Available online at www.sciencedirect.com
SciVerse ScienceDirect
Stochastic Processes and their Applications 123 (2013) 2899–2920
stochastic processes and their applications
www.elsevier.com/locate/spa

Asymptotic theory for maximum deviations of sample covariance matrix estimates

Han Xiao[a,*], Wei Biao Wu[b]

[a] Department of Statistics and Biostatistics, Rutgers University, 501 Hill Center, 110 Frelinghuysen Road, Piscataway, NJ 08854, United States
[b] Department of Statistics, The University of Chicago, 5734 S. University Ave, Chicago, IL 60637, United States

Available online 28 March 2013

Abstract

We consider asymptotic distributions of maximum deviations of sample covariance matrices, a fundamental problem in high-dimensional inference of covariances. Under mild dependence conditions on the entries of the data matrices, we establish the Gumbel convergence of the maximum deviations. Our result substantially generalizes earlier ones where the entries are assumed to be independent and identically distributed, and it provides a theoretical foundation for high-dimensional simultaneous inference of covariances.

MSC: primary 62H15; 62H10; secondary 62E20

Keywords: Covariance matrix; High dimensional analysis; Maximal deviation; Tapering; Test for bandedness; Test for covariance structure; Test for stationarity

1. Introduction

Let $\boldsymbol{X}_n = \left(X_{ij}\right)_{1\le i\le n, 1\le j\le m}$ be a data matrix whose n rows are independent and identically distributed (i.i.d.) as some population distribution with mean vector $\boldsymbol{\mu}_n$ and covariance matrix Σ_n. High dimensional data increasingly occur in modern statistical applications in biology, finance and wireless communication, where the dimension m may be comparable to the number of

* Corresponding author.
E-mail addresses: hxiao@stat.rutgers.edu (H. Xiao), wbwu@galton.uchicago.edu (W.B. Wu).

http://dx.doi.org/10.1016/j.spa.2013.03.012

observations n, or even much larger than n. Therefore, it is necessary to study the asymptotic behavior of statistics of $\boldsymbol{X}_n$ under the setting that $m = m_n$ grows to infinity as n goes to infinity.

In many empirical examples, it is often assumed that $\Sigma_n = I_m$, where I_m is the $m \times m$ identity matrix, so it is important to perform the test

$$H_0 : \Sigma_n = I_m \tag{1}$$

before carrying out further estimation or inference procedures. Due to high dimensionality, conventional tests often do not work well or cannot be implemented. For example, when $m > n$, the likelihood ratio test (LRT) cannot be used because the sample covariance matrix is singular; and even when $m < n$, the LRT is drifted to infinity and leads to many false rejections if m is also large [1]. Ledoit and Wolf [16] found that the empirical distance test [21] is not consistent when both m and n are large. The problem has been studied by several authors under the "large n, large m" paradigm. Bai et al. [1] and Ledoit and Wolf [16] proposed corrections to the LRT and the empirical distance test respectively. Assuming that the population distribution is Gaussian with $\boldsymbol{\mu}_n = 0$, [14] used the largest eigenvalue of the sample covariance matrix $\boldsymbol{X}_n^\top \boldsymbol{X}_n$ as the test statistic, and proved that its limiting distribution follows the Tracy–Widom law [27]. Here we use the superscript $^\top$ to denote the transpose of a matrix or a vector. His work was extended to the non-Gaussian case by Soshnikov [24] and Péché [22], where they assumed the entries of $\boldsymbol{X}_n$ are i.i.d. with sub-Gaussian tails.

Let $x_1, x_2, \ldots, x_m$ be the m columns of $\boldsymbol{X}_n$. In practice, the entries of the mean vector $\boldsymbol{\mu}_n$ are often unknown, and are estimated by $\bar{x}_i = (1/n) \sum_{k=1}^n X_{ki}$. Write $x_i - \bar{x}_i$ for the vector $x_i - \bar{x}_i \mathbf{1}_n$, where $\mathbf{1}_n$ is the n-dimensional vector with all entries being one. Let $\sigma_{ij} = \mathrm{Cov}(X_{1i}, X_{1j})$, $1 \le i, j \le m$, be the covariance function, namely, the (i, j)th entry of Σ_n. The sample covariance between columns x_i and x_j is defined as

$$\hat{\sigma}_{ij} = \frac{1}{n}(x_i - \bar{x}_i)^\top (x_j - \bar{x}_j).$$

In high-dimensional covariance inference, a fundamental problem is to establish an asymptotic distributional theory for the maximum deviation

$$M_n = \max_{1 \le i < j \le m} |\hat{\sigma}_{ij} - \sigma_{ij}|.$$

With such a distributional theory, one can perform statistical inference for structures of covariance matrices. For example, one can use M_n to test the null hypothesis $H_0 : \Sigma_n = \Sigma^{(0)}$, where $\Sigma^{(0)}$ is a pre-specified matrix. Here the null hypothesis can be that the population distribution is a stationary process so that Σ_n is Toeplitz, or that Σ_n has a banded structure.

It is very challenging to derive an asymptotic theory for M_n if we allow dependence among $X_{11}, \ldots, X_{1m}$. Many of the earlier results assume that the entries of the data matrix $\boldsymbol{X}_n$ are i.i.d.. In this case $\sigma_{ij} = 0$ if $i \neq j$. The quantity

$$L_n = \max_{1 \le i < j \le m} |\hat{\sigma}_{ij}|$$

is referred to as the *mutual coherence* of the matrix $\mathbf{X}_n$, and is related to compressed sensing (see for example [9]). Jiang [13] derived the asymptotic distribution of L_n.

Theorem 1 (*[13]*)**.** *Suppose $X_{i,j}, i, j = 1, 2, \ldots$ are independent and identically distributed as ξ which has variance one. Suppose $\mathbb{E}|\xi|^{30+\epsilon} < \infty$ for some $\epsilon > 0$. If $n/m \to c \in (0, \infty)$, then* *for any* $y \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left(nL_n^2 - 4\log m + \log(\log m) + \log(8\pi) \le y\right) = \exp\left(-e^{-y/2}\right).$$

Jiang's work has attracted considerable attention, and been followed by Li et al. [17], Liu et al. [19], Zhou [30] and Li and Rosalsky [18]. Under the same setup that $\boldsymbol{X}_n$ consists of i.i.d. entries, these works focus on three directions (i) reduce the moment condition; (ii) allow a wider range of m; and (iii) show that some moment condition is necessary. In a recent article, [5] extended those results in two ways: (i) the dimension m could grow exponentially as the sample size n provided exponential moment conditions; and (ii) they showed that the test statistic $\max_{|i-j|>s_n} |\hat{\sigma}_{ij}|$ also converges to the Gumbel distribution if each row of $\boldsymbol{X}_n$ is Gaussian and is s_n-dependent. The latter generalization is important since it is one of the very few results that allow dependent entries.

In this paper we shall show that a self-normalized version of M_n converges to the Gumbel distribution under mild dependence conditions on the vector $(X_{11}, \ldots, X_{1m})$. Thus our result provides a theoretical foundation for high-dimensional simultaneous inference of covariances.

Besides testing covariance structure and simultaneous inference, the limiting behavior of M_n is also useful in several other applications. Liu et al. [19] and Tony Cai et al. [26] discussed the connection with the compressed sensing matrices. Kramer et al. [15] proposed to use the maximum cross correlation between a pair of time series to identify the edge between the corresponding nodes for electrocorticogram data. They employed the false discovery rate procedure to control for multiple testing, whilst the family-wise error rate is related to a quantity similar to M_n. Fan et al. [10] showed that the distance between theoretical and empirical risks of minimum variance portfolios is controlled by M_n, and thus provided a mathematical understanding of the finding of [12]. Cai et al. [7] studied a related test for the equality of two high dimensional covariance matrices.

The rest of this article is organized as follows. We present the main result in Section 2. In Section 3, we use two examples on linear processes and nonlinear processes to demonstrate that the technical conditions are easily satisfied. We discuss three tests for the covariance structure using our main result in Section 4. The proof is given in Section 5, and some auxiliary results are collected in Section 6. There is a supplementary file, which contains the technical proofs of several lemmas.

2. Main result

We consider a general situation where population distribution can depend on n. Recall that the dimension $m = m_n$ depends on n, but we will suppress the subscript and use m for ease of notation. Let $\mathbf{X}_n = (X_{n,k,i})_{1\le k\le n, 1\le i\le m}$ be a data matrix whose n rows are i.i.d. m-dimensional random vectors with mean $\boldsymbol{\mu}_n = (\mu_{n,i})_{1\le i\le m}$ and covariance matrix $\Sigma_n = (\sigma_{n,i,j})_{1\le i,j\le m}$. Let $x_1, x_2, \ldots, x_m$ be the m columns of $\boldsymbol{X}_n$. Let $\bar{x}_i = (1/n)\sum_{k=1}^n X_{n,k,i}$, and write $x_i - \bar{x}_i$ for the vector $x_i - \bar{x}_i \mathbf{1}_n$. The sample covariance between x_i and x_j is defined as

$$\hat{\sigma}_{n,i,j} = \frac{1}{n}(x_i - \bar{x}_i)^\top (x_j - \bar{x}_j).$$

It is unnatural to study the maximum of a collection of random variables which are on different scales, so we consider the normalized version $|\hat{\sigma}_{n,i,j} - \sigma_{n,i,j}|/\sqrt{\tau_{n,i,j}}$, where

$$\tau_{n,i,j} = \mathrm{Var}\left[(X_{n,1,i} - \mu_{n,i})(X_{n,1,j} - \mu_{n,j})\right].$$

In practice, $\tau_{n,i,j}$ are usually unknown, and can be estimated by

$$\hat{\tau}_{n,i,j} = \frac{1}{n}\left|(x_i - \bar{x}_i) \circ (x_j - \bar{x}_j) - \hat{\sigma}_{n,i,j} \cdot \mathbf{1}_n\right|^2$$

where $\circ$ denotes the Hadamard product defined as $A \circ B := (a_{ij}b_{ij})$ for two matrices $A = (a_{ij})$ and $B = (b_{ij})$ with the same dimensions. We thus consider

$$M_n = \max_{1\le i<j\le m} \frac{|\hat{\sigma}_{n,i,j} - \sigma_{n,i,j}|}{\sqrt{\hat{\tau}_{n,i,j}}}. \tag{2}$$

Due to the normalization procedure, we can assume without loss of generality that $\sigma_{n,i,i} = 1$ and $\mu_{n,i} = 0$ for each $1 \le i \le m$.

