Accepted Manuscript

More powerful tests for sparse high-dimensional covariances matrices

Liu-hua Peng, Song Xi Chen, Wen Zhou

PII: S0047-259X(16)30004-5
DOI: http://dx.doi.org/10.1016/j.jmva.2016.03.008
Reference: YJMVA 4106

To appear in: *Journal of Multivariate Analysis*

Received date: 2 September 2015

Please cite this article as: L.-h. Peng, S.X. Chen, W. Zhou, More powerful tests for sparse high-dimensional covariances matrices, *Journal of Multivariate Analysis* (2016), http://dx.doi.org/10.1016/j.jmva.2016.03.008

More Powerful Tests for Sparse High-Dimensional Covariances Matrices

Liu-hua Peng[a,*], Song Xi Chen[b], Wen Zhou[c]

[a] *Department of Statistics, Iowa State University, Ames, Iowa 50011, U.S.A.*
[b] *Department of Statistics, Iowa State University, Ames, Iowa 50011, U.S.A.; Department of Business Statistics and Econometrics, Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, P.R. China*
[c] *Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA.*

Abstract

This paper considers improving the power of tests for the identity and sphericity hypotheses regarding high dimensional covariance matrices. The power improvement is achieved by employing the banding estimator for the covariance matrices, which leads to significant reduction in the variance of the test statistics in high dimension. Theoretical justification and simulation experiments are provided to ensure the validity of the proposed tests. The tests are used to analyze a dataset from an acute lymphoblastic leukemia gene expression study for an illustration.

Keywords:
Banding estimation; High-dimensional inference; Hypothesis testing; Sparse Covariance matrix.

1. Introduction

This paper is interested in testing hypothesis for high-dimensional covariance matrices, $\mathbf{\Sigma}$, of a p-dimensional random vector $\mathbf{X}$. In practice, it is often of scientific interest to test whether or not a prescribed dependence structure is supported by data, for instance

$$H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0 \quad vs. \quad H_1 : \mathbf{\Sigma} \neq \mathbf{\Sigma}_0 \tag{1.1}$$

and

$$H_0 : \mathbf{\Sigma} = \sigma^2\mathbf{\Sigma}_0 \quad vs. \quad H_1 : \mathbf{\Sigma} \neq \sigma^2\mathbf{\Sigma}_0 \quad \text{for some unknown } \sigma^2 > 0, \tag{1.2}$$

for a known non-degenerate covariance matrix $\mathbf{\Sigma}_0$. Among many practical applications, genomic studies usually motivate (1.1) or (1.2): it is not uncommon to postulate a grouping structure among

*Corresponding author
Email addresses: penglh@iastate.edu (Liu-hua Peng), songchen@iastate.edu,csx@gsm.pku.edu.cn (Song Xi Chen), riczw@stat.colostate.edu (Wen Zhou)

genes of interest such that genes are not correlated across groups (Katsani et al. [23]), i.e. $\mathbf{\Sigma}$ is presumed in a diagonal block shape upon permutations. Additionally, in the fields of image segmentation, epidemiology and ecology, large numbers of pixels or population abundances are collected across the spatial domain. Certain spatial autocorrelations are usually prescribed for fitting data to a parametric or semiparametric model for predictions (Bolker [9], Cressie [17]). It is important to verify whether or not these hypothetical dependence structures are supported by data.

For identically and independently distributed data $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with unknown common mean μ and covariance $\mathbf{\Sigma}_0$, linear transform $\mathbf{\Sigma}_0^{-1/2}\mathbf{X}_i$ reduces (1.1) and (1.2) to

$$H_0 : \mathbf{\Sigma} = \mathbf{I}_p \quad vs. \quad H_1 : \mathbf{\Sigma} \neq \mathbf{I}_p, \tag{1.3}$$

and

$$H_0 : \mathbf{\Sigma} = \sigma^2\mathbf{I}_p \quad vs. \quad H_1 : \mathbf{\Sigma} \neq \sigma^2\mathbf{I}_p, \tag{1.4}$$

where $\mathbf{I}_p$ is the p-dimensional identity matrix. Hypotheses (1.3) and (1.4) are called the identity and sphericity hypothesis, respectively. For fixed p, likelihood ratio test has been developed and widely applied. We refer to Anderson [1] for more details. Let $\widehat{\mathbf{\Sigma}}$ be the sample covariance matrix. John [25, 26] and Nagao [28] showed that for a fixed p, test statistics

$$\mathrm{V}_n = p^{-1}\mathrm{tr}\{(\widehat{\mathbf{\Sigma}} - \mathbf{I}_p)^2\} \quad \text{and} \quad \mathrm{U}_n = p^{-1}\mathrm{tr}[\{p\widehat{\mathbf{\Sigma}}/\mathrm{tr}(\widehat{\mathbf{\Sigma}}) - \mathbf{I}_p\}^2] \tag{1.5}$$

provide the most powerful invariant tests for both the identity and sphericity hypotheses against the local alternatives. Traditional tests, however, are not applicable to the large p, small n paradigm since the sample covariance matrix is singular with probability one whenever $p > n$ and is no longer a consistent estimator if p is not a smaller order of n (Bai and Yin [4], Bai et al. [5]).

Tests for covariance matrices suited for the high dimensionality have been developed over the recent years. Ledoit and Wolf [27] established the asymptotic properties of statistics in (1.5) for $p/n \to c \in (0, +\infty)$ and proposed tests for identity (1.4) and sphericity (1.4) under the Gaussian assumption. Jiang [24] developed a sphericity test based on the max-type statistic $L_n = \max_{1\leq i<j\leq p} |\hat{\rho}_{ij}|$, where $\hat{\rho}_{ij}$ denotes the sample correlation coefficient between the i-th and j-th components of $\mathbf{X}$. With the aid of the random matrix theory, Bai et al. [2] proposed a modified likelihood ratio statistic for testing (1.3) for $p/n \to y \in (0, 1)$. To avoid the issue of inconsistency of $\widehat{\mathbf{\Sigma}}$ when $p > n$, Chen et al. [14]

proposed U-statistic based testing procedures for both the identity and sphericity hypotheses. Their tests require much relaxed assumptions on the data distribution, and allow p diverges in n in any rates. See, for example, Cai and Jiang [10], Hallin and Paindaveine [22], Srivastava and Yanagihara [35], Srivastava and Reid [36], Srivastava et al. [37], Zou et al., [43] for alternative test formulations, and Bai et al. [2], Schott [33], Zheng et al., [42] for related works. One limitation of these high dimensional tests is a loss of power under sparse high dimension situations, largely due to a rapid increase in the variance of the test statistic as the p gets larger. For instance, in the formulation of the identity test, estimation of the discrepancy measure $p^{-1}\text{tr}\{(\widehat{\mathbf{\Sigma}} - \mathbf{I}_p)^2\}$ involves all the entries of the sample covariance. As a result, the test statistics incurs larger variation as the dimension gets larger. The increased variance dilutes the signal $p^{-1}\text{tr}\{(\widehat{\mathbf{\Sigma}} - \mathbf{I}_p)^2\}$ of the test and hence brings down its power.

While we are gathering more dimensions in the data as more features are recorded, the information content of the data is not necessarily increasing at the same rate as the dimension. Indeed, it is commonly acknowledged that parameters associated with high dimensional data can be sparse in the sense of that many of the parameters are either zero or taking small values. This was the rationale behind the proposal of LASSO in Tibshirani [38] as well as other regularization-based estimations in regression and covariance matrices; see Bickel and Levina [8], Cai et al. [11], Fan and Li [19], Rothman et al. [32]. We consider in this paper tests for covariance matrices by utilizing the regularization-based estimation constructed for a specific class of sparse covariance matrices, the so-called bandable covariances, introduced by Bickel and Levina [7]. The bandable class is naturally suited as alternative hypotheses to the null identity and the sphericity hypotheses. Specifically, we formulate the test statistics by employing the banded covariance estimator proposed in Bickel and Levina [7]. This allows us to take advantage of the knowledge of sparsity in the $\mathbf{\Sigma}$. We demonstrate in this paper that the new test formulations have a remarkable power enhancement over the existing high dimensional tests for the covariance which do not utilize the sparsity information.

The rest of the paper is organized as follows. We introduce our motivations in Section 2 and present the testing procedures in Section 3. The theoretical properties of the proposed tests are also investigated in Section 3. Section 4 is devoted to a discussion on the selection of k for the proposed tests. Numerical results are displayed in Section 5 to investigate the performance of the tests in practice. Both simulation studies and applications of the proposed tests to an acute lymphoblastic leukemia gene expression data set are reported. The last section concludes the article with a brief discussion, and technical proofs are given in the Appendix. Supplementary material contains more details on the numerical studies.

2. Motivations and Preliminaries

Our investigation is motivated by the notion of the bandable covariance class introduced by Bickel and Levina [7], which is defined as

$$\mathcal{U}(\varepsilon_0, \mathrm{C}, \alpha) = \Big\{ \boldsymbol{\Sigma} = (\sigma_{ij})_{1\le i,j\le p} : \max_j \sum_{|i-j|>k_0} |\sigma_{ij}| \le \mathrm{C}k_0^{-\alpha} \text{ for all } k_0 \ge 0, \\ 0 < \varepsilon_0 \le \lambda_{\min}(\boldsymbol{\Sigma}) \le \lambda_{\max}(\boldsymbol{\Sigma}) \le 1/\varepsilon_0 \Big\}, \tag{2.1}$$

where ε_0, C and α are positive constants, $\lambda_{\min}(\boldsymbol{\Sigma})$ and $\lambda_{\max}(\boldsymbol{\Sigma})$ are the smallest and the largest eigenvalues of $\boldsymbol{\Sigma}$. The bandable covariance prescribes a general decaying pair-wise covariances σ_{ij} for large $|i-j|$. The increasing sparsity as the pair-wise covariance moves away from the main diagonal is ideally suited as alternative models for the identity and sphericity hypotheses. It is noted that the $\boldsymbol{\Sigma}$ of the original random vector $\mathbf{X}$ may not be bandable. We assume there is a permutation of $\mathbf{X}$ such that the corresponding covariance is bandable. There are algorithms, for instance the angle-based ordering algorithm (Friendly [20]) or the Isoband algorithm (Wagaman and Levina [41]), which may be used to permute the data so that $\boldsymbol{\Sigma}$ under the permutation is more bandable.

For the above bandable covariance class, Bickel and Levina [7] proposes the banding covariance estimator. Let $k < p$ be a positive integer and write the sample covariance matrix $\widehat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{ij})_{1\le i,j\le p}$. The banding estimator of $\boldsymbol{\Sigma}$ with bandwidth k is $\hat{\boldsymbol{\Sigma}}_{k,p} \equiv \hat{\boldsymbol{\Sigma}}_k = \mathrm{B}_k(\widehat{\boldsymbol{\Sigma}}) = (\hat{\sigma}_{ij}\mathrm{I}\{|i-j| \le k\})_{1\le i,j\le p}$. For $\boldsymbol{\Sigma} \in \mathcal{U}(\varepsilon_0, \mathrm{C}, \alpha)$, Bickel and Levina [7] established the consistency of $\mathrm{B}_k(\widehat{\boldsymbol{\Sigma}})$ to $\boldsymbol{\Sigma}$ under the spectral norm by letting k divergence at the rate $(n^{-1}\ln p)^{-1/\{2(\alpha+1)\}}$ for sub-Gaussian distributed data when $\ln p/n \to 0$.

Encouraged by this important advance in high dimensional covariance estimation, we consider replacing $\widehat{\boldsymbol{\Sigma}}$ in (1.5) by $\mathrm{B}_k(\widehat{\boldsymbol{\Sigma}})$ leads to the following test statistics

$$p^{-1}\mathrm{tr}[\{\mathrm{B}_k(\widehat{\boldsymbol{\Sigma}}) - \mathbf{I}_p\}^2] \ \text{ and } \ p^{-1}\mathrm{tr}[\{p\mathrm{B}_k(\widehat{\boldsymbol{\Sigma}})/\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}) - \mathbf{I}_p\}^2]. \tag{2.2}$$

Comparing with the statistics V_n and U_n given in (5), the above formulations based on the banding estimator $\mathrm{B}_k(\widehat{\boldsymbol{\Sigma}})$ are expected to be less variable since those pair-wise sample covariances $\hat{\sigma}_{ij}$ located further away from the k-th superdiagonals (subdiagonals) are excluded due to the banding operation.

Indeed, for $\boldsymbol{\Sigma}$ in the bandable class, most of the signals (those larger σ_{ij}) are located closer to the main diagonals. This form of sparsity suggests us to discount covariances which are far away from the main diagonals. It is obvious that the formulation is critically dependent on the choice of the banding width k. Recently, Qiu and Chen [31] have proposed a data driven method for choosing k by minimizing an empirical version of $||\mathrm{B}_k(\boldsymbol{\Sigma}) - \boldsymbol{\Sigma}||_F$, where $||\cdot||_F$ denotes the Frobenius norm.

The seemingly natural constructions given in (2.2) need to be refined in order to be suitable for more relaxed relationship between p and n and without the sub-Gaussian assumption in Bickel and Levina [7]. Our aim here is to obtain unbiased estimators for $\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\}$ and $\mathrm{tr}[\{\mathrm{B}_k(\boldsymbol{\Sigma})\}^2]$ for $\boldsymbol{\Sigma} \in \mathcal{U}(\varepsilon_0, \mathrm{C}, \alpha)$. Denote $\mathbf{X}_i = (x_{i1}, \ldots, x_{ip})^\top$ for each i, and let

$$\mathrm{L}_{n_1}(i,j) = \frac{1}{P_n^2} \sum_{l_1 \neq l_2} x_{l_1 i} x_{l_1 j} x_{l_2 i} x_{l_2 j},\ \mathrm{L}_{n_2}(i,j) = \frac{1}{P_n^3} \sum_{l_1,l_2,l_3}^{*} x_{l_1 i} x_{l_2 j} x_{l_3 i} x_{l_3 j},$$

$$\mathrm{L}_{n_3}(i,j) = \frac{1}{P_n^4} \sum_{l_1,l_2,l_3,l_4}^{*} x_{l_1 i} x_{l_2 j} x_{l_3 i} x_{l_4 j},\ \mathrm{L}_{n_4}(i) = \frac{1}{n} \sum_{l=1}^{n} x_{li}^2,\ \mathrm{L}_{n_5}(i) = \frac{1}{P_n^2} \sum_{l_1 \neq l_2} x_{l_1 i} x_{l_2 i},$$

where $P_n^r = n!/(n-r)!$ for $r = 2, 3, 4$ and $\sum^{*}$ denotes the summation over mutually different indices. Notice that $\sum_{i=1}^{p} \{\mathrm{L}_{n_4}(i) - \mathrm{L}_{n_5}(i)\}$ and $\sum_{|i-j| \leq k} \{\mathrm{L}_{n_1}(i,j) - 2\mathrm{L}_{n_2}(i,j) + \mathrm{L}_{n_3}(i,j)\}$ are unbiased estimators of $\mathrm{tr}(\boldsymbol{\Sigma})$ and $\mathrm{tr}[\{\mathrm{B}_k(\boldsymbol{\Sigma})\}^2]$, respectively.

We consider two discrepancy measures for the identity and sphericity hypotheses:

$$p^{-1}\mathrm{tr}[\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}^2] = p^{-1}\mathrm{tr}[\{\mathrm{B}_k(\boldsymbol{\Sigma})\}^2] - 2p^{-1}\mathrm{tr}(\boldsymbol{\Sigma}) + 1,$$

and

$$\frac{1}{p}\mathrm{tr}\left[\left\{\frac{\mathrm{B}_k(\boldsymbol{\Sigma})}{(1/p)\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\}} - \mathbf{I}_p\right\}^2\right] = \frac{p\mathrm{tr}[\{\mathrm{B}_k(\boldsymbol{\Sigma})\}^2]}{\{\mathrm{tr}(\boldsymbol{\Sigma})\}^2} - 1,$$

upon which, we propose the following two unbiased estimators to these two measures

$$\mathrm{V}_{n,k} = p^{-1} \sum_{|i-j| \leq k} \{\mathrm{L}_{n_1}(i,j) - 2\mathrm{L}_{n_2}(i,j) + \mathrm{L}_{n_3}(i,j)\} - 2p^{-1} \sum_{i=1}^{p} \{\mathrm{L}_{n_4}(i) - \mathrm{L}_{n_5}(i)\} + 1, \tag{2.3}$$

and

$$\mathrm{U}_{n,k} = p\left[\frac{\sum_{|i-j| \leq k} \{\mathrm{L}_{n_1}(i,j) - 2\mathrm{L}_{n_2}(i,j) + \mathrm{L}_{n_3}(i,j)\}}{\{\sum_{i=1}^{p} (\mathrm{L}_{n_4}(i) - \mathrm{L}_{n_5}(i))\}^2}\right] - 1. \tag{2.4}$$

It is noted that $\mathrm{V}_{n,k}$ and $\mathrm{U}_{n,k}$ reduce to the statistics in Chen et al. [14] when the banding width

$k = p - 1$ which involves all the sample covariances. Thus, $\mathrm{V}_{n,k}$ and $\mathrm{U}_{n,k}$ are regularized versions of those proposed by Chen et al. [14]. By utilizing the sparse information that $\mathbf{\Sigma} \in \mathcal{U}(\varepsilon_0, \mathrm{C}, \alpha)$, those $\hat{\sigma}_{ij}$ beyond the k-th superdiagonal are avoided which makes $\mathrm{V}_{n,k}$ and $\mathrm{U}_{n,k}$ have less variations and hence more powerful tests as we will demonstrate later.

