The Annals of Statistics
2025, Vol. 53, No. 3, 1319–1346
https://doi.org/10.1214/25-AOS2507

SELF-NORMALIZED CRAMÉR TYPE MODERATE DEVIATION THEOREM FOR GAUSSIAN APPROXIMATION

BY JINGKUN QIU[1,a], SONG XI CHEN[2,b] AND QI-MAN SHAO[3,c]

[1] *Guanghua School of Management, Peking University,* [a] *jkqiu@stu.pku.edu.cn*
[2] *Department of Statistics and Data Science, Tsinghua University,* [b] *sxchen@tsinghua.edu.cn*
[3] *Department of Statistics and Data Science, Shenzhen International Center for Mathematics, National Center for Applied Mathematics Shenzhen, Southern University of Science and Technology,* [c] *shaoqm@sustech.edu.cn*

Berry–Esseen type bounds for Gaussian approximation of standardized sums have been extensively studied under exponential type moment conditions. In this paper, a Cramér type moderate deviation theorem is established for self-normalized Gaussian approximation under finite moment conditions. More specifically, let $X_1, X_2, \dots, X_n$ be i.i.d. $\mathbb{R}^p$-valued random vectors with zero means. Let $S_{n,j} = \sum_{i=1}^n X_{ij}$ and $V_{n,j}^2 = \sum_{i=1}^n X_{ij}^2$. We show that if the correlation matrix of X_1 is I_p and the third moment of X_1 is finite, then

$$\frac{\mathbb{P}(\max_{1\le j\le p} S_{n,j}/V_{n,j} > x)}{\mathbb{P}(\max_{1\le j\le p} Z_j > x)} \to 1$$

uniformly for $0 \le x \le o(n^{1/6})$ and for all $p \ge 1$, where $Z_1, \dots, Z_p$ are independent standard normal random variables. Similar result is also established for large x when X_1 has a general correlation matrix. The proof is based on a new Cramér type moderate deviation theorem for the minimum of several self-normalized sums. As an application, we propose a high dimensional one-sample t-test that allows for an exponential growth of p without requiring the commonly used sub-Gaussian assumption.

1. Introduction. Let $X_1, \dots, X_n$ be a sequence of independent and identically distributed (i.i.d.) $\mathbb{R}^p$-valued random vectors with zero means and finite variances. Let R be the correlation matrix of $X_1 = (X_{11}, \dots, X_{1p})^T$ and $Z^R = (Z_1^R, \dots, Z_p^R)^T \sim \mathcal{N}(0, R)$. Let $W_n = (W_{n,1}, \dots, W_{n,p})^T$ be the standardized sum, where

$$W_{n,j} = \frac{\sum_{i=1}^n X_{ij}}{(\sum_{i=1}^n \mathbb{E}X_{ij}^2)^{1/2}} \quad \text{for } 1 \le j \le p. \tag{1.1}$$

Gaussian approximation of $\max_{1\le j\le p} W_{n,j}$ by $\max_{1\le j\le p} Z_j^R$ has been extensively studied in recent years. In a seminal paper, Chernozhukov, Chetverikov and Kato [10] proved that if $\mathbb{E}X_{ij}^2 \ge c_1$ and $\mathbb{E}\exp(|X_{ij}|/c_2) \le 2$ for some positive constants c_1 and c_2, then

$$\left|\mathbb{P}\Big(\max_{1\le j\le p} W_{n,j} > x\Big) - \mathbb{P}\Big(\max_{1\le j\le p} Z_j^R > x\Big)\right| \le C\frac{\log^{7/8}(pn)}{n^{1/8}} \tag{1.2}$$

for $x \in \mathbb{R}$, where C is a positive constant depending only on c_1 and c_2.

Since then, the Gaussian approximation result in (1.2) and its extensions have been used in many aspects of modern statistical methods for high dimensional and/or complex data analysis, including testing mean vectors [7, 55], testing covariance matrices [24, 57], testing regression coefficients [15, 34, 58], testing independence or conditional independence [31,

Received October 2023; revised February 2025.
MSC2020 subject classifications. Primary 62E17, 62E20; secondary 62G10.
Key words and phrases. Gaussian approximations, moderate deviations, self-normalized.

46], change point detection [2, 8, 19, 23, 53, 56], penalized regression models [47, 59], post-selection inference [3, 4, 13, 36, 40], factor model analysis [20, 21], and graphical models [9, 27, 28, 37, 50], among others.

The Berry–Esseen type bound in (1.2) has been improved to $Cn^{-1/6}\log^{7/6}(pn)$ by Chernozhukov, Chetverikov and Kato [11] and $Cn^{-1/4}\log^{5/4}(pn)$ by Chernozhuokov et al. [14] under the same sub-exponential condition. When R is non-degenerate in the sense that $\lambda_{\min}(R) \geq c_3 > 0$ where $\lambda_{\min}(R)$ is the minimum eigenvalue, Fang and Koike [22] provided a bound of $Cn^{-1/3}\log^{4/3}(pn)$ under a sub-Gaussian condition, which has also been improved to $Cn^{-1/2}\log^4(pn)\log n$ by Lopes [35] and Kuchibhotla and Rinaldo [30] under the same sub-Gaussian condition and to $Cn^{-1/2}(\log p)^{3/2}\log n$ by Chernozhukov, Chetverikov and Koike [12] for bounded variables. Furthermore, it is shown in Fang and Koike [22] and Chernozhukov, Chetverikov and Koike [12] that the bound $Cn^{-1/2}(\log p)^{3/2}$ cannot be attained in general. For a related direction called multiplier bootstrap approximations, Deng and Zhang [18] improved the bound in (1.2) to $Cn^{-1/6}\log^{5/6}(pn)$ under the sub-Gaussian condition, which has been further improved to $Cn^{-1/2}\log^{3/2}(pn)$ by Chernozhuokov et al. [14] for sub-exponential symmetric variables.

Note that Berry–Esseen type bound like (1.2) does not imply

$$\frac{\mathbb{P}(\max_{1\leq j\leq p} W_{n,j} > x)}{\mathbb{P}(\max_{1\leq j\leq p} Z_j^R > x)} \to 1 \tag{1.3}$$

uniformly for $x \geq 0$. To see this, let Φ and ϕ be, respectively, the distribution function and the density of $\mathcal{N}(0,1)$. Let $I = I_p$ be the $p \times p$ identity matrix. Let $a \ll b$ denote $a/b \to 0$ and $a \sim b$ denote $a/b \to 1$. Since $1 - \Phi(x) \sim \phi(x)/x$ as $x \to \infty$, when $R = I_p$, $x = (2q\log p)^{1/2}$, $q \geq 1$, $p \to \infty$ and $n \ll p^{2(q-1)}(\log p)^4$, we have $p(1 - \Phi(x)) \to 0$ and therefore

$$\mathbb{P}\Big(\max_{1\leq j\leq p} Z_j^I > x\Big) \sim p\big(1 - \Phi(x)\big) \sim \frac{p^{1-q}}{2(\pi q\log p)^{1/2}} \ll \frac{(\log p)^{3/2}}{n^{1/2}}. \tag{1.4}$$

Hence, any of the aforementioned Berry–Esseen bounds does not yield (1.3). When

$$\frac{\mathbb{P}(\max_{1\leq j\leq p} W_{n,j} > x)}{\mathbb{P}(\max_{1\leq j\leq p} Z_j^R > x)} \nrightarrow 1,$$

there is no justification to use $\mathbb{P}(\max_{1\leq j\leq p} Z_j^R > x)$ to estimate $\mathbb{P}(\max_{1\leq j\leq p} W_{n,j} > x)$.

It is worthy to note that the variances $\mathbb{E}X_{ij}^2$ are typically unknown, the sub-exponential or sub-Gaussian assumption is very restrictive and so the standardized sum W_n cannot be used directly in practice. It is natural to seek the Student's t-statistics, or equivalently, the self-normalized sums. Let $T_n = (T_{n,1}, \ldots, T_{n,p})^T$ be the self-normalized sum, where

$$T_{n,j} = \frac{\sum_{i=1}^n X_{ij}}{(\sum_{i=1}^n X_{ij}^2)^{1/2}} \quad \text{for } 1 \leq j \leq p. \tag{1.5}$$

The main purpose of this paper is to establish a Cramér type moderate deviation theorem for $\max_{1\leq j\leq p} T_{n,j}$ as well as $\max_{1\leq j\leq p} |T_{n,j}|$ under finite $(2+\tau)$-th moments, $0 < \tau \leq 1$.

For classical Cramér type moderate deviation theorems concerning standardized sums, we refer to [17, 33, 38] for the univariate case and [1, 41, 43–45, 51] for the multivariate case; see Saulis and Statulevičius [42] for a comprehensive review for related topics. In particular, a Cramér type moderate deviation result for $\max_{1\leq j\leq p} |W_{n,j}|$ under the sub-exponential condition was obtained by Kuchibhotla, Mukherjee and Banerjee [29]:

$$\left|\frac{\mathbb{P}(\max_{1\leq j\leq p} |W_{n,j}| > x)}{\mathbb{P}(\max_{1\leq j\leq p} |Z_j^R| > x)} - 1\right| \leq C\frac{(1+x)(\log p)^{8/3}}{n^{1/6}} \tag{1.6}$$

for $n \geq (\log p)^{64/15}(\log n)^{32/5}(1+\mu)^{-34/5}$ and $n^{-1/6}(1+x) \leq c \log^{-28/9}(p+n)$, where c and C are positive constants depending only on c_1 and c_2 defined above, and μ is the median of $\max_{1\leq j\leq p}|Z_j^R|$. On the other hand, Cramér type moderate deviations for the self-normalized sums provide a completely novel picture. Let $X_1, \dots, X_n$ be i.i.d. random variables with zero means and finite $(2+\tau)$-th moments, $0 < \tau \leq 1$. Let $S_n = \sum_{i=1}^n X_i$ and $V_n^2 = \sum_{i=1}^n X_i^2$. Shao [48] established that

$$\frac{\mathbb{P}(S_n/V_n \geq x)}{1-\Phi(x)} \to 1 \tag{1.7}$$

uniformly for $0 \leq x \ll n^{\tau/(4+2\tau)}$. Jing, Shao and Wang [26] obtained that

$$\left|\frac{\mathbb{P}(S_n/V_n \geq x)}{1-\Phi(x)} - 1\right| \leq A \frac{(1+x)^{2+\tau}\mathbb{E}|X_1|^{2+\tau}}{n^{\tau/2}(\mathbb{E}X_1^2)^{1+\tau/2}} \tag{1.8}$$

for $0 \leq x \leq n^{\tau/(4+2\tau)}(\mathbb{E}X_1^2)^{1/2}/(\mathbb{E}|X_1|^{2+\tau})^{1/(2+\tau)}$, where A is an absolute constant. We refer to Jing, Shao and Wang [26] for general independent random variables and Shao and Zhou [49] for Studentized non-linear statistics.

The rest of this paper is organized as follows. Section 2 provides Cramér type moderate deviation theorems for self-normalized Gaussian approximation. Section 3 presents an application to high dimensional test for means. The proofs of main results are provided in Section 4, where we shall establish a new Cramér type moderate deviation theorem for the minimum of several self-normalized sums, whose proof is postponed to Section 5. Some preliminary lemmas and propositions are presented in Section 6 and the Supplementary Material [39].

2. Main results.

We first introduce the notations to be used in the paper. Let Φ and ϕ be, respectively, the distribution function and density of $\mathcal{N}(0,1)$. Let $a_n \ll b_n$ denote $a_n/b_n \to 0$ as $n \to \infty$, and $a_n \sim b_n$ denote $a_n/b_n \to 1$ for $\{a_n\}$ and $\{b_n\}$ being two sequences of real numbers. We write $a_n = o(b_n)$ if $a_n/b_n \to 0$, and $a_n = O(b_n)$ if $\limsup_{n\to\infty}|a_n/b_n| < \infty$. For $a, b \in \mathbb{R}$, let $a \wedge b$ denote $\min\{a,b\}$ and $a \vee b$ denote $\max\{a,b\}$. Let $\mathbb{R}_+ = (0,\infty)$ be the set of positive real numbers. Let I_k be the $k \times k$ identity matrix, 1_k be the k dimensional vector of ones, and $e_{k,i}$ be the ith canonical basis vector in $\mathbb{R}^k$. The matrix $R = (\rho_{ij})_{1\leq i,j\leq p}$ is reserved for the correlation matrix of X_1 and $\Omega = (\omega_{ij})_{1\leq i,j\leq p}$ for the inverse of R.

Let $Z^I = (Z_1^I, \dots, Z_p^I)^T \sim \mathcal{N}(0, I_p)$ and $Z^R = (Z_1^R, \dots, Z_p^R)^T \sim \mathcal{N}(0,R)$. For $1 \leq j_1 < \cdots < j_k \leq p$ and $1 \leq k \leq p$, let $R(j_1, \dots, j_k) = (\rho_{j_{\ell_1} j_{\ell_2}})_{1\leq \ell_1,\ell_2\leq k}$ be a $k \times k$ principal submatrix of R and $\Omega(j_1,\dots,j_k)$ be the corresponding inverse. For $\omega_0 > 0$, let $\mathbb{D}^k(\omega_0) = \{A = (a_{ij}) \in \mathbb{R}^{k\times k} : |a_{ii}| - \sum_{j\neq i}|a_{ij}| \geq \omega_0 \text{ for } 1 \leq i \leq k\}$ be the class of $k \times k$ diagonally dominant matrix within a tolerance ω_0. Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ be, respectively, the minimum and maximum eigenvalues of a square matrix A.

Let $X_1, \dots, X_n$ be a sequence of i.i.d. $\mathbb{R}^p$-valued random vectors with zero means and finite variances. Recall that R is the correlation matrix of $X_1 = (X_{11}, \dots, X_{1p})^T$ and $Z^R = (Z_1^R, \dots, Z_p^R)^T \sim \mathcal{N}(0,R)$. When $R = I_p$, we have our first main theorem as follows.

THEOREM 2.1. *Assume that* $R = I_p$ *and*

$$\max_{1\leq j\leq p} \frac{\mathbb{E}|X_{ij}|^{2+\tau}}{(\mathbb{E}X_{ij}^2)^{1+\tau/2}} \leq \gamma_\tau \tag{2.1}$$

for some constants $0 < \tau \leq 1$ *and* $\gamma_\tau < \infty$. *Then, as* $n \to \infty$,

$$\frac{\mathbb{P}(\max_{1\leq j\leq p} T_{n,j} > x)}{\mathbb{P}(\max_{1\leq j\leq p} Z_j^I > x)} \to 1 \tag{2.2}$$

and

$$\frac{\mathbb{P}(\max_{1\le j\le p}|T_{n,j}|>x)}{\mathbb{P}(\max_{1\le j\le p}|Z_j^I|>x)}\to 1 \tag{2.3}$$

uniformly for $p\ge 1$ *and* $0\le x\ll n^{\tau/(4+2\tau)}$.

Theorem 2.1 is a nontrivial multivariate extension of the univariate result (1.7) in Shao [48]. The range of $0\le x\ll n^{\tau/(4+2\tau)}$ is the best possible even when $p=1$; see Shao [48] and Chistyakov and Götze [16]. Note that both (2.2) and (2.3) are valid uniformly for all $p\ge 1$, which allows for the potential use in arbitrarily large dimensional data analysis.

For general correlation matrix $R=(\rho_{ij})_{1\le i,j\le p}$ with an additional assumption that $p(1-\Phi(x))\to 0$, we have the following theorem.

THEOREM 2.2. *Assume that Condition* (2.1) *is satisfied for some constants* $0<\tau\le 1$ *and* $\gamma_\tau<\infty$, *and the correlation matrix* $R=(\rho_{ij})_{1\le i,j\le p}$ *satisfies*

$$\max_{1\le i<j\le p}|\rho_{ij}|\le\rho_0 \tag{2.4}$$

and

$$\sum_{1\le i<j\le p}|\rho_{ij}|\le A_1 p^{c_0} \tag{2.5}$$

for some constants $\rho_0<1$, $A_1<\infty$ *and* $1\le c_0<2/(1+\rho_0)$. *Then, as* $n\to\infty$,

$$\frac{\mathbb{P}(\max_{1\le j\le p}T_{n,j}>x)}{p(1-\Phi(x))}\to 1 \tag{2.6}$$

and

$$\frac{\mathbb{P}(\max_{1\le j\le p}|T_{n,j}|>x)}{2p(1-\Phi(x))}\to 1 \tag{2.7}$$

uniformly for $p\ge 1$, $p(1-\Phi(x))\to 0$ *and* $1\ll x\ll n^{\tau/(4+2\tau)}$.