Define the index set $\mathcal{I}_n = \{(i, j) : 1 \le i < j \le m\}$, and for $\alpha = (i, j) \in \mathcal{I}_n$, let $X_{n,\alpha} := X_{n,1,i}X_{n,1,j}$. Define

$$\mathcal{K}_n(t, p) = \sup_{1\le i\le m} \mathbb{E}\exp\left(t|X_{n,1,i}|^p\right),$$
$$\mathcal{M}_n(p) = \sup_{1\le i\le m} \mathbb{E}(|X_{n,1,i}|^p),$$
$$\tau_n = \inf_{1\le i<j\le m} \tau_{n,i,j},$$
$$\gamma_n = \sup_{\alpha,\beta\in\mathcal{I}_n \text{ and } \alpha\ne\beta} \left|\mathrm{Cor}(X_{n,\alpha}, X_{n,\beta})\right|,$$
$$\gamma_n(b) = \sup_{\alpha\in\mathcal{I}_n} \sup_{\mathcal{A}\subset\mathcal{I}_n, |\mathcal{A}|=b} \inf_{\beta\in\mathcal{A}} \left|\mathrm{Cor}(X_{n,\alpha}, X_{n,\beta})\right|.$$

We need the following technical conditions.

(A1) $\liminf_{n\to\infty} \tau_n > 0$.
(A2) $\limsup_{n\to\infty} \gamma_n < 1$.
(A3) $\gamma_n(b_n)\log b_n = o(1)$ for any sequence (b_n) such that $b_n \to \infty$.
(A3′) $\gamma_n(b_n) = o(1)$ for any sequence (b_n) such that $b_n \to \infty$, and for some $\epsilon > 0$,

$$\sum_{\alpha,\beta\in\mathcal{I}_n} \left[\mathrm{Cov}(X_{n,\alpha}, X_{n,\beta})\right]^2 = O(m^{4-\epsilon}).$$

(A4) For some constants $t > 0$ and $0 < p \le 2$, $\limsup_{n\to\infty} \mathcal{K}_n(t, p) < \infty$, and

$$\log m = \begin{cases} o\left(n^{p/(4+p)}\right) & \text{when } 0 < p < 2 \\ o\left(n^{1/3}(\log n)^{-2/3}\right) & \text{when } p = 2. \end{cases}$$

(A4′) $\log m = o\left(n^{p/(4+3p)}\right)$ and $\limsup_{n\to\infty} \mathcal{K}_n(t, p) < \infty$ for some constants $t > 0$ and $p > 0$.
(A4″) $m = O(n^q)$ and $\limsup_{n\to\infty} \mathcal{M}_n(4q + 4 + \delta) < \infty$ for some constants $q > 0$ and $\delta > 0$.

The two conditions (A3) and (A3′) require that the dependence among $X_{n,\alpha}, \alpha \in \mathcal{I}_n$, are not too strong. They are translations of (B1) and (B2) in Section 6.1 (see Remark 2 for some equivalent versions), and either of them will make our results valid. We use (A2) to get rid of the case where there may be lots of pairs $(\alpha, \beta) \in \mathcal{I}_n$ such that $X_{n,\alpha}$ and $X_{n,\beta}$ are perfectly correlated. Assumptions (A4), (A4′) and (A4″) connect the growth speed of m relative to n and the moment conditions. They are typical in the context of high dimensional covariance matrix estimation. Condition (A1) excludes the case that $X_{n,\alpha}$ is a constant.

Theorem 2. *Suppose that $\mathbf{X}_n = (X_{n,k,i})_{1\le k\le n, 1\le i\le m}$ is a data matrix whose n rows are i.i.d. m-dimensional random vectors, and whose entries have mean zero and variance one. Assume the dimension $m = m_n$ grows to infinity as $n \to \infty$, and* (A1), (A2)*, then under any one of the following conditions:*

(i) (A3) *and* (A4),
(ii) (A3′) *and* (A4′),
(iii) (A3) *and* (A4″),
(iv) *(A3′) and (A4″);*

we have for any $y \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left(nM_n^2 - 4\log m + \log(\log m) + \log(8\pi) \le y\right) = \exp\left(-e^{-y/2}\right).$$

3. Examples

Except for (A4) and (A4′), which put conditions on every single entry of the random vector $(X_{n,1,i})_{1\le i\le m}$, all the other conditions of Theorem 2 are related to the dependence among these entries, which can be arbitrarily complicated. In this section we shall provide examples which satisfy the four conditions (A1)–(A3′). Observe that if each row of $\mathbf{X}_n$ is a random vector with uncorrelated entries (specifically, the entries are independent), then all these conditions are automatically satisfied. They are also satisfied if the number of non-zero covariances is bounded.

3.1. Stationary processes

Suppose $(X_{n,k,i}) = (X_{k,i})$, and each row of $(X_{k,i})_{1\le i\le m}$ is distributed as a stationary process $(X_i)_{1\le i\le m}$ of the form

$$X_i = g(\epsilon_i, \epsilon_{i-1}, \ldots)$$

where ϵ_i's are i.i.d. random variables, and g is a measurable function such that X_i is well defined. Let $(\epsilon_i')_{i\in\mathbb{Z}}$ be an i.i.d. copy of $(\epsilon_i)_{i\in\mathbb{Z}}$, and $X_i' = g(\epsilon_i, \ldots, \epsilon_1, \epsilon_0', \epsilon_{-1}, \epsilon_{-2}, \ldots)$. Following [2], define the *physical dependence measure* of order p by

$$\delta_p(i) = \|X_i - X_i'\|_p.$$

Define the squared tail sum

$$\Psi_p(k) = \left[\sum_{i=k}^{\infty} (\delta_p(i))^2\right]^{1/2},$$

and use Ψ_p as a shorthand for $\Psi_p(0)$.

We give sufficient conditions for (A1)–(A3′) in the following lemma and leave its proof to the supplementary file.

Lemma 3. (i) *If $0 < \Psi_4 < \infty$ and $\mathrm{Var}(X_iX_j) > 0$ for all $i, j \in \mathbb{Z}$, then* (A1) *holds.*
(ii) *If in addition, $|\mathrm{Cor}(X_iX_j, X_kX_l)| < 1$ for all i, j, k, l such that they are not all the same, then* (A2) *holds.*
(iii) *Assume that the conditions of (i) and (ii) hold. If $\Psi_p(k) = o(1/\log k)$ as $k \to \infty$, then* (A3) *holds. If $\sum_{j=0}^{m} (\Psi_4(j))^2 = O(m^{1-\delta})$ for some $\delta > 0$, then (A3′) holds.*

Remark 1. Let g be a linear function with $g(\epsilon_i, \epsilon_{i-1}, \ldots) = \sum_{j=0}^{\infty} a_j \epsilon_{i-j}$, where ϵ_j are i.i.d. with mean 0 and $\mathbb{E}(|\epsilon_j|^p) < \infty$ and a_j are real coefficients with $\sum_{j=0}^{\infty} a_j^2 < \infty$. Then the physical dependence measure $\delta_p(i) = |a_i| \, \|\epsilon_0 - \epsilon_0'\|_p$. If $a_i = i^{-\beta}\ell(i)$, where $1/2 < \beta < 1$ and ℓ is a slowly varying function, then (X_i) is a long memory process. Smaller β indicates stronger dependence. Condition (iii) holds for all $\beta \in (1/2, 1)$. Moreover, if $a_i = i^{-1/2}(\log(i))^{-2}$, $i \geq 2$, which corresponds to the extremal case with very strong dependence $\beta = 1/2$, we also have $\Psi_p(k) = O((\log k)^{-3/2}) = o(1/\log k)$. So our dependence conditions are actually quite mild.

If (X_i) is a linear process which is not identically zero, then the following regularity conditions are automatically satisfied: $\Psi_4 > 0$, $\mathrm{Var}(X_i X_j) > 0$ for all $i, j \in \mathbb{Z}$, and $|\mathrm{Cor}(X_i X_j, X_k X_l)| < 1$ for all i, j, k, l such that they are not all the same.

