*The Annals of Statistics*
2012, Vol. 40, No. 3, 1285–1314
DOI: 10.1214/12-AOS1002


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

BY YUMOU QIU AND SONG XI CHEN[1]

*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 $\Sigma$, we propose a test for $\Sigma$ 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 $\Sigma$ 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 (1988), Johnstone (2001) and El Karoui (2011). These mean that the sample covariance $S_n$ is no longer consistent to $\Sigma$, 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 been a focus in high-dimensional covariance estimation. Wu and Pourahmadi (2003) considered

Received November 2011; revised April 2012.

[1]Supported by NSFC key Grant 11131002.

*MSC2010 subject classifications.* Primary 62H15; secondary 62G10, 62G20.

*Key words and phrases.* Banded covariance matrix, bandwidth estimation, high data dimension, large $p$, small $n$, nonparametric.

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 $\Sigma$ and banding the Cholesky factor matrix for estimating $\Sigma^{-1}$. They demonstrated that both estimators are consistent to $\Sigma$ and $\Sigma^{-1}$, 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 $\Sigma$ is banded. The rationale for developing such a test is to check a $\Sigma$ 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 $\Sigma$ is within the "bandable" class so that the banding and tapering can be applied. Hence, a direct testing on $\Sigma$ 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 $\Sigma$ 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 (2011) under high dimensionality. For normally distributed data, Jiang (2004) proposed testing for diagonal $\Sigma$ 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 $i$th and the $j$th 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 (2011) extended the test of Jiang (2004) for the bandedness of $\Sigma$, 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 $\Sigma$ 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 $\mu$ 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 $k \in \{0, \dots, p-1\}$ 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 $\Sigma$ with bandwidth $k$. Specifically, $B_0(\Sigma)$ is the diagonal version of $\Sigma$. We intend to test

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

for $k = o(p^{1/4})$. Hence, the bandwidth $k$ of $\Sigma$ 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, Zhang and Zhong (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, Zhang and Zhou (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 $\alpha$ is an index value for a "bandable" class of covariances

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

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

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 $\operatorname{tr}\{\Sigma - B_k(\Sigma)\}^2 = 0$. A strategy is to construct an unbiased estimator of $\operatorname{tr}\{\Sigma - B_k(\Sigma)\}^2$ and use it to develop the test statistic. Let $D_q := \sum_{l=1}^{p-q} \sigma_{ll+q}^2$ be the sum of squares of the $q$th sub-diagonal of $\Sigma$. Then, $\operatorname{tr}\{\Sigma - B_k(\Sigma)\}^2 = 2\sum_{q=k+1}^{p-1} D_q$. It can be checked that an unbiased estimator of $D_q$ is

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

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

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

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

(ii) Write $Z_1 = (z_{11}, \dots, z_{1m})^T$. Each $z_{1l}$ has uniformly bounded 8th moment, and there exist finite constants $\Delta$ and $\omega$ such that for $l = 1, \dots, m$, $\mathrm{E}(z_{1l}^4) = 3 + \Delta$, $\mathrm{E}(z_{1l}^3) = \omega$ and for any integers $\ell_\nu \ge 0$ 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 $i_1, i_2, \dots, i_q$ 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 $\Sigma$ 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 $p \gg n$, 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 $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_p$ be the eigenvalues of $\Sigma$. If the smallest and largest eigenvalues are bounded away from 0 and $\infty$ respectively, then

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

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

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

Thus, $\Sigma$ 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, Zhang and Zhou (2010), which allows the smallest eigenvalue to diminish to zero. It can be also checked that the following two covariances,

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

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

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

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 $H_{k,0}$,

$$\mathrm{E}(W_{nk}) = 0 \quad \text{and} \quad \nu_{nk} = 2\,\mathrm{tr}[\{B_k(\Sigma)\}^2]/n,$$

and $\nu_{nk}^2$ is the leading order variance of $W_{nk}$. It can be shown that $\mathrm{tr}\{\Sigma(\Sigma - B_k(\Sigma))\}^2 \leq 4(k+1)^2\mathrm{tr}(\Sigma^4)$. Since