REMARK 2.1. When $\rho_{ij}=\rho(i-j)$ and $\sum_{j=-\infty}^{\infty}|\rho(j)|<\infty$, Condition (2.5) holds with $c_0=1$. Note also that we allow for the possibility that $c_0>1$ in (2.5) to cover the case of long-range dependence in R. From the proof of Theorem 2.3, one can replace (2.5) by the following assumption to cover stronger dependence: there exist positive constants $A_1<\infty$, $0\le\delta_0\le 1$ and $1\le c_0<2/(1+\rho_0)$ such that

$$\sum_{1\le i<j\le p}|\rho_{ij}|\le A_1 p^{2-\delta_0} \tag{2.8}$$

and

$$\left|\left\{(i,j):1\le i<j\le p,|\rho_{ij}|>\frac{\delta_0}{2}\right\}\right|\le A_1 p^{c_0}. \tag{2.9}$$

REMARK 2.2. Since $1-\Phi(x)\sim\phi(x)/x$ as $x\to\infty$, it is easy to see that $p(1-\Phi(x))\to 0$ is equivalent to $p\ll xe^{x^2/2}$. We also have $\mathbb{P}(\max_{1\le j\le p}Z_j^R>x)\sim\mathbb{P}(\max_{1\le j\le p}Z_j^I>x)\sim p(1-\Phi(x))$ under (2.4) and (2.5) when $p(1-\Phi(x))\to 0$. Furthermore, when $\tau=1$, the following exponential type non-uniform bound is an immediate consequence of (2.6):

$$\left|\mathbb{P}\Big(\max_{1\le j\le p}T_{n,j}>x\Big)-\mathbb{P}\Big(\max_{1\le j\le p}Z_j^R>x\Big)\right|\ll\frac{p}{xe^{x^2/2}} \tag{2.10}$$

uniformly for $1\le p\ll xe^{x^2/2}$ and $1\ll x\ll n^{1/6}$. In particular, we have $(p/x)e^{-x^2/2}\ll n^{-1/2}(\log p)^{3/2}$ when $x\ge(2\log p+\log n)^{1/2}$. This means that (2.6) recovers the best possible Berry–Esseen bounds when $x\ge(2\log p+\log n)^{1/2}$.

As pointed out by Remark 2.2, Theorem 2.2 requires $1 \le p \ll xe^{x^2/2}$, which is implied by $x \ge (2\log p)^{1/2}$. To obtain a more general result that can be valid uniformly for all $p \ge 1$ in the presence of correlation, we define $\mathbb{S}_+^p = \mathbb{S}_+^p(\rho_0, A_1, A_2, c_0, \lambda_{0,1}, \dots, \lambda_{0,p}, \omega_{0,1}, \dots, \omega_{0,p})$ as a subset of symmetric positive definite $p \times p$ matrices such that for any $R \in \mathbb{S}_+^p$, Conditions (2.4) and (2.5) are satisfied for some constants $\rho_0 < 1$, $A_1 < \infty$ and $1 \le c_0 < 2/(1+\rho_0)$, and

$$\max_{1 \le i \le p} \sum_{j=1}^{p} \rho_{ij}^2 \le A_2, \tag{2.11}$$

$$\lambda_{\min}\big(R(j_1, \dots, j_k)\big) \ge \lambda_{0,k} \tag{2.12}$$

and

$$\Omega(j_1, \dots, j_k) \in \mathbb{D}^k(\omega_{0,k}) \tag{2.13}$$

for some constants $A_2 < \infty$ and $\lambda_{0,1}, \dots, \lambda_{0,p}, \omega_{0,1}, \dots, \omega_{0,p} > 0$.

REMARK 2.3. Condition (2.11) is taken from Cai, Liu and Xia [6]. Conditions (2.12) and (2.13) are required for the use of the Cramér type moderate deviation theorem for the minimum of self-normalized sums in Section 4. It is noteworthy that we do not require a uniform lower bound for the minimum eigenvalues or diagonally dominant properties, but allow for the possibility that $\lambda_{0,k} \to 0$ and $\omega_{0,k} \to 0$ as $k \to \infty$. We also do not impose any condition on the maximum eigenvalues. The following is an example of $R \in \mathbb{S}_+^p$ in the presence of long-range dependence with $c_0 > 1$ and $\omega_{0,k} \to 0$.

EXAMPLE 1. Let $1_p = (1, \dots, 1)^T \in \mathbb{R}^p$ and $R = (1-\rho)I_p + \rho 1_p 1_p^T$ with $\rho = (8p)^{-1/2}$. Then, it follows that Conditions (2.4)–(2.5) and (2.11)–(2.13) are satisfied for all $p \ge 1$ with $\rho_0 = 1/4$, $A_1 = \sqrt{2}/8$, $A_2 = 9/8$, $c_0 = 3/2$, $R(j_1, \dots, j_k) = (1-\rho)I_k + \rho 1_k 1_k^T$, $\lambda_{0,k} = 3/4$, $\Omega(j_1, \dots, j_k) = (1-\rho)^{-1} I_k - \rho(1-\rho)^{-1}(1+(k-1)\rho)^{-1} 1_k 1_k^T$ and $\omega_{0,k} = 4/(3+k)$.

THEOREM 2.3. *Assume that Condition* (2.1) *is satisfied for some constants* $0 < \tau \le 1$ *and* $\gamma_\tau < \infty$, *and the correlation matrix* $R \in \mathbb{S}_+^p$ *for some constants* $\rho_0 < 1$, $A_1, A_2 < \infty$, $1 \le c_0 < 2/(1+\rho_0)$ *and* $\lambda_{0,1}, \dots, \lambda_{0,p}, \omega_{0,1}, \dots, \omega_{0,p} > 0$ *in Conditions* (2.4)–(2.5) *and* (2.11)–(2.13). *Then, as* $n \to \infty$,

$$\frac{\mathbb{P}(\max_{1 \le j \le p} T_{n,j} > x)}{\mathbb{P}(\max_{1 \le j \le p} Z_j^I > x)} \to 1 \tag{2.14}$$

and

$$\frac{\mathbb{P}(\max_{1 \le j \le p} |T_{n,j}| > x)}{\mathbb{P}(\max_{1 \le j \le p} |Z_j^I| > x)} \to 1 \tag{2.15}$$

uniformly for $p \ge 1$ *and* $1 \ll x \ll n^{\tau/(4+2\tau)}$.

REMARK 2.4. Assume that Condition (2.1) is satisfied for some constants $0 < \tau \le 1$ and $\gamma_\tau < \infty$. For any correlation matrix R, we conjecture that as $n \to \infty$,

$$\frac{\mathbb{P}(\max_{1 \le j \le p} T_{n,j} > x)}{\mathbb{P}(\max_{1 \le j \le p} Z_j^R > x)} \to 1 \tag{2.16}$$

and

$$\frac{\mathbb{P}(\max_{1 \le j \le p} |T_{n,j}| > x)}{\mathbb{P}(\max_{1 \le j \le p} |Z_j^R| > x)} \to 1 \tag{2.17}$$

uniformly for $p \ge 1$ and $0 \le x \ll n^{\tau/(4+2\tau)}$.

3. High dimensional one-sample t-test. Theorem 2.3 can be readily applied to construct a one-sample test for high dimensional means. Let $X_1, \dots, X_n$ be a sequence of i.i.d. $\mathbb{R}^p$-valued random vectors with mean vector μ and a positive-definite covariance matrix Σ. Let A be an invertible matrix in $\mathbb{R}^{p\times p}$ and $W_n^A = (W_{n,1}^A, \dots, W_{n,p}^A)^T$ be the standardized sum of the transformed data $AX_1, \dots, AX_n$, where

$$W_{n,j}^A = \frac{\sum_{i=1}^n e_{p,j}^T AX_i}{(ne_{p,j}^T A\Sigma A^T e_{p,j})^{1/2}} \quad \text{for } 1 \le j \le p. \tag{3.1}$$

To test $H_0 : \mu = 0$ versus $H_1 : \mu \ne 0$ or equivalently $H_0 : A\mu = 0$ versus $H_1 : A\mu \ne 0$, the test statistic considered in Cai, Liu and Xia [6] can be written as

$$M(A) = \max_{1\le j\le p} \left|W_{n,j}^A\right|^2, \tag{3.2}$$

where the choices of A can be I_p, $\Sigma^{-1/2}$ and Σ^{-1}. In particular, under $H_0 : \mu = 0$, some weak conditions on Σ, and either

$$\log p \ll n^{1/4} \quad \text{and} \quad \mathbb{E}\exp\left(\left(e_{p,j}^T\Sigma^{-1}X_i\right)^2/c_1\right) \le 2 \quad \text{for a positive constant } c_1, \tag{3.3}$$

or

$$p = O\left(n^{\tau/2-\varepsilon}\right) \quad \text{and} \quad \mathbb{E}\left|e_{p,j}^T\Sigma^{-1}X_i\right|^{2+\tau} \le \gamma_\tau \\ \text{for some } \tau > 0, \gamma_\tau < \infty \text{ and } \varepsilon > 0, \tag{3.4}$$

they proved that for any $y \in \mathbb{R}$, as $p, n \to \infty$,

$$\mathbb{P}\left(M\left(\Sigma^{-1}\right) \le 2\log p - \log\log p + y\right) \to \exp\left(-\frac{1}{\sqrt{\pi}}e^{-y/2}\right). \tag{3.5}$$

Consequently, the null hypothesis H_0 should be rejected with asymptotic level $\alpha \in (0, 1)$ if

$$M\left(\Sigma^{-1}\right) > 2\log p - \log\log p + q_\alpha, \tag{3.6}$$

where

$$q_\alpha = -\log\pi - 2\log\log\left(1/(1-\alpha)\right). \tag{3.7}$$

REMARK 3.1. Assume that Condition (3.3) is satisfied for a positive constant c_1 and that the minimum eigenvalue of the correlation matrix of $\Sigma^{-1}X_1$ is bounded below by a positive constant c_2. Then, (3.5) can also be obtained by applying the Berry–Esseen type Gaussian approximation result Corollary 1.3 in Fang and Koike [22].

REMARK 3.2. Similar conditions like (3.3) or (3.4) were also used in Chang et al. [7] and Xue and Yao [55] in which the test statistics are based on the maximum statistic and the Berry–Esseen type Gaussian or multiplier bootstrap approximation results are required.

Observe that the aforementioned conditions for (3.5) requires either a sub-Gaussian assumption or a polynomial rate restriction of p in n, both of which are very restrictive. To reduce those conditions, we shall seek for the Student's t-statistics to replace the standardized sums. Let $t_n^A = (t_{n,1}^A, \dots, t_{n,p}^A)^T$ be the t-statistic of the transformed data $AX_1, \dots, AX_n$, where

$$t_{n,j}^A = \frac{1}{\sqrt{n}\hat{\sigma}_{n,j}^A}\sum_{i=1}^n e_{p,j}^T AX_i \quad \text{for } 1 \le j \le p, \tag{3.8}$$

and

$$(\hat{\sigma}^A_{n,j})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(e^T_{p,j}AX_i - \bar{X}^A_{n,j})^2, \quad \bar{X}^A_{n,j} = \frac{1}{n}\sum_{i=1}^{n} e^T_{p,j}AX_i. \tag{3.9}$$

Then, the proposed test statistic is defined by

$$\hat{M}(A) = \max_{1\le j\le p} |t^A_{n,j}|^2. \tag{3.10}$$

The following proposition provides the asymptotic distribution of (3.10) by applying the self-normalized Gaussian approximation result in Theorem 2.3.

PROPOSITION 3.1. *Let A be an invertible matrix in $\mathbb{R}^{p\times p}$. Assume that Condition (2.1) is satisfied with X_{ij} replaced by $e^T_{p,j}AX_i$ for some constants $0<\tau\le 1$ and $\gamma_\tau<\infty$, and the correlation matrix of AX_1 is contained in $\mathbb{S}^p_+$ for some constants $\rho_0<1$, $A_1, A_2<\infty$, $1\le c_0<2/(1+\rho_0)$ and $\lambda_{0,1},\dots,\lambda_{0,p},\omega_{0,1},\dots,\omega_{0,p}>0$ in Conditions (2.4)–(2.5) and (2.11)–(2.13). Then, under $H_0:\mu=0$, for any $y\in\mathbb{R}$, as $p,n\to\infty$ and $\log p \ll n^{\tau/(2+\tau)}$,*

$$\mathbb{P}(\hat{M}(A)\le 2\log p - \log\log p + y) \to \exp\biggl(-\frac{1}{\sqrt{\pi}}e^{-y/2}\biggr). \tag{3.11}$$

Proposition 3.1 allows us to reject H_0 with asymptotic level $\alpha\in(0,1)$ if

$$\hat{M}(A) > 2\log p - \log\log p + q_\alpha, \tag{3.12}$$

where q_α is defined in (3.7). Assume that Σ is known under some weak conditions and let $A=\Sigma^{-1}$. Then, we notice that Condition (3.3) or (3.4) required to ensure the asymptotic distribution in (3.5) can be greatly reduced in Proposition 3.1 where (3.2) is replaced by (3.10). For the high dimensional scenario with $\log p\ll n^{1/4}$, the testing procedure (3.6) needs a sub-Gaussian assumption while the proposed test (3.12) only requires a $(2+\tau)$-th moment condition with $\tau\ge 2/3$. For the heavy-tailed scenario such that $\mathbb{E}|e^T_{p,j}\Sigma^{-1}X_i|^{2+\tau}\le\gamma_\tau$ for some constants $0<\tau\le 1$ and $\gamma_\tau<\infty$, the testing procedure (3.6) needs a polynomial rate restriction of $p=O(n^{\tau/2-\varepsilon})$ while the proposed test (3.12) allows for an exponential growth of $\log p\ll n^{\tau/(2+\tau)}$. Table 1 summarizes the aforementioned comparisons. Therefore, the proposed testing procedure (3.12) based on t-statistics is more suitable for higher dimensional setting and/or more heavy-tailed data, comparing to the exiting method (3.6) based on the standardized sums in Cai, Liu and Xia [6]. Furthermore, (3.12) can still be valid without requiring any knowledge of Σ if we take $A=I_p$. All of these demonstrate the power of the newly developed self-normalized Cramér type moderate deviation theorem.

TABLE 1
Comparisons between the required conditions of the testing procedure (3.6) in Cai, Liu and Xia [6] and the proposed test (3.12) under high dimensional and heavy-tailed scenarios

Scenario	High dimensional data $\log p\ll n^{1/4}$	Heavy-tailed data $\mathbb{E}\lvert e^T_{p,j}\Sigma^{-1}X_i\rvert^{2+\tau}\le\gamma_\tau$
(3.6) requires	$\mathbb{E}\exp((e^T_{p,j}\Sigma^{-1}X_i)^2/c_1)\le 2$	$p=O(n^{\tau/2-\varepsilon})$
(3.12) requires	$\mathbb{E}\lvert e^T_{p,j}\Sigma^{-1}X_i\rvert^{8/3}\le\gamma_{2/3}$	$\log p\ll n^{\tau/(2+\tau)}$

4. Proof of main results. In this section, we provide the proofs of Theorems 2.1–2.3 and Proposition 3.1. The proofs of Theorems 2.1–2.3 are based on the following new Cramér type moderate deviation theorem for $\min_{1\le\ell\le k} T_{n,j_\ell}$, where $1 \le j_1 < \cdots < j_k \le p$ and $1 \le k \le p$.

THEOREM 4.1. *Assume that*

$$\lambda_{\min}\big(R(j_1,\dots,j_k)\big) \ge c_1, \tag{4.1}$$

$$\min_{1\le i\le k} \sum_{j=1}^{k} e_{k,i}^T \Omega(j_1,\dots,j_k) e_{k,j} \ge c_2 \tag{4.2}$$

and

$$\max_{1\le\ell\le k} \frac{\mathbb{E}|X_{ij_\ell}|^{2+\tau}}{(\mathbb{E}X_{ij_\ell}^2)^{1+\tau/2}} \le \gamma_\tau \tag{4.3}$$

for some constants $c_1, c_2 > 0$, $0 < \tau \le 1$ *and* $\gamma_\tau < \infty$. *Then, there is a constant* $C \ge 1$ *depending only on* c_1, c_2, γ_τ *and* k *such that for* $0 \le x \le n^{\tau/(4+2\tau)}$,

$$\left| \frac{\mathbb{P}(\min_{1\le\ell\le k} T_{n,j_\ell} > x)}{\mathbb{P}(\min_{1\le\ell\le k} Z_{j_\ell}^R > x)} - 1 \right| \le C\left(\frac{x^{2+\tau}}{n^{\tau/2}} + \frac{1}{n^{\tau/4}} \right). \tag{4.4}$$

REMARK 4.1. In view of the proof of Theorem 4.1 which is postponed to Section 5, a key step is to solve the following optimization problem:

$$\text{find } \lambda \in \mathbb{R}_+^k \text{ subject to } \lambda^T R(j_1,\dots,j_k)\lambda = \lambda^T 1_k = 1_k^T \Omega(j_1,\dots,j_k) 1_k. \tag{4.5}$$

It can be shown that (4.5) has a unique solution $\lambda = \Omega(j_1,\dots,j_k)1_k$ if and only if $\Omega(j_1,\dots,j_k)1_k \in \mathbb{R}_+^k$, which is ensured by Condition (4.2) in Theorem 4.1. We note that Conditions (4.1) and (4.2) are satisfied for Theorem 2.1 when $R = I_p$, for Theorem 2.2 since only the results with $k \le 2$ are used in this case, and for Theorem 2.3 under (2.12) and (2.13).