3.2. Non-stationary linear processes

Assume that each row of $(X_{n,k,i})$ is distributed as $(X_{n,i})_{1 \leq i \leq m}$, which is of the form

$$X_{n,i} = \sum_{t \in \mathbb{Z}} f_{n,i,t} \epsilon_{i-t},$$

where ϵ_i, $i \in \mathbb{Z}$ are i.i.d. random variables with mean zero, variance one and finite fourth moment, and the sequence $(f_{n,i,t})$ satisfies $\sum_{t \in \mathbb{Z}} f_{n,i,t}^2 = 1$. Denote by κ_4 the fourth cumulant of ϵ_0. For $1 \leq i, j, k, l \leq m$, we have

$$\sigma_{n,i,j} = \sum_{t \in \mathbb{Z}} f_{n,i,i-t} f_{n,j,j-t},$$
$$\mathrm{Cov}(X_{n,i} X_{n,j}, X_{n,k} X_{n,l}) = \mathrm{Cum}(X_{n,i}, X_{n,j}, X_{n,k}, X_{n,l}) + \sigma_{n,i,k}\sigma_{n,j,l} + \sigma_{n,i,l}\sigma_{n,j,k},$$

where $\mathrm{Cum}(X_{n,i}, X_{n,j}, X_{n,k}, X_{n,l})$ is the fourth order joint cumulant of the random vector $(X_{n,i}, X_{n,j}, X_{n,k}, X_{n,l})^\top$, which can be expressed as

$$\mathrm{Cum}(X_{n,i}, X_{n,j}, X_{n,k}, X_{n,l}) = \sum_{t \in \mathbb{Z}} f_{n,i,i-t} f_{n,j,j-t} f_{n,k,k-t} f_{n,l,l-t} \kappa_4,$$

by the multilinearity of cumulants. In particular, we have

$$\mathrm{Var}(X_i X_j) = 1 + \sigma_{n,i,j}^2 + \kappa_4 \cdot \sum_{t \in \mathbb{Z}} f_{n,i,t}^2 f_{n,j,t}^2.$$

Since $\kappa_4 = \mathrm{Var}(\epsilon_0^2) - 2\left(\mathbb{E}\epsilon_0^2\right)^2 \geq -2$, the condition

$$\kappa_4 > -2 \tag{3}$$

guarantees (A1) in view of

$$\mathrm{Var}(X_i X_j) \geq (1 + \sigma_{n,i,j}^2)(1 + \min\{\kappa/2, 0\}) \geq \min\{1, 1 + \kappa/2\} > 0.$$

To ensure the validity of (A2), it is natural to assume that no pairs $X_{n,i}$ and $X_{n,j}$ are strongly correlated, *i.e.*

$$\limsup_{n \to \infty} \sup_{1 \leq i < j \leq m} \left| \sum_{t \in \mathbb{Z}} f_{n,i,i-t} f_{n,j,j-t} \right| < 1. \tag{4}$$

We need the following lemma, whose proof is elementary and will be given in the supplementary file.

Lemma 4. *The condition (4) suffices for* (A2) *if ϵ_i's are i.i.d.* $N(0, 1)$.

As an immediate consequence, when ϵ_i's are i.i.d. $N(0, 1)$, we have

$$\ell := \limsup_{n\to\infty} \inf_{*} \inf_{\rho\in\mathbb{R}} \operatorname{Var}\left(X_{n,i}X_{n,j} - \rho X_{n,k}X_{n,l}\right) > 0,$$

where $\inf_*$ is taken over all $1 \le i, j, k, l \le m$ such that $i < j, k < l$ and $(i, j) \ne (k, l)$. Observe that when ϵ_i's are i.i.d. $N(0, 1)$,

$$\begin{aligned} \operatorname{Var}\left(X_{n,i}X_{n,j} - \rho X_{n,k}X_{n,l}\right) = 2 \cdot & \sum_{t\in\mathbb{Z}} (f_{n,i,i-t} f_{n,j,j-t} - \rho f_{n,k,k-t} f_{n,l,l-t})^2 \\ & + \sum_{s<t} \big(f_{n,i,i-t} f_{n,j,j-s} + f_{n,i,i-s} f_{n,j,j-t} \\ & - \rho f_{n,k,k-t} f_{n,l,l-s} - \rho f_{n,k,k-s} f_{n,l,l-t}\big)^2; \end{aligned} \tag{5}$$

and when ϵ_i's are arbitrary variables, the variance is given by the same formula with the number 2 in (5) being replaced by $2 + \kappa_4$. Therefore, if (3) holds, then

$$\limsup_{n\to\infty} \inf_{*} \inf_{\rho\in\mathbb{R}} \operatorname{Var}\left(X_{n,i}X_{n,j} - \rho X_{n,k}X_{n,l}\right) \ge \min\{1, 1 + \kappa_4/2\} \cdot \ell > 0,$$

which implies (A2) holds. To summarize, we have shown that (3) and (4) suffice for (A2).

Now we turn to Conditions (A3) and (A3′). Set

$$h_n(k) = \sup_{1\le i\le m} \left(\sum_{|t|=\lfloor k/2 \rfloor}^{\infty} f_{n,i,t}^2 \right)^{1/2},$$

where $\lfloor x \rfloor = \max\{y \in \mathbb{Z} : y \le x\}$ for any $x \in \mathbb{E}$, then we have

$$|\sigma_{n,i,j}| \le 2h_n(0)h_n(|i-j|) = 2h_n(|i-j|).$$

Fixing a subset $\{i, j\}$, for any integer $b > 0$, there are at most $8b^2$ subsets $\{k, l\}$ such that $\{k, l\} \subset B(i; b) \cup B(j; b)$, where $B(x; r)$ is the open ball $\{y : |x - y| < r\}$. For all other subsets $\{k, l\}$, we have

$$|\operatorname{Cov}(X_{n,i}X_{n,j}, X_{n,k}X_{n,l})| \le (4 + 2\kappa_4)h_n(b),$$

and hence (A3) holds if we assume $h_n(k_n)\log k_n = o(1)$ for any positive sequence (k_n) such that $k_n \to \infty$. The condition (A3′) holds if we assume

$$\sum_{k=1}^{m} [h_n(k)]^2 = O\left(m^{1-\delta}\right)$$

for some $\delta > 0$, because

$$\left|\operatorname{Cov}(X_{n,i}X_{n,j}, X_{n,k}X_{n,l})\right| \le 2\kappa_4 h_n(|i-j|) + 2h_n(|i-k|) + 2h_n(|i-l|).$$

4. Testing for covariance structures

The asymptotic distribution given in Theorem 2 has several statistical applications. One of them is in high dimensional covariance matrix regularization, because Theorem 2 implies a uniform convergence rate for all sample covariances. Recently, [6] explored this direction, and proposed a thresholding procedure for sparse covariance matrix estimation, which is adaptive to the variability of each individual entry. Their method is superior to the uniform thresholding approach studied by Bickel and Levina [3].

Testing structures of covariance matrices is also a very important statistical problem. As mentioned in the introduction, when the data dimension is high, conventional tests often cannot be implemented or do not work well. Let Σ_n and R_n be the covariance matrix and correlation matrix of the random vector $(X_{n,1,i})_{1\le i\le m}$ respectively. Two types of tests have been studied under the large n, large m paradigm. Chen et al. [8], Bai et al. [1], Ledoit and Wolf [16] and Johnstone [14] considered the test

$$H_0 : \Sigma_n = I_m; \tag{6}$$

and [19,23,25,13] studied the problem of testing for complete independence

$$H_0 : R_n = I_m. \tag{7}$$

Their testing procedures are all based on the critical assumption that the entries of the data matrix $\mathbf{X}_n$ are i.i.d., while the hypotheses themselves only require the entries of $(X_{n,1,i})_{1\le i\le m}$ to be uncorrelated. Evidently, we can use M_n in (2) to test (7), and we only require the uncorrelatedness for the validity of the limiting distribution established in Theorem 2, as long as the mild conditions of the theorem are satisfied. On the other hand, we can also take the sample variances into consideration, and use the following test statistic

$$M_n' = \max_{1\le i\le j\le m} \frac{|\hat{\sigma}_{n,i,j} - \sigma_{n,i,j}|}{\sqrt{\hat{\tau}_{n,i,j}}}$$

to test the identity hypothesis (6), where $\sigma_{n,i,j} = I\{i = j\}$. It is not difficult to verify that M_n' has the same asymptotic distribution as M_n under the same conditions with the only difference being that we now have to take sample variances into account as well, namely, the index set $\mathcal{I}_n$ in Section 2 is redefined as $\mathcal{I}_n = \{(i, j) : 1 \le i \le j \le m\}$. Clearly, we can also use M_n' to test $H_0 : \Sigma_n = \Sigma^0$ for some known covariance matrix Σ^0.

By checking the proof of Theorem 2, it can be seen that if instead of taking the maximum over the set $\mathcal{I}_n = \{(i, j) : 1 \le i < j \le m\}$, we only take the maximum over some subset $A_n \subset \mathcal{I}_n$ whose cardinality $|A_n|$ approaches infinity, then the maximum also has the Gumbel type convergence with normalization constants which are functions of the cardinality of the set A_n. Based on this observation, we are able to consider three more testing problems.