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

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

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

for a constant $C_0 \geq 4$ 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 $\alpha$ that rejects $H_{k,0}$ if $T_{nk} \geq 2z_\alpha$ where $z_\alpha$ is the upper $\alpha$ 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 $\Sigma$ and its $k$-banded version, and $\nu_{nk}$ 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

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

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,

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

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 nondecreasing with respect to $k$. If $\Sigma$ is banded with bandwidth $k_0$, then

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

From the bounds for $\nu_{nk}$ in (3.2), it follows that

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

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 $\delta_{nk}$ 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) $\liminf_n \beta_{nk} \geq 1 - \Phi(z_\alpha - \liminf_n \delta_{nk})$;
(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 nontrivial 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 $\alpha$. 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 $\Sigma$. 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 $\Sigma$. 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 $\alpha$, we consider an estimator based on the difference between successive statistics, $d_{nk} = \tilde{T}_{nk} - \tilde{T}_{nk+1}$.

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

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

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

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

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

$$\mathrm{E}(b_{nk}) = 0 \quad \text{and} \quad \operatorname{Var}(b_{nk}) = O(a_{np} r_k^{-2}). \tag{4.2}$$

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

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

It can be checked that $a_{np}^{1/2}k_0 \to 0$ 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 $k \ge k_0$, since $r_k = 1$ for $k \ge k_0$.

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 $n^\delta$ on $\tilde{T}_{nk}$ for a small positive $\delta$ and let $d_{nk}^{(\delta)} = n^\delta(\tilde{T}_{nk} - \tilde{T}_{nk+1})$. The proposed bandwidth estimator is

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

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* $\infty$.

We would like to remark that the multiplier $n^\delta$ in $d_{nk}^{(\delta)}$'s formation leads to $\theta$ being "free ranged" as long as $\theta > 0$. If such multiplication is not administrated, namely by setting $\delta = 0$, the range of $\theta$ 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 $\Sigma$ has segments of zero sub-diagonals followed by nonzero 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 $\delta$ 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 $k \ge 5$. 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 $\delta$ and $\theta$. 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 $k \ge k_0$. Such a pattern, as displayed in Figure 1, indicates that $k_0$ is a change

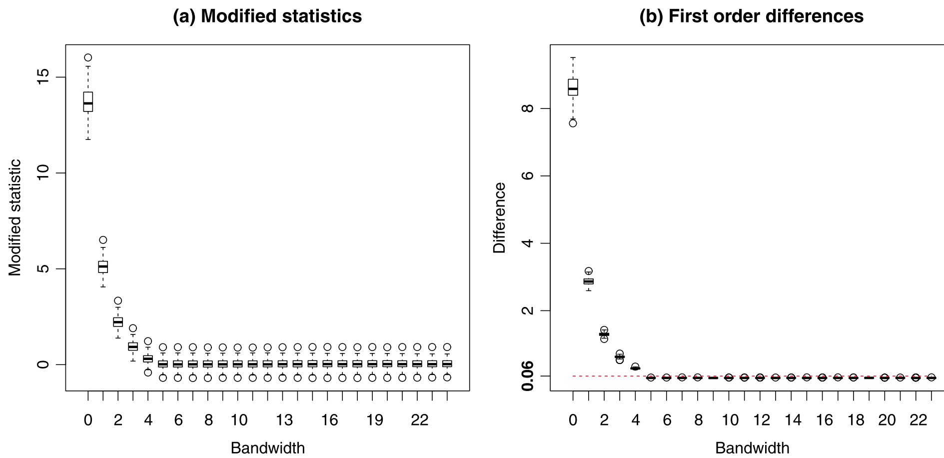


FIG. 1. *Box-plots of the modified statistics $n^{\delta}\tilde{T}_{nk}$ and their first order differences of the simulated data. The dashed line in the right panel is $\theta = 0.06$. The true bandwidth is 5.*

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 $j = 1, \dots, M$, 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 $j \in L_k = \{l : 0 \le l \le k\}$ 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 $j \in R_k = \{l : k+1 \le l \le M\}$ 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

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

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

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

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 $v$th split, for $v = 1, \dots, N$, 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 $A = (a_{ij})$, $\|A\|_{(1,1)} = \max_{1\le j\le p_2} \sum_{i=1}^{p_1} |a_{ij}|$. 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

$$\hat{k}_{\mathrm{BL}} = \operatorname*{arg\,min}_{0\le k\le p-1} \hat{R}(k).$$

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 $\Sigma$ 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 $X_i = (X_{i1}, \dots, X_{ip})'$ according to a model

