Test for Bandedness of High Dimensional Covariance Matrices with Bandwidth Estimation

Qiu, Yumou and Chen, Songxi

2012

Submitted to the Annals of Statistics

TEST FOR BANDEDNESS OF HIGH-DIMENSIONAL COVARIANCE MATRICES AND BANDWIDTH ESTIMATION

BY YUMOU QIU[*] AND SONG XI CHEN[†]

Iowa State University[*], *and Peking University and Iowa State University*[†]

Motivated by the latest effort to employ banded matrices to estimate a high-dimensional covariance Σ, we propose a test for Σ being banded with possible diverging bandwidth. The test is adaptive to the "large p, small n" situations without assuming a specific parametric distribution for the data. We also formulate a consistent estimator for the bandwidth of a banded high-dimensional covariance matrix. The properties of the test and the bandwidth estimator are investigated by theoretical evaluations and simulation studies, as well as an empirical analysis on a protein mass spectroscopy data.

1. Introduction. High-dimensional data are increasingly collected in statistical applications, which include biological experiments, climate and environmental studies, financial observations and others. The high dimensionality calls for new statistical methodologies which are adaptive to this new feature of the modern statistical data. The covariance matrix $\Sigma = \mathrm{Var}(X)$ for a p-dimensional random vector X is an important measure on the dependence among components of X. The sample covariance S_n, constructed based on n independent copies of X, is a key ingredient in many statistical procedures in the conventional multivariate analysis [Anderson (2003) and Muirhead (1982)] where the data dimension p is regarded as fixed. The widespread use of S_n in the conventional multivariate procedures is largely due to S_n being a consistent estimator of Σ when p is fixed or small relative to the sample size n. However, for high-dimensional data such that $p/n \to c \in (0, \infty]$, it is known that the eigenvalues of the sample covariance matrix are no longer consistent to their population counterpart, as demonstrated in Bai and Yin (1993), Bai, Silverstein and Yin (1998), Johnstone (2001) and El Karoui (2011). These mean that the sample covariance S_n is no longer consistent to Σ, which hinders applications of many conventional multivariate statistical procedures for high-dimensional data.

To overcome the problem with the sample covariance, constructing covariance estimators via banding or tapering the sample covariance matrix has

Keywords and phrases: Banded covariance matrix, Bandwidth estimation, High data dimension, Large p, small n, Nonparametric.

been a focus in high-dimensional covariance estimation. Wu and Pourahmadi (2003) considered banding the Cholesky factor matrix via the kernel smoothing estimation, which was further developed by Rothman, Levina and Zhu (2010). Bickel and Levina (2008a) proposed banding the sample covariance matrix directly for estimating Σ and banding the Cholesky factor matrix for estimating Σ^{-1}. They demonstrated that both estimators are consistent to Σ and Σ^{-1}, respectively, for some "bandable" classes of covariance matrices. Cai, Zhang and Zhou (2010) proposed a tapering estimator, which can be viewed as a soft banding on the sample covariance, which was designed to improve the banding estimator of Bickel and Levina. They demonstrated that the tapering estimator attains the optimal minimax rates of convergence for estimating the covariance matrix. Wagaman and Levina (2009) developed a method for discovering meaningful orderings of variables such that banding and tapering can be applied. Both the banding and tapering methods for covariance estimation are well connected to the regularization method considered in Huang et al (2006), Bickel and Levina (2008b), Fan, Fan and Lv (2008) and Rothman, Levina and Zhu (2009).

Motivated by the promising results regarding banding and tapering the sample covariance, we develop in this paper a test procedure on the hypothesis that Σ is banded. The rationale for developing such a test is to check a Σ in the so-called "bandable" class outlined in Bickel and Levina (2008a) such that the banding or the tapering estimators are consistent. There is yet a practical guideline to confirm or otherwise if a Σ is within the "bandable" class so that the banding and tapering can be applied. Hence, a direct testing on Σ being banded provides a path of advance to gain knowledge on the structure of the covariance. If the banded hypothesis is confirmed by the test, the banding and tapering estimators may be employed.

Diagonal matrices are the simplest among banded matrices. Given the importance commanded by covariance matrices in high-dimensional multivariate analysis, directly testing for Σ being diagonal and the so-called sphericity hypothesis in classical multivariate analysis [John (1972) and Nagao (1973)], have been considered in a set of studies including Ledoit and Wolf (2002), Jiang (2004), Schott (2005), Chen, Zhang and Zhong (2010) and Cai and Jiang (2012) under high-dimensionality. For normally distributed data, Jiang (2004) proposed testing for diagonal Σ by considering a coherence statistic $L_n = \max_{1 \le i < j \le p} |\hat{\rho}_{ij}|$, where $\hat{\rho}_{ij}$ is the sample correlation coefficient between the ith and the jth components of the random vector X. Jiang established the asymptotic distribution of L_n under the null diagonal hypothesis, which was used to derive a sphericity test. As L_n is an extreme value type, its convergence to its limiting distribution can be slow. Liu, Lin and Shao (2008) proposed a modification which is shown to be able to speed up the convergence. Cai and Jiang (2012) extended the test of Jiang (2004) for the bandedness of Σ, which is shown to be applicable for the "large p, small n" situations such that $\log(p) = o(n^{1/3})$.

In this paper, we propose a nonparametric test for Σ being banded without assuming a parametric distribution for the high-dimensional data. The test is formulated to allow the dimension to be much larger than the sample size. Based on the test statistic for bandedness, we propose a consistent estimator for the bandwidth of a banded high-dimensional covariance. The properties of the test and bandwidth estimator are demonstrated by theoretical evaluation, simulation studies and empirical analysis on a protein mass spectroscopy data for prostate cancer.

The paper is organized as follows. Section 2 introduces the hypotheses, the assumptions and the test statistic. In Section 3, we present the properties of the test statistic and the test, and evaluate its power properties. Estimation of the bandwidth is considered in Section 4. Section 5 reports simulation results. An empirical analysis on a prostate cancer spectroscopy data is outlined in Section 6. All technical details are relegated to the Appendix.

2. Preliminary. Let $X_1, X_2, \dots X_n$ be independent and identically distributed p-dimensional random vectors with mean μ and covariance matrix $\Sigma = (\sigma_{ij})_{p\times p}$. A matrix $A = (a_{ij})_{p\times p}$ is said to be banded if there exists an integer $k \in \{0, \cdots, p-1\}$ such that $a_{ij} = 0$ for $|i-j| > k$. The smallest k such that A is banded is called the bandwidth of A. Banding of A at a bandwidth k refers to setting $a_{ij} = 0$ for all $|i-j| > k$.

Let $B_k(\Sigma) = (\sigma_{ij}\mathrm{I}\{|i-j| \le k\})_{p\times p}$ be a banded version of Σ with bandwidth k. Specifically, $B_0(\Sigma)$ is the diagonal version of Σ. We intend to test

$$H_{k,0} : \Sigma = B_k(\Sigma) \quad \text{v.s.} \quad H_{k,1} : \Sigma \neq B_k(\Sigma) \tag{2.1}$$

for $k = o(p^{1/4})$. Hence, the bandwidth k of Σ to be tested can be either fixed or diverging to infinite as long as it is slower than $p^{1/4}$. Allowing divergent bandwidth in the hypothesis is an improvement over the sphericity test as considered in Ledoit and Wolf (2002) and Chen et al. (2010). It also connects to the latest works on high-dimensional covariance estimation with banded or tapered versions of the sample covariance as in Bickel and Levina (2008a) and Cai, Zhang and Zhou (2010). In particular, Cai et al. (2010) showed that the optimal minimax rates for the bandwidth of the banded covariance estimator of Bickel and Levina (2008a) is $k = O[\{n/\log(p)\}^{1/(2\alpha+1)}]$, and that for the tapering estimator is $k = O(n^{1/(2\alpha+1)})$, where α is an index value for a "bandable" class of covariances

$$\mathfrak{U}(\varepsilon_0, \alpha, C) = \Big\{\Sigma : \max_j \sum_{|i-j|>k} |\sigma_{ij}| \le Ck^{-\alpha} \quad \text{for all} \quad k > 0, \tag{2.2}$$
$$\text{and} \quad 0 < \varepsilon_0 \le \lambda_{min}(\Sigma) \le \lambda_{max}(\Sigma) \le \varepsilon_0^{-1}\Big\}.$$

The range of bandwidths $k = o(p^{1/4})$ in the hypothesis (2.1) should cover the above optimal rates when $p >> n$.

We note that $H_{k,0}$ is valid if and only if $\sum_{|i-j|>k_p} \sigma_{ij}^2 = 0$, and the latter implies that $\mathrm{tr}\{\Sigma - B_k(\Sigma)\}^2 = 0$. A strategy is to construct an unbiased estimator of $\mathrm{tr}\{\Sigma - B_k(\Sigma)\}^2$ and use it to develop the test statistic. Let $D_q := \sum_{l=1}^{p-q} \sigma_{l\,l+q}^2$ be the sum of squares of the qth sub-diagonal of Σ. Then, $\mathrm{tr}\{\Sigma - B_k(\Sigma)\}^2 = 2\sum_{q=k+1}^{p-1} D_q$. It can be checked that an unbiased estimator of D_q is

$$\hat{D}_{nq} = \sum_{l=1}^{p-q} \Big\{ \frac{1}{P_n^2} \sum_{i,j}^{*} (X_{il}X_{i\ l+q})(X_{jl}X_{j\ l+q}) - 2\frac{1}{P_n^3} \sum_{i,j,k}^{*} X_{il}X_{k\ l+q}(X_{jl}X_{j\ l+q})$$
$$+ \frac{1}{P_n^4} \sum_{i,j,k,m}^{*} X_{il}X_{j\ l+q}X_{kl}X_{m\ l+q} \Big\},$$

where $\sum^{*}$ denotes summation over mutually different subscripts shown and $P_n^b = n!/(n-b)!$. The reason to sum over different indices is for easier manipulations with the mean and variance of the final test statistic and to establish the asymptotic normality. The latter leads to a test procedure for the bandedness.

We consider the following statistic:

$$W_{nk} := 2 \sum_{q=k+1}^{p-1} \hat{D}_{nq}. \tag{2.3}$$

As each $\hat{D}_{nq}$ is invariant under the location shift, W_{nk} is also location shift invariant. Hence, without loss of generality, we assume $\mu = \mathrm{E}(X) = 0$.

To facilitate our analysis, as Bai and Saranadasa (1996) and Chen, Zhang and Zhong (2010), we assume a multivariate model for the high-dimensional data.

Assumption 1. *(i) $X_1, X_2, \cdots, X_n$ are independent and identically distributed (i.i.d.) p-dimensional random vectors such that*

$$X_i = \Gamma Z_i \qquad \textit{for } i = 1, 2, \cdots, n, \tag{2.4}$$

where Γ is a $p\times m$ constant matrix with $m \geq p$, $\Gamma\Gamma' = \Sigma$, and $Z_1, \cdots, Z_n$ are i.i.d. m-dimensional random vectors such that $\mathrm{E}(Z_1) = 0$ and $\mathrm{Var}(Z_1) = I_m$.

(ii) Write $Z_1 = (z_{11}, \ldots, z_{1m})^T$. Each z_{1l} has uniformly bounded 8th moment, and there exist finite constants Δ and ω such that for $l = 1, \cdots, m$, $\mathrm{E}(z_{1l}^4) = 3 + \Delta$, $\mathrm{E}(z_{1l}^3) = \omega$ and for any integers $\ell_\nu \geq 0$ with $\sum_{\nu=1}^q \ell_\nu = 8$

$$\mathrm{E}(z_{i_1}^{\ell_1} z_{i_2}^{\ell_2} \cdots z_{i_q}^{\ell_q}) = \mathrm{E}(z_{1i_1}^{\ell_1})\mathrm{E}(z_{1i_2}^{\ell_2}) \cdots \mathrm{E}(z_{1i_q}^{\ell_q}) \tag{2.5}$$

whenever i_1, i_2, $\cdots$, i_q are distinct subscripts.

The requirement of common third and fourth moments of z_{1l} is not essential and is purely for the sake of simpler notation. Our theory allows different third and fourth moments as long as they are uniformly bounded, which are actually assured by z_{1l} having uniformly bounded 8th moment.

The asymptotic framework that regulates the sample size n, the dimensionality p and the covariance Σ is the following.

Assumption 2. *As $n \to \infty$, $p = p(n) \to \infty$, $n = O(p)$ and $\mathrm{tr}(\Sigma^4)/\mathrm{tr}^2(\Sigma^2) = O(p^{-1})$.*

We note that $n = O(p)$ includes $p >> n$, the "large p, small n" paradigm, but may not imply $p = O(n)$. Different from the usual approach of specifying an explicit growth rate of p with respect to n, Assumption 2 requires ratio of $\mathrm{tr}(\Sigma^4)$ to $\mathrm{tr}^2(\Sigma^2)$ shrinks at the rate of p^{-1} or smaller. The latter is stronger than $\mathrm{tr}(\Sigma^4)/\mathrm{tr}^2(\Sigma^2) = o(1)$. It is needed due to possible diverging bandwidths.