4.1. Test for stationarity

Suppose we want to test whether the population is a stationary time series. Under the null hypothesis, each row of the data matrix $\mathbf{X}_n$ is distributed as a stationary process $(X_i)_{1\le i\le m}$. Let $\gamma_l = \mathrm{Cov}(X_0, X_l)$ be the autocovariance at lag l. In principle, we can use the following test statistic

$$\tilde{T}_n = \max_{1\le i\le j\le m} \frac{|\hat{\sigma}_{n,i,j} - \gamma_{i-j}|}{\sqrt{\hat{\tau}_{n,i,j}}}.$$

The problem is that γ_l are unknown. Fortunately, they can be estimated with higher accuracy than $\sigma_{n,i,j}$

$$\hat{\gamma}_{n,l} = \frac{1}{nm} \sum_{k=1}^{n} \sum_{i=|l|+1}^{m} (X_{n,k,i-|l|} - \hat{\mu}_n)(X_{n,k,i} - \hat{\mu}_n),$$

where $\hat{\mu}_n = (1/nm)\sum_{k=1}^{n}\sum_{i=1}^{m} X_{n,k,i}$, and we are lead to the test statistic

$$T_n = \max_{1\le i\le j\le m} \frac{|\hat{\sigma}_{n,i,j} - \hat{\gamma}_{i-j}|}{\sqrt{\hat{\tau}_{n,i,j}}}.$$

Using similar arguments of Theorem 2 of [28], under suitable conditions, we have

$$\max_{0\le l\le m-1} |\hat{\gamma}_{n,l} - \gamma_l| = O_P(\sqrt{\log m/nm}).$$

Therefore, the limiting distribution for M_n in Theorem 2 also holds for T_n.

4.2. Test for bandedness

In time series and longitudinal data analysis, it can be of interest to test whether Σ_m has the banded structure. The hypothesis to be tested is

$$H_0 : \sigma_{n,i,j} = 0 \quad \text{if } |i-j| > B, \tag{8}$$

where $B = B_n$ may depend on n. Cai and Jiang [5] studied this problem under the assumption that each row of the data matrix $\mathbf{X}_n$ is a Gaussian random vector. They proposed to use the maximum sample correlation outside the band

$$\tilde{T}_n = \max_{|i-j|>B} \frac{\hat{\sigma}_{n,i,j}}{\sqrt{\hat{\sigma}_{n,i,i}\hat{\sigma}_{n,j,j}}}$$

as the test statistic, and proved that T_n also has the Gumbel type convergence provided that $B_n = o(m)$ and several other technical conditions hold.

Apparently, our Theorem 2 can be employed to test (8). If all the conditions of the theorem are satisfied, the test statistic

$$T_n = \max_{|i-j|>B_n} \frac{|\hat{\sigma}_{n,i,j}|}{\sqrt{\hat{\tau}_{n,i,j}}}$$

has the same asymptotic distribution as M_n as long as $B_n = o(m)$. Our theory does not need the normality assumption.

4.3. Assess the tapering procedure

Banding and tapering are commonly used regularization procedures in high dimensional covariance matrix estimation. Convergence rates were first obtained by Bickel and Levina [4], and later on improved by Cai et al. [26]. Let us introduce a weaker version of the latter result. Suppose each row of $\mathbf{X}_n$ is distributed as the random vector $X = (X_i)_{1\le i\le m}$ with mean μ and covariance matrix $\Sigma = (\sigma_{ij})$. Let K_0, K and t be positive constants, and $\mathscr{C}_\eta(K_0, K, t)$ be the class of m-dimensional distributions which satisfy the following conditions

$$\max_{|i-j|=k} |\sigma_{ij}| \le Kk^{-(1+\eta)} \quad \text{for all } k; \tag{9}$$

$$\lambda_{\max}(\Sigma) \le K_0;$$

$$P\left[|v^\top(X-\mu)| > x\right] \le e^{-tx^2/2} \quad \text{for all } x > 0 \quad \text{and} \quad \|v\| = 1;$$

where $\lambda_{\max}(\Sigma)$ is the largest eigenvalue of Σ. For a given even integer $1 \le B \le m$, define the tapered estimate of the covariance matrix Σ

$$\hat{\Sigma}_{n,B_n} = \left(w_{ij}\hat{\sigma}_{n,i,j}\right),$$

where the weights correspond to a flat top kernel and are given by

$$w_{ij} = \begin{cases} 1, & \text{when } |i-j| \le B_n/2, \\ 2 - 2|i-j|/B_n, & \text{when } B_n/2 < |i-j| \le B_n, \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 5 (*[26]*). *If $m \ge n^{1/(2\eta+1)}$, $\log m = o(n)$ and $B_n = n^{1/(2\eta+1)}$, then there exists a constant $C > 0$ such that*

$$\sup_{\mathscr{C}_\eta} \mathbb{E}\left[\lambda(\hat{\Sigma}_{n,B_n} - \Sigma)\right]^2 \le Cn^{-2\eta/(2\eta+1)} + C\frac{\log m}{n}.$$

We see that it is the parameter η that decides the convergence rate under the operator norm. After such a tapering procedure has been applied, it is important to ask whether it is appropriate, and in particular, whether (9) is satisfied. We propose to use

$$T_n = \max_{|i-j|>B_n} \frac{|\hat{\sigma}_{n,i,j}|}{\sqrt{\hat{\tau}_{n,i,j}}}$$

as the test statistic. According to the observation made at the beginning of Section 4, if the conditions of Theorem 2 are satisfied, then

$$T_n' = \max_{|i-j|>B_n} \frac{|\hat{\sigma}_{n,i,j} - \sigma_{i,j}|}{\sqrt{\hat{\tau}_{n,i,j}}}$$

has the same limiting law as M_n. On the other hand, (9) implies that

$$\max_{|i-j|>B_n} |\sigma_{i,j}| = O\left(n^{-(1+\eta)/(2\eta+1)}\right),$$

so T_n has the same limiting distribution as T_n' if we further assume $\log m = o\left(n^{2/(4\eta+2)}\right)$.

5. Proof

The proofs of Theorem 2 under various conditions are similar, and they share a common Poisson approximation step, which we will formulate in Section 5.1 under a more general context, where the limiting distribution of the maximum of sample means is obtained. Since the proof of (i) is more involved, we provide the detailed proof under this assumption in Section 5.2. The proof of (ii) is almost the same, which we point out in Section 5.3. The proofs of (iii) and (iv) are provided in Section 5.4.

5.1. Maximum of sample means: an intermediate step

In this section we provide a general result on the maximum of sample means. Let $\mathbf{Y}_n = (Y_{n,k,i})_{1\le k\le n,\, i\in\mathcal{I}_n}$ be a data matrix whose n rows are i.i.d., and whose entries have mean zero and variance one, where $\mathcal{I}_n$ is an index set with cardinality $|\mathcal{I}_n| = s_n$. For each $i \in \mathcal{I}_n$, let y_i be the i-th column of $\mathbf{Y}_n$, $\bar{y}_i = (1/n) \sum_{k=1}^{n} Y_{n,k,i}$.

Define

$$W_n = \max_{i \in \mathcal{I}_n} |\bar{y}_i|. \tag{10}$$

Let Σ_n be the covariance matrix of the s_n-dimensional random vector $(Y_{n,1,i})_{i \in \mathcal{I}_n}$.

Lemma 6. *Assume Σ_n satisfies either (B1) or (B2) of Section 6.1 and $\log s_n = o(n^{1/3})$. Suppose there is a constant $C > 0$ such that $Y_{n,k,i} \in \mathscr{B}(1, Ct_n)$ for each $1 \le k \le n, i \in \mathcal{I}_n$, with*

$$t_n = \frac{\sqrt{n}\delta_n}{(\log s_n)^{3/2}},$$

where (δ_n) is a sequence of positive numbers such that $\delta_n = o(1)$, and the definition of the collection $\mathscr{B}(d, \tau)$ is given in (27) below. Then

$$\lim_{n\to\infty} P\left(nW_n^2 - 2\log s_n + \log(\log s_n) + \log \pi \le z\right) = \exp\left(-e^{-z/2}\right). \tag{11}$$

We remark that if $|Y_{n,k,i}| \le K$, then $Y_{n,k,i} \in \mathscr{B}(1, K)$. The condition $\log s_n = o(n^{1/3})$ is implicitly used to guarantee the existence of δ_n such that $\delta_n = o(1)$ and $t_n^{-1} = O(1)$.

Proof. For each $z \in \mathbb{R}$, let $z_n = (2\log s_n - \log(\log s_n) - \log \pi + z)^{1/2}$. Let $(Z_{n,i})_{i \in \mathcal{I}_n}$ be a mean zero normal random vector with covariance matrix Σ_n. For any subset $A = \{i_1, i_2, \dots, i_d\} \subset \mathcal{I}_n$, let $y_A = \sqrt{n}(\bar{y}_{i_1}, \bar{y}_{i_2}, \dots, \bar{y}_{i_d})^\top$ and $Z_A = (Z_{i_1}, Z_{i_2}, \dots, Z_{i_d})$. For a vector $\boldsymbol{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$, define $|\boldsymbol{x}|_\bullet := \min\{|x_j| : 1 \le j \le d\}$. By Lemma 9, we have for $\theta_n = \delta_n^{1/2}/\sqrt{\log s_n}$ that

$$\begin{aligned} P\left(|y_A|_\bullet > z_n\right) &\le P(|Z_A|_\bullet > z_n - \theta_n) + C_d \exp\left\{-\frac{\theta_n}{C_d \delta_n (\log s_n)^{-3/2}}\right\} \\ &\le P(|Z_A|_\bullet > z_n - \theta_n) + C_d \exp\left\{-(\log s_n)\delta_n^{-1/2}\right\}. \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{A \subset \mathcal{I}_n, |A|=d} P\left(|y_A|_\bullet > z_n\right) \le &\sum_{A \subset \mathcal{I}_n, |A|=d} P(|Z_A|_\bullet > z_n - \theta_n) \\ &+ C_d s_n^d \exp\left\{-(\log s_n)\delta_n^{-1/2}\right\}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} \sum_{A \subset \mathcal{I}_n, |A|=d} P\left(|y_A|_\bullet > z_n\right) \ge &\sum_{A \subset \mathcal{I}_n, |A|=d} P(|Z_A|_\bullet > z_n + \theta_n) \\ &- C_d s_n^d \exp\left\{-(\log s_n)\delta_n^{-1/2}\right\}. \end{aligned}$$

Since $(z_n \pm \theta_n)^2 = 2\log s_n - \log(\log s_n) - \log \pi + z + o(1)$, by Lemma 7, under either of (B1) and (B2), we have

$$\lim_{n\to\infty} \sum_{A \subset \mathcal{I}_n, |A|=d} P(|Z_A|_\bullet > z_n \pm \theta_n) = \frac{e^{-dz/2}}{d!},$$

and hence

$$\lim_{n\to\infty} \sum_{A\subset \mathcal{I}_n, |A|=d} P\left(|y_A|_\bullet > z_n\right) = \frac{e^{-dz/2}}{d\,!}.$$

The proof is complete in view of Lemma 10. □

5.2. *Proof of (i)*

We divide the proof into three steps. The first one is a truncation step, which will make the Gaussian approximation result Lemma 9 and the Bernstein inequality applicable, so that we can prove Theorem 2 under the assumption that all the involved mean and variance parameters are known. In the next two steps we show that plugging in estimated mean and variance parameters does not change the limiting distribution.