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

where $k_0$ is the bandwidth of the covariance, $\gamma_0 = 1$ in all settings and the other coefficients $\gamma_l$ 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 $\gamma_l$ 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 $\gamma_l$ 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 (2011), 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 (2011) 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 $p$s, 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 $\gamma_l$ for $l > 0$ are:

Bandwidth 3: $\gamma_i = 1$, for $i = 1, 2, 3$;
Bandwidth 5: $\gamma_i = 0.4$ for $1 \le i \le 5$;
Bandwidth 10: $\gamma_i = 0.2$ for $1 \le i \le 5$ and $\gamma_i = 0.4$ for $6 \le i \le 10$;
Bandwidth 15: $\gamma_i = 0.2$ for $1 \le i \le 10$ and $\gamma_i = 0.4$ for $11 \le i \le 15$.

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

TABLE 1

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

| | **Normal** | | | | **Gamma** | | | |
|---|---|---|---|---|---|---|---|---|
| | ***p*** | | | | ***p*** | | | |
| ***n*** | **50** | **100** | **300** | **600** | **50** | **100** | **300** | **600** |
| | (a) $H_0 : \Sigma = B_0(\Sigma)$ | | | | | | | |
| 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 |
| | $H_0 : \Sigma = B_1(\Sigma)$ | | | | | | | |
| | $\gamma_1 = 1$ | | | | | | | |
| 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 |

candidate $k$s 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 $\Sigma$, 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, 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 2
*Empirical power of the proposed test at $\alpha = 5\%$ for the normal and gamma random vectors generated according to model (5.1)*

| | **Normal** | | | | **Gamma** | | | |
|---|---|---|---|---|---|---|---|---|
| | ***p*** | | | | ***p*** | | | |
| ***n*** | **50** | **100** | **300** | **600** | **50** | **100** | **300** | **600** |
| | (a) $H_0 : \Sigma = B_1(\Sigma)$ when $\Sigma = B_2(\Sigma)$ $\gamma_1 = \gamma_2 = 1$, $1 - r_1 = 1/14$ | | | | | | | |
| 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 |

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 48,538 $m/z$-sites and the full data set consisted of 157 healthy patients and 167 with cancer.

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)*

| | | | Bandwidth | | | |
|---|---|---|---|---|---|---|
| ***n*** | ***p*** | **Method** | **3** | **5** | **10** | **15** |
| 20 | 40 | 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) |
| | | 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) |
| | 100 | 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) |
| | | 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) |
| | 200 | 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) |
| | | 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) |
| 40 | 80 | 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) |
| | | 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) |
| | 200 | 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) |
| | | 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) |
| | 400 | Fixed | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| | | Change-point | 0.56 (0.499) | 0 (0) | 0 (0) | 0 (0) |
| | | 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) |

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

TABLE 3
(*Continued*)

| **n** | **p** | **Method** | **Bandwidth** 3 | 5 | 10 | 15 |
|---|---|---|---|---|---|---|
| 60 | 120 | 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) |
| | | 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) |
| | 300 | Fixed | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| | | Change-point | 0.58 (0.496) | 0 (0) | 0 (0) | 0 (0) |
| | | 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) |
| | 600 | Fixed | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| | | Change-point | 0.54 (0.501) | 0 (0) | 0 (0) | 0 (0) |
| | | 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) |