Let

$$\mathcal{U}_p = \left\{\Sigma : \frac{\mathrm{tr}(\Sigma^4)}{\mathrm{tr}^2(\Sigma^2)} = O(p^{-1})\right\}$$

be the class of covariances satisfying the last part of Assumption 2. The class includes the "bandable" class $\mathfrak{U}(\varepsilon_0, \alpha, C)$ of Bickel and Levina (2008a) given in (2.2) for the banding estimation. To appreciate this, let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$ be the eigenvalues of Σ. If the smallest and largest eigenvalues are bounded away from 0 and ∞ respectively, then

$$\frac{\mathrm{tr}(\Sigma^4)}{\mathrm{tr}^2(\Sigma^2)} = \frac{\sum_{i=1}^p \lambda_i^4}{\left(\sum_{i=1}^p \lambda_i^2\right)^2} \leq \frac{\lambda_p^4}{p\lambda_1^4} = O(p^{-1}).$$

Therefore, the "bandable" covariances are contained in $\mathcal{U}_p$. Now suppose that Σ has exactly m_p zero eigenvalues and λ_{m_p+1} being the smallest nonzero eigenvalue. Then

$$\frac{\mathrm{tr}(\Sigma^4)}{\mathrm{tr}^2(\Sigma^2)} \leq \frac{\lambda_p^4}{(p - m_p)\lambda_{m_p+1}^4}.$$

Thus, Σ is in $\mathcal{U}_p$ as long as $\lambda_p/\lambda_{m_p+1}$ is bounded and $m_p \le cp$ for some $c \in (0,1)$ as $p \to \infty$. The latter means that the class $\mathcal{U}_p$ is likely to contain the class considered in Cai et al. (2010), which allows the smallest eigenvalue to diminish to zero. It can be also checked that the following two covariances

$$\Sigma = \left(\sigma_i\sigma_j\rho^{|j-i|}\right)_{p\times p} \quad \text{or} \quad \Sigma = \left(\sigma_i\sigma_j\rho^{|j-i|}I(|j-i|\le d)\right)_{p\times p}$$

are members of $\mathcal{U}_p$ if $\{\sigma_l^2\}_{l=1}^p$ are uniformly bounded from infinity and zero respectively.

3. Main results. We first describe the basic properties of the statistic W_{nk} defined in (2.3). Let

$$\begin{aligned}\nu_{nk}^2 = {}&\frac{4}{n^2}\mathrm{tr}^2(\Sigma^2) + \frac{8}{n}\mathrm{tr}\left\{\Sigma(\Sigma - B_k(\Sigma))\right\}^2 \\ &+ \frac{4}{n}\Delta\mathrm{tr}\{\Gamma'(\Sigma - B_k(\Sigma))\Gamma \circ \Gamma'(\Sigma - B_k(\Sigma))\Gamma\},\end{aligned} \tag{3.1}$$

where $\Omega \circ \Lambda = (\omega_{ij}\lambda_{ij})$ for two matrices $\Omega = (\omega_{ij})$ and $\Lambda = (\lambda_{ij})$.

Proposition 1. *Under Assumptions 1 and 2,*

$$\mathrm{E}(W_{nk}) = \mathrm{tr}[\{\Sigma - B_k(\Sigma)\}^2] \quad \textit{and} \quad \mathrm{Var}(W_{nk}) = \nu_{nk}^2 + o(\nu_{nk}^2).$$

The proposition indicates that under $H_{k,0}$, $\mathrm{E}(W_{nk}) = 0$ and $\nu_{nk} = 2\mathrm{tr}[\{B_k(\Sigma)\}^2]/n$, and ν_{nk}^2 is the leading order variance of W_{nk}. It can be shown that $\mathrm{tr}\{\Sigma(\Sigma - B_k(\Sigma))\}^2 \le 4(k+1)^2\mathrm{tr}(\Sigma^4)$. Since

$$\mathrm{tr}\{\Gamma'(\Sigma - B_k(\Sigma))\Gamma \circ \Gamma'(\Sigma - B_k(\Sigma))\Gamma\} \le \mathrm{tr}\{\Sigma(\Sigma - B_k(\Sigma))\}^2,$$

$\Delta \ge -2$ and $\mathrm{tr}(\Sigma^4)/\mathrm{tr}^2(\Sigma^2) = O(p^{-1})$, we have

$$4n^{-2}\mathrm{tr}^2(\Sigma^2) \le \nu_{nk}^2 \le C_0 a_{np}\mathrm{tr}^2(\Sigma^2) \tag{3.2}$$

for a constant $C_0 \ge 4$ and $a_{np} = n^{-2} + k^2(np)^{-1}$. We note that $a_{np} \to 0$ as $n \to \infty$ since $k = o(p^{1/4})$. In particular, if k is fixed, $a_{np} = O(n^{-2})$.

The following theorem establishes the asymptotic normality of W_{nk}.

Theorem 1. *Under Assumptions 1 and 2, and if $k = o(p^{1/4})$,*

$$\frac{W_{nk} - \mathrm{tr}[\{\Sigma - B_k(\Sigma)\}^2]}{\nu_{nk}} \xrightarrow{D} N(0,1).$$

In order to formulate a test procedure based on the asymptotic normality, we need to estimate $\mathrm{tr}[\{B_k(\Sigma)\}^2]$ since $\nu_{nk} = 2\mathrm{tr}[\{B_k(\Sigma)\}^2]/n$ under $H_{k,0}$. Let $V_{nk} := \hat{D}_{n0} + 2\sum_{q=1}^{k} \hat{D}_{nq}$ be the estimator, whose consistency to $\mathrm{tr}[\{B_k(\Sigma)\}^2]$ is implied in the following proposition.

Proposition 2. *Under Assumptions 1 and 2,* $\mathrm{Var}\{V_{nk}/\mathrm{tr}(\Sigma^2)\} = O(a_{np})$, *where* $a_{np} = n^{-2} + k^2(np)^{-1}$.

Since $\mathrm{E}(V_{nk}) = \mathrm{tr}[\{B_k(\Sigma)\}^2]$ and $a_{np} \to 0$, Proposition 2 means that, under $H_{k,0}$, $V_{nk}/\mathrm{tr}[\{B_k(\Sigma)\}^2] \xrightarrow{p} 1$ as $n \to \infty$. This together with Theorem 1 indicates that under $H_{k,0}$

$$T_{nk} =: n\frac{W_{nk}}{V_{nk}} \xrightarrow{D} N(0,4).$$

This leads to our choice of T_{nk} as the test statistic and the proposed test of size α that rejects $H_{k,0}$ if $T_{nk} \geq 2z_\alpha$ where z_α is the upper α quantile of $N(0,1)$.

As Theorem 1 prescribes the asymptotic normality under both $H_{k,0}$ and $H_{k,1}$, it permits a power evaluation of the test. Let

$$\delta_{nk} = \frac{\mathrm{tr}(\Sigma^2) - \mathrm{tr}[\{B_k(\Sigma)\}^2]}{\nu_{nk}}, \tag{3.3}$$

which may be viewed as a signal to noise ratio for the testing problem. This is because $\mathrm{tr}[\{\Sigma - B_k(\Sigma)\}^2]$ is the square of Frobenius norm of the difference between Σ and its k-banded version, and ν_{nk} measures the level of noise in the statistic W_{nk}. Then, the power of the test under $H_{k,1} : \Sigma \neq B_k(\Sigma)$ is

$$\begin{aligned}\beta_{nk} &= P\{nW_{nk}/V_{nk} \geq 2z_\alpha | \Sigma \neq B_k(\Sigma)\} \\ &= P\left(\frac{W_{nk} - \mathrm{tr}(\Sigma^2) + \mathrm{tr}[\{B_k(\Sigma)\}^2]}{\nu_{nk}} \geq \frac{2z_\alpha V_{nk}}{n\nu_{nk}} - \delta_{nk}\right).\end{aligned}$$

Since $\nu_{nk} \geq 2n^{-1}\mathrm{tr}(\Sigma^2)$, then $2V_{nk}/(n\nu_{nk}) \leq V_{nk}/\mathrm{tr}(\Sigma^2)$ for n large. Hence, asymptotically,

$$\beta_{nk} \geq P\left(\frac{W_{nk} - \mathrm{tr}(\Sigma^2) + \mathrm{tr}[\{B_k(\Sigma)\}^2]}{\nu_{nk}} \geq z_\alpha \frac{V_{nk}}{\mathrm{tr}(\Sigma^2)} - \delta_{nk}\right). \tag{3.4}$$

To gain more insight on the power, let $r_k = \mathrm{tr}[\{B_k(\Sigma)\}^2]/\mathrm{tr}(\Sigma^2)$. Clearly, $r_k \leq 1$ and is monotone non-decreasing with respect to k. If Σ is banded with bandwidth k_0, then

$$r_k < 1 \quad \text{for } k < k_0 \quad \text{and} \quad r_k = 1 \quad \text{for } k \geq k_0. \tag{3.5}$$

From the bounds for ν_{nk} in (3.2), it follows that

$$(C_0 a_{np})^{-1/2}(1-r_k) \le \delta_{nk} \le \tfrac{1}{2}n(1-r_k), \tag{3.6}$$

which indicates that $a_{np}^{-1/2}(1-r_k) = O(\delta_{nk})$. When k is fixed, $a_{np} = O(n^{-2})$ and $\delta_{nk} \sim n(1-r_k)$, indicating that δ_{nk} is at the exact order of $n(1-r_k)$.

Theorem 2. *Under Assumptions 1 and 2, $H_{k,1}$, and if $k = o(p^{1/4})$, then*
(i) $\liminf_n \beta_{nk} \ge 1 - \Phi\left(z_\alpha - \liminf_n \delta_{nk}\right)$*;*
(ii) If $a_{np}^{-1/2}(1-r_k) \to \infty$, then $\beta_{nk} \to 1$ as $n \to \infty$.

Theorem 2 indicates that the proposed test is consistent as long as the speed of $1 - r_k \to 0$ under $H_{k,1}$ is not faster than $a_{np}^{1/2}$. The test will have non-trivial power as long as $\liminf_n \delta_{nk} > 0$. If $n(1-r_k) \to 0$, the test will have no power beyond the significant level α. We note that this happens when $H_{k,0}$ and $H_{k,1}$ are extremely close to each other, so that $1 - r_k$ decays to zero faster than n^{-1}. We are actually a little amazed by the fact that the test is powerful as long as $\liminf_n a_{np}^{-1/2}(1-r_k) > 0$ or equivalently $(1-r_k)$ does not shrink to zero faster than $a_{np}^{1/2}$, despite the high-dimensionality and a possible diverging bandwidth k. Theorem 2 and (3.6) together imply that if r_k does not vary much as p increases, the power of the test will be largely determined by n, as confirmed by our simulation study in Section 5.

Our proposed test is targeted on the covariance matrix Σ. A test for the correlation matrix can be developed by modifying the test statistic by first standardizing each data dimension via its sample standard deviation. The theoretical justification would be quite involved, and would require extra effort. In addition to be invariant under the location shift, the test statistic is invariant if all the variables among the high-dimensional data vector are transformed by a common scale. However, the proposed test statistic is not invariant under variable-specific scale transformation. The above mentioned test for the correlation matrix would be invariant under variable-specific scale transformation.

4. Bandwidth estimation. We propose in this section an estimator to the bandwidth of banded covariance Σ. Estimating the bandwidth of a banded covariance matrix is an important and practical issue, given the latest advances on covariance estimation by banding [Bickel and Levina (2008a)] or tapering [Cai, Zhang and Zhou (2010)] sample covariance matrices. Indeed, finding an adequate bandwidth is a pre-requisite for applying either the banding or tapering estimators.

The proposed estimator is motivated by the test procedure developed in the previous section. Let k_0 be the true bandwidth. As the proposed test is consistent as long as $r_k \to 1$ not too fast, and the sample size is large enough (can still be much less than p), the proposed test would reject (not reject) $H_{k,0}$ for k less (larger) than k_0. An immediate but rather naive strategy would be to use the smallest integer k such that $H_{k,0}$ is not rejected as the bandwidth estimator. However, this strategy may be insufficient to counter "abnormal" samples which can produce larger (smaller) values of the statistic $\tilde{T}_{nk} := W_{nk}/V_{nk}$ consistently for a wide range of k values, when in fact $H_{k,0}$ ($H_{k,1}$) is true. And yet these "abnormal" samples are expected within the normal range of variations. To make the estimator robust against these "abnormal" samples and not so much dependent on the significant level α, we consider an estimator based on the difference between successive statistics, $d_{nk} = \tilde{T}_{nk} - \tilde{T}_{n\,k+1}$.

We assume the true bandwidth k_0 be either fixed or diverging as long as

$$k_0(n^{-1/2} + p^{-1/4}) \to 0 \quad \text{as } n \to \infty, \tag{4.1}$$

which covers a quite wide range for the bandwidth. Note that

$$\tilde{T}_{nk} \;=\; \frac{W_{nk} - \mathrm{E}(W_{nk})}{\nu_{nk}} \frac{\nu_{nk}}{V_{nk}} + \frac{\mathrm{E}(W_{nk})}{V_{nk}}.$$

For $k \le M$, where $M = o(p^{1/4})$ is a pre-chosen sufficiently large integer, $\{W_{nk} - \mathrm{E}(W_{nk})\}/\nu_{nk}$ is stochastically bounded (Theorem 1) and from (3.2), we have

$$\tilde{T}_{nk} \;=\; O_p\left(\frac{a_{np}^{\frac{1}{2}} r_k^{-1} \mathrm{tr}[\{B_k(\Sigma)\}^2]}{V_{nk}} \right) + (r_k^{-1} - 1) \frac{\mathrm{tr}[\{B_k(\Sigma)\}^2]}{V_{nk}}.$$

Let $b_{n\,k} = V_{n\,k}/\mathrm{tr}[\{B_k(\Sigma)\}^2] - 1$. From Propositions 1 and 2,

$$\mathrm{E}(b_{n\,k}) = 0 \;\text{ and }\; \mathrm{Var}(b_{n\,k}) = O(a_{np} r_k^{-2}). \tag{4.2}$$

Since $\Sigma = B_{k_0}(\Sigma)$ is non-negative definite, $\mathrm{tr}(\Sigma^2) \le (2k_0 + 1)\mathrm{tr}[\{B_0(\Sigma)\}^2]$. Hence, for any k, $r_k \ge (2k_0 + 1)^{-1}$. These imply that

$$\tilde{T}_{nk} \;=\; O_p(a_{np}^{\frac{1}{2}} k_0) + (r_k^{-1} - 1)\{1 + o_p(1)\}. \tag{4.3}$$

It can be checked that $a_{np}^{\frac{1}{2}} k_0 \to 0$ under (4.1), which makes the first term on the right of the above equation negligible relative to the second term. And the second term is quite indicative between $k < k_0$ and $k \geq k_0$, since $r_k = 1$ for $k \geq k_0$.