Step 1: *Truncation*. Let

$$M_{n,0} = \max_{1\le i<j\le m} \frac{1}{\sqrt{\tau_{n,i,j}}} \left| \frac{1}{n}\sum_{k=1}^{n} X_{n,k,i}X_{n,k,j} - \sigma_{n,i,j} \right|.$$

In this step, we show that

$$\lim_{n\to\infty} P\left(nM_{n,0}^2 - 4\log m + \log(\log m) + \log(8\pi) \le y\right) = \exp\left(-e^{-y/2}\right). \tag{12}$$

Let us define the operator $\mathbb{E}_0$ as $\mathbb{E}_0(X) := X - \mathbb{E}(X)$ for any random variable X. Set $\varepsilon_n = n^{-(2-p)/[4(p+4)]}$ when $0 < p < 2$, and $\varepsilon_n = n^{-1/6}(\log n)^{1/3}(\log m)^{1/2}$ when $p = 2$. Observe that (ε_n) converges to zero because of (A4). Define

$$\tilde{X}_{n,k,i} = \mathbb{E}_0\left\{X_{n,k,i} I\left[|X_{n,k,i}| \le T_n\right]\right\}, \quad \text{where } T_n = \varepsilon_n\left[n/(\log m)^3\right]^{1/4}$$

where $I[\cdot]$ denotes the indicator function. Define $\tilde{\sigma}_{n,i,j} = \mathbb{E}\left(\tilde{X}_{n,1,i}\tilde{X}_{n,1,j}\right)$, and $\tilde{\tau}_{n,i,j} = \mathrm{Var}\left(\tilde{X}_{n,1,i}\tilde{X}_{n,1,j}\right)$, and

$$M_{n,1} = \max_{1\le i<j\le m} \frac{1}{\sqrt{\tau_{n,i,j}}} \left| \frac{1}{n}\sum_{k=1}^{n} \tilde{X}_{n,k,i}\tilde{X}_{n,k,j} - \tilde{\sigma}_{n,i,j} \right|;$$

$$M_{n,2} = \max_{1\le i<j\le m} \frac{1}{\sqrt{\tilde{\tau}_{n,i,j}}} \left| \frac{1}{n}\sum_{k=1}^{n} \tilde{X}_{n,k,i}\tilde{X}_{n,k,j} - \tilde{\sigma}_{n,i,j} \right|.$$

For $\alpha = (i, j) \in \mathcal{I}_n$, let $\tilde{X}_{n,\alpha} = \tilde{X}_{n,1,i}\tilde{X}_{n,1,j}$. Elementary calculation shows that for some constant C

$$\max_{\alpha,\beta\in\mathcal{I}_n} \left|\mathrm{Cov}(\tilde{X}_{n,\alpha}, \tilde{X}_{n,\beta}) - \mathrm{Cov}(X_{n,\alpha}, X_{n,\beta})\right| \le C\exp\left\{-C^{-1}T_n^p\right\}. \tag{13}$$

Because of (A3), (13) and the assumption $\log m = o(n^{p/(p+4)})$, we know the covariance matrix of $(\tilde{X}_{n,\alpha})_{\alpha\in\mathcal{I}_n}$ satisfies (B1). On the other hand, since

$$\left|\tilde{X}_{n,\alpha}\right| \le 4T_n^2 = 4\varepsilon_n^2\sqrt{n/(\log m)^3},$$

the condition of Lemma 6 is satisfied. It follows that (12) holds if we replace $M_{n,0}$ therein by $M_{n,2}$. Furthermore, by (13) we know $M_{n,1}$ and $M_{n,2}$ have the same limiting distribution. Therefore, in order to obtain (12), it suffices to show

$$M_{n,0} - M_{n,1} = o_P\left[(n\log m)^{-1/2}\right]. \tag{14}$$

For notational simplicity, we let $Y_{n,k,i} = X_{n,k,i} - \tilde{X}_{n,k,i}$. Write

$$\begin{aligned}&\sum_{k=1}^{n}(X_{n,k,i}X_{n,k,j} - \sigma_{n,i,j}) - \sum_{k=1}^{n}(\tilde{X}_{n,k,i}\tilde{X}_{n,k,j} - \tilde{\sigma}_{n,i,j})\\&\quad = \sum_{k=1}^{n}\mathbb{E}_0(Y_{n,k,i}X_{n,k,j}) + \sum_{k=1}^{n}\mathbb{E}_0(\tilde{X}_{n,k,i}Y_{n,k,j}) =: \mathfrak{I}_{n,i,j} + \mathfrak{N}_{n,i,j}.\end{aligned}$$

For any $s \le t/4$ (t is used in the definition of (A4)), we have

$$\begin{aligned}&\sum_{r=1}^{\infty}\frac{s^r}{r!}\left|X_{n,k,i}I\left[|X_{n,k,i}| > T_n\right]X_{n,k,j}\right|^{pr/2} \le \sum_{r=1}^{\infty}\frac{s^r}{r!}\left|X_{n,k,i}X_{n,k,j}\right|^{pr/2}\frac{e^{s|X_{n,k,i}|^p}}{e^{sT_n^p}}\\&\quad\le e^{-sT_n^p}\cdot\exp\left\{s|X_{n,k,i}X_{n,k,j}|^{p/2} + s|X_{n,k,i}|^p\right\}\\&\quad\le e^{-sT_n^p}\cdot\exp\left\{2s|X_{n,k,i}|^p + s|X_{n,k,j}|^p\right\},\end{aligned}$$

and it follows that for some constant C,

$$\mathbb{E}\exp\left\{t/8\cdot|\mathbb{E}_0(Y_{n,k,i}X_{n,k,j})|^{p/2}\right\} \le \exp\left\{Ce^{-C^{-1}T_n^p}\right\}.$$

Let (δ_n) be a sequence of positive numbers which converges to zero, we have

$$\begin{aligned}\max_{1\le i<j\le m} P\left(|\mathfrak{I}_{n,i,j}| > \delta_n\sqrt{n/\log m}\right) &\le \exp\left\{n\cdot Ce^{-C^{-1}T_n^p} - C^{-1}\delta_n^{p/2}(n/\log m)^{p/4}\right\}\\&\le C\exp\left\{-(C\delta_n)^{-1}\log m\right\},\end{aligned}$$

where the last inequality is obtained by letting (δ_n) converge to zero slowly enough, which is possible because we have assumed that $\log m = o(n^{p/(p+4)})$ and $\log m = o(n^{1/3})$. It follows that

$$\max_{1\le i<j\le m}|\mathfrak{I}_{n,i,j}| = o_P\left(\sqrt{n/\log m}\right),$$

which together with a similar result on $\max_{1\le i<j\le m}|\mathfrak{N}_{n,i,j}|$ implies (14), and hence the proof of (12) is complete.

Step 2: *Effect of estimated means*. Set $\bar{X}_{n,i} = (1/n)\sum_{k=1}^{n}X_{n,k,i}$. Define

$$M_{n,3} = \max_{1\le i<j\le m}\frac{1}{\sqrt{\tau_{n,i,j}}}\left|\frac{1}{n}\sum_{k=1}^{n}(X_{n,k,i} - \bar{X}_{n,i})(X_{n,k,j} - \bar{X}_{n,j}) - \sigma_{n,i,j}\right|.$$

In this step we show that (12) also holds for $M_{n,3}$. Observe that

$$\left|M_{n,3} - M_{n,2}\right| \le \max_{1\le i<j\le m}\frac{|\bar{X}_{n,i}\bar{X}_{n,j}|}{\sqrt{\tau_{n,i,j}}} \le \max_{1\le i\le m}|\bar{X}_{n,i}|^2\cdot\left(\min_{1\le i<j\le m}\tau_{n,i,j}\right)^{-1/2}.$$

By Lemma 8 and the Bernstein inequality, for any constant $K > 0$, there is a constant C which does not depend on K such that

$$\max_{1\le i\le m} P\left(|\bar{X}_{n,i}| > K\sqrt{\frac{\log m}{n}}\right) \le C\exp\left\{-\frac{C^{-1}K^2 n\log m}{n + (K^2 n\log m)^{1-p/2} + K\sqrt{n\log m}}\right\} + Cn\exp\left\{-C^{-1}K^p(n\log m)^{p/2}\right\} \le Cm^{-K^2/C},$$

and hence

$$\max_{1\le i\le m} |\bar{X}_{n,i}| = O_P\left(\sqrt{\frac{\log m}{n}}\right), \tag{15}$$

which implies that

$$\left|M_{n,3} - M_{n,2}\right| = O_P\left(\frac{\log m}{n}\right) = o_P\left(\sqrt{\frac{1}{n\log m}}\right).$$

Therefore, (12) also holds for $M_{n,3}$.

