5.07	5.33	4.77	5.02

4. Conclusions

This paper develops a new method for testing the high dimensional mean vector and covariance matrix simultaneously. The asymptotic null distribution is derived and the theoretical power function is provided. Our method not only works for $n > p$ but also works for $p \geq n$ where n is the sample size and p is the dimension. Simulation results show that the proposed test behaves well for both normal data and non-normal data. In this paper, we only consider a single population. It will be our future work to test the high dimensional mean vectors and covariance matrices simultaneously for several populations.

Table 6: Empirical test powers (in percentage) of T_n under Model II with non-Gaussian data based on 10,000 replications.

ϵ	n	Dimension p											
		Gamma(4,2)						$\chi^2(3)$					
		50	100	200	300	500	1,000	50	100	200	300	500	1,000
0.3	55	14.9	20.2	33.0	43.8	63.2	89.6	12.0	14.6	21.7	29.4	42.6	70.2
	105	34.8	51.6	76.0	88.5	98.0	99.9	27.4	37.5	56.2	71.0	88.6	99.2
	205	79.5	95.6	99.7	100	100	100	68.9	86.7	98.3	99.8	99.9	100
	305	97.4	99.9	100	100	100	100	93.9	99.3	99.9	100	100	100
0.5	55	82.1	97.1	99.9	100	100	100	67.7	89.4	99.3	99.9	100	100
	105	99.7	100	100	100	100	100	99.0	100	100	100	100	100
	205	100	100	100	100	100	100	100	100	100	100	100	100
	305	100	100	100	100	100	100	100	100	100	100	100	100

Table 7: Empirical test powers (in percentage) of T_n under Model III with non-Gaussian data based on 10,000 replications.

p_1	n	Dimension p											
		Gamma(4,2)						$\chi^2(3)$					
		50	100	200	300	500	1,000	50	100	200	300	500	1,000
0	55	44.0	31.6	20.9	16.1	12.5	9.4	33.9	23.7	15.5	13.1	10.5	8.10
	105	95.1	88.6	72.0	58.5	42.3	25.9	88.6	76.1	55.6	42.8	29.6	19.1
	205	100	100	100	99.9	99.1	89.9	100	100	99.8	99.1	93.8	74.2
	305	100	100	100	100	100	99.9	100	100	100	100	100	99.4
$[p/2]$	55	52.6	41.6	32.3	27.5	25.2	25.3	41.6	31.4	23.6	21.1	18.1	18.6
	105	97.3	94.1	86.3	79.5	71.8	66.2	92.8	84.8	71.3	62.8	54.4	48.6
	205	100	100	100	100	99.9	99.8	100	100	99.9	99.8	99.5	97.6
	305	100	100	100	100	100	100	100	100	100	100	100	100

Table 8: Empirical test powers (in percentage) of T_n under Model IV with non-Gaussian data based on 10,000 replications.

p_1	n	Dimension p											
		Gamma(4,2)						$\chi^2(3)$					
		50	100	200	300	500	1,000	50	100	200	300	500	1,000
0	55	39.3	59.9	77.5	85.4	91.1	95.8	31.1	47.4	67.1	77.5	85.9	93.2
	105	84.1	97.2	99.6	99.9	99.9	100	75.5	93.4	99.1	99.7	99.9	100
	205	99.9	100	100	100	100	100	99.6	100	100	100	100	100
	305	100	100	100	100	100	100	100	100	100	100	100	100
$[p/2]$	55	47.2	67.7	82.9	89.4	94.4	97.7	38.3	55.4	73.9	82.5	89.7	95.3
	105	89.3	98.3	99.8	99.9	100	100	81.6	95.4	99.3	99.8	99.9	100
	205	99.9	100	100	100	100	100	99.8	100	100	100	100	100
	305	100	100	100	100	100	100	100	100	100	100	100	100

Appendix A. Proofs of Theorems 2.1, 2.2 and 2.3

A.1 Proofs of Theorems 2.1 and 2.2

Recall that Theorem 2.1 is a special case of Theorem 2.2 when $\boldsymbol{\mu} = \mathbf{0}_p$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$. So, we first give the proof of Theorem 2.2. The whole proof is break into two steps. In Step 1, we will prove $\bar{\mathbf{x}}^T\bar{\mathbf{x}} - \mathrm{E}\bar{\mathbf{x}}^T\bar{\mathbf{x}} \to 0$ in probability when $\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}$ tends to a constant and $n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2$ converges as $n, p \to \infty$ with $p/n \to y \in (0, \infty)$. In Step 2, we will prove the central limit theorem (CLT) of $\mathrm{tr}(\mathbf{S}_n - \mathbf{I}_p)^2 - \mathrm{Etr}(\mathbf{S}_n - \mathbf{I}_p)^2$.