To amplify the second term when $k < k_0$ while not inflicting the first term on the right of (4.3) too much, we consider multiplying n^δ on $\tilde{T}_{nk}$ for a small positive δ and let $d_{nk}^{(\delta)} = n^\delta(\tilde{T}_{nk} - \tilde{T}_{n\,k+1})$. The proposed bandwidth estimator is

$$\hat{k}_{\delta,\theta} = \min\{k : |d_{n\,k}^{(\delta)}| < \theta\}, \tag{4.4}$$

for a pair of tuning parameters $\delta > 0$ and $\theta > 0$. The following theorem gives the consistency of the bandwidth estimator for both fixed or diverging k_0.

Theorem 3. *Under Assumptions 1 and 2, if* $\liminf_n\{\inf_{k<k_0}(r_{k+1} - r_k)\} > 0$*, then for any* $\theta > 0$*,* $\hat{k}_{\delta,\theta} - k_0 \xrightarrow{p} 0$ *under either of the two settings: (i) for any* $\delta \in (0,1)$ *if* k_0 *is bounded; (ii) for any* $\delta \in (0,1/2]$ *if* k_0 *is diverging but satisfies (4.1), and* $\{\sigma_{ll}\}_{l=1}^p$ *are uniformly bounded away from* 0 *and* ∞*.*

We would like to remark that the multiplier n^δ in $d_{nk}^{(\delta)}$'s formation leads to θ being "free ranged" as long as $\theta > 0$. If such multiplication is not administrated, namely by setting $\delta = 0$, the range of θ needs to be restricted properly to ensure convergence. The requirement of $\liminf_n\{\inf_{k<k_0}(r_{k+1} - r_k)\} > 0$ is to avoid situations where Σ has segments of zero sub-diagonals followed by non-zero sub-diagonals when one moves away from the main diagonal. Our estimator can be modified to suit such situations. However, we would not elaborate here for the sake of simplicity in the presentation. Attaining the consistency of $\hat{k}_{\delta,\theta}$ with diverging k_0 requires a smaller δ value.

To better understand the theorem and the bandwidth estimator, we conducted a simulation study for $k_0 = 5$, $n = 60$ and $p = 600$ with X_i generated from Model (5.1) with a multivariate normal distribution. The detailed simulation setting will be provided in Section 5. Figure 1 presents box-plots of the modified statistics $n^\delta\tilde{T}_{nk}$ (Left panel) and its first-order difference $d_{nk}^{(\delta)}$ (right panel), with $\delta = 0.5$. We see from the right panel that the first five boxes are relatively large, and $d_{nk}^{(\delta)}$ is close to 0 while for $k \geq 5$. This indicates that five would be the bandwidth estimate.

In practical implementations with finite samples, the bandwidth estimator may be sensitive to the tuning parameters δ and θ. Note that, as revealed a few paragraphs earlier, d_{nk} should be significantly larger than 0 for $k < k_0$ and close to 0 for $k \geq k_0$. Such a pattern, as displayed in Figure 1, indicates that k_0 is a change point for $\{d_{nk}\}_{k=0}^{M}$. This motivates us to consider a regression change-point detection algorithm for bandwidth estimation. Consider d_{nj}, the difference between successive statistics T_{nj}, for $j = 1, \cdots, M$, for a sufficiently large M that covers the true bandwidth k_0. The idea is to fit, at each candidate k, a regression function $g_k(j)$ to $\{d_{nj}\}_{j=0}^{M}$ such that $g(j) \equiv g(k)$ for all $j > k$. We may fit a nonparametric, locally weighted linear regression [Cleveland and Devlin (1988); Fan and Gijbels (1996)] on $j \in L_k = \{l : 0 \leq l \leq k\}$ to the left of k with the smoothing window-width hk, where h is a smoothing parameter, and fit a flat line at the level d_{nk} for $j \in R_k = \{l : k+1 \leq l \leq M\}$ to the right of k. If k is too small for the above nonparametric regression, a parametric polynomial regression may be conducted. Let $\hat{g}_k(j)$ be the regression estimate, nonparametric or parametric, obtained over the set L_k, and let

$$err(k) = \sum_{j \in L_k} |\hat{g}_k(j) - d_{nj}| + \sum_{j \in R_k} |d_{nk} - d_{nj}|$$

be the absolute deviation of the fitted errors. Then a bandwidth estimator, as we call the change-Point estimator, is

$$\hat{k} = \arg\min_{k} \{err(k) : 1 \leq k \leq M\}. \tag{4.5}$$

Our empirical studies reported in Section 5 show this estimator worked quite well.

Bickel and Levina (2008a, 2008b) proposed a method to select the bandwidth based on a repeated random splitting of the original sample to two sub-samples of sizes n_1 and $n_2 = n - n_1$. Let $\hat{\Sigma}_1^v$ and $\hat{\Sigma}_2^v$ be the sample covariances based the sub-samples of sizes n_1 and n_2 respectively, where v denotes the vth split, for $v = 1, \cdots, N$, where N is the total numbers of sample splitting. The risk for each candidate k is defined to be $R(k) = \mathrm{E}||B_k(\hat{\Sigma}) - \Sigma||_{(1,1)}$, where for a $p_1 \times p_2$ matrix $A = (a_{ij})$, $||A||_{(1,1)} = \max_{1 \leq j \leq p_2} \sum_{i=1}^{p_1} |a_{ij}|$. An empirical version of the risk is

$$\hat{R}(k) = \frac{1}{N} \sum_{v=1}^{N} ||B_k(\hat{\Sigma}_1^v) - \hat{\Sigma}_2^v||_{(1,1)} \tag{4.6}$$

and the bandwidth estimator is

$$\hat{k}_{BL} = \arg\min_{0 \leq k \leq p-1} \hat{R}(k).$$

Bickel and Levina (2008a) recommended n_1 to be $n/3$, and the number of random splits $N = 50$, while Bickel and Levina (2008b) suggested $n_1 =$

$n(1-1/\log n)$ and using the Frobenius norm instead of the $||\cdot||_{(1,1)}$ norm. Rothman, Levina and Zhu (2010) considered a similar method to select the bandwidth in their estimator. We note that these approaches can be adversely impacted by high-dimensionality, due to the fact that $\hat{\Sigma}_2$ may be a poor estimator of Σ if p is much larger than n, as found in early works [Johnstone (2001); Bai and Silverstein (2005)].

5. Simulation results. In this section, we report results from simulation studies to verify the proposed test for the bandedness and the bandwidth estimator. We evaluate the performance of the proposed test under several different structures of covariance matrix for normal and gamma random vectors. We generate p-dimensional independent and identical multivariate random vectors $X_i = (X_{i1}, \cdots, X_{ip})'$ according to a model

$$X_{i\,j} = \sum_{l=0}^{k_0} \gamma_l Z_{i\,j+l}, \tag{5.1}$$

where k_0 is the bandwidth of the covariance, $\gamma_0 = 1$ in all settings and the other coefficients γ_l will be specified shortly. Two distributions are assigned to the i.i.d. Z_{ij}: (i) the normal distribution $N(0,1)$; (ii) the standardized $Gamma(1, 0.5)$ distribution so that it has zero mean and unit variance. To mimic the "large p, small n" paradigm, we choose $n = 20, 40, 60$ and $p = 50, 100, 300, 600$, respectively.

We first evaluate the size of the proposed test under the null hypothesis $H_{k,0} : \Sigma = B_k(\Sigma)$ for $k = 0$ (diagonal), $1, 2$ and 5. The coefficients γ_l for $l > 0$ are: $\gamma_1 = 1$ and 0.5, respectively, for $k = 1$; $\gamma_1 = \gamma_2 = 1$, and $\gamma_1 = 0.5$ and $\gamma_2 = 0.25$, respectively, for $k = 2$; and $\gamma_1 = \cdots = \gamma_5 = 0.4$ for $k = 5$. To assess the power, we generate data according to (5.1) so that $\Sigma = B_k(\Sigma)$ and test for $H_{k-1,0} : \Sigma = B_{k-1}(\Sigma)$ for $k = 2$ and 5, respectively, with the γ_l values being the same with those in the corresponding k in the simulation for the size reported above. We note that this design, having the bandwidth of the null hypothesis adjacent to the true bandwidth, is the hardest for the test, as the null and the alternative is the closest, given the setting of the parameters $\{\gamma_l\}$. All the simulation results are based on 1000 simulations.

We also evaluate the test proposed in Cai and Jiang (2012), based on the asymptotic distribution of the coherence statistic L_n under the same simulation settings used for the proposed test. The test encountered a very severe size distortion in that the real sizes are much less than the nominal level of 5%, which also caused the power of the test to be unfavorably low. For these reasons, we will not report the simulation results of the test. The coherence statistic is the largest Pearson correlation coefficients among all pairs of different components in X, and is an extreme value-type statistic. Extreme value statistics are known to be slowly converging, and a computing intensive method is needed to speed up its convergence. The asymptotic distribution established in Cai and Jiang (2012) may be the foundation to justify such a method.

Table 1 reports the empirical sizes of the proposed test at the 5% nominal significance for $H_{k,0}$ with $k = 0, 1, 2$ and 5, respectively, under both the normal and gamma distributions. Table 2 summarizes the empirical power of the tests whose sizes are reported in Table 1. To understand the power results, Table 2 also contains the values of $1 - r_k$ for each simulation setting. We observe from Table 1 that the test has reasonably empirical sizes, around 5%, and that the test is not sensitive to the dimensionality indicated by its robust performance. There is some size inflation, which is due to a number of factors, mainly to the dimensionality p, the sample size n and the approximation error of the finite sample distribution of the test statistic by the limiting normal distribution. We recall that the test statistic is a linear combination of U-statistics, whose convergence to the limiting normal distribution can be slow. In the simulations for power evaluation (reported in Table 2), we designed the simulation so that a constant r_k was maintained for a set of different ps, while n was held fixed. The empirical powers reported in Table 2 show that the power is quite reflective to the sample size n and $1 - r_k$, namely larger n or large $1 - r_k$ leads to higher power. This is because as r_k decreases, the signal of the test increases. So it becomes easier to distinguish the null hypothesis from the alternative. And after we controlled n and $1 - r_k$, the power was not sensitive to p at all, confirming a remark made at the end of Section 3.

For bandwidth estimation, we generate $\{X_i\}_{i=1}^n$ according to (5.1). While we keep $\gamma_0 = 1$, the other coefficients γ_l for $l > 0$ are:

$$\begin{aligned}
&\text{Bandwidth 3:} \quad \gamma_i = 1, \quad \text{for} \quad i = 1, 2, 3;\\
&\text{Bandwidth 5:} \quad \gamma_i = 0.4 \quad \text{for} \quad 1 \le i \le 5;\\
&\text{Bandwidth 10:} \quad \gamma_i = 0.2 \quad \text{for} \quad 1 \le i \le 5 \quad \text{and} \quad \gamma_i = 0.4 \quad \text{for} \quad 6 \le i \le 10;\\
&\text{Bandwidth 15:} \quad \gamma_i = 0.2 \quad \text{for} \quad 1 \le i \le 10 \quad \text{and} \quad \gamma_i = 0.4 \quad \text{for} \quad 11 \le i \le 15.
\end{aligned}$$

The covariances have bandwidth 3, 5, 10 and 15 respectively. We evaluate two bandwidth estimators. One is $\hat{k}_{\delta,\theta}$ given in (4.4) with $\delta = 0.5$ and $\theta = 0.06$, namely $\hat{k}_{0.5,0.06}$, and the other is the change-point estimator given in (4.5), applied on candidate ks whose p-values for H_{0k} are larger than 10^{-10}. We employ the LOESS algorithm in R to carry our the nonparametric regression estimation to the left of a k, with a default smoothing parameter $h = 0.75$.

For each Σ, we compare the proposed bandwidth estimators with the estimators advocated in Bickel and Levina (2008a, 2008b) and Rothman, Levina and Zhu (2010). We choose n to be 20, 40 and 60. For each n, p is chosen 2 times, 5 times and 10 times of n, respectively. Following the settings of Bickel and Levina (2008a) and (2008b), n_1 is chosen to be $n/3$ and $n(1 - 1/\log n)$, respectively, and the number of random splits in (4.6) is $N = 50$.