It is easy to see that $\mathrm{V}_{n,k}$ and $\mathrm{U}_{n,k}$ are invariant under the location shift. Hence, without loss of generality, we assume data has been centered such that $\mu = 0$.

Assumption 1. $\mathbf{\Sigma} \in \mathcal{U}(\varepsilon_0, \mathrm{C}, \alpha)$ for some constants $\varepsilon_0, \mathrm{C}$ and α which are unrelated to p.

Similarly to Bai and Saranadasa [3] and Chen et al. [14], we assume the following multivariate model for $\mathbf{X}_i$.

Assumption 2. Data $X_1, \ldots, X_n$ are independent and identically distributed p-dimensional random vectors such that

$$\mathbf{X}_i = \mathbf{\Gamma}\mathbf{Z}_i \qquad \text{for} \quad i = 1, \ldots, n, \tag{2.5}$$

where $\mathbf{\Gamma} = (\mathbf{\Gamma}_{ij})_{1 \le i \le p, 1 \le j \le m}$ is a constant loading matrix with $p \le m$ and $\mathbf{\Gamma}\mathbf{\Gamma}^\top = \mathbf{\Sigma}$, $\mathbf{Z}_i = (z_{i1}, \ldots, z_{im})$'s are independent and identically p-dimensional random vectors with zero mean and identity covariance. Furthermore, we assume $\sup_j \mathrm{E}(z_{1j}^8) < C_1$ for some constant $C_1 > 0$ and there exists a constant $\Delta < \infty$ such that $\mathrm{E}(z_{1j}^4) = 3 + \Delta$ for each j. For any integer $\ell_v \ge 0$ with $\sum_{v=1}^q \ell_v \le 8$,

$$\mathrm{E}(z_{1i_1}^{\ell_1} \ldots z_{1i_q}^{\ell_q}) = \mathrm{E}(z_{1i_1}^{\ell_1}) \ldots \mathrm{E}(z_{1i_q}^{\ell_q}) \tag{2.6}$$

whenever $i_1, \ldots, i_q$ are distinct.

This model, first employed by Bai and Saranadasa [3] for testing high-dimensional mean vectors, ensures that the observations $\mathbf{X}_i$ are linearly generated by m-variate $\mathbf{Z}_i$ consisted of white noises. The dimension m of $\mathbf{Z}_i$ is finite but diverge to infinity as p and n both go to infinity. So the dimension of $\mathbf{Z}_i$ is arbitrary as long as $m \ge p$ that equips the model flexibility in generating $\mathbf{X}_i$ with covariance $\mathbf{\Sigma}$. The distribution of $\mathbf{Z}_i$ is not restricted to particular families, and is therefore nonparametric. Assumption 2 has been extensively employed in high dimensional multivariate testing problems, for example see Chen et al. [14], Touloumis et al. [39].

3. Testing Procedures

3.1. Identity test

We first derive the mean and variance of the test statistic $\mathrm{V}_{n,k}$ for the identity hypothesis. Derivations given in Lemma A.1 in the appendix show that under Assumptions 1 and 2, as $n, p \to \infty$ and $k \to \infty$, if $k = o(\min(n^{1/2}, p^{1/2}))$,

$$\mathrm{E}(\mathrm{V}_{n,k}) = p^{-1}\mathrm{tr}[\{\mathrm{B}_k(\mathbf{\Sigma}) - \mathbf{I}_p\}^2] \quad \text{and} \quad \mathrm{var}(\mathrm{V}_{n,k}) = p^{-2}\sigma^2_{\mathrm{V}_{n,k}}\{1 + o(1)\} \tag{3.1}$$

where

$$\begin{aligned}\sigma^2_{\mathrm{V}_{n,k}} =& \tau^2_{n,k}(\mathbf{\Sigma}) + 8n^{-1}\mathrm{tr}\left[\mathbf{\Sigma}\left\{\mathrm{B}_k(\mathbf{\Sigma}) - \mathbf{I}_p\right\}\right]^2 \\ &+ 4\Delta n^{-1}\mathrm{tr}\left[\left\{\mathbf{\Gamma}^\top(\mathrm{B}_k(\mathbf{\Sigma}) - \mathbf{I}_p)\mathbf{\Gamma}\right\} \circ \left\{\mathbf{\Gamma}^\top(\mathrm{B}_k(\mathbf{\Sigma}) - \mathbf{I}_p)\mathbf{\Gamma}\right\}\right],\end{aligned} \tag{3.2}$$

$$\begin{aligned}\tau^2_{n,k}(\mathbf{\Sigma}) =& \frac{4}{n(n-1)}\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k}\sigma^2_{i_1i_2}\sigma^2_{j_1j_2} + \frac{4}{n(n-1)}\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k}\sigma_{i_1i_2}\sigma_{j_1j_2}\sigma_{i_1j_2}\sigma_{j_1i_2} \\ &+ \frac{8\Delta}{n(n-1)}\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k}\sigma_{i_1i_2}\sigma_{j_1j_2}f_{i_1j_1i_2j_2}(\mathbf{\Sigma}) + \frac{2\Delta^2}{n(n-1)}\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k}f^2_{i_1j_1i_2j_2}(\mathbf{\Sigma}),\end{aligned} \tag{3.3}$$

and $\circ$ denotes the Hadamard product of matrices. In (3.3), $f_{i_1j_1i_2j_2}(\mathbf{\Sigma}) = \sum_{r=1}^m \mathbf{\Gamma}_{i_1r}\mathbf{\Gamma}_{j_1r}\mathbf{\Gamma}_{i_2r}\mathbf{\Gamma}_{j_2r}$ for the loading matrix $\mathbf{\Gamma} = (\mathbf{\Gamma}_{jl})_{p\times m}$ in Assumption 2.

Based on (3.1), the following theorem establishes the asymptotic normality of $\mathrm{V}_{n,k}$.

Theorem 1. *Under Assumptions 1 and 2, as $n \to \infty$, $p \to \infty$ and $k \to \infty$, if $k = o(\min(n^{1/2}, p^{1/2}))$,*

$$\sigma^{-1}_{\mathrm{V}_{n,k}}\{p\mathrm{V}_{n,k} - \mathrm{tr}[\{\mathrm{B}_k(\mathbf{\Sigma}) - \mathbf{I}_p\}^2]\} \to \mathcal{N}(0,1) \tag{3.4}$$

in distribution.

In Theorem 1, no explicit restrictions on the growth rates of p and n are imposed. We note that the banding width prescribed in Bickel and Levina [7] that allows consistent estimation was $k = (n^{-1}\mathrm{ln}p)^{-1/\{2(\alpha+1)\}}$ for the sub-Gaussian distributed data and $k = (n^{-1/2}p^{2/\beta})^{-1/(1+\alpha+2/\beta)}$ for data with bounded β-th moment, where α is the sparsity index in the bandable class. For both cases, the condition $k = o(\min(n^{1/2}, p^{1/2}))$ assumed in Theorem 1 allows wider range of the banding width than that in Bickel and Levina [7]. This is because testing hypotheses usually requires less stringent assumptions than the estimation. It is also noted that the asymptotic normality holds even for a fixed k, but $\sigma^2_{\mathrm{V}_{n,k}}$ will be in a more involved form with more terms than those in (3.2).

Under the null identity hypothesis H_0 in (1.3), $\mathrm{E}(\mathrm{V}_{n,k}) = 0$ and $\mathrm{var}(\mathrm{V}_{n,k}) = p^{-2}\sigma^2_{\mathrm{V}_{n,k}0} + o(p^{-2}\sigma^2_{\mathrm{V}_{n,k}0})$ where

$$\sigma^2_{\mathrm{V}_{n,k}0} = \tau^2_{n,k}(I_p) = 4\{n(n-1)\}^{-1}(2pk + 2p - k^2 - k) + 8\Delta\{n(n-1)\}^{-1}\sum_{|i-j|\le k} f_{iijj}(\mathbf{I}_p)$$
$$+ 2\Delta^2\{n(n-1)\}^{-1}\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k} f^2_{i_1j_1i_2j_2}(\mathbf{I}_p).$$

From Theorem 1, the asymptotic null distribution of $\mathrm{V}_{n,k}$ is

$$p\sigma^{-1}_{\mathrm{V}_{n,k}0}\mathrm{V}_{n,k} \to \mathcal{N}(0,1) \quad \text{in distribution.} \tag{3.5}$$

To facilitate testing, we need to estimate $\sigma^2_{\mathrm{V}_{n,k}0}$. Given that the loading matrix $\boldsymbol{\Gamma}$ is not observable, it is difficult to estimate $f_{iijj}(\mathbf{I}_p)$ and $f^2_{i_1j_1i_2j_2}(\mathbf{I}_p)$ directly from data. However, we note that under the H_0

$$\sigma^2_{\mathrm{V}_{n,k}0} = 2\{n(n-1)\}^{-1}\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k}\{\mathrm{E}(x_{li_1}x_{lj_1}x_{li_2}x_{lj_2}) - \sigma_{i_1j_1}\sigma_{i_2j_2}\}^2,$$

which suggests that $\sigma^2_{\mathrm{V}_{n,k}0}$ can be estimated by

$$\begin{aligned}\hat\sigma^2_{\mathrm{V}_{n,k}0} =& 2\{n(n-1)\}^{-1}\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k}\Bigg\{\frac{1}{P_n^2}\sum_{l_1\ne l_2} x_{l_1i_1}x_{l_1j_1}x_{l_1i_2}x_{l_1j_2}\times \\ &\times x_{l_2i_1}x_{l_2j_1}x_{l_2i_2}x_{l_2j_2} - \frac{2}{P_n^3}\sum^{*}_{l_1,l_2,l_3} x_{l_1i_1}x_{l_1j_1}x_{l_2i_2}x_{l_2j_2}x_{l_3i_1}x_{l_3j_1}x_{l_3i_2}x_{l_3j_2} \\ &+ \frac{1}{P_n^4}\sum^{*}_{l_1,l_2,l_3,l_4} x_{l_1i_1}x_{l_1j_1}x_{l_2i_2}x_{l_2j_2}x_{l_3i_1}x_{l_3j_1}x_{l_4i_2}x_{l_4j_2}\Bigg\}.\end{aligned} \tag{3.6}$$

The consistency of $\hat\sigma^2_{\mathrm{V}_{n,k}0}$ is implied from the following proposition.

Proposition 1. Under Assumptions 1 and 2, and if $\mu = 0$, then

$$\mathrm{E}\left(\hat\sigma^2_{\mathrm{V}_{n,k}0}\right) = \sigma^2_{\mathrm{V}_{n,k}0} \quad \text{and} \quad \mathrm{var}\left(\hat\sigma^2_{\mathrm{V}_{n,k}0}/\sigma^2_{\mathrm{V}_{n,k}0}\right) = O(k^2n^{-1} + n^{-1}).$$

In practice, to cater for the case of $\boldsymbol{\mu} \ne 0$, we can replace $x_{i\ell}$ by $x_{i\ell} - \bar x_\ell$ where $\bar x_\ell = n^{-1}\sum_{\ell=1}^n x_{i\ell}$

for each ℓ to center the data so that $\mu = 0$ is satisfied. As Proposition 1 implies $\hat{\sigma}^2_{\mathrm{V}_{n,k}0}/\sigma^2_{\mathrm{V}_{n,k}0} \to 1$ in probability under the H_0. we have

$$p\hat{\sigma}^{-1}_{\mathrm{V}_{n,k}0}\mathrm{V}_{n,k} \to \mathcal{N}(0,1) \quad \text{in distribution} \tag{3.7}$$

under the H_0. Therefore, a regularised identity test with a nominal significant level α rejects $H_0 : \mathbf{\Sigma} = \mathbf{I}_p$ if

$$p\hat{\sigma}^{-1}_{\mathrm{V}_{n,k}0}\mathrm{V}_{n,k} > z_\alpha$$

where z_α is the α upper-quantile of $\mathcal{N}(0,1)$. As long as $k = o(\min(n^{1/2}, p^{1/2}))$, the asymptotic normality given in (3.7) ensures the nominal level of significance asymptotically.

3.2. Power of the identity test

To evaluate the power of the test for the identity hypothesis, let $\delta_{\mathrm{V}_{n,k}} = \mathrm{tr}[\{\mathrm{B}_k(\mathbf{\Sigma}) - \mathbf{I}_p\}^2]$, which can be viewed as the signal of the test problem under the alternative. The power of the regularised identity test

$$\begin{aligned}
\beta_{\mathrm{V}_{n,k}}(\alpha) &= \Pr\left(p\sigma^{-1}_{\mathrm{V}_{n,k}0}\mathrm{V}_{n,k} \geq z_\alpha\right) \\
&= \Pr\left\{\sigma^{-1}_{\mathrm{V}_{n,k}}(p\mathrm{V}_{n,k} - \mathrm{tr}[\{\mathrm{B}_k(\mathbf{\Sigma}) - \mathbf{I}_p\}^2]) \geq \sigma^{-1}_{\mathrm{V}_{n,k}}\sigma_{\mathrm{V}_{n,k}0}z_\alpha - \sigma^{-1}_{\mathrm{V}_{n,k}}\delta_{\mathrm{V}_{n,k}}\right\} \\
&= 1 - \Phi\left(\sigma^{-1}_{\mathrm{V}_{n,k}}\sigma_{\mathrm{V}_{n,k}0}z_\alpha - \sigma^{-1}_{\mathrm{V}_{n,k}}\delta_{\mathrm{V}_{n,k}}\right),
\end{aligned}$$

where Φ is the standard normal distribution. It can be shown that $\sigma^{-1}_{\mathrm{V}_{n,k}}\sigma_{\mathrm{V}_{n,k}0}$ is bounded. Hence, the power of the proposed identity test is largely determined by $\sigma^{-1}_{\mathrm{V}_{n,k}}\delta_{\mathrm{V}_{n,k}}$, which may be regarded as the signal to noise ratio of the test.

We now discuss the role of the banding width on the power of the test. We note that both $\delta_{\mathrm{V}_{n,k}}$ and $\sigma^2_{\mathrm{V}_{n,k}}$ are increasing with respect to k. For two banding widths $k_2 > k_1$, suppose that $\delta_{\mathrm{V}_{n,k_1}} > 0$, it may be shown that $\sigma^{-1}_{\mathrm{V}_{n,k_2}}\delta_{\mathrm{V}_{n,k_2}} \geq \sigma^{-1}_{\mathrm{V}_{n,k_1}}\delta_{\mathrm{V}_{n,k_1}}$ if and only if

$$(\delta_{\mathrm{V}_{n,k_2}} - \delta_{\mathrm{V}_{n,k_1}})/\delta_{\mathrm{V}_{n,k_1}} + 1 \geq \{(\sigma^2_{\mathrm{V}_{n,k_2}} - \sigma^2_{\mathrm{V}_{n,k_1}})/\sigma^2_{\mathrm{V}_{n,k_1}} + 1\}^{1/2}.$$

Therefore, if the relative signal increment $(\delta_{\mathrm{V}_{n,k_2}} - \delta_{\mathrm{V}_{n,k_1}})/\delta_{\mathrm{V}_{n,k_1}}$ can off-set the relative increase in the noise as specified above, the test with the larger k_2 is more powerful than that with the smaller k_1. On the contrary, if the relative increase in the signal can not off-set the relative increase in the noise, using the large k leads to a loss of power.