Step 1: We have

$$\bar{\mathbf{x}}^T\bar{\mathbf{x}} = n^{-2}\sum_{i\neq j}\mathbf{x}_i^T\mathbf{x}_j + n^{-2}\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{x}_i,$$

$$\mathrm{Var}(n^{-2}\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{x}_i) = n^{-3}\mathrm{Var}(\mathbf{x}_1^T\mathbf{x}_1) = n^{-3}\mathrm{E}(\mathbf{x}_1^T\mathbf{x}_1 - \mathrm{tr}\boldsymbol{\Sigma})^2 = n^{-3}[2\mathrm{tr}\boldsymbol{\Sigma}^2 + \beta_w\sum_{\ell=1}^{p}(\mathbf{e}_\ell^T\boldsymbol{\Sigma}\mathbf{e}_\ell)^2].$$

Then $\mathrm{Var}(n^{-2}\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{x}_i) \to 0$ as $p/n \to y \in (0, \infty)$. That is

$$n^{-2}\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{x}_i - n^{-1}\mathrm{E}\mathbf{x}_1^T\mathbf{x}_1 \xrightarrow{D} 0.$$

Moreover, when $\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}$ tends to a constant and $n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2$ converges, we have

$$\begin{aligned}
&\mathrm{Var}(n^{-2}\sum_{i\neq j}\mathbf{x}_i^T\mathbf{x}_j)\\
= \;& n^{-4}\sum_{i\neq j}\sum_{k\neq \ell}\mathrm{E}(\mathbf{x}_i^T\mathbf{x}_j - \boldsymbol{\mu}^T\boldsymbol{\mu})(\mathbf{x}_k^T\mathbf{x}_\ell - \boldsymbol{\mu}^T\boldsymbol{\mu})\\
= \;& 2n^{-4}n(n-1)\mathrm{E}(\mathbf{x}_1^T\mathbf{x}_2 - \boldsymbol{\mu}^T\boldsymbol{\mu})^2 + 4n^{-4}n(n-1)(n-2)\mathrm{E}(\mathbf{x}_1^T\mathbf{x}_2 - \boldsymbol{\mu}^T\boldsymbol{\mu})(\mathbf{x}_2^T\mathbf{x}_3 - \boldsymbol{\mu}^T\boldsymbol{\mu})\\
= \;& 2n^{-3}n\mathrm{tr}\boldsymbol{\Sigma}^2 + 4n^{-3}(n-1)(n-2)\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}\\
\to \;& 0.
\end{aligned}$$

This implies that,

$$n^{-2}\sum_{i\neq j}\mathbf{x}_i^T\mathbf{x}_j - n^{-1}(n-1)\boldsymbol{\mu}^T\boldsymbol{\mu} \xrightarrow{D} 0.$$

By now, we proved that

$$\bar{\mathbf{x}}^T\bar{\mathbf{x}} - \mathrm{E}\bar{\mathbf{x}}^T\bar{\mathbf{x}} \xrightarrow{D} 0.$$

Step 2: We will prove the central limit theorem of $\mathrm{tr}(\mathbf{S}_n-\mathbf{I}_p)^2-\mathrm{Etr}(\mathbf{S}_n-\mathbf{I}_p)^2$. The basic idea of this step is to use the theories of the martingale difference sequence. From the fact that

$$\mathbf{x}_i\mathbf{x}_i^T = (\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^T + (\mathbf{x}_i-\boldsymbol{\mu})\boldsymbol{\mu}^T + \boldsymbol{\mu}(\mathbf{x}_i-\boldsymbol{\mu})^T + \boldsymbol{\mu}\boldsymbol{\mu}^T,$$

we have

$$\mathrm{E}\bar{\mathbf{x}}^T\bar{\mathbf{x}} = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}\mathbf{x}_i^T\mathbf{x}_j = \boldsymbol{\mu}^T\boldsymbol{\mu} + n^{-1}\mathrm{tr}\boldsymbol{\Sigma},$$

and

$$\begin{aligned}\mathrm{Etr}\mathbf{S}_n^2 &= \mathrm{tr}\boldsymbol{\Sigma}^2 + n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2 + \beta_w n^{-1}\sum_{k=1}^{p}(\mathbf{e}_k^T\boldsymbol{\Sigma}\mathbf{e}_k)^2 + n^{-1}(\mathrm{tr}\boldsymbol{\Sigma})^2\\ &\quad +2(n+1)n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu} + 4n^{-1}\mathrm{E}(\mathbf{x}_1-\boldsymbol{\mu})^T(\mathbf{x}_1-\boldsymbol{\mu})(\mathbf{x}_1-\boldsymbol{\mu})^T\boldsymbol{\mu}\\ &\quad +2\boldsymbol{\mu}^T\boldsymbol{\mu}(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) + (\boldsymbol{\mu}^T\boldsymbol{\mu})^2.\end{aligned}$$

