TESTING FOR HIGH-DIMENSIONAL WHITE NOISE

Long Feng[1], Binghui Liu[2,*] and Yanyuan Ma[3]

[1]*Nankai University,* [2]*Northeast Normal University and* [2]*Pennsylvania State University*

Supplementary Material

Supplementary Material presents the technical details of Remark 2, some additional simulation results and the technical proofs.

S1. Technical details of Remark 2

We consider a special case to investigate the power function of the max-type test when $c_0 \in [1, 3]$. Let $\varepsilon_{t1} = z_{t1} + \rho z_{t-1,1}$, where $z_{t1} \sim \mathcal{N}(0, 1)$ and $\rho = O(\sqrt{\log p/n})$. For $i \in \{2, \cdots, p\}$, ε_{ti}'s are all i.i.d. from $\mathcal{N}(0, 1)$. $\{\varepsilon_{t1}\}_{t=1,\cdots,n}$ are independent of $\{\varepsilon_{ti}\}_{t=1,\cdots,n},\ i = 2, \cdots, p$. By the Central Limit Theorem, $\sqrt{n-k}\hat{\rho}_{ij}(k) \xrightarrow{d} \mathcal{N}(0, 1)$ for $k > 1$; $\sqrt{n-1}\hat{\rho}_{ij}(1) \xrightarrow{d} \mathcal{N}(0, 1)$ for $i \neq j$ and $\frac{\sqrt{n-1}}{\sqrt{1-\frac{3\rho^2+2\rho^4+3\rho^6}{(1+\rho^2)^4}}}\{\hat{\rho}_{11}(1) - \frac{\rho}{1+\rho^2}\} \xrightarrow{d} \mathcal{N}(0, 1)$ for $k = 1$. Define $x_\alpha = 2\log(Kp^2) -$

*Corresponding author.

$\log\log(Kp^2) + q_\alpha$. Define $\mathcal{A} = \{(i,j,k) | 1 \le i,j \le p, 1 \le k \le K\}$. Thus,

$$
\begin{aligned}
&\mathrm{P}\left\{T_{\mathrm{MAX}}^2 - 2\log(Kp^2) + \log\log(Kp^2) \ge q_\alpha\right\}\\
=&\mathrm{P}\left\{\max_{1\le i,j\le p, 1\le k\le K}(n-k)\hat{\rho}_{ij}^2(k) \ge x_\alpha\right\} \ge \mathrm{P}\left\{(n-1)\hat{\rho}_{11}^2(1) \ge x_\alpha\right\}\\
=&\mathrm{P}\left\{\left|\mathcal{N}\left(\frac{\sqrt{n}\rho}{1+\rho^2}, 1 - \frac{3\rho^2+2\rho^4+3\rho^6}{(1+\rho^2)^4}\right)\right| \ge \sqrt{x_\alpha}\right\} + o(1)\\
=&\mathrm{P}\left\{\left|\mathcal{N}\left(\sqrt{n}\rho, 1\right)\right| \ge \sqrt{x_\alpha}\right\} + o(1)\\
=&\Phi(\sqrt{n}\rho - \sqrt{x_\alpha}) + \Phi(-\sqrt{n}\rho - \sqrt{x_\alpha}) + o(1)
\end{aligned}
$$

for sufficiently small ρ and diverging p, where $\mathcal{N}(\mu, \sigma^2)$ denotes a random variable that follows the normal distribution with mean μ and variance σ^2. On the other hand,

$$
\begin{aligned}
&\mathrm{P}\left\{T_{\mathrm{MAX}}^2 - 2\log(Kp^2) + \log\log(Kp^2) \ge q_\alpha\right\}\\
=&\mathrm{P}\left\{\max_{1\le i,j\le p, 1\le k\le K}(n-k)\hat{\rho}_{ij}^2(k) \ge x_\alpha\right\}\\
\le&\mathrm{P}\left\{(n-1)\hat{\rho}_{11}^2(1) \ge x_\alpha\right\} + \mathrm{P}\left\{\max_{(i,j,k)\in\mathcal{A}/(1,1,1)}(n-k)\hat{\rho}_{ij}^2(k) \ge x_\alpha\right\}\\
=&\Phi(\sqrt{n}\rho - \sqrt{x_\alpha}) + \Phi(-\sqrt{n}\rho - \sqrt{x_\alpha}) + \alpha + o(1).
\end{aligned}
$$

In the last equality, we used the following derivation. By the proof of Theorem 1 in Chang et al. (2018), we have

$$\mathrm{P}\left\{\max_{(i,j,k)\in\mathcal{A}/(1,1,1)}(n-k)\hat{\rho}_{ij}^2(k)\geq x_\alpha\right\}-\mathrm{P}\left(\max_{1\leq l\leq Kp^2-1}\xi_l^2\geq x_\alpha\right)\to 0,$$

where $(\xi_1,\cdots,\xi_{Kp^2-1})$ follows the multivariate normal distribution with zero mean and the same correlation matrix of $\{(n-k)\hat{\rho}_{ij}^2(k)\}_{(i,j,k)\in\mathcal{A}/(1,1,1)}$. After some calculations, we have $\mathrm{cor}\left\{(n-k)\hat{\rho}_{11}(k),(n-k-1)\hat{\rho}_{11}(k+1)\right\}\to 2\rho$, $\mathrm{cor}\left\{(n-k)\hat{\rho}_{11}(k),(n-k-2)\hat{\rho}_{11}(k+2)\right\}\to\frac{\rho^2}{1+\rho^2}$ and the other correlations between $(n-k)\hat{\rho}_{ij}^2(k)$ are all zeros. By Theorem 1 in the supplementary and condition $\rho=O(\sqrt{\log p/n})$, we have $\mathrm{P}\left(\max_{1\leq l\leq Kp^2-1}\xi_l^2\geq x_\alpha\right)\to\alpha$.

Hence, the power function of the max-type test is

$$\begin{aligned}\lim_{n,p\to\infty}\Phi(\sqrt{n}\rho-\sqrt{x_\alpha})+\lim_{n,p\to\infty}\Phi(-\sqrt{n}\rho-\sqrt{x_\alpha})\leq\beta_{\mathrm{MAX}}(\rho)\\ \leq\alpha+\lim_{n,p\to\infty}\Phi(\sqrt{n}\rho-\sqrt{x_\alpha})+\lim_{n,p\to\infty}\Phi(-\sqrt{n}\rho-\sqrt{x_\alpha}).\end{aligned}$$

Note that $x_\alpha\sim 2\sqrt{\log p}$. When $\rho=c_0\sqrt{\log p/n}$, we have: (1) if $0<c_0<2$, then $c_0\sqrt{\log p}-\sqrt{x_\alpha}\to-\infty$, $-c_0\sqrt{\log p}-\sqrt{x_\alpha}\to-\infty$ and $\beta_{\mathrm{MAX}}(\rho)\in(0,\alpha)$; (2) if $c_0>2$, then $c_0\sqrt{\log p}-\sqrt{x_\alpha}\to\infty$, $-c_0\sqrt{\log p}-\sqrt{x_\alpha}\to-\infty$ and $\beta_{\mathrm{MAX}}(\rho)=1$. On the other hand, when $\sqrt{n}\rho=\sqrt{4\log p+c_1\sqrt{\log p}}$, we have $\sqrt{n}\rho-\sqrt{x_\alpha}\to c_1/4$ and $-\sqrt{n}\rho-\sqrt{x_\alpha}\to-\infty$, then $\beta_{\mathrm{MAX}}(\rho)\in$

$(\Phi(c_1/4), \alpha + \Phi(c_1/4))$. The above analysis indicates that when $c_0 \in [1, 3]$, different power result can occur depending on the specific alternative situation and no universal power conclusion can be drawn.