To answer the question that when will the increase of the banding width not bring in more power, we note that

$$\delta_{\mathrm{V}_{n,k_2}} - \delta_{\mathrm{V}_{n,k_1}} = \sum_{k_1<|i-j|\le k_2} \sigma_{ij}^2$$

and for a positive constant c,

$$\sigma^2_{\mathrm{V}_{n,k_2}} - \sigma^2_{\mathrm{V}_{n,k_1}} \ge 4n(n-1)^{-1} \sum_{k_1<|i-j|\le k_2} \sigma_{ii}^2\sigma_{jj}^2 \ge c(k_2-k_1)pn^{-2}.$$

This implies that a necessary condition for $\sigma^{-1}_{\mathrm{V}_{n,k_2}}\delta_{\mathrm{V}_{n,k_2}} \ge \sigma^{-1}_{\mathrm{V}_{n,k_1}}\delta_{\mathrm{V}_{n,k_1}}$ is

$$(\delta_{\mathrm{V}_{n,k_2}} - \delta_{\mathrm{V}_{n,k_1}})/\delta_{\mathrm{V}_{n,k_1}} \ge c_1(k_2-k_1)n^{-1}$$

for a positive constant c_1. Hence, if the relative signal increment is smaller than $c_1(k_2-k_1)n^{-1}$, using the larger k_2 will result in a loss in the power.

To gain further insight, let us consider the case of banded covariance in which $\mathbf{\Sigma} = \mathrm{B}_{\tilde{k}}(\mathbf{\Sigma})$ for a $\tilde{k}$. In this case, $\delta_{\mathrm{V}_{n,k}} = \delta_{\mathrm{V}_{n,\tilde{k}}}$ while $\sigma^2_{\mathrm{V}_{n,k}}$ keeps increasing for all $k \ge \tilde{k}$. This means that $\sigma^{-1}_{\mathrm{V}_{n,k}}\delta_{\mathrm{V}_{n,k}}$ gets smaller and a loss of power occurs as k gets larger beyond $\tilde{k}$. Since our proposed test is identical to the one in Chen et al. [14] if $k = p-1$, the proposed identity test is asymptotically more powerful for a properly selected k under the banded scenario.

The following theorem establishes the consistency of the proposed identity test.

Theorem 2. *Under Assumptions 1 and 2, $p \to \infty$, $k \to \infty$ and $k = o(\min(n^{1/2}, p^{1/2}))$ as $n \to \infty$, $\beta_{\mathrm{V}_{n,k}} \to 1$ provided $\{\tau_{n,k}(\mathbf{\Sigma})\}^{-1}\delta_{\mathrm{V}_{n,k}} \to \infty$.*

We note here that $\tau_{n,k}(\mathbf{\Sigma})$ defined in (3.3) is the first term of $\sigma^2_{V_n,k}$. Theorem 2 implies that the proposed identity test is able to detect alternatives with power tending to 1 when $\{\tau_{n,k}(\mathbf{\Sigma})\}^{-1}\delta_{\mathrm{V}_{n,k}} \to \infty$. We note that when $\{\tau_{n,k}(\mathbf{\Sigma})\}^{-1}\delta_{\mathrm{V}_{n,k}} \to \infty$, $\tau^2_{n,k}(\mathbf{\Sigma})$ dominates the other two terms of $\sigma^2_{\mathrm{V}_{n,k}}$ in (3.2), which means $\sigma^{-1}_{\mathrm{V}_{n,k}}\delta_{\mathrm{V}_{n,k}} \to \infty$. It may be shown that $\tau^2_{n,k}(\mathbf{\Sigma})$ is at most $O(k^2pn^{-2})$ under Assumptions 1 and 2. Hence, the test has power approaching to one as long as $p^{-1}\delta_{\mathrm{V}_{n,k}}$ is a larger order of $kp^{-1/2}n^{-1}$, which is much weaker than n^{-1}, the corresponding lower limit for the test of Chen et al. [14].

3.3. Sphericity test

We firstly establish the asymptotic properties of $\mathrm{U}_{n,k}$. Let

$$\sigma^2_{\mathrm{U}_{n,k}} = \frac{\tau^2_{n,k}(\boldsymbol{\Sigma})}{\mathrm{tr}^2\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}} + \frac{2}{n}\mathrm{tr}\left[\left\{\frac{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}}{\mathrm{tr}(\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma})} - \frac{\boldsymbol{\Sigma}}{\mathrm{tr}(\boldsymbol{\Sigma})}\right\}^2\right] + \frac{\Delta}{n}\mathrm{tr}\left(\left[\boldsymbol{\Gamma}^\top\left\{\frac{\mathrm{B}_k(\boldsymbol{\Sigma})}{\mathrm{tr}(\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma})} - \frac{\mathbf{I}_p}{\mathrm{tr}(\boldsymbol{\Sigma})}\right\}\boldsymbol{\Gamma}\right] \circ \left[\boldsymbol{\Gamma}^\top\left\{\frac{\mathrm{B}_k(\boldsymbol{\Sigma})}{\mathrm{tr}(\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma})} - \frac{\mathbf{I}_p}{\mathrm{tr}(\boldsymbol{\Sigma})}\right\}\boldsymbol{\Gamma}\right]\right).$$

Theorem 3. *Under Assumptions 1 and 2, as $n \to \infty$, $p \to \infty$ and $k \to \infty$, if $k = o(\min(n^{1/2}, p^{1/2}))$,*

$$\sigma^{-1}_{\mathrm{U}_{n,k}}\left[\left\{\frac{\mathrm{tr}^2(\boldsymbol{\Sigma})}{\mathrm{tr}(\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma})}\right\}\left(\frac{\mathrm{U}_{n,k}+1}{p}\right) - 1\right] \to \mathcal{N}(0,1) \quad \textit{in distribution.} \tag{3.8}$$

The asymptotic variance under the null hypothesis is

$$\sigma^2_{\mathrm{U}_{n,k}0} = \sigma^{-8}p^{-2}\Big[4\{n(n-1)\}^{-1}(2pk - k^2 - k + 2p) + 8\Delta\{n(n-1)\}^{-1}\sum_{|i-j|\le k} f_{ijij}(\mathbf{I}_p) + 2\Delta^2\{n(n-1)\}^{-1}\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k} f^2_{i_1j_1i_2j_2}(\mathbf{I}_p)\Big].$$

Then, Theorem 3 implies that under the null sphericity hypothesis

$$\sigma^{-1}_{\mathrm{U}_{n,k}0}\mathrm{U}_{n,k} \to \mathcal{N}(0,1) \quad \text{in distribution.} \tag{3.9}$$

Using a similar approach to estimating the null variance $\sigma^2_{\mathrm{V}_{n,k}0}$ in Section 3.1, an estimator of the null variance $\sigma^2_{\mathrm{U}_{n,k}0}$ is

$$\hat\sigma^2_{\mathrm{U}_{n,k}0} = \hat\sigma^2_{\mathrm{V}_{n,k}0}\left\{\sum_{|i-j|\le k}(\mathrm{L}_{1,ij} - 2\mathrm{L}_{2,ij} + \mathrm{L}_{3,ij})\right\}^{-2}.$$

It may be shown that $\hat\sigma^2_{\mathrm{U}_{n,k}0} \to \sigma^2_{\mathrm{U}_{n,k}0}$ in probability and $\hat\sigma^{-1}_{\mathrm{U}_{n,k}0}\mathrm{U}_{n,k}$ converges to $\mathcal{N}(0,1)$ in distribution. These lead to a sphericity test with a nominal significance level α that rejects $H_0 : \boldsymbol{\Sigma} = \sigma^2\mathbf{I}_p$ if

$$\hat\sigma^{-1}_{\mathrm{U}_{n,k}0}\mathrm{U}_{n,k} > z_\alpha. \tag{3.10}$$

Let $\delta_{\mathrm{U}_{n,k}} = 1 - \{\mathrm{tr}^2(\boldsymbol{\Sigma})\}/[p\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}]$. The power of the sphericity test is

$$\begin{aligned}\beta_{\mathrm{U}_{n,k}}(\alpha) &= \Pr(\sigma^{-1}_{\mathrm{U}_{n,k}0}\mathrm{U}_{n,k} \geq z_\alpha)\\ &= 1 - \Phi\left[\left\{\frac{\mathrm{tr}^2(\boldsymbol{\Sigma})}{p\mathrm{tr}(\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma})}\right\}\left(\frac{\sigma_{\mathrm{U}_{n,k}0}}{\sigma_{\mathrm{U}_{n,k}}}\right) z_\alpha - \frac{\delta_{\mathrm{U}_{n,k}}}{\sigma_{\mathrm{U}_{n,k}}}\right].\end{aligned}$$

As $\mathrm{tr}^2(\boldsymbol{\Sigma})/[p\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}]$ and $\sigma_{\mathrm{U}_{n,k}0}/\sigma_{\mathrm{U}_{n,k}}$ are both bounded, the power is largely influenced by $\delta_{\mathrm{U}_{n,k}}/\sigma_{\mathrm{U}_{n,k}}$, which can be viewed as the signal to noise ratio of the testing problem.

To gain insight on the power, we study the signal to noise ratio as what we did for the identity test in section 3.1. It can be shown that for $k_2 > k_1$, $\sigma^{-1}_{\mathrm{U}_{n,k_2}}\delta_{\mathrm{U}_{n,k_2}} \geq \sigma^{-1}_{\mathrm{U}_{n,k_1}}\delta_{\mathrm{U}_{n,k_1}}$ if and only if

$$(\delta_{\mathrm{U}_{n,k_2}} - \delta_{\mathrm{U}_{n,k_1}})/\delta_{\mathrm{U}_{n,k_1}} + 1 \geq \{(\sigma^2_{\mathrm{U}_{n,k_2}} - \sigma^2_{\mathrm{U}_{n,k_1}})/\sigma^2_{\mathrm{U}_{n,k_1}} + 1\}^{1/2}.$$

We also note that, for $k_2 > k_1$

$$\delta_{\mathrm{U}_{n,k_2}} - \delta_{\mathrm{U}_{n,k_1}} = p^{-1}\mathrm{tr}^2(\boldsymbol{\Sigma})\frac{\sum_{k_1<|i-j|\leq k_2}\sigma_{ij}^2}{\mathrm{tr}\{\mathrm{B}^2_{k_1}(\boldsymbol{\Sigma})\}\mathrm{tr}\{\mathrm{B}^2_{k_2}(\boldsymbol{\Sigma})\}},$$

which indicates that the increase in the signal is largely driven by $\sum_{k_1<|i-j|\leq k_2}\sigma_{ij}^2$, those σ_{ij} between the k_1 and k_2-th superdiagonals. At the same time, it can be shown that $\sigma^2_{\mathrm{U}_{n,k}}$ is increasing with respect to k at a rate at least $p^{-1}n^{-2}$. Hence, a power-enhancing strategy is to use the smallest k that captures the most signals.

The following theorem establishes the consistency of the proposed sphericity test (3.10).

Theorem 4. *Under Assumptions 1 and 2, $p \to \infty$, $k \to \infty$ and $k = o(\min(n^{1/2}, p^{1/2}))$ as $n \to \infty$, if $\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}\{\tau_{n,k}(\boldsymbol{\Sigma})\}^{-1}\delta_{\mathrm{U}_{n,k}} \to \infty$ then $\beta_{\mathrm{U}_{n,k}} \to 1$.*

It can be shown that $\tau_{n,k}(\boldsymbol{\Sigma})/\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}$ is at most $O(kp^{-1/2}n^{-1})$. Hence, the proposed sphericity test is consistent as long as $\delta_{\mathrm{U}_{n,k}}$ is a larger order of $kp^{-1/2}n^{-1}$, which is much lower than the corresponding lower limit of the test of Chen et al. [14] since $k = o(p^{1/2})$. The latter test requires $\delta_{\mathrm{U}_{n,p-1}}$ is a larger order of n^{-1}.

4. Selection of k

Given the beneficial power property of the tests based on the banding operation, we report numerical results of the proposed tests with respect to k in this section.

We start with evaluating the impacts of k on the size of the tests. Clearly, under the null hypotheses for both (1.3) and (1.4), $\mathbf{\Sigma} \in \mathcal{U}(\varepsilon_0, \mathrm{C}, \alpha)$. Given the established asymptotic normality for the two test statistics, the size of the proposed tests are expected to be close to the nominal significance level for a wider range of k as long as $k = o(\min(n^{1/2}, p^{1/2}))$. To confirm this, we ran simulations for the sphericity test. We generated independent and identically distributed random vectors $\mathbf{X}_i$ from $\mathcal{N}(0, \mathbf{\Sigma})$ where $\mathbf{\Sigma} = 2\mathbf{I}_p$ and considered $k = \lceil a_i n^{1/3} \rceil$ for $\{a_i\}_{i=1}^{10} = \{0, 0.1, 0.3, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0\}$. The sample size and the dimensions considered were $n = 20, 40$ and 60, and $p = 38, 89$ and 181, respectively. Figure 1 displays the empirical size of the proposed sphericity test with respect to k's for the combinations of n and p based on 1000 simulations. The figure shows that the empirical size was largely close to the nominal 5% level for the wider choices of k's.

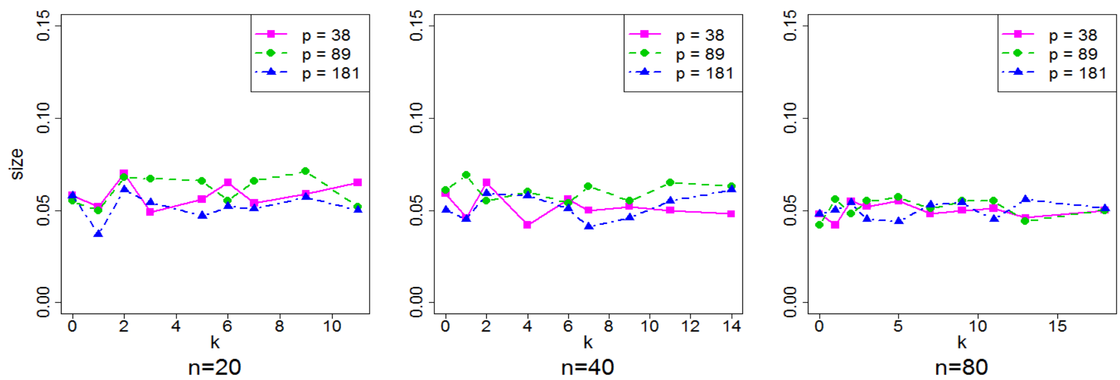


Figure 1: Empirical size of the 5% sphericity test with respect to the banding width k for various sample sizes and dimensions.

To gain information on the role of k on the power of the tests, we considered the identity test at 5% level of significance under the alternative where $\mathbf{\Sigma}_1 = \mathbf{I}_p + 0.6^2\mathbf{\Omega}$, $\mathbf{\Omega} = (\omega_{ij})_{1\le i,j\le p}$ with $\omega_{ij} = \mathrm{I}(|i-j| = \ell)$ for a fixed integer ℓ. According to Theorem 2, the proposed identity test is powerful if $\{\tau_{n,k}(\mathbf{\Sigma}_1)\}^{-1}\delta_{\mathrm{V}_{n,k}} \to \infty$. For the given covariance structure, $\delta_{\mathrm{V}_{n,k}} = 0$ for $k < \ell$ and $\delta_{\mathrm{V}_{n,k}} = 0.2592(p-\ell)$ for $k \ge \ell$. In the meanwhile, $\tau^2_{n,k}(\mathbf{\Sigma}_1)$ is strictly increasing along with k. Thus, $\{\tau_{n,k}(\mathbf{\Sigma}_1)\}^{-1}\delta_{\mathrm{V}_{n,k}}$ is maximized at $k = \ell$ for any given n and p, namely the power would be maximised if k agreed with the underlying bandwidth ℓ.

We carried out a simulation experiment for $\mathcal{N}(0, \mathbf{\Sigma}_1)$ distributed data with $\mathbf{\Sigma}_1$ defined as above and $\ell = 2$, and $n = 20$ and $p = 20$. Table 1 reports the empirical power of the proposed identity test with banding widths ranging from 0 to 7 based on 1000 replications. It shows that the test gained powers as k was increased to $\ell = 2$, peaked at $k = \ell = 2$, and then declined afterward. This power profile was highly consistent with the discussion made toward the end of Section 3.1 regarding the signals and the noise of the test statistic.

Table 1: Empirical power with respect to k for the identity test. Data were multivariate Gaussian with $\mu = 0$ and $\mathbf{\Sigma}_1 = \mathbf{I}_p + 0.6^2\mathbf{\Omega}$, where $\mathbf{\Omega} = (\omega_{ij})_{1\le i,j\le p}$ that $\omega_{ij} = \mathrm{I}(|i-j| = 2)$, $n = 20$ and $p = 20$.

k	0	1	2	3	4	5	6	7
power	0.072	0.077	0.942	0.902	0.856	0.831	0.798	0.766

While the above results were assuring, we need a practical way to select the banding width k in order to carry out the tests. Qiu and Chen [31] proposed a selection method by minimising an empirical estimate of $\mathrm{E}||\mathrm{B}_k(\mathbf{S}_n) - \mathbf{\Sigma}||_F^2$. They demonstrated that the approach has superior performance than the cross-validation approach based on random sample splitting proposed in Bickel and Levina [7]. The cross-validation was formulated based on a score function of k that measured the discrepancy between $\mathrm{B}_k(\mathbf{S}_n)$ based on one part of the split sample, and the sample covariance based on the remaining sample. The issue with the cross-validation was that the use of the inconsistent sample covariance makes the approach unreliable in high dimension.