Then the mean of the statistic $\bar{\mathbf{x}}^T\bar{\mathbf{x}} + \mathrm{tr}(\mathbf{S}_n - \mathbf{I}_p)^2$ is

$$\begin{aligned}
\mu_A &= \mathrm{E}\bar{\mathbf{x}}^T\bar{\mathbf{x}} + \mathrm{E}\mathrm{tr}\mathbf{S}_n^2 - 2\mathrm{E}\mathrm{tr}\mathbf{S}_n + p \\
&= \boldsymbol{\mu}^T\boldsymbol{\mu} + n^{-1}\mathrm{tr}\boldsymbol{\Sigma} - 2\mathrm{tr}\boldsymbol{\Sigma} - 2\boldsymbol{\mu}^T\boldsymbol{\mu} + p \\
&\quad +\mathrm{tr}\boldsymbol{\Sigma}^2 + n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2 + \beta_w n^{-1}\sum_{k=1}^{p}(\mathbf{e}_k^T\boldsymbol{\Sigma}\mathbf{e}_k)^2 + n^{-1}(\mathrm{tr}\boldsymbol{\Sigma})^2 \\
&\quad +2(n+1)n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu} + 4n^{-1}\mathrm{E}(\mathbf{x}_1-\boldsymbol{\mu})^T(\mathbf{x}_1-\boldsymbol{\mu})(\mathbf{x}_1-\boldsymbol{\mu})^T\boldsymbol{\mu} \\
&\quad +2\boldsymbol{\mu}^T\boldsymbol{\mu}(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) + (\boldsymbol{\mu}^T\boldsymbol{\mu})^2 \\
&= n^{-1}\mathrm{tr}\boldsymbol{\Sigma} + \mathrm{tr}(\boldsymbol{\Sigma} - \mathbf{I}_p)^2 - \boldsymbol{\mu}^T\boldsymbol{\mu} \\
&\quad +n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2 + \beta_w n^{-1}\sum_{k=1}^{p}(\mathbf{e}_k^T\boldsymbol{\Sigma}\mathbf{e}_k)^2 + n^{-1}(\mathrm{tr}\boldsymbol{\Sigma})^2 \\
&\quad +2(n+1)n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu} + 4n^{-1}\mathrm{E}(\mathbf{x}_1-\boldsymbol{\mu})^T(\mathbf{x}_1-\boldsymbol{\mu})(\mathbf{x}_1-\boldsymbol{\mu})^T\boldsymbol{\mu} \\
&\quad +2\boldsymbol{\mu}^T\boldsymbol{\mu}(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) + (\boldsymbol{\mu}^T\boldsymbol{\mu})^2.
\end{aligned}$$

Let $\mathbf{r}_\ell = n^{-1/2}\mathbf{w}_\ell$ with $\mathbf{w}_\ell = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x}_\ell - \boldsymbol{\mu})$ for $\ell = 1, \ldots, n$, then we have

$$\begin{aligned}
&\mathrm{tr}(\mathbf{S}_n - \mathbf{I}_p)^2 - \mathrm{E}\mathrm{tr}(\mathbf{S}_n - \mathbf{I}_p)^2 \\
= &\sum_{j=1}^{n}(\mathrm{E}_j - \mathrm{E}_{j-1})(\mathrm{tr}\mathbf{S}_n^2 - 2\mathrm{tr}\mathbf{S}_n + p) \\
= &\sum_{j=1}^{n}\Big[(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n^2 - 2(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n\Big],
\end{aligned}$$

where E_j is the conditional expectation given $\{\mathbf{r}_1, \ldots, \mathbf{r}_j\}$,

$$(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n = \mathrm{E}_j\mathrm{tr}\mathbf{S}_n - \mathrm{E}_{j-1}\mathrm{tr}\mathbf{S}_n = \mathbf{r}_j^T\boldsymbol{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma} + 2n^{-1/2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_j,$$

and

$$
\begin{aligned}
&(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n^2 \\
= \;& 2n^{-1}(n-\ell)(\mathbf{r}_\ell^T\boldsymbol{\Sigma}^2\mathbf{r}_\ell - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2) + [(\mathbf{e}_\ell^T\boldsymbol{\Sigma}\mathbf{e}_\ell)^2 - \mathrm{E}(\mathbf{e}_\ell^T\boldsymbol{\Sigma}\mathbf{e}_\ell)^2] \\
&+2\mathbf{r}_\ell^T\boldsymbol{\Sigma}(\sum_{j=1}^{\ell-1}\mathbf{r}_j\mathbf{r}_j^T)\boldsymbol{\Sigma}\mathbf{r}_\ell - 2\mathrm{tr}\boldsymbol{\Sigma}^2(\sum_{j=1}^{\ell-1}\mathbf{r}_j\mathbf{r}_j^T) + 4n^{-3/2}(n-\ell)\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{3/2}\mathbf{r}_\ell \\
&+4n^{-1/2}\mathbf{r}_\ell^T\boldsymbol{\Sigma}(\sum_{j=1}^{\ell-1}\mathbf{r}_j\mathbf{r}_j^T)\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_\ell - 4n^{-1/2}\mathrm{tr}\boldsymbol{\Sigma}^{3/2}(\sum_{j=1}^{\ell-1}\mathbf{r}_j\mathbf{r}_j^T)\boldsymbol{\mu}^T \\
&+4n^{-1/2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}(\sum_{j=1}^{\ell-1}\mathbf{r}_j\mathbf{r}_j^T)\boldsymbol{\Sigma}\mathbf{r}_\ell + 4n^{-1/2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_\ell(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) \\
&+4n^{-1/2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_\ell(\mathbf{r}_\ell^T\boldsymbol{\Sigma}\mathbf{r}_\ell - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) - 4n^{-1/2}\mathrm{E}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_\ell(\mathbf{r}_\ell^T\boldsymbol{\Sigma}\mathbf{r}_\ell - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) \\
&+2(1+n^{-1})\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_\ell\mathbf{r}_\ell^T\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu} - 2(1+n^{-1})n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu} \\
&+4n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_\ell\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}(\sum_{j=1}^{\ell-1}\mathbf{r}_j) + 4n^{-1/2}\boldsymbol{\mu}^T\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_\ell \\
&+2(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\mu})(\mathbf{r}_\ell^T\boldsymbol{\Sigma}\mathbf{r}_\ell - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) + 4(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\mu})\mathbf{r}_\ell^T\boldsymbol{\Sigma}(\sum_{j=1}^{\ell-1}\mathbf{r}_j).
\end{aligned}
$$