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

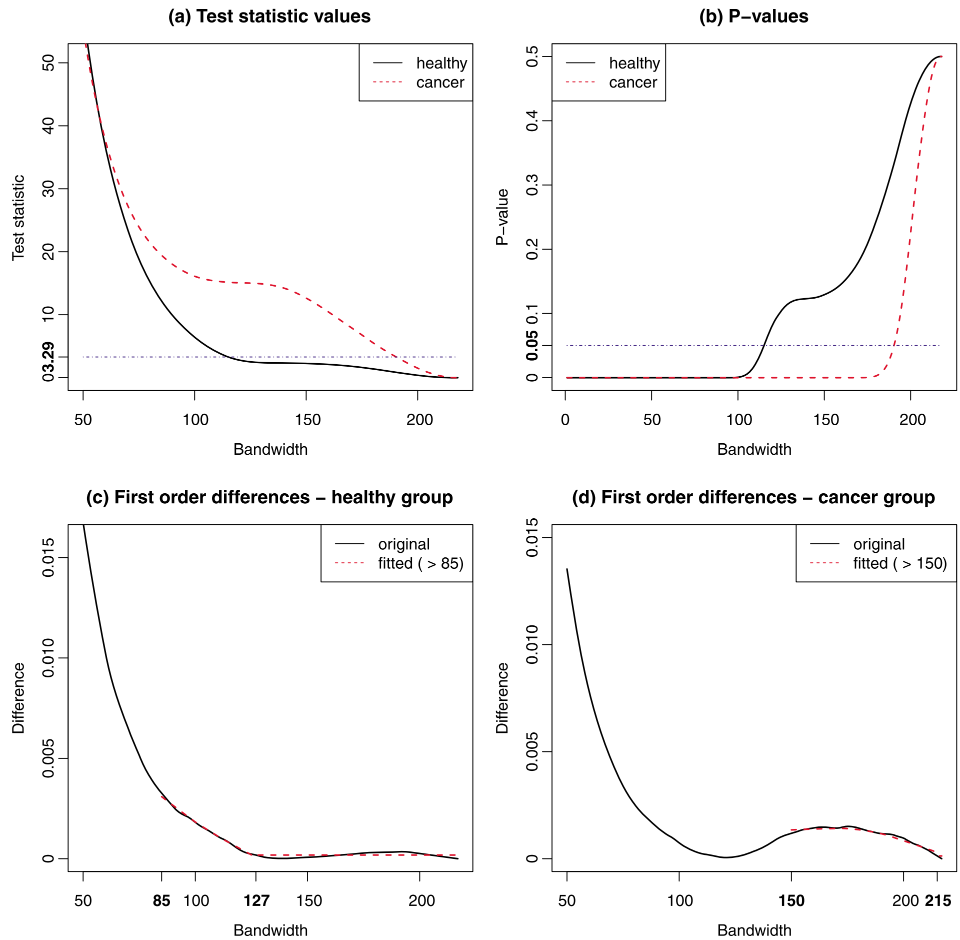


FIG. 2. *Test statistics, p-values and the first order differences $d_{nk}$ for the healthy and the cancer groups for bandwidths larger than* 50. *The p-values of the test for $H_{0k}$ for $k < 50$ are too small to be considered for bandwidth estimation.*

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 suggests 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.

## APPENDIX

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

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

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

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

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 $\sigma_{l_1 l_2} \sigma_{l_3 l_4}[3] = \sigma_{l_1 l_2} \sigma_{l_3 l_4} + \sigma_{l_1 l_3} \sigma_{l_2 l_4} + \sigma_{l_1 l_4} \sigma_{l_2 l_3}$. Then, $\mathrm{E}(C_{nk}^2) = (P_n^2)^{-2}(L_{n1} + L_{n2} + L_{n3})$, where

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

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

and

$$L_{n3} = 2P_n^2 \sum_{|l_1-l_2|>k} \sum_{|l_3-l_4|>k} (\Delta f_{l_1l_2l_3l_4} + \sigma_{l_1l_2}\sigma_{l_3l_4}[3])^2.$$

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

$$\sum_{|l_1-l_2|>k} \sum_{|l_3-l_4|>k} f_{l_1l_2l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4} = \operatorname{tr}(A^2 \circ A^2) - 2 \sum_{|l_1-l_2|\le k} \sum_{l_3,l_4} f_{l_1l_2l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4} + \sum_{|l_1-l_2|\le k} \sum_{|l_3-l_4|\le k} f_{l_1l_2l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4}.$$

By the Cauchy–Schwarz inequality,

$$\left| \sum_{|l_1-l_2|\le k} \sum_{l_3,l_4} f_{l_1l_2l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4} \right| \le \operatorname{tr}^{1/2}(T^2)\operatorname{tr}^{1/2}[(\Sigma\Gamma \circ \Sigma\Gamma)\{(\Sigma\Gamma)' \circ (\Sigma\Gamma)'\}]$$

and

$$\left| \sum_{|l_1-l_2|\le k} \sum_{|l_3-l_4|\le k} f_{l_1l_2l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4} \right| \le (2k+1)^2 \operatorname{tr}\{(\Gamma \circ \Gamma)(\Gamma' \circ \Gamma')(\Sigma \circ \Sigma)\},$$