Step 3: *Effect of estimated variances*. In this step we show that (12) holds for $\tilde{M}_n$. Since

$$n\left|M_{n,3}^2 - M_n^2\right| \le nM_{n,3}^2 \cdot \max_{1\le i<j\le m} |1 - \tau_{n,i,j}/\hat{\tau}_{n,i,j}|,$$

it suffices to show that

$$\max_{1\le i<j\le m} \left|\hat{\tau}_{n,i,j} - \tau_{n,i,j}\right| = o_P(1/\log m). \tag{16}$$

Set

$$\hat{\tau}_{n,i,j,1} = \frac{1}{n}\sum_{k=1}^n \left[(X_{n,k,i} - \bar{X}_{n,i})(X_{n,k,j} - \bar{X}_{n,j}) - \sigma_{n,i,j}\right]^2$$
$$\hat{\tau}_{n,i,j,2} = \frac{1}{n}\sum_{k=1}^n \left(X_{n,k,i}X_{n,k,j} - \sigma_{n,i,j}\right)^2.$$

Observe that

$$\hat{\tau}_{n,i,j,1} - \hat{\tau}_{n,i,j} = (\hat{\sigma}_{n,i,j} - \sigma_{n,i,j})^2$$

which together with (12) implies that

$$\max_{1\le i<j\le m} \left|\hat{\tau}_{n,i,j,1} - \hat{\tau}_{n,i,j}\right| = O_P\left(\log m/n\right). \tag{17}$$

Let (δ_n) be a sequence of positive numbers which converges to zero slowly, by Lemma 8 and the Bernstein inequality, there exist a constant C such that

$$\max_{1\le i<j\le m} P\left(|\hat{\tau}_{n,i,j,2} - \tau_{n,i,j}| \ge \delta_n/\log m\right)$$
$$\le \exp\left\{-C^{-1}\frac{(n\delta_n/\log m)^2}{n + (n\delta_n/\log m)^{2-p/4} + n\delta_n/\log m}\right\}$$

$$
\begin{aligned}
&+ Cn \exp\left\{-C^{-1}\left(\frac{n\delta_n}{\log m}\right)^{p/4}\right\} \\
&\leq C \exp\left\{-C^{-1}\left(\frac{n\delta_n}{\log m}\right)^{p/4}\right\} + C \exp\left\{-\frac{n\delta_n^2}{C(\log m)^2}\right\} \\
&\leq C \exp\left\{-(C\delta_n)^{-1} \log m\right\},
\end{aligned}
$$

where the last inequality is obtained by letting (δ_n) converge to zero slowly enough, which is possible because we have assumed that $\log m = o(n^{p/(p+4)})$ and $\log m = o(n^{1/3})$. It follows that

$$
\max_{1 \leq i < j \leq m} \left|\hat{\tau}_{n,i,j,2} - \tau_{n,i,j}\right| = o_P(1/\log m). \tag{18}
$$

In view of (17) and (18), and the assumption $\log m = o(n^q)$, we know to show (16), it remains to prove

$$
\max_{1 \leq i < j \leq m} \left|\hat{\tau}_{n,i,j,1} - \hat{\tau}_{n,i,j,2}\right| = o_P(1/\log m). \tag{19}
$$

Elementary calculations show that

$$
\max_{1 \leq i < j \leq m} \left|\hat{\tau}_{n,i,j,1} - \hat{\tau}_{n,i,j,2}\right| \leq 4h_{n,1}^2 h_{n,2} + 3h_{n,1}^4 + 4h_{n,4}^{1/2} h_{n,2}^{1/2} h_{n,1} + 2h_{n,3} h_{n,1}^2,
$$

where

$$
\begin{aligned}
h_{n,1} &= \max_{1 \leq i \leq m} |\bar{X}_{n,i}| \\
h_{n,2} &= \max_{1 \leq i \leq m} \frac{1}{n} \sum_{k=1}^{n} X_{n,k,i}^2 \\
h_{n,3} &= \max_{1 \leq i \leq j \leq m} \left| \frac{1}{n} \sum_{k=1}^{n} X_{n,k,i} X_{n,k,j} - \sigma_{n,i,j} \right| \\
h_{n,4} &= \max_{1 \leq i \leq j \leq m} \hat{\tau}_{n,i,j,2}.
\end{aligned}
$$

By (15), we know $h_{n,1} = O_P(\sqrt{\log m/n})$. By (18) we have $h_{n,4} = O_P(1)$. Using Lemma 8 and the Bernstein inequality, we can show that

$$
h_{n,3} = O_P\left(\sqrt{\log m/n}\right).
$$

As an immediate consequence, we know $h_{n,2} = O_P(1)$. Therefore,

$$
\max_{1 \leq i < j \leq m} \left|\hat{\tau}_{n,i,j,1} - \hat{\tau}_{n,i,j,2}\right| = O_P\left(\sqrt{\log m/n}\right),
$$

and (19) holds by using the assumption $\log m = o(n^{1/3})$. The proof of Theorem 2 under (A3) and (A4) is now complete.

5.3. *Proof of (ii)*

The same proof from Section 5.2 applies with the following modification. In the definition of the truncation threshold T_n, we now update ε_n as $\varepsilon_n = (\log m)^{1/2} n^{-p/(6p+8)}$. We also need

(A3′), (13) and the assumption $\log m = o(n^{p/(3p+4)})$, which is given in (A4′), to guarantee that the covariance matrix of $(\tilde{X}_{n,\alpha})_{\alpha\in\mathcal{I}_n}$ satisfies (B2).

5.4. Proofs of (iii) and (iv)

For notational simplicity, we let $p = 4(1+q)+\delta$.

Step 1: *Truncation*. We truncate $X_{n,k,i}$ by

$$\tilde{X}_{n,k,i} = X_{n,k,i} I\left\{|X_{n,k,i}| \le n^{1/4}/\log n\right\}.$$

Define $\tilde{M}_n$ similarly as M_n with $X_{n,k,i}$ being replaced by its truncated version $\tilde{X}_{n,k,i}$, we have

$$P\left(\tilde{M}_n \ne M_n\right) \le nm\mathcal{M}_n(p)n^{-p/4}(\log n)^p \le C\mathcal{M}_n(p)n^{-\delta/4}(\log n)^p = o(1).$$

Therefore, in the rest of the proof, it suffices to consider $\tilde{X}_{n,k,i}$. For notational simplicity, we still use $\tilde{X}_{n,k,i}$ to denote its centered version with mean zero.

Define $\tilde{\sigma}_{n,i,j} = \mathbb{E}\left(\tilde{X}_{n,1,i}\tilde{X}_{n,1,j}\right)$, and $\tilde{\tau}_{n,i,j} = \mathrm{Var}\left(\tilde{X}_{n,1,i}\tilde{X}_{n,1,j}\right)$. Set

$$M_{n,1} = \max_{1\le i<j\le m} \frac{1}{\sqrt{\tilde{\tau}_{n,i,j}}}\left|\frac{1}{n}\sum_{k=1}^{n}\tilde{X}_{n,k,i}\tilde{X}_{n,k,j} - \tilde{\sigma}_{n,i,j}\right|;$$

$$M_{n,2} = \max_{1\le i<j\le m} \frac{1}{\sqrt{\tilde{\tau}_{n,i,j}}}\left|\frac{1}{n}\sum_{k=1}^{n}\tilde{X}_{n,k,i}\tilde{X}_{n,k,j} - \sigma_{n,i,j}\right|.$$

Elementary calculations show that

$$\max_{1\le i\le j\le m}|\tilde{\sigma}_{n,i,j} - \sigma_{n,i,j}| \le Cn^{-(p-2)/4}(\log n)^{p-2}; \tag{20}$$

$$\max_{\alpha,\beta\in\mathcal{I}_n}\left|\mathrm{Cov}(\tilde{X}_{n,\alpha},\tilde{X}_{n,\beta}) - \mathrm{Cov}(X_{n,\alpha},X_{n,\beta})\right| \le Cn^{-(p-4)/4}(\log n)^{p-4}. \tag{21}$$

By (21), we know the covariance matrix of $(\tilde{X}_{n,\alpha})_{\alpha\in\mathcal{I}_n}$ satisfies either (B1) or (B2) if Σ_n satisfies (B1) or (B2) correspondingly. Since

$$\mathbb{E}_0\tilde{X}_{n,\alpha} \in \mathscr{B}\left[1, 8\sqrt{n}/(\log n)^2\right],$$

we know all the conditions of Lemma 6 are satisfied, and hence (12) holds if we replace $M_{n,0}$ therein by $M_{n,1}$. Combining (20) and (21), we know (12) also holds with $M_{n,0}$ being replaced by $M_{n,2}$.