The $\{(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n^2, j = 1, \ldots, n\}$ and $\{(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n, j = 1, \ldots, n\}$ are two martingale difference sequences where $E_j(\cdot) = E(\cdot|\mathcal{F}_j)$ and $\mathcal{F}_0 = \sigma\{\Phi, \Omega\}$, $\mathcal{F}_j = \sigma\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_j\}, j \geq 1$. To derive the CLT of $\mathrm{tr}(\mathbf{S}_n - \mathbf{I}_p)^2$, we only need to verify the Lyapunov conditions for the above two martingale difference sequences. In fact, Bai and Silverstein (2004) has verified the Lyapunov conditions for the martingale difference sequence. Thus we only need to derive the variances of the martingale difference sequence. Defining

$$\begin{aligned}\sigma_{11A} &= \sum_{j=1}^{n}[(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n]^2,\\ \sigma_{22A} &= \sum_{j=1}^{n}[(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n^2]^2,\\ \sigma_{12A} &= \sum_{j=1}^{n}[(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n][(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n^2],\end{aligned}$$

then we have

$$\sigma_A^2 = \sigma_{11A} + 4\sigma_{22A} - 4\sigma_{12A}.$$

By the CLT of the martingale difference sequence, we obtain that

$$\sigma_A^{-1}(T_n - \mu_A) \xrightarrow{D} N(0,1).$$

Let $o_p(1)$ denote the convergence to zero in probability. In fact, we have

$$\sum_{j=1}^{n}\mathrm{E}_{j-1}(\mathbf{r}_j^T\boldsymbol{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2 = 2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^T\boldsymbol{\Sigma}\mathbf{e}_\ell)^2,$$

$$4\sum_{j=1}^{n}\mathrm{E}_{j-1}(n^{-1/2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_j)^2 = 4(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}),$$

$$\sum_{\ell=1}^{n}4n^{-2}(n-\ell)^2\mathrm{E}_{\ell-1}(\mathbf{r}_\ell^T\boldsymbol{\Sigma}^2\mathbf{r}_\ell - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2)^2 = (4/3)[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^4 + \beta_w n^{-1}\sum_{j=1}^{p}(\mathbf{e}_j^T\boldsymbol{\Sigma}^2\mathbf{e}_j)^2],$$

$$4\sum_{\ell=1}^{n}(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\mu})^2\mathrm{E}_{\ell-1}(\mathbf{r}_\ell^T\boldsymbol{\Sigma}\mathbf{r}_\ell - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2 = 4(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\mu})^2[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2 + \beta_w n^{-1}\sum_{j=1}^{p}(\mathbf{e}_j^T\boldsymbol{\Sigma}\mathbf{e}_j)^2],$$

$$4\sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}\Big(\mathbf{r}_\ell^T\boldsymbol{\Sigma}\sum_{i=1}^{\ell-1}\mathbf{r}_i\mathbf{r}_i^T\boldsymbol{\Sigma}\mathbf{r}_\ell - n^{-1}\sum_{i=1}^{\ell-1}\mathbf{r}_i^T\boldsymbol{\Sigma}^2\mathbf{r}_i\Big)^2$$
$$- \quad [4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2)^2+(8/3)n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^4+(4/3)n^{-1}\sum_{j=1}^{p}(\mathbf{e}_j^T\boldsymbol{\Sigma}^2\mathbf{e}_j)^2] = o_p(1),$$

$$16n^{-2}\sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}[\mathbf{r}_\ell^T\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}\Big(\sum_{i=1}^{\ell-1}\mathbf{r}_i\Big)^T\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}]^2 - 8(n^{-1}\boldsymbol{\mu}\boldsymbol{\Sigma}\boldsymbol{\mu})^2 = o_p(1),$$

$$16n^{-1}\sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}[\mathbf{r}_\ell^T\boldsymbol{\Sigma}\Big(\sum_{i=1}^{\ell-1}\mathbf{r}_i\Big)\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_\ell - n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{3/2}\Big(\sum_{i=1}^{\ell-1}\mathbf{r}_i\Big)]^2$$
$$- \quad 8[n^{-2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^3\boldsymbol{\mu}+(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu})(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2)+n^{-2}\sum_{k=1}^{p}\mathbf{e}_k^T\boldsymbol{\Sigma}^2\mathbf{e}_k(\mathbf{e}_k^T\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu})^2] = o_p(1),$$