S2. Some additional simulation results

S2.1 Results under a polynomial-type tail distribution

Here, we consider the situation where the distribution of z_{it} has a polynomial-type tail, named distribution (iii). Specifically, we set $z_{it} \overset{i.i.d}{\sim} t(8)/\sqrt{4/3}$ for $i \in \{1, \cdots, p\}$ and $t \in \{1, \cdots, n\}$, keeping all other simulation parameters consistent with those outlined in Section 3 of the main text. The empirical sizes of the tests MAX, SUM, LY, and FC are reported in Table 1. The power curves of these four tests are illustrated in Figures 1 and 2, respectively. These results are very similar to those presented in Section 3 of the main text, which demonstrate that the tests proposed in this paper has a certain robustness for distributions with different types of tails.

S2.2 Results when K is relatively large

In this subsection, we explore the performance of the proposed tests when the value of K is relatively large. We set $K = 10$, keeping all other simulation parameters identical to those outlined in Section 3 of the main text. The empir-

Table 1: Size performance in the case of $\boldsymbol{\varepsilon}_t = \mathbf{A}\boldsymbol{z}_t$ with distribution (iii): $z_{it} \overset{i.i.d}{\sim} t(8)/\sqrt{4/3}$.

		$K=1$				$K=2$				$K=3$			
n	p	MAX	LY	SUM	FC	MAX	LY	SUM	FC	MAX	LY	SUM	FC
		Setting (I)											
100	30	1.0(0.3)	4.2(0.6)	5.2(0.7)	4.3(0.6)	1.5(0.4)	3.6(0.6)	5.7(0.7)	3.9(0.6)	1.2(0.3)	2.5(0.5)	6.0(0.8)	3.6(0.6)
100	60	1.2(0.3)	3.3(0.6)	4.7(0.7)	3.4(0.6)	1.0(0.3)	1.3(0.4)	4.9(0.7)	3.1(0.5)	0.9(0.3)	1.0(0.3)	4.9(0.7)	2.9(0.5)
100	90	1.4(0.4)	2.6(0.5)	4.7(0.7)	2.9(0.5)	1.5(0.4)	0.6(0.2)	4.8(0.7)	3.2(0.6)	0.4(0.2)	0.2(0.1)	4.0(0.6)	2.3(0.5)
100	120	1.0(0.3)	1.1(0.3)	4.3(0.6)	2.1(0.5)	0.6(0.2)	0.4(0.2)	3.8(0.6)	2.0(0.4)	0.6(0.2)	0.1(0.1)	5.0(0.7)	2.0(0.4)
200	30	1.3(0.4)	5.1(0.7)	5.9(0.7)	4.4(0.6)	1.8(0.4)	4.3(0.6)	5.1(0.7)	4.2(0.6)	1.4(0.4)	3.9(0.6)	5.7(0.7)	4.2(0.6)
200	60	1.6(0.4)	4.5(0.7)	5.9(0.7)	4.3(0.6)	1.0(0.3)	3.3(0.6)	5.2(0.7)	3.6(0.6)	1.1(0.3)	1.7(0.4)	5.0(0.7)	2.8(0.5)
200	90	1.3(0.4)	2.8(0.5)	4.6(0.7)	3.3(0.6)	1.2(0.3)	1.5(0.4)	4.6(0.7)	2.5(0.5)	1.0(0.3)	0.4(0.2)	4.2(0.6)	2.6(0.5)
200	120	1.2(0.3)	2.5(0.5)	4.8(0.7)	3.0(0.5)	1.6(0.4)	1.7(0.4)	5.2(0.7)	3.5(0.6)	1.9(0.4)	0.5(0.2)	4.4(0.6)	2.4(0.5)
		Setting (II)											
100	30	2.1(0.5)	4.1(0.6)	5.3(0.7)	4.5(0.7)	1.0(0.3)	3.8(0.6)	6.1(0.8)	4.4(0.6)	1.6(0.4)	2.5(0.5)	5.2(0.7)	4.3(0.6)
100	60	0.9(0.3)	3.0(0.5)	5.0(0.7)	2.8(0.5)	0.4(0.2)	1.4(0.4)	4.4(0.6)	2.8(0.5)	0.6(0.2)	1.3(0.4)	6.5(0.8)	2.9(0.5)
100	90	0.8(0.3)	2.0(0.4)	4.4(0.6)	2.2(0.5)	0.8(0.3)	0.8(0.3)	5.0(0.7)	2.0(0.4)	0.7(0.3)	0.4(0.2)	4.3(0.6)	2.0(0.4)
100	120	0.9(0.3)	1.3(0.4)	3.9(0.6)	2.8(0.5)	0.8(0.3)	0.9(0.3)	5.9(0.7)	3.7(0.6)	1.1(0.3)	0.2(0.1)	5.4(0.7)	2.7(0.5)
200	30	1.7(0.4)	5.1(0.7)	5.5(0.7)	5.2(0.7)	1.3(0.4)	4.2(0.6)	5.3(0.7)	3.9(0.6)	1.9(0.4)	4.4(0.6)	6.6(0.8)	5.7(0.7)
200	60	1.4(0.4)	3.3(0.6)	5.4(0.7)	4.8(0.7)	1.6(0.4)	2.2(0.5)	3.7(0.6)	4.2(0.6)	1.8(0.4)	1.6(0.4)	4.0(0.6)	2.6(0.5)
200	90	1.2(0.3)	3.6(0.6)	5.4(0.7)	3.2(0.6)	1.5(0.4)	1.8(0.4)	4.8(0.7)	3.4(0.6)	1.9(0.4)	1.0(0.3)	5.6(0.7)	3.5(0.6)
200	120	1.6(0.4)	3.3(0.6)	5.0(0.7)	4.1(0.6)	1.3(0.4)	1.6(0.4)	6.0(0.8)	3.3(0.6)	2.0(0.4)	0.6(0.2)	5.9(0.7)	3.5(0.6)
		Setting (III)											
100	30	0.8(0.3)	4.7(0.7)	5.2(0.7)	4.4(0.6)	0.3(0.2)	4.1(0.6)	6.1(0.8)	3.7(0.6)	1.1(0.3)	2.9(0.5)	5.5(0.7)	3.2(0.6)
100	60	0.9(0.3)	3.7(0.6)	5.0(0.7)	2.6(0.5)	0.6(0.2)	2.3(0.5)	5.4(0.7)	3.1(0.5)	0.9(0.3)	1.1(0.3)	5.7(0.7)	2.9(0.5)
100	90	0.9(0.3)	1.7(0.4)	3.9(0.6)	1.6(0.4)	0.8(0.3)	1.2(0.3)	5.1(0.7)	2.8(0.5)	1.0(0.3)	0.3(0.2)	3.9(0.6)	2.5(0.5)
100	120	0.4(0.2)	2.1(0.5)	5.6(0.7)	2.1(0.5)	0.6(0.2)	1.0(0.3)	5.2(0.7)	2.6(0.5)	0.6(0.2)	0.2(0.1)	4.9(0.7)	2.2(0.5)
200	30	1.3(0.4)	4.1(0.6)	5.1(0.7)	5.0(0.7)	1.3(0.4)	4.9(0.7)	5.8(0.7)	4.6(0.7)	2.1(0.5)	3.4(0.6)	5.0(0.7)	4.4(0.6)
200	60	1.7(0.4)	4.0(0.6)	5.3(0.7)	4.5(0.7)	0.9(0.3)	3.2(0.6)	5.3(0.7)	3.0(0.5)	1.1(0.3)	3.1(0.5)	6.1(0.8)	3.6(0.6)
200	90	1.6(0.4)	4.3(0.6)	5.3(0.7)	4.5(0.7)	0.9(0.3)	2.5(0.5)	5.8(0.7)	3.3(0.6)	1.3(0.4)	2.1(0.5)	5.4(0.7)	4.0(0.6)
200	120	1.7(0.4)	3.2(0.6)	4.8(0.7)	3.6(0.6)	1.1(0.3)	2.0(0.4)	5.2(0.7)	3.0(0.5)	1.1(0.3)	0.9(0.3)	5.0(0.7)	3.1(0.5)