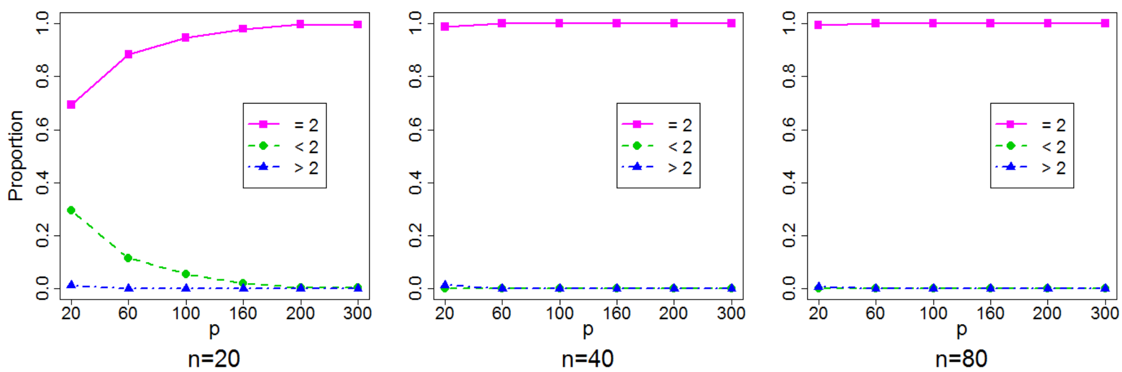


Figure 2: Proportions of the banding width selected by the method of Qiu and Chen [31]. Red solid lines with squares report the proportion of times that the selected k agrees with the true value $\ell = 2$; the green dashed lines with circles report the proportion that the selected $k < 2$; and the blue dot-dash lines with triangles report the proportion that the $k > 2$.

We carried out 1000 simulation experiments to investigate the performance of the banding width denoted as k_{qc} prescribed by Qiu and Chen [31] for the same Gaussian model with covariance matrix $\mathbf{\Sigma}_1$ in the setting for Table 1 for the identity test. We considered $n = 20, 40, 80$ and $p = 20, 60, 100, 160, 200, 300$ in the simulations. Figure 2 reports the empirical proportion that the selected banding width k_{qc} agreed with the true value, which was $\ell = 2$, and otherwise. The performance of the banding width selection algorithm was satisfactory. Figure 2 shows that even when n was small at 20, the algorithm could still identify the true banding width with sufficient probability, and the precision improved as the dimension p was increased. When n was 40, the proportion of correct selection started to be close to 100 percent. The corresponding empirical powers of the proposed identity test using k_{qc} were close to one for most combinations of n and p. Hence, this numerical study provided support to using k_{qc} for the regularised tests for high dimensional covariance matrices.

5. Numerical results

5.1. Simulation studies

We report results of simulation experiments which were designed to evaluate the performance of the proposed tests for identity and sphericity of high dimensional $\mathbf{\Sigma}$. To demonstrate the improvement in power of the proposed tests, we also experimented two existing high dimensional tests: the one by Chen et al. [14] (CZZ hereafter) and the test by Ledoit and Wolf [27] (LW hereafter). The LW test is applicable for Gaussian data only. The test statistics of CZZ and LW tests include all the components of the sample covariance matrix in their formulation and were expected to have lower power as the test statistics can bear larger variation.

We considered two types of innovations for generating data according to Assumption 2: the Gaussian innovation where $\mathbf{Z}_i$ were $\mathcal{N}(0_m, \mathbf{I}_m)$ distributed; and the Gamma innovation where $\mathbf{Z}_i$ had m independent and identically distributed components, and each component was the standardized Gamma random variable with parameters 4 and 0.5 such that it had zero mean and unit variance. In the simulations for the sphericity test, the null hypothesis was $H_0 : \mathbf{\Sigma} = \sqrt{2}\mathbf{I}_p$.

When evaluating the powers of the tests for both identity and sphericity hypotheses, we considered the following three forms of $\mathbf{\Sigma}$.

- Diagonal Form: Set $\mathbf{\Sigma} = \text{diag}(4 \times 1_{[vp]}, 2 \times 1_{p-[vp]}])$ where $[x]$ denotes the integer truncation of x and v characterizes the sparsity of the signals. We chose $v = 0.05$ and 0.1.
- Banded Form: $\mathbf{\Sigma} = (\sigma_{ij})_{1\le i,j\le p}$ with $\sigma_{ij} = \rho^{|i-j|}\text{I}\{|i-j| \le 1\}$ and $\rho = 0.10$.
- Bandable Form: Take $\mathbf{\Sigma} = (\sigma_{ij})_{1\le i,j\le p}$ with $\sigma_{ij} = \text{I}(i = j) + \theta|i-j|^{-\rho}\text{I}(i \ne j)$ with $\theta = 1.2$ and $\rho = 0.16$, which prescribed a polynomial decay as considered by Qiu and Chen [31].

To mimic the large p, small n scenario, we set $n = 20, 40, 60$ and 80, and for each n let $p = 38, 55, 89, 159, 181, 331, 343, 642$, also set $m = p$ when generating the data according to Assumption 2. All the simulation results were based on 1000 iterations with nominal significance level at 5%.

The parameter k was determined by $k = k_{qc} + 1$ where k_{qc} was the banding width by the method of Qiu and Chen [31]. We add 1 to k_{qc} was to ensure the banding estimator $B_k(\hat{\Sigma})$ contains enough signals.

Table 2: Empirical sizes of the proposed test for the identity hypothesis (denoted by IT), along with those of the tests by Chen et al. [14] (CZZ) and Ledoit and Wolf [27] (LW) at 5% nominal significance.

	IT	CZZ $n = 20$	LW	IT	CZZ $n = 40$	LW	IT	CZZ $n = 60$	LW	IT	CZZ $n = 80$	LW
p	Gaussian distributed innovation											
38	0.077	0.075	0.059	0.072	0.057	0.064	0.057	0.056	0.051	0.062	0.064	0.053
55	0.080	0.058	0.059	0.056	0.053	0.047	0.052	0.046	0.058	0.061	0.066	0.063
89	0.076	0.084	0.057	0.072	0.057	0.045	0.063	0.071	0.045	0.072	0.071	0.048
159	0.073	0.065	0.057	0.064	0.062	0.049	0.060	0.057	0.049	0.068	0.056	0.050
181	0.077	0.087	0.049	0.057	0.057	0.050	0.046	0.056	0.056	0.053	0.051	0.053
331	0.081	0.071	0.063	0.066	0.055	0.061	0.060	0.051	0.057	0.060	0.054	0.062
343	0.069	0.072	0.054	0.062	0.065	0.053	0.056	0.056	0.051	0.060	0.042	0.044
642	0.074	0.071	0.054	0.056	0.060	0.055	0.056	0.059	0.053	0.057	0.045	0.045
	Gamma distributed innovation											
38	0.094	0.081	0.198	0.078	0.071	0.209	0.055	0.067	0.221	0.052	0.059	0.199
55	0.081	0.074	0.182	0.067	0.072	0.212	0.065	0.060	0.192	0.058	0.062	0.191
89	0.085	0.080	0.213	0.066	0.055	0.195	0.059	0.051	0.195	0.066	0.065	0.185
159	0.080	0.073	0.168	0.075	0.061	0.178	0.067	0.064	0.207	0.056	0.046	0.189
181	0.087	0.081	0.176	0.071	0.059	0.200	0.059	0.056	0.212	0.062	0.052	0.228
331	0.088	0.071	0.198	0.067	0.074	0.193	0.059	0.073	0.182	0.060	0.049	0.188
343	0.083	0.079	0.182	0.063	0.064	0.198	0.073	0.065	0.172	0.062	0.063	0.189
642	0.076	0.065	0.183	0.074	0.056	0.173	0.061	0.060	0.184	0.056	0.065	0.178

Table 2 displays the empirical sizes of the proposed identity test, the CZZ and LW tests for testing (1.3). It shows that for the Gaussian data the LW test maintained the size better than the other two tests as the LW test was designed for Gaussian data. It is observed that as the sample size was increased, both the proposed identity test and the CZZ test had empirical sizes approaching to the nominal significance level. For the Gamma distributed innovation, as displayed in Table 2, the proposed identity test and the CZZ test had the empirical sizes close to the nominal level, while the LW test failed to control the size. The empirical sizes of the three tests for testing the sphericity hypothesis are reported in the supplementary materials Chen et al. [15], which were quite similar to those of the identity tests in Table 2.

To compare the powers of the tests, we considered $n = 40, 60, 80$ for p set as above. For the Gamma distributed data, only the proposed identity and the CZZ tests were considered since the LW test was no longer applicable. Figures 3 to 5 display the empirical powers of the proposed identity test, the CZZ and LW tests. It was very clear that the proposed identity test outperformed

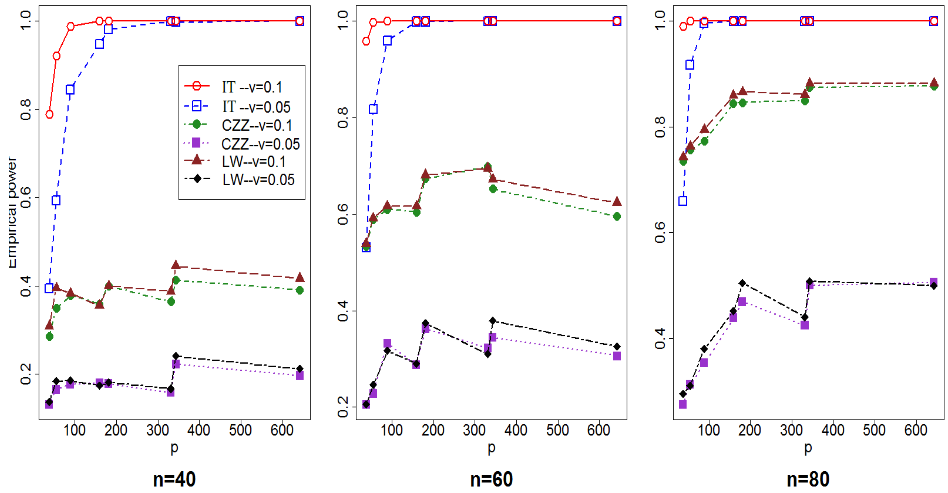


Figure 3: Empirical powers against the alternative in the diagonal form, with $v = 0.05$ and $v = 0.1$ respectively, of the proposed test for the identity hypothesis (IT), along with those of the tests by Chen et al. [14] (CZZ) and Ledoit and Wolf [27] (LW) at 5% nominal significance with the Gaussian distributed innovation.

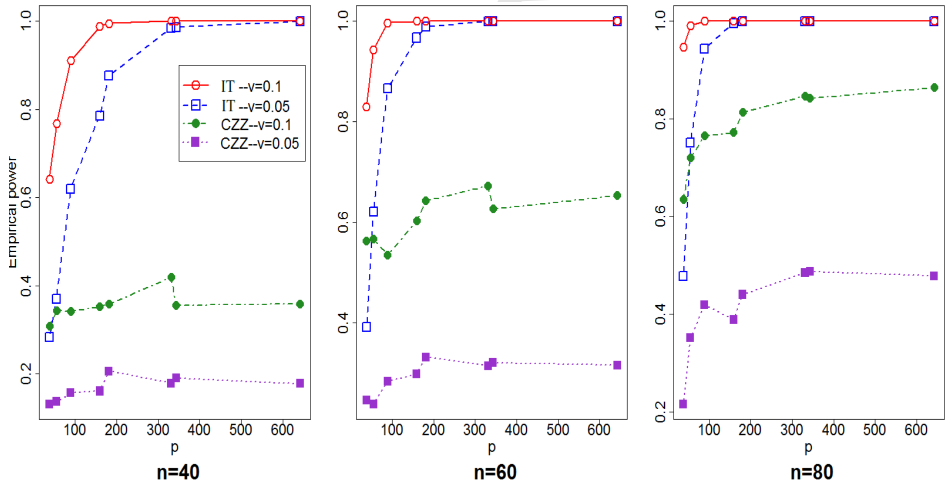
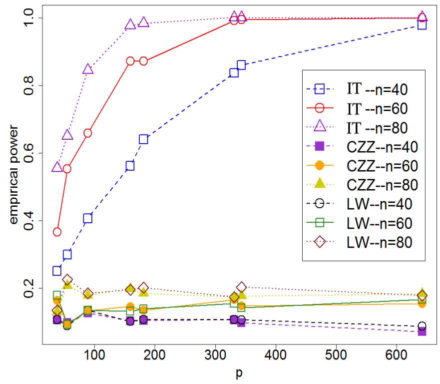
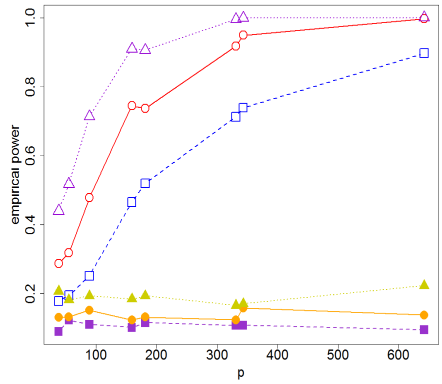


Figure 4: Empirical powers against the alternative in the diagonal form, with $v = 0.05$ and $v = 0.1$ respectively, of the proposed test for the identity hypothesis (IT), along with those of the tests by Chen et al. [14] (CZZ) at 5% nominal significance with the Gamma distributed innovation.

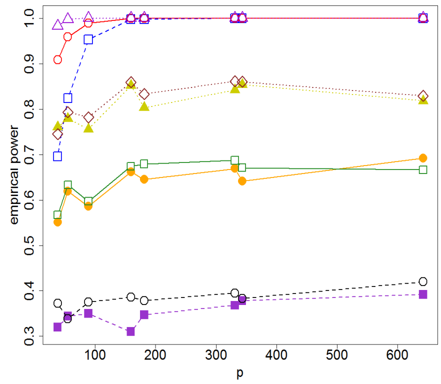
the other two tests for both data generating distributions and the covariance models considered in the simulation. For the alternative $\mathbf{\Sigma}$ with the diagonal form (in Figures 3 and 4), the powers of all three tests were improved for all tests when v is large as expected; the LW and CZZ tests have comparable powers for the Gaussian data (Figure 3). Figure 5 reports the empirical power for the bandable and banded alternative. It shows that the proposed identity test outperformed the other two tests. For the alternative in the bandable form, neither the CZZ test nor the LW test gained extra powers as p was increased, while the proposed identity test had its power increased as p was increased as shown in panels (a) and (b) of Figure 5). As n was increased, all three tests gained



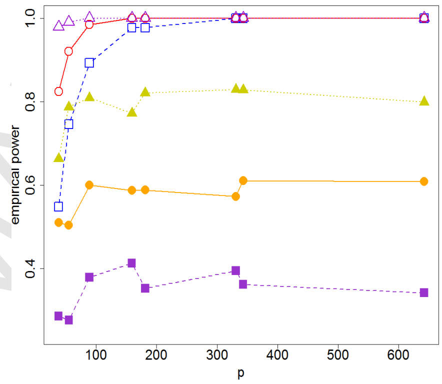
(a) Bandable alternative with Gaussian innovations.



(b) Bandable alternative with Gaussian innovations.



(c) Banded alternative with Gaussian innovations.



(d) Banded alternative with Gamma innovations.

Figure 5: Empirical powers against the alternatives in either the bandable or banded forms of the proposed test for the identity hypothesis (IT), along with those of the tests by Chen et al. [14] (CZZ) and Ledoit and Wolf [27] (LW) at 5% nominal significance with either the Gaussian or Gamma distributed innovation. Different sample sizes are compared ($n = 40, 60, 80$).

powers as expected.

Given the slow decay rate of off-diagonal entries, the alternative in the banded form had a relatively larger banding width than other two types of alternatives. For the banded alternative, all tests gain powers in growing p or n and the CZZ test has power approaching to the proposed identity test as n increasing (panels (c) and (d) in Figure 5).