Table 3 reports the average empirical bias and standard deviation of the five bandwidth estimators based on 100 replications. We observe from Table 3 that the overall performance of the proposed estimators is better than those of Bickel and Levina (2008a, 2008b) and Rothman, Levina and Zhu (2010), with smaller standard deviation and bias. Moreover, as n is increased, both the bias and standard deviation of the proposed estimators decreased, and are quite robust to p, which is a nice property to have. For the estimators of Bickel and Levina (2008a, 2008b) and Rothman, Levina and Zhu (2010), the bias and the standard deviation could increase along with the increase of p, and are much larger than those of the proposed estimators. These are likely caused by the problems associated with the sample covariance matrix when the data dimension is high.

6. Empirical study. In this section, we report an empirical study on a prostate cancer data set [Adam et al. (2003)] from protein mass spectroscopy, which was aimed to distinguish the healthy people from the ones with the cancer by analyzing the constituents of the proteins in the blood. Adam et al. (2003) recorded for each blood serum sample i, the intensity X_{ij} for a large number of time-of-flight values t_j. The time of flight is related to the mass over charge ratio m/z of the constituent proteins. They collected the intensity in the total of 48538 m/z-sites and the full data set consisted of 157 healthy patients and 167 with cancer.

Tibshirani et al. (2005) analyzed the data by the fused Lasso. They ignored m/z-ratios below 2000 to avoid chemical artifacts, and averaged the intensity recordings in consecutive blocks of 20. These gave rise to a total of 2181 dimensions per observation. Levina, Rothman and Zhu (2008) estimated the inverse of the covariance matrix of the intensities by an adaptive banding approach with a nested Lasso penalty. They carried out additional averaging of the data of Tibshirani et al. (2005) in consecutive blocks of 10, resulting in a total of 218 dimensions. We considered the standardized data of Levina, Rothman and Zhu (2008), and tested for the banded structure of the covariance matrix of the intensities.

The test statistics, p-values and the first order differences d_{nk} for the healthy and cancer groups are displayed in Figure 2 for bandwidths $k \geq 50$. We do not display in the figure for bandwidths less than 50 since the values of the test statistics are too large, and the associated p-values for H_{0k} are too small for $k < 50$. These bandwidth estimates together with the shapes of the curves for the test statistics and the p-values in Figure 2 suggest that the covariance matrix of the healthy group is likely to be banded, while the covariance of the cancer group may not be banded at all, given the very large bandwidth and the shape of the curve. For the cancer group, as shown in Figure 2, the test statistics are relatively flat for $120 \leq k \leq 140$, and then fall sharply afterward, which indicates relatively small values in the covariance matrix from sub-diagonal 120 to 140. However, there is a substantial contribution from sub-diagonals for $k > 140$. These are echoed in the p-values displayed in panel (b) with almost stationary p-values within the above mentioned range, followed by a sharp increase. Panel (d) of Figure 2 displays a rather unsettled curve for d_{nk}, the difference between successive statistics $\tilde{T}_{nk}$. These are all in sharp contrasts to those of the healthy group, indicating rather different covariance structures between the two groups.

At $\alpha = 5\%$, we reject a $H_{k,0}$ when the statistic is larger than 3.29. For the healthy group, the smallest k such that $H_{k,0}$ is not rejected is $k = 116$, while for the cancer group is 191. We apply the bandwidth estimator (4.4) with $\delta = 0.5$ and $\theta = 0.005$. The estimated bandwidth for the health group is 121 and for the cancer group is 212. At the same time, the bandwidth estimates, by employing Bickel and Levina's (2008a) approach, are 144 for the healthy group and 193 for the cancer group. The one for the healthy group is much larger than the 121 we obtained earlier, using the estimator (4.4). We then apply the proposed regression change-point bandwidth estimator over a range of bandwidths whose associated p-values for testing H_{0k} are larger than 10^{-10}. For the healthy group, the bandwidth range is $k \geq 85$; for the cancer group the range is $k \geq 150$. We set the smoothing parameter $h = 0.75$ in the LOESS procedure in R. The regression bandwidth estimator is $\hat{k}_h = 127$ for the healthy group, which is slightly larger than the 121 obtained from the estimator (4.4). For the cancer group, the estimated bandwidth is 215. This rather large estimated bandwidth suggest that, compared to the healthy group, there is substantially more dependence among the protein mass spectroscopy measurements among the cancer patients, and, in particular, the covariance may not be banded at all for this group of patients.

Acknowledgements We thank an Associate Editor and two referees for constructive comments and suggestions which have lead to improvements in the presentation of the paper. Chen acknowledges the support of a NSFC key Grant 11131002. We also thank Professor Ji Zhu for sharing the data.

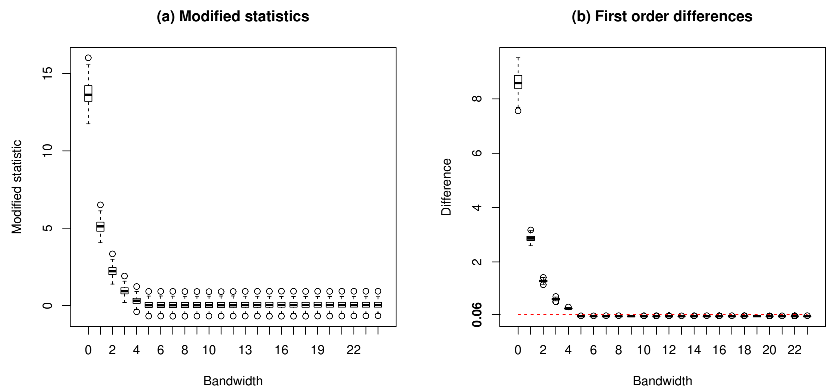

Fig 1: Box-plots of the modified statistics $n^{\delta}\tilde{T}_{nk}$ and their first order differences of the simulated data. The dash line in the right panel is $\theta = 0.06$. The true bandwidth is 5.

Appendix. We first introduce some notation. For $q = 0, \cdots, p$, define

$$B_{1,q} = \frac{1}{P_n^2} \sum_{l=1}^{p-q} \sum_{i,j}^{*} (X_{il} X_{il+q})(X_{jl} X_{jl+q}),$$

$$B_{2,q} = \frac{1}{P_n^3} \sum_{l=1}^{p-q} \sum_{i,j,k}^{*} X_{il} X_{kl+q}(X_{jl} X_{jl+q}) \quad \text{and}$$

$$B_{3,q} = \frac{1}{P_n^4} \sum_{l=1}^{p-q} \sum_{i,j,k,m}^{*} X_{il} X_{jl+q} X_{kl} X_{ml+q}.$$

Then, $V_{nk} = B_{1,0} - 2B_{2,0} + B_{3,0} + 2\sum_{q=1}^{k}(B_{1,q} - 2B_{2,q} + B_{3,q})$, and $W_{nk} = 2\sum_{q=k+1}^{p-1}(B_{1,q} - 2B_{2,q} + B_{3,q})$. Let $C_{nk} = 2\sum_{q=k+1}^{p-1} B_{1,q}$ and $U_i = B_{i,0} + 2\sum_{q=1}^{p-1} B_{i,q}$ for $i = 1, 2, 3$. We first establish some lemmas for later use.

Lemma 1. *Under Assumptions 1 and 2,* $\mathrm{Var}(C_{nk}) = \nu_{nk}^2 + o\{n^{-2}\mathrm{tr}^2(\Sigma^2)\}$.

Table 1
Empirical sizes of the proposed test at 5% significance for the normal and gamma random vectors generated according to model (5.1).

(a) $H_0 : \Sigma = B_0(\Sigma)$

	Normal				Gamma			
	p				p			
n	50	100	300	600	50	100	300	600
20	0.069	0.065	0.061	0.066	0.055	0.056	0.065	0.075
40	0.067	0.049	0.047	0.060	0.056	0.054	0.055	0.059
60	0.066	0.064	0.045	0.051	0.068	0.039	0.065	0.049

(b) $H_0 : \Sigma = B_1(\Sigma)$

	$\gamma_1 = 1$							
20	0.069	0.061	0.056	0.060	0.062	0.058	0.069	0.069
40	0.061	0.048	0.048	0.069	0.059	0.049	0.069	0.075
60	0.045	0.053	0.056	0.067	0.048	0.061	0.068	0.059
	$\gamma_1 = 0.5$							
20	0.065	0.069	0.058	0.067	0.063	0.061	0.057	0.061
40	0.063	0.052	0.047	0.068	0.059	0.055	0.066	0.071
60	0.050	0.056	0.057	0.061	0.050	0.070	0.068	0.060

(c) $H_0 : \Sigma = B_2(\Sigma)$

	$\gamma_1 = \gamma_2 = 1$							
20	0.058	0.050	0.055	0.058	0.056	0.046	0.062	0.062
40	0.049	0.042	0.051	0.058	0.059	0.048	0.076	0.071
60	0.050	0.043	0.065	0.064	0.040	0.063	0.065	0.052
	$\gamma_1 = 0.5$, $\gamma_2 = 0.25$							
20	0.060	0.055	0.056	0.061	0.059	0.054	0.062	0.062
40	0.055	0.047	0.055	0.059	0.058	0.046	0.071	0.064
60	0.044	0.043	0.058	0.060	0.042	0.060	0.067	0.061

(d) $H_0 : \Sigma = B_5(\Sigma)$ with $\gamma_1 = \cdots = \gamma_5 = 0.4$

20	0.045	0.058	0.067	0.059	0.050	0.061	0.054	0.064
40	0.043	0.054	0.049	0.061	0.041	0.052	0.065	0.064
60	0.031	0.046	0.065	0.069	0.034	0.040	0.053	0.048

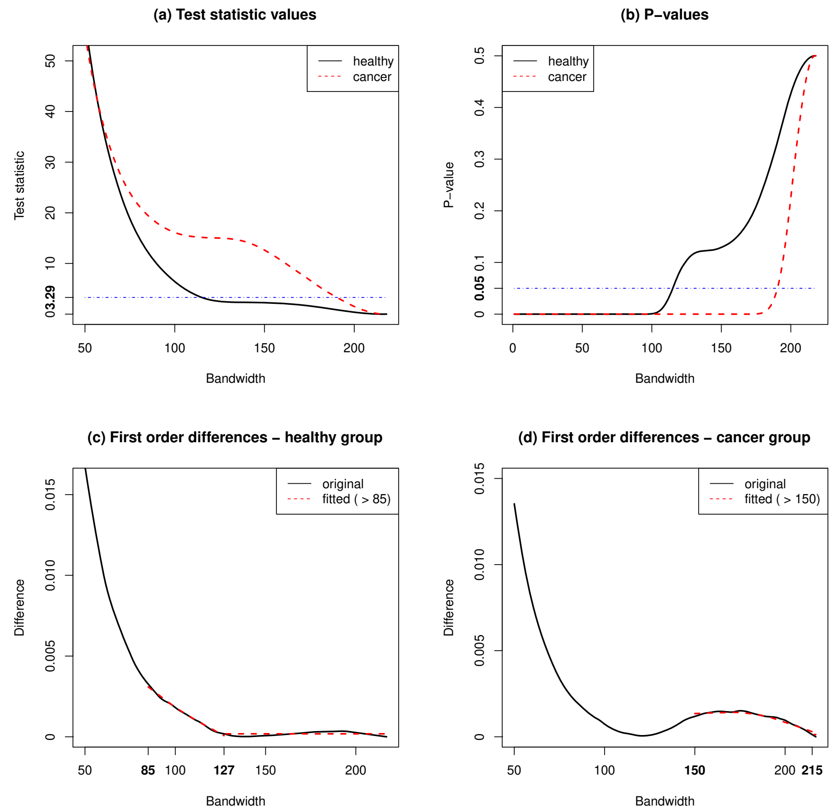

Fig 2: Test statistics, p-values and the first order differences d_{nk} for the healthy and cancer groups for bandwidths larger than 50. The p-values of the test for H_{0k} for $k < 50$ were too small to be considered for bandwidth estimation.

Proof. Since $C_{nk} = (P_n^2)^{-1} \sum_{i,j}^{*} \sum_{|l_1 - l_2| > k} X_{il_1} X_{il_2} X_{jl_1} X_{jl_2}$, by the independence between different observations, we have

$$\mathrm{E}(C_{nk}) = (P_n^2)^{-1} \sum_{i,j}^{*} \sum_{|l_1 - l_2| > k} \mathrm{E}(X_{il_1} X_{il_2}) \mathrm{E}(X_{jl_1} X_{jl_2}) = \sum_{|l_1 - l_2| > k} \sigma_{l_1 l_2}^2.$$

Note that

$$C_{nk}^2 = (P_n^2)^{-2} \sum_{i_1, j_1}^{*} \sum_{i_2, j_2}^{*} \sum_{|l_1 - l_2| > k} \sum_{|l_3 - l_4| > k} X_{i_1 l_1} X_{i_1 l_2} X_{i_2 l_3} X_{i_2 l_4} X_{j_1 l_1} X_{j_1 l_2} X_{j_2 l_3} X_{j_2 l_4}.$$

Let $f_{l_1 l_2 l_3 l_4} = \sum_m \Gamma_{l_1 m} \Gamma_{l_2 m} \Gamma_{l_3 m} \Gamma_{l_4 m}$ and $\sigma_{l_1 l_2} \sigma_{l_3 l_4}[3] = \sigma_{l_1 l_2} \sigma_{l_3 l_4} + \sigma_{l_1 l_3} \sigma_{l_2 l_4} +$

Table 2
Empirical power of the proposed test at $\alpha = 5\%$ for the normal and gamma random vectors generated according to model (5.1).

