Testing Block-Diagonal Covariance Structure for High-Dimensional Data

MASASHI HYODO[1*], NOBUMICHI SHUTOH[2], TAKAHIRO NISHIYAMA[3], AND TATJANA PAVLENKO[4]

[1] Department of Mathematical Sciences, Graduate School of Engineering, Osaka Prefecture University

[2] Graduate School of Maritime Sciences, Kobe University

[3] Department of Business Administration, Senshu University

[4] Department of Mathematics, KTH Royal Institute of Technology

A test statistic is developed for making inference about a block-diagonal structure of the covariance matrix when the dimensionality, p exceeds n, where $n = N - 1$ and N denotes the sample size. The suggested procedure extends the complete independence results derived by Srivastava (2005) and Schott (2005). Since the classical hypothesis testing methods based on the likelihood ratio degenerate when $p > n$, the main idea is to turn instead to a distance function between the null and alternative hypothesis. The test statistic is then constructed using a consistent estimator of this function, where consistency is considered in an asymptotic framework that allows p to grow together with n. By adapting some technical results of Srivastava (2005) and Schott (2005), the suggested statistic is also shown to have an asymptotic normality under the null hypothesis. Some auxiliary results on the moments of products of multivariate normal random vectors and higher order moments of Wishart matrices, important for our evaluation of the test statistic, are derived. We perform empirical power analysis for a number of alternative covariance structures.

Key Words Block-diagonal covariance structure; High dimensionality.

Mathematics Subject Classification 62H12; 62H15.

1 Introduction

In recent years, a number of statistical methods has been developed to meet the challenge of analyzing high-dimensional data, such as e.g., DNA microarray gene expressions, where the number of feature variables, p typically greatly exceeds n. A particular attention has been paid to estimation of the $p \times p$ covariance matrix, Σ, and its inverse, Σ^{-1}, as they play a major role in many statistical procedures, examples include e.g., linear and quadratic discriminant analysis, etc.

Because of the difficult nature of the task of reliable estimation of Σ and Σ^{-1} when $p \gg n$, there has been a flurry of recent activity in the literature on deriving test procedures for rather restrictive covariance structures. Examples include e.g., procedures for testing identity ($H_{01} : \Sigma = I_p$), sphericity ($H_{02} : \Sigma = \sigma^2 I_p$), diagonal structure ($H_{03} : \Sigma = \text{diag}(d_1, \cdots, d_p)$) and independence of two sub-vectors ($H_{04} : \Sigma = \text{diag}(\Sigma_1, \Sigma_2)$) of the covariance matrix have been developed and studied in a high-dimensional setting. The problem of testing the hypothesis in H_{01} is treated in Ledoit and Wolf (2002), Schott (2005), Srivastava (2005) and Fisher (2012). Srivastava (2005) also derived distributions of test statistics under H_{02} and H_{03}, respectively in a growing dimensions, i.e., assuming that $n = O(p^\delta), 0 < \delta \leq 1$. In addition, Srivastava and Reid (2012) proposed test statistic for H_{04}. Recently, Srivastava, Kollo and Rosen (2011) discussed H_{01}, H_{02} and H_{03} under non-normal assumptions.

Apparently, testing the above hypotheses can essentially simplify the analyses of data sets where $p \gg n$; for instance, if Σ satisfies either of H_{01}, H_{02} or H_{03}, then the univariate statistical approaches are applicable. Observe however, that many practical real world problems are *structured*, meaning that the variables can be clustered or grouped into blocks, which share similar connectivity or correlation patterns. For instance, genes can be partitioned into pathways, i.e., into blocks of variables, where the connection within a pathway might be more stronger than connections between pathways, and where the block-diagonal covariance structure naturally comes. Hence, the pathway-level analysis can provide biologically meaningful hypotheses on the block-diagonality, for which it would be desirable obtain the significance tests that can accommodate $p \gg n$ settings.

When $p \gg n$ the verification of the block-diagonal structure is very important; if the null hypothesis holds then the total number of unknown parameters to be estimated for Σ reduces to $\sum_{i=1}^{b} p_i(p_i+1)/2$ instead of $p(p+1)/2$ and, assuming that the block size $p_i < n$, a local estimation of each block-diagonal entry of Σ or Σ^{-1} can be obtained using the standard maximum likelihood approach.

It is also interesting to note, that the block-diagonal covariance structure attracts much attention across the fields of convex optimization and machine learning, with the main focus on the Gaussian graph structure learning procedures. With the same motivation as in parameter estimation, a special attraction in the context of covariance graph structure learning hinges on the essential reduction of the computational complexity in high dimensions. As it is shown in e.g., Pavlenko et al. (2012) and Danaher et al. (2014), the decomposition of the sample covariance graphs into a set of connected components induced by the block-diagonality, makes it possible to solve otherwise computationally infeasible large-scale problems.

Our objective in this paper is to derive a new test procedure for testing hypothesis that Σ has a block-diagonal structure. Due to the failure of the classical likelihood ratio based procedures when $p > n$, we turn to a distance function between the null and alternative hypothesis, and suggest a test statistic based on a consistent estimator of this function where consistency is stated in the asymptotic framework which allows p to grow to ∞ along with n. To evaluate properties of our test, we prove a number of important theoretical results. In particular, to prove the unbiasedness and consistency of the estimator for a distance function, some higher order moments of Wishart matrix and moments of products of multivariate normal vectors are established. Further, to derive the asymptotic null distribution, the martingale difference central limit theorem is proven for the scaled test statistic.

The rest of the paper is organized as follows. Section 2 provides description of our new test statistic along with the asymptotic framework, and states main asymptotic properties of the suggested test statistic. Section 3 provides auxiliary lemmas followed by proofs of main theorems stated in Section 2. In Section 4, we provide simulation experiments to study the finite sample test performance with a variety of alternative covariance structures. At last, we give some concluding remarks.

2 Description of the test

For the development of the test we have i.i.d. observations, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n+1}$ coming from $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, and our goal is to derive a test for a candidate covariance structure to be block-diagonal, i.e., this test is equivalent to test for mutual independence of k components of the observed vector $\boldsymbol{x}$.

Now suppose that $p = \sum_{i=1}^{k} p_i$, where p_i is the size of the i-th block and k is the number of blocks. Let us partition $\boldsymbol{x}$, $\boldsymbol{\mu}$ and Σ into k components, as

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}^{(1)} \\ \boldsymbol{x}^{(2)} \\ \vdots \\ \boldsymbol{x}^{(k)} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \\ \vdots \\ \boldsymbol{\mu}^{(k)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \cdots & \Sigma_{kk} \end{pmatrix},$$

where $\boldsymbol{x}^{(i)}$ and $\boldsymbol{\mu}^{(i)}$ are $p_i \times 1$ and Σ_{ij} is $p_i \times p_j$, $i, j = 1, \cdots, k$. Then the corresponding sample estimators

$$\bar{\boldsymbol{x}} = \frac{1}{n+1} \sum_{i=1}^{n+1} \boldsymbol{x}_i, \; S = \frac{1}{n} \sum_{i=1}^{n+1} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$$

can be partitioned accordingly, as

$$\bar{\boldsymbol{x}} = \begin{pmatrix} \bar{\boldsymbol{x}}^{(1)} \\ \bar{\boldsymbol{x}}^{(2)} \\ \vdots \\ \bar{\boldsymbol{x}}^{(k)} \end{pmatrix}, \qquad S = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ S_{21} & S_{22} & \cdots & S_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{k1} & S_{k2} & \cdots & S_{kk} \end{pmatrix}.$$

We are interested to test the hypothesis

$$H_0 : \Sigma_{ij} = O_{ij} \quad \text{vs.} \quad H_A : \Sigma_{ij} \neq O_{ij} \quad \text{for at least one pair } (i, j), \; i < j, \tag{2.1}$$

where O_{ij} is $p_i \times p_j$ zero matrix for $i, j = 1, \ldots, k$. Note that $\Sigma = \Sigma_d$ if H_0 is true, where $\Sigma_d = \mathrm{diag}(\Sigma_{11}, \Sigma_{22}, \cdots, \Sigma_{kk})$.

We define the distance measure between Σ_d and Σ as the normalized Frobenious matrix norm, which can be re-expressed by

$$\begin{aligned}\|\Sigma_d - \Sigma\|_F^2 &= \frac{\mathrm{tr}(\Sigma_d - \Sigma)(\Sigma_d - \Sigma)'}{p} = \frac{2\sum_{j=2}^{k}\sum_{i=1}^{j-1}\mathrm{tr}\,\Sigma_{ji}\Sigma_{ij}}{p} \\ &= \frac{\sum_{j=1}^{k}\sum_{i=1}^{k}\mathrm{tr}\,\Sigma_{ji}\Sigma_{ij}}{p} - \frac{\sum_{j=1}^{k}\mathrm{tr}\,\Sigma_{jj}^2}{p} = a_2 - b_2,\end{aligned}$$

where $a_2 = \mathrm{tr}\,\Sigma^2/p,\ b_2 = \sum_{j=1}^{k} a_2^{(j)}/p$, and $a_2^{(j)} = \mathrm{tr}\,\Sigma_{jj}^2$ $(j = 1, \ldots, k)$.