REMARK 4.2. A key step of the proof of Theorem 4.1 relies on a non-uniform multivariate Berry–Esseen inequality in von Bahr [52], which involves a constant whose dependency on k was not explicitly provided. This causes difficulty in specifying the dependency of the constant C in (4.4) on k. However, a careful examination of other steps of the proof of Theorem 4.1 suggests that C has at least an iterated exponential growth in k based on the current proof techniques. As will be seen below, the growth rate does not affect the main Gaussian approximation results in Theorems 2.1–2.3 since we shall prove by truncation.

4.1. *Proof of Theorem* 2.1. We shall only prove (2.3) since the proof of (2.2) is similar. To prove (2.3), it suffices to show that for any $\varepsilon \in (0,1)$, there are constants $N \ge 1$ and $\delta \in (0,1)$ depending only on ε, τ and γ_τ such that for $p \ge 1$, $n \ge N$ and $0 \le x \le \delta n^{\tau/(4+2\tau)}$,

$$\left| \frac{\mathbb{P}(\max_{1\le j\le p} |T_{n,j}| > x)}{\mathbb{P}(\max_{1\le j\le p} |Z_j^I| > x)} - 1 \right| \le \varepsilon. \tag{4.6}$$

Since $|T_{n,j}| = T^+_{n,j} \vee T^-_{n,j}$, we have $\{|T_{n,j}| > x\} = \{T_{n,j} > x\} \cup \{-T_{n,j} > x\}$ while $\{T_{n,j} > x\} \cap \{-T_{n,j} > x\} = \varnothing$ for $x \geq 0$. Hence, by $R = I_p$ and the inclusion-exclusion formula,

$$\mathbb{P}\Big(\max_{1\le j\le p} |T_{n,j}| > x\Big) = \mathbb{P}\Bigg(\bigcup_{j=1}^{p}\{|T_{n,j}| > x\}\Bigg) = \mathbb{P}\Bigg(\bigcup_{j=1}^{p}\bigcup_{\sigma\in\{-1,1\}}\{\sigma T_{n,j} > x\}\Bigg)$$

$$= \sum_{k=1}^{p}(-1)^{k-1}\sum_{1\le j_1<\cdots<j_k\le p}\ \sum_{\sigma_1,\dots,\sigma_k\in\{-1,1\}}\mathbb{P}\Bigg(\bigcap_{1\le \ell\le k}\{\sigma_\ell T_{n,j_\ell} > x\}\Bigg) \tag{4.7}$$

$$= \sum_{k=1}^{p}(-1)^{k-1}\sum_{1\le j_1<\cdots<j_k\le p}\ \sum_{\sigma_1,\dots,\sigma_k\in\{-1,1\}}\mathbb{P}\Big(\min_{1\le \ell\le k}\sigma_\ell T_{n,j_\ell} > x\Big).$$

The key idea of this proof is to apply Theorem 4.1 so that each $\mathbb{P}(\min_{1\le\ell\le k}\sigma_\ell T_{n,j_\ell} > x)$ in (4.7) can be replaced by $\mathbb{P}(\min_{1\le\ell\le k}\sigma_\ell Z^I_{j_\ell} > x)$ with a uniform rate of convergence, where we may take $c_1 = c_2 = 1$ due to $R(j_1,\dots,j_k) \equiv I_k$. Since the constant C appeared in (4.4) depends on k, to deal with the situation where p can become arbitrarily large (and so can k), we simply truncate the inclusion-exclusion formula and use Bonferroni inequalities instead. Let $\psi_p(x) = 2p(1-\Phi(x))$. It turns out that this strategy works well for $\psi_p(x)$ small. When $\psi_p(x)$ is large, on the other hand, we notice that $\mathbb{P}(\max_{1\le j\le p}|Z^I_j| > x)$ will be close to 1 because

$$\mathbb{P}\Big(\max_{1\le j\le p}|Z^I_j| > x\Big) = 1 - \bigg(1 - \frac{\psi_p(x)}{p}\bigg)^p \geq 1 - e^{-\psi_p(x)}. \tag{4.8}$$

Hence, it suffices to show that $\mathbb{P}(\max_{1\le j\le p}|T_{n,j}| > x)$ is also close to 1. Therefore, we formulate the detailed proof into two steps.

Step 1. Here we prove (4.6) for the case of $\psi_p(x) \geq \psi_\varepsilon = 384/\varepsilon$. Let

$$\mathcal{I}_n = \sum_{j=1}^{p}\mathbb{1}(T_{n,j} > x) =: \sum_{j=1}^{p}\mathcal{I}_{n,j}, \qquad \mathcal{J}_n = \sum_{j=1}^{p}\mathbb{1}(-T_{n,j} > x) =: \sum_{j=1}^{p}\mathcal{J}_{n,j},$$

$\mathcal{P}_n = \mathbb{E}(\mathcal{I}_n + \mathcal{J}_n)$ and $\delta_n(x) = n^{-\tau/2}x^{2+\tau} + n^{-\tau/4}$. By Theorem 4.1 for $k = 1$, there is a constant $C_1 \geq 1$ depending only on γ_τ such that for $0 \le x \le n^{\tau/(4+2\tau)}$,

$$\psi_p(x)\big(1 - C_1\delta_n(x)\big) \le \mathcal{P}_n \le \psi_p(x)\big(1 + C_1\delta_n(x)\big).$$

Observe that $\psi_p(x)/2 \le \mathcal{P}_n \le 2\psi_p(x)$ and $\mathcal{P}_n - 1 \ge \mathcal{P}_n/2 \ge \psi_p(x)/4 \ge 1$ whenever $\psi_p(x) \ge 4$ and $\delta_n(x) \le 1/2C_1$. Then, by Chebyshev's inequality and Theorem 4.1 for $k = 2$, there is a constant $C_2 \ge C_1$ depending only on γ_τ such that for $p \ge 1$, $\psi_p(x) \ge \psi_\varepsilon$ and $\delta_n(x) \le \varepsilon/288C_2$, we have

$$\mathbb{P}\Big(\max_{1\le j\le p}|T_{n,j}| > x\Big) \ge 1 - \frac{\varepsilon}{3}. \tag{4.9}$$

Indeed,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1\le j\le p}|T_{n,j}| \le x\Big)\\
&\quad = \mathbb{P}(\mathcal{I}_n + \mathcal{J}_n = 0) = \mathbb{P}(\mathcal{I}_n + \mathcal{J}_n < 1)\\
&\quad = \mathbb{P}\big(\mathcal{P}_n - (\mathcal{I}_n + \mathcal{J}_n) > \mathcal{P}_n - 1\big) \le \frac{\operatorname{Var}(\mathcal{I}_n + \mathcal{J}_n)}{(\mathcal{P}_n - 1)^2} \le \frac{2\operatorname{Var}(\mathcal{I}_n) + 2\operatorname{Var}(\mathcal{J}_n)}{(\mathcal{P}_n - 1)^2}\\
&\quad \le \frac{2\sum_{j=1}^{p}(\mathbb{E}\mathcal{I}^2_{n,j} + \mathbb{E}\mathcal{J}^2_{n,j}) + 2\sum_{1\le i\ne j\le p}(\operatorname{Cov}(\mathcal{I}_{n,i},\mathcal{I}_{n,j}) + \operatorname{Cov}(\mathcal{J}_{n,i},\mathcal{J}_{n,j}))}{(\mathcal{P}_n - 1)^2}
\end{aligned}$$

$$\leq \frac{1}{(\mathcal{P}_n - 1)^2}\Bigl(2\mathcal{P}_n + 4\sum_{1\leq i\neq j\leq p}(1-\Phi(x))^2\bigl(1+C_2\delta_n(x) - \bigl(1-C_1\delta_n(x)\bigr)^2\bigr)\Bigr)$$
$$\leq \frac{16}{\psi_p^2(x)}\bigl(4\psi_p(x) + 3\psi_p^2(x)C_2\delta_n(x)\bigr) \leq \frac{\varepsilon}{3}.$$

Combining (4.8) and (4.9), we have for $\psi_p(x) \geq \psi_\varepsilon (\geq \log(3/\varepsilon))$ and $\delta_n(x) \leq \varepsilon/288C_2$,

$$\left|\frac{\mathbb{P}(\max_{1\leq j\leq p}|T_{n,j}| > x)}{\mathbb{P}(\max_{1\leq j\leq p}|Z_j^I| > x)} - 1\right| \leq \frac{\varepsilon/3 + e^{-\psi_p(x)}}{1-e^{-\psi_p(x)}} \leq \varepsilon. \tag{4.10}$$

Step 2. Assume that $\psi_p(x) \leq \psi_\varepsilon$. Then there exists a positive constant C_ε depending only on ε and γ_τ such that for $p \geq 1$ and $0 \leq x \leq n^{\tau/(4+2\tau)}$,

$$\frac{\mathbb{P}(\max_{1\leq j\leq p}|T_{n,j}| > x)}{\mathbb{P}(\max_{1\leq j\leq p}|Z_j^I| > x)} \leq 1 + \frac{\varepsilon}{2} + C_\varepsilon\delta_n(x) \tag{4.11}$$

and

$$\frac{\mathbb{P}(\max_{1\leq j\leq p}|T_{n,j}| > x)}{\mathbb{P}(\max_{1\leq j\leq p}|Z_j^I| > x)} \geq 1 - \frac{\varepsilon}{2} - C_\varepsilon\delta_n(x). \tag{4.12}$$

By taking $C_\varepsilon\delta_n(x) \leq \varepsilon/2$, (4.6) is an immediate consequence of (4.11) and (4.12). Note also that the constants N and δ in (4.6) can then be determined by the definition of $\delta_n(x)$. In particular, we may take $N = (4C_\varepsilon/\varepsilon)^{4/\tau}$ and $\delta = (\varepsilon/4C_\varepsilon)^{1/(2+\tau)}$ to obtain (4.6). Consequently, it suffices to prove (4.11) since the proof of (4.12) is similar.

PROOF OF (4.11). Let $m = m_\varepsilon$ be the smallest integer such that

$$\frac{e^{\psi_\varepsilon}\psi_\varepsilon^{2m-1}}{(2m)!} \leq \frac{\varepsilon}{2}, \tag{4.13}$$

where it can be shown that $m \sim c_0/\varepsilon$ as $\varepsilon \to 0$ for a constant $c_0 = 689.495\cdots > 384$. Let C_k be the positive constant appeared in (4.4) of Theorem 4.1 for the k-dimensional case and suppose without loss of generality that $C_1 \leq C_2 \leq \cdots \leq C_{2m}$. Let $C_\varepsilon = e^{\psi_\varepsilon}C_{2m}$.

When $p \geq 2m$, by Bonferroni inequalities, Theorem 4.1, (4.8) and (4.13),

$$\begin{aligned}
&\mathbb{P}\Bigl(\max_{1\leq j\leq p}|T_{n,j}| > x\Bigr)\\
&\leq \sum_{k=1}^{2m-1}(-1)^{k-1}\sum_{1\leq j_1<\cdots<j_k\leq p}\ \sum_{\sigma_1,\ldots,\sigma_k\in\{-1,1\}}\mathbb{P}\Bigl(\min_{1\leq \ell\leq k}\sigma_\ell T_{n,j_\ell} > x\Bigr)\\
&\leq \sum_{k=1}^{2m-1}\bigl((-1)^{k-1} + C_{2m-1}\delta_n(x)\bigr)\sum_{1\leq j_1<\cdots<j_k\leq p}\ \sum_{\sigma_1,\ldots,\sigma_k\in\{-1,1\}}\mathbb{P}\Bigl(\min_{1\leq \ell\leq k}\sigma_\ell Z_{j_\ell}^I > x\Bigr)\\
&\leq \mathbb{P}\Bigl(\max_{1\leq j\leq p}|Z_j^I| > x\Bigr) + \sum_{1\leq j_1<\cdots<j_{2m}\leq p}\ \sum_{\sigma_1,\ldots,\sigma_{2m}\in\{-1,1\}}\mathbb{P}\Bigl(\min_{1\leq \ell\leq 2m}\sigma_\ell Z_{j_\ell}^I > x\Bigr)\\
&\quad + C_{2m-1}\delta_n(x)\sum_{k=1}^{2m-1}\binom{p}{k}2^k(1-\Phi(x))^k\\
&\leq \mathbb{P}\Bigl(\max_{1\leq j\leq p}|Z_j^I| > x\Bigr) + \binom{p}{2m}2^{2m}(1-\Phi(x))^{2m} + C_{2m-1}\delta_n(x)\sum_{k=1}^{2m-1}\frac{\psi_p^k(x)}{k!}
\end{aligned}$$

$$
\begin{aligned}
&\le \mathbb{P}\Big(\max_{1\le j\le p}|Z_j^I|>x\Big)\Big(1+\frac{\psi_p^{2m}(x)}{(2m)!(1-e^{-\psi_p(x)})}+\frac{C_{2m-1}\delta_n(x)(e^{\psi_p(x)}-1)}{1-e^{-\psi_p(x)}}\Big)\\
&\le \mathbb{P}\Big(\max_{1\le j\le p}|Z_j^I|>x\Big)\Big(1+\frac{e^{\psi_\varepsilon}\psi_\varepsilon^{2m-1}}{(2m)!}+e^{\psi_\varepsilon}C_{2m-1}\delta_n(x)\Big)\\
&\le \mathbb{P}\Big(\max_{1\le j\le p}|Z_j^I|>x\Big)\Big(1+\frac{\varepsilon}{2}+C_\varepsilon\delta_n(x)\Big).
\end{aligned}
$$

Thus, we establish (4.11).

When $1\le p\le 2m$, we notice that by (4.7) and a similar argument,

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{1\le j\le p}|T_{n,j}|>x\Big)\\
&\quad=\sum_{k=1}^{p}(-1)^{k-1}\sum_{1\le j_1<\cdots<j_k\le p}\ \sum_{\sigma_1,\dots,\sigma_k\in\{-1,1\}}\mathbb{P}\Big(\min_{1\le \ell\le k}\sigma_\ell T_{n,j_\ell}>x\Big)\\
&\quad\le\sum_{k=1}^{p}\big((-1)^{k-1}+C_p\delta_n(x)\big)\sum_{1\le j_1<\cdots<j_k\le p}\ \sum_{\sigma_1,\dots,\sigma_k\in\{-1,1\}}\mathbb{P}\Big(\min_{1\le \ell\le k}\sigma_\ell Z_{j_\ell}^I>x\Big)\\
&\quad=\mathbb{P}\Big(\max_{1\le j\le p}|Z_j^I|>x\Big)+C_p\delta_n(x)\sum_{k=1}^{p}\binom{p}{k}2^k\big(1-\Phi(x)\big)^k\\
&\quad\le\mathbb{P}\Big(\max_{1\le j\le p}|Z_j^I|>x\Big)+C_p\delta_n(x)\sum_{k=1}^{p}\frac{\psi_p^k(x)}{k!}\\
&\quad\le\mathbb{P}\Big(\max_{1\le j\le p}|Z_j^I|>x\Big)\Big(1+\frac{C_p\delta_n(x)(e^{\psi_p(x)}-1)}{1-e^{-\psi_p(x)}}\Big)\\
&\quad\le\mathbb{P}\Big(\max_{1\le j\le p}|Z_j^I|>x\Big)\big(1+e^{\psi_\varepsilon}C_p\delta_n(x)\big)\le\mathbb{P}\Big(\max_{1\le j\le p}|Z_j^I|>x\Big)\big(1+C_\varepsilon\delta_n(x)\big).
\end{aligned}
$$

This completes the proof of (4.11). Proof of (4.12) is similar. □

4.2. *Proof of Theorems* 2.2 *and* 2.3. We shall only prove Theorem 2.3 since the proof of Theorem 2.2 will be contained in Step 1 below. For Theorem 2.3, we shall only prove (2.15) since the proof of (2.14) is similar. In the light of the proof of Theorem 2.1, we prove (2.15) in the following three steps corresponding to the cases of $\psi_p(x)\to 0$, $\psi_p(x)\to\infty$ and $0<\liminf\psi_p(x)\le\limsup\psi_p(x)<\infty$, respectively.