The power performance of the proposed test for the sphericity along with the CZZ and LW tests are reported in Figures 6 and 7, which suggest that the proposed sphericity test was much more powerful than the other two tests under the three forms of the alternatives $\mathbf{\Sigma}$. More simulation results are reported in the supplementary material, Chen et al. [15].

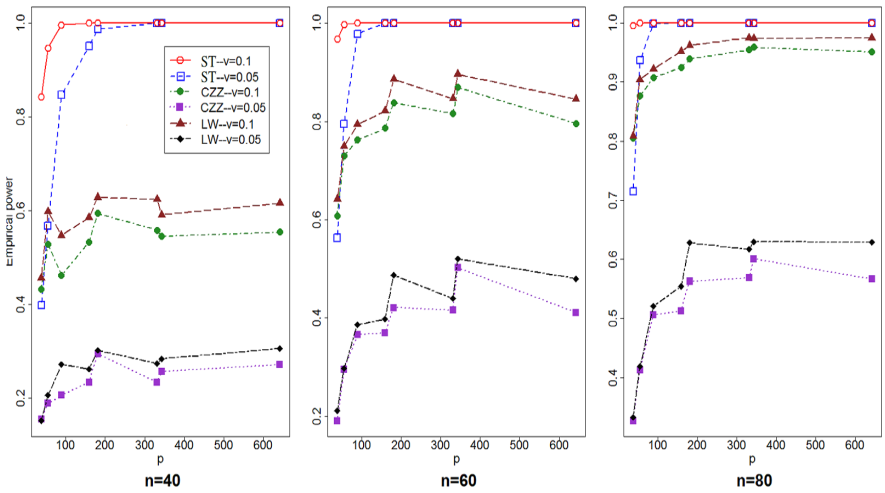


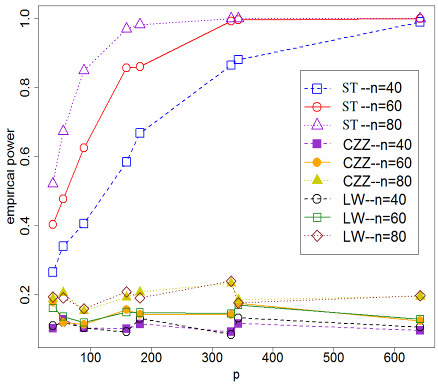
Figure 6: Empirical powers against the alternative in the diagonal form, with $v = 0.05$ and $v = 0.1$ respectively, of the proposed test for the sphericity hypothesis (ST), along with those of the tests by Chen et al. [14] (CZZ) and Ledoit and Wolf [27] (LW) at 5% nominal significance with the Gaussian distributed innovation.

5.2. Empirical study

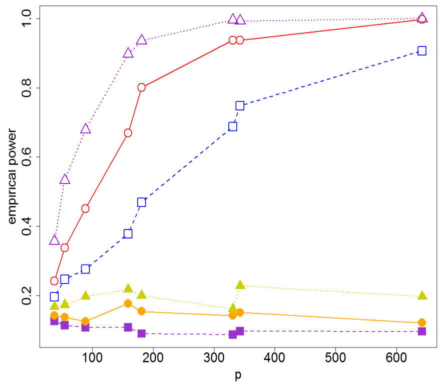
We analyzed an acute lymphoblastic leukemia (ALL) data reported in Chiaretti et al. [16] to demonstrate the proposed regularized tests for identity and sphericity. The data contain microarray expressions for patients having acute lymphoblastic leukemia of either T-lymphocyte type or B-lymphocyte type. We focused on the sub-sample of B-lymphocyte type leukemia in this analysis. The 78 patients of B-lymphocyte type leukemia were classified into two groups: the BCR/ABL fusion (36 patients) and cytogenetically normal NEG (42 patients). The original dataset has been analyzed by Chen and Qin [13], Chiaretti et al. [16], and Dudoit et al. [18], using different methodologies.

Our analysis is to study the covariance structures for sets of genes defined within the gene ontology (GO) framework. It is known that genes tend to work collectively to achieve certain biological tasks, which gave rise to the identification of gene-sets (also called GO terms) with respect to three broader categories of biological functions: biological processes (BP), cellular components (CC) and molecular functions (MF). The gene-sets are technically defined in the gene ontology (GO) system via structured vocabularies which produce unique name for a gene-set. After a preliminary screening with the gene-filtering approach advocated in Gentleman et al. [21], there left 2694 unique gene-sets in the BP category, 352 in the CC category and 419 in the MF category for the ALL data. The largest gene-set had 3048, 3140 and 303 genes in BP, CC and MF, respectively.

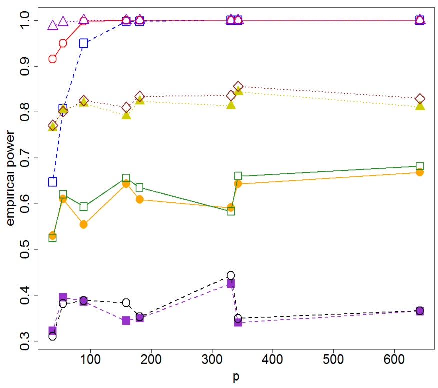
Our aim was to study the dependence structures in the expression levels of gene-sets between the BCR/ABL and NEG groups for each of the three functional categories by testing hypotheses (1.3) and (1.4) for appropriately transformed data. The procedure is described as following. To attain



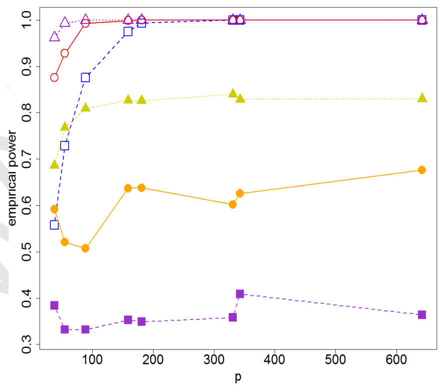
(a) Bandable alternative with Gaussian innovations.



(b) Bandable alternative with Gaussian innovations.



(c) Banded alternative with Gaussian innovations.



(d) Banded alternative with Gamma innovations.

Figure 7: Empirical powers against the alternatives in either the bandable or banded forms of the proposed test for the sphericity hypothesis (ST), along with those of the tests by Chen et al. [14] (CZZ) and Ledoit and Wolf [27] (LW) at 5% nominal significance with either the Gaussian or Gamma distributed innovation. Different sample sizes are compared ($n = 40, 60, 80$).

bandable covariance structure so that we can apply the proposed tests, we employed the re-ordering algorithm in Friendly [20] to each gene-set in both the NEG and the BCR/ABL fusion groups to obtain a permutation of the genes in that gene-set so that the covariance was more bandable. As a demonstration of this data re-ordering algorithm, we compare the heat maps of the correlation matrices before and after the re-ordering based on samples from the NEG group for the gene-set GO:0000086 (G2/M transition of mitotic cell cycle), which has 63 genes, in the supplementary materials, Chen et al. [15].

For a gene-set of a functional category with the NEG sample, say the g-th gene-set, its covariance $\mathbf{\Sigma}_{\text{neg},g}$ was estimated by using the banding estimator, denoted as $\widehat{\mathbf{\Sigma}}_{\text{neg},g}$ with the banding width determined by the method in Qiu and Chen [31]. And $\widehat{\mathbf{\Sigma}}_{\text{neg},g}^{-1/2}$ was used to transform the same gene-set in the BCR/ABL group. For each transformed gene-set in the BCR/ABL group, we tested the hypotheses (1.3) and (1.4) using the proposed regularized identity and sphericity tests when $p > 10$ and use the tests by John [25, 26] for those gene-sets with smaller dimensions. For the proposed tests, k was chosen as $k_{qc} + 1$. We essentially tested hypotheses $H_0 : \mathbf{\Sigma}_{\text{neg},g} = \mathbf{\Sigma}_{\text{BCR/ABL},g}$ or $\mathbf{\Sigma}_{\text{neg},g} = \sigma_g^2 \mathbf{\Sigma}_{\text{BCR/ABL},g}$ for some $\sigma_g^2 > 0$ for each $g = 1, \ldots, \mathcal{G}$ where $\mathcal{G}$ is the total number of gene-sets in the category. The test of Chen et al. [14] was also performed to serve as a comparison.

By controlling the false discovery rate (FDR) (Benjamini and Hochberg [6]) at 0.001, we have identified gene-sets that have significantly different covariance structures between the NEG and BCR/ABL groups. Table 3 provides a broad classification for the gene-sets identified by the proposed test and the test of Chen et al. [14]. The table shows that the dependence structure between NEG and BCR/ABL were largely different with quite a large number of significantly differential expressed gene-sets, which was due to a large number of very small p-values (Figure S3 in Chen et al. [15]). Biologically speaking, the NEG and BCR/ABL cases have different genetic mechanisms that cytogenetically normal leukemia is not associated with large chromosomal abnormalities while the BCR/ABL leukemia involves fusion of BCR and ABL genes in Philadelphia chromosome (Pakakasama et al. [29]).

Table 3 reveals that the proposed tests identified more gene-sets than the CZZ test. It is interesting to notice that the proposed sphericity test has identified GO:0004527 and GO:0004869 as diseases-associated gene-sets in the MF category while they were missed by the CZZ test, and biologically these two gene-sets correspond to exonuclease activity and endopeptidase inhibitor activity, which have been recognized associated to the disease development of different types of leukemia recently (Shi et al. [34], Tsakou et al. [40]).

Table 3: Number of identified gene-sets against null hypotheses by controlling FDR at 0.001. ST and IT stand for the proposed sphericity and identity tests, respectively. The last column reports the means and standard deviations for k_{qc}.

GO Category	Total	Sphericity hypothesis (1.4)			Identity hypothesis (1.3)			k after re-ordering
		ST only	Both	CZZ only	IT only	Both	CZZ only	
BP	2694	21	2338	0	19	2450	1	(14.11, 10.03)
CC	352	0	317	0	4	325	2	(14.57, 9.51)
MF	418	7	363	1	8	372	2	(13.57, 9.33)

6. Discussion

In this paper, we introduced two powerful tests for the identity and sphericity hypotheses of large covariance matrices and showed that the proposed testing procedures perform particularly well against sparse alternatives from some particular classes. The proposed tests leverage the sparsity information of the alternatives in high dimensional settings that leads to significant reduction in the variance of the test statistics. The theoretical properties of the proposed tests were established. We also explored how the proposed tests improve the powers comparing to the test by Chen et al. [14]. Furthermore, we discussed the selection of k for the proposed tests in practice. Finally, we examined the proposed tests by numerical studies and illustrated its applications in real data analysis.

The forms of the identity hypothesis that $\mathbf{\Sigma} = \mathrm{I}_p$ and the spherecity hypothesis that $\mathbf{\Sigma} = \sigma^2 \mathrm{I}_p$ are two idealized hypotheses. Despite being idealized, they play central roles in either the classical multivariate analysis where the dimension of data p is fixed or the high dimensional multivariate analysis where p is diverging and can be larger than the sample size as treated in this paper. The literature of the classical multivariate analysis include those of Anderson [1], John [25, 26] and Nagao [28], while the contemporary literature includes Chen et al. [14] and Ledoit and Wolf [27] among others.

The identity hypothesis in (1.3) actually covers the hypothesis $H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0$ for a known invertible covariance matrix $\mathbf{\Sigma}_0$. By transforming the data via left multiplying $\mathbf{\Sigma}_0^{-1/2}$, the identity hypothesis can be carried out for the transformed data. The spherecity hypothesis can be treated similarly. In practice, the hypothesed $\mathbf{\Sigma}_0$ can be postulated based on the empirical estimates of $\mathbf{\Sigma}$. This was what we have done in the case study by first permuting the data to rearrange the data components so that those with high correlations are grouped closer than those with low correlation. After the permutation, we employed the banding estimator of the high dimensional covariance matrix of Bickel and Levina [7] to attain a banded form for $\mathbf{\Sigma}$, which are used to standardize the data. We are aware that using the estimated $\mathbf{\Sigma}_0$ would introduce issues of inference as the estimation error may affect the asymptotic distribution. While this would not be a big issue in the classical fixed dimensional context, it would be an issue when p diverges. We would consider this issue in a future study.

Another issue is a concern on the computational costs required in carrying out the proposed test procedures. The main cost of computation is in computing the raw test statistics and in estimating their variance. To gain information on the computation time needed, we report in Table 4 the time needed to accomplish a single spherecity test under the alternative bandable form $\mathbf{\Sigma}$ in the simulation study for different n and p on a PC with Intel(R) Core(MT) i7-4790K processor and CPU speed of 4.0GHz. The computing times in the table suggest that the computation burden for carrying out the test procedure is manageable even for relatively larger values of n and p with a quite standard computational capacity.

Table 4: Computational time (in seconds) on a PC equipped with Intel(R) Core(MT) i7-4790K and the CPU of 4.0GHz for carrying out a single spherecity test under the alternative $\mathbf{\Sigma} = (\sigma_{ij})_{1\le i,j\le p}$ with $\sigma_{ij} = \mathrm{I}(i=j) + \theta|i-j|^{-\rho}\mathrm{I}(i\neq j)$ with different n and p.

		p							
		38	55	89	159	181	331	343	641
	20	0.172	0.354	0.932	3.051	3.961	13.408	14.830	51.652
n	40	0.299	0.633	1.659	5.403	7.023	23.863	25.926	91.187
	60	0.425	0.901	2.399	7.747	10.109	34.104	37.088	131.226
	80	0.550	1.182	3.133	10.119	13.158	44.805	48.658	186.654

Appendix

We begin this appendix by presenting some notation and technical preliminaries that will be used in the proof of the main results. For a matrix $\mathbf{M} = (m_{ij})_{1\le i,j\le p}$, we denote $\lambda(\mathbf{M})$ the eigenvalues of $\mathbf{M}$ with $\lambda_{\min}(\mathbf{M})$ and $\lambda_{\max}(\mathbf{M})$ the smallest and largest eigenvalues of $\mathbf{M}$ respectively, and denote the matrix norms by $\|\mathbf{M}\|_1 \equiv \max_j \sum_i |m_{ij}|$, $\|\mathbf{M}\| \equiv \left\{\lambda_{\max}(\mathbf{M}^\top\mathbf{M})\right\}^{1/2}$, and $\|\mathbf{M}\|_\infty \equiv \max_i \sum_j |m_{ij}|$. For symmetric $\mathbf{\Sigma} \in \mathcal{U}(\varepsilon_0, \mathrm{C}, \alpha)$, we have the following properties

(1) $\|\mathbf{\Sigma}\| \le 1/\varepsilon_0$ and $\|\mathrm{B}_k(\mathbf{\Sigma}) - \mathbf{\Sigma}\|_1 \le \mathrm{C}k^{-\alpha}$;

(2) $\|\mathrm{B}_k(\mathbf{\Sigma})\| \le 1/\varepsilon_0 + \mathrm{C}k^{-\alpha}$;

(3) $|\lambda_{\min}\{\mathrm{B}_k(\mathbf{\Sigma})\} - \lambda_{\min}(\mathbf{\Sigma})| \le \mathrm{C}k^{-\alpha}$; and

(4) there exist $C_1, C_2 > 0$ such that $C_1 \le \min_i \sigma_{ii} \le \max_i \sigma_{ii} \le C_2$ for $\mathbf{\Sigma} = (\sigma_{ij})_{1\le i,j\le p}$.