Under H_0, $\|\Sigma_d - \Sigma\|_F^2$ equals zero, thus (2.1) is equivalent to the following test

$$H_0 : \|\Sigma_d - \Sigma\|_F^2 = 0 \quad \text{vs.} \quad H_A : \|\Sigma_d - \Sigma\|_F^2 > 0. \tag{2.2}$$

A test for the above hypothesis H_0 vs. H_A can be based on the (n, p)-consistent estimator of $\|\Sigma_d - \Sigma\|_F^2$, where consistency is in high-dimensional framework which is defined more precisely below.

Let

$$\hat{a}_2 = \frac{n^2}{(n-1)(n+2)}\frac{1}{p}\left\{\mathrm{tr}\,S^2 - \frac{1}{n}(\mathrm{tr}\,S)^2\right\}, \tag{2.3}$$

$$\hat{a}_2^{(i)} = \frac{n^2}{(n-1)(n+2)}\left\{\mathrm{tr}\,S_{ii}^2 - \frac{1}{n}(\mathrm{tr}\,S_{ii})^2\right\}. \tag{2.4}$$

These estimators are proposed in Bai and Saranadasa (1996) and Srivastava (2005). By using these estimators, we consider the estimator T of $\|\Sigma_d - \Sigma\|_F^2$ by

$$T = \hat{a}_2 - \hat{b}_2, \ \text{where } \hat{b}_2 = \frac{1}{p}\sum_{i=1}^{k}\hat{a}_2^{(i)}. \tag{2.5}$$

We make the following assumptions for deriving the null distribution of T:

(A1) p_j $(j = 1, 2, \ldots)$ is fixed and $k \to \infty$,

(A2) $n = O(k^\delta)$ $(0 < \delta \leq 1)$,

(A3) $\sum_{i \neq j}^{k} a_2^{(i)} a_2^{(j)} \asymp k^2$,

(A4) $\sum_{j=1}^{k} a_4^{(j)} = o(k^2), \ \sum_{j=1}^{k} a_2^{(j)^2} = o(k^2)$, where $a_4^{(j)} = \operatorname{tr}\Sigma_{jj}^4$ $(j = 1, \ldots, k)$. The assumption "p_j is fixed" in (A1) means that the number of small block size matrix is large. Since $p = \sum_{i=1}^k p_i$, we note that if (A1) holds, then it holds that $p \asymp k$. From Assumption (A2) and $p \asymp k$, we note that $n = O(p^\delta)$.

In the next theorem, we propose an estimator of $\|\Sigma_d - \Sigma\|_F^2$ and show that it is unbiased and consistent.

Theorem 2.1. (Unbiasedness and consistency)

In addition to (A1)-(A4)*, we assume that*

$$\text{(A5)}\ a_2 = O(1),\ a_4 = \frac{\operatorname{tr}\Sigma^4}{p} = O(k),\ \sum_{i\neq j}^{k} \operatorname{tr}\Sigma_{ji}\Sigma_{ii}\Sigma_{ij}\Sigma_{jj} = O(k^2),$$

$$\sum_{i\neq j}^{k} (\operatorname{tr}\Sigma_{ij}\Sigma_{ji})^2 = O(k^2).$$

Then for the estimator T it holds that $\mathrm{E}[T] = \|\Sigma_d - \Sigma\|_F^2$ and $T \xrightarrow{P} \|\Sigma_d - \Sigma\|_F^2$.

Proof. See Section 3.

From Theorem 2.1, a test for the block diagonal structure can be based on the statistic T. In the next theorem we show that the null distribution of T is asymptotically normal.

Theorem 2.2. (Asymptotic normality)

Under H_0 and Assumptions (A1)-(A4)*,*

$$\frac{nT}{\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where

$$\sigma^2 = \lim_{k\to\infty} \sum_{j=2}^{k}\sum_{i=1}^{j-1} \frac{8(n-1)(n+2)a_2^{(j)}a_2^{(i)}}{n^2p^2}.$$

Proof. See Section 3.

Clearly, to make T in (2.5) practically workable, we need to estimate its variance, i.e., we need to estimate quantities $a_2^{(j)}$ and $a_2^{(i)}$ in σ^2 and these estimators must be consistent under high-dimensionality.

Lemma 2.3. *Suppose that*

$$\hat{\sigma}^2 = \sum_{j=2}^{k}\sum_{i=1}^{j-1} \frac{8(n-1)(n+2)\hat{a}_2^{(j)}\hat{a}_2^{(i)}}{n^2p^2}.$$

Then, under H_0 and Assumptions (A1)-(A4), $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$.

Proof. See Section 3.

By using Slutzky's theorem and the asymptotic set up given in Theorem 2.2 and Lemma 2.3, we obtain the following theorem.

Theorem 2.4. *Under H_0 and Assumptions* (A1)-(A4), *it holds that*

$$\frac{nT}{\hat{\sigma}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Thus, by using Theorem 2.4, we propose a test based on the test statistic T and reject H_0 at the approximate level α when

$$\frac{nT}{\hat{\sigma}} > z_\alpha,$$

where z_α is the upper critical value of standard normal distribution. Note that according to (2.1) the test is also one-sided.

We will also consider the special case of H_0 in which we set $p_i = 1$. In this case, a hypothesis H_0 reduces to the test of completely independence hypothesis H_{03}, see Section 1. Srivastava (2005) proposed the test statisticfor testing H_{03} as

$$T_3 = \frac{n}{2}\frac{(\hat{a}_2/\hat{a}_{20}) - 1}{\{1 - (\hat{a}_{40}/\hat{a}_{20}^2)/p\}^{1/2}},$$

where

$$\hat{a}_{20} = \frac{n}{p(n+2)}\sum_{i=1}^{p} s_{ii}^2,\ \hat{a}_{40} = \frac{1}{p}\sum_{i=1}^{p} s_{ii}^4,$$

and s_{ij} is the (i,j) element of the sample covariance matrix S. The asymptotic distribution of T_3 has been derived under both null and local alternative hypotheses under assumptions. Observe that our results of Theorem 2.4 allow to derive the asymptotic null distribution of T_3 under weaker assumptions. We relax

(A1)′ $p \to \infty$,

(A2)′ $n = O(p^\delta)$ $(0 < \delta \leq 1)$, and

$$\text{(A) } 0 < \lim_{k\to\infty} a_i < \infty, \ i = 1, 2, \ldots, 8.$$

to $(A1)'$, $(A2)'$,

$$\begin{aligned}
&\text{(A3)}' \ \sum_{i\neq j}^{p} \sigma_{ii}^2\sigma_{jj}^2 \asymp p^2, \\
&\text{(A4)}' \ a_4 = o(p).
\end{aligned}$$

where σ_{ij} denotes the (i, j) element of sample covariance matrix Σ. Now we see that under H_{03} the following holds

$$\begin{aligned}
\frac{nT}{\hat{\sigma}} &= \frac{n}{2}\frac{(\hat{a}_2/\hat{a}_{20}) - 1}{\{(n-1)(n+2)/n^2 - (n-1)/(n+2)(\hat{a}_{40}/\hat{a}_{20}^2)/p\}^{1/2}} \\
&= T_3 + o_p(1). \qquad (2.6)
\end{aligned}$$

Thus, from Theorem 2.4 and (2.4), the asymptotic normality of T_3 immediately follows under H_{03} and Assumptions (A1)′-(A4)′. We note that Assumptions (A3)′ and (A4)′ are weaker assumption than (A). Thus, (A1)′–(A4)′ is the special case of our asymptotic framework (A1)–(A4) when $p_i = 1, i = 1, \ldots, k$. Using (A1)′–(A4)′, Theorem 2.4 and (2.6), we observe that the asymptotic normality of T_3 immediately follows.

3 Proofs

For the full proofs of Theorem 2.1, 2.2 and Lemma 2.3, see supplementary material featured on the Statistica Neerlandica website.

3.1 Preliminary results

To prove Theorems 2.1 and 2.2 and Lemma 2.3, we need to evaluate some moments of products of multivariate normal vectors and some higher order moments of Wishart matrix. Our results for these moments are summarized below.