ical size results of this setting are presented in Table 2. The findings indicate that the LY test tends to exhibit excessive conservatism under this setting, while the proposed SUM test demonstrates effective control over the empirical size in most cases. In addition, the MAX test appears to lean towards conservatism, a trait attributed to its slow rate of convergence. FC is reasonably conservative

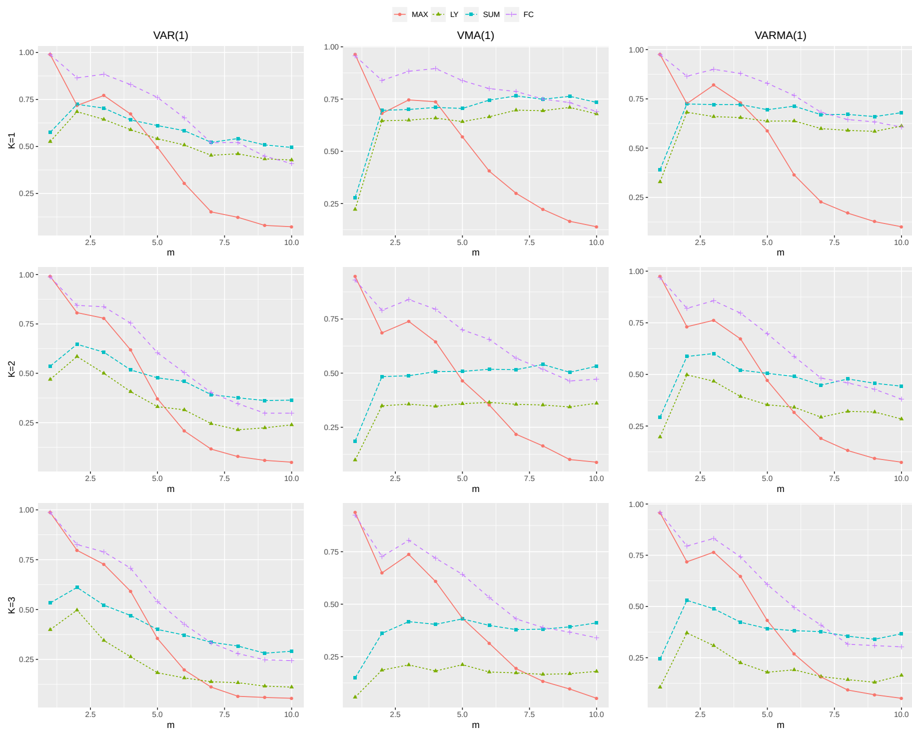

Figure 1: Power curves of the involved tests in situation where $z_{it} \overset{i.i.d}{\sim} t(8)/\sqrt{4/3}$, $m \in \{1, 2, \cdots, 10\}$ and $(n, p) = (200, 60)$.

and maintains the size well. Figure 3 presents the power curves of the involved tests under distribution (i) with $K = 10$ and $(n, p) = (200, 60)$. Comparing it with the figures obtained in the main text for smaller values of K, it can be observed that the performance of the MAX test is not sensitive to the choice of K. In contrast, the proposed SUM test is sensitive to the value of K. It exhibits poor power performance when K is large, possibly because the signal may be weakened as K increases. Therefore, we recommend not using an excessively large K in the proposed tests. FC is clear winner in terms of power in all cases.

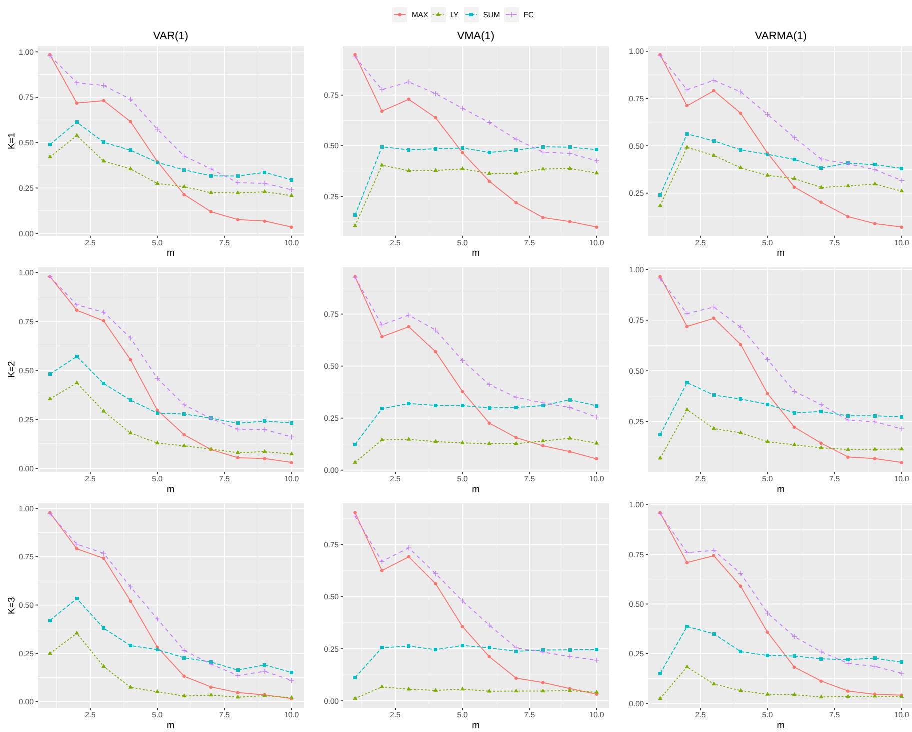

Figure 2: Power curves of the involved tests in situation where $z_{it} \overset{i.i.d}{\sim} t(8)/\sqrt{4/3}$, $m \in \{1, 2, \cdots, 10\}$ and $(n, p) = (200, 90)$.