Properties (2) and (3) imply that for sufficiently large k, there exits a positive constant δ_0 such that

$$0 < \delta_0 \le \lambda_{\min}\{\mathrm{B}_k(\mathbf{\Sigma})\} \le \lambda_{\max}\{\mathrm{B}_k(\mathbf{\Sigma})\} \le 1/\delta_0,$$

which means $\mathrm{tr}(\mathbf{\Sigma}) = O(p)$ and $\mathrm{tr}\{\mathrm{B}_k(\mathbf{\Sigma})\} = O(p)$ for sufficiently large k. In addition, some algebraic computations yield following useful results in remaining derivations:

(i) $\sum_{|i_1-j_1|\le k}\sum_{i_2=1}^{p} \sigma_{i_1j_1} f_{i_1j_1i_2i_2}(\mathbf{\Sigma}) = \mathrm{tr}\left[\left\{\mathbf{\Gamma}^\top \mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Gamma}\right\} \circ (\mathbf{\Gamma}^\top\mathbf{\Gamma})\right] \le \mathrm{tr}\{\mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Sigma}^2\}$;

(ii) $\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k} \sigma_{i_1j_1}\sigma_{i_2j_2} f_{i_1j_1i_2j_2}(\mathbf{\Sigma}) = \mathrm{tr}\left[\left\{\mathbf{\Gamma}^\top \mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Gamma}\right\} \circ \left\{\mathbf{\Gamma}^\top \mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Gamma}\right\}\right] \le \mathrm{tr}\left[\{\mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Sigma}\}^2\right]$;

(iii) $\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k} \sigma_{i_1i_2}^2\sigma_{j_1j_2}^2 \le (2k+1)^2\mathrm{tr}\{(\mathbf{\Sigma}\circ\mathbf{\Sigma})^2\}$, $\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k} \sigma_{i_1i_2}\sigma_{j_1j_2}\sigma_{i_1j_2}\sigma_{j_1i_2} \le (2k+1)^2\mathrm{tr}\{(\mathbf{\Sigma}\circ\mathbf{\Sigma})^2\}$;

(v) $\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k} \sigma_{i_1i_2}\sigma_{j_1j_2} f_{i_1j_1i_2j_2}(\mathbf{\Sigma}) \le (2k+1)^2\mathrm{tr}\{(\mathbf{\Gamma}\circ\mathbf{\Gamma})(\mathbf{\Gamma}\circ\mathbf{\Gamma})^\top(\mathbf{\Sigma}\circ\mathbf{\Sigma})\}$; and,

(vi) $\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k} f_{i_1j_1i_2j_2}^2(\mathbf{\Sigma}) \le (2k+1)^2\mathrm{tr}\left[\left\{(\mathbf{\Gamma}\circ\mathbf{\Gamma})(\mathbf{\Gamma}\circ\mathbf{\Gamma})^\top\right\}^2\right]$.

A.1. Critical lemmas

In this section, we collect some technical lemmas that will be used in the proofs. The first lemma provides algebraic representations of the variances and covariances of $\mathrm{L}_{n_1}, \mathrm{L}_{n_2}, \mathrm{L}_{n_3}, \mathrm{L}_{n_4}$ and L_{n_5}.

Lemma A.1. *Under Assumptions 1 and 2, for $\tau^2_{n,k}(\mathbf{\Sigma})$ defined in (3.3) and $n^* = n(n-1)(n-2)(n-3)$*

$$\mathrm{var}\left\{\sum_{|i-j|\le k} \mathrm{L}_{n_1}(i,j)\right\} = \tau^2_{n,k}(\mathbf{\Sigma}) + 8n^{-1}\mathrm{tr}\left[\{\mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Sigma}\}^2\right] + 4\Delta n^{-1}\mathrm{tr}\left[\left\{\mathbf{\Gamma}^\top \mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Gamma}\right\}\circ\left\{\mathbf{\Gamma}^\top \mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Gamma}\right\}\right],$$

$$\mathrm{var}\left\{\sum_{|i-j|\le k} \mathrm{L}_{n_2}(i,j)\right\} = \frac{2}{n(n-1)}\mathrm{tr}\left\{\mathrm{B}_k(\mathbf{\Sigma})\Sigma \mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Sigma}\right\} + \frac{1}{n(n-1)(n-2)}\sum_{|i_1-j_1|\le k}$$

$$\sum_{|i_2-j_2|\le k}\left\{(\sigma_{i_1i_2}\sigma_{j_1j_2}+\sigma_{i_1j_2}\sigma_{j_1i_2})^2 + \Delta(\sigma_{i_1i_2}\sigma_{j_1j_2}+\sigma_{i_1j_2}\sigma_{j_1i_2})f_{i_1j_1i_2j_2}(\mathbf{\Sigma})\right\},$$

$$\mathrm{var}\left\{\sum_{|i-j|\le k} \mathrm{L}_{n_3}(i,j)\right\} = \frac{8}{n^*}\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k}\sigma^2_{i_1i_2}\sigma^2_{j_1j_2} + \frac{16}{n^*}\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k}\sigma_{i_1i_2}\sigma_{i_1j_2}\sigma_{j_1i_2}\sigma_{j_1j_2},$$

$$\mathrm{var}\left\{\sum_{i=1}^p \mathrm{L}_{n_4}(i)\right\} = 2n^{-1}\mathrm{tr}(\mathbf{\Sigma}^2) + \Delta n^{-1}\mathrm{tr}\left\{\left(\mathbf{\Gamma}^\top\mathbf{\Gamma}\right)\circ\left(\mathbf{\Gamma}^\top\mathbf{\Gamma}\right)\right\},$$

$$\mathrm{var}\left\{\sum_{i=1}^p \mathrm{L}_{n_5}(i)\right\} = 2\{n(n-1)\}^{-1}\mathrm{tr}(\mathbf{\Sigma}^2),$$

and furthermore,

$$\mathrm{cov}\left\{\sum_{|i-j|\le k} \mathrm{L}_{n_1}(i,j), \sum_{i=1}^p \mathrm{L}_{n_4}(i)\right\} = 4n^{-1}\mathrm{tr}\left\{\mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Sigma}^2\right\} + 2\Delta n^{-1}\mathrm{tr}\left[\left\{\mathbf{\Gamma}^\top\mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Gamma}\right\}\circ(\mathbf{\Gamma}^\top\mathbf{\Gamma})\right],$$

$$\mathrm{cov}\left\{\sum_{i=1}^p \mathrm{L}_{n_4}(i), \sum_{i=1}^p \mathrm{L}_{n_5}(i)\right\} = 0.$$

The proof of Lemma A.1 is based on standard yet tedious computations that we omit here. Lemma A.1 implies that $\mathrm{var}\{\sum_{|i-j|\le k}\mathrm{L}_{n_3}(i,j)\} = O\left(n^{-2}\mathrm{var}\{\sum_{|i-j|\le k}\mathrm{L}_{n_1}(i,j)\}\right)$, and under the assumption that $k\to\infty$ and $k = o(\min(n^{1/2}, p^{1/2}))$, it yields

$$\mathrm{var}\left\{\sum_{|i-j|\le k}\mathrm{L}_{n_2}(i,j)\right\} = o\left(\mathrm{var}\left\{\sum_{|i-j|\le k}\mathrm{L}_{n_1}(i,j)\right\}\right)$$

and

$$\mathrm{var}\left\{\sum_{i=1}^p \mathrm{L}_{5,i}\right\} = o\left(\mathrm{var}\left\{\sum_{|i-j|\le k}\mathrm{L}_{n_1}(i,j)\right\}\right).$$

The next lemma is on the asymptotic normality of statistics $T_{n,k}$.

Lemma A.2. *Under Assumptions 1 and 2, for any real sequences* $\{a_n\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$, $T_{n,k} = a_n \sum_{|i-j|\le k} \mathrm{L}_{n_1}(i,j) + b_n \sum_{i=1}^{p} \mathrm{L}_{n_4}(i)$ *satisfies*

$$\{\mathrm{var}(T_{n,k})\}^{-1/2}\{T_{n,k} - \mathrm{E}(T_{n,k})\} \to \mathcal{N}(0,1) \tag{A.1}$$

in distribution provided $k = o(min(n^{1/2}, p^{1/2}))$.

Proof. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_t = \sigma\{\mathbf{X}_1, \dots, \mathbf{X}_t\}$ for $t = 1, \dots, n$ be the sequence of σ-fields generated by data, and denote $\mathrm{E}_t(\cdot) \equiv \mathrm{E}(\cdot|\mathcal{F}_t)$ and $\mathrm{E}(\cdot) \equiv \mathrm{E}_0(\cdot)$. Write $T_{n,k} - \mathrm{E}(T_{n,k}) = \sum_{t=1}^{n} \mathrm{D}_{t,k}$, where $\mathrm{D}_{t,k} = \mathrm{E}_t(T_{n,k}) - \mathrm{E}_{t-1}(T_{n,k})$. It is easy to see that $\mathrm{D}_{t,k}$ is $\mathcal{F}_t$ measurable and $\mathrm{E}_{t-1}(\mathrm{D}_{t,k}) = 0$ for each $t \ge 1$ so that for every n, $\{\mathrm{D}_{t,k}, \mathcal{F}_t\}_{1\le t\le n}$ is a martingale difference array. By the martingale central limit theorem, (A.1) is straightforward once one can show that, as $n \to \infty$,

$$\frac{\sum_{t=1}^{n} \sigma_{t,k}^2}{\mathrm{var}(T_{n,k})} \xrightarrow{p} 1 \quad \text{and} \quad \frac{\sum_{t=1}^{n} \mathrm{E}(\mathrm{D}_{t,k}^4)}{\mathrm{var}^2(T_{n,k})} \to 0, \tag{A.2}$$

with $\sigma_{t,k}^2 = \mathrm{E}_{t-1}(\mathrm{D}_{t,k}^2)$.

As $\mathrm{E}(\sum_{t=1}^{n} \sigma_{t,k}^2) = \mathrm{var}(T_{n,k})$, it suffices to show that $\mathrm{var}(\sum_{t=1}^{n} \sigma_{t,k}^2) = o(\mathrm{var}^2(T_{n,k}))$ to derive the first part of (A.2). By Lemma A.1,

$$\begin{aligned}\mathrm{var}(T_{n,k}) =& a_n^2 \tau_{n,k}^2(\mathbf{\Sigma}) + 2n^{-1}\mathrm{tr}\left[\{2a_n \mathrm{B}_k(\mathbf{\Sigma}) + b_n \mathbf{I}_p\}\, \mathbf{\Sigma}\right]^2 + \Delta n^{-1}\mathrm{tr}\Big[\left\{\mathbf{\Gamma}^\top (2a_n \mathrm{B}_k(\mathbf{\Sigma}) + b_n \mathbf{I}_p)\, \mathbf{\Gamma}\right\} \\ & \circ \left\{\mathbf{\Gamma}^\top (2a_n \mathrm{B}_k(\mathbf{\Sigma}) + b_n \mathbf{I}_p)\, \mathbf{\Gamma}\right\}\Big].\end{aligned}$$

Also, notice that

$$\begin{aligned}\mathrm{D}_{t,k} =& 2a_n\{n(n-1)\}^{-1}\left[\mathbf{X}_t^\top \mathrm{B}_k(\mathbf{Q}_{t-1}) X_t - \mathrm{tr}\{\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma}\}\right] \\ & + 2a_n n^{-1}\left[\mathbf{X}_t^\top \mathrm{B}_k(\mathbf{\Sigma})\mathbf{X}_t - \mathrm{tr}\{\mathrm{B}_k(\mathbf{\Sigma})\mathbf{\Sigma}\}\right] + b_n n^{-1}\left\{\mathbf{X}_t^\top \mathbf{X}_t - \mathrm{tr}(\mathbf{\Sigma})\right\},\end{aligned}$$

where $\mathbf{Q}_{t-1} = (\mathbf{Q}_{t-1}^{i,j})_{1\le i,j\le p} = \sum_{s=1}^{t-1}(\mathbf{X}_s\mathbf{X}_s^\top - \mathbf{\Sigma})$ with $\mathrm{Q}_{t-1}^{i,j} = \sum_{s=1}^{t-1}(x_{si}x_{sj} - \sigma_{ij})$, so

$$\sum_{t=1}^{n} \sigma_{t,k}^2 = \mathrm{R}_{1,n} + \mathrm{R}_{2,n} + \mathrm{R}_{3,n} + \mathrm{R}_{4,n} + nC,$$

for some constant C and

$$\begin{aligned}
\mathrm{R}_{1,n} =& 8a_n\{n^2(n-1)\}^{-1}\sum_{t=1}^{n}\mathrm{tr}\left[\mathrm{B}_k(\mathbf{Q}_{t-1})\boldsymbol{\Sigma}\left\{2a_n\mathrm{B}_k(\boldsymbol{\Sigma})+b_n\mathbf{I}_p\right\}\boldsymbol{\Sigma}\right],\\
\mathrm{R}_{2,n} =& 4\Delta a_n\{n^2(n-1)\}^{-1}\sum_{t=1}^{n}\mathrm{tr}\left[\left\{\boldsymbol{\Gamma}^\top\mathrm{B}_k(\mathbf{Q}_{t-1})\boldsymbol{\Gamma}\right\}\circ\left\{\boldsymbol{\Gamma}^\top\left(2a_n\mathrm{B}_k(\boldsymbol{\Sigma})+b_n\mathbf{I}_p\right)\boldsymbol{\Gamma}\right\}\right],\\
\mathrm{R}_{3,n} =& 8a_n^2\{n^2(n-1)^2\}^{-1}\sum_{t=1}^{n}\mathrm{tr}\left[\{\mathrm{B}_k(\mathbf{Q}_{t-1})\boldsymbol{\Sigma}\}^2\right],\\
\mathrm{R}_{4,n} =& 4\Delta a_n^2\{n^2(n-1)^2\}^{-1}\sum_{t=1}^{n}\mathrm{tr}\left[\left\{\boldsymbol{\Gamma}^\top\mathrm{B}_k(\mathbf{Q}_{t-1})\boldsymbol{\Gamma}\right\}\circ\left\{\boldsymbol{\Gamma}^\top\mathrm{B}_k(\mathbf{Q}_{t-1})\boldsymbol{\Gamma}\right\}\right],
\end{aligned}$$

It is sufficient to show $\mathrm{var}(\mathrm{R}_{i,n}) = o(\mathrm{var}^2(T_{n,k}))$ for $i = 1, 2, 3, 4$ to obtain $\mathrm{var}(\sum_{t=1}^n \sigma_{t,k}^2) = o(\mathrm{var}^2(T_{n,k}))$.

We first study $\mathrm{R}_{1,n}$. Denote $\boldsymbol{\Omega}_k = \boldsymbol{\Sigma}\left\{2a_n\mathrm{B}_k(\boldsymbol{\Sigma})+b_n\mathbf{I}_p\right\}\boldsymbol{\Sigma}$. Recall the fact that $\mathrm{tr}\{\mathrm{B}_k(\mathbf{A})\mathbf{C}\} = \mathrm{tr}\{\mathbf{A}\mathrm{B}_k(\mathbf{C})\}$ for symmetric matrices $\mathbf{A}$ and $\mathbf{C}$ with conformable sizes, we have

$$\mathrm{tr}\left\{\mathrm{B}_k(\mathbf{Q}_{t-1})\boldsymbol{\Omega}_k\right\} = \sum_{s=1}^{t-1}\left[\mathbf{X}_s^\top\mathrm{B}_k(\boldsymbol{\Omega}_k)\mathbf{X}_s - \mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Omega}_k\}\right].$$

Therefore, for each $s \geq 1$

$$\mathrm{var}\left[\mathbf{X}_s^\top\mathrm{B}_k(\boldsymbol{\Omega}_k)\mathbf{X}_s - \mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Omega}_k\}\right] = 2\mathrm{tr}\{\boldsymbol{\Sigma}\mathrm{B}_k(\boldsymbol{\Omega}_k)\}^2 + \Delta\mathrm{tr}\left[\left\{\boldsymbol{\Gamma}^\top\mathrm{B}_k(\boldsymbol{\Omega})\boldsymbol{\Gamma}\right\}\circ\left\{\boldsymbol{\Gamma}^\top\mathrm{B}_k(\boldsymbol{\Omega})\boldsymbol{\Gamma}\right\}\right].$$

By the algebraic properties summarized before, it holds for some constant $\gamma > 0$ that

$$\mathrm{tr}\{\boldsymbol{\Sigma}\mathrm{B}_k(\boldsymbol{\Omega}_k)\}^2 \leq \gamma\mathrm{tr}[\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}^2]\mathrm{tr}\left[\left\{2a_n\mathrm{B}_k(\boldsymbol{\Sigma})+b_n\mathbf{I}_p\right\}\boldsymbol{\Sigma}\right]^2 = o\left\{n^3\mathrm{var}^2(T_{n,k})\right\},$$

and also $\mathrm{tr}\left[\left\{\boldsymbol{\Gamma}^\top\mathrm{B}_k(\boldsymbol{\Omega}_k)\boldsymbol{\Gamma}\right\}\circ\left\{\boldsymbol{\Gamma}^\top\mathrm{B}_k(\boldsymbol{\Omega}_k)\boldsymbol{\Gamma}\right\}\right] \leq \mathrm{tr}\{\boldsymbol{\Sigma}\mathrm{B}_k(\boldsymbol{\Omega}_k)\}^2$. We therefore conclude that there exist constant $C > 0$ such that

$$\mathrm{var}(\mathrm{R}_{1,n}) \leq Cn^{-3}\mathrm{var}\left[\mathrm{tr}\left\{\mathrm{B}_k(\mathbf{Q}_{t-1})\boldsymbol{\Omega}_k\right\}\right] = o(\mathrm{var}^2(T_{n,k})).$$