Lemma 3.1 (Moments of products of multivariate normal random vector)**.** *Let* $\boldsymbol{z} = (z_1, \cdots, z_p)'$ *be distributed as* $\mathcal{N}_p(\mathbf{0}, I_p)$ *and let* A_i *be an arbitrary* $p \times p$ *non-random matrix,* $i = 1, 2, 3$*. Then, the following assertions hold:*

(a) $\mathrm{E}[\boldsymbol{z}\boldsymbol{z}'] = I_p$,

(b) $\mathrm{E}[\boldsymbol{z}\boldsymbol{z}'A_1\boldsymbol{z}\boldsymbol{z}'] = \mathbb{A}_1 + \mathrm{tr}\, A_1 \cdot I_p$,

(c)
$$\mathrm{E}[\boldsymbol{z}\boldsymbol{z}'A_1\boldsymbol{z}\boldsymbol{z}'A_2\boldsymbol{z}\boldsymbol{z}'] = \mathbb{A}_1\mathbb{A}_2 + \mathbb{A}_2\mathbb{A}_1 + \mathrm{tr}\, A_1 \cdot \mathbb{A}_2 + \mathrm{tr}\, A_2 \cdot \mathbb{A}_1 + \{\mathrm{tr}\, A_1\mathbb{A}_2 + \mathrm{tr}\, A_1 \cdot \mathrm{tr}\, A_2\}I_p,$$

(d)
$$\begin{aligned}\mathrm{E}[\boldsymbol{z}\boldsymbol{z}'A_1\boldsymbol{z}\boldsymbol{z}'A_2\boldsymbol{z}\boldsymbol{z}'A_3\boldsymbol{z}\boldsymbol{z}'] = &\sum^3 \sum^3_{i\neq j\neq k\neq i} \sum^3 \mathbb{A}_i\mathbb{A}_j\mathbb{A}_k \\ &+ \mathrm{tr}\, A_1 \cdot \{\mathbb{A}_2\mathbb{A}_3 + \mathbb{A}_3\mathbb{A}_2\} \\ &+ \mathrm{tr}\, A_2 \cdot \{\mathbb{A}_3\mathbb{A}_1 + \mathbb{A}_1\mathbb{A}_3\} \\ &+ \mathrm{tr}\, A_3 \cdot \{\mathbb{A}_1\mathbb{A}_2 + \mathbb{A}_2\mathbb{A}_1\} \\ &+ \{\mathrm{tr}\, A_1\mathbb{A}_2 + \mathrm{tr}\, A_1 \cdot \mathrm{tr}\, A_2\}\mathbb{A}_3 \\ &+ \{\mathrm{tr}\, A_2\mathbb{A}_3 + \mathrm{tr}\, A_2 \cdot \mathrm{tr}\, A_3\}\mathbb{A}_1 \\ &+ \{\mathrm{tr}\, A_3\mathbb{A}_1 + \mathrm{tr}\, A_3 \cdot \mathrm{tr}\, A_1\}\mathbb{A}_2 \\ &+ \{\mathrm{tr}\, \mathbb{A}_1\mathbb{A}_2\mathbb{A}_3 + \mathrm{tr}\, A_1\mathbb{A}_2 \cdot \mathrm{tr}\, A_3 + \mathrm{tr}\, A_2\mathbb{A}_3 \cdot \mathrm{tr}\, A_1 \\ &\quad + \mathrm{tr}\, A_3\mathbb{A}_1 \cdot \mathrm{tr}\, A_2 + \mathrm{tr}\, A_1 \cdot \mathrm{tr}\, A_2 \cdot \mathrm{tr}\, A_3\}I_p,\end{aligned}$$

where $\mathbb{A}_i = A_i + A_i'$ *for* $i = 1, 2, 3$.

Proof. Throughout the proof, we use the following property: for $\nu \sim \mathcal{N}(0, 1)$,

$$\mathrm{E}[\nu^i] \;=\; (i-1)!!$$

if $i \geq 2$ is an even number, and $\mathrm{E}[\nu^i] = 0$ otherwise. Here,

$$x!! = \begin{cases} x(x-2)\cdots 3 \cdot 1 \text{ if } x \text{ is odd} \\ x(x-2)\cdots 4 \cdot 2 \text{ if } x \text{ is even.} \end{cases}$$

At first, we begin by evaluating $\mathrm{E}[B\boldsymbol{z}\boldsymbol{z}'C]$ where B and C are $p \times p$ arbitrary non-random matrices. Then for the (i,j)-th element of this matrix:

$$
\begin{aligned}
\mathrm{E}\left[(B\boldsymbol{z}\boldsymbol{z}'C)_{ij}\right] &= \mathrm{E}\left[\sum_{i_1,i_2}(B)_{ii_1}z_{i_1}z_{i_2}(C)_{i_2j}\right] \\
&= \sum_{i_1,i_2}(B)_{ii_1}(C)_{i_2j}\mathrm{E}[z_{i_1}z_{i_2}] \\
&= \sum_{i_1}(B)_{ii_1}(C)_{i_1j}\mathrm{E}(z_{i_1}^2) + \sum_{i_1,i_2,i_1\neq i_2}(B)_{ii_1}(C)_{i_2j}\mathrm{E}[z_{i_1}]\mathrm{E}[z_{i_2}] \\
&= \sum_{i_1}(B)_{ii_1}(C)_{i_1j},
\end{aligned}
$$

where $(X)_{ij}$ is the (i,j)-th element of matrix X from which it follows that $\mathrm{E}[B\boldsymbol{z}\boldsymbol{z}'C] = BC$.

Now by applying the same technique to the (i,j)-th element of $B\boldsymbol{z}\boldsymbol{z}'A_1\boldsymbol{z}\boldsymbol{z}'C$ in (b), we obtain

$$
\begin{aligned}
& \mathrm{E}[(B\boldsymbol{z}\boldsymbol{z}'A_1\boldsymbol{z}\boldsymbol{z}'C)_{ij}] \\
= \ & \mathrm{E}[\sum_{i_1,i_2,i_3,i_4}(B)_{ii_1}z_{i_1}z_{i_2}(A_1)_{i_2i_3}z_{i_3}z_{i_4}(C)_{i_4j}] \\
= \ & \sum_{i_1,i_2,i_3,i_4}(B)_{ii_1}(A_1)_{i_2i_3}(C)_{i_4j}\mathrm{E}[z_{i_1}z_{i_2}z_{i_3}z_{i_4}] \\
= \ & 3\sum_{i_1}(B)_{ii_1}(A_1)_{i_1i_1}(C)_{i_1j} \\
& + \sum_{i_1,i_2,i_1\neq i_2}\left[(B)_{ii_1}\left\{(A_1)_{i_1i_2}+(A_1)_{i_2i_1}\right\}(C)_{i_2j}+(A_1)_{i_2i_2}(B)_{ii_1}(C)_{i_1j}\right] \\
= \ & \sum_{i_1,i_2}\left[(B)_{ii_1}\left\{(A_1)_{i_1i_2}+(A_1)_{i_2i_1}\right\}(C)_{i_2j}+(A_1)_{i_2i_2}(B)_{ii_1}(C)_{i_1j}\right],
\end{aligned}
$$

which implies that $\mathrm{E}[B\boldsymbol{z}\boldsymbol{z}'A_1\boldsymbol{z}\boldsymbol{z}'C] = B\mathbb{A}_1C + \mathrm{tr}\,A_1 \cdot BC$.

Proceeding in the same way for (c) and (d) and setting $B = C = I_p$ completes the proof. □

By using Lemma 3.1, we provide results on the higher order moments of Wishart matrix.

Lemma 3.2 (Higher moments of Wishart matrix)**.** *Let* $W \sim \mathcal{W}_p(n,\Sigma)$*, where* $\mathcal{W}_p(n,\Sigma)$ *denotes Wishart distribution with* n *degrees of freedom and the scale parameter* Σ*. Let* G*,* H *be any non-random* $p \times p$ *symmetric matrices. Then the following assertions hold:*