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

$$\operatorname{tr}(T) \le \operatorname{tr}(\Sigma^2), \qquad \operatorname{tr}\{(\Gamma \circ \Gamma)(\Gamma' \circ \Gamma')(\Sigma \circ \Sigma)\} \le \operatorname{tr}(\Sigma^4)$$

and

$$\operatorname{tr}[(\Sigma\Gamma \circ \Sigma\Gamma)\{(\Sigma\Gamma)' \circ (\Sigma\Gamma)'\}] \le \operatorname{tr}(\Sigma^6).$$

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

$$\sum_{|l_1-l_2|\le k} \sum_{l_3,l_4} f_{l_1l_2l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4} = o\{\operatorname{tr}^2(\Sigma^2)\}$$

and

$$\sum_{|l_1-l_2|>k} \sum_{|l_3-l_4|>k} f_{l_1l_2l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4} = o\{\operatorname{tr}^2(\Sigma^2)\}.$$

Similarly, it can be shown that

$$\sum_{|l_1-l_2|\le k} (\Sigma^2)^2_{l_1l_2} = o\{\operatorname{tr}^2(\Sigma^2)\},$$

$$\sum_{|l_1-l_2|\le k} \sum_{|l_3-l_4|\le k} \sigma^2_{l_2l_4}\sigma^2_{l_1l_3} = o\{\operatorname{tr}^2(\Sigma^2)\},$$

$$\sum_{|l_1-l_2|\le k} (\Sigma^2)_{l_1l_1}(\Sigma^2)_{l_2l_2} = o\{\operatorname{tr}^2(\Sigma^2)\},$$

$$\sum_{|l_1-l_2|>k} \sum_{|l_3-l_4|>k} f^2_{l_1l_2l_3l_4} = o\{\operatorname{tr}^2(\Sigma^2)\}$$

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,

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

It can be checked that

$$\sum_{|l_1-l_2|>k}\sum_{|l_3-l_4|>k}\sigma_{l_1l_3}\sigma_{l_2l_4}\sigma_{l_1l_2}\sigma_{l_3l_4}=\mathrm{tr}\big\{\Sigma\big(\Sigma-B_k(\Sigma)\big)\big\}^2$$

and

$$\sum_{|l_1-l_2|>k}\sum_{|l_3-l_4|>k}f_{l_1l_2l_3l_4}\sigma_{l_1l_2}\sigma_{l_3l_4}=\mathrm{tr}\big(\Gamma'\big(\Sigma-B_k(\Sigma)\big)\Gamma\circ\Gamma'\big(\Sigma-B_k(\Sigma)\big)\Gamma\big).$$

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* $q=0,\dots,k$,

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

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

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

where

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

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

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

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

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

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

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 $n^4 \operatorname{tr}^{1/2}(\Sigma^4) \operatorname{tr}(\Sigma^2)$. First, note that $\mathrm{E}(B_{2,q,a_1}) = P_n^4 \sum_{l_1,l_2=1}^{p-q} \sigma_{l_1 l_2} \sigma_{l_1+q l_2+q} \times \sigma_{l_1 l_1+q} \sigma_{l_2 l_2+q}$. By the Cauchy–Schwarz inequality, it can be shown that

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

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, $\operatorname{Var}(U_i) = o\{n^{-2} \operatorname{tr}^2(\Sigma^2)\}$ *for* $i = 2, 3$.