For $\mathrm{R}_{3,n}$, denote for any $1 \leq s_1, s_2 \leq n$,

$$\mathrm{Y}_{s_1s_2} = \sum_{|i_1-j_1|\leq k}\sum_{|i_2-j_2|\leq k}\sigma_{i_1i_2}\sigma_{j_1j_2}(x_{s_1i_1}x_{s_1j_1}-\sigma_{i_1j_1})(x_{s_2i_2}x_{s_2j_2}-\sigma_{i_2j_2}).$$

Then

$$\mathrm{tr}\left[\{\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma}\}^2\right] = \sum_{s=1}^{t-1} \mathrm{Y}_{ss} + \sum_{s_1 \neq s_2} \mathrm{Y}_{s_1 s_2}.$$

Notice that $\mathrm{E}(\mathrm{Y}_{s_1 s_2}) = 0$ for any $s_1 \neq s_2$ and $\mathrm{E}(\mathrm{Y}_{s_1 s_2}\mathrm{Y}_{s_3 s_4}) = 0$ for any (s_1, s_2, s_3, s_4) except $s_1 = s_2 = s_3 = s_4$, $s_1 = s_3$ and $s_2 = s_4$ or $s_1 = s_4$ and $s_2 = s_3$. Hence, for any $t \leq l$

$$\mathrm{cov}\left(\mathrm{tr}\left[\{\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma}\}^2\right], \mathrm{tr}\left[\{\mathrm{B}_k(\mathbf{Q}_{l-1})\mathbf{\Sigma}\}^2\right]\right) = (t-1)\mathrm{var}(\mathrm{Y}_{11}) + 2(t-1)(t-2)\mathrm{var}(\mathrm{Y}_{12}).$$

It is therefore sufficiently to show $\mathrm{var}(\mathrm{Y}_{11}) = o\left(n^5 \mathrm{var}^2(T_{n,k})\right)$ and $\mathrm{var}(\mathrm{Y}_{12}) = o\left(n^4 \mathrm{var}^2(T_{n,k})\right)$. As $k = o(n^{1/2})$, then for constant $C_1 > 0$

$$\begin{aligned}\mathrm{E}(\mathrm{Y}_{11}^2) = &\sum_{|i_1-j_1|\leq k} \sum_{|i_2-j_2|\leq k} \sum_{|i_3-j_3|\leq k} \sum_{|i_4-j_4|\leq k} \Big[\sigma_{i_1 i_2}\sigma_{j_1 j_2}\sigma_{i_3 i_4}\sigma_{j_3 j_4} \\ &\times \mathrm{E}\{(x_{1i_1}x_{1j_1} - \sigma_{i_1 j_1})(x_{1i_2}x_{1j_2} - \sigma_{i_2 j_2})(x_{1i_3}x_{1j_3} - \sigma_{i_3 j_3})(x_{1i_4}x_{1j_4} - \sigma_{i_4 j_4})\}\Big] \\ \leq & C_1(2k+1)^4 \mathrm{tr}^2(\mathbf{\Sigma}^4) = o\left(n^5 \mathrm{var}^2(T_{n,k})\right),\end{aligned}$$

so that $\mathrm{var}(\mathrm{Y}_{11}) = o\left\{n^5 \mathrm{var}^2(T_{n,k})\right\}$. Similarly, for some constant $C_2 > 0$

$$\begin{aligned}\mathrm{var}(\mathrm{Y}_{12}) = &\sum_{|i_1-j_1|\leq k} \sum_{|i_2-j_2|\leq k} \sum_{|i_3-j_3|\leq k} \sum_{|i_4-j_4|\leq k} \Big[\sigma_{i_1 i_2}\sigma_{j_1 j_2}\sigma_{i_3 i_4}\sigma_{j_3 j_4} \\ &\times \{\sigma_{i_1 i_3}\sigma_{j_1 j_3} + \sigma_{i_1 j_3}\sigma_{j_1 i_3} + \Delta f_{i_1 j_1 i_3 j_3}(\mathbf{\Sigma})\}\{\sigma_{i_2 i_4}\sigma_{j_2 j_4} + \sigma_{i_2 j_4}\sigma_{j_2 i_4} + \Delta f_{i_2 j_2 i_4 j_4}(\mathbf{\Sigma})\}\Big] \\ \leq & C_2(2k+1)^4 \mathrm{tr}(\mathbf{\Sigma}^8) = o\left(n^4 \mathrm{var}^2(T_{n,k})\right).\end{aligned}$$

Thus,

$$\mathrm{var}\left[\mathrm{tr}\{(\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma})^2\}\right] = t^2 o\left(n^4 \mathrm{var}^2(T_{n,k})\right),$$

which implies that for some positive constants γ, γ_1 and γ_2

$$\begin{aligned}\mathrm{var}(\mathrm{R}_{3,n}) \leq \gamma n^{-8} \mathrm{var}\left[\sum_{t=1}^{n} \mathrm{tr}\left\{(\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma})^2\right\}\right] &\leq \gamma_1 n^{-5}\mathrm{var}(\mathrm{Y}_{11}) + \gamma_2 n^{-4}\mathrm{var}(\mathrm{Y}_{12}) \\ &= o(\mathrm{var}^2(T_{n,k})).\end{aligned}$$

Likewise, we can show that $\mathrm{var}(\mathrm{R}_{i,n}) = o(\mathrm{var}^2(T_{n,k}))$ for $i = 2, 4$, by which we obtain the first part of (A.2) using the concentration property.

It remains to show the second part of (A.2). By standard algebraic computations, one can show that $\mathrm{D}_{t,k} = \mathrm{M}_{t,1} + \mathrm{M}_{t,2}$ where

$$\begin{aligned}\mathrm{M}_{t,1} =& 2a_n\{n(n-1)\}^{-1}\left[\mathbf{X}_t^\top \mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{X}_t - \mathrm{tr}\{\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma}\}\right],\\ \mathrm{M}_{t,2} =& n^{-1}\left[\mathbf{X}_t^\top\{2a_n\mathrm{B}_k(\mathbf{\Sigma}) + b_n\mathbf{I}_p\}\mathbf{X}_t - \mathrm{tr}\left\{(2a_n\mathrm{B}_k(\mathbf{\Sigma}) + b_n\mathbf{I}_p)\Sigma\right\}\right].\end{aligned}$$

Since $\sum_{t=1}^n \mathrm{E}(\mathrm{D}_{t,k}^4) \le C\left\{\sum_{t=1}^n \mathrm{E}(\mathrm{M}_{t,1}^4) + \sum_{t=1}^n \mathrm{E}(\mathrm{M}_{t,2}^4)\right\}$ for constant $C > 0$, it suffices to show $\mathrm{E}(\mathrm{M}_{t,i}^4) = o(\mathrm{var}^2(T_{n,k}))$ for $i = 1, 2$. By Cauchy-Schwartz inequality, for some constant $c > 0$, $\mathrm{E}\left[\mathbf{X}_t^\top \mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{X}_t - \mathrm{tr}\left\{\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma}\right\}\right]^4 \le c\mathrm{E}\left[\mathrm{tr}^2\{(\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma})^2\}\right]$. On the other hand,

$$\begin{aligned}&\mathrm{E}\left[\mathrm{tr}\{(\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma})^2\}\right]\\ =&(t-1)\sum_{|i_1-j_1|\le k}\sum_{|i_2-j_2|\le k}\left\{\sigma_{i_1i_2}\sigma_{j_1j_2} + \sigma_{i_1j_2}\sigma_{i_2j_1} + \Delta f_{i_1j_1i_2j_2}(\mathbf{\Sigma})\right\}\sigma_{i_1i_2}\sigma_{j_1j_2},\end{aligned}$$

and

$$\mathrm{var}\left[\mathrm{tr}\{(\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma})^2\}\right] = t^2 o\left(n^2\mathrm{var}(T_{n,k})\right)^2,$$

we have $\mathrm{E}\left[\mathrm{tr}^2\{(\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma})^2\}\right] = t^2 O\left(n^2\mathrm{var}(T_{n,k})\right)^2$. Therefore, it yields

$$\begin{aligned}\sum_{t=1}^n \mathrm{E}(\mathrm{M}_{t,1}^4) =& 16a_n^4 n^{-8}\sum_{t=1}^n \mathrm{E}\left[X_t^\top \mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{X}_t - \mathrm{tr}\left\{\mathrm{B}_k(\mathbf{Q}_{t-1})\mathbf{\Sigma}\right\}\right]^4\\ \le& n^{-5}O\left(n^2\mathrm{var}(T_{n,k})\right)^2 = o(\mathrm{var}(T_{n,k})).\end{aligned}$$

Similarly, for some constant $C > 0$,

$$\begin{aligned}\sum_{t=1}^n \mathrm{E}(\mathrm{M}_{t,2}^4) =& n^{-4}\sum_{t=1}^n \mathrm{E}\left[\mathbf{X}_t^\top\{2a_n\mathrm{B}_k(\mathbf{\Sigma}) + b_n\mathbf{I}_p\}\mathbf{X}_t - \mathrm{tr}\left\{(2a_n\mathrm{B}_k(\mathbf{\Sigma}) + b_n\mathbf{I}_p)\mathbf{\Sigma}\right\}\right]^4\\ \le& Cn^{-3}\mathrm{tr}^2\left[\{2a_n\mathrm{B}_k(\mathbf{\Sigma}) + b_n\mathbf{I}_p\}\mathbf{\Sigma}\right]^2,\end{aligned}$$

which implies $\sum_{t=1}^n \mathrm{E}(\mathrm{M}_{t,2}^4) = o\left(\mathrm{var}^2(T_{n,k})\right)$. We therefore obtain the second part in (A.2), and derive the assertion of Lemma A.2. □

A.2. Proof of Proposition 1

Simple algebraic computations yield

$$\hat{\sigma}^2_{\mathrm{V}_{n,k}0} = 2\{n(n-1)\}^{-1} \sum_{|i_1-j_1|\le k} \sum_{|i_2-j_2|\le k} \{\Pi_1 - 2\Pi_2 + \Pi_3\},$$

where

$$\Pi_1 = \frac{1}{P_n^2} \sum_{l_1\neq l_2} \{(x_{1_1 i_1}x_{1_1 j_1} - \sigma_{i_1 j_1})(x_{1_1 i_2}x_{1_1 j_2} - \sigma_{i_2 j_2})(x_{1_2 i_1}x_{1_2 j_1} - \sigma_{i_1 j_1})(x_{1_2 i_2}x_{1_2 j_2} - \sigma_{i_2 j_2})\},$$

$$\Pi_2 = \frac{1}{P_n^3} \sum^{*}_{l_1,l_2,l_3} \{(x_{1_1 i_1}x_{1_1 j_1} - \sigma_{i_1 j_1})(x_{1_2 i_2}x_{1_2 j_2} - \sigma_{i_2 j_2})(x_{1_3 i_1}x_{1_3 j_1} - \sigma_{i_1 j_1})(x_{1_3 i_2}x_{1_3 j_2} - \sigma_{i_2 j_2})\},$$

$$\Pi_3 = \frac{1}{P_n^4} \sum^{*}_{l_1,l_2,l_3,l_4} \{(x_{1_1 i_1}x_{1_1 j_1} - \sigma_{i_1 j_1})(x_{1_2 i_2}x_{1_2 j_2} - \sigma_{i_2 j_2})(x_{1_3 i_1}x_{1_3 j_1} - \sigma_{i_1 j_1})(x_{1_4 i_2}x_{1_4 j_2} - \sigma_{i_2 j_2})\}.$$

Thus, $\mathrm{E}(\hat{\sigma}^2_{\mathrm{V}_{n,k}0}) = \sigma^2_{\mathrm{V}_{n,k}0}$. Similar to Lemma A.1,

$$\begin{aligned}
&\mathrm{var}\left(\sum_{|i_1-j_1|\le k} \sum_{|i_2-j_2|\le k} \Pi_1\right) \\
=&\Bigg(-4n^{-1}\Bigg[\sum_{|i_1-j_1|\le k} \sum_{|i_2-j_2|\le k} \{\mathrm{E}(x_{li_1}x_{lj_1}x_{li_2}x_{lj_2}) - \sigma_{i_1j_1}\sigma_{i_2j_2}\}^2\Bigg]^2 + 4n^{-1} \sum_{|i_1-j_1|\le k} \sum_{|i_2-j_2|\le k} \sum_{|i_3-j_3|\le k} \cdots \\
&\cdots \sum_{|i_4-j_4|\le k} \Bigg[\{\mathrm{E}(x_{li_1}x_{lj_1}x_{li_2}x_{lj_2}) - \sigma_{i_1j_1}\sigma_{i_2j_2}\} \{\mathrm{E}(x_{li_3}x_{lj_3}x_{li_4}x_{lj_4}) - \sigma_{i_3j_3}\sigma_{i_4j_4}\} \times \cdots \\
&\cdots \mathrm{E}\{(x_{1i_1}x_{1j_1} - \sigma_{i_1j_1})(x_{1i_2}x_{1j_2} - \sigma_{i_2j_2})(x_{1i_3}x_{1j_3} - \sigma_{i_3j_3})(x_{1i_4}x_{1j_4} - \sigma_{i_4j_4})\}\Bigg]\Bigg) \{1 + O(n^{-1})\} \\
=&O\left((n^3 + k^2n^3)\sigma^2_{\mathrm{V}_{n,k}0}\right).
\end{aligned}$$

The variances for the remaining terms of $\hat{\sigma}^2_{\mathrm{V}_{n,k}0}$ can be estimated similarly and the proposition follows.

A.3. Proof of Theorem 1

Denote

$$\widetilde{V}_{n,k} = \sum_{|i-j|\le k} \mathrm{L}_{n_1}(i,j) - 2\sum_{i=1}^{p} \mathrm{L}_{n_4}(i) + p$$

Under Assumptions 1 and 2, $\mathrm{E}(\widetilde{V}_{n,k}) = \mathrm{tr}[\{\mathrm{B}_k(\mathbf{\Sigma}) - \mathbf{I}_p\}^2]$. From Lemma A.1, we have $\mathrm{var}(\widetilde{V}_{n,k}) = \sigma^2_{\mathrm{V}_{n,k}} + o(\sigma^2_{\mathrm{V}_{n,k}})$, $\mathrm{var}\{\sum_{|i-j|\le k} \mathrm{L}_{n_2}(i,j)\} = o(\sigma^2_{\mathrm{V}_{n,k}})$, $\mathrm{var}\{\sum_{|i-j|\le k} \mathrm{L}_{n_3}(i,j)\} = o(\sigma^2_{\mathrm{V}_{n,k}})$, and $\mathrm{var}\,\{\sum_{i=1}^{p} \mathrm{L}_{n_5}(i)\} =$

$o(\sigma^2_{\mathrm{V}_{n,k}})$. Theorem 1 therefore follows Lemma A.2 and Slutsky's Theorem.

A.4. Proof of Theorem 2

It follows Theorems 1 and (3.5) that

$$\beta_{\mathrm{V}_{n,k}} = 1 - \Phi\left(\sigma^{-1}_{\mathrm{V}_{n,k}}\sigma_{\mathrm{V}_{n,k}0} z_\alpha - \sigma^{-1}_{\mathrm{V}_{n,k}}[\mathrm{tr}\{(\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p)^2\}]\right).$$

As discussed before, it can be seen that

$$\sigma_{\mathrm{V}_{n,k}} \leq \mathrm{tr}\left[\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}^2\right]\left(\frac{\tau^2_{n,k}(\boldsymbol{\Sigma})}{\mathrm{tr}^2\left[\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}^2\right]} + \frac{8+4\Delta}{n}\frac{\mathrm{tr}[\Sigma\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}]^2}{\mathrm{tr}^2\left[\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}^2\right]}\right)^{1/2},$$

so that under the condition $\{\tau_{n,k}(\boldsymbol{\Sigma})\}^{-1}\mathrm{tr}\left[\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}^2\right] \to \infty$, it suffices to show

$$\frac{\mathrm{tr}[\Sigma\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}]^2}{n\mathrm{tr}^2\left[\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}^2\right]} \to 0, \tag{A.3}$$

provided $\sigma^{-1}_{\mathrm{V}_{n,k}}\sigma_{\mathrm{V}_{n,k}0}$ is bounded.