$$16n^{-1}\sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}[\mathbf{r}_\ell^T\boldsymbol{\Sigma}\sum_{i=1}^{\ell-1}\mathbf{r}_i\mathbf{r}_i^T\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}]^2 - [(16/3)n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^3\boldsymbol{\mu}+8(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu})^2] = o_p(1),$$

$$16(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\mu})^2\sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}[\sum_{i=1}^{\ell-1}\mathbf{r}_i^T\boldsymbol{\Sigma}\mathbf{r}_\ell]^2 - 8(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\mu})^2\mathrm{tr}\boldsymbol{\Sigma}^2 = o_p(1),$$

$$4\sum_{\ell=1}^{n}\mathrm{E}_\ell[\mathbf{r}_\ell^T\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_\ell - n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}] = 4n^{-1}[2(\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu})^2+\beta_w\sum_{k=1}^{p}(\mathbf{e}_k^T\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{e}_k)^2],$$

$$\sum_{\ell=1}^{n}\mathrm{E}[(\mathbf{r}_\ell^T\boldsymbol{\Sigma}\mathbf{r}_\ell)^2 - \mathrm{E}(\mathbf{r}_\ell^T\boldsymbol{\Sigma}\mathbf{r}_\ell)^2]^2 = 4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2+\beta_w n^{-1}\sum_{k=1}^{p}(\mathbf{e}_k^T\boldsymbol{\Sigma}\mathbf{e}_k)^2],$$

$$16n^{-1}(\boldsymbol{\mu}^T\boldsymbol{\mu})^2\sum_{\ell=1}^{n}\mathrm{E}(\mathbf{r}_\ell^T\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{1/2}\mathbf{r}_\ell) = 16n^{-1}(\boldsymbol{\mu}^T\boldsymbol{\mu})^2\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu},$$

$$16n^{-1}(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2\sum_{\ell=1}^{n}\mathrm{E}(\mathbf{r}_\ell^T\boldsymbol{\Sigma}\boldsymbol{\mu})^2 = 16(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}),$$

$$16n^{-1}\sum_{\ell=1}^{n}\mathrm{E}[\mathbf{r}_\ell^T\mathbf{\Sigma}^{1/2}\boldsymbol{\mu}(\mathbf{r}_\ell^T\mathbf{\Sigma}\mathbf{r}_\ell - n^{-1}\mathrm{tr}\mathbf{\Sigma}) - \mathrm{E}\mathbf{r}_\ell^T\mathbf{\Sigma}^{1/2}\boldsymbol{\mu}(\mathbf{r}_\ell^T\mathbf{\Sigma}\mathbf{r}_\ell - n^{-1}\mathrm{tr}\mathbf{\Sigma})]^2$$
$$\le 16n^{-1}\sum_{\ell=1}^{n}\mathrm{E}[\mathbf{r}_\ell^T\mathbf{\Sigma}^{1/2}\boldsymbol{\mu}(\mathbf{r}_\ell^T\mathbf{\Sigma}\mathbf{r}_\ell - n^{-1}\mathrm{tr}\mathbf{\Sigma})]^2$$
$$\le 16(n^{-1}\boldsymbol{\mu}^T\mathbf{\Sigma}\boldsymbol{\mu}),$$

$$16n^{-3}\sum_{\ell=1}^{n}(n-\ell)^2\mathrm{E}[\mathbf{r}_\ell^T\mathbf{\Sigma}^{3/2}\boldsymbol{\mu}]^2 = (16/3)n^{-1}\boldsymbol{\mu}^T\mathbf{\Sigma}^3\boldsymbol{\mu},$$

$$16(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\mu})^2\sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}(\mathbf{r}_\ell^T\mathbf{\Sigma}\sum_{j=1}^{\ell-1}\mathbf{r}_j)^2 = 8(n^{-1}\boldsymbol{\mu}^T\boldsymbol{\mu})\mathrm{tr}\mathbf{\Sigma}^2.$$

Then we have

$$\begin{aligned}
&\sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n^2]^2 + \sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n]^2 \\
= \;& C_0\Big[n^{-1}\mathrm{tr}\mathbf{\Sigma}^2 + n^{-1}\sum_{k=1}^{p}(\mathbf{e}_k^T\mathbf{\Sigma}\mathbf{e}_k)^2 + n^{-1}\boldsymbol{\mu}^T\mathbf{\Sigma}\boldsymbol{\mu} + (n^{-1}\boldsymbol{\mu}^T\mathbf{\Sigma}\boldsymbol{\mu})^2 + n^{-1}\boldsymbol{\mu}^T\mathbf{\Sigma}^3\boldsymbol{\mu} \\
&+(n^{-1}\mathrm{tr}\mathbf{\Sigma}^2)^2 + n^{-1}\mathrm{tr}\mathbf{\Sigma}^4 + n^{-1}\sum_{k=1}^{p}(\mathbf{e}_k^T\mathbf{\Sigma}^2\mathbf{e}_k)^2 + (n^{-1}\mathrm{tr}\mathbf{\Sigma})^3 + n^{-1}(\boldsymbol{\mu}^T\boldsymbol{\mu})^2\mathrm{tr}\mathbf{\Sigma}^2 \\
&+(n^{-1}\boldsymbol{\mu}^T\mathbf{\Sigma}\boldsymbol{\mu})(n^{-1}\mathrm{tr}\mathbf{\Sigma}^2) + (n^{-1}\boldsymbol{\mu}^T\mathbf{\Sigma}\boldsymbol{\mu})n^{-1}\sum_{k=1}^{p}(\mathbf{e}_k^T\mathbf{\Sigma}\mathbf{e}_k)^2 + \boldsymbol{\mu}^T\boldsymbol{\mu} \\
&+n^{-1}\sum_{k=1}^{p}(\mathbf{e}_k^T\mathbf{\Sigma}^{1/2}\boldsymbol{\mu})^4 + (n^{-1}\boldsymbol{\mu}^T\boldsymbol{\mu})^2(n^{-1}\boldsymbol{\mu}^T\mathbf{\Sigma}\boldsymbol{\mu}) + n^{-1}(\boldsymbol{\mu}^T\mathbf{\Sigma}\boldsymbol{\mu})^2\Big] + o_p(1),
\end{aligned}$$