PROOF. The proof is similar to Lemma 2. □

LEMMA 4. *Under Assumptions* 1 *and* 2, $\operatorname{Var}(\sum_{q=k+1}^{p-1} B_{i,q}) = o\{n^{-2} \operatorname{tr}^2(\Sigma^2)\}$ *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 Lemmas 2, 3, $k = o(p^{1/4})$ and Assumption 2. □

In the following, we provide the proof of Propositions 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 $\mathrm{E}(C_{nk}) = \sum_{|i-j|>k} \sigma_{ij}^2 = \operatorname{tr}[\{\Sigma - B_k(\Sigma)\}^2]$ and $\mathrm{E}(B_{i,q}) = 0$ for $i = 2, 3$ and any $q = 0, 1, \ldots, p-1$, the first statement is readily obtained. The second statement follows by applying Lemmas 1, 4 and the fact that $\nu_{nk}^2 \geq 4n^{-2} \operatorname{tr}^2(\Sigma^2)$. □

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

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 $\mathscr{F}_0 = \{\varnothing, \Omega\}$, and $\mathscr{F}_t = \sigma\{X_1, \dots, X_t\}$ for $t = 1, 2, \dots, n$, be a sequence of $\sigma$-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 $n$, $D_{tk}$, $1 \le t \le n$, is a martingale difference sequence with respect to the $\sigma$-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 $n \to \infty$,*

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

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

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

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 $\gamma$ is a constant and

$$R_{11} = \frac{4}{n^2(n-1)^2}\sum_{t=1}^n \mathrm{tr}(M_{t-1}^2),$$

$$R_{12} = -\frac{8}{n^2(n-1)^2}\sum_{t=1}^n\sum_{|l_1-l_2|\le k} Q_{t-1}^{l_1l_2}(\Sigma Q_{t-1}\Sigma)_{l_1l_2},$$

$$R_{13} = \frac{4}{n^2(n-1)^2}\sum_{t=1}^n\sum_{|l_1-l_2|\le k}\sum_{|l_3-l_4|\le k} Q_{t-1}^{l_1l_2}Q_{t-1}^{l_3l_4}\sigma_{l_1l_3}\sigma_{l_2l_4},$$

$$R_{21} = \frac{4}{n^2(n-1)^2} \sum_{t=1}^{n} \operatorname{tr}(M_{t-1} \circ M_{t-1}),$$

$$R_{22} = -\frac{8}{n^2(n-1)^2} \sum_{t=1}^{n} \sum_{m} \sum_{|l_1-l_2|\le k} Q_{t-1}^{l_1l_2} M_{t-1}^{mm} \Gamma_{l_1m}\Gamma_{l_2m},$$

$$R_{23} = \frac{4}{n^2(n-1)^2} \sum_{t=1}^{n} \sum_{m} \sum_{|l_1-l_2|\le k} \sum_{|l_3-l_4|\le k} Q_{t-1}^{l_1l_2} Q_{t-1}^{l_3l_4} \Gamma_{l_1m}\Gamma_{l_2m}\Gamma_{l_3m}\Gamma_{l_4m},$$

$$R_{31} = \frac{8}{n^2(n-1)} \sum_{t=1}^{n} \operatorname{tr}(\Sigma Q_{t-1} \Sigma^2),$$

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

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

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

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

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

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

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

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

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

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 $t < l$,

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

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

$$\begin{aligned}
\mathrm{E}(Y_{11}^{12})^2 &= \mathrm{E}\sum_{|l_1-l_2|\le k}\sum_{|l_3-l_4|\le k}(X_{1l_1}X_{1l_2} - \sigma_{l_1 l_2})(X_{1l_3}X_{1l_4} - \sigma_{l_3 l_4})\\
&\qquad \times \{(\Sigma X_1 X_1'\Sigma)_{l_1 l_2} - (\Sigma^3)_{l_1 l_2}\}\\
&\qquad \times \{(\Sigma X_1 X_1'\Sigma)_{l_3 l_4} - (\Sigma^3)_{l_3 l_4}\}\\
&\le \gamma_{12}\sum_{|l_1-l_2|\le k}\sum_{|l_3-l_4|\le k}(\sigma_{l_1 l_2}^2 + \sigma_{l_1 l_1}\sigma_{l_2 l_2})^{1/2}(\sigma_{l_3 l_4}^2 + \sigma_{l_3 l_3}\sigma_{l_4 l_4})^{1/2}\\
&\qquad \times \{(\Sigma^3)_{l_1 l_2}^2 + (\Sigma^3)_{l_1 l_1}(\Sigma^3)_{l_2 l_2}\}^{1/2}\\
&\qquad \times \{(\Sigma^3)_{l_3 l_4}^2 + (\Sigma^3)_{l_3 l_3}(\Sigma^3)_{l_4 l_4}\}^{1/2}\\
&\le \gamma_{12}\sum_{|l_1-l_2|\le k}(\sigma_{l_1 l_2}^2 + \sigma_{l_1 l_1}\sigma_{l_2 l_2})\sum_{|l_1-l_2|\le k}\{(\Sigma^3)_{l_1 l_2}^2 + (\Sigma^3)_{l_1 l_1}(\Sigma^3)_{l_2 l_2}\}\\
&\le \gamma_{12}(2k+1)^2\,\mathrm{tr}(\Sigma^2)\,\mathrm{tr}(\Sigma^6),
\end{aligned}$$