(a) $H_0 : \Sigma = B_1(\Sigma)$ when $\Sigma = B_2(\Sigma)$

	Normal				Gamma			
	p				p			
n	50	100	300	600	50	100	300	600
	$\gamma_1 = \gamma_2 = 1,\ 1 - r_1 = 1/14$							
20	0.300	0.313	0.330	0.336	0.315	0.312	0.340	0.312
40	0.683	0.722	0.711	0.702	0.710	0.721	0.752	0.741
60	0.962	0.964	0.952	0.954	0.958	0.955	0.950	0.949
	$\gamma_1 = 0.5,\ \gamma_2 = 0.25,\ 1 - r_1 = 1/35$							
20	0.146	0.144	0.139	0.152	0.148	0.140	0.147	0.143
40	0.269	0.253	0.258	0.279	0.256	0.281	0.311	0.311
60	0.406	0.443	0.455	0.451	0.438	0.449	0.458	0.441

(b) $H_0 : \Sigma = B_4(\Sigma)$ when $\Sigma = B_5(\Sigma)$ with $\gamma_1 = \cdots = \gamma_5 = 0.4,\ 1 - r_4 = 1/38.05$

20	0.090	0.112	0.119	0.123	0.096	0.112	0.108	0.118
40	0.149	0.181	0.178	0.200	0.161	0.169	0.218	0.196
60	0.261	0.284	0.328	0.314	0.246	0.297	0.290	0.284

$\sigma_{l_1 l_4}\sigma_{l_2 l_3}$. Then, $\mathrm{E}(C_{nk}^2) = (P_n^2)^{-2}(L_{n1} + L_{n2} + L_{n3})$, where

$$L_{n1} = P_n^4 \sum_{|l_1-l_2|>k} \sum_{|l_3-l_4|>k} \sigma_{l_1 l_2}^2 \sigma_{l_3 l_4}^2,$$

$$L_{n2} = 4P_n^3 \sum_{|l_1-l_2|>k} \sum_{|l_3-l_4|>k} (\Delta f_{l_1 l_2 l_3 l_4} + \sigma_{l_1 l_2}\sigma_{l_3 l_4}[3])\sigma_{l_1 l_2}\sigma_{l_3 l_4} \quad \text{and}$$

$$L_{n3} = 2P_n^2 \sum_{|l_1-l_2|>k} \sum_{|l_3-l_4|>k} (\Delta f_{l_1 l_2 l_3 l_4} + \sigma_{l_1 l_2}\sigma_{l_3 l_4}[3])^2.$$

We compute L_{n2} and L_{n3} part by part. First, note that

$$\sum_{|l_1-l_2|>k} \sum_{|l_3-l_4|>k} f_{l_1,l_2,l_3,l_4}\sigma_{l_1 l_3}\sigma_{l_2 l_4}$$
$$= \mathrm{tr}(A^2 \circ A^2) - 2 \sum_{|l_1-l_2|\le k} \sum_{l_3,l_4} f_{l_1 l_2 l_3 l_4}\sigma_{l_1 l_3}\sigma_{l_2 l_4} + \sum_{|l_1-l_2|\le k} \sum_{|l_3-l_4|\le k} f_{l_1 l_2 l_3 l_4}\sigma_{l_1 l_3}\sigma_{l_2 l_4}.$$

TABLE 3
Averaged empirical bias (standard deviation) of the five bandwidth estimators: estimator (4.4) with $\delta = 0.5$ and $\theta = 0.06$ (Fixed), the change-point estimator (4.5) (change-Point) with $h = 0.75$ and the estimators proposed in Bickel and Levina (2008a) (BLa), Bickel and Levina (2008b) (BLb) and Rothman, Levina and Zhu (RLZ).

n	p	method	Bandwidth 3	5	10	15
		Fixed	0.58(1.465)	0.07(0.946)	-0.5(1.114)	-1.63(1.931)
		Change-Point	0.60(0.569)	-0.21(0.518)	-1.48(2.134)	0.06(1.734)
	40	BLa	-0.66(0.855)	-0.86(1.287)	-4.72(2.202)	-9.19(2.246)
		BLb	0.59(1.036)	-0.53(1.460)	-3.97(2.932)	-6.63(4.403)
		RLZ	0.11(1.363)	-0.18(1.855)	-2.55(2.732)	-8.02(2.760)
		Fixed	0.14(0.636)	0.1(0.659)	-0.22(0.440)	-0.96(0.875)
		Change-Point	0.56(0.499)	-0.07(0.293)	-0.52(0.882)	0.18(0.968)
20	100	BLa	-0.09(1.272)	0.45(1.617)	-2.33(2.010)	-6.14(2.686)
		BLb	0.7 (1.219)	-0.26(1.561)	-3.88(2.772)	-7.29(3.506)
		RLZ	0.45(1.861)	-0.59(1.799)	-3.79(2.203)	-8.8(2.103)
		Fixed	0.01(0.1)	0(0)	-0.12(0.327)	-0.66(0.728)
		Change-Point	0.67(0.473)	0(0)	-0.18(0.435)	0.09(0.379)
	200	BLa	0.78(2.077)	1.14(2.327)	-0.58(2.637)	-2.55(3.560)
		BLb	1.18(1.935)	-0.1(2.302)	-2.91(2.878)	-6.14(3.579)
		RLZ	0.55(1.641)	-0.29(1.719)	-4.6(1.928)	-9.51(1.823)
		Fixed	0.14(0.551)	0.08(0.464)	-0.01(0.1)	-0.10(0.302)
		Change-Point	0.47(0.502)	-0.01(0.100)	-0.12(0.383)	0.08(0.273)
	80	BLa	-0.24(0.780)	0.23(1.014)	-1.32(1.663)	-3.55(2.907)
		BLb	1.5(1.514)	0.94(1.427)	0.06(2.210)	-0.17(3.260)
		RLZ	1.05(1.629)	0.71(2.222)	0.72(2.374)	1.28(3.229)
		Fixed	0(0)	0(0)	0(0)	-0.04(0.197)
		Change-Point	0.55(0.500)	0(0)	-0.04(0.281)	0.02(0.141)
40	200	BLa	0.29(1.200)	1.03(1.322)	0.28(1.633)	-1.30(2.285)
		BLb	1.64(1.605)	1.24(1.837)	0.58(2.833)	-0.1(2.976)
		RLZ	1.36(2.435)	1.16(2.465)	2.07(3.647)	1.07(2.861)
		Fixed	0(0)	0(0)	0(0)	0(0)
		Change-Point	0.56(0.499)	0(0)	0(0)	0(0)
	400	BLa	0.88(1.754)	1.5(1.962)	1.25(2.240)	0.22(2.642)
		BLb	2.61(2.457)	1.74(2.493)	0.68(3.396)	0.09(3.715)
		RLZ	2.19(2.943)	1.98(3.369)	1.17(3.420)	-0.39(2.821)
		Fixed	0.02(0.141)	0.08(0.706)	0.02(0.2)	-0.01(0.1)
		Change-Point	0.52(0.502)	0(0)	0(0)	-0.01(0.1)
	120	BLa	0.22(0.938)	0.85(0.989)	0.14(1.363)	-0.88(1.659)
		BLb	1.71(1.458)	1.52(1.541)	1.67(2.108)	1.49(2.615)
		RLZ	1.24(1.753)	0.71(1.431)	2.03(2.683)	2.13(2.845)
		Fixed	0(0)	0(0)	0(0)	0(0)
		Change-Point	0.58(0.496)	0(0)	0(0)	0(0)
60	300	BLa	0.47(1.439)	1.56(1.683)	1.06(2.136)	0.70(2.452)
		BLb	2.15(2.017)	2.04(2.474)	1.73(2.877)	1.74(2.922)
		RLZ	1.68(2.188)	1.02(2.383)	2.45(3.686)	2.75(3.331)
		Fixed	0(0)	0(0)	0(0)	0(0)
		Change-Point	0.54(0.501)	0(0)	0(0)	0(0)
	600	BLa	1.05(1.702)	1.92(2.102)	2.01(2.393)	1.06(2.490)
		BLb	3.16(2.631)	2.87(2.699)	2.97(3.532)	1.33(3.254)
		RLZ	3.3(3.721)	3.23(3.787)	3.82(4.029)	2.7(3.506)

By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&|\sum_{|l_1-l_2|\le k}\sum_{l_3,l_4} f_{l_1l_2l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4}| \le \mathrm{tr}^{\frac{1}{2}}(T^2)\mathrm{tr}^{\frac{1}{2}}[(\Sigma\Gamma\circ\Sigma\Gamma)\{(\Sigma\Gamma)'\circ(\Sigma\Gamma)'\}] \quad \text{and} \\
&|\sum_{|l_1-l_2|\le k}\sum_{|l_3-l_4|\le k} f_{l_1l_2l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4}| \le (2k+1)^2\mathrm{tr}\{(\Gamma\circ\Gamma)(\Gamma'\circ\Gamma')(\Sigma\circ\Sigma)\},
\end{aligned}
$$

where $T=(\Gamma\circ\Gamma)(\Gamma'\circ\Gamma')$. Note that

$$
\begin{aligned}
\mathrm{tr}(T)\le\mathrm{tr}(\Sigma^2),\ \ \mathrm{tr}\{(\Gamma\circ\Gamma)(\Gamma'\circ\Gamma')(\Sigma\circ\Sigma)\}\le\mathrm{tr}(\Sigma^4)) \quad \text{and} \\
\mathrm{tr}\big[(\Sigma\Gamma\circ\Sigma\Gamma)\{(\Sigma\Gamma)'\circ(\Sigma\Gamma)'\}\big]\le\mathrm{tr}(\Sigma^6).
\end{aligned}
$$

Since $\mathrm{tr}(\Sigma^6)\le\mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma^4)$, $k=o(p^{1/4})$ and from Assumption 2, it follows that

$$
\begin{aligned}
\sum_{|l_1-l_2|\le k}\sum_{l_3,l_4} f_{l_1l_2l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4} &= o\{\mathrm{tr}^2(\Sigma^2)\} \quad \text{and} \\
\sum_{|l_1-l_2|>k}\sum_{|l_3-l_4|>k} f_{l_1l_2l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4} &= o\{\mathrm{tr}^2(\Sigma^2)\}.
\end{aligned}
$$

Similarly, it can be shown that

$$
\begin{aligned}
&\sum_{|l_1-l_2|\le k}(\Sigma^2)^2_{l_1l_2}=o\{\mathrm{tr}^2(\Sigma^2)\}, &&\sum_{|l_1-l_2|\le k}\sum_{|l_3-l_4|\le k}\sigma^2_{l_2l_4}\sigma^2_{l_1l_3}=o\{\mathrm{tr}^2(\Sigma^2)\},\\
&\sum_{|l_1-l_2|\le k}(\Sigma^2)_{l_1l_1}(\Sigma^2)_{l_2l_2}=o\{\mathrm{tr}^2(\Sigma^2)\}, &&\sum_{|l_1-l_2|>k}\sum_{|l_3-l_4|>k}f^2_{l_1,l_2,l_3,l_4}=o\{\mathrm{tr}^2(\Sigma^2)\},
\end{aligned}
$$

and $\sum_{|l_1-l_2|\le k}\sum_{|l_3-l_4|\le k}\sigma_{l_1l_3}\sigma_{l_2l_4}\sigma_{l_1l_4}\sigma_{l_2l_3}=o\{\mathrm{tr}^2(\Sigma^2)\}$. By combining these together,

$$
\begin{aligned}
\mathrm{Var}(C_{nk}) =& 4n^{-2}\mathrm{tr}^2(\Sigma^2)+8n^{-1}\sum_{|l_1-l_2|>k}\sum_{|l_3-l_4|>k}\sigma_{l_1l_3}\sigma_{l_2l_4}\sigma_{l_1l_2}\sigma_{l_3l_4} \\
&+4\Delta n^{-1}\sum_{|l_1-l_2|>k}\sum_{|l_3-l_4|>k}f_{l_1l_2l_3l_4}\sigma_{l_1l_2}\sigma_{l_3l_4}+o\left(n^{-2}\mathrm{tr}^2(\Sigma^2)\right).
\end{aligned}
$$

It can be checked that

$$
\begin{aligned}
&\sum_{|l_1-l_2|>k}\sum_{|l_3-l_4|>k}\sigma_{l_1l_3}\sigma_{l_2l_4}\sigma_{l_1l_2}\sigma_{l_3l_4}=\mathrm{tr}\{\Sigma(\Sigma-B_k(\Sigma))\}^2 \quad \text{and} \\
&\sum_{|l_1-l_2|>k}\sum_{|l_3-l_4|>k}f_{l_1l_2l_3l_4}\sigma_{l_1l_2}\sigma_{l_3l_4}=\mathrm{tr}(\Gamma'(\Sigma-B_k(\Sigma))\Gamma\circ\Gamma'(\Sigma-B_k(\Sigma))\Gamma).
\end{aligned}
$$

Therefore, $\mathrm{Var}(C_{nk}) = \nu_{nk}^2 + o\{n^{-2}\mathrm{tr}^2(\Sigma^2)\}$. $\square$

Lemma 2. *Under Assumptions 1 and 2, for $q = 0, \cdots, k$,*

$$\mathrm{Var}(B_{2,q}) = O\{n^{-2}\mathrm{tr}^{\frac{1}{2}}(\Sigma^4)\mathrm{tr}(\Sigma^2)\} \qquad \textit{and} \qquad \mathrm{Var}(B_{3,q}) = O\{n^{-4}\mathrm{tr}(\Sigma^4)\}.$$