$$
\begin{aligned}
\text{(i) } \mathrm{E}[\mathrm{tr}(GW)^2] &= n(n+1)\,\mathrm{tr}(G\Sigma)^2 + n(\mathrm{tr}\, G\Sigma)^2,\\
\text{(ii) } \mathrm{E}[(\mathrm{tr}\, GW)^2] &= n^2(\mathrm{tr}\,\Sigma G)^2 + 2n\,\mathrm{tr}(\Sigma G)^2,\\
\text{(iii) } \mathrm{E}[\mathrm{tr}(GW)^2\,\mathrm{tr}(HW)^2] &= n^2(n+1)^2\,\mathrm{tr}(\Sigma G)^2\,\mathrm{tr}(\Sigma H)^2\\
&\quad + n^2(n+1)\,\mathrm{tr}(\Sigma G)^2(\mathrm{tr}\,\Sigma H)^2\\
&\quad + n^2(n+1)(\mathrm{tr}\,\Sigma G)^2\,\mathrm{tr}(\Sigma H)^2\\
&\quad + n^2(\mathrm{tr}\,\Sigma G)^2(\mathrm{tr}\,\Sigma H)^2\\
&\quad + 8n(n+1)^2\,\mathrm{tr}(\Sigma G)^2(\Sigma H)^2\\
&\quad + 4n(n+1)(\mathrm{tr}\,\Sigma G\Sigma H)^2\\
&\quad + 8n(n+1)\,\mathrm{tr}\,\Sigma H\,\mathrm{tr}(\Sigma G)^2\Sigma H\\
&\quad + 8n(n+1)\,\mathrm{tr}\,\Sigma G\,\mathrm{tr}\,\Sigma G(\Sigma H)^2\\
&\quad + 8n\,\mathrm{tr}\,\Sigma G\,\mathrm{tr}\,\Sigma H\,\mathrm{tr}\,\Sigma G\Sigma H\\
&\quad + 4n(n+3)\,\mathrm{tr}(\Sigma G\Sigma H)^2,\\
\text{(iv) } \mathrm{E}[(\mathrm{tr}\, GW)^2\,\mathrm{tr}(HW)^2] &= n^3(\mathrm{tr}\,\Sigma G)^2(\mathrm{tr}\,\Sigma H)^2\\
&\quad + (n+1)n^3(\mathrm{tr}\,\Sigma G)^2\,\mathrm{tr}(\Sigma H)^2\\
&\quad + 2n^2\,\mathrm{tr}(\Sigma G)^2(\mathrm{tr}\,\Sigma H)^2\\
&\quad + 8n^2\,\mathrm{tr}\,\Sigma G\,\mathrm{tr}\,\Sigma H\,\mathrm{tr}\,\Sigma G\Sigma H\\
&\quad + 2(n+1)n^2\,\mathrm{tr}(\Sigma G)^2\,\mathrm{tr}(\Sigma H)^2\\
&\quad + 8(n+1)n^2\,\mathrm{tr}\,\Sigma G\,\mathrm{tr}\,\Sigma G(\Sigma H)^2\\
&\quad + 8n(\mathrm{tr}\,\Sigma G\Sigma H)^2
\end{aligned}
$$

$$
\begin{aligned}
&+16n \operatorname{tr}\Sigma H \operatorname{tr}(\Sigma G)^2\Sigma H\\
&+16(n+1)n \operatorname{tr}(\Sigma G)^2(\Sigma H)^2\\
&+8(n+1)n \operatorname{tr}(\Sigma G\Sigma H)^2,
\end{aligned}
$$

$$
\begin{aligned}
\text{(v) } \mathrm{E}[(\operatorname{tr} GW)^2(\operatorname{tr} HW)^2] \;=\; & n^4(\operatorname{tr}\Sigma G)^2(\operatorname{tr}\Sigma H)^2\\
&+2n^3 \operatorname{tr}(\Sigma G)^2(\operatorname{tr}\Sigma H)^2\\
&+8n^3 \operatorname{tr}\Sigma G \operatorname{tr}\Sigma H \operatorname{tr}\Sigma G\Sigma H\\
&+2n^3(\operatorname{tr}\Sigma G)^2 \operatorname{tr}(\Sigma H)^2\\
&+8n^2(\operatorname{tr}\Sigma G\Sigma H)^2\\
&+4n^2 \operatorname{tr}(\Sigma G)^2 \operatorname{tr}(\Sigma H)^2\\
&+16n^2 \operatorname{tr}\Sigma H \operatorname{tr}(\Sigma G)^2\Sigma H\\
&+16n^2 \operatorname{tr}\Sigma G \operatorname{tr}\Sigma G(\Sigma H)^2\\
&+32n \operatorname{tr}(\Sigma G)^2(\Sigma H)^2\\
&+16n \operatorname{tr}(\Sigma G\Sigma H)^2.
\end{aligned}
$$

PROOF. Let $\mathbb{U} = (\boldsymbol{u}_1\ \boldsymbol{u}_2\ \ldots\ \boldsymbol{u}_n)$, where $\boldsymbol{u}_i$ are iid $\mathcal{N}_p(\mathbf{0}, I_p)$. By re-writing $W = \Sigma^{1/2}\mathbb{U}\mathbb{U}'\Sigma^{1/2}$, we obtain

$$
\begin{aligned}
\mathrm{E}[(GW)^2] &= \mathrm{E}[G\Sigma^{1/2}\mathbb{U}\mathbb{U}'\Sigma^{1/2}G\Sigma^{1/2}\mathbb{U}\mathbb{U}'\Sigma^{1/2}]\\
&= \mathrm{E}\left[G\Sigma^{1/2}\left(\sum_{i=1}^{n}\boldsymbol{u}_i\boldsymbol{u}_i'\right)\Sigma^{1/2}G\Sigma^{1/2}\left(\sum_{i=1}^{n}\boldsymbol{u}_i\boldsymbol{u}_i'\right)\Sigma^{1/2}\right]\\
&= G\Sigma^{1/2}\left(\sum_{i=1}^{n}\mathrm{E}\left[\boldsymbol{u}_i\boldsymbol{u}_i'\Sigma^{1/2}G\Sigma^{1/2}\boldsymbol{u}_i\boldsymbol{u}_i'\right]\right.\\
&\quad\left.+\sum_{i\neq j}^{n}\mathrm{E}\left[\boldsymbol{u}_i\boldsymbol{u}_i'\right]\Sigma^{1/2}G\Sigma^{1/2}\,\mathrm{E}\left[\boldsymbol{u}_j\boldsymbol{u}_j'\right]\right)\Sigma^{1/2}. \qquad (3.1)
\end{aligned}
$$

By applying the technique derived in Lemma 3.1 to the moments $\mathrm{E}\left[\boldsymbol{u}_i\boldsymbol{u}_i'\Sigma^{1/2}G\Sigma^{1/2}\boldsymbol{u}_i\boldsymbol{u}_i'\right]$ and $\mathrm{E}\left[\boldsymbol{u}_i\boldsymbol{u}_i'\right]$ in (3.1), we obtain

$$
\mathrm{E}\left[\boldsymbol{u}_i\boldsymbol{u}_i'\Sigma^{1/2}G\Sigma^{1/2}\boldsymbol{u}_i\boldsymbol{u}_i'\right] = 2\Sigma^{1/2}G\Sigma^{1/2} + (\operatorname{tr}\Sigma^{1/2}G\Sigma^{1/2})I_p, \qquad (3.2)
$$

$$
\mathrm{E}\left[\boldsymbol{u}_i\boldsymbol{u}_i'\right] = I_p. \qquad (3.3)
$$

By inserting (3.2) and (3.3) into (3.1), we have

$$\mathrm{E}[(GW)^2] \quad = \quad n(n+1)(G\Sigma)^2 + n(\operatorname{tr} G\Sigma)G\Sigma,$$

which gives claim (i).

To show (ii), we consider

$$\begin{aligned}
\mathrm{E}[(\operatorname{tr} GW)^2] &= \mathrm{E}[(\operatorname{tr} G\Sigma^{1/2}\mathbb{U}\mathbb{U}'\Sigma^{1/2})^2] \\
&= \mathrm{E}\left[\left(\sum_{i=1}^{n} \operatorname{tr} G\Sigma^{1/2}\boldsymbol{u}_i\boldsymbol{u}_i'\Sigma^{1/2}\right)^2\right] \\
&= \sum_{i=1}^{n} \mathrm{E}\left[\left(\operatorname{tr} G\Sigma^{1/2}\boldsymbol{u}_i\boldsymbol{u}_i'\Sigma^{1/2}\right)^2\right] \\
&\quad + \sum_{i\neq j}^{n} \mathrm{E}\left[\operatorname{tr} G\Sigma^{1/2}\boldsymbol{u}_i\boldsymbol{u}_i'\Sigma^{1/2}\right] \mathrm{E}\left[\operatorname{tr} G\Sigma^{1/2}\boldsymbol{u}_j\boldsymbol{u}_j'\Sigma^{1/2}\right]. \qquad (3.4)
\end{aligned}$$

In the same way as in (i), we use (a) and (b), Lemma 3.1 which yields

$$\begin{aligned}
\mathrm{E}\left[\left(\operatorname{tr} G\Sigma^{1/2}\boldsymbol{u}_i\boldsymbol{u}_i'\Sigma^{1/2}\right)^2\right] &= \mathrm{E}[\operatorname{tr}(\boldsymbol{u}_i\boldsymbol{u}_i'\Sigma^{1/2}G\Sigma^{1/2})^2] \\
&= 2\operatorname{tr}(\Sigma G)^2 + (\operatorname{tr}\Sigma G)^2, \qquad (3.5) \\
\mathrm{E}\left[\operatorname{tr} G\Sigma^{1/2}\boldsymbol{u}_i\boldsymbol{u}_i'\Sigma^{1/2}\right] &= \operatorname{tr} G\Sigma. \qquad (3.6)
\end{aligned}$$

By inserting (3.5) and (3.6) into (3.4), we have

$$\mathrm{E}[(\operatorname{tr} GW)^2] \quad = \quad n^2(\operatorname{tr}\Sigma G)^2 + 2n\operatorname{tr}(\Sigma G)^2,$$

which gives claim (ii). Proceeding in the same way , with applying correspondent results of Lemma 3.1, we obtain claims (iii)-(v). □

3.2 Proof of Theorem 2.1

We prove consistency and unbiasedness of the suggested estimator T.