Denote $\lambda_1 \leq \ldots \leq \lambda_p$ the eigenvalues of $\mathrm{B}_k(\boldsymbol{\Sigma})$ that for some δ_0, $0 < \delta_0 \leq \lambda_1 \leq \lambda_p \leq \delta_0^{-1}$ for sufficiently large k. Standard algebraic computations give

$$\begin{aligned}\mathrm{tr}\left[\Sigma\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}\right]^2 &= \mathrm{tr}\left[\mathrm{B}_k(\boldsymbol{\Sigma})\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}\right]^2 + \mathrm{tr}\left[\{\boldsymbol{\Sigma} - \mathrm{B}_k(\boldsymbol{\Sigma})\}\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}\right]^2 \\ &\quad + 2\mathrm{tr}\left[\mathrm{B}_k(\boldsymbol{\Sigma})\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}\{\boldsymbol{\Sigma} - \mathrm{B}_k(\boldsymbol{\Sigma})\}\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}\right] \\ &\equiv I_1 + I_2 + I_3.\end{aligned}$$

By the fact that $\|\boldsymbol{\Sigma} - \mathrm{B}_k(\boldsymbol{\Sigma})\| \leq Ck^{-\alpha}$, $I_i = o\left(n\mathrm{tr}^2\left[\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}^2\right]\right)$ for $i = 2, 3$. Notice that

$$\mathrm{tr}\left[\mathrm{B}_k(\boldsymbol{\Sigma})\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}\right]^2 = \sum_{i=1}^p \lambda_i^2(\lambda_i - 1)^2 \leq \delta_0^{-2}\sum_{i=1}^p(\lambda_i - 1)^2 = \delta_0^{-2}\mathrm{tr}\left[\{\mathrm{B}_k(\boldsymbol{\Sigma}) - \mathbf{I}_p\}^2\right],$$

from which (A.3) follows. We therefore obtain Theorem 2.

A.5. Proof of Theorem 3

Denote

$$\widetilde{\mathrm{U}}_{n,k} = \frac{\sum_{|i-j|\leq k}\mathrm{L}_{n_1}(i,j)}{\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}} - \frac{2\sum_{i=1}^p \mathrm{L}_{n_4}(i)}{\mathrm{tr}(\boldsymbol{\Sigma})} + 1.$$

By Lemmas A.1 and A.2, we have $\mathrm{var}(\widetilde{\mathrm{U}}_{n,k}) = \sigma^2_{\mathrm{U}_{n,k}}\{1+o(1)\}$ and $\sigma^{-1}_{\mathrm{U}_{n,k}}\widetilde{\mathrm{U}}_{n,k} \to \mathcal{N}(0,1)$ in distribution. Together with results that $\mathrm{var}\left[\sum_{|i-j|\le k}\mathrm{L}_{n_2}(i,j)/\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}\right] = o(\sigma^2_{\mathrm{U}_{n,k}})$, $\mathrm{var}\left\{\sum_{i=1}^p \mathrm{L}_{n_5}(i)/\mathrm{tr}(\boldsymbol{\Sigma})\right\} = o(\sigma^2_{\mathrm{U}_{n,k}})$ and $\mathrm{var}\left[\sum_{|i-j|\le k}\mathrm{L}_{n_3}(i,j)/\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}\right] = o(\sigma^2_{\mathrm{U}_{n,k}})$, we have

$$\sigma^{-1}_{\mathrm{U}_{n,k}}\widehat{\mathrm{U}}_{n,k} \to \mathcal{N}(0,1)$$

in distribution, where

$$\widehat{\mathrm{U}}_{n,k} = \frac{\sum_{|i-j|\le k}\{\mathrm{L}_{n_1}(i,j) - 2\mathrm{L}_{n_2}(i,j) + \mathrm{L}_{n_3}(i,j)\}}{\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}} - \frac{2\sum_{i=1}^p\{\mathrm{L}_{n_4}(i) - \mathrm{L}_{n_5}(i)\}}{\mathrm{tr}(\boldsymbol{\Sigma})} + 1.$$

Denote $\epsilon_n = \left[\sum_{i=1}^p\{\mathrm{L}_{n_4}(i) - \mathrm{L}_{n_5}(i)\} - \mathrm{tr}(\boldsymbol{\Sigma})\right]/\mathrm{tr}(\boldsymbol{\Sigma})$ with $\mathrm{E}(\epsilon_n) = 0$, then

$$\begin{aligned}\mathrm{var}(\epsilon_n) &= \mathrm{tr}^{-2}(\boldsymbol{\Sigma})\left[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}^2) + \Delta n^{-1}\mathrm{tr}\left\{\left(\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}\right)\circ\left(\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}\right)\right\} + 2\{n(n-1)\}^{-1}\mathrm{tr}(\boldsymbol{\Sigma}^2)\right]\\ &\le \left[(2+\Delta)n^{-1} + 2\{n(n-1)\}^{-1}\right]\mathrm{tr}^{-2}(\boldsymbol{\Sigma})\mathrm{tr}(\boldsymbol{\Sigma}^2) = o(\sigma_{\mathrm{U}_n}).\end{aligned}$$

Thus, Theorem 3 follows from

$$\left(\frac{\mathrm{tr}^2(\boldsymbol{\Sigma})}{\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}}\right)\left(\frac{\mathrm{U}_{n,k}+1}{p}\right) - 1 = \frac{\widehat{\mathrm{U}}_{n,k} - \epsilon_n^2}{(1+\epsilon_n^2)^2}.$$

A.6. Proof of Theorem 4

It follows Theorems 3 and (3.10)

$$\beta_{\mathrm{U}_{n,k}} = 1 - \Phi\left(\left[\frac{\mathrm{tr}^2(\boldsymbol{\Sigma})}{p\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}}\right]\left(\frac{\sigma_{\mathrm{U}_{n,k}0}}{\sigma_{\mathrm{U}_{n,k}}}\right)z_\alpha - \sigma^{-1}_{\mathrm{U}_{n,k}}\left[1 - \frac{\mathrm{tr}^2(\boldsymbol{\Sigma})}{p\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}}\right]\right).$$

Since $\sigma_{\mathrm{U}_{n,k}0}/\sigma_{\mathrm{U}_{n,k}}$ and $p^{-1}\mathrm{tr}^2(\boldsymbol{\Sigma})/\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\}$ are bounded and

$$\sigma_{\mathrm{U}_{n,k}} \le \left[\frac{\tau^2_{n,k}(\boldsymbol{\Sigma})}{\mathrm{tr}^2\{(\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma})\}} + \frac{8+4\Delta}{n}\mathrm{tr}\left\{\left(\frac{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma}}{\mathrm{tr}(\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma})} - \frac{\boldsymbol{\Sigma}}{\mathrm{tr}(\boldsymbol{\Sigma})}\right)^2\right\}\right]^{1/2},$$

similar to Theorem 2, under the condition $\{\tau_{n,k}(\boldsymbol{\Sigma})\}^{-1}[\mathrm{tr}\{\mathrm{B}_k(\boldsymbol{\Sigma})\boldsymbol{\Sigma} - \mathrm{tr}^2(\boldsymbol{\Sigma})/p\}] \to \infty$ it is sufficient to show

$$n^{-1}\mathrm{tr}\left[\left\{\frac{\boldsymbol{\Sigma}^2}{\mathrm{tr}(\boldsymbol{\Sigma}^2)} - \frac{\boldsymbol{\Sigma}}{\mathrm{tr}(\boldsymbol{\Sigma})}\right\}^2\right]\left\{1 - \frac{\mathrm{tr}^2(\boldsymbol{\Sigma})}{p\mathrm{tr}(\boldsymbol{\Sigma}^2)}\right\}^{-2} \to 0. \tag{A.4}$$

Standard algebra and definition in (2.1) implies that

$$
\begin{aligned}
&\operatorname{tr}\left[\left\{\boldsymbol{\Sigma}^2-\operatorname{tr}^{-1}(\boldsymbol{\Sigma})\operatorname{tr}(\boldsymbol{\Sigma}^2)\Sigma\right\}^2\right]\\
\leq&\epsilon_0^{-2}\sum_{i=1}^{p}\left\{\lambda_i(\boldsymbol{\Sigma})-\operatorname{tr}^{-1}(\boldsymbol{\Sigma})\operatorname{tr}(\boldsymbol{\Sigma}^2)\right\}^2\\
=&\epsilon_0^{-2}\operatorname{tr}^2(\boldsymbol{\Sigma})\sum_{i=1}^{p}\left\{\operatorname{tr}^{-1}(\boldsymbol{\Sigma})\lambda_i(\boldsymbol{\Sigma})-p^{-1}+p^{-1}-\operatorname{tr}^{-2}(\boldsymbol{\Sigma})\operatorname{tr}(\boldsymbol{\Sigma}^2)\right\}^2\\
=&\epsilon_0^{-2}\operatorname{tr}^2(\boldsymbol{\Sigma})\left[\operatorname{tr}\left\{\left(\operatorname{tr}^{-1}(\boldsymbol{\Sigma})\boldsymbol{\Sigma}-p^{-1}\mathbf{I}_p\right)^2\right\}+p\left\{p^{-1}-\operatorname{tr}^{-2}(\boldsymbol{\Sigma})\operatorname{tr}(\boldsymbol{\Sigma}^2)\right\}^2\right],
\end{aligned}
$$

and $\{\operatorname{tr}(\boldsymbol{\Sigma}^2)-p^{-1}\operatorname{tr}^2(\boldsymbol{\Sigma})\}^2=\operatorname{tr}^2(\boldsymbol{\Sigma})\operatorname{tr}\left\{\operatorname{tr}^{-1}(\boldsymbol{\Sigma})\boldsymbol{\Sigma}-p^{-1}\mathbf{I}_p\right\}^2$, which implies (A.4) and Theorem 4 follows.

Acknowledgments

The authors thanks two anonymous referees and the associate editor for constructive comments and suggestions which have improved the presentation of the paper. Chen was partially supported by National Science Foundation grant DMS-1309210, National Natural Science Foundation of China Grants 11131002, G0113 and 71532001, National Key Basic Research Program of China grant 2015CB856000, and Center for Statistical Science at Peking University. Zhou was supported in part by NSF Grant IIS-1545994.

[1] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, third ed., Wiley-Interscience, New York, 2003.

[2] Z. Bai, D. Jiang, J. Yao, S. Zheng, Corrections to LRT on large-dimensional covariance matrix by RMT, Ann. Statist. 37(6) (2009) 3822-3840.

[3] Z. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem, Statist. Sinica 6 (1996) 311-329.

[4] Z. Bai, Y. Yin, Limit of the smallest eigenvalue of a large dimensional sample covariance matrix, Ann. Probab. 21 (1993) 1276-1294.

[5] Z. Bai, J.W. Silverstein, Y. Yin, A note on the largest eigenvalue of a large-dimensional sample covariance matrix, J. Multivariate Anal. 26 (1988) 166-168.

[6] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Roy. Statist. Soc. Ser. B 57 (1995) 289-300.

[7] P.J. Bickel, E. Levina, Regularized estimation of large covariance matrices, Ann. Statist. 36 (2008) 199-227.

[8] P.J. Bickel, E. Levina, Covariance regularization by thresholding, Ann. Statist. 36 (2008) 2577-2604.

[9] B.M. Bolker, Ecological Models and Data in R, Princeton University Press, 2008.

[10] T.T. Cai, T. Jiang, Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices, Ann. Statist. 39 (2011) 1496-1525.

[11] T.T. Cai, C.H. Zhang, H.H. Zhou, Optimal rates of convergence for covariance matrix estimation, Ann. Statist. 38 (2010) 2118-2144.

[12] Cai, T.T. and Liu, W.D. (2011), "Adaptive thresholding for sparse covariance matrix estimation", *Journal of the American Statistical Association*, **494**, 672-684.

[13] S. X. Chen, Y. Qin, A two sample test for high dimensional data with applications to gene-set testing, Ann. Statist. 38 (2010) 808-835.

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

[15] S. X. Chen, L.-H. Peng, W. Zhou, Supplement to More powerful tests for sparse high-dimensional covariances, Technical report, (2015).

[16] S. Chiaretti, X. Li, R. Gentleman, A. Vitale, M. Vignetti, F. Mandelli, J. Ritz, R. Foa, Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival, Blood 103 (2004) 2771-2778.

[17] N. Cressie, Statistics for spatial data, Wiley-Interscience, New York, 1993.

[18] S. Dudoit, S. Keles, M.J. van der Laan, Multiple tests of associations with biological annotation metadata, Institute of Mathematical Statistics Collections 2 (2008) 153-218.

[19] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Amer. Statist. Assoc. 96 (2001) 1348-1360.

[20] M. Friendly, Corrgrams: Exploratory displays for correlation matrices, Amer. Statist. 56 (2002) 316-324.

[21] R. Gentleman, R.A. Irizarry, V.J. Carey, S. Dudoit, W. Huber, Bioinformtics and computational biology solutions using R and Bioconductor, Springer-Verlag, New York, 2005.

[22] M. Hallin, D. Paindaveine, Semiparametrically efficient rank-based inference for shape: optimal rank-based tests for sphericity, Ann. Statist. 34 (2006) 2707-2756.

[23] K.R. Katsani, M. Irimia, C. Karapiperis, Z.G. Scouras, B.J. Blencowe, V.J. Promponas, C.A. Ouzounis, Functional genomics evidence unearths new moonlighting roles of outer ring coat Nucleoporins, Scientific Reports 4 (2014) 4655.

[24] T. Jiang, The asymptotic distribution of the largest entries of sample correlation matrices, Ann. Appl. Probab. 14 (2004) 865-880.

[25] S. John, Some optimal multivariate tests, Biometrika 59 (1971) 123-127.

[26] S. John, The distribution of a statistic used for testing sphericity of normal distributions, Biometrika 59 (1972) 169-173.

[27] O. Ledoit, M. Wolf, Some hypothesis tests for the covariance matrix when the dimension is large compare to the sample size, Ann. Statist. 30 (2002) 1081-1102.

[28] H. Nagao, On some test criteria for covariance matrix, Ann. Statist. 1 (1973) 700-709.

[29] S. Pakakasama, S. Kajanachumpol, S. Kanjanapongkul, N. Sirachainan, A. Meekaewkunchorn, V. Ningsanond, S. Hongeng, Simple multiplex RT-PCR for identifying common fusion transcripts in childhood acute leukemia, Intl. J. Lab. Hemato. 30 (2008) 286-291.

[30] Y.M. Qiu, S. X. Chen, Test for bandedness of high-dimensional covariance matrices and bandwidth estimation, Ann. Statist. 40 (2012) 1285-1314.

[31] Y.M. Qiu, S. X. Chen, Band width selection for high dimensional covariance matrix estimation, J. Amer. Statist. Assoc., to appear.

[32] A.J. Rothman, E. Levina, J. Zhu, Generalized thresholding of large covariance matrices, J. Amer. Statist. Assoc. 104 (2009) 177-186.

[33] J.R. Schott, Testing for complete independence in high dimensions, Biometrika 92 (2005) 951-956.

[34] Q. Shi, S. Gao, L. Song, Y. Zhao, X. Li, J. Wu, C. He, H. Li, H. Zhao, Comparative proteomics analysis of differential proteins in respond to doxorubicin resistance in myelogenous leukemia cell lines, Proteome Science 13 (2015) doi:10.1186/s12953-014-0057-y.

[35] M.S. Srivastava, H. Yanagihara, Testing the equality of several covariance matrices with fewer observations than the dimension, J. Multivariate Anal. 101 (2010) 1319-1329.

[36] M.S. Srivastava, N. Reid, Testing the structure of the covariance matrix with fewer observations than the dimension, J. Multivariate Anal. 112 (2012) 156–171.

[37] M.S. Srivastava, H. Yanagihara, T. Kubokawa, Tests for covariance matrices in high dimension with less sample size, J. Multivariate Anal. 130 (2014) 289-309.

[38] R. Tibshirani, Regression shrinkage and selection via the lasso, J. Roy. Statist. Soc. Ser. B 58 (1996) 267-288.

[39] A. Touloumis, S. Tavaré, J.C. Marioni, Testing the mean matrix in high-dimensional transposable data, Biometrics (2015) doi: 10.1111/biom.12257.

[40] E. Tsakou, A. Agathagelidis, M. Boudjoghra, T. Raff, A. Dagklis, M. Chatzouli, T. Smilevska, G. Bourikas, H. Merle-Beral, E. Manioudaki-Kavallieratou, A. Anagnostopoulos, M. Bruggemann, F. Davi, K. Stamatopoulos, C. Belessi, Partial versus productive immunoglobulin heavy locus rearrangements in chronic lymphocytic leukemia: implications for b-cell receptor stereotypy, Molecular Medicine 18 (2012) 138-145.

[41] A.S. Wagaman, E. Levina, Discovering sparse covariance structures with the Isomap, J. Comp. Graph. Statist. 18 (2008) 551-572.

[42] S. Zheng, D. Jiang, Z. Bai, X. He, Inference on multiple correlation coefficients with moderately high dimensional data, Biometrika 101 (2014) 748-754.

[43] C.L. Zou, L.-H. Peng, L. Feng, Z.J. Wang, Multivariate sign-based high-dimensional tests for sphericity, Biometrika 101 (2014) 229-236.