Proof. First consider $B_{2,q}$. Since $\mathrm{E}B_{2,q} = 0$ for any $q = 0, \cdots, k$, we only need to calculate $\mathrm{E}B_{2,q}^2$. Note that we can decompose $B_{2,q}^2$ as

$$B_{2,q}^2 = (P_n^3)^{-2}(\sum_{i=1}^{2} B_{2,q,a_i} + \sum_{i=1}^{3} B_{2,q,b_i} + \sum_{i=1}^{2} B_{2,q,c_i}),$$

where

$$B_{2,q,a_1} = \sum_{l_1,l_2=1}^{p-q} \sum_{i,k,j1,j2}^{*} (X_{il_1}X_{il_2})(X_{kl_1+q}X_{kl_2+q})(X_{j_1l_1}X_{j_1l_1+q})(X_{j_2l_2}X_{j_2l_2+q}),$$

$$B_{2,q,a_2} = \sum_{l_1,l_2=1}^{p-q} \sum_{i,k,j1,j2}^{*} (X_{il_1}X_{il_2+q})(X_{kl_1+q}X_{kl_2})(X_{j_1l_1}X_{j_1l_1+q})(X_{j_2l_2}X_{j_2l_2+q}),$$

$$B_{2,q,b_1} = \sum_{l_1,l_2=1}^{p-q} \sum_{i,j,k}^{*}(X_{il_1}X_{il_2}X_{il_2+q})(X_{jl_1}X_{jl_1+q}X_{jl_2})X_{kl_1+q}X_{kl_2+q},$$

$$B_{2,q,b_2} = 2\sum_{l_1,l_2=1}^{p-q} \sum_{i,j,k}^{*}(X_{il_1}X_{il_2}X_{il_2+q})(X_{jl_1}X_{jl_1+q}X_{jl_2+q})X_{kl_1+q}X_{kl_2},$$

$$B_{2,q,b_3} = \sum_{l_1,l_2=1}^{p-q} \sum_{i,j,k}^{*}(X_{il_1+q}X_{il_2}X_{il_2+q})(X_{jl_1}X_{jl_1+q}X_{jl_2+q})X_{kl_1}X_{kl_2},$$

$$B_{2,q,c_1} = \sum_{l_1,l_2=1}^{p-q} \sum_{i,j,k}^{*}(X_{il_1}X_{il_2})(X_{kl_1+q}X_{kl_2+q})(X_{jl_1}X_{jl_1+q}X_{jl_2}X_{jl_2+q}) \quad \text{and}$$

$$B_{2,q,c_2} = \sum_{l_1,l_2=1}^{p-q} \sum_{i,j,k}^{*}(X_{il_1}X_{il_2+q})(X_{kl_1+q}X_{kl_2})(X_{jl_1}X_{jl_1+q}X_{jl_2}X_{jl_2+q}).$$

We need to show that the expectations of all the terms above are controlled by the order $n^4\mathrm{tr}^{\frac{1}{2}}(\Sigma^4)\mathrm{tr}(\Sigma^2)$. First, note that

$$\mathrm{E}(B_{2,q,a_1}) = P_n^4 \sum_{l_1,l_2=1}^{p-q} \sigma_{l_1l_2}\sigma_{l_1+q\ l_2+q}\sigma_{l_1l_1+q}\sigma_{l_2l_2+q}.$$

By the Cauchy-Schwarz inequality, it can be shown that

$$|\mathrm{E}(B_{2,q,a_1})| = P_n^4 O(\mathrm{tr}^{\frac{1}{2}}(\Sigma^4)\mathrm{tr}(\Sigma^2)).$$

Employing a similar derivation, we can show that the same result holds for all the other terms, which lead to the first part of Lemma 2. The second part can be proved following the same track. □

Lemma 3. *Under Assumptions 1 and 2,* $\mathrm{Var}(U_i) = o\{n^{-2}\mathrm{tr}^2(\Sigma^2)\}$ *for* $i = 2, 3$.

Proof. The proof is similar to Lemma 2. □

Lemma 4. *Under Assumptions 1 and 2,* $\mathrm{Var}\big(\sum_{q=k+1}^{p-1} B_{i,q}\big) = o\{n^{-2}\mathrm{tr}^2(\Sigma^2)\}$ *for* $i = 2, 3$.

Proof. Noting that $\sum_{q=k+1}^{p} B_{i,q} = U_i - \sum_{q=1}^{k} B_{i,q}$, the lemma follows by applying Lemma 2, Lemma 3, $k = o(p^{1/4})$ and Assumption 2. □

In the following, we provide the proof of Proposition 1 and 2.

Proof of Proposition 1. Rewrite W_{nk} as

$$W_{nk} = C_{nk} - 2\sum_{q=k+1}^{p} B_{2,q} + \sum_{q=k+1}^{p} B_{3,q}.$$

Since $\mathrm{E}(C_{nk}) = \sum_{|i-j|>k}\sigma_{ij}^2 = \mathrm{tr}[\{\Sigma - B_k(\Sigma)\}^2]$ and $\mathrm{E}(B_{i,q}) = 0$ for $i = 2, 3$ and any $q = 0, 1, \cdots, p-1$, the first statement is readily obtained. The second statement follows by applying Lemma 1, Lemma 4 and the fact that $\nu_{nk}^2 \geq 4n^{-2}\mathrm{tr}^2(\Sigma^2)$.

Proof of Proposition 2. It can be carried out following the same routes as those in Lemma 1 and 2. Specifically, it can be shown that $\mathrm{Var}(V_{nk}) = O\{a_{np}\mathrm{tr}^2(\Sigma^2)\}$. Hence, $\mathrm{Var}\{V_{nk}/\mathrm{tr}(\Sigma^2)\} = O(a_{np}) \to 0$. □

It is clear from the proof of Proposition 1 that $W_{nk} = C_{nk} + o_p(\nu_{nk})$. Therefore, in order to derive the asymptotical distribution of the statistic, we only need to consider the asymptotical normality of C_{nk}. Let $\mathscr{F}_0 = \{\emptyset, \Omega\}$, and $\mathscr{F}_t = \sigma\{X_1, \cdots, X_t\}$ for $t = 1, 2, \cdots, n$, be a sequence of σ-field generated by the data sequence. Let $\mathrm{E}_t(\cdot)$ denote the conditional expectation with respect to $\mathscr{F}_t$. Write $C_{nk} - \mathrm{E}(C_{nk}) = \sum_{t=1}^{n} D_{tk}$, where $D_{tk} = (\mathrm{E}_t - \mathrm{E}_{t-1})C_{nk}$. Then for every n, $D_{tk}, 1 \leq t \leq n$ is a martingale difference sequence with respect to the σ-fields $\{\mathscr{F}_t\}_{t=0}^{\infty}$.

Lemma 5. *Let $\sigma_{tk}^2 = \mathrm{E}_{t-1}(D_{tk}^2)$. Under Assumptions 1 and 2, as $n \to \infty$,*

$$\frac{\sum_{t=1}^n \sigma_{tk}^2}{\mathrm{Var}(C_{nk})} \xrightarrow{p} 1 \quad \text{and} \quad \frac{\sum_{t=1}^n \mathrm{E}(D_{tk}^4)}{\mathrm{Var}^2(C_{nk})} \to 0. \tag{A.1}$$

Proof. We first establish the first part of (A.1). Noting that $\mathrm{E}(\sum_{t=1}^n \sigma_{tk}^2) = \mathrm{Var}(C_{nk})$, we need only to show $\mathrm{Var}(\sum_{t=1}^n \sigma_{tk}^2) = o(\mathrm{Var}^2(C_{nk}))$. Note that

$$\begin{aligned} D_{tk} =& \frac{2}{n(n-1)}\Big[\sum_{|l_1-l_2|>k}(X_{tl_1}X_{tl_2}-\sigma_{l_1l_2})\Big\{\sum_{i=1}^{t-1}(X_{il_1}X_{il_2}-\sigma_{l_1l_2})\Big\}\Big] \\ &+ \frac{2}{n}\Big(\sum_{|l_1-l_2|>k} X_{tl_1}X_{tl_2}\sigma_{l_1l_2} - \sum_{|l_1-l_2|>k}\sigma_{l_1l_2}^2\Big). \end{aligned}$$

Denote $Q_{t-1}^{l_1l_2} = \sum_{i=1}^{t-1}(X_{il_1}X_{il_2} - \sigma_{l_1l_2})$. Let Q_{t-1} be the matrix with the (l_1, l_2)th entry being $Q_{t-1}^{l_1l_2}$ and $M_{t-1} = \Gamma' Q_{t-1}\Gamma$; then

$$\sum_{t=1}^n \sigma_{tk}^2 = \sum_{i=1}^3 R_{1i} + \Delta\sum_{i=1}^3 R_{2i} + \sum_{i=1}^4 R_{3i} + \Delta\sum_{i=1}^4 R_{4i} + n\gamma,$$

where γ is a constant and

$$\begin{aligned} R_{11} =& \frac{4}{n^2(n-1)^2}\sum_{t=1}^n \mathrm{tr}(M_{t-1}^2), \\ R_{12} =& -\frac{8}{n^2(n-1)^2}\sum_{t=1}^n\sum_{|l_1-l_2|\le k} Q_{t-1}^{l_1l_2}(\Sigma Q_{t-1}\Sigma)_{l_1l_2}, \\ R_{13} =& \frac{4}{n^2(n-1)^2}\sum_{t=1}^n\sum_{|l_1-l_2|\le k}\sum_{|l_3-l_4|\le k} Q_{t-1}^{l_1l_2}Q_{t-1}^{l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4}, \\ R_{21} =& \frac{4}{n^2(n-1)^2}\sum_{t=1}^n \mathrm{tr}(M_{t-1}\circ M_{t-1}), \\ R_{22} =& -\frac{8}{n^2(n-1)^2}\sum_{t=1}^n\sum_m\sum_{|l_1-l_2|\le k} Q_{t-1}^{l_1l_2}M_{t-1}^{mm}\Gamma_{l_1m}\Gamma_{l_2m}, \\ R_{23} =& \frac{4}{n^2(n-1)^2}\sum_{t=1}^n\sum_m\sum_{|l_1-l_2|\le k}\sum_{|l_3-l_4|\le k} Q_{t-1}^{l_1l_2}Q_{t-1}^{l_3l_4}\Gamma_{l_1m}\Gamma_{l_2m}\Gamma_{l_3m}\Gamma_{l_4m}, \\ R_{31} =& \frac{8}{n^2(n-1)}\sum_{t=1}^n \mathrm{tr}(\Sigma Q_{t-1}\Sigma^2), \end{aligned}$$

$$R_{32} = -\frac{8}{n^2(n-1)}\sum_{t=1}^{n}\sum_{|l_1-l_2|\le k} Q_{t-1}^{l_1l_2}(\Sigma^3)_{l_1l_2},$$

$$R_{33} = -\frac{8}{n^2(n-1)}\sum_{t=1}^{n}\sum_{|l_1-l_2|\le k} (\Sigma Q_{t-1}\Sigma)_{l_1l_2}\sigma_{l_1l_2},$$

$$R_{34} = \frac{8}{n^2(n-1)}\sum_{t=1}^{n}\sum_{|l_1-l_2|\le k}\sum_{|l_3-l_4|\le k} Q_{t-1}^{l_1l_2}\sigma_{l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4},$$

$$R_{41} = \frac{8}{n^2(n-1)}\sum_{t=1}^{n}\mathrm{tr}(M_{t-1}\circ A^2),$$

$$R_{42} = -\frac{8}{n^2(n-1)}\sum_{t=1}^{n}\sum_{m}\sum_{|l_1-l_2|\le k} Q_{t-1}^{l_1l_2}\Gamma_{l_1m}\Gamma_{l_2m}(A^2)_{mm},$$

$$R_{43} = -\frac{8}{n^2(n-1)}\sum_{t=1}^{n}\sum_{m}\sum_{|l_1-l_2|\le k} \sigma_{l_1l_2}\Gamma_{l_1m}\Gamma_{l_2m}M_{t-1}^{mm} \quad \text{and}$$

$$R_{44} = \frac{8}{n^2(n-1)}\sum_{t=1}^{n}\sum_{m}\sum_{|l_1-l_2|\le k}\sum_{|l_3-l_4|\le k} Q_{t-1}^{l_1l_2}\sigma_{l_3l_4}\Gamma_{l_1m}\Gamma_{l_2m}\Gamma_{l_3m}\Gamma_{l_4m}.$$

To prove $\mathrm{Var}(\sum_{t=1}^{n}\sigma_{tk}^2) = o(\mathrm{Var}^2(C_{nk}))$, we intend to prove the variance of each R_{ij} is of small order of $n^{-4}\mathrm{tr}^4(\Sigma^2)$.