Step 1. We first prove (2.15) for the case of $\psi_p(x)\to 0$. By Bonferroni inequalities, (2.4) and Theorem 4.1 with $c_1=1-\rho_0$ and $c_2=1/(1+\rho_0)$, it follows that

$$
\mathbb{P}\Big(\max_{1\le j\le p}|T_{n,j}|>x\Big)\le\sum_{j=1}^{p}\sum_{\sigma\in\{-1,1\}}\mathbb{P}(\sigma T_{n,j}>x)\le\psi_p(x)\big(1+C_1\delta_n(x)\big)\tag{4.14}
$$

and

$$
\begin{aligned}
\mathbb{P}\Big(\max_{1\le j\le p}|T_{n,j}|>x\Big)&\ge\sum_{j=1}^{p}\sum_{\sigma\in\{-1,1\}}\mathbb{P}(\sigma T_{n,j}>x)\\
&\quad-\sum_{1\le i<j\le p}\ \sum_{\sigma_1,\sigma_2\in\{-1,1\}}\mathbb{P}(\sigma_1T_{n,i}>x,\sigma_2T_{n,j}>x)\\
&\ge\psi_p(x)\big(1-C_1\delta_n(x)\big)
\end{aligned}\tag{4.15}
$$

$$- \sum_{1\le i<j\le p} \sum_{\sigma_1,\sigma_2\in\{-1,1\}} \mathbb{P}(\sigma_1 Z_i^R > x, \sigma_2 Z_j^R > x)\big(1 + C_2\delta_n(x)\big),$$

where $\delta_n(x) = n^{-\tau/2}x^{2+\tau} + n^{-\tau/4}$ and $C_2 \ge C_1 \ge 1$ are constants depending only on ρ_0 and γ_τ. Observe that $\delta_n(x) \to 0$ for $0 \le x \ll n^{\tau/(4+2\tau)}$ as $n \to \infty$. By the normal comparison inequality in Corollary 2.1 of Li and Shao [32], we notice that for $x \in \mathbb{R}$,

$$\mathbb{P}(\sigma_1 Z_i^R > x, \sigma_2 Z_j^R > x) \le \big(1 - \Phi(x)\big)^2 + \frac{1}{4}|\rho_{ij}|\exp\biggl(-\frac{x^2}{1+|\rho_{ij}|}\biggr). \tag{4.16}$$

Then, by using (2.4), (2.8), (2.9) and the fact that $1 - \Phi(x) \ge \phi(x)/2x$ for $x \ge 1$, we have

$$\begin{aligned}
&\sum_{1\le i<j\le p} \sum_{\sigma_1,\sigma_2\in\{-1,1\}} \mathbb{P}(\sigma_1 Z_i^R > x, \sigma_2 Z_j^R > x) \\
&\quad\le \frac{\psi_p^2(x)}{2} + \biggl(\sum_{1\le i<j\le p, |\rho_{ij}|>\delta_0/2} + \sum_{1\le i<j\le p, |\rho_{ij}|\le\delta_0/2}\biggr)|\rho_{ij}|\exp\biggl(-\frac{x^2}{1+|\rho_{ij}|}\biggr) \qquad (4.17)\\
&\quad\le \frac{\psi_p^2(x)}{2} + A_1 p^{c_0}\exp\biggl(-\frac{x^2}{1+\rho_0}\biggr) + \operatorname{sgn}(\delta_0)A_1 p^{2-\delta_0}\exp\biggl(-\frac{2x^2}{2+\delta_0}\biggr),
\end{aligned}$$

where

$$\begin{aligned}
\frac{p^{c_0}}{\psi_p(x)}\exp\biggl(-\frac{x^2}{1+\rho_0}\biggr) &\le (2\pi)^{1/2}p^{c_0-1}x\exp\biggl(-\frac{(1-\rho_0)x^2}{2(1+\rho_0)}\biggr) \\
&\ll x^{c_0}\exp\biggl(\frac{(c_0(1+\rho_0)-2)x^2}{2(1+\rho_0)}\biggr) \to 0
\end{aligned} \tag{4.18}$$

and when $\delta_0 \in (0, 1]$,

$$\begin{aligned}
\frac{p^{2-\delta_0}}{\psi_p(x)}\exp\biggl(-\frac{2x^2}{2+\delta_0}\biggr) &\le (2\pi)^{1/2}p^{1-\delta_0}x\exp\biggl(-\frac{(2-\delta_0)x^2}{2(2+\delta_0)}\biggr) \\
&\ll x^{2-\delta_0}\exp\biggl(-\frac{\delta_0^2 x^2}{2(2+\delta_0)}\biggr) \to 0
\end{aligned} \tag{4.19}$$

uniformly for $1 \le p \ll xe^{x^2/2}$ as $x \to \infty$. Hence, it follows from (4.14)–(4.15) and (4.17)–(4.19) that $\mathbb{P}(\max_{1\le j\le p}|T_{n,j}| > x) \sim \psi_p(x)$. Note also that $p \ll xe^{x^2/2}$ is equivalent to $\psi_p(x) \to 0$ and we have $\mathbb{P}(\max_{1\le j\le p}|Z_j^R| > x) \sim \mathbb{P}(\max_{1\le j\le p}|Z_j^I| > x) \sim \psi_p(x)$ in this case by applying a similar argument. Thus, we establish (2.15).

Step 2. In view of (4.8) and Bonferroni inequalities, we can see that $\psi_p(x) \to 0$ if and only if $\mathbb{P}(\max_{1\le j\le p}|Z_j^I| > x) \to 0$. Hence, when $\liminf \psi_p(x) > 0$, it suffices to show that

$$\Bigl|\mathbb{P}\Bigl(\max_{1\le j\le p}|T_{n,j}| > x\Bigr) - \mathbb{P}\Bigl(\max_{1\le j\le p}|Z_j^I| > x\Bigr)\Bigr| \to 0. \tag{4.20}$$

Recall that when $x \to \infty$, $\psi_p(x) \to 0$ is equivalent to $p \ll xe^{x^2/2}$, which can be implied by $x \ge (2\log p)^{1/2}$. Then when $x \to \infty$ and $\liminf \psi_p(x) > 0$, it follows that $x \le (2\log p)^{1/2}$. In this step, we prove (4.20) for the case of $\psi_p(x) \to \infty$. Since $\mathbb{P}(\max_{1\le j\le p}|Z_j^I| > x) \to 1$ in this case, it suffices to prove $\mathbb{P}(\max_{1\le j\le p}|T_{n,j}| > x) \to 1$.

By the fact that $x \le (2\log p)^{1/2}$, following the argument in (4.17)–(4.19), we have

$$\frac{p^{c_0}}{\psi_p^2(x)}\exp\biggl(-\frac{x^2}{1+\rho_0}\biggr) \le 2\pi p^{c_0-2}x^2\exp\biggl(\frac{\rho_0 x^2}{1+\rho_0}\biggr) \le 4\pi p^{c_0-2/(1+\rho_0)}\log p \to 0$$

and when $\delta_0 \in (0, 1]$,

$$\frac{p^{2-\delta_0}}{\psi_p^2(x)} \exp\Big(-\frac{2x^2}{2+\delta_0}\Big) \le 2\pi p^{-\delta_0} x^2 \exp\Big(\frac{\delta_0 x^2}{2+\delta_0}\Big) \le 4\pi p^{-\delta_0^2/(2+\delta_0)} \log p \to 0.$$

Then, by a similar argument as that in Step 1 of the proof of Theorem 2.1 as well as the normal comparison inequality in (4.16), we have as $\psi_p(x) \to \infty$, $\delta_n(x) \to 0$ and $x \to \infty$,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1\le j\le p} |T_{n,j}| \le x\Big) \\
&\quad\le \frac{2\sum_{j=1}^{p}(\mathbb{E}\mathcal{I}_{n,j}^2 + \mathbb{E}\mathcal{J}_{n,j}^2) + 2\sum_{1\le i\ne j\le p}(\mathrm{Cov}(\mathcal{I}_{n,i}, \mathcal{I}_{n,j}) + \mathrm{Cov}(\mathcal{J}_{n,i}, \mathcal{J}_{n,j}))}{(\mathcal{P}_n - 1)^2} \\
&\quad\le \frac{1}{(\mathcal{P}_n - 1)^2}\bigg(2\mathcal{P}_n + 2\sum_{1\le i\ne j\le p}\sum_{\sigma\in\{-1,1\}} \mathbb{P}(\sigma Z_i^R > x, \sigma Z_j^R > x)(1 + C_2\delta_n(x)) \\
&\qquad - 4\sum_{1\le i\ne j\le p}(1-\Phi(x))^2(1 - C_1\delta_n(x))^2\bigg) \\
&\quad\le \frac{16}{\psi_p^2(x)}\bigg(4\psi_p(x) + 3\psi_p^2(x)C_2\delta_n(x) + \sum_{1\le i\ne j\le p}|\rho_{ij}|\exp\Big(-\frac{x^2}{1+|\rho_{ij}|}\Big)\bigg) \\
&\quad\le \frac{64}{\psi_p(x)} + 48C_2\delta_n(x) + \frac{16A_1 p^{c_0}}{\psi_p^2(x)e^{x^2/(1+\rho_0)}} + \frac{16\mathrm{sgn}(\delta_0)A_1 p^{2-\delta_0}}{\psi_p^2(x)e^{2x^2/(2+\delta_0)}} \to 0.
\end{aligned}$$

This establishes (4.20).

Step 3. It suffices to prove (4.20) for the case of $0 < \liminf \psi_p(x) \le \limsup \psi_p(x) < \infty$. We use a similar argument as that in proving Lemma 6 of Cai, Liu and Xia [6]. Let m be a positive integer such that $p \ge 2m$. We shall first let $\delta_n(x) \to 0$ and $x \to \infty$ (so that $p \to \infty$) and then take $m \to \infty$ according to the strategy used in Step 2 of the proof of Theorem 2.1. Let k be an integer such that $1 \le k \le 2m$. Let $\vartheta = \vartheta_m > 0$ be a sufficiently small positive constant that will be specified later. Following Cai, Liu and Xia [6], we define

$$\begin{aligned}
\mathcal{J}_k &= \Big\{1 \le j_1 < \cdots < j_k \le p : \max_{1\le \ell_1<\ell_2\le k} |\rho_{j_{\ell_1}}\rho_{j_{\ell_2}}| > p^{-\vartheta}\Big\}, \\
\mathcal{J}_{k|d} &= \Big\{1 \le j_1 < \cdots < j_k \le p : \mathrm{card}(S) = d,\ \text{where } S \text{ is the largest subset} \\
&\qquad \text{of } \{1,\dots,k\} \text{ such that } \max_{\ell_1\ne\ell_2\in S}|\rho_{j_{\ell_1}}\rho_{j_{\ell_2}}| \le p^{-\vartheta}\Big\}, \\
\mathcal{J}_{k|d,1} &= \Big\{1 \le j_1 < \cdots < j_k \le p : \mathrm{card}(S) = d,\ \text{where } S \text{ is the largest subset} \\
&\qquad \text{of } \{1,\dots,k\} \text{ such that } \max_{\ell_1\ne\ell_2\in S}|\rho_{j_{\ell_1}}\rho_{j_{\ell_2}}| \le p^{-\vartheta}, \\
&\qquad \text{and there exists a } \ell_0 \in \{1,\dots,k\}\backslash S \text{ such that } \max_{\ell_1\ne\ell_2\in S}\min_{i\in\{1,2\}}|\rho_{j_{\ell_0}j_{\ell_i}}| > p^{-\vartheta}\Big\}
\end{aligned}$$

and $\mathcal{J}_{k|d,2} = \mathcal{J}_{k|d}\backslash\mathcal{J}_{k|d,1}$, where $1 \le d \le k-1$. Then, it can be shown that $\mathcal{J}_k = \bigcup_{d=1}^{k-1}\mathcal{J}_{k|d}$, $\mathrm{card}(\mathcal{J}_{k|d}) \le Cp^{d+2\vartheta k}$, $\mathrm{card}(\mathcal{J}_k^c) \sim \binom{p}{k}$ and $\mathrm{card}(\mathcal{J}_{k|d,1}) \le Cp^{d-1+2\vartheta k}$, where $\vartheta = \vartheta_m < 1/4m$ and C is a positive constant depending only on m. Note also that by a similar argument as that in the proof of Lemma 6 in Cai, Liu and Xia [6], we have (i)

$$\mathbb{P}\Big(\min_{1\le\ell\le k}\sigma_\ell Z_{j_\ell}^R > x\Big) = \mathbb{P}\Big(\min_{1\le\ell\le k}\sigma_\ell Z_{j_\ell}^I > x\Big)\bigg(1 + O\Big(\frac{(\log p)^2}{p^\vartheta}\Big)\bigg) + O\Big(\frac{1}{p^{2k}}\Big) \tag{4.21}$$

uniformly for $(j_1, \dots, j_k) \in \mathcal{J}_k^c$ and $\sigma_1, \dots, \sigma_k \in \{-1, 1\}$; (ii)

$$
\begin{aligned}
\mathbb{P}\Bigl(\min_{1\le \ell\le k} \sigma_\ell Z_{j_\ell}^R > x\Bigr) &\le \mathbb{P}\Bigl(\min_{\ell\in S} \sigma_\ell Z_{j_\ell}^R > x\Bigr) \\
&= \mathbb{P}\Bigl(\min_{\ell\in S} \sigma_\ell Z_{j_\ell}^I > x\Bigr)\Bigl(1 + O\Bigl(\frac{(\log p)^2}{p^\vartheta}\Bigr)\Bigr) + O\Bigl(\frac{1}{p^{2d}}\Bigr) \qquad (4.22)\\
&= O\bigl(\psi_p^d(x) p^{-d}\bigr) + O\bigl(p^{-2d}\bigr) = O\bigl(p^{-d}\bigr)
\end{aligned}
$$

uniformly for $(j_1, \dots, j_k) \in \mathcal{J}_{k|d,1}$ and $\sigma_1, \dots, \sigma_k \in \{-1, 1\}$, where S is the largest subset of $\{1, \dots, k\}$ such that $\operatorname{card}(S) = d$ and $\max_{\ell_1\ne\ell_2\in S} |\rho_{j_{\ell_1}}\rho_{j_{\ell_2}}| \le p^{-\vartheta}$; and (iii) by the normal comparison inequality in (4.16) and the facts that $\limsup \psi_p(x) < \infty$ and $x \le (2\log p)^{1/2}$,

$$
\begin{aligned}
&\mathbb{P}\Bigl(\min_{1\le \ell\le k} \sigma_\ell Z_{j_\ell}^R > x\Bigr) \\
&\quad\le \mathbb{P}\Bigl(\min_{\ell\in S\cup\{\ell_0\}} \sigma_\ell Z_{j_\ell}^R > x\Bigr) \\
&\quad= \mathbb{P}\Bigl(\min_{\ell\in S\cup\{\ell_0\}} \sigma_\ell Z_{j_\ell}^{\tilde R} > x\Bigr)\Bigl(1 + O\Bigl(\frac{(\log p)^2}{p^\vartheta}\Bigr)\Bigr) + O\Bigl(\frac{1}{p^{2d+2}}\Bigr) \\
&\quad= \mathbb{P}\Bigl(\min_{\ell\in S\setminus\{\ell_1\}} \sigma_\ell Z_{j_\ell}^{\tilde R} > x\Bigr)\mathbb{P}\bigl(\sigma_{\ell_0} Z_{j_{\ell_0}}^{\tilde R} > x, \sigma_{\ell_1} Z_{j_{\ell_1}}^{\tilde R} > x\bigr)\bigl(1 + o(1)\bigr) + O\bigl(p^{-2d-2}\bigr) \qquad (4.23)\\
&\quad\le O\bigl(\psi_p^{d-1}(x) p^{1-d}\bigr)\mathbb{P}\bigl(\sigma_{\ell_0} Z_{j_{\ell_0}}^{\tilde R} > x, \sigma_{\ell_1} Z_{j_{\ell_1}}^{\tilde R} > x\bigr) + O\bigl(p^{-2d-2}\bigr) \\
&\quad\le O\bigl(p^{1-d}\bigr)\bigl(\bigl(1 - \Phi(x)\bigr)^2 + e^{-x^2/(1+\rho_0)}\bigr) + O\bigl(p^{-2d-2}\bigr) \\
&\quad\le O\bigl(p^{-d-1} x^2 e^{\rho_0 x^2/(1+\rho_0)}\bigr) + O\bigl(p^{-2d-2}\bigr) \le O\bigl(p^{-d-(1-\rho_0)/(1+\rho_0)} \log p\bigr)
\end{aligned}
$$

uniformly for $(j_1, \dots, j_k) \in \mathcal{J}_{k|d,2}$ and $\sigma_1, \dots, \sigma_k \in \{-1, 1\}$, where $\ell_0 \in \{1, \dots, k\}\setminus S$ and $\ell_1 \in S$ satisfy $|\rho_{j_{\ell_0} j_{\ell_1}}| > p^{-\vartheta}$, and $\tilde R = (\tilde\rho_{ij})_{1\le i,j\le p}$ is a correlation matrix such that $\tilde\rho_{j_{\ell_0} j_{\ell_1}} = \rho_{j_{\ell_0} j_{\ell_1}}$ and $\tilde\rho_{ij} = 0$ for other off-diagonal elements. Hence, by (4.21)–(4.23) and a similar argument as that in the proof of (4.11), there are constants $C_{2m} \ge \cdots \ge C_2 \ge C_1 \ge 1$ with C_k depending only on C_{k-1}, $\lambda_{0,k}$ and $\omega_{0,k}$ such that for $C_{2m-1}\delta_n(x) \le 1$,