with a constant C_0. Under Assumption [B], we have

$$2\sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n^2]^2 + 2\sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n]^2 \le C_1 + C_2(\boldsymbol{\mu}^T\boldsymbol{\mu})^2 + C_3\boldsymbol{\mu}^T\boldsymbol{\mu} + o_p(1),$$

where C_1, C_2, C_3 are positive constants. By now, we have proved $\sigma_A^2 \le C_1 + C_2(\boldsymbol{\mu}^T\boldsymbol{\mu})^2 + C_3\boldsymbol{\mu}^T\boldsymbol{\mu}$.

Especially, under H_0, we have $\boldsymbol{\mu} = \mathbf{0}_p$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$. It is trivial to obtain the following result:

$$\sigma_0^{-1}(T_n - \mu_0) \xrightarrow{D} N(0,1),$$

where

$$\mu_0 = p^2/n + (\beta_w + 2)p/n,$$

and

$$\begin{aligned} \sigma_0^2 &= 4n^{-1}(2p + \beta_w p) + 4y^2 + 8(y-1)n^{-1}(2p + \beta_w p) + 4(y-1)^2 n^{-1}(2p + \beta_w p) \\ &= 4y^3(2 + \beta_w) + 4y^2. \end{aligned}$$

The proofs of Theorem 2.1 and 2.2 are completed.

A.2 Proof of Theorem 2.3

We consider the case of $\boldsymbol{\mu} = \mathbf{0}_p$ and $\boldsymbol{\Sigma} = \mathbf{I}_p + \epsilon \mathbf{I}_p$ with $\epsilon > 0$. We have

$$\begin{aligned} & \mu_A - \mu_0 \\ = \; & n^{-1}\mathrm{tr}\boldsymbol{\Sigma} + \mathrm{tr}(\boldsymbol{\Sigma} - \mathbf{I}_p)^2 + n^{-1}\beta_w \sum_{\ell=1}^{p} (\mathbf{e}_\ell^T \boldsymbol{\Sigma} \mathbf{e}_\ell)^2 + n^{-1}(\mathrm{tr}\boldsymbol{\Sigma})^2 + n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2 \\ & -[p^2/n + (\beta_w + 2)p/n] \\ = \; & (n^{-1}\mathrm{tr}\boldsymbol{\Sigma} - p/n) + \mathrm{tr}(\boldsymbol{\Sigma} - \mathbf{I}_p)^2 + (n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2 - p/n) \\ & +[n^{-1}\beta_w \sum_{\ell=1}^{p} (\mathbf{e}_\ell^T \boldsymbol{\Sigma} \mathbf{e}_\ell)^2 - p/n] + [n^{-1}(\mathrm{tr}\boldsymbol{\Sigma})^2 - p^2/n] \\ \to \; & +\infty, \quad \text{as } p/n \to y \in (0, \infty). \end{aligned}$$

Besides, we have

$$\begin{aligned}
\sigma_A^2 &= 4[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^4 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^T\boldsymbol{\Sigma}^2\mathbf{e}_\ell)^2] + 4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2)^2 \\
&\quad +8(n^{-1}\mathrm{tr}\boldsymbol{\Sigma} - 1)[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^3 + \beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^T\boldsymbol{\Sigma}^2\mathbf{e}_\ell\mathbf{e}_\ell^T\boldsymbol{\Sigma}\mathbf{e}_\ell] \\
&\quad +4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma} - 1)^2[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^T\boldsymbol{\Sigma}\mathbf{e}_\ell)^2] \\
&= 4(1+\epsilon)^4(2p/n + \beta_w p/n) + 4(p/n)^2(1+\epsilon)^4 \\
&\quad +8(1+\epsilon)^3[(p/n)(1+\epsilon) - 1](2n^{-1}p + \beta_w n^{-1}p) \\
&\quad +4(1+\epsilon)^2[n^{-1}p(1+\epsilon) - 1]^2(2p/n + \beta_w p/n) \\
&\to 4(1+\epsilon)^4(2+\beta_w)y + 4y^2(1+\epsilon)^4 \\
&\quad +8(1+\epsilon)^3[y(1+\epsilon) - 1](2+\beta_w)y + 4(1+\epsilon)^2[y(1+\epsilon) - 1]^2(2+\beta_w)y,
\end{aligned}$$