For R_{12}, denote for any $1\le i,j\le n$,

$$Y_{ij}^{12} = \sum_{|l_1-l_2|\le k}(X_{il_1}X_{il_2}-\sigma_{l_1l_2})\{(\Sigma X_jX_j'\Sigma)_{l_1l_2}-(\Sigma^3)_{l_1l_2}\}.$$

Then $\sum_{|l_1-l_2|\le k} Q_{t-1}^{l_1l_2}(\Sigma Q_{t-1}\Sigma)_{l_1l_2} = \sum_{i=1}^{t-1} Y_{ii}^{12} + \sum_{i\ne j}^{t-1} Y_{ij}^{12}$. Note that $\mathrm{E}Y_{ij}^{12}=0$ for any $i\ne j$ and $\mathrm{E}(Y_{i_1j_1}^{12}Y_{i_2j_2}^{12}) = 0$ for any (i_1,i_2,j_1,j_2), except $\{i_1=i_2, j_1=j_2\}$ and $\{i_1=j_1, i_2=j_2\}$. Thus for any $t<l$,

$$\begin{aligned}&\mathrm{Cov}(\sum_{|l_1-l_2|\le k} Q_{t-1}^{l_1l_2}(\Sigma Q_{t-1}\Sigma)_{l_1l_2}, \sum_{|l_1-l_2|\le k} Q_{l-1}^{l_1l_2}(\Sigma Q_{l-1}\Sigma)_{l_1l_2})\\ =&(t-1)\mathrm{Var}(Y_{11}^{12}) + (t-1)(t-2)\mathrm{Var}(Y_{12}^{12}).\end{aligned}$$

We only need to verify that $\mathrm{Var}(y_{11}^{12})$ and $\mathrm{Var}(y_{12}^{12})$ are of small orders of $\mathrm{tr}^4(\Sigma^2)$. Note that

$$\begin{aligned}\mathrm{E}(Y_{11}^{12})^2 =&\mathrm{E}\sum_{|l_1-l_2|\le k}\sum_{|l_3-l_4|\le k}(X_{1l_1}X_{1l_2}-\sigma_{l_1l_2})(X_{1l_3}X_{1l_4}-\sigma_{l_3l_4})\\ &\times\{(\Sigma X_1X_1'\Sigma)_{l_1l_2}-(\Sigma^3)_{l_1l_2}\}\{(\Sigma X_1X_1'\Sigma)_{l_3l_4}-(\Sigma^3)_{l_3l_4}\}\end{aligned}$$

$$
\begin{aligned}
\leq & \gamma_{12} \sum_{|l_1-l_2|\leq k} \sum_{|l_3-l_4|\leq k} (\sigma_{l_1l_2}^2 + \sigma_{l_1l_1}\sigma_{l_2l_2})^{\frac{1}{2}} (\sigma_{l_3l_4}^2 + \sigma_{l_3l_3}\sigma_{l_4l_4})^{\frac{1}{2}} \\
& \times \{(\Sigma^3)_{l_1l_2}^2 + (\Sigma^3)_{l_1l_1}(\Sigma^3)_{l_2l_2}\}^{\frac{1}{2}} \{(\Sigma^3)_{l_3l_4}^2 + (\Sigma^3)_{l_3l_3}(\Sigma^3)_{l_4l_4}\}^{\frac{1}{2}} \\
\leq & \gamma_{12} \sum_{|l_1-l_2|\leq k} (\sigma_{l_1l_2}^2 + \sigma_{l_1l_1}\sigma_{l_2l_2}) \sum_{|l_1-l_2|\leq k} \{(\Sigma^3)_{l_1l_2}^2 + (\Sigma^3)_{l_1l_1}(\Sigma^3)_{l_2l_2}\} \\
\leq & \gamma_{12}(2k+1)^2 \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma^6),
\end{aligned}
$$

where γ_{12} is a constant. Since $\mathrm{tr}(\Sigma^6) \leq \mathrm{tr}^{\frac{3}{2}}(\Sigma^4)$,

$$
(2k+1)^2\mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma^6) = O\{k^2\mathrm{tr}(\Sigma^2)\mathrm{tr}^{\frac{3}{2}}(\Sigma^4)\} = O\{k^2p^{-\frac{3}{2}}\mathrm{tr}^4(\Sigma^2)\} = o\{\mathrm{tr}^4(\Sigma^2)\},
$$

which indicates that $\mathrm{Var}(Y_{11}^{12}) = o\{\mathrm{tr}^4(\Sigma^2)\}$. Similarly, we can also show that $\mathrm{Var}(Y_{12}^{12}) = o\{\mathrm{tr}^4(\Sigma^2)\}$. Thus

$$
\mathrm{Var}(R_{12}) = \frac{64}{n^4(n-1)^4}\mathrm{Var}\left\{\sum_{t=1}^{n}\sum_{|l_1-l_2|\leq k} Q_{t-1}^{l_1l_2}(\Sigma Q_{t-1}\Sigma)_{l_1l_2}\right\} = o\{n^{-4}\mathrm{tr}^4(\Sigma^2)\}.
$$

Following the same procedure, we can prove that for all the other R_{ij}, $\mathrm{Var}(R_{ij}) = o\{n^{-4}\mathrm{tr}^4(\Sigma^2)\}$. Since $\mathrm{Var}^2(C_{nk}) \geq n^{-4}\mathrm{tr}^4(\Sigma^2)$, we have $\mathrm{Var}(R_{ij}) = o\{\mathrm{Var}^2(C_{nk})\}$. Thus we have $\mathrm{Var}(\sum_{t=1}^n \sigma_{tk}^2) = o(\mathrm{Var}^2(C_{nk}))$, and hence the first part of (A.1).

For the second part of (A.1), by simple algebra, we can rewrite D_{tk} as $D_{tk} = S_{t1} - S_{t2} + S_{t3} - S_{t4}$, where

$$
\begin{aligned}
S_{t1} =& \frac{2}{n(n-1)}\left\{X_t'Q_{t-1}X_t - \mathrm{tr}(Q_{t-1}\Sigma)\right\}, \\
S_{t2} =& \frac{2}{n(n-1)}\left[X_t'B_k(Q_{t-1})X_t - \mathrm{tr}\{B_k(Q_{t-1})\Sigma\}\right], \\
S_{t3} =& \frac{2}{n}\left\{X_t'\Sigma X_t - \mathrm{tr}(\Sigma^2)\right\} \quad \text{and} \\
S_{t4} =& \frac{2}{n}\left[X_t'B_k(\Sigma)X_t - \mathrm{tr}\{B_k(\Sigma)\Sigma\}\right].
\end{aligned}
$$

Since $D_{tk}^4 \leq \tilde{\gamma}(S_{t1}^4 + S_{t2}^4 + S_{t3}^4 + S_{t4}^4)$, we have for a positive constant $\tilde{\gamma}$,

$$
\sum_{t=1}^{n}\mathrm{E}(D_{tk}^4) \leq \tilde{\gamma}\left\{\sum_{t=1}^{n}\mathrm{E}(S_{t1}^4) + \sum_{t=1}^{n}\mathrm{E}(S_{t2}^4) + \sum_{t=1}^{n}\mathrm{E}(S_{t3}^4) + \sum_{t=1}^{n}\mathrm{E}(S_{t4}^4)\right\}.
$$

In the following, we will prove the four terms on the right are of small orders of $\mathrm{Var}^2(C_{nk})$, respectively. To this end, note that

$$\mathrm{E}\left\{X_t' Q_{t-1} X_t - \mathrm{tr}(Q_{t-1}\Sigma)\right\}^4 \le \tilde{\gamma_1}\mathrm{E}\{\mathrm{tr}^2(M_{t-1}^2)\},$$

where $\tilde{\gamma_1}$ is a positive constant. Since $\mathrm{E}\{\mathrm{tr}(M_{t-1}^2)\} = (t-1)O\{\mathrm{tr}^2(\Sigma^2)\}$, and $\mathrm{Var}\{\mathrm{tr}(M_{t-1}^2)\} = t^2 O(\mathrm{tr}^2(\Sigma^2)\mathrm{tr}(\Sigma^4))$, then we have $\mathrm{E}\{\mathrm{tr}^2(M_{t-1}^2)\} = t^2 O\{\mathrm{tr}^4(\Sigma^2)\}$. Thus,

$$\begin{aligned}\sum_{t=1}^{n}\mathrm{E}(S_{t1}^4) =& \frac{16}{n^4(n-1)^4}\sum_{t=1}^{n}\mathrm{E}\left\{X_t' Q_{t-1} X_t - \mathrm{tr}(Q_{t-1}\Sigma)\right\}^4 \\ \le& \frac{16}{n^4(n-1)^4}\sum_{t=1}^{n} t^2 O\{\mathrm{tr}^4(\Sigma^2)\} = \frac{1}{n^5}O\{\mathrm{tr}^4(\Sigma^2)\} = o\{\mathrm{Var}^2(C_n)\}.\end{aligned}$$

Similarly, we can show that for $i = 2, 3$ and 4, $\sum_{t=1}^{n}\mathrm{E}(S_{ti}^4) = o\{\mathrm{Var}^2(C_n)\}$. Combining all the four parts together, we have $\sum_{t=1}^{n}\mathrm{E}(D_{k,t}^4) = o\{\mathrm{Var}^2(C)\}$, which leads to the second part of (A.1). $\square$

Denote $I_{nk} = \{W_{nk} - \mathrm{E}(W_{nk})\}/V_{nk}$ and $J_{nk} = \mathrm{E}(W_{nk})/V_{nk}$. Then $\tilde{T}_{nk} = I_{nk} + J_{nk}$. For k_0 diverging, but satisfying (4.1), we intend to prove $n^{\delta}(J_{nk} - J_{n\,k+1})$ diverging to ∞ uniformly on $k < k_0$ for any $\delta > 0$. And $n^{\delta} I_{nk}$ uniformly converges to 0 in probability for any $\delta \le 1/2$ and $k \le M$, where $M > k_0$ and $M = o(p^{1/4})$.

Lemma 6. *Under Assumptions 1, 2, and (4.1), if* $\liminf_n\{\inf_{k<k_0}(r_{k+1} - r_k)\} > 0$ *and* $\{\sigma_{ll}\}_{l=1}^{p}$ *are uniformly bounded away from* 0 *and* ∞*, for any* $\delta \le 0.5$*, as* $n \to \infty$*:*

(a). $P\big(n^{\delta}(J_{nk} - J_{n\,k+1}) > \xi, \quad$ *for any* $k < k_0\big) \to 1$ *for any* $\xi > 0$*;*

(b). $P\big(n^{\delta}|I_{nk}| \le \varepsilon, \quad$ *for any* $k \le k_0\big) \to 1$ *for any* $0 < \varepsilon < 1$*;*

(c). $P\big(n^{\delta}|I_{nk}| \le \varepsilon, \quad$ *for any* $k_0 < k \le M\big) \to 1$ *for any* $0 < \varepsilon < 1$*, where* $k_0 < M$ *and* $M = o(p^{1/4})$*.*

Proof. (a). If $\{\sigma_{ll}\}_{l=1}^{p}$ is bounded away from ∞, similarly to the proof of Lemma 1 and Lemma 2, it can be checked that $\mathrm{Var}(V_{nk}) = O(k^2\mathrm{tr}(\Sigma^2)/n)$. Therefore, by Chebyshev's inequality, for any $\varepsilon > 0$,

$$P\left(\left|\frac{V_{nk} - \mathrm{E}(V_{nk})}{\mathrm{tr}(\Sigma^2)}\right| > \varepsilon r_k^2\right) \le \frac{\mathrm{Var}(V_{nk})}{\varepsilon^2\mathrm{tr}^2(\Sigma^2)r_k^4} \le \frac{Ck^2}{\varepsilon^2 npr_k^3} \le \frac{Ck^2k_0^3}{\varepsilon^2 np},$$

where the last inequality comes from the fact that $r_k^{-1} \le 2k_0 + 1$. Hence,

$$P\left(\max_{0\le k\le k_0}\left|\frac{V_{nk} - \mathrm{E}(V_{nk})}{\mathrm{tr}(\Sigma^2)r_k^2}\right| \le \varepsilon\right) \ge 1 - \sum_{k=0}^{k_0}\frac{Ck^2k_0^3}{\varepsilon^2 np} \ge 1 - \frac{Ck_0^6}{\varepsilon^2 np},$$

which converge to 1 since k_0 satisfies (4.1). Consider $\varepsilon < 1/2$, and denote

$$\Omega = \big\{\omega : |V_{nk} - \mathrm{E}(V_{nk})| \le \varepsilon r_k^2 \mathrm{tr}(\Sigma^2), \quad \text{for any } k \le k_0\big\}.$$

By the above argument, $P(\Omega) \to 1$ as $n \to \infty$. For any $\omega \in \Omega$, we have

$$1 - \varepsilon r_k \le 1/(1+\varepsilon r_k) \le \mathrm{tr}[\{B_k(\Sigma)\}^2]/V_{nk} \le 1/(1-\varepsilon r_k) \le 1 + 2\varepsilon r_k$$

for any $k < k_0$. Hence, for any $\omega \in \Omega$,

$$\begin{aligned} n^{\delta}(J_{nk} - J_{n\,k+1}) &\ge n^{\delta}(r_{k+1} - r_k) + n^{\delta}(\varepsilon r_k + 2\varepsilon r_{k+1} - 3\varepsilon) \\ &\ge n^{\delta}(r_{k+1} - r_k) - 3n^{\delta}\varepsilon, \end{aligned}$$

which implies that $n^{\delta}(J_{nk} - J_{n\,k+1})$ diverge uniformly on $k < k_0$, by choosing ε small enough. Therefore, for any $\xi > 0$, by choosing ε small enough, there exists a $N > 0$ such that for any $n > N$,

$$P\big(n^{\delta}(J_{nk} - J_{n\,k+1}) > \xi \ \text{ for any } k < k_0\big) \ge P(\Omega).$$

The conclusion follows by noting that $P(\Omega) \to 1$ as $n \to \infty$. The other two parts of the conclusion can be obtained similarly. For simplicity in the presentation, we omit them here. □

Proof of Theorem 1. By Lemma 5, Lemma 1 and the martingale central limit theorem [Billingsley (1995)], it is readily shown that as $n \to \infty$,

$$\frac{C_{nk} - \mathrm{E}(C_{nk})}{\nu_{nk}} \xrightarrow{D} N(0,1).$$

Substituting C_{nk} for W_{nk}, Theorem 1 follows by noting $W_{nk} = C_{nk} + o_p(\nu_{nk})$. □