S2.3 Comparison with the test proposed by Chang et al. (2017)

Table 3 presents the empirical size performance of the test proposed by Chang et al. (2017), abbreviated as CYZ, using the same parameters as in the main text. In addition, the simulation results under distribution (iii) that was considered in Section S2.1 is included in this table. It suggests that CYZ is generally conservative.

Figure 4 illustrates the power curves of MAX and CYZ, under identical conditions as in the main text, with $(n, p) = (200, 60)$ and distribution (i). It suggest-

Table 2: Size performance in the case of $\boldsymbol{\varepsilon}_t = \mathbf{A}\boldsymbol{z}_t$ with $K = 10$ and $(n, p) = (200, 60)$.

		(i) $N(0,1)$				(ii) $Ga(4,0.5)-2$				(iii) $t(8)/\sqrt{4/3}$			
n	p	MAX	LY	SUM	FC	MAX	LY	SUM	FC	MAX	LY	SUM	FC
		Setting (I)											
100	30	0.8(0.3)	0.0(0.0)	4.3(0.6)	2.5(0.5)	1.7(0.4)	0.1(0.1)	4.5(0.7)	4.5(0.7)	1.2(0.3)	0.1(0.1)	4.6(0.7)	2.6(0.5)
100	60	0.7(0.3)	0.0(0.0)	3.8(0.6)	1.5(0.4)	2.1(0.5)	0.0(0.0)	4.8(0.7)	3.7(0.6)	0.9(0.3)	0.0(0.0)	3.2(0.6)	1.7(0.4)
100	90	0.3(0.2)	0.0(0.0)	4.8(0.7)	1.9(0.4)	1.5(0.4)	0.0(0.0)	5.3(0.7)	4.0(0.6)	0.5(0.2)	0.0(0.0)	4.7(0.7)	2.5(0.5)
100	120	0.4(0.2)	0.0(0.0)	3.9(0.6)	1.2(0.3)	1.8(0.4)	0.0(0.0)	4.5(0.7)	3.2(0.6)	0.4(0.2)	0.0(0.0)	4.4(0.6)	2.1(0.5)
200	30	0.7(0.3)	0.3(0.2)	5.6(0.7)	2.6(0.5)	3.6(0.6)	0.0(0.0)	3.9(0.6)	5.6(0.7)	1.3(0.4)	0.2(0.1)	5.3(0.7)	3.7(0.6)
200	60	1.5(0.4)	0.1(0.1)	5.0(0.7)	2.3(0.5)	3.6(0.6)	0.0(0.0)	3.7(0.6)	3.7(0.6)	1.3(0.4)	0.0(0.0)	4.6(0.7)	3.0(0.5)
200	90	1.7(0.4)	0.0(0.0)	4.6(0.7)	3.5(0.6)	3.1(0.5)	0.0(0.0)	4.4(0.6)	4.5(0.7)	1.0(0.3)	0.0(0.0)	3.8(0.6)	2.3(0.5)
200	120	0.8(0.3)	0.0(0.0)	5.0(0.7)	3.1(0.5)	2.8(0.5)	0.0(0.0)	4.9(0.7)	4.7(0.7)	1.6(0.4)	0.0(0.0)	4.8(0.7)	3.2(0.6)
		Setting (II)											
100	30	0.8(0.3)	0.1(0.1)	5.1(0.7)	2.9(0.5)	1.7(0.4)	0.0(0.0)	4.6(0.7)	3.2(0.6)	1.1(0.3)	0.1(0.1)	4.5(0.7)	2.8(0.5)
100	60	0.8(0.3)	0.0(0.0)	5.7(0.7)	3.4(0.6)	1.9(0.4)	0.0(0.0)	6.1(0.8)	4.6(0.7)	0.7(0.3)	0.0(0.0)	4.6(0.7)	2.1(0.5)
100	90	0.6(0.2)	0.0(0.0)	4.1(0.6)	1.9(0.4)	1.7(0.4)	0.0(0.0)	4.9(0.7)	2.6(0.5)	0.6(0.2)	0.0(0.0)	4.0(0.6)	1.7(0.4)
100	120	0.4(0.2)	0.0(0.0)	4.8(0.7)	2.3(0.5)	1.9(0.4)	0.0(0.0)	3.9(0.6)	2.8(0.5)	0.3(0.2)	0.0(0.0)	4.2(0.6)	1.9(0.4)
200	30	0.8(0.3)	0.5(0.2)	4.3(0.6)	3.2(0.6)	2.2(0.5)	0.1(0.1)	4.3(0.6)	4.0(0.6)	1.6(0.4)	0.3(0.2)	5.4(0.7)	3.7(0.6)
200	60	0.7(0.3)	0.1(0.1)	4.0(0.6)	2.5(0.5)	3.5(0.6)	0.0(0.0)	4.7(0.7)	4.0(0.6)	1.0(0.3)	0.0(0.0)	5.2(0.7)	2.4(0.5)
200	90	0.8(0.3)	0.0(0.0)	6.0(0.8)	3.7(0.6)	3.3(0.6)	0.0(0.0)	4.0(0.6)	4.3(0.6)	1.5(0.4)	0.0(0.0)	4.7(0.7)	2.7(0.5)
200	120	0.6(0.2)	0.0(0.0)	4.5(0.7)	2.1(0.5)	3.9(0.6)	0.0(0.0)	5.2(0.7)	4.8(0.7)	1.4(0.4)	0.0(0.0)	5.8(0.7)	3.0(0.5)
		Setting (III)											
100	30	1.0(0.3)	0.0(0.0)	3.9(0.6)	2.6(0.5)	1.9(0.4)	0.1(0.1)	5.3(0.7)	4.5(0.7)	0.7(0.3)	0.0(0.0)	4.4(0.6)	2.3(0.5)
100	60	0.6(0.2)	0.0(0.0)	5.3(0.7)	2.2(0.5)	2.5(0.5)	0.0(0.0)	4.2(0.6)	3.8(0.6)	0.6(0.2)	0.0(0.0)	5.4(0.7)	2.1(0.5)
100	90	0.5(0.2)	0.0(0.0)	3.9(0.6)	1.4(0.4)	1.6(0.4)	0.0(0.0)	3.8(0.6)	2.9(0.5)	0.6(0.2)	0.0(0.0)	4.5(0.7)	2.5(0.5)
100	120	0.2(0.1)	0.0(0.0)	4.8(0.7)	1.1(0.3)	0.8(0.3)	0.0(0.0)	4.5(0.7)	3.1(0.5)	0.4(0.2)	0.0(0.0)	4.1(0.6)	1.7(0.4)
200	30	1.3(0.4)	0.3(0.2)	5.3(0.7)	3.1(0.5)	2.4(0.5)	0.4(0.2)	5.1(0.7)	5.4(0.7)	1.9(0.4)	0.3(0.2)	3.8(0.6)	3.1(0.5)
200	60	1.1(0.3)	0.0(0.0)	5.7(0.7)	3.5(0.6)	2.4(0.5)	0.0(0.0)	5.3(0.7)	4.4(0.6)	1.8(0.4)	0.0(0.0)	4.1(0.6)	3.1(0.5)
200	90	1.0(0.3)	0.0(0.0)	4.4(0.6)	2.3(0.5)	2.1(0.5)	0.0(0.0)	3.7(0.6)	3.3(0.6)	1.4(0.4)	0.0(0.0)	4.1(0.6)	2.4(0.5)
200	120	1.4(0.4)	0.0(0.0)	4.6(0.7)	1.9(0.4)	2.4(0.5)	0.0(0.0)	4.1(0.6)	4.3(0.6)	1.8(0.4)	0.0(0.0)	5.2(0.7)	3.4(0.6)

s that CYZ outperforms MAX, possibly due to the application of bootstrapping, which tends to provide more precise estimation of the variances. However, the computational time consumed by CYZ is about several hundred times that of MAX. Therefore, the decision to opt for either of these two methods hinges on the specific computational time constraints.