as $p/n \to y \in (0, \infty)$. Then the power function satisfies

$$\begin{aligned}
\beta_{T_n} &= 1 - \Phi\left(\frac{\mu_0 - \mu_A + \sigma_0 q_{1-\alpha/2}}{\sigma_A}\right) + \Phi\left(\frac{\mu_0 - \mu_A + \sigma_0 q_{\alpha/2}}{\sigma_A}\right) \\
&\to 1, \quad \text{as } p/n \to y \in (0, \infty)
\end{aligned}$$

given the constant ϵ. Similarly, when $\boldsymbol{\Sigma} = \mathbf{I}_p$ and $\boldsymbol{\mu} = \epsilon\mathbf{1}_p$ with $\epsilon > 0$, it can be proved that

$$\beta_{T_n} = 1 - \Phi\left(\frac{\mu_0 - \mu_A + \sigma_0 q_{1-\alpha/2}}{\sigma_A}\right) + \Phi\left(\frac{\mu_0 - \mu_A + \sigma_0 q_{\alpha/2}}{\sigma_A}\right) \to 1,$$

as $p/n \to y \in (0, \infty)$.

The proof of Theorem 2.3 is completed.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis.* 3rd Edition. John Wiley & Sons.

Bai, Z. D. and Hewa, S. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica* **6**, 311-329.

Bai, Z., Jiang, D., Yao, J. F. and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37**, 3822-3840.

Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. *Ann. Probab.* **32(1)**, 553-605.

Chen, S. X., Zhang, L. X. and Zhong, P. S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105**, 810-819.

Chou, C. Y., Chen, C. H., Liu, H. R., and Huang, X. R. (2003). Economic-statistical design of multivariate control charts for monitoring the mean vector and covariance matrix. *Journal of Loss Prevention in the Process Industries* **16**, 9-18.

Das Gupta, S. (1969). Properties of power functions of some tests concerning dispersion matrices of multivariate normal distributions. *Ann. Math. Statist.* **40**, 697-701.

Jiang, T. F. and Qi, Y. C. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scandinavian Journal of Statistics* **42**, 988-1009.

Jiang, T. F. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Stat.* **41**, 2029-2074.

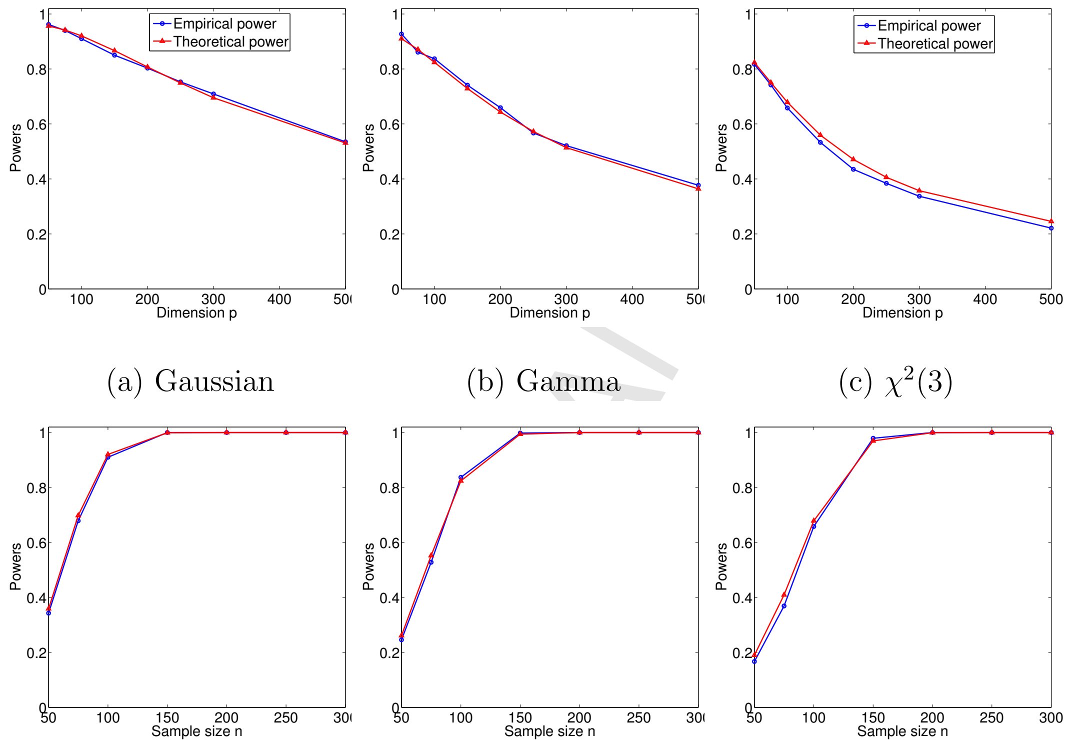
Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)* **72**, 507-536.