where $\gamma_{12}$ is a constant. Since $\mathrm{tr}(\Sigma^6) \le \mathrm{tr}^{3/2}(\Sigma^4)$,

$$\begin{aligned}
(2k+1)^2\,\mathrm{tr}(\Sigma^2)\,\mathrm{tr}(\Sigma^6) &= O\{k^2\,\mathrm{tr}(\Sigma^2)\,\mathrm{tr}^{3/2}(\Sigma^4)\}\\
&= O\{k^2 p^{-3/2}\,\mathrm{tr}^4(\Sigma^2)\}\\
&= o\{\mathrm{tr}^4(\Sigma^2)\},
\end{aligned}$$

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

$$\begin{aligned}
\mathrm{Var}(R_{12}) &= \frac{64}{n^4(n-1)^4}\mathrm{Var}\Bigg\{\sum_{t=1}^{n}\sum_{|l_1-l_2|\le k} Q_{t-1}^{l_1 l_2}(\Sigma Q_{t-1}\Sigma)_{l_1 l_2}\Bigg\}\\
&= o\{n^{-4}\,\mathrm{tr}^4(\Sigma^2)\}.
\end{aligned}$$

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 $\mathrm{Var}^2(C_{nk}) \ge n^{-4}\,\mathrm{tr}^4(\Sigma^2)$, 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

$$S_{t1} = \frac{2}{n(n-1)}\{X_t' Q_{t-1} X_t - \mathrm{tr}(Q_{t-1}\Sigma)\},$$

$$S_{t2} = \frac{2}{n(n-1)}[X_t' B_k(Q_{t-1}) X_t - \mathrm{tr}\{B_k(Q_{t-1})\Sigma\}],$$

$$S_{t3} = \frac{2}{n}\{X_t'\Sigma X_t - \mathrm{tr}(\Sigma^2)\}$$

and

$$S_{t4} = \frac{2}{n}[X_t' B_k(\Sigma) X_t - \mathrm{tr}\{B_k(\Sigma)\Sigma\}].$$

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

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

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

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

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 $\mathrm{E}\{\mathrm{tr}^2(M_{t-1}^2)\} = t^2 \times O\{\mathrm{tr}^4(\Sigma^2)\}$. Thus,

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

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). □

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 $n^\delta(J_{nk} - J_{nk+1})$ diverging to $\infty$ 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* $\infty$, *for any* $\delta \le 0.5$, *as* $n \to \infty$:

(a) $P(n^{\delta}(J_{nk} - J_{nk+1}) > \xi, \textit{for any } k < k_0) \to 1$ *for any* $\xi > 0$;
(b) $P(n^{\delta}|I_{nk}| \le \varepsilon, \textit{for any } k \le k_0) \to 1$ *for any* $0 < \varepsilon < 1$;
(c) $P(n^{\delta}|I_{nk}| \le \varepsilon, \textit{for any } k_0 < k \le M) \to 1$ *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 $\infty$, similarly to the proof of Lemmas 1 and 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

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

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 $\omega \in \Omega$,

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

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

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

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 Lemmas 1, 5 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 $\operatorname{Var}\{V_{nk}/\operatorname{tr}(\Sigma^2)\} \to 0$, $\mathrm{E}\{V_{nk}/\operatorname{tr}(\Sigma^2)\} = r_k$ 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 $B_{n,\eta} = \{V_{nk} < (1+\eta)\operatorname{tr}(\Sigma^2)\}$. 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),