We begin by proving unbiasedness of T. Using Lemma 3.2 with $G = H = I_p$, we get $\mathrm{E}[\hat{a}_2] = a_2$. Denote $\mathcal{P}_i = \operatorname{diag}(\mathbf{0}'_{p_1}, \cdots, \mathbf{0}'_{p_{i-1}}, \mathbf{1}'_{p_i}, \mathbf{0}'_{p_{i+1}}, \cdots, \mathbf{0}'_{p_k})$, and re-write $\hat{a}_2^{(i)}$ as

$$\hat{a}_2^{(i)} = \frac{n^2}{(n-1)(n+2)}\left\{\operatorname{tr}(\mathcal{P}_i S\mathcal{P}_i)^2 - \frac{1}{n}(\operatorname{tr}(\mathcal{P}_i S\mathcal{P}_i))^2\right\}.$$

Then by using Lemma 3.2 and setting $\mathcal{P}_i = G = H$, we get $\mathrm{E}[\hat{a}_2^{(i)}] = a_2^{(i)}$ and $\mathrm{E}[\hat{b}_2] = b_2$. This proves unbiasedness of T.

Next, we prove consistency of T under Assumptions (A1)-(A5). From Chebychev's inequality, for any $\varepsilon > 0$

$$\Pr\left(|T - \|\Sigma_d - \Sigma\|^2| \geq \varepsilon\right) \leq \frac{\sigma_T^2}{\varepsilon^2},$$

where $\sigma_T^2 = \mathrm{Var}[T]$. σ_T^2 can be expressed as

$$\begin{aligned}\sigma_T^2 &= \mathrm{Var}[\hat{a}_2] + \mathrm{Var}[\hat{b}_2] - 2\mathrm{Cov}(\hat{b}_2, \hat{a}_2) \\ &\leq \mathrm{Var}[\hat{a}_2] + \mathrm{Var}[\hat{b}_2] + 2\sqrt{\mathrm{Var}[\hat{a}_2]\,\mathrm{Var}[\hat{b}_2]},\end{aligned}$$

and it is now sufficient to show that

$$\mathrm{Var}[\hat{a}_2] \to 0, \ \ \mathrm{Var}[\hat{b}_2] \to 0$$

under Assumptions (A1)-(A5). By Lemma 3.2, it follows that

$$\begin{aligned}\mathrm{Var}[\hat{a}_2] &= \frac{4a_2^2}{(n-1)(n+2)} + \frac{4\,(2n^2+3n-6)}{n(n-1)(n+2)}\frac{a_4}{p} \\ &= O\left(n^{-1}\right), \\ \mathrm{Var}[\hat{b}_2] &= \frac{4}{(n-1)(n+2)}\frac{\sum_{i=1}^k a_2^{(i)2}}{p^2} + \frac{4\,(2n^2+3n-6)}{n(n-1)(n+2)}\frac{\sum_{i=1}^k a_4^{(i)}}{p^2} \\ &\quad + \frac{8}{n}\frac{\sum_{i\neq j}^k \mathrm{tr}\,\Sigma_{ji}\Sigma_{ii}\Sigma_{ij}\Sigma_{jj}}{p^2} + \frac{4}{(n+2)(n-1)}\frac{\sum_{i\neq j}^k(\mathrm{tr}\,\Sigma_{ij}\Sigma_{ji})^2}{p^2} \\ &\quad + \frac{4(n-2)}{n(n+2)(n-1)}\frac{\sum_{i\neq j}^k \mathrm{tr}(\Sigma_{ij}\Sigma_{ji})^2}{p^2} \\ &= O(n^{-1}).\end{aligned}$$

By combining the last two results, gives $\sigma_T^2 \to 0$ under Assumptions (A1)-(A5). This proves the consistency claim. □

3.3 Proof of Theorem 2.2

Let $\Gamma = \mathrm{diag}(\Gamma_1, \cdots, \Gamma_k)$ such that $\Gamma_i\Sigma_{ii}\Gamma_i' = \Lambda_i$, $\Gamma_i\Gamma_i' = I_{p_i}$ $(i = 1, 2, \cdots, k)$, $\Lambda_i = \mathrm{diag}(\lambda_1^{(i)}, \lambda_2^{(i)}, \cdots, \lambda_{p_i}^{(i)})$. Under H_0, $\Gamma(nS)\Gamma' \sim \mathcal{W}_p(n, \mathrm{diag}(\Lambda_1, \cdots, \Lambda_k))$. Then we can write

$$\Gamma(nS)\Gamma'$$
$$= \mathrm{diag}(\Lambda_1^{1/2},\cdots,\Lambda_k^{1/2}) \begin{pmatrix} \boldsymbol{u}_1^{(1)} \cdots \boldsymbol{u}_n^{(1)} \\ \boldsymbol{u}_1^{(2)} \cdots \boldsymbol{u}_n^{(2)} \\ \vdots \\ \boldsymbol{u}_1^{(k)} \cdots \boldsymbol{u}_n^{(k)} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_1^{(1)'} \cdots \boldsymbol{u}_1^{(k)'} \\ \boldsymbol{u}_2^{(1)'} \cdots \boldsymbol{u}_2^{(k)'} \\ \vdots \\ \boldsymbol{u}_n^{(1)'} \cdots \boldsymbol{u}_n^{(k)'} \end{pmatrix} \mathrm{diag}(\Lambda_1^{1/2},\cdots,\Lambda_k^{1/2}),$$

where $\boldsymbol{u}_i^{(j)} \sim \mathcal{N}_{p_j}(\mathbf{0}, I_{p_j})$. Denote $U_i = (\boldsymbol{u}_1^{(i)} \ \cdots \ \boldsymbol{u}_n^{(i)})'$. Then the statistic nT can be expressed as

$$nT \ = \ \sum_{j=1}^{k} \varepsilon_j + o_p(1),$$

where

$$\begin{aligned} \varepsilon_j &= \sum_{i=0}^{j-1} \eta_{ij}, \\ \eta_{ij} &= \frac{2}{n\sqrt{np}} \Big\{ \mathrm{tr}(\Lambda_i^{1/2} U_i' U_j \Lambda_j^{1/2})(\Lambda_j^{1/2} U_j' U_i \Lambda_i^{1/2}) \\ &\quad -\frac{1}{n} (\mathrm{tr}\,\Lambda_i^{1/2} U_i' U_i \Lambda_i^{1/2})(\mathrm{tr}\,\Lambda_j^{1/2} U_j' U_j \Lambda_j^{1/2}) \Big\}. \end{aligned}$$

Here, $\Lambda_0 = O$.

Now we prove the normality of T under H_0 by applying the martingale central limit theorem to $\sum_{j=1}^k \varepsilon_j$. Let $\mathcal{F}_j$ be the σ-algebra generated by the sequence of random matrices $U_0, \cdots, U_j$. Then by letting $U_0 = O$, and $\mathcal{F}_0 = (\emptyset, \Omega)$, where $\emptyset$ is the empty set and Ω is the entire space, we find $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_\infty$. Also, we have

$$\mathrm{E}\left[\varepsilon_j | \mathcal{F}_{j-1}\right] = \mathrm{E}\left[\varepsilon_j | U_0, \ldots, U_{j-1}\right] = 0, \ (j = 1, 2, \cdots).$$

By applying Lemma 3.1 to $\mathrm{E}[\varepsilon_j^2]$, we obtain

$$0 < \sigma^2 = \lim_{k\to\infty} \sum_{j=1}^{k} \mathrm{E}[\varepsilon_j^2] = \lim_{k\to\infty} \sum_{j=2}^{k} \sum_{i=1}^{j-1} \frac{8(n-1)(n+2) a_2^{(i)} a_2^{(j)}}{n^2 p^2} < \infty$$

under Assumptions (A1)-(A4). To apply the theorem, we need to check the following two conditions:

$$\text{(I)} \qquad \sum_{j=1}^{k} \mathrm{E}[\varepsilon_j^2 | \mathcal{F}_{j-1}] \xrightarrow{P} \sigma^2 \text{ as } k \to \infty,$$

$$\text{(II)} \qquad \sum_{j=1}^{k} \mathrm{E}[\varepsilon_j^2 \mathrm{I}(|\varepsilon_j| > \delta) | \mathcal{F}_{j-1}] \xrightarrow{P} 0 \text{ as } k \to \infty \text{ for any } \delta > 0.$$