$$
\begin{aligned}
&\mathbb{P}\Bigl(\max_{1\le j\le p} |T_{n,j}| > x\Bigr) \\
&\quad\le \sum_{k=1}^{2m-1} (-1)^{k-1} \sum_{1\le j_1<\cdots<j_k\le p} \sum_{\sigma_1,\dots,\sigma_k\in\{-1,1\}} \mathbb{P}\Bigl(\min_{1\le\ell\le k} \sigma_\ell T_{n,j_\ell} > x\Bigr) \\
&\quad\le \sum_{k=1}^{2m-1} \bigl((-1)^{k-1} + C_{2m-1}\delta_n(x)\bigr) \sum_{1\le j_1<\cdots<j_k\le p} \sum_{\sigma_1,\dots,\sigma_k\in\{-1,1\}} \mathbb{P}\Bigl(\min_{1\le\ell\le k} \sigma_\ell Z_{j_\ell}^R > x\Bigr) \\
&\quad\le \sum_{k=1}^{2m-1} \bigl((-1)^{k-1} + C_{2m-1}\delta_n(x)\bigr) \sum_{(j_1,\dots,j_k)\in\mathcal{J}_k^c} \sum_{\sigma_1,\dots,\sigma_k\in\{-1,1\}} \mathbb{P}\Bigl(\min_{1\le\ell\le k} \sigma_\ell Z_{j_\ell}^R > x\Bigr) \\
&\qquad + 2\sum_{k=1}^{2m-1}\sum_{d=1}^{k-1} \sum_{(j_1,\dots,j_k)\in\mathcal{J}_{k|d,1}} \sum_{\sigma_1,\dots,\sigma_k\in\{-1,1\}} \mathbb{P}\Bigl(\min_{1\le\ell\le k} \sigma_\ell Z_{j_\ell}^R > x\Bigr) \\
&\qquad + 2\sum_{k=1}^{2m-1}\sum_{d=1}^{k-1} \sum_{(j_1,\dots,j_k)\in\mathcal{J}_{k|d,2}} \sum_{\sigma_1,\dots,\sigma_k\in\{-1,1\}} \mathbb{P}\Bigl(\min_{1\le\ell\le k} \sigma_\ell Z_{j_\ell}^R > x\Bigr) =: I + II + III,
\end{aligned}
$$

where by first letting $\delta_n(x) \to 0$ and $x \to \infty$ (so that $p \to \infty$) and then $m \to \infty$, we have

$$II \le \sum_{k=1}^{2m-1} \sum_{d=1}^{k-1} O\bigl(p^{d-1+2\vartheta k} 2^k p^{-d}\bigr) = O\bigl(p^{4\vartheta m-1} 2^{2m}\bigr) \to 0,$$

$$\begin{aligned} III &\le \sum_{k=1}^{2m-1} \sum_{d=1}^{k-1} O\bigl(p^{d+2\vartheta k} 2^k p^{-d-(1-\rho_0)/(1+\rho_0)} \log p\bigr) \\ &= O\bigl(p^{4\vartheta m-(1-\rho_0)/(1+\rho_0)} 2^{2m} \log p\bigr) \to 0 \end{aligned}$$

and

$$\begin{aligned} I &= \sum_{k=1}^{2m-1} \bigl((-1)^{k-1} + o(1)\bigr) \sum_{(j_1,\dots,j_k)\in \mathcal{J}_k^c} \sum_{\sigma_1,\dots,\sigma_k \in \{-1,1\}} \mathbb{P}\Bigl(\min_{1\le \ell \le k} \sigma_\ell Z^I_{j_\ell} > x\Bigr) \\ &\quad + \sum_{k=1}^{2m-1} \sum_{(j_1,\dots,j_k)\in \mathcal{J}_k^c} \sum_{\sigma_1,\dots,\sigma_k \in \{-1,1\}} O\bigl(p^{-2k}\bigr) \\ &\le \sum_{k=1}^{2m-1} \bigl((-1)^{k-1} + o(1)\bigr) \sum_{1\le j_1 < \cdots < j_k \le p} \sum_{\sigma_1,\dots,\sigma_k \in \{-1,1\}} \mathbb{P}\Bigl(\min_{1\le \ell \le k} \sigma_\ell Z^I_{j_\ell} > x\Bigr) \\ &\quad + 2 \sum_{k=1}^{2m-1} \sum_{d=1}^{k-1} \sum_{(j_1,\dots,j_k)\in \mathcal{J}_{k|d,1}} \sum_{\sigma_1,\dots,\sigma_k \in \{-1,1\}} \mathbb{P}\Bigl(\min_{1\le \ell \le k} \sigma_\ell Z^I_{j_\ell} > x\Bigr) \\ &\quad + 2 \sum_{k=1}^{2m-1} \sum_{d=1}^{k-1} \sum_{(j_1,\dots,j_k)\in \mathcal{J}_{k|d,2}} \sum_{\sigma_1,\dots,\sigma_k \in \{-1,1\}} \mathbb{P}\Bigl(\min_{1\le \ell \le k} \sigma_\ell Z^I_{j_\ell} > x\Bigr) \\ &\quad + \sum_{k=1}^{2m-1} O\left(\binom{p}{k} 2^k p^{-2k}\right) \\ &\le \mathbb{P}\Bigl(\max_{1\le j\le p} |Z^I_j| > x\Bigr) + o(1) + O\bigl(e^{2/p} - 1\bigr) = \mathbb{P}\Bigl(\max_{1\le j\le p} |Z^I_j| > x\Bigr) + o(1). \end{aligned}$$

Here we have taken $\vartheta = \vartheta_m < (1-\rho_0)/4m(1+\rho_0)$. Consequently, we have

$$\mathbb{P}\Bigl(\max_{1\le j\le p} |T_{n,j}| > x\Bigr) \le \mathbb{P}\Bigl(\max_{1\le j\le p} |Z^I_j| > x\Bigr) + o(1).$$

Similarly, we can show that

$$\mathbb{P}\Bigl(\max_{1\le j\le p} |T_{n,j}| > x\Bigr) \ge \mathbb{P}\Bigl(\max_{1\le j\le p} |Z^I_j| > x\Bigr) + o(1).$$

This completes the proof of (4.20) and therefore (2.15).

4.3. *Proof of Proposition* 3.1. Let $T^A_n = (T^A_{n,1}, \dots, T^A_{n,p})^T$ be the self-normalized sum of the transformed data $AX_1, \dots, AX_n$, where

$$T^A_{n,j} = \frac{\sum_{i=1}^n e^T_{p,j} AX_i}{(\sum_{i=1}^n (e^T_{p,j} AX_i)^2)^{1/2}} \quad \text{for } 1 \le j \le p. \tag{4.24}$$

Observe that

$$\{t^A_{n,j} > x\} = \left\{T^A_{n,j} > x\left(\frac{n}{n-1+x^2}\right)^{1/2}\right\} \quad \text{for } x > 0, \tag{4.25}$$

and that

$$x - x\left(\frac{n}{n-1+x^2}\right)^{1/2} \sim \frac{x^3}{2n} \quad \text{for } 1 \ll x \ll n^{1/2}. \tag{4.26}$$

By taking $x = (2\log p - \log\log p + y)^{1/2}$ and applying Theorem 2.3, we have

$$\begin{aligned}
&\mathbb{P}(\hat{M}(A) > 2\log p - \log\log p + y) \\
&\quad = \mathbb{P}\left(\max_{1\le j\le p} |T^A_{n,j}| > \left(2\log p - \log\log p + y + O\left(\frac{(\log p)^2}{n}\right)\right)^{1/2}\right) \\
&\quad = \mathbb{P}\left(\max_{1\le j\le p} |Z^A_j| > \left(2\log p - \log\log p + y + O\left(\frac{(\log p)^2}{n}\right)\right)^{1/2}\right)(1+o(1)) \\
&\quad \to 1 - \exp\left(-\frac{1}{\sqrt{\pi}} e^{-y/2}\right),
\end{aligned}$$

as desired.

5. Proof of Theorem 4.1. We now proceed to prove (4.4). Throughout this proof as well as the proofs of the propositions below, we shall simply write X_{ij_ℓ} as X_{ij}, write T_{n,j_ℓ} as $T_{n,j}$, write $Z^R_{j_\ell}$ as Z_j, write $R(j_1,\dots,j_k)$ as R and use $C, C_1, C_2, \dots \ge 1$ to stand for constants depending only on c_1, c_2, γ_τ and k. These constants may change from case to case. Let

$$W_{n,j} = \frac{\sum_{i=1}^n X_{ij}}{(\sum_{i=1}^n \mathbb{E}X_{ij}^2)^{1/2}} \quad \text{and} \quad V_{n,j} = \left(\frac{\sum_{i=1}^n X_{ij}^2}{\sum_{i=1}^n \mathbb{E}X_{ij}^2}\right)^{1/2} \quad \text{for } 1 \le j \le k.$$

Clearly, we have $T_{n,j} = W_{n,j}/V_{n,j}$. By the Cauchy inequality, we have

$$V_{n,j} \le (1 + V_{n,j}^2)/2,$$

and therefore

$$\{T_{n,j} > x\} \supset \{xW_{n,j} - x^2V_{n,j}^2/2 > x^2/2\} = \{S_{n,j} > x^2/2\}, \tag{5.1}$$

where $S_{n,j} = \sum_{i=1}^n \eta_{ij}$, $\eta_{ij} = \xi_{ij} - \xi_{ij}^2/2$ and $\xi_{ij} = xX_{ij}/(\sum_{i=1}^n \mathbb{E}X_{ij}^2)^{1/2}$. This implies that

$$\mathbb{P}\left(\min_{1\le j\le k} T_{n,j} > x\right) \ge \mathbb{P}\left(\min_{1\le j\le k} S_{n,j} > \frac{x^2}{2}\right). \tag{5.2}$$

Hence, a lower bound of $\mathbb{P}(\min_{1\le j\le k} T_{n,j} > x)$ follows from the proposition below.

PROPOSITION 5.1. *We have for* $C_1 \le x \le n^{\tau/(4+2\tau)}$,

$$\left|\frac{\mathbb{P}(\min_{1\le j\le k} S_{n,j} > x^2/2)}{\mathbb{P}(\min_{1\le j\le k} Z_j > x)} - 1\right| \le C_2 \frac{x^{2+\tau}}{n^{\tau/2}}. \tag{5.3}$$

To derive an upper bound, following Shao and Zhou [49], we note that

$$V_{n,j} = (1 + (V_{n,j}^2 - 1))^{1/2} \ge 1 + (V_{n,j}^2 - 1)/2 - (V_{n,j}^2 - 1)^2,$$

and therefore

$$\{T_{n,j} > x\} \subset \{S_{n,j} + D_{n,j} > x^2/2\}, \tag{5.4}$$

where

$$D_{n,j} = x^2(V_{n,j}^2 - 1)^2. \tag{5.5}$$

Consequently, we have

$$\begin{aligned}\mathbb{P}\Big(\min_{1\le j\le k} T_{n,j} > x\Big) \le \mathbb{P}\Big(&\min_{1\le j\le k}(S_{n,j}+D_{n,j}) > \frac{x^2}{2}, \max_{1\le j\le k} D_{n,j} \le \varepsilon_n\Big)\\ &+\mathbb{P}\Big(\min_{1\le j\le k} T_{n,j} > x, \max_{1\le j\le k} D_{n,j} > \varepsilon_n, \max_{1\le j\le k} V_{n,j} \le M\Big)\\ &+\mathbb{P}\Big(\min_{1\le j\le k} T_{n,j} > x, \max_{1\le j\le k} V_{n,j} > M\Big),\end{aligned}\tag{5.6}$$

where $\varepsilon_n = (1/4)n^{-\tau/4} \le 1/4$ and $M = 4(k/c_1)^{1/2} \ge 4$. The reason for the choice of M will be given in the end of the proof of Proposition 5.4 below. In view of (5.6), the upper bound of $\mathbb{P}(\min_{1\le j\le k} T_{n,j} > x)$ can be derived from the following propositions.

PROPOSITION 5.2. *We have for $C_1 \le x \le n^{\tau/(4+2\tau)}$,*

$$\left|\frac{\mathbb{P}(\min_{1\le j\le k} S_{n,j} > x^2/2-\varepsilon_n)}{\mathbb{P}(\min_{1\le j\le k} Z_j > x)} - 1\right| \le C_2\Big(\frac{x^{2+\tau}}{n^{\tau/2}} + \frac{1}{n^{\tau/4}}\Big).\tag{5.7}$$

PROPOSITION 5.3. *We have for $C_1 \le x \le n^{\tau/(4+2\tau)}$,*

$$\frac{\mathbb{P}(\min_{1\le j\le k} T_{n,j} > x, \max_{1\le j\le k} D_{n,j} > \varepsilon_n, \max_{1\le j\le k} V_{n,j} \le M)}{\mathbb{P}(\min_{1\le j\le k} Z_j > x)} \le \frac{C_2}{n^{\tau/4}}.\tag{5.8}$$

PROPOSITION 5.4. *We have for $C_1 \le x \le n^{\tau/(4+2\tau)}$,*

$$\frac{\mathbb{P}(\min_{1\le j\le k} T_{n,j} > x, \max_{1\le j\le k} V_{n,j} > M)}{\mathbb{P}(\min_{1\le j\le k} Z_j > x)} \le \frac{C_2}{n^{\tau/2}}.\tag{5.9}$$

To prove (4.4) for $0 \le x \le C_1$, we need to prove a rough estimate for absolute errors.

PROPOSITION 5.5. *We have for $x \ge 0$,*

$$\Big|\mathbb{P}\Big(\min_{1\le j\le k} T_{n,j} > x\Big) - \mathbb{P}\Big(\min_{1\le j\le k} Z_j > x\Big)\Big| \le C\frac{1+x^{2+\tau}}{n^{\tau/2}}.\tag{5.10}$$

Now let us prove (4.4) for $0 \le x \le C_1$. By (4.1), (4.2) and a lower bound for multivariate Mills ratios in Proposition 3.2 of Hashorva and Hüsler [25], we have for $x \ge 0$,

$$\begin{aligned}\mathbb{P}\Big(\min_{1\le j\le k} Z_j > x\Big) &\ge \frac{\lambda_{\max}^{k/2}(R)e^{-(x^2/2)1_k^T R^{-1}1_k}}{\det^{1/2}(2\pi R)}\prod_{j=1}^k \frac{1}{1+x\lambda_{\min}^{1/2}(R)(R^{-1}1_k)_j}\\ &\ge \frac{e^{-(x^2/2)1_k^T R^{-1}1_k}}{(2\pi)^{k/2}(1+x1_k^T R^{-1}1_k)^k} \ge \frac{e^{-(k/2c_1)x^2}}{(2\pi)^{k/2}(1+kx/c_1)^k},\end{aligned}\tag{5.11}$$

and therefore $\mathbb{P}(\min_{1\le j\le k} Z_j > x) \ge 1/C_2$ when $0 \le x \le C_1$. Hence, it follows from (5.10) that for $0 \le x \le C_1$, we have

$$\left|\frac{\mathbb{P}(\min_{1\le j\le k} T_{n,j} > x)}{\mathbb{P}(\min_{1\le j\le k} Z_j > x)} - 1\right| \le C_2\frac{1+x^{2+\tau}}{n^{\tau/2}}.\tag{5.12}$$

This completes the proof.

6. Proof of propositions. We provide the proofs of Propositions 5.1–5.4. The proof of Proposition 5.5 is deferred to the Supplementary Material [39] due to the space limit. For brevity, we shall introduce a set of notations that will be used in the rest of this paper. Let $\|x\|_1 = \sum_{j=1}^k |x_j|$ and $\|x\|_\infty = \max_{1\le j\le k}|x_j|$ be, respectively, the L_1 and L_∞ norms for a vector $x \in \mathbb{R}^k$. Let $\|A\|$ be the operation norm for a matrix $A = (a_{ij}) \in \mathbb{R}^{k\times k}$. For vectors $x = (x_1, \dots, x_k)^T$, $y = (y_1, \dots, y_k)^T \in \mathbb{R}^k$ and a scalar $a \in \mathbb{R}$, we shall write $x \circ y$ and $x^{\circ a}$ according to, respectively, $(x \circ y)_j = x_j y_j$ and $(x^{\circ a})_j = x_j^a$ for $1 \le j \le k$. Sometimes we shall write $x^{\otimes 2} = xx^T$ for a vector $x \in \mathbb{R}^k$. Let $A_1 = A_2 + \Theta(1)A_3$ be a shorthand for $A_2 - A_3 \le A_1 \le A_2 + A_3$, where A_1 and A_2 are some real quantities in $\mathbb{R}$, $\mathbb{R}^k$, or $\mathbb{R}^{k\times k}$, while A_3 is a conformable quantity with non-negative elements. Here, "$\le$" is understood as an elementwise operation. We shall write $\mathbb{E}(X; A) = \mathbb{E}(X\mathbb{1}_A)$ for a random vector X and an indicator function $\mathbb{1}_A$ with respect to an event A.