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

Therefore, from Theorem 1,

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

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

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

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

$$P(|d_{nk}^{(\delta)}| > \theta) \to 1 \qquad \text{for } k < k_0 \quad \text{and} \quad P(|d_{nk}^{(\delta)}| > \theta) \to 0 \qquad \text{for } k \ge k_0.$$

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

$$P(|d_{nk}^{(\delta)}| < \theta) < \varepsilon/(M+1) \qquad \text{for any } k < k_0$$

and

$$P\big(|d_{nk}^{(\delta)}| \ge \theta\big) < \varepsilon/(M+1) \qquad \text{for any } k_0 \le k \le M.$$

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 $k \le M$, $B_{nk} := \{|d_{nk}^{(\delta)}| < \theta\}$ and $B_n := (\bigcap_{i=0}^{k_0-1} B_{ni}^c) \cup (\bigcap_{i=k_0}^{M} B_{ni})$ for $n > N$. Then, for any $\omega \in B_n$, $\hat{k}_{\delta,\theta}(\omega) = k_0$.

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

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

$$U_1 = \{\omega : n^{\delta}|I_{nk}| \le \varepsilon, \text{ for any } k \le k_0\},$$
$$U_2 = \{\omega : n^{\delta}|I_{nk}| \le \varepsilon, \text{ for any } k_0 < k \le M\}$$

and

$$U_3 = \{\omega : n^{\delta}(J_{nk} - J_{nk+1}) > \xi, \text{ for any } k < k_0\}.$$

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

$$d_{nk}^{(\delta)} = n^{\delta}(I_{nk} - I_{nk+1}) + n^{\delta}(J_{nk} - J_{nk+1}) > \theta \qquad \text{for any } k < k_0$$

and

$$\big|d_{nk}^{(\delta)}\big| \le n^{\delta}|I_{nk} - I_{nk+1}| < \theta \qquad \text{for any } k_0 \le k < M.$$

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 $P(\bigcap_{i=1}^3 U_i) \to 1$ as $n \to \infty$ by Lemma 6, we have $\hat{k}_{\delta,\theta} - k_0 \xrightarrow{p} 0$. $\square$

**Acknowledgments.** We thank an Associate Editor and two referees for constructive comments and suggestions which have lead to improvements in the presentation of the paper. We also thank Professor Ji Zhu for sharing the data.

## REFERENCES

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, G. L. W. JR. (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.

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ. MR1990662

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

BAI, Z. D. and SILVERSTEIN, J. W. (2005). *Spectral Analysis of Large Dimensional Random Matrices*. Scientific Press, Beijing.

BAI, Z. D., SILVERSTEIN, J. W. and YIN, Y. Q. (1988). A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.* **26** 166–168. MR0963829

BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

BICKEL, P. J. and LEVINA, E. (2008a). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

BICKEL, P. J. and LEVINA, E. (2008b). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York. MR1324786

CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210

CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885

CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863

CLEVELAND, W. and DEVLIN, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.* **83** 596–610.

EL KAROUI, N. (2011). On the largest eigenvalue of Wishart matrices with identity covariance when $n$, $p$ and $n/p$ tend to infinity. Unpublished manuscript.

FAN, J., FAN, Y. and LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics* **147** 186–197. MR2472991

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Smoothing*. Chapman & Hall, London.

HUANG, J. Z., LIU, N., POURAHMADI, M. and LIU, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* **93** 85–98. MR2277742

JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906

JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59** 169–173. MR0312619

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169

LEVINA, E., ROTHMAN, A. and ZHU, J. (2008). Sparse estimation of large covariance matrices via a nested Lasso penalty. *Ann. Appl. Stat.* **2** 245–263. MR2415602

LIU, W.-D., LIN, Z. and SHAO, Q.-M. (2008). The asymptotic distribution and Berry–Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18** 2337–2366. MR2474539

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. MR0652932

NAGAO, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. MR0339405

ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2009). Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.* **104** 177–186. MR2504372

ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2010). A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika* **97** 539–550. MR2672482

SCHOTT, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956. MR2234197

TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 91–108. MR2136641

WAGAMAN, A. S. and LEVINA, E. (2009). Discovering sparse covariance structures with the isomap. *J. Comput. Graph. Statist.* **18** 551–572. MR2751641

WU, W. B. and POURAHMADI, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90** 831–844. MR2024760

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