At first, check (I). Now we can rewrite

$$\mathrm{E}\left[\varepsilon_j^2 | \mathcal{F}_{j-1}\right] - \frac{8(n-1)(n+2)a_2^{(i)}a_2^{(j)}}{n^2p^2} = \frac{8 \operatorname{tr} \Lambda_j^2}{n^2p^2} \sum_{i=0}^{j-1} y_i^{(1)} + \frac{8 \operatorname{tr} \Lambda_j^2}{n^2p^2} \sum_{i \neq i'}^{j-1} y_{ii'}^{(2)}, \qquad (3.7)$$

where

$$\begin{aligned} y_i^{(1)} &= \operatorname{tr}(U_i\Lambda_iU_i')^2 - \frac{1}{n}(\operatorname{tr} U_i\Lambda_iU_i')^2 - (n-1)(n+2) \operatorname{tr} \Lambda_i^2, \\ y_{ii'}^{(2)} &= \left\{ \operatorname{tr} U_i\Lambda_iU_i'U_{i'}\Lambda_{i'}U_{i'}' - \frac{1}{n} \operatorname{tr} U_i\Lambda_iU_i' \operatorname{tr} U_{i'}\Lambda_{i'}U_{i'}' \right\}. \end{aligned}$$

Here, $\Lambda_0 = O$. By using (3.7) and Cauchy-Schwarz inequality, we obtain that

$$\mathrm{E}\left[\left(\sum_{j=1}^{k} \mathrm{E}\left[\varepsilon_j^2 | \mathcal{F}_{j-1}\right] - \sigma^2 \right)^2 \right] = \mathrm{E}\left[(v_1 + v_2)^2\right] \leq \sum_{i=1}^{2} \mathrm{E}[v_i^2] + 2\sqrt{\mathrm{E}[v_1^2]\,\mathrm{E}[v_2^2]},$$

where

$$v_1 = \sum_{j=2}^{k} \frac{8 \operatorname{tr} \Lambda_j^2}{n^3p} \sum_{i=1}^{j-1} y_i^{(1)}, \;\; v_2 = \sum_{j=3}^{k} \frac{16 \operatorname{tr} \Lambda_j^2}{n^3p} \sum_{i=2}^{j-1} \sum_{i'=1}^{i-1} y_{ii'}^{(2)}.$$

Under H_0 and Assumptions (A1)-(A4), it holds that $\mathrm{E}[v_1^2] \to 0$ and $\mathrm{E}[v_2^2] \to 0$, which completes the proof of (I). (See details of evaluation of $\mathrm{E}[v_1^2]$ and $\mathrm{E}[v_2^2]$ in supplementary material featured on the Statistica Neerlandica website.)

Next, we check (II), for which it is sufficient to show that

$$\sum_{j=1}^{k} \mathrm{E}[\varepsilon_j^4] \to 0$$

for the proof of (II). We first note that

$$\sum_{j=1}^{k} \mathrm{E}[\varepsilon_j^4] \;=\; \sum_{j=2}^{k} \left(\sum_{i=1}^{j-1} \mathrm{E}[\eta_{ij}^4] + \sum_{i_1 \neq i_2} \mathrm{E}[\eta_{i_1j}^2 \eta_{i_2j}^2] \right).$$

From the evaluation of $\mathrm{E}[\eta_{i_1j}^2 \eta_{i_2j}^2]$ and $\mathrm{E}[\eta_{ij}^4]$, we get

$$\sum_{j=1}^{k} \mathrm{E}[\varepsilon_j^4] \;=\; o(1),$$

which completes the proof of (II). (See details of evaluation of $\sum_{j=1}^{k} \mathrm{E}[\varepsilon_j^4]$ in supplementary material featured on the Statistica Neerlandica website.)

By combining conditions (I) and (II), the normal convergence of $\sum_{j=1}^{k} \varepsilon_j$ follows, which in turn provides the normal convergence of nT. $\square$

3.4 Proof of Lemma 2.3

We note that

$$\hat{\sigma}^2 - \sigma^2 = \frac{4}{p^2} \sum_{j=2}^{k} \sum_{i=1}^{j-1} (\hat{a}_2^{(i)} \hat{a}_2^{(j)} - a_2^{(i)} a_2^{(j)}) + o_p(1).$$

By using Lemma 3.2, under H_0, we obtain

$$\mathrm{E}[(\hat{\sigma}^2 - \sigma^2)^2] \to 0. \tag{3.8}$$

(See details of evaluation of $\mathrm{E}[(\hat{\sigma}^2 - \sigma^2)^2]$ in supplementary material featured on the Statistica Neerlandica website.) From (3.8) and Chebyshev's inequality, we obtain that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ under H_0 and Assumptions (A1)-(A4). $\square$

4 Simulation study

We now turn to the numerical evaluation of the properties of the new test procedure. The goal of the simulation study is threefold: to investigate the effect of large p on the size of the test, to investigate power of the test under certain alternative hypotheses, and to evaluate the effect of the block size, p_i and covariance parameter on the test accuracy.

Monte Carlo simulation are used to find the size and power of T for sample size fixed to $n = 100$ and the dimensionality $p = cn$ where $c = 1, \dots, 4$. An assortment of p_i's is considered for each p, which qualitatively represents "small-", "large-" and "mix-"size blocks cases.

"small-" and "large-" size blocks cases, i.e., all the Σ_{ii}'s (p/p_i blocks) have equal order $p_i = 5, 10$ or 25, respectively.

"mix-" size blocks cases, i.e., two types of order are included in $\Sigma_{11}, \dots, \Sigma_{kk}$: $p/(2p_i)$ blocks have equal order $p_i = 5$ or 10, and $p/(2p_j)$ block(s) have equal order $p_j = 25$ or 50.

We test with the nominal level $\alpha = 0.05$ and reject H_0 when the test statistics exceeds $(\hat{\sigma}/n)z_\alpha$, where z_α is the upper $100\alpha\%$ critical point of the standard normal distribution. All the results reported are based on $r = 10,000$ Monte Carlo simulations.

To evaluate the size of the proposed test we calculate the empirical Type I error or empirical power as follows: draw a sample of N independent observations from p-dimensional normal distribution under the null hypothesis or alternative hypotheses. Replicate this r times, and using T from (2.4) calculate

$$\frac{\#\left(T > (\hat{\sigma}/n)z_\alpha\right)}{r},$$

where $\hat{\sigma}$ is defined in (2.5).

The simulation study analyzes the empirical Type I error of T under the following within-block structures: $\Sigma = \text{diag}(\Sigma_{11}, \dots, \Sigma_{kk})$, $\Sigma_{ii} = \Sigma_i(\omega^2, \rho_0)$ has an intraclass structure, i.e.,

$$\Sigma_i(\omega^2, \rho_0) = \omega^2[(1 - \rho_0)I_{p_i} + \rho_0 \mathbf{1}_{p_i}\mathbf{1}'_{p_i}],$$

and a special case for $\Sigma_i(\omega^2, \rho_0)$ with $\omega^2 = 1$ and $\rho_0 = 0$ turns out to be an identity matrix. Further, ρ_0's are chosen so that Σ_{ii} is positive definite. On the other hand, the empirical power of T is similarly calculated for the block structures $\Sigma = \text{diag}(\Sigma_{11}, \dots, \Sigma_{kk})$ with some contamination.

4.1 The empirical Type I error rate

Tables 1 and 2 present the empirical type I error rate with "small-" and "large-" size blocks case and the same with "mix-" size blocks cases, respectively. The parameters for the covariance matrix are set as $\omega^2 = 1.0, 5.0$ and $\rho_0 = 0.1, 0.5, 0.9$. Although we note that most cases of the empirical Type I error rate are close to 0.05, unfortunately, in the few cases in Tables 1 and 2, we observe that the empirical type I error rate is slightly higher than the desired level of 0.05. This can be explained by the sensitivity of the test to the choice of model parameter: the empirical type I error rate becomes higher for large values of ρ_0, and this effect is most pronounced for blocks of both sizes with large p_i or p_j, see Tables 1 and 2. However, we also observe that the empirical type I error rate becomes close to the significance level when p grows. Further it can be observed that ω^2 does not depend on the empirical type I error rate for both size blocks cases.

Please insert Tables 1 and 2 around here.

4.2 The empirical power

A number of simulations are performed to analyze power of the test under certain alternative structures of Σ. We use the notations of (2.1) and consider the following local alternatives:

Σ has ν-contaminated block structure, i.e.,

$$H_{A_1}: \quad \Sigma_{ii} = I_{p_i},\, \Sigma_{ij} = [\omega_{\ell m}],\, \omega_{\ell m} = \rho_1 \neq 0$$

for ν pairs of (ℓ, m) and otherwise $\omega_{\ell m} = 0$ where $\ell = 1, \dots, p_i$, $m = 1, \dots, p_j$, $i, j = 1, \dots, k$, and locations of ρ_1 are chosen at random under the constraint that Σ is symmetric.

Σ has p_i-banded structure, i.e.,

$$H_{A_2}: \Sigma = \Sigma(\rho_0, \rho_1)$$

where the elements of $\Sigma(\rho_0, \rho_1)$ are defined as follows: $|\ell - m| > p_i$ implies that $\omega_{ij} = 0$, the within-band structure is designed so that $p_i \times p_i$ blocks are represented by intraclass correlation model with $\omega^2 = 1$, i.e., $(1-\rho_0)I_{p_i} + \rho_0 \mathbf{1}_{p_i}\mathbf{1}'_{p_i}$, and the remaining off-blocks elements are set to ρ_1.