Proof of Theorem 2. Note that $\mathrm{Var}\{V_{nk}/\mathrm{tr}(\Sigma^2)\} \to 0$, $\mathrm{E}\{V_{nk}/\mathrm{tr}(\Sigma^2)\} = r_k$ and $\limsup_n r_k \le 1$. It can be shown that for any $\eta > 0$, $\lim_{n\to\infty} P(B_{n,\eta}) = 1$ where $B_{n,\eta} = \{V_{nk} < (1+\eta)\mathrm{tr}(\Sigma^2)\}$. This means that for any $\varepsilon > 0$, there exists a positive integer N, such that for all $n > N$, $P(B_{n,\eta}) > 1 - \varepsilon$. Then from (3.4),

$$\begin{aligned} \beta_{nk} &\ge P\left(\frac{W_{nk} - \mathrm{tr}(\Sigma^2) + \mathrm{tr}[\{B_k(\Sigma)\}^2]}{\nu_{nk}} \ge z_{\alpha}\frac{V_{nk}}{\mathrm{tr}(\Sigma^2)} - \delta_{nk}, B_{n,\eta}\right) \\ &\ge P\left(\frac{W_{nk} - \mathrm{tr}(\Sigma^2) + \mathrm{tr}[\{B_k(\Sigma)\}^2]}{\nu_{nk}} \ge z_{\alpha}(1+\eta) - \delta_{nk}, B_{n,\eta}\right) \\ &\ge P\left(\frac{W_{nk} - \mathrm{tr}(\Sigma^2) + \mathrm{tr}[\{B_k(\Sigma)\}^2]}{\nu_{nk}} \ge z_{\alpha}(1+\eta) - \delta_{nk}\right) - P(B_{n,\eta}^{c}). \end{aligned}$$

Therefore, from Theorem 1,

$$\begin{aligned}\liminf_{n\to\infty}\beta_{nk}\geq&\liminf_{n\to\infty}P\left\{\frac{W_{nk}-\operatorname{tr}(\Sigma^2)+\operatorname{tr}[\{B_k(\Sigma)\}^2]}{\nu_{nk}}\geq z_\alpha(1+\eta)-\delta_{nk}\right\}\\&-\limsup_{n\to\infty}P(B^c_{n,\eta})\\\geq&1-\Phi\{z_\alpha(1+\eta)-\liminf_{n\to\infty}\delta_{nk}\}-\varepsilon.\end{aligned}$$

The first part of the theorem follows by taking $\varepsilon\to 0$ and $\eta\to 0$.

(ii) The condition $a_{np}^{-1/2}(1-r_k)\to\infty$ implies that $\delta_{nk}\to\infty$ as $n\to\infty$. Hence, $\beta_{nk}\to 1$. $\square$

Proof of Theorem 3. First consider the case where k_0 is bounded. Consider M to be a fixed sufficiently large integer. Recall that $\tilde{T}_{nk}=I_{nk}+J_{nk}$. where

$$I_{nk}=\{W_{nk}-\mathrm{E}(W_{nk})\}/V_{nk}\quad\text{and}\quad J_{nk}=\mathrm{E}(W_{nk})/V_{nk}.$$

By (4.3), since $a_{np}^{\frac{1}{2}}=O(n^{-1})$, we have $n^\delta I_{nk}=O_p(n^\delta a_{np}^{\frac{1}{2}})\to 0$, for any $k\leq M$. Note that

$$n^\delta(r_k^{-1}-r_{k+1}^{-1})=n^\delta\frac{r_{k+1}-r_k}{r_{k+1}r_k}\geq n^\delta(r_{k+1}-r_k).$$

Thus, from (4.3), for $k<k_0$, the condition $\liminf_n(r_{k+1}-r_k)>0$ implies that $n^\delta(J_{nk}-J_{n\,k+1})\sim n^\delta\to\infty$ in probability, where $\delta\in(0,1)$. Therefore, $d^{(\delta)}_{n\,k}\to\infty$ for $k<k_0$ and $d^{(\delta)}_{n\,k}=o_p(1)$ for $k\geq k_0$. Hence, for any $\theta>0$, as $n\to\infty$,

$$P(|d^{(\delta)}_{n\,k}|>\theta)\to 1\quad\text{for } k<k_0\quad\text{and}\quad P(|d^{(\delta)}_{n\,k}|>\theta)\to 0\quad\text{for } k\geq k_0.$$

Therefore, for any $\theta>0$ and any $\varepsilon>0$, for each k, there exists a positive integer N_k such that for all $n\geq N_k$,

$$\begin{aligned}&P(|d^{(\delta)}_{n\,k}|<\theta)<\varepsilon/(M+1)\quad\text{for any } k<k_0\ \text{ and}\\&P(|d^{(\delta)}_{n\,k}|\geq\theta)<\varepsilon/(M+1)\quad\text{for any } k_0\leq k\leq M.\end{aligned}$$

Note that both k_0 and M are finite, we can set an N, which is larger than all N_k such that the above are satisfied. Define, for $k\leq M$, $B_{n\,k}:=\{|d^{(\delta)}_{n\,k}|<\theta\}$

and $B_n := \left(\bigcap_{i=0}^{k_0-1} B_{n\,i}^c\right) \bigcap \left(\bigcap_{i=k_0}^{M} B_{n\,i}\right)$ for $n > N$. Then, for any $\omega \in B_n$, $\hat{k}_{\delta,\theta}(\omega) = k_0$.

$$P(B_n^c) \le \sum_{i=0}^{k_0-1} P(B_{n\,i}) + \sum_{k_0}^{M} P(B_{n\,i}^c) \le \varepsilon.$$

Hence, for any $0 < \delta < 1$ and $\theta > 0$, $\hat{k}_{\delta,\theta} \xrightarrow{p} k_0$.

For the case of diverging k_0, consider $k_0 < M$ and $M = o(p^{1/4})$. For any $\theta > 0$ and $\delta \le 1/2$, let $\varepsilon < \theta/2$ and $\xi > 2\theta$. Denote

$$\begin{aligned}
U_1 &= \big\{\omega : n^\delta |I_{nk}| \le \varepsilon, \quad \text{for any } k \le k_0\big\},\\
U_2 &= \big\{\omega : n^\delta |I_{nk}| \le \varepsilon, \quad \text{for any } k_0 < k \le M\big\} \quad \text{and}\\
U_3 &= \big\{\omega : n^\delta (J_{nk} - J_{n\,k+1}) > \xi, \quad \text{for any } k < k_0\big\}.
\end{aligned}$$

Then for any $\omega \in \bigcap_{i=1}^3 U_i$, we have $n^\delta (J_{nk} - J_{n\,k+1}) > \xi > 2\theta$ for any $k < k_0$ and $n^\delta |I_{nk}| \le \varepsilon < \theta/2$ for any $k \le M$, which lead to $n^\delta |I_{nk} - I_{n\,k+1}| < \theta$ for any $k \le M$. Therefore,

$$\begin{aligned}
d_{n\,k}^{(\delta)} &= n^\delta (I_{nk} - I_{n\,k+1}) + n^\delta (J_{nk} - J_{n\,k+1}) > \theta \quad \text{for any } k < k_0 \text{ and}\\
|d_{n\,k}^{(\delta)}| &\le n^\delta |I_{nk} - I_{n\,k+1}| < \theta \quad \text{for any } k_0 \le k < M.
\end{aligned}$$

From (4.4), we have $\hat{k}_{\delta,\theta} - k_0 = 0$. It follows that $\bigcap_{i=1}^3 U_i \subset \{\omega : \hat{k}_{\delta,\theta} - k_0 = 0\}$. Since $P\big(\bigcap_{i=1}^3 U_i\big) \to 1$ as $n \to \infty$ by Lemma 6, we have $\hat{k}_{\delta,\theta} - k_0 \xrightarrow{p} 0$. $\square$

References.

[1] Adam, B.-L., Qu, Y, Davis, J. W, Ward, M. D., Clements, M. A., Cazares, L. H., Semmes, O. J., Schellhamm, P F., Yasui, Y, Feng, Z. and Wright, Jr, G. L. W (2003). Serum Protein Fingerprinting Coupled With a Pattern-matching Algorithm Distinguishes Prostate Cancer from Benign Prostate Hyperplasia and Healthy mean. *Cancer Research* **63** 3609-3614.

[2] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[3] Bai, Z. and Saranadasa, H. (1996). Effect of High Dimension: by an Example of a Two Sample Problem. *Statistica Sinica* **6** 311-329.

[4] Bai, Z. D. and Silverstein, J.W. (2005). *Spectral Analysis of Large Dimensional Random Matrices.* Scientific Press, Beijing.

[5] Bai, Z.D., Silverstein, J.W. and Yin, Y.Q. (1988). A Note on the Largest Eigenvalue of a Large-dimensional Sample Covariance Matrix. *Journal of Multivariate Analysis* **26** 166-168.

[6] Bai, Z.D. and Yin, Y.Q. (1993). Limit of the Smallest Eigenvalue of a Large Dimensional Sample Covariance Matrix. *The Annals of Probability* **21** 1276-1294.

[7] Bickel, P. J. and Levina, E. (2008a). Regularized Estimation of Large Covariance Matrices. *The Annals of Statistics* **36** 199-227.

[8] Bickel, P. and Levina, E. (2008b). Covariance Regularization by Thresholding. *The Annals of Statistics* **36** 2577-2604.

[9] Billingsley, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York.

[10] Cai, T. and Jiang T. (2011). Limiting Laws of Coherence of Random Matrices with Applications to Testing Covariance Structure and Construction of Compressed Sensing Matrices. *The Annals of Statistics* **39** 1496-1525.

[11] Cai, T. and Jiang T. (2011). Phase Transition in Limiting Distributions of Coherence of High-Dimensional Random Matrices. *technical report*.

[12] Cai, T., Zhang, C.H. and Zhou, H. (2010). Optimal Rates of Convergence for Covariance Matrix Estimation. *The Annals of Statistics* **38** 2118-2144.

[13] Chen, S. X., Zhang L.-X. and Zhong, P.-S. (2010). Tests for High Dimensional Covariance Matrices. *Journal of the American Statistical Association* **105** 810-819.

[14] Cleveland, W. and Devlin, S. J. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of American Statistical Association* **83** 596-610.

[15] El Karoui, N. (2011). On the largest eigenvalue of Wishart matrices with identity covariance when n, p and n/p tend to infinity. *manuscript*.

[16] Fan, J., Fan, Y. and Lv, J. (2008). High Dimensional Covariance Matrix Estimation Using a Factor Model. *Journal of Econometrics* **147** 186-197.

[17] Fan, J. and Gijbles, I. (1996). *Local Polynomial Smoothing*. Chapman and Hall, London.

[18] Huang, J., Liu, N., Pourahmadi, M., and Liu, L. (2006). Covariance Matrix Selection and Estimation via Penalised Normal Likelihood. *Biometrika* **93** 85-98.

[19] Jiang, T. (2004). The Asymptotic Distributions of the Largest Entries of Sample Correlation Matrices. *The Annals of Applied Probability* **14** 865-880.

[20] John, S. (1972). The Distribution of a Statistic Used for Testing Sphericity of Normal Distributions. *Biometrika* **59** 169-173.

[21] Johnstone, I. M. (2001). On the Distribution of the Largest Eigenvalue in Principal Components Analysis. *The Annals of Statistics* **29** 295-327.

[22] Ledoit, O. and Wolf, M. (2002). Some Hypothesis Tests for the Covariance Matrix When the Dimension is Large Compare to the Sample Size. *The Annals of Statistics* **30** 1081-1102.

[23] Levina, E., Rothman, A. and Zhu, J. (2008). Sparse Estimation of Large Covariance Matrices Via a Nested Lasso Penalty. *The Annals of Applied Statistics* **2** 245-263.

[24] Liu, W.-D., Lin, Z. and Shao, Q.-M. (2008). The Asymptotic Distribution and Berry-Essen Bound of a New Test for Independence in High Dimension with an Application to Stochastic Optimization. *The Annals of Applied Probability* **18** 2337-2366.

[25] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

[26] Nagao, H. (1973). On Some Test Criteria for Covariance Matrix. *The Annals of Statistics* **1** 700-709.

[27] Rothman, A. J., Levina, E. and Zhu, J. (2009). Generalized Thresholding of Large Covariance Matrices. *Journal of the American Statistical Association* **104** 177-186.

[28] Rothman, A. J., Levina, E. and Zhu, J. (2010). A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika* **97** 539-550.

[29] Schott, J. R. (2005). Testing for Complete Independence in High Dimensions. *Biometrika* **92** 951-956.

[30] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and Smoothness via the Fused Lasso. *Journal of Royal Statistical Society Ser. B* **67** 91-108.

[31] Wagaman, A. S. and Levina, E. (2009). Discovering Sparse Covariance Structures With the Isomap. *Journal of Computational and Graphical Statistics* **18** 551-572.

[32] Wu, W. B. and Pourahmadi, M. (2003). Nonparametric Estimation of Large Covariance Matrices of Longitudinal Data. *Biometrika* **93** 831-844.

Department of Statistics
Iowa State University
Ames, Iowa 50011-1210, USA
E-mail: yumouqiu@iastate.edu

Department of Business Statistics
and Econometrics
Guanghua School of Management and
Center for Statistical Science
Peking University
Beijing 100871, China
AND
Department of Statistics
Iowa State University
Ames, Iowa 50011-1210, USA
E-mail: songchen@iastate.edu