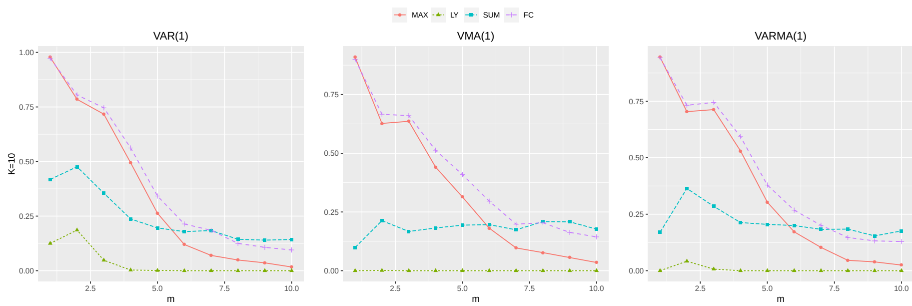

Figure 3: Power curves of the involved tests under distribution (i) with $K = 10$, $m \in \{1, 2, \cdots, 10\}$ and $(n, p) = (200, 60)$.

S2.4 Comparison with the test proposed by Chang et al. (2023)

Chang et al. (2023) proposed a method for testing the martingale difference in high dimensions, which can be applied to testing white noise. We implement this method in the following way, and compare it with the methods proposed in this paper. We set $\phi(\boldsymbol{x}) = \boldsymbol{x}$, then the corresponding test statistic in Chang et al. (2023) takes the form

$$T_{\text{CJS}} = \sum_{k=1}^{K} T_{n,k} = \sum_{k=1}^{K} \max_{1 \le i,j \le p} n^{1/2} |\hat{\rho}_{ij}(k)|.$$

In a manner akin to Chang et al. (2017), the test based on T_{CJS}, abbreviated as CJS, employed a simulation-based method to compute the critical value. Hence, it is also time-intensive. Table 4 reports the empirical size results of CJS under the same settings as in Section S2.3. It suggests that CJS is conservative in general.

Table 3: Size performance of the max-type test proposed by Chang et al. (2017) in the case of $\varepsilon_t = \mathbf{A}z_t$.

		(i) $N(0,1)$			(ii) $Ga(4,0.5)-2$			(iii) $t(8)/\sqrt{4/3}$		
n	p	$K=1$	$K=2$	$K=3$	$K=1$	$K=2$	$K=3$	$K=1$	$K=2$	$K=3$
		Setting (I)								
100	30	1.3(0.4)	0.3(0.2)	0.6(0.2)	0.0(0.0)	0.0(0.0)	0.1(0.1)	0.1(0.1)	0.3(0.2)	0.1(0.1)
100	60	0.5(0.2)	0.0(0.0)	0.3(0.2)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
100	90	0.4(0.2)	0.3(0.2)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
100	120	0.1(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
200	30	1.8(0.4)	2.1(0.5)	1.5(0.4)	1.4(0.4)	0.7(0.3)	0.4(0.2)	0.7(0.3)	1.0(0.3)	0.3(0.2)
200	60	1.4(0.4)	1.0(0.3)	0.9(0.3)	0.3(0.2)	0.4(0.2)	0.1(0.1)	0.5(0.2)	0.1(0.1)	0.1(0.1)
200	90	1.0(0.3)	1.0(0.3)	0.6(0.2)	0.2(0.1)	0.2(0.1)	0.0(0.0)	0.3(0.2)	0.1(0.1)	0.0(0.0)
200	120	1.1(0.3)	0.8(0.3)	0.3(0.2)	0.3(0.2)	0.0(0.0)	0.0(0.0)	0.1(0.1)	0.0(0.0)	0.2(0.1)
		Setting (II)								
100	30	0.7(0.3)	0.4(0.2)	0.3(0.2)	0.2(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.2(0.1)	0.1(0.1)
100	60	0.3(0.2)	0.5(0.2)	0.2(0.1)	0.1(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
100	90	0.0(0.0)	0.0(0.0)	0.2(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
100	120	0.1(0.1)	0.0(0.0)	0.1(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
200	30	2.0(0.4)	1.9(0.4)	1.5(0.4)	0.7(0.3)	0.9(0.3)	0.5(0.2)	1.7(0.4)	0.4(0.2)	0.2(0.1)
200	60	1.4(0.4)	1.1(0.3)	0.5(0.2)	0.4(0.2)	0.0(0.0)	0.2(0.1)	0.3(0.2)	0.2(0.1)	0.0(0.0)
200	90	1.4(0.4)	0.9(0.3)	0.8(0.3)	0.2(0.1)	0.0(0.0)	0.1(0.1)	0.3(0.2)	0.2(0.1)	0.0(0.0)
200	120	0.4(0.2)	0.5(0.2)	0.5(0.2)	0.2(0.1)	0.1(0.1)	0.1(0.1)	0.1(0.1)	0.2(0.1)	0.0(0.0)
		Setting (III)								
100	30	1.1(0.3)	0.5(0.2)	0.3(0.2)	0.2(0.1)	0.1(0.1)	0.1(0.1)	0.2(0.1)	0.0(0.0)	0.0(0.0)
100	60	0.2(0.1)	0.3(0.2)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
100	90	0.1(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
100	120	0.1(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)
200	30	1.8(0.4)	1.8(0.4)	0.7(0.3)	0.7(0.3)	0.7(0.3)	0.7(0.3)	1.2(0.3)	0.3(0.2)	0.2(0.1)
200	60	1.2(0.3)	0.6(0.2)	0.9(0.3)	0.5(0.2)	0.4(0.2)	0.2(0.1)	0.4(0.2)	0.3(0.2)	0.0(0.0)
200	90	0.9(0.3)	0.6(0.2)	0.5(0.2)	0.3(0.2)	0.0(0.0)	0.0(0.0)	0.4(0.2)	0.0(0.0)	0.0(0.0)
200	120	0.4(0.2)	0.8(0.3)	0.5(0.2)	0.2(0.1)	0.1(0.1)	0.1(0.1)	0.1(0.1)	0.0(0.0)	0.0(0.0)

Next, we will compare the empirical power performance of CJS with the proposed combination test procedure FC. The simulation settings are the same as those outlined in Section 3 of the main text. Figure 5 illustrates the power curves of these test procedures under different sparsity levels with $(n,p) = (200,60)$. The curves indicate that under sparse alternatives, CJS performs better than FC

Table 4: Empirical size performance of CJS in the case of $\varepsilon_t = \mathbf{A}\boldsymbol{z}_t$.