H_{A_1} is designed to challenge the test procedure for some near block-diagonal structures with sparsely distributed non-zero off-blocks entries, whereas H_{A_2} represents a dense alternative.

Under the ν-contaminated block structure for $\nu = 50, 100, 150, 200, 250, 300$ and $\rho_1 = 0, 0.02, 0.04, 0.06, 0.08, 0.1$, all the simulation results in this simulation study are summarized in Figures 1–5 in order to investigate the empirical power.

For a fixed dimensionality $p = 100$, Figures 1 and 2 shows the power in each $p_i = 5, 10, 25$. ν is set as 100 and 300 in Figures 1 and 2, respectively. In the both figures, it could be observed that the power for the large size blocks was slightly higher than the same for the small size blocks. The power turns out to be higher when ν, i.e., the number of the pairs of contamination grows. Further, Figure 3 describes the relation between the empirical power and ρ_1, and it implies that the larger ρ_1 makes the power higher. The effect is obviously observed for the larger ν.

Please insert Figures 1–3 around here.

The simulation results for a fixed $p_i = 10$ and $p = 100, 200, 300, 400$ are described in Figures 4 and 5. ν is set as 100 and 300 in Figures 4 and 5, respectively. Similarly to Figures 1 and 2, the higher power is also observed when $p = 100$ and ρ_1 grows. This effect is most pronounced for $\nu = 300$, refer to Figure 5.

Please insert Figures 4 and 5 around here.

Next, we investigate the power of T for p_i-banded structure. The parameters in covariance structure are set as $\rho_0 = 0, 0.1, 0.3, 0.5, 0.7$ and $\rho_1 = 0, 0.02, 0.04, 0.06, 0.08$ except for some cases such that Σ is not positive definite. All the simulation results we conducted are summarized in the rest of the figures in order to analyze the power of T for p_i-banded structure.

Figure 6 shows the empirical power of T for a fixed $\rho_0 = 0.1$, $p_i = 5, 10, 25$ and $p = 100$. It implies that the large p_i increases the number of ρ_1 in p_i-banded structure and the empirical power grows. This effect could be also observed for $p = 200, 300, 400$.

Please insert Figure 6 around here.

Further, Figures 7 and 8 and Figures 9 and 10 are drawn in order to investigate the relation between ρ_0 and the power of T in each $p_i = 5$ and $p_i = 10$. In two sets of the two figures, ρ_0 is set as 0.1 and 0.5, respectively. It could be observed that the larger ρ_0 makes the power of T lower in Figures 7 and 8 and Figures 9 and 10. By setting the horizontal axis as ρ_0 and plotting the empirical power for $\rho_0 = 0, 0.1, 0.3, 0.5, 0.7$, the effect is obviously observed and is most pronounced for the large ρ_1 in Figure 11 which shows the empirical power of T for $p = 100$ and $p_i = 10$. It seems that the reason why the procedure does not work well for large ρ_0 is related to Assumptions (A3)-(A5). In fact, we observed that a_4 obviously grows as ρ_0 is larger (For example, a_4/k equals 0.316135 for $p = 100,\ p_i = 10,\ \rho_0 = 0.1$ and a_4/k equals 10.021735 for $p = 100,\ p_i = 10,\ \rho_0 = 0.5$). Further, it could be also observed that the effects of value of p are little under the fixed p_i's.

Please insert Figures 7–11 around here.

5 Conclusion

We have proposed a new test statistic for testing hypothesis that the covariance matrix has a block-diagonal structure. We have also presented higher order moments of multivariate normal random vector which are needed for the derivation of the proposed test statistic. Simulation results indicate that the proposed test statistic has good performance in a sense of detecting deviation from block-diagonal covariance structure. In conclusion, the test can be recommended when p is much larger than n and when a small deviation from H_0 is suspected.

Finally, we address the future works related to the proposed test in this paper. Unfortunately, we also observed that one of the assumptions stated in (A5) was not realistic under the strong correlations in block-diagonal structure. Therefore, the assumption should be relaxed in order to modify the testing procedure. Further, we derived only the null distribution of the test statistic in this paper. Thus, the discussion for the unbiasedness of the test statistic should be also noted as one of the important problems.

Acknowledgments

We thank two reviewers and Associate Editor for useful comments that substantially improved this paper. We are greatly indebted to Professor Takashi Seo, Tokyo University of Science, for helpful conversations and valuable comments. Tatjana Pavlenko is supported by the grant 2013-45266 of National Research Council of Sweden (VR).

References

Bai, Z. and Saranadasa, H. (1992). Effect od high dimension by an example of a two sample problem. *Statistica Sinica*, **6**, 311–329.

Danaher, P., Wang, P. and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B*, **76**, 373–397.

Fisher, T. J. (2012). On testing for an identity covariance matrix when the dimensionality equals or exceeds the sample size. *Journal of Statistical Planning and Inference*, **142**, 312–326.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when dimension is large compared to the sample size. *The Annals of Statistics*, **30**, 1081–1102.

Pavlenko, T., Bjorkstrom, A. and Tillander, A. (2012). Covariance structure approximation via gLasso in high-dimensional supervised classification. *Journal of Applied Statistics*, **39**, 1643–1666.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, **92**, 951–956.

Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of Japanese Statistical Society*, **35**, 251–272.

Srivastava, M. S., Kollo, T. and Rosen, D. (2011). Some tests for the covariance matrix with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis*, **102**, 1090–1103.

Srivastava, M. S. and Reid, N. (2012). Testing the structure of the covariance matrix with fewer observations than the dimension. *Journal of Multivariate Analysis*, **112**, 156–171.

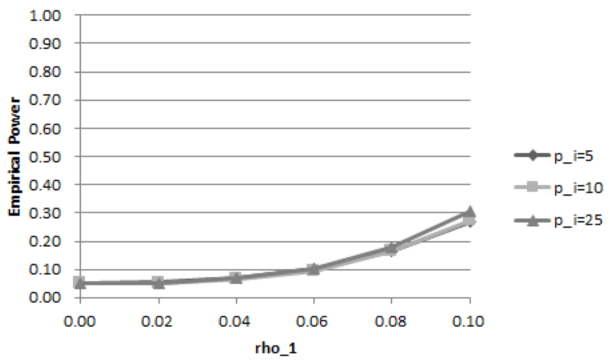

Figure 1: The empirical power of T for $p = 100$ and $\nu = 100$ under ν-contaminated block structure

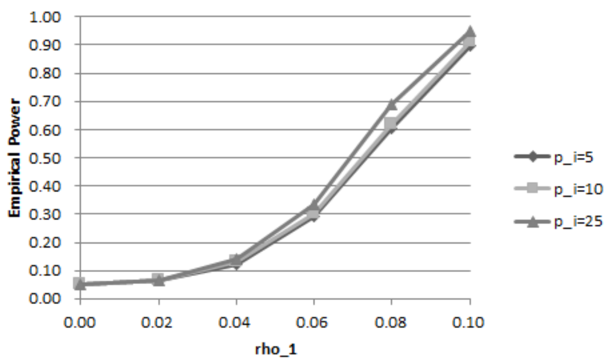

Figure 2: The empirical power of T for $p = 100$ and $\nu = 300$ under ν-contaminated block structure

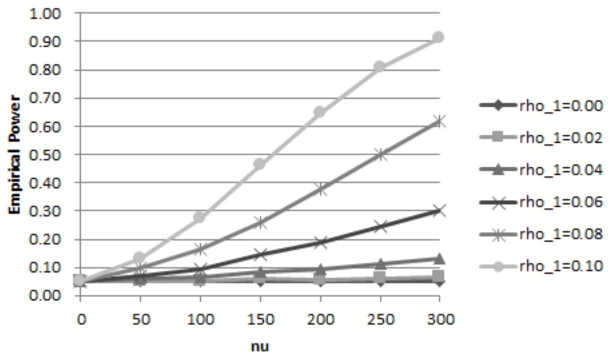

Figure 3: The empirical power of T for $p = 100$ and $p_i = 10$ under ν-contaminated block structure

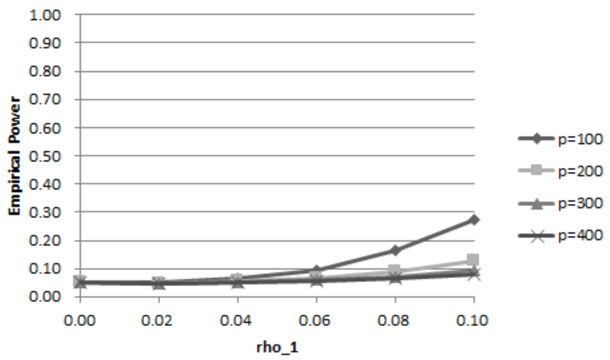

Figure 4: The empirical power of T for $p_i = 10$ and $\nu = 100$ under ν-contaminated block structure