6.1. *Preliminary lemmas.* In this section, we present three preliminary lemmas which are, respectively, straightforward multivariate generalizations of Lemmas 6.1 and 6.2 in Jing, Shao and Wang [26] and Lemma 5.3 in Shao and Zhou [49]. We refer to the Supplementary Material [39] for their proofs. These lemmas will be intensively used in the following subsections when proving Propositions 5.1–5.5, which are building blocks of the proof of Theorem 4.1. Let X be an $\mathbb{R}^k$-valued random vector with $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|_\infty^2 < \infty$. Let $Y = X - (1/2)X^{\circ 2}$, $Z = X^{\circ 2} - \mathbb{E}X^{\circ 2}$ and

$$\delta_1 = \mathbb{E}(\|X\|_\infty^2; \|X\|_\infty > 1) + \mathbb{E}(\|X\|_\infty^3; \|X\|_\infty \le 1).$$

LEMMA 6.1. *For λ and $\theta \in \mathbb{R}_+^k$, we have*

$$\mathbb{E}e^{\lambda^T X - \theta^T X^{\circ 2}} = 1 + \mathbb{E}\left(\frac{1}{2}(\lambda^T X)^2 - \theta^T X^{\circ 2}\right) + \Theta(1)C_{\lambda,\theta}\delta_1,$$

where

$$C_{\lambda,\theta} = \|\lambda\|_1 + \|\theta\|_1 + \frac{(\|\lambda\|_1 + \|\theta\|_1)^2}{2} + \left(1 + \frac{(\|\lambda\|_1 + \|\theta\|_1)^3}{6}\right)e^{(1/4)\|\lambda^{\circ 2}\circ\theta^{\circ(-1)}\|_1}.$$

LEMMA 6.2. *For $\lambda \in \mathbb{R}_+^k$, we have*

$$\mathbb{E}e^{\lambda^T Y} = 1 + (1/2)\mathbb{E}((\lambda^T X)^2 - \lambda^T X^{\circ 2}) + \Theta(1)C_{\lambda,0}\delta_1, \tag{6.1}$$

$$\mathbb{E}(Ye^{\lambda^T Y}) = \mathbb{E}(XX^T\lambda - (1/2)X^{\circ 2}) + \Theta(1)C_{\lambda,1}\delta_1 1_k, \tag{6.2}$$

$$\mathbb{E}(YY^T e^{\lambda^T Y}) = \mathbb{E}(XX^T) + \Theta(1)C_{\lambda,2}\delta_1 1_k 1_k^T, \tag{6.3}$$

$$\mathbb{E}(\|Y\|_\infty^3 e^{\lambda^T Y}) = \Theta(1)C_{\lambda,3}\delta_1, \tag{6.4}$$

$$(\mathbb{E}Ye^{\lambda^T Y})^{\otimes 2} = \Theta(1)C_{\lambda,4}\delta_1 1_k 1_k^T, \tag{6.5}$$

where

$$C_{\lambda,0} = (3/2)\|\lambda\|_1 + (9/8)\|\lambda\|_1^2 + (1 + (9/16)\|\lambda\|_1^3)e^{(1/2)\|\lambda\|_1},$$

$$C_{\lambda,1} = 3/2 + (9/4)\|\lambda\|_1 + ((27/16)\|\lambda\|_1^2 + (1/2)(1 \vee \|\lambda^{\circ(-1)}\|_\infty))e^{(1/2)\|\lambda\|_1},$$

$$C_{\lambda,2} = 9/4 + ((27/8)\|\lambda\|_1 + (4/e^2)(1 \vee \|\lambda^{\circ(-1)}\|_\infty)^2)e^{(1/2)\|\lambda\|_1},$$

$$C_{\lambda,3} = (27/8)(1 \vee \|\lambda^{\circ(-1)}\|_\infty)^3 e^{(1/2)\|\lambda\|_1},$$

$$C_{\lambda,4} = (3/2 + ((9/4)\|\lambda\|_1 + (1/2)(1 \vee \|\lambda^{\circ(-1)}\|_\infty))e^{(1/2)\|\lambda\|_1})^2.$$

LEMMA 6.3. *For $\lambda \in \mathbb{R}_+^k$ and $1 \le j \le k$, we have*

$$|\mathbb{E}(Z_j e^{\lambda^T Y})| \le C_{\lambda,5}\delta_1, \tag{6.6}$$

$$\mathbb{E}(Z_j^2 e^{\lambda^T Y}) \le C_{\lambda,6}\delta_1(1+\delta_1), \tag{6.7}$$

$$\mathbb{E}(|Y_j Z_j| e^{\lambda^T Y}) \le C_{\lambda,7}\delta_1, \tag{6.8}$$

$$\mathbb{E}(|Y_j Z_j^2| e^{\lambda^T Y}) \le C_{\lambda,8}\delta_1(1+\delta_1), \tag{6.9}$$

where

$$\begin{aligned}
C_{\lambda,5} &= 7(1 \vee \|\lambda\|_1) e^{(1/2)\|\lambda\|_1},\\
C_{\lambda,6} &= 6(1 \vee \|\lambda^{\circ(-1)}\|_\infty) e^{(1/2)\|\lambda\|_1},\\
C_{\lambda,7} &= 4(1 \vee \|\lambda^{\circ(-1)}\|_\infty) e^{(1/2)\|\lambda\|_1},\\
C_{\lambda,8} &= (2+16/e^2)(1 \vee \|\lambda^{\circ(-1)}\|_\infty)^2 e^{(1/2)\|\lambda\|_1}.
\end{aligned}$$

6.2. *Proof of Propositions* 5.1 *and* 5.2. We shall only prove (5.7) since (5.3) is a special case of (5.7) with $\varepsilon_n = 0$. To prove (5.7), we shall use a multivariate conjugate method (cf. (2) in von Bahr [51]) which can date back to Cramér [17]. Let $\xi_1, \dots, \xi_n$ be a sequence of i.i.d. $\mathbb{R}^k$-valued random vectors and $g : \mathbb{R}^k \to \mathbb{R}$ be a measurable function satisfying $\mathbb{E}e^{g(\xi_i)} < \infty$. Let $\hat{\xi}_1, \dots, \hat{\xi}_n$ be a sequence of i.i.d. $\mathbb{R}^k$-valued random vectors induced by the following distribution function:

$$\mathbb{P}(\hat{\xi}_i \le t) = \frac{\mathbb{E}(e^{g(\xi_i)}; \xi_i \le t)}{\mathbb{E}e^{g(\xi_i)}}. \tag{6.10}$$

Note that for every measurable function $f : \mathbb{R}^{k\times n} \to \mathbb{R}$ and Borel set A, we have

$$\mathbb{P}(f(\xi_1, \dots, \xi_n) \in A) = \mathbb{E}e^{\sum_{i=1}^n g(\xi_i)} \mathbb{E}(e^{-\sum_{i=1}^n g(\hat{\xi}_i)}; f(\hat{\xi}_1, \dots, \hat{\xi}_n) \in A). \tag{6.11}$$

In view of Lemma 6.2, we may take $\xi_{ij} = xX_{ij}/(\sum_{i=1}^n \mathbb{E}X_{ij}^2)^{1/2}$ and $g(\xi_i) = \lambda^T \eta_i$, where $\lambda = R^{-1}1_k$ and $\eta_{ij} = \xi_{ij} - \xi_{ij}^2/2$. We notice that Conditions (4.1) and (4.2) ensure that $\lambda \in [c_2, k/c_1]^k \subset \mathbb{R}_+^k$ and therefore $\|\lambda\|_1 \vee \|\lambda^{\circ(-1)}\|_\infty \le C$. Then, by letting $X = \xi_i$ and $Y = \eta_i$ in Lemma 6.2 with Remark 4.1 and $\delta_1 \le \mathbb{E}\|\xi_i\|_\infty^{2+\tau} \le Cn^{-(1+\tau/2)}x^{2+\tau}$, we have

$$\begin{aligned}
\mathbb{E}e^{\lambda^T \eta_i} &= 1 + \Theta(1)C_1 n^{-(1+\tau/2)} x^{2+\tau},\\
\mathbb{E}(\eta_i e^{\lambda^T \eta_i}) &= (1/2)n^{-1}x^2 1_k + \Theta(1)C_2 n^{-(1+\tau/2)} x^{2+\tau} 1_k,\\
\mathbb{E}(\eta_i \eta_i^T e^{\lambda^T \eta_i}) &= n^{-1}x^2 R + \Theta(1)C_3 n^{-(1+\tau/2)} x^{2+\tau} 1_k 1_k^T,\\
\mathbb{E}(\|\eta_i\|_\infty^{2+\tau} e^{\lambda^T \eta_i}) &= \Theta(1)C_4 n^{-(1+\tau/2)} x^{2+\tau},\\
(\mathbb{E}\eta_i e^{\lambda^T \eta_i})^{\otimes 2} &= \Theta(1)C_5 n^{-(1+\tau/2)} x^{2+\tau} 1_k 1_k^T.
\end{aligned}$$

Let $S_n = \sum_{i=1}^n \eta_i$. Since $e^{s-s^2} \le 1 + s \le e^s$ for $|s| \le 1/2$, it follows that

$$\mathbb{E}e^{\lambda^T S_n} = \exp\left(\Theta(1)C_6 \frac{x^{2+\tau}}{n^{\tau/2}}\right) \tag{6.12}$$

for $(2C_1)^{\tau/(4+2\tau)} \le (2C_1)^{1/6} \le x \le n^{\tau/(4+2\tau)}$.

Let $\hat S_n = \sum_{i=1}^n \hat\eta_i$. Then, by the change of measure in (6.10), we have

$$\hat\mu_n := \mathbb{E}\hat S_n = \sum_{i=1}^n \frac{\mathbb{E}(\eta_i e^{\lambda^T\eta_i})}{\mathbb{E}e^{\lambda^T\eta_i}} = \frac{x^2}{2}1_k + \Theta(1)C_7\frac{x^{2+\tau}}{n^{\tau/2}}1_k, \tag{6.13}$$

$$\hat\Sigma_n := \operatorname{Var}(\hat S_n) = \sum_{i=1}^n \frac{\mathbb{E}(\eta_i\eta_i^T e^{\lambda^T\eta_i}) - (\mathbb{E}\eta_i e^{\lambda^T\eta_i})^{\otimes 2}}{\mathbb{E}e^{\lambda^T\eta_i}} = x^2R + \Theta(1)C_8\frac{x^{2+\tau}}{n^{\tau/2}}1_k1_k^T, \tag{6.14}$$

and

$$\mathbb{E}\left\|\frac{n^{1/2}(\hat\eta_i - \mathbb{E}\hat\eta_i)}{x}\right\|_\infty^{2+\tau} \le \frac{Cn^{1+\tau/2}\mathbb{E}\|\hat\eta_i\|_\infty^{2+\tau}}{x^{2+\tau}} = \frac{Cn^{1+\tau/2}\mathbb{E}(\|\eta_i\|_\infty^{2+\tau}e^{\lambda^T\eta_i})}{x^{2+\tau}\mathbb{E}e^{\lambda^T\eta_i}} \le C_9. \tag{6.15}$$

Let $\hat F_n(t) = \mathbb{P}(\hat S_n - \hat\mu_n \le t)$ and $\hat G_n(t) = \mathbb{P}(\mathcal{N}(0, \hat\Sigma_n) \le t)$. Then, by (6.11)–(6.14), we have

$$\begin{aligned}
&\mathbb{P}\Big(\min_{1\le j\le k} S_{n,j} > \frac{x^2}{2} - \varepsilon_n\Big)\\
&\quad = \mathbb{E}e^{\lambda^T S_n}\mathbb{E}\Big(e^{-\lambda^T\hat S_n}; \min_{1\le j\le k}\hat S_{n,j} > \frac{x^2}{2} - \varepsilon_n\Big)\\
&\quad = e^{-(x^2/2)1_k^TR^{-1}1_k}\mathbb{E}\big(e^{-\lambda^T(\hat S_n - \hat\mu_n)}; \hat S_n - \hat\mu_n > \iota_n\big)\Big(1 + \Theta(1)C_{10}\frac{x^{2+\tau}}{n^{\tau/2}}\Big)\\
&\quad = e^{-(x^2/2)1_k^TR^{-1}1_k}(I + II)\Big(1 + \Theta(1)C_{10}\frac{x^{2+\tau}}{n^{\tau/2}}\Big),
\end{aligned} \tag{6.16}$$

where $\iota_n = (x^2/2 - \varepsilon_n)1_k - \hat\mu_n = \Theta(1)C_{11}(n^{-\tau/2}x^{2+\tau} + \varepsilon_n)1_k$,

$$I = \int_{t>\iota_n} e^{-\lambda^T t}\hat G_n(dt) \quad\text{and}\quad II = \int_{t>\iota_n} e^{-\lambda^T t}(\hat F_n - \hat G_n)(dt).$$

Notice that for $(2C_{12})^{1/2} \le x \le n^{\tau/(4+2\tau)}$, we have

$$\|(x^2R)^{-1}\|\|\hat\Sigma_n - x^2R\| \le C_{12}\frac{x^\tau}{n^{\tau/2}} \le \frac{1}{2},$$

and therefore by Theorem 8.1.2 in Wang, Wei and Qiao [54],

$$\begin{aligned}
\left|\frac{t^T\hat\Sigma_n^{-1}t}{2} - \frac{t^TR^{-1}t}{2x^2}\right| &\le \|\hat\Sigma_n^{-1} - (x^2R)^{-1}\|\frac{t^Tt}{2} \le \|\hat\Sigma_n^{-1} - (x^2R)^{-1}\|\|x^2R\|\frac{t^TR^{-1}t}{2x^2}\\
&\le 2\|(x^2R)^{-1}\|^2\|\hat\Sigma_n - x^2R\|\|x^2R\|\frac{t^TR^{-1}t}{2x^2} \le \Big(C_{13}\frac{x^\tau}{n^{\tau/2}}\Big)\frac{t^TR^{-1}t}{2x^2}.
\end{aligned}$$

Note also that for $t \in \mathbb{R}_+^k$,

$$\begin{aligned}
|\iota_n^T\hat\Sigma_n^{-1}t| &\le C\|(x^2R)^{-1}\|\|\iota_n\|_\infty\|t\|_1 \le \frac{C}{x^2}\Big(\frac{x^{2+\tau}}{n^{\tau/2}} + \varepsilon_n\Big)\|t\|_1\\
&\le \frac{C_{14}}{x^2}\Big(\frac{x^{2+\tau}}{n^{\tau/2}} + \varepsilon_n\Big)\lambda^Tt,
\end{aligned}$$

$$\lambda^T\iota_n + \frac{\iota_n^T\hat\Sigma_n^{-1}\iota_n}{2} \le C\|\iota_n\|_\infty \le C_{15}\Big(\frac{x^{2+\tau}}{n^{\tau/2}} + \varepsilon_n\Big),$$

and by an inequality in Problem I.6.11 of Bhatia [5],

$$\begin{aligned}\bigl|\det(\hat{\Sigma}_n)-\det(x^2R)\bigr| &\le k\bigl(\|\hat{\Sigma}_n\|\vee\|x^2R\|\bigr)^{k-1}\|\hat{\Sigma}_n-x^2R\|\\ &\le C\frac{x^{2k+\tau}}{n^{\tau/2}}\le\Bigl(C_{16}\frac{x^\tau}{n^{\tau/2}}\Bigr)\det(x^2R).\end{aligned}$$

Hence, for $C_{17}\le x\le n^{\tau/(4+2\tau)}$, we have

$$\frac{C_{13}\vee C_{14}\vee C_{15}}{x^2}\Bigl(\frac{x^{2+\tau}}{n^{\tau/2}}+\varepsilon_n\Bigr)\le\frac{1}{2},$$

and hence

$$\begin{aligned}I&=\int_{t>\iota_n}e^{-\lambda^Tt}\hat{G}_n(dt)=\frac{1}{\det^{1/2}(2\pi\hat{\Sigma}_n)}\int_{t>\iota_n}e^{-\lambda^Tt-t^T\hat{\Sigma}_n^{-1}t/2}\,dt\\ &=\frac{e^{-\lambda^T\iota_n-\iota_n^T\hat{\Sigma}_n^{-1}\iota_n/2}}{\det^{1/2}(2\pi\hat{\Sigma}_n)}\int_{\mathbb{R}_+^k}e^{-\lambda^Tt-t^T\hat{\Sigma}_n^{-1}t/2-\iota_n^T\hat{\Sigma}_n^{-1}t}\,dt\\ &=\frac{1}{\det^{1/2}(2\pi x^2R)}\int_{\mathbb{R}_+^k}e^{-1_k^TR^{-1}t-t^TR^{-1}t/2x^2}\,dt\Bigl(1+\Theta(1)C_{18}\Bigl(\frac{x^{2+\tau}}{n^{\tau/2}}+\varepsilon_n\Bigr)\Bigr)\\ &=e^{(x^2/2)1_k^TR^{-1}1_k}\mathbb{P}\Bigl(\min_{1\le j\le k}Z_j>x\Bigr)\Bigl(1+\Theta(1)C_{18}\Bigl(\frac{x^{2+\tau}}{n^{\tau/2}}+\varepsilon_n\Bigr)\Bigr),\end{aligned}\tag{6.17}$$

where both of the second and the third lines follow from a change of variables.