Step 2: *Effect of estimated means*. Set $\bar{X}_{n,i} = (1/n)\sum_{k=1}^{n}\tilde{X}_{n,k,i}$. Define

$$M_{n,3} = \max_{1\le i<j\le m} \frac{1}{\sqrt{\tilde{\tau}_{n,i,j}}}\left|\frac{1}{n}\sum_{k=1}^{n}(\tilde{X}_{n,k,i} - \bar{X}_{n,i})(\tilde{X}_{n,k,j} - \bar{X}_{n,j}) - \sigma_{n,i,j}\right|.$$

In this step we show that (12) also holds for $M_{n,3}$. Observe that

$$\left|M_{n,3} - M_{n,2}\right| \le \max_{1\le i<j\le m}\frac{|\bar{X}_{n,i}\bar{X}_{n,j}|}{\sqrt{\tilde{\tau}_{n,i,j}}} \le \max_{1\le i\le m}|\bar{X}_{n,i}|^2\cdot\left(\min_{1\le i<j\le m}\tilde{\tau}_{n,i,j}\right)^{-1/2}.$$

Using Bernstein's inequality, we can show

$$\max_{1\le i\le m} |\bar{X}_{n,i}| = O_P\left(\sqrt{\frac{\log n}{n}}\right),$$

which in together with (21) implies that

$$\left|M_{n,3} - M_{n,2}\right| = O_P\left(\frac{\log n}{n}\right)$$

and hence (12) also holds for $M_{n,3}$.

Step 3: *Effect of estimated variances*. Denote by $\check{\sigma}_{n,i,j}$ the estimate of $\tilde{\sigma}_{n,i,j}$

$$\check{\sigma}_{n,i,j} = \frac{1}{n}\sum_{k=1}^{n}(\tilde{X}_{n,k,i} - \bar{X}_{n,i})(\tilde{X}_{n,k,j} - \bar{X}_{n,j}).$$

In the definition of $\tilde{M}_n$, $\tilde{\tau}_{n,i,j}$ is unknown, and is estimated by

$$\check{\tau}_{n,i,j} = \frac{1}{n}\sum_{k=1}^{n}\left[(\tilde{X}_{n,k,i} - \bar{X}_{n,i})(\tilde{X}_{n,k,j} - \bar{X}_{n,j}) - \check{\sigma}_{n,i,j}\right]^2.$$

In order to show that (12) holds for $\tilde{M}_n$, it suffices to verify

$$\max_{1\le i<j\le m}\left|\check{\tau}_{n,i,j} - \tilde{\tau}_{n,i,j}\right| = o_P(1/\log n). \tag{22}$$

Set

$$\check{\tau}_{n,i,j,1} = \frac{1}{n}\sum_{k=1}^{n}\left[(\tilde{X}_{n,k,i} - \bar{X}_{n,i})(\tilde{X}_{n,k,j} - \bar{X}_{n,j}) - \tilde{\sigma}_{n,i,j}\right]^2$$

$$\check{\tau}_{n,i,j,2} = \frac{1}{n}\sum_{k=1}^{n}\left(\tilde{X}_{n,k,i}\tilde{X}_{n,k,j} - \tilde{\sigma}_{n,i,j}\right)^2.$$

Using (12), we know

$$\max_{1\le i<j\le m}\left|\check{\tau}_{n,i,j,1} - \check{\tau}_{n,i,j}\right| = O_P\left(\log n/n\right). \tag{23}$$

Since

$$\left(\tilde{X}_{n,k,i}\tilde{X}_{n,k,j} - \tilde{\sigma}_{n,i,j}\right)^2 \le 64n/(\log n)^4.$$

By Corollary 1.6 of [20] (with $x = n/(\log n)^2$ and $y = n/[2(\log n)^3]$ in their inequality (1.22)), we have

$$\max_{1\le i<j\le m} P\left(|\check{\tau}_{n,i,j,2} - \tilde{\tau}_{n,i,j}| \ge (\log n)^{-2}\right) \le \left[\frac{Cn}{n(\log n)^{-2}\cdot[n(\log n)^{-3}]^{q\wedge 1}}\right]^{\log n}$$
$$\le \left[\frac{C(\log n)^5}{n^{q\wedge 1}}\right]^{\log n},$$

where $x \wedge y := \min\{x, y\}$ for any $x, y \in \mathbb{R}$. It follows that

$$\max_{1\le i<j\le m} \left|\check{\tau}_{n,i,j,2} - \tilde{\tau}_{n,i,j}\right| = O_P\left[(\log n)^{-2}\right]. \tag{24}$$

In view of (23) and (24), we know to show (22), it remains to prove

$$\max_{1\le i<j\le m} \left|\check{\tau}_{n,i,j,1} - \check{\tau}_{n,i,j,2}\right| = o_P(1/\log n). \tag{25}$$

Elementary calculations show that

$$\max_{1\le i<j\le m} \left|\check{\tau}_{n,i,j,1} - \check{\tau}_{n,i,j,2}\right| \le 4h_{n,1}^2 h_{n,2} + 3h_{n,1}^4 + 4h_{n,4}^{1/2} h_{n,2}^{1/2} h_{n,1} + 2h_{n,3} h_{n,1}^2,$$

where

$$\begin{aligned}
h_{n,1} &= \max_{1\le i\le m} |\bar{X}_{n,i}| \\
h_{n,2} &= \max_{1\le i\le m} \frac{1}{n}\sum_{k=1}^{n} \tilde{X}_{n,k,i}^2 \\
h_{n,3} &= \max_{1\le i\le j\le m} \left|\frac{1}{n}\sum_{k=1}^{n} \tilde{X}_{n,k,i}\tilde{X}_{n,k,j} - \tilde{\sigma}_{n,i,j}\right| \\
h_{n,4} &= \max_{1\le i\le j\le m} \check{\tau}_{n,i,j,2}.
\end{aligned}$$

We know $h_{n,1} = O_P(\sqrt{\log n/n})$ and $h_{n,4} = O_P(1)$. Using Bernstein's inequality, we can show that

$$h_{n,3} = O_P\left(\sqrt{\log n/n}\right),$$

and it follows that $h_{n,2} = O_P(1)$. Therefore,

$$\max_{1\le i<j\le m} \left|\check{\tau}_{n,i,j,1} - \check{\tau}_{n,i,j,2}\right| = O_P\left(\sqrt{\log n/n}\right),$$

and (25) holds. The proofs of (iii) and (iv) of Theorem 2 are now complete.

6. Some auxiliary results

In this section we provide a normal comparison principle and a Gaussian approximation result, and a Poisson convergence theorem.

6.1. *A normal comparison principle*

Suppose for each $n \ge 1$, $(X_{n,i})_{i\in\mathcal{I}_n}$ is a Gaussian random vector whose entries have mean zero and variance one, where $\mathcal{I}_n$ is an index set with cardinality $|\mathcal{I}_n| = s_n$. Let $\Sigma_n = (r_{n,i,j})_{i,j\in\mathcal{I}_n}$ be the covariance matrix of $(X_{n,i})_{i\in\mathcal{I}_n}$. Assume that $s_n \to \infty$ as $n \to \infty$.

We impose either of the following two conditions.

(B1) For any sequence (b_n) such that $b_n \to \infty$, $\gamma(n, b_n) = o\left(1/\log b_n\right)$;

and $\limsup_{n\to\infty} \gamma_n < 1$.

(B2) For any sequence (b_n) such that $b_n \to \infty$, $\gamma(n, b_n) = o(1)$;

$$\sum_{i \neq j \in \mathcal{I}_n} r_{n,i,j}^2 = O\left(s_n^{2-\delta}\right) \quad \text{for some } \delta > 0; \quad \text{and} \quad \limsup_{n\to\infty} \gamma_n < 1$$

where

$$\gamma(n, b_n) := \sup_{i \in \mathcal{I}_n} \sup_{\mathcal{A} \subset \mathcal{I}_n, |\mathcal{A}| = b_n} \inf_{j \in \mathcal{A}} \left| r_{n,i,j} \right|$$

$$\text{and} \quad \gamma_n := \sup_{i,j \in \mathcal{I}_n;\, i \neq j} |r_{n,i,j}|.$$

Lemma 7. *Assume either (B1) or (B2). For a fixed* $z \in \mathbb{R}$ *and a sequence* (z_n) *satisfying* $z_n^2 = 2\log s_n - \log\log s_n - \log \pi + 2z + o(1)$, *define*

$$A'_{n,i} = \{|X_{n,i}| > z_n\} \quad \textit{and} \quad Q'_{n,d} = \sum_{\mathcal{A} \subset \mathcal{I}_n, |\mathcal{A}| = d} P\left(\bigcap_{i \in \mathcal{A}} A'_{n,i}\right);$$

then for all $d \geq 1$, *it holds that*

$$\lim_{n\to\infty} Q'_{n,d} = \frac{e^{-dz}}{d\,!}.$$

Lemma 7 is a refined version of Lemma 20 in [28], so we omit the proof and put the details in a supplementary file.

Remark 2. The conditions imposed on $\gamma(n, b_n)$ seem a little involved. We have the following equivalent versions. Define

$$G_n(t) = \max_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_n} I\{|r_{n,i,j}| > t\}.$$

Then (i) $\gamma(n, b_n) = o(1)$ for any sequence $b_n \to \infty$ if and only if the sequence $[G_n(t)]_{n\geq 1}$ is bounded for all $t > 0$; and (ii) $\gamma(n, b_n)(\log b_n) = o(1)$ for any sequence $b_n \to \infty$ if and only if $G_n(t_n) = \exp\{o(1/t_n)\}$ for any positive sequence (t_n) converging to zero.