		(i) $N(0,1)$			(ii) $Ga(4,0.5)-2$			(iii) $t(8)/\sqrt{4/3}$		
n	p	$K=1$	$K=2$	$K=3$	$K=1$	$K=2$	$K=3$	$K=1$	$K=2$	$K=3$
		Setting (I)								
100	30	0.7(0.3)	0.4(0.2)	0.2(0.1)	0.8(0.3)	0.2(0.1)	0.1(0.1)	0.3(0.2)	0.2(0.1)	0.0(0.0)
100	60	0.6(0.2)	0.2(0.1)	0.1(0.1)	0.1(0.1)	0.0(0.0)	0.0(0.0)	0.1(0.1)	0.0(0.0)	0.0(0.0)
100	90	0.5(0.2)	0.3(0.2)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.2(0.1)	0.1(0.1)	0.0(0.0)
100	120	0.2(0.1)	0.1(0.1)	0.0(0.0)	0.2(0.1)	0.0(0.0)	0.0(0.0)	0.1(0.1)	0.0(0.0)	0.0(0.0)
200	30	1.8(0.4)	2.0(0.4)	0.6(0.2)	1.1(0.3)	0.5(0.2)	0.1(0.1)	1.1(0.3)	0.3(0.2)	0.2(0.1)
200	60	1.4(0.4)	0.7(0.3)	0.2(0.1)	0.5(0.2)	0.0(0.0)	0.0(0.0)	0.4(0.2)	0.2(0.1)	0.0(0.0)
200	90	0.9(0.3)	0.6(0.2)	0.0(0.0)	0.4(0.2)	0.0(0.0)	0.0(0.0)	0.3(0.2)	0.1(0.1)	0.0(0.0)
200	120	1.1(0.3)	0.3(0.2)	0.2(0.1)	0.4(0.2)	0.0(0.0)	0.0(0.0)	0.1(0.1)	0.1(0.1)	0.0(0.0)
		Setting (II)								
100	30	1.1(0.3)	0.6(0.2)	0.3(0.2)	0.3(0.2)	0.1(0.1)	0.0(0.0)	0.3(0.2)	0.3(0.2)	0.0(0.0)
100	60	0.6(0.2)	0.5(0.2)	0.3(0.2)	0.2(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.2(0.1)	0.0(0.0)
100	90	0.7(0.3)	0.1(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.1(0.1)	0.1(0.1)	0.0(0.0)	0.0(0.0)
100	120	0.4(0.2)	0.1(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.1(0.1)	0.1(0.1)	0.0(0.0)
200	30	2.6(0.5)	1.5(0.4)	1.0(0.3)	1.8(0.4)	0.4(0.2)	0.1(0.1)	1.2(0.3)	0.9(0.3)	0.2(0.1)
200	60	1.7(0.4)	0.8(0.3)	0.6(0.2)	0.7(0.3)	0.2(0.1)	0.0(0.0)	0.6(0.2)	0.3(0.2)	0.0(0.0)
200	90	1.8(0.4)	0.5(0.2)	0.0(0.0)	0.2(0.1)	0.1(0.1)	0.0(0.0)	0.1(0.1)	0.1(0.1)	0.1(0.1)
200	120	1.4(0.4)	0.4(0.2)	0.0(0.0)	0.7(0.3)	0.0(0.0)	0.1(0.1)	0.2(0.1)	0.0(0.0)	0.0(0.0)
		Setting (III)								
100	30	1.8(0.4)	0.3(0.2)	0.2(0.1)	0.9(0.3)	0.1(0.1)	0.2(0.1)	0.4(0.2)	0.1(0.1)	0.0(0.0)
100	60	0.7(0.3)	0.3(0.2)	0.0(0.0)	0.1(0.1)	0.0(0.0)	0.0(0.0)	0.4(0.2)	0.0(0.0)	0.0(0.0)
100	90	0.4(0.2)	0.0(0.0)	0.0(0.0)	0.1(0.1)	0.0(0.0)	0.2(0.1)	0.1(0.1)	0.0(0.0)	0.0(0.0)
100	120	0.3(0.2)	0.1(0.1)	0.0(0.0)	0.1(0.1)	0.0(0.0)	0.1(0.1)	0.0(0.0)	0.1(0.1)	0.0(0.0)
200	30	2.1(0.5)	0.8(0.3)	0.7(0.3)	1.0(0.3)	0.4(0.2)	0.2(0.1)	1.0(0.3)	0.2(0.1)	0.3(0.2)
200	60	2.7(0.5)	0.3(0.2)	0.4(0.2)	0.8(0.3)	0.2(0.1)	0.0(0.0)	0.5(0.2)	0.0(0.0)	0.0(0.0)
200	90	2.0(0.4)	0.4(0.2)	0.1(0.1)	0.4(0.2)	0.0(0.0)	0.0(0.0)	0.3(0.2)	0.0(0.0)	0.0(0.0)
200	120	1.1(0.3)	0.2(0.1)	0.0(0.0)	0.2(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)

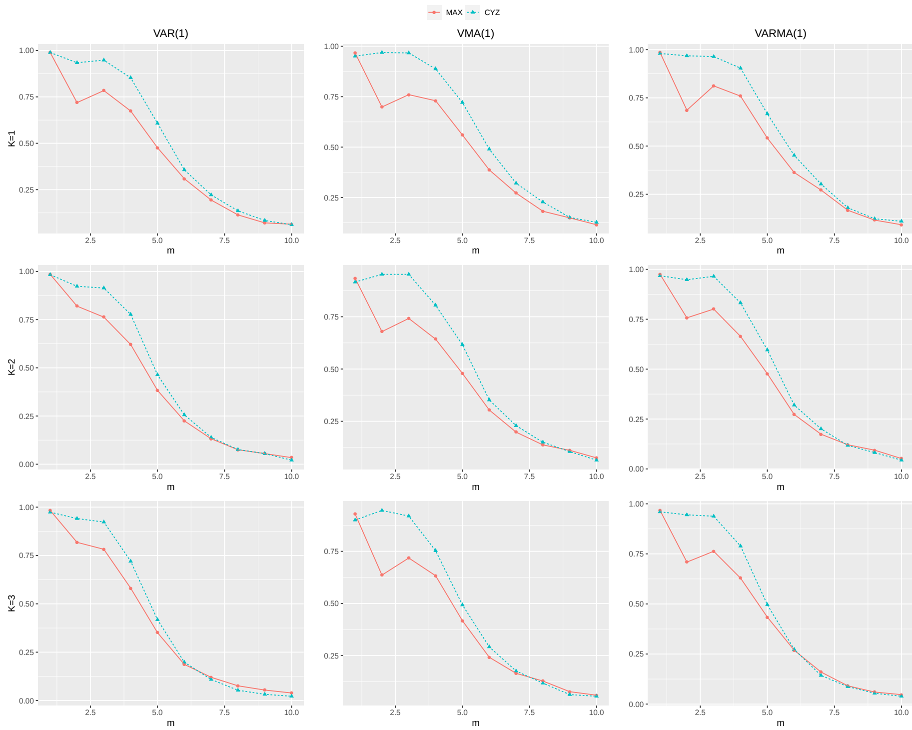

Figure 4: Power curves of our proposed MAX tests and Chang et al. (2017)'s test under distribution (i) with $m \in \{1, 2, \cdots, 10\}$ and $(n, p) = (200, 60)$.

in power comparison, whereas under dense alternatives, FC performs better than CJS. These results inspire us to consider how to establish new combination tests based on statistics similar to CJS in future work, which may significantly enhance the power of the combination tests.