Nagarsenker, B. N. and Pillai, K. C. S. (1974). Distribution of the likelihood ratio criterion for testing $\mathbf{\Sigma} = \mathbf{\Sigma}_0$, $\boldsymbol{\mu} = \boldsymbol{\mu}_0$. *J. Multivariate Anal.* **4**, 114-122.

Pham-Gia, T. and Turkkan, N. (2009). Testing a covariance matrix: exact null distribution of its likelihood criterion. *Journal of Statistical Computation and Simulation* **79**, 1331-1340.

Sugiura, N. and Nagao, H. (1968). Unbiasedness of some test criteria for the equality of one or two covariance matrices. *Ann. Math. Statist.* **39**, 1686-1692.

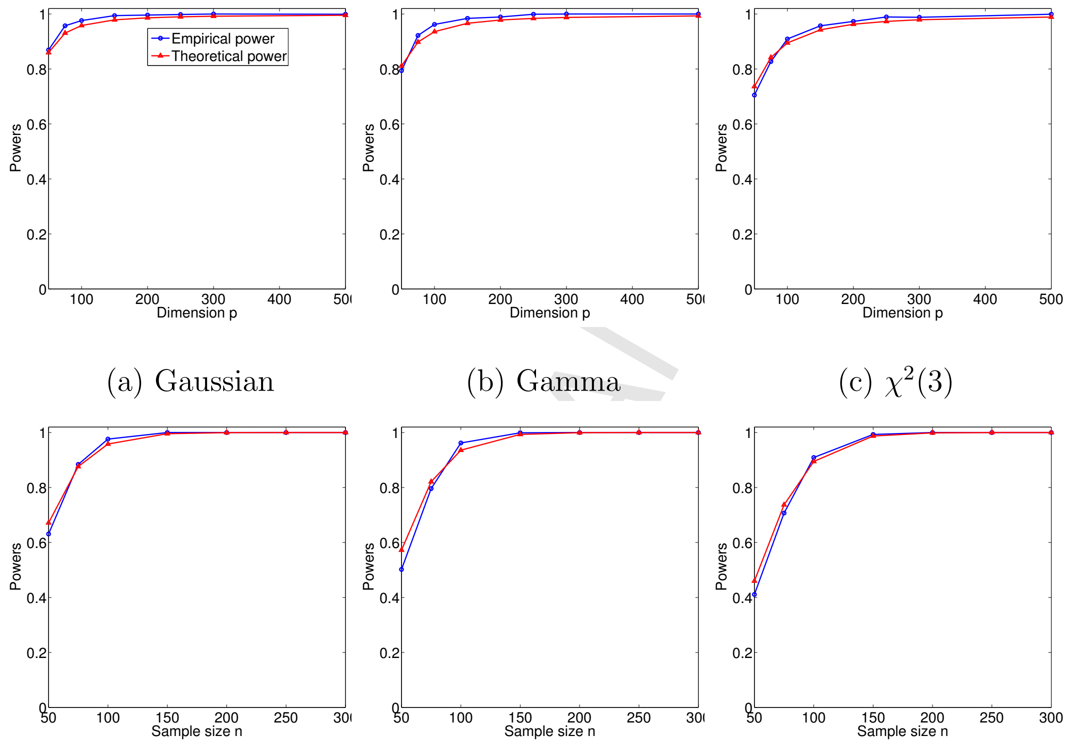
Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78**, 509-521.



(a) Gaussian (b) Gamma (c) $\chi^2(3)$

(d) Gaussian (e) Gamma (f) $\chi^2(3)$

Figure 1: Empirical powers and theoretical powers of T_n under the alternative hypothesis $H_1 : \boldsymbol{\mu} = \mathbf{0}_p, \boldsymbol{\Sigma} = (\sigma_{ij})_{p\times p}$ based on 1,000 simulation iterations, where $\sigma_{ij} = 1$ for $i = j$, $\sigma_{ij} = 0.1$ for $0 < |i-j| \leq 3$, and $\sigma_{ij} = 0$ for $|i-j| > 3$. The sample size of the top panels is fixed as $n = 100$ and the dimension of the bottom panels is fixed as $p = 100$.



(a) Gaussian (b) Gamma (c) $\chi^2(3)$

(d) Gaussian (e) Gamma (f) $\chi^2(3)$

Figure 2: Empirical powers and theoretical powers of of T_n under the alternative hypothesis $H_1 : \boldsymbol{\mu} = \mathbf{0}_p, \boldsymbol{\Sigma} = 0.95\mathbf{I}_p + 0.05\mathbf{J}_p$ based on 1,000 simulation iterations, where $\mathbf{J}_p$ denotes the $p \times p$ matrix with all elements being 1. The sample size of the top panels is fixed as $n = 100$ and the dimension of the bottom panels is fixed as $p = 100$.

Highlights

- Developing a new method for simultaneously testing the mean vector and covariance matrix of high-dimensional data.
- Establishing the central limit theorem of the proposed test under the asymptotic regime that the dimension of data and the sample size tend to infinity proportionally.
- The new method is applicable for the case of large dimension and small sample size, and requires no any assumptions of population distribution except the population fourth moment existing.