Next, we estimate the remainder term II in (6.16). Let $\hat{H}_n(s)=(\hat{F}_n-\hat{G}_n)(xA_s)$, where $A_s=\{t\in\mathbb{R}^k:\lambda^Tt\le s/x, t>\iota_n/x\}$. Then, by applying a non-uniform multivariate Berry–Esseen inequality in Theorem 4 of von Bahr [52] with (6.15), we have

$$\bigl|\hat{H}_n(s)\bigr|\le\frac{C_{19}}{n^{\tau/2}}\bigl(\mathcal{S}\bigl(A_s^{C_{19}n^{-1/2}}\bigr)+\mathcal{V}\bigl(A_s^{C_{19}n^{-1/2}}\bigr)\bigr),$$

where $\mathcal{S}(A)$ and $\mathcal{V}(A)$ are, respectively, the surface area and the volume of a set $A\subset\mathbb{R}^k$ and $A^\varepsilon=\bigcup_{t^Tt<1}(A+\varepsilon t)$ is the exterior parallel set of A for $\varepsilon>0$. As a k-dimensional simplex, the length of each orthogonal edge of A_s can be bounded above by

$$\max_{1\le j\le k}\Bigl(\frac{s}{\lambda_jx}-\frac{\iota_{n,j}}{x}\Bigr)\le C_{20}\frac{1+s}{x}.$$

Then, we have

$$\mathcal{S}\bigl(A_s^{C_{17}n^{-1/2}}\bigr)\le C_{21}\frac{(1+s)^{k-1}}{x^{k-1}},\qquad\mathcal{V}\bigl(A_s^{C_{17}n^{-1/2}}\bigr)\le C_{21}\frac{(1+s)^k}{x^k},$$

and therefore by Fubini's theorem and (5.11),

$$\begin{aligned}|II|&=\Bigl|\int_{t>\iota_n}e^{-\lambda^Tt}(\hat{F}_n-\hat{G}_n)(dt)\Bigr|=\Bigl|\int_{t>\iota_n}\int_{\lambda^Tt}^\infty e^{-s}\,ds(\hat{F}_n-\hat{G}_n)(dt)\Bigr|\\ &=\Bigl|\int_{\lambda^T\iota_n}^\infty\int_{xA_s}e^{-s}(\hat{F}_n-\hat{G}_n)(dt)\,ds\Bigr|=\Bigl|\int_{\lambda^T\iota_n}^\infty e^{-s}\hat{H}_n(s)\,ds\Bigr|\\ &\le\int_{\lambda^T\iota_n}^\infty e^{-s}\bigl|\hat{H}_n(s)\bigr|\,ds\le\int_{\lambda^T\iota_n}^\infty\frac{Ce^{-s}(1+s)^k}{n^{\tau/2}x^{k-1}}\,ds\\ &\le\frac{C}{n^{\tau/2}x^{k-1}}\le C_{23}\frac{x}{n^{\tau/2}}e^{(x^2/2)1_k^TR^{-1}1_k}\mathbb{P}\Bigl(\min_{1\le j\le k}Z_j>x\Bigr).\end{aligned}\tag{6.18}$$

By (6.16)–(6.18), we have (5.7) holds for $C_{24}\le x\le n^{\tau/(4+2\tau)}$. This completes the proof.

6.3. *Proof of Proposition* 5.3. Observe that $\max_{1\le j\le k} D_{n,j} \le \varepsilon_n$ if and only if $\underline{v}_n \le V_{n,j} \le \overline{v}_n$, where $\underline{v}_n = (1-\varepsilon_n^{1/2}/x)^{1/2}$ and $\overline{v}_n = (1+\varepsilon_n^{1/2}/x)^{1/2}$. The key of this proof is to show $\{V_{n,j} \notin [\underline{v}_n, \overline{v}_n]\}$ is a much rarer event than its complement and therefore makes the event $\{\max_{1\le j\le k} D_{n,j} > \varepsilon_n\}$ rare even if there is only one $V_{n,j}$ outside the interval. Indeed

$$\begin{aligned}
&\mathbb{P}\Big(\min_{1\le j\le k} T_{n,j} > x, \max_{1\le j\le k} D_{n,j} > \varepsilon_n, \max_{1\le j\le k} V_{n,j} \le M\Big) \\
&\quad\le \sum_{J=\{J_0,J_1,J_2\}\text{ is a partition of }\{1,\dots,k\}\text{ with }|J_0|\le k-1} \mathbb{P}\Big(A_J \cap \Big\{\max_{1\le j\le k} D_{n,j} > \varepsilon_n\Big\}\Big) \\
&\quad\le \max_{J=\{J_0,J_1,J_2\}\text{ is a partition of }\{1,\dots,k\}\text{ with }|J_0|\le k-1} \max_{1\le j\le k} \frac{C_1}{\varepsilon_n}\mathbb{E}(D_{n,j}; A_J),
\end{aligned} \tag{6.19}$$

where

$$A_J = \bigcap_{\ell=0}^{2}\bigcap_{j\in J_\ell}\{(W_{n,j}, V_{n,j}) \in \mathcal{E}_{j,\ell}\},$$

and

$$\begin{aligned}
\mathcal{E}_{j,0} &= \{(u,v)\in\mathbb{R}\times\mathbb{R}_+ : u/v > x, \underline{v}_n \le v \le \overline{v}_n\}, \\
\mathcal{E}_{j,1} &= \{(u,v)\in\mathbb{R}\times\mathbb{R}_+ : u/v > x, \overline{v}_n < v \le M\}, \\
\mathcal{E}_{j,2} &= \{(u,v)\in\mathbb{R}\times\mathbb{R}_+ : u/v > x, 0 < v < \underline{v}_n\}.
\end{aligned}$$

Let $\lambda = (\lambda_1,\dots,\lambda_k)^T$ and $\theta = (\theta_1,\dots,\theta_k)^T$, where

$$\lambda_j = (R^{-1}1_k)_j \quad\text{and}\quad \theta_j = \frac{\lambda_j}{2}\Big(\mathbb{1}_{j\in J_0} + \frac{1}{M}\mathbb{1}_{j\in J_1} + 4\mathbb{1}_{j\in J_2}\Big).$$

Since

$$\sum_{i=1}^{n}\lambda^T\xi_i = \sum_{j=1}^{k}\lambda_j W_{n,j} x \quad\text{and}\quad \sum_{i=1}^{n}\theta^T\xi_i^{\circ 2} = \sum_{j=1}^{k}\theta_j V_{n,j}^2 x^2,$$

by Markov's inequality, we have

$$\mathbb{E}(D_{n,j}; A_J) \le \frac{\mathbb{E}(D_{n,j}e^{\sum_{i=1}^n(\lambda^T\xi_i - \theta^T\xi_i^{\circ 2})})}{e^{\sum_{\ell=0}^2\sum_{j\in J_\ell}\inf_{(u,v)\in\mathcal{E}_{j,\ell}}(\lambda_j ux - \theta_j v^2x^2)}}. \tag{6.20}$$

Let $\zeta_{ij} = \xi_{ij}^2 - \mathbb{E}\xi_{ij}^2$ so that $\sum_{i=1}^n \zeta_{ij} = x^2(V_{n,j}^2 - 1)$. Then by (5.5) and a minor modification as that in proving Lemmas 6.1 and 6.3, we have for $(2C_2)^{1/2} \le x \le n^{\tau/(4+2\tau)}$,

$$\mathbb{E}e^{\lambda^T\xi_i - \theta^T\xi_i^{\circ 2}} = 1 + \Theta(1)C_2\frac{x^2}{n} \in \Big[\frac{1}{2}, \frac{3}{2}\Big],$$

and therefore

$$\begin{aligned}
&\frac{\mathbb{E}(x^2 D_{n,j}e^{\sum_{i=1}^n(\lambda^T\xi_i - \theta^T\xi_i^{\circ 2})})}{\mathbb{E}e^{\sum_{i=1}^n(\lambda^T\xi_i - \theta^T\xi_i^{\circ 2})}} \\
&\quad= \frac{\mathbb{E}((\sum_{i=1}^n\zeta_{ij})^2 e^{\sum_{i=1}^n(\lambda^T\xi_i - \theta^T\xi_i^{\circ 2})})}{\mathbb{E}e^{\sum_{i=1}^n(\lambda^T\xi_i - \theta^T\xi_i^{\circ 2})}} \\
&\quad= \sum_{i=1}^{n}\frac{\mathbb{E}(\zeta_{ij}^2 e^{\lambda^T\xi_i - \theta^T\xi_i^{\circ 2}})}{\mathbb{E}e^{\lambda^T\xi_i - \theta^T\xi_i^{\circ 2}}} + \sum_{i_1\ne i_2}\frac{\mathbb{E}(\zeta_{i_1 j}e^{\lambda^T\xi_{i_1} - \theta^T\xi_{i_1}^{\circ 2}})\mathbb{E}(\zeta_{i_2 j}e^{\lambda^T\xi_{i_2} - \theta^T\xi_{i_2}^{\circ 2}})}{\mathbb{E}e^{\lambda^T\xi_{i_1} - \theta^T\xi_{i_1}^{\circ 2}}\mathbb{E}e^{\lambda^T\xi_{i_2} - \theta^T\xi_{i_2}^{\circ 2}}} \\
&\quad\le C\Big(\frac{x^{2+\tau}}{n^{\tau/2}} + \frac{x^{4+2\tau}}{n^{\tau}}\Big) \le C_3\frac{x^{2+\tau}}{n^{\tau/2}},
\end{aligned} \tag{6.21}$$

and

$$\mathbb{E}e^{\sum_{i=1}^n(\lambda^T\xi_i-\theta^T\xi_i^{\circ 2})}=e^{\sum_{j=1}^k(\lambda_j x^2/2-\theta_j x^2)+\Theta(1)C_4 n^{-\tau/2}x^{2+\tau}}. \tag{6.22}$$

Consequently, by combining (6.20)–(6.22), we have

$$\mathbb{E}(D_{n,j};A_J)\le C_5\frac{x^{2+\tau}}{n^{\tau/2}}e^{\sum_{\ell=0}^2\sum_{j\in J_\ell}(\lambda_j x^2/2-\theta_j x^2-\inf_{(u,v)\in\mathcal{E}_{j,\ell}}(\lambda_j ux-\theta_j v^2x^2))}. \tag{6.23}$$

Since for $j\in J_\ell$,

$$\inf_{(u,v)\in\mathcal{E}_{j,\ell}}(\lambda_j ux-\theta_j v^2x^2)=\begin{cases}\lambda_j x^2(\underline{v}_n-\underline{v}_n^2/2) & \ell=0,\\ \lambda_j x^2(\overline{v}_n-\overline{v}_n^2/2M) & \ell=1,\\ \lambda_j x^2(\underline{v}_n-2\underline{v}_n^2) & \ell=2,\end{cases}$$

we have

$$\frac{\lambda_j x^2}{2}-\theta_j x^2-\inf_{(u,v)\in\mathcal{E}_{j,\ell}}(\lambda_j ux-\theta_j v^2x^2)\le\begin{cases}-\lambda_j x^2/2+C_6 & \ell=0,\\ -\lambda_j x^2/2-x/C_7 & \ell=1,\\ -\lambda_j x^2/2-x/C_8 & \ell=2,\end{cases}$$

and therefore by Remark 4.1 and the fact that $|J_0|\le k-1$ in (6.19),

$$\begin{aligned}&\sum_{\ell=0}^2\sum_{j\in J_\ell}\Big(\frac{\lambda_j x^2}{2}-\theta_j x^2-\inf_{(u,v)\in\mathcal{E}_{j,\ell}}(\lambda_j ux-\theta_j v^2x^2)\Big)\\ &\quad\le-\frac{x^2}{2}1_k^TR^{-1}1_k+(k-1)C_6-\frac{x}{C_7\vee C_8}.\end{aligned} \tag{6.24}$$

Combining (5.11), (6.19) and (6.23)–(6.24), we have for $C_9\le x\le n^{\tau/(4+2\tau)}$,

$$\begin{aligned}&\frac{\mathbb{P}(\min_{1\le j\le k}T_{n,j}>x,\max_{1\le j\le k}D_{n,j}>\varepsilon_n,\max_{1\le j\le k}V_{n,j}\le M)}{\mathbb{P}(\min_{1\le j\le k}Z_j>x)}\\ &\quad\le C\frac{x^{2+\tau}e^{-(x^2/2)1_k^TR^{-1}1_k-x/(C_7\vee C_8)}}{\varepsilon_n n^{\tau/2}\mathbb{P}(\min_{1\le j\le k}Z_j>x)}\le C\frac{x^{k+2+\tau}e^{-x/(C_7\vee C_8)}}{\varepsilon_n n^{\tau/2}}\le\frac{C_{10}}{\varepsilon_n n^{\tau/2}},\end{aligned}$$

as desired.

6.4. *Proof of Proposition* 5.4. Observe that

$$\begin{aligned}\mathbb{P}\Big(\min_{1\le j\le k}T_{n,j}>x,\max_{1\le j\le k}V_{n,j}>M\Big)&\le\sum_{j=1}^k\mathbb{P}\Big(\min_{1\le j\le k}T_{n,j}>x,V_{n,j}>M\Big)\\ &\le\sum_{j=1}^k\mathbb{P}(T_{n,j}>x,V_{n,j}>M).\end{aligned} \tag{6.25}$$

The key of this proof is to apply a similar truncation argument as that in proving Proposition 5.2 of Shao and Zhou [49] to bound each term in the right-hand side of (6.25).

Let $m\ge 1$ be a constant to be specified later depending only on M. Define

$$\tilde{W}_{n,j}=\frac{\sum_{i=1}^n X_{ij}\mathbb{1}(\xi_{ij}\le m)}{(\sum_{i=1}^n\mathbb{E}X_{ij}^2)^{1/2}}=:\sum_{i=1}^n\tilde{X}_{ij}\quad\text{and}\quad\bar{V}_{n,j}^2=\frac{\sum_{i=1}^n X_{ij}^2\mathbb{1}(|\xi_{ij}|\le 1)}{\sum_{i=1}^n\mathbb{E}X_{ij}^2}=:\sum_{i=1}^n\bar{X}_{ij}^2.$$

Then, we have

$$
\begin{aligned}
(6.26)\qquad & \mathbb{P}(T_{n,j} > x, V_{n,j} > M) \\
&= \mathbb{P}(W_{n,j} > xV_{n,j}, V_{n,j}^2 > M^2) \\
&\le \mathbb{P}(\tilde{W}_{n,j} > Mx/2, \bar{V}_{n,j}^2 > 3) + \mathbb{P}(\tilde{W}_{n,j} > Mx/2, V_{n,j}^2 - \bar{V}_{n,j}^2 > M^2 - 3) \\
&\quad + \mathbb{P}(W_{n,j} - \tilde{W}_{n,j} > xV_{n,j}/2).
\end{aligned}
$$

Next, by the facts that

$$\mathbb{E}(\xi_{ij}; \xi_{ij} \le m) = -\mathbb{E}(\xi_{ij}; \xi_{ij} > m) \le 0,$$

and for $s \in \mathbb{R}$,

$$e^s \le 1 + s + \frac{s^2}{2} + \frac{(s \vee 0)^3 e^s}{6},$$

we have

$$
\begin{aligned}
\mathbb{E}e^{(Mx/2)\tilde{W}_{n,j}} &= \prod_{i=1}^n \mathbb{E}e^{(M/2)\xi_{ij}\mathbb{1}(\xi_{ij}\le m)} \\
&\le \prod_{i=1}^n \left(1 + \frac{M^2}{8}\mathbb{E}\xi_{ij}^2 + C_1 \frac{x^{2+\tau}}{n^{1+\tau/2}}\right) \le e^{M^2x^2/8 + C_1 n^{-\tau/2} x^{2+\tau}}.
\end{aligned}
$$

Note also that

$$x^2 \mathbb{E}(V_{n,j}^2 - \bar{V}_{n,j}^2) = \sum_{i=1}^n \mathbb{E}(|\xi_{ij}|^2; |\xi_{ij}| > 1) \le \sum_{i=1}^n \mathbb{E}|\xi_{ij}|^{2+\tau} \le \frac{x^{2+\tau}}{n^{\tau/2}}.$$

Hence, by Markov's inequality and an argument as that in deriving (6.21),

$$\mathbb{P}(\tilde{W}_{n,j} > Mx/2, \bar{V}_{n,j}^2 > 3) \le \frac{C_2 x^{\tau-2}}{n^{\tau/2} e^{M^2x^2/8}}. \qquad (6.27)$$