S3. Technical proofs

S3.1 Proof of Theorem 1

First, we present some technical results for the proof of Theorem 1.

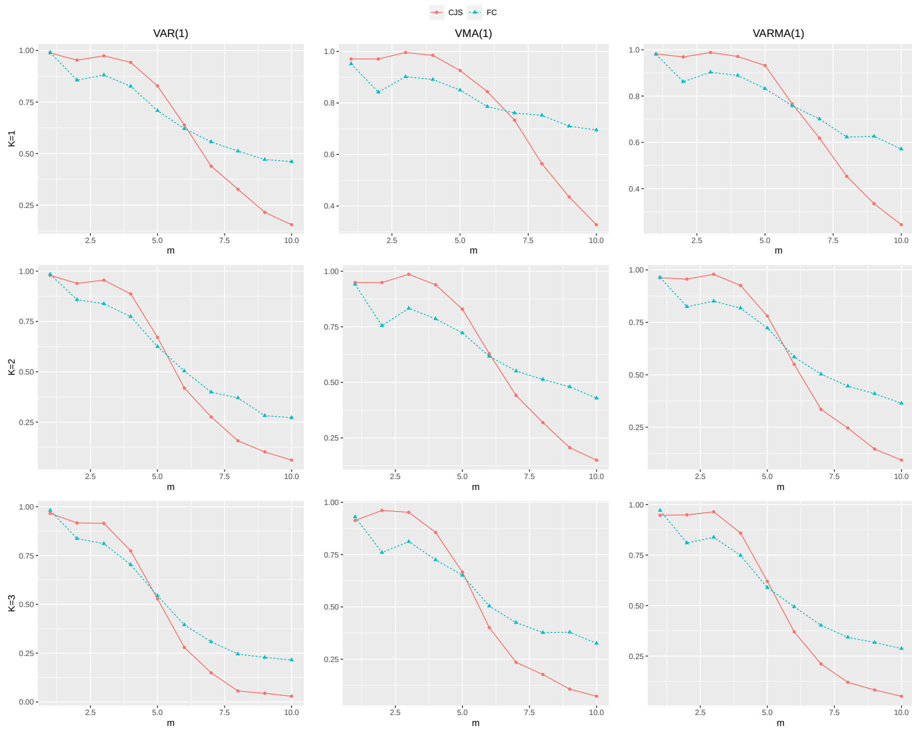

Figure 5: Power curves of our proposed combination test and Chang et al. (2023)'s test under distribution (i) with $m \in \{1, 2, \cdots, 10\}$ and $(n, p) = (200, 60)$.

We restate Theorem 2 in Feng et al. (2024) as the following Proposition 1, in which the following condition is imposed.

(CA1) Let $\boldsymbol{\Sigma} = \{\sigma_{ij}\}_{1 \le i,j \le p}$. For some $\varrho \in (0, 1)$, assume $|\sigma_{ij}| \le \varrho$ for all $1 \le i < j \le p$ and $p \ge 2$. Suppose $\{\delta_p : p \ge 1\}$ and $\{\kappa_p : p \ge 1\}$ are positive constants with $\delta_p = o(1/\log p)$ and $\kappa = \kappa_p \to 0$ as $p \to \infty$. For $1 \le i \le p$, define $B_{p,i} = \big\{1 \le j \le p : |\sigma_{ij}| \ge \delta_p\big\}$ and $C_p = \big\{1 \le i \le p : |B_{p,i}| \ge p^{\kappa}\big\}$. Assume that $|C_p|/p \to 0$ as $p \to \infty$.

Proposition 1. *Suppose* $(Z_1, \cdots, Z_p)^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ *and Condition (CA1) holds.*

Then, we have $\max_{1\le i\le p} Z_i^2 - 2\log p + \log\log p$ *converges to a Gumbel distribution with cdf* $G(x) = \exp(-\frac{1}{\sqrt{\pi}}e^{-x/2})$ *as* $p \to \infty$.

The following lemma is from the proof of Theorem 2 in Feng et al. (2024).

Lemma 1. *Suppose* $(Z_1, \cdots, Z_p)^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ *and Condition (CA1) holds. For any* $x \in \mathbb{R}$ *and any* $1 \le t \le p$*, let*

$$\alpha_t = \sum \mathrm{P}\{|Z_{i_1}| > z, \cdots, |Z_{i_t}| > z\},\ z = \left(2\log p - \log\log p + x\right)^{1/2},$$

where the sum runs over all $i_1 < \cdots < i_t$ *with* $i_1 \cdots, i_t \in D_p = \{1, \cdots, p\}\backslash C_p$*. Then,*

$$\lim_{p\to\infty} \alpha_t = \frac{1}{t!}\pi^{-t/2}e^{-tx/2}. \tag{S3.1}$$

Lemma 2. *(Bernstein's inequality) Let* $X_1, \ldots, X_n$ *be independent centered random variables a.s. bounded by* $A < \infty$ *in absolute value. Let* $\sigma^2 = n^{-1}\sum_{i=1}^n \mathbb{E}\left(X_i^2\right)$. *Then for all* $x > 0$*,*

$$\mathrm{P}\left(\sum_{i=1}^n X_i \ge x\right) \le \exp\left(-\frac{x^2}{2n\sigma^2 + 2Ax/3}\right).$$

Define $\hat{\sigma}_i^2 = \frac{1}{n}\sum_{t=1}^n \varepsilon_{ti}^2$ and $\sigma_i^2 = \mathrm{var}(\varepsilon_{ti})$.

Lemma 3. *Suppose Condition (C2) holds. Then, under* H_0*,* *(1) if (C1)-(i) holds, we have*

$$P\left(\max_{1\le i\le p}\left|\hat{\sigma}_i^2-\sigma_i^2\right|\ge C\frac{\epsilon_n}{\log p}\right)=O(p^{-1}), \tag{S3.2}$$

(2) if (C1)-(ii) holds,

$$P\left(\max_{1\le i\le p}\left|\hat{\sigma}_i^2-\sigma_i^2\right|\ge C\frac{\epsilon_n}{\log p}\right)=O(n^{-\epsilon/8}), \tag{S3.3}$$

as $\epsilon_n \doteq \max\left\{(\log p)^{1/6}/n^{1/2}, (\log p)^{-1}\right\}\to 0.$

Proof. We first assume that (C1)-(i) holds. It suffices to show that, for any $\delta>0$,

$$P\left\{\max_{1\le i\le p}\left|\frac{1}{n}\sum_{k=1}^{n}\left(\varepsilon_{ki}^2-\mathbb{E}\varepsilon_{ki}^2\right)\right|\ge C\sqrt{\frac{\log p}{n}}\right\}=O\left(p^{-\delta}\right). \tag{S3.4}$$