6.2. *Bernstein inequality under fractal exponential moments*

The following inequality, taken from [11], is an extension of the Bernstein inequality.

Lemma 8. *Let* $X, X_1, \ldots, X_n$ *be i.i.d. random variables with mean zero and unit variance. Assume that for some* $0 < \alpha < 1$,

$$\mathbb{E}\left(|X|^{3(1-\alpha)} e^{t|X|^\alpha}\right) \leq A, \quad \textit{for } 0 \leq t < T. \tag{26}$$

Let $S_n = X_1 + \cdots + X_n$. *If* $x^{1-\alpha} \geq 2A/T^2$, *then we have*

$$P[S_n \geq x] \leq \exp\left\{-\frac{x^2}{2(n + x^{2-\alpha}/T)}\right\} + nP(X \geq x).$$

6.3. A Gaussian approximation result

For a positive integer d, let $\mathfrak{B}_d$ be the Borel σ-field on the Euclidean space $\mathbb{R}^d$. Denote by $|x|$ the Euclidean norm of a vector $x \in \mathbb{R}^d$. For two probability measures P and Q on $\left(\mathbb{R}^d, \mathfrak{B}_d\right)$ and $\lambda > 0$, define the quantity

$$\pi(P, Q; \lambda) = \sup_{A \in \mathfrak{B}_d} \left\{\max\left[P(A) - Q\left(A^\lambda\right), Q(A) - P\left(A^\lambda\right)\right]\right\},$$

where A^λ is the λ-neighborhood of A

$$A^\lambda := \left\{x \in \mathbb{R}^d : \inf_{y \in A} |x - y| < \lambda\right\}.$$

For $\tau > 0$, let $\mathscr{B}(d, \tau)$ be the collection of d-dimensional random variables which satisfy the multivariate analogue of the Bernstein's condition. Denote by (x, y) the inner product of two vectors x and y.

$$\mathscr{B}(d, \tau) = \Bigg\{\xi \text{ is a random variable} : \mathbb{E}\xi = 0, \text{ and } \left|\mathbb{E}\left[(\xi, t)^2 (\xi, u)^{m-2}\right]\right| \\ \le \frac{1}{2} m! \tau^{m-2} |u|^{m-2} \mathbb{E}\left[(\xi, t)^2\right] \text{ for every } m = 3, 4, \ldots \text{ and for all } t, u \in \mathbb{R}^d\Bigg\}. \tag{27}$$

The following Lemma on the Gaussian approximation is taken from [29].

Lemma 9. *Let $\tau > 0$, and $\xi_1, \xi_2, \ldots, \xi_n \in \mathbb{R}^d$ be independent random vectors such that $\xi_i \in \mathscr{B}(d, \tau)$ for $i = 1, 2, \ldots, n$. Let $S = \xi_1 + \xi_2 + \cdots + \xi_n$, and $\mathscr{L}(S)$ be the induced distribution on $\mathbb{R}^d$. Let Φ be the Gaussian distribution with the zero mean and the same covariance matrix as that of S. Then for all $\lambda > 0$*

$$\pi[\mathscr{L}(S), \Phi; \lambda] \le c_{1,d} \exp\left(-\frac{\lambda}{c_{2,d}\tau}\right),$$

where the constants $c_{j,d}$, $j = 1, 2$ may be taken in the form $c_{j,d} = c_j d^{5/2}$.

6.4. Poisson approximation: moment method

Lemma 10. *Suppose for each $n \ge 1$, $(A_{n,i})_{i \in \mathcal{I}_n}$ is a finite collection of events. Let $I_{A_{n,i}}$ be the indicator function of $A_{n,i}$, and $W_n = \sum_{i \in \mathcal{I}} I_{A_{n,i}}$. For each $d \ge 1$, define*

$$Q_{n,d} = \sum_{\mathcal{A} \subset \mathcal{I}_n, |\mathcal{A}| = d} P\left(\bigcap_{i \in \mathcal{A}} A_{n,i}\right).$$

Suppose there exists a $\lambda > 0$ such that

$$\lim_{n \to \infty} Q_{n,d} = \lambda^d / d! \quad \text{for each } d \ge 1.$$

Then

$$\lim_{n \to \infty} P(W_n = k) = \lambda^k e^{-\lambda} / k! \quad \text{for each } k \ge 0.$$

Observe that for each $d \geq 1$, the d-th factorial moment of W_n is given by

$$\mathbb{E}\left[W_n(W_n - 1)\cdots(W_n - d + 1)\right] = d\,! \cdot Q_{n,d},$$

so Lemma 10 is essentially the moment method. The proof is elementary, and we omit details.

Acknowledgments

Han Xiao's research was supported in part by the US National Science Foundation (DMS-1209091). Wei Biao Wu's research was supported in part by the US National Science Foundation (DMS-0906073 and DMS-1106970).

Appendix. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.spa.2013.03.012.

References

[1] Zhidong Bai, Dandan Jiang, Jian-Feng Yao, Shurong Zheng, Corrections to LRT on large-dimensional covariance matrix by RMT, Ann. Statist. 37 (6B) (2009) 3822–3840.
[2] Wei Biao Wu, Nonlinear system theory: another look at dependence, Proc. Natl. Acad. Sci. USA 102 (40) (2005) 14150–14154 (electronic).
[3] Peter J. Bickel, Elizaveta Levina, Covariance regularization by thresholding, Ann. Statist. 36 (6) (2008) 2577–2604.
[4] Peter J. Bickel, Elizaveta Levina, Regularized estimation of large covariance matrices, Ann. Statist. 36 (1) (2008) 199–227.
[5] Tony Cai, Tiefeng Jiang, Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices, Ann. Statist. 39 (3) (2011) 1496–1525.
[6] Tony Cai, Weidong Liu, Adaptive thresholding for sparse covariance matrix estimation, J. Amer. Statist. Assoc. 106 (494) (2011) 672–684.
[7] T. Cai, W. Liu, Y. Xia, Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings, J. Amer. Statist. Assoc. 108 (2013) 265–277.
[8] Song Xi Chen, Li-Xin Zhang, Ping-Shou Zhong, Tests for high-dimensional covariance matrices, J. Amer. Statist. Assoc. 105 (490) (2010) 810–819.
[9] David L. Donoho, Michael Elad, Vladimir N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Inform. Theory 52 (1) (2006) 6–18.
[10] J. Fan, J. Zhang, K. Yu, Vast portfolio selection with gross-exposure constraints, J. Amer. Statist. Assoc. 107 (498) (2012) 592–606.
[11] Ning Hao, Hao H. Zhang, Interaction screening for ultra-high dimensional data, 2012, Preprint.
[12] R. Jagannathan, T. Ma, Risk reduction in large portfolios: why imposing the wrong constraints helps, J. Finance 58 (2003) 1651–1683.
[13] Tiefeng Jiang, The asymptotic distributions of the largest entries of sample correlation matrices, Ann. Appl. Probab. 14 (2) (2004) 865–880.
[14] Iain M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2) (2001) 295–327.
[15] Mark A. Kramer, Uri T. Eden, Sydney S. Cash, Eric D. Kolaczyk, Network inference with confidence from multivariate time series, Phys. Rev. E (3) 79 (6) (2009) 061916. 13.
[16] Olivier Ledoit, Michael Wolf, Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size, Ann. Statist. 30 (4) (2002) 1081–1102.
[17] Deli Li, Wei-Dong Liu, Andrew Rosalsky, Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix, Probab. Theory Related Fields 148 (1–2) (2010) 5–35.
[18] Deli Li, Andrew Rosalsky, Some strong limit theorems for the largest entries of sample correlation matrices, Ann. Appl. Probab. 16 (1) (2006) 423–447.

[19] Wei-Dong Liu, Zhengyan Lin, Qi-Man Shao, 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 (6) (2008) 2337–2366.

[20] S.V. Nagaev, Large deviations of sums of independent random variables, Ann. Probab. 7 (5) (1979) 745–789.

[21] Hisao Nagao, On some test criteria for covariance matrix, Ann. Statist. 1 (1973) 700–709.

[22] Sandrine Péché, Universality results for the largest eigenvalues of some sample covariance matrix ensembles, Probab. Theory Related Fields 143 (3–4) (2009) 481–516.

[23] James R. Schott, Testing for complete independence in high dimensions, Biometrika 92 (4) (2005) 951–956.

[24] Alexander Soshnikov, A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices, J. Stat. Phys. 108 (5–6) (2002) 1033–1056. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

[25] Muni S. Srivastava, Some tests concerning the covariance matrix in high dimensional data, J. Japan Statist. Soc. 35 (2) (2005) 251–272.

[26] T. Tony Cai, Cun-Hui Zhang, Harrison H. Zhou, Optimal rates of convergence for covariance matrix estimation, Ann. Statist. 38 (4) (2010) 2118–2144.

[27] Craig A. Tracy, Harold Widom, Level-spacing distributions and the Airy kernel, Comm. Math. Phys. 159 (1) (1994) 151–174.

[28] Han Xiao, Wei Biao Wu, Asymptotic inference of autocovariances of stationary processes, 2011, Preprint. Available at http://arxiv.org/abs/1105.3423.

[29] A. Yu Zaĭtsev, On the Gaussian approximation of convolutions under multidimensional analogues of S.N. Bernstein's inequality conditions, Probab. Theory Related Fields 74 (4) (1987) 535–566.

[30] Wang Zhou, Asymptotic distribution of the largest off-diagonal entry of correlation matrices, Trans. Amer. Math. Soc. 359 (11) (2007) 5345–5363.