Indeed

$$
\begin{aligned}
& \mathbb{P}(\tilde{W}_{n,j} > Mx/2, \bar{V}_{n,j}^2 > 3) \\
&\quad \le \frac{\mathbb{E}((\bar{V}_{n,j}^2 - 1)^2 e^{(Mx/2)\tilde{W}_{n,j}})}{4e^{M^2x^2/4}} \\
&\quad \le \frac{\mathbb{E}((\sum_{i=1}^n (\bar{X}_{ij}^2 - \mathbb{E}\bar{X}_{ij}^2))^2 e^{(Mx/2)\tilde{W}_{n,j}}) + (\mathbb{E}\bar{V}_{n,j}^2 - 1)^2 \mathbb{E}e^{(Mx/2)\tilde{W}_{n,j}}}{2e^{M^2x^2/4}} \\
&\quad =: \frac{\mathbb{E}((\sum_{i=1}^n \bar{\zeta}_{ij})^2 e^{(Mx/2)\tilde{W}_{n,j}}) + x^4(\mathbb{E}\bar{V}_{n,j}^2 - \mathbb{E}V_{n,j}^2)^2 \mathbb{E}e^{(Mx/2)\tilde{W}_{n,j}}}{2x^4 e^{M^2x^2/4}} \\
&\quad \le \frac{C}{x^4 e^{M^2x^2/4}} \left(\frac{x^{2+\tau}}{n^{\tau/2}} e^{M^2x^2/8} + \frac{x^{4+2\tau}}{n^\tau} e^{M^2x^2/8}\right) \le \frac{C_2 x^{\tau-2}}{n^{\tau/2} e^{M^2x^2/8}}.
\end{aligned}
$$

Notice that by letting $\tilde{W}_{n,j}^{(i)} = \tilde{W}_{n,j} - \tilde{X}_{ij}$ and employing a similar argument,

$$
\begin{aligned}
(6.28)\qquad & \mathbb{P}(\tilde{W}_{n,j} > Mx/2, V_{n,j}^2 - \bar{V}_{n,j}^2 > M^2 - 3) \\
&\quad \le \frac{\sum_{i=1}^n \mathbb{E}(X_{ij}^2 e^{(Mx/2)\tilde{X}_{ij}}; |\xi_{ij}| > 1)\mathbb{E}e^{(Mx/2)\tilde{W}_{n,j}^{(i)}}}{(M^2 - 3)e^{M^2x^2/4} \sum_{i=1}^n \mathbb{E}X_{ij}^2} \le C_3 \frac{x^\tau}{n^{\tau/2} e^{M^2x^2/8}}.
\end{aligned}
$$

Next, by Cauchy–Schwarz inequality and Markov's inequality, we have for $s > 0$,

$$\begin{aligned}
&\mathbb{P}(W_{n,j} - \tilde{W}_{n,j} > xV_{n,j}/2)\\
&= \mathbb{P}\biggl(\frac{\sum_{i=1}^{n} X_{ij}\mathbb{1}(\xi_{ij} > m)}{(\sum_{i=1}^{n}\mathbb{E}X_{ij}^2)^{1/2}} > \frac{xV_{n,j}}{2}\biggr)\\
&\le \mathbb{P}\biggl(\frac{(\sum_{i=1}^{n} X_{ij}^2)^{1/2}(\sum_{i=1}^{n}\mathbb{1}(\xi_{ij} > m))^{1/2}}{(\sum_{i=1}^{n}\mathbb{E}X_{ij}^2)^{1/2}} > \frac{xV_{n,j}}{2}\biggr)\\
&\le \mathbb{P}\Biggl(\sum_{i=1}^{n} s\mathbb{1}(\xi_{ij} > m) > \frac{sx^2}{4}\Biggr) \le \sum_{i=1}^{n}\frac{4\mathbb{E}(e^{s\mathbb{1}(\xi_{ij}>m)}; \xi_{ij} > m)\prod_{i_1\ne i}\mathbb{E}e^{s\mathbb{1}(\xi_{i_1 j}>m)}}{x^2 e^{sx^2/4}}\\
&\le \sum_{i=1}^{n}\frac{4e^{s}\mathbb{P}(\xi_{ij} > m)\prod_{i_1\ne i}(1 + e^{s}\mathbb{P}(\xi_{i_1 j} > m))}{x^2 e^{sx^2/4}} \le \frac{4e^{s+e^{s}x^2/m^2}x^{\tau}}{m^2 n^{\tau/2}e^{sx^2/4}}.
\end{aligned}$$

Consequently, by taking $s = 2\log(m/2) = 1 + M^2/2$, we have

$$\mathbb{P}(W_{n,j} - \tilde{W}_{n,j} > xV_{n,j}/2) \le \frac{x^{\tau}}{n^{\tau/2}e^{M^2x^2/8}}, \tag{6.29}$$

and therefore by combining (5.11) and (6.25)–(6.29), we have for $C_4 \le x \le n^{\tau/(4+2\tau)}$,

$$\begin{aligned}
&\frac{\mathbb{P}(\min_{1\le j\le k} T_{n,j} > x, \max_{1\le j\le k} V_{n,j} > M)}{\mathbb{P}(\min_{1\le j\le k} Z_j > x)}\\
&\quad\le C\frac{x^{\tau}}{n^{\tau/2}e^{M^2x^2/8}\mathbb{P}(\min_{1\le j\le k} Z_j > x)} \le C_5\frac{e^{(k/2c_1)x^2}x^{k+\tau}}{n^{\tau/2}e^{M^2x^2/8}}.
\end{aligned}$$

By taking $M = 4(k/c_1)^{1/2}$, we complete the proof.

Acknowledgments. The authors sincerely thank the Editor, Associate Editor and anonymous referees for their constructive comments that led to substantial improvement of the paper.

Qi-Man Shao is the corresponding author.

Funding. Chen was partially supported by National Natural Science Foundation of China Grants 12292983 and 12071013.

Shao was partially supported by National Natural Science Foundation of China NSFC 12031005 and Shenzhen Outstanding Talents Training Fund, China.

SUPPLEMENTARY MATERIAL

Supplement to "Self-normalized Cramér type moderate deviation theorem for Gaussian approximation" (DOI: 10.1214/25-AOS2507SUPP; .pdf). The proofs of Proposition 5.5 and Lemmas 6.1–6.3 can be found in the Supplementary Material [39].

REFERENCES

[1] ALESHKYAVICHENE, A. K. (1983). Multidimensional integral limit theorems for probabilities of large deviations. *Teor. Veroyatn. Primen.* **28** 62–82. MR0691468

[2] AVANESOV, V. and BUZUN, N. (2018). Change-point detection in high-dimensional covariance structure. *Electron. J. Stat.* **12** 3254–3294. MR3861282 https://doi.org/10.1214/18-EJS1484

[3] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and WEI, Y. (2018). Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework. *Ann. Statist.* **46** 3643–3675. MR3852664 https://doi.org/10.1214/17-AOS1671

[4] BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika* **102** 77–94. MR3335097 https://doi.org/10.1093/biomet/asu056

[5] BHATIA, R. (1997). *Matrix Analysis*. *Graduate Texts in Mathematics* **169**. Springer, New York. MR1477662 https://doi.org/10.1007/978-1-4612-0653-8

[6] CAI, T. T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870 https://doi.org/10.1111/rssb.12034

[7] CHANG, J., ZHENG, C., ZHOU, W.-X. and ZHOU, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73** 1300–1310. MR3744543 https://doi.org/10.1111/biom.12695

[8] CHEN, L., WANG, W. and WU, W. B. (2022). Inference of breakpoints in high-dimensional time series. *J. Amer. Statist. Assoc.* **117** 1951–1963. MR4528482 https://doi.org/10.1080/01621459.2021.1893178

[9] CHEN, P., GAO, C. and ZHANG, A. Y. (2022). Partial recovery for top-k ranking: Optimality of MLE and suboptimality of the spectral method. *Ann. Statist.* **50** 1618–1652. MR4441134 https://doi.org/10.1214/21-aos2166

[10] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

[11] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

[12] CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* **33** 2374–2425. MR4583674 https://doi.org/10.1214/22-aap1870

[13] CHERNOZHUKOV, V., HÄRDLE, W. K., HUANG, C. and WANG, W. (2021). LASSO-driven inference in time and space. *Ann. Statist.* **49** 1702–1735. MR4298878 https://doi.org/10.1214/20-aos2019

[14] CHERNOZHUOKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *Ann. Statist.* **50** 2562–2586. MR4500619 https://doi.org/10.1214/22-aos2193

[15] CHETVERIKOV, D., LARSEN, B. and PALMER, C. (2016). IV quantile regression for group-level treatments, with an application to the distributional effects of trade. *Econometrica* **84** 809–833. MR3481380 https://doi.org/10.3982/ECTA12121

[16] CHISTYAKOV, G. P. and GÖTZE, F. (2003). On bounds for moderate deviations for Student's statistic. *Teor. Veroyatn. Primen.* **48** 609–615. MR2141355 https://doi.org/10.1137/S0040585X97980610

[17] CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actual. Sci. Ind.* **736** 5–23.

[18] DENG, H. and ZHANG, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48** 3643–3671. MR4185823 https://doi.org/10.1214/20-AOS1946

[19] DETTE, H. and GÖSMANN, J. (2018). Relevant change points in high dimensional time series. *Electron. J. Stat.* **12** 2578–2636. MR3849896 https://doi.org/10.1214/18-EJS1464

[20] FAN, J., LOU, Z. and YU, M. (2024). Are latent factor regression and sparse regression adequate? *J. Amer. Statist. Assoc.* **119** 1076–1088. MR4765971 https://doi.org/10.1080/01621459.2023.2169700

[21] FAN, J., MASINI, R. P. and MEDEIROS, M. C. (2023). Bridging factor and sparse models. *Ann. Statist.* **51** 1692–1717. MR4658573 https://doi.org/10.1214/23-aos2304

[22] FANG, X. and KOIKE, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31** 1660–1686. MR4312842 https://doi.org/10.1214/20-aap1629

[23] GÖSMANN, J., STOEHR, C., HEINY, J. and DETTE, H. (2022). Sequential change point detection in high dimensional time series. *Electron. J. Stat.* **16** 3608–3671. MR4444665 https://doi.org/10.1214/22-ejs2027

[24] GUO, X. and TANG, C. Y. (2021). Specification tests for covariance structures in high-dimensional statistical models. *Biometrika* **108** 335–351. MR4259135 https://doi.org/10.1093/biomet/asaa073

[25] HASHORVA, E. and HÜSLER, J. (2003). On multivariate Gaussian tails. *Ann. Inst. Statist. Math.* **55** 507–522. MR2007795 https://doi.org/10.1007/BF02517804

[26] JING, B.-Y., SHAO, Q.-M. and WANG, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann. Probab.* **31** 2167–2215. MR2016616 https://doi.org/10.1214/aop/1068646382

[27] KIM, B., LIU, S. and KOLAR, M. (2021). Two-sample inference for high-dimensional Markov networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 939–962. MR4349123 https://doi.org/10.1111/rssb.12446

[28] KLAASSEN, S., KUECK, J., SPINDLER, M. and CHERNOZHUKOV, V. (2023). Uniform inference in high-dimensional Gaussian graphical models. *Biometrika* **110** 51–68. MR4565443 https://doi.org/10.1093/biomet/asac030

[29] KUCHIBHOTLA, A. K., MUKHERJEE, S. and BANERJEE, D. (2021). High-dimensional CLT: Improvements, non-uniform extensions and large deviations. *Bernoulli* **27** 192–217. MR4177366 https://doi.org/10.3150/20-BEJ1233

[30] KUCHIBHOTLA, A. K. and RINALDO, A. (2020). High-dimensional CLT for sums of non-degenerate random vectors: $n^{-1/2}$-rate. Available at arXiv:2009.13673.

[31] LEUNG, D. and DRTON, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.* **46** 280–307. MR3766953 https://doi.org/10.1214/17-AOS1550

[32] LI, W. V. and SHAO, Q.-M. (2002). A normal comparison inequality and its applications. *Probab. Theory Related Fields* **122** 494–508. MR1902188 https://doi.org/10.1007/s004400100176

[33] LINNIK, J. V. (1961). On the probability of large deviations for the sums of independent variables. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. II* 289–306. Univ. California Press, Berkeley, CA. MR0137142

[34] LIU, Y. and XIE, J. (2019). Accurate and efficient P-value calculation via Gaussian approximation: A novel Monte-Carlo method. *J. Amer. Statist. Assoc.* **114** 384–392. MR3941262 https://doi.org/10.1080/01621459.2017.1407776

[35] LOPES, M. E. (2022). Central limit theorem and bootstrap approximation in high dimensions: Near $1/\sqrt{n}$ rates via implicit smoothing. *Ann. Statist.* **50** 2492–2513. MR4505371 https://doi.org/10.1214/22-aos2184

[36] LU, J., KOLAR, M. and LIU, H. (2020). Kernel meets sieve: Post-regularization confidence bands for sparse additive model. *J. Amer. Statist. Assoc.* **115** 2084–2099. MR4189778 https://doi.org/10.1080/01621459.2019.1689984

[37] NEYKOV, M., LU, J. and LIU, H. (2019). Combinatorial inference for graphical models. *Ann. Statist.* **47** 795–827. MR3909951 https://doi.org/10.1214/17-AOS1650

[38] PETROV, V. V. (1954). Generalization of Cramér's limit theorem. *Uspekhi Mat. Nauk* **9** 195–202. MR0065058

[39] QIU, J., CHEN, S. X. and SHAO, Q.-M. (2025). Supplement to "Self-normalized Cramér type moderate deviation theorem for Gaussian approximation." https://doi.org/10.1214/25-AOS2507SUPP

[40] RINALDO, A., WASSERMAN, L. and G'SELL, M. (2019). Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *Ann. Statist.* **47** 3438–3469. MR4025748 https://doi.org/10.1214/18-AOS1784

[41] RUDZKIS, R. and BAKSHAEV, A. (2017). General theorems on large deviations for random vectors. *Lith. Math. J.* **57** 367–390. MR3685166 https://doi.org/10.1007/s10986-017-9367-y

[42] SAULIS, L. and STATULEVIČIUS, V. A. (1991). *Limit Theorems for Large Deviations. Mathematics and Its Applications* (*Soviet Series*) **73**. Kluwer Academic, Dordrecht. Translated and revised from the 1989 Russian original. MR1171883 https://doi.org/10.1007/978-94-011-3530-6

[43] SAULIS, L. I. (1987). General lemmas on large deviations for a random vector with regular cumulant behavior. I. *Liet. Mat. Rink.* **27** 535–549. MR0925358

[44] SAULIS, L. I. (1987). General lemmas on large deviations for a random vector with regular cumulant behavior. II. *Liet. Mat. Rink.* **27** 747–758. MR0928543

[45] SAULIS, L. I. (1988). General lemmas on large deviations for a random vector with regular cumulant behavior. III. *Liet. Mat. Rink.* **28** 99–111. MR0949646 https://doi.org/10.1007/BF00972252

[46] SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. MR4124333 https://doi.org/10.1214/19-AOS1857

[47] SHANG, Z. and CHENG, G. (2015). Nonparametric inference in generalized functional linear models. *Ann. Statist.* **43** 1742–1773. MR3357877 https://doi.org/10.1214/15-AOS1322

[48] SHAO, Q.-M. (1999). A Cramér type large deviation result for Student's t-statistic. *J. Theoret. Probab.* **12** 385–398. MR1684750 https://doi.org/10.1023/A:1021626127372

[49] SHAO, Q.-M. and ZHOU, W.-X. (2016). Cramér type moderate deviation theorems for self-normalized processes. *Bernoulli* **22** 2029–2079. MR3498022 https://doi.org/10.3150/15-BEJ719

[50] SHI, C. and LI, L. (2022). Testing mediation effects using logic of Boolean matrices. *J. Amer. Statist. Assoc.* **117** 2014–2027. MR4528486 https://doi.org/10.1080/01621459.2021.1895177

[51] VON BAHR, B. (1967). Multi-dimensional integral limit theorems for large deviations. *Ark. Mat.* **7** 89–99. MR0224129 https://doi.org/10.1007/BF02591679

[52] VON BAHR, B. (1967). Multi-dimensional integral limit theorems. *Ark. Mat.* **7** 71–88. MR0228044 https://doi.org/10.1007/BF02591678

[53] WANG, G. and FENG, L. (2023). Computationally efficient and data-adaptive changepoint inference in high dimension. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 936–958. MR4735036 https://doi.org/10.1093/jrsssb/qkad048

[54] WANG, G., WEI, Y. and QIAO, S. (2018). *Generalized Inverses: Theory and Computations*, 2nd ed. *Developments in Mathematics* **53**. Springer, Singapore. MR3793648 https://doi.org/10.1007/978-981-13-0146-9

[55] XUE, K. and YAO, F. (2020). Distribution and correlation-free two-sample test of high-dimensional means. *Ann. Statist.* **48** 1304–1328. MR4124324 https://doi.org/10.1214/19-AOS1848

[56] YU, M. and CHEN, X. (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 247–270. MR4250275 https://doi.org/10.1111/rssb.12406

[57] YU, X., LI, D. and XUE, L. (2024). Fisher's combined probability test for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **119** 511–524. MR4713910 https://doi.org/10.1080/01621459.2022.2126781

[58] ZHANG, X. and CHENG, G. (2017). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.* **112** 757–768. MR3671768 https://doi.org/10.1080/01621459.2016.1166114

[59] ZHANG, Y. and POLITIS, D. N. (2022). Ridge regression revisited: Debiasing, thresholding and bootstrap. *Ann. Statist.* **50** 1401–1422. MR4441125 https://doi.org/10.1214/21-aos2156