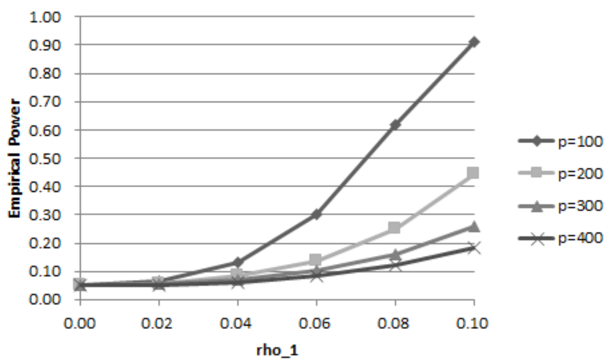

Figure 5: The empirical power of T for $p_i = 10$ and $\nu = 300$ under ν-contaminated block structure

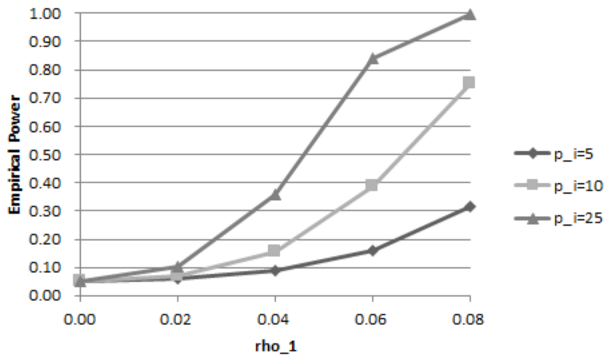

Figure 6: The empirical power of T for $p = 100$ and $\rho_0 = 0.1$ under p_i-banded block structure

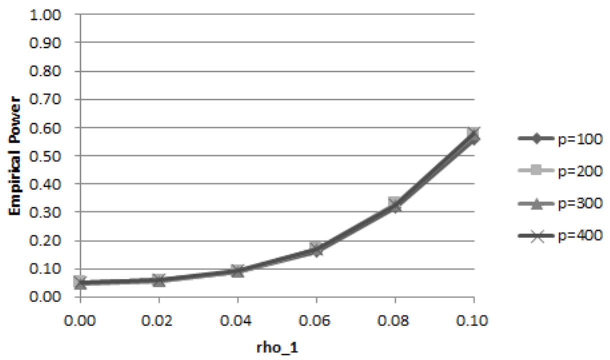

Figure 7: The empirical power of T for $p_i = 5$ and $\rho_0 = 0.1$ under p_i-banded block structure

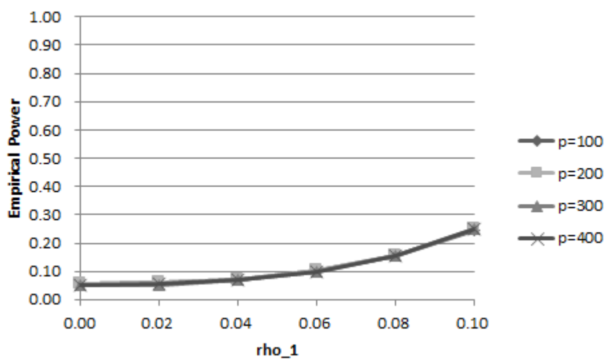

Figure 8: The empirical power of T for $p_i = 5$ and $\rho_0 = 0.5$ under p_i-banded block structure

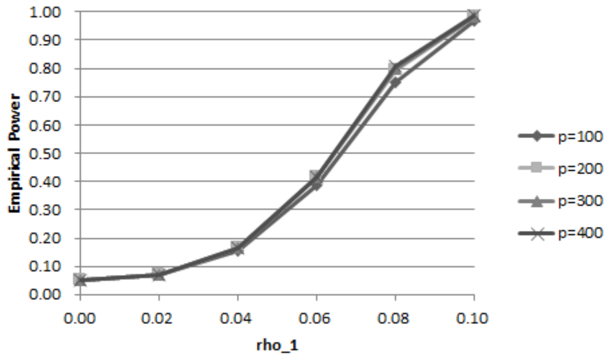

Figure 9: The empirical power of T for $p_i = 10$ and $\rho_0 = 0.1$ under p_i-banded block structure

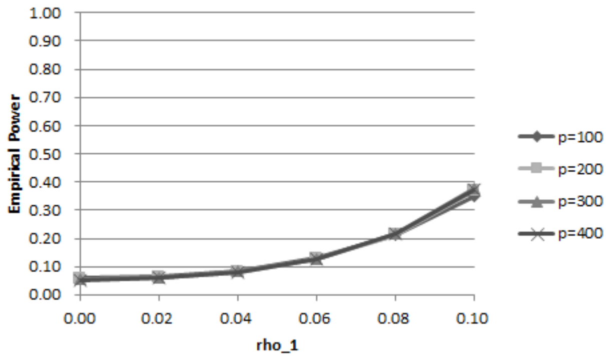

Figure 10: The empirical power of T for $p_i = 10$ and $\rho_0 = 0.5$ under p_i-banded block structure

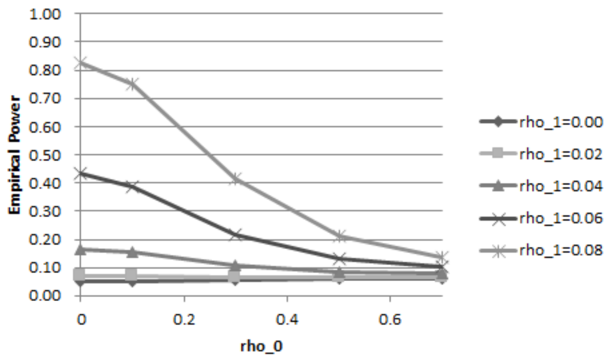

Figure 11: The empirical power of T for $p = 100$ and $p_i = 10$ under p_i-banded block structure

Table 1: The empirical type I error rate for "small-" and "large-" size blocks cases

p_i	σ^2	ρ_0	$p = 100$	$p = 200$	$p = 300$	$p = 400$
5	1.0	0.1	0.0506	0.0492	0.0473	0.0507
		0.5	0.0547	0.0550	0.0490	0.0515
		0.9	0.0539	0.0532	0.0538	0.0491
	5.0	0.1	0.0506	0.0492	0.0473	0.0507
		0.5	0.0547	0.0550	0.0490	0.0515
		0.9	0.0539	0.0532	0.0538	0.0491
10	1.0	0.1	0.0496	0.0502	0.0498	0.0519
		0.5	0.0590	0.0560	0.0552	0.0527
		0.9	0.0584	0.0510	0.0540	0.0531
	5.0	0.1	0.0496	0.0502	0.0498	0.0519
		0.5	0.0590	0.0560	0.0552	0.0527
		0.9	0.0584	0.0510	0.0540	0.0531
25	1.0	0.1	0.0527	0.0505	0.0489	0.0461
		0.5	0.0642	0.0581	0.0504	0.0500
		0.9	0.0668	0.0566	0.0540	0.0509
	5.0	0.1	0.0527	0.0505	0.0489	0.0461
		0.5	0.0642	0.0581	0.0504	0.0500
		0.9	0.0668	0.0566	0.0540	0.0509

Table 2: The empirical type I error rate for "mix-" size blocks cases

(p_i, p_j)	σ^2	ρ_0	$p = 100$	$p = 200$	$p = 300$	$p = 400$
(5, 25)	1.0	0.1	0.0506	0.0492	0.0470	0.0474
		0.5	0.0621	0.0542	0.0530	0.0488
		0.9	0.0641	0.0608	0.0562	0.0549
	5.0	0.1	0.0506	0.0492	0.0470	0.0474
		0.5	0.0621	0.0542	0.0530	0.0488
		0.9	0.0641	0.0608	0.0562	0.0549
(5, 50)	1.0	0.1	0.0519	0.0515	0.0512	0.0504
		0.5	0.0636	0.0615	0.0592	0.0621
		0.9	0.0666	0.0626	0.0645	0.0591
	5.0	0.1	0.0519	0.0515	0.0512	0.0504
		0.5	0.0636	0.0615	0.0592	0.0621
		0.9	0.0666	0.0626	0.0645	0.0591
(10, 25)	1.0	0.1	0.0501	0.0507	0.0485	0.0481
		0.5	0.0640	0.0535	0.0531	0.0507
		0.9	0.0659	0.0565	0.0550	0.0536
	5.0	0.1	0.0501	0.0507	0.0485	0.0481
		0.5	0.0640	0.0535	0.0531	0.0507
		0.9	0.0659	0.0565	0.0550	0.0536
(10, 50)	1.0	0.1	0.0507	0.0518	0.0529	0.0479
		0.5	0.0656	0.0632	0.0573	0.0603
		0.9	0.0693	0.0588	0.0614	0.0574
	5.0	0.1	0.0507	0.0518	0.0529	0.0479
		0.5	0.0656	0.0632	0.0573	0.0603
		0.9	0.0693	0.0588	0.0614	0.0574