Define

$$\tilde{\varepsilon}_{ki}\doteq\varepsilon_{ki}\mathbb{I}\left\{|\varepsilon_{ki}|\le\tau\sqrt{\log(p+n)}\right\},$$

where τ is sufficiently large. We have

$$\begin{aligned}\left|\mathbb{E}\varepsilon_{ki}^2-\mathbb{E}\tilde{\varepsilon}_{ki}^2\right| &\le C\left(\mathbb{E}\varepsilon_{ki}^4\mathbb{E}\left[I\left\{|\varepsilon_{ki}|\ge\tau\sqrt{\log(p+n)}\right\}\right]\right)^{1/2}\\ &\le C(n+p)^{-\tau^2\eta/2}\left\{\mathbb{E}\varepsilon_{ki}^4\exp\left(2^{-1}\eta\varepsilon_{ki}^2\right)\right\}^{1/2}\\ &\le C(n+p)^{-\tau^2\eta/2}, \end{aligned} \tag{S3.5}$$

where C does not depend on n and p. Thus, it follows that

$$\begin{aligned}
&\mathrm{P}\left\{\max_{1\le i\le p}\left|\frac{1}{n}\sum_{k=1}^{n}\left(\varepsilon_{ki}^2-\mathbb{E}\varepsilon_{ki}^2\right)\right|\ge C\sqrt{\frac{\log p}{n}}\right\}\\
\le&\mathrm{P}\left\{\max_{1\le i\le p}\left|\frac{1}{n}\sum_{k=1}^{n}\left(\tilde\varepsilon_{ki}^2-\mathbb{E}\tilde\varepsilon_{ki}^2\right)\right|\ge 2^{-1}C\sqrt{\frac{\log p}{n}\Big)}\right\}\\
&+np\mathrm{P}\left\{|\varepsilon_{ki}|\ge\tau\sqrt{\log(p+n)}\right\},
\end{aligned}$$

where

$$np\mathrm{P}\left\{|\varepsilon_{ki}|\ge\tau\sqrt{\log(p+n)}\right\}\le np(n+p)^{-\tau^2\eta}\mathbb{E}\exp\left(\eta\varepsilon_{ki}^2\right)=O\left(p^{-\delta}\right).$$

Let $t=\eta\left(8\tau^2\right)^{-1}\sqrt{\log p/n}$ and $\tilde Z_{ki}=\tilde\varepsilon_{ki}^2-\mathbb{E}\tilde\varepsilon_{ki}^2$. Then, we have

$$\begin{aligned}
&\mathrm{P}\left\{\frac{1}{n}\sum_{k=1}^{n}\left(\tilde\varepsilon_{ki}^2-\mathbb{E}\tilde\varepsilon_{ki}^2\right)\ge C\sqrt{\frac{\log p}{n}}\right\}\\
\le&\exp\left(-Ct\sqrt{n\log p}\right)\prod_{k=1}^{n}\mathbb{E}\exp\left(t\tilde Z_{ki}\right)\\
\le&\exp\left(-Ct\sqrt{n\log p}\right)\prod_{k=1}^{n}\left\{1+\mathbb{E}t^2\tilde Z_{ki}^2\exp\left(t\left|\tilde Z_{ki}\right|\right)\right\}\\
\le&\exp\left\{-Ct\sqrt{n\log p}+\sum_{k=1}^{n}\mathbb{E}t^2\tilde Z_{ki}^2\exp\left(t\left|\tilde Z_{ki}\right|\right)\right\}\\
\le&\exp\left\{-C\eta\left(8\tau^2\right)^{-1}\log p+c_{\tau,\eta}\log p\right\}\le Cp^{-\delta},
\end{aligned}$$

where $c_{\tau,\eta}$ is a positive constant depending only on τ and η. Similarly, we can show that

$$\mathrm{P}\left\{\frac{1}{n}\sum_{k=1}^{n}\left(\tilde{\varepsilon}_{ki}^2-\mathbb{E}\tilde{\varepsilon}_{ki}^2\right)\le -C\sqrt{\frac{\log p}{n}}\right\}\le Cp^{-\delta},$$

which leads to (S3.4).

It remains to prove this lemma under (C2)-(ii). Define

$$\hat{\varepsilon}_{ki}^2=\varepsilon_{ki}^2\mathbb{I}\left\{\left|\varepsilon_{ki}^2\right|\le n/(\log p)^8\right\}.$$

Then, as in $(S3.5)$, we can show that $|\mathbb{E}\varepsilon_{ki}^2-\mathbb{E}\hat{\varepsilon}_{ki}^2|\le Cn^{-\gamma_0/4}$. It follows that

$$\begin{aligned}
&\mathrm{P}\left\{\max_{1\le i\le p}\left|\sum_{k=1}^{n}\left(\varepsilon_{ki}^2-\mathbb{E}\varepsilon_{ki}^2\right)\right|\ge\frac{n\epsilon_n}{\log p}\right\}\\
\le&\mathrm{P}\left\{\max_{1\le i\le p}\left|\sum_{k=1}^{n}\left(\hat{\varepsilon}_{ki}^2-\mathbb{E}\hat{\varepsilon}_{ki}^2\right)\right|\ge 2^{-1}\frac{n\epsilon_n}{\log p}\right\}+\mathrm{P}\left\{\max_{i,k}\left|\varepsilon_{ki}^2\right|\ge\frac{n}{(\log p)^8}\right\}\\
\le&Cp^2\exp\left\{-C(\log p)^4\right\}+Cn^{-\epsilon/8},
\end{aligned}$$

where the last inequality follows from Lemma 2 and (C2)-(ii). Then, the proof of this lemma is completed. □

Define $\tilde{\Gamma}(k)=\{\tilde{\rho}_{ij}(k)\}_{1\leqslant i,j\leqslant p}\doteq\operatorname{diag}\{\mathbf{\Sigma}(0)\}^{-1/2}\hat{\mathbf{\Sigma}}(k)\operatorname{diag}\{\mathbf{\Sigma}(0)\}^{-1/2}$. We have the following results for $\tilde{\rho}_{ij}(k)$.

Lemma 4. *Under H_0, we have*

(1) if (C1)-(i) holds, we have

$$\mathrm{P}\left\{\max_{(i,j,k)\in\Lambda} n\tilde{\rho}_{ij}^2(k) \geq x^2\right\} \leq C|\Lambda|\{1-\Phi(x)\} + O(p^{-2}).$$

(2) if (C1)-(ii) holds, we have

$$\mathrm{P}\left\{\max_{(i,j,k)\in\Lambda} n\tilde{\rho}_{ij}^2(k) \geq x^2\right\} \leq C|\Lambda|\{1-\Phi(x)\} + O(n^{-\epsilon/8})$$

uniformly for $0 \leq x \leq \sqrt{8\log p}$ *and* $\Lambda \subset \{(i,j,k) : 1 \leq i,j \leq p, 1 \leq k \leq K\}$.

Proof. Rewrite

$$\begin{aligned} n\tilde{\rho}_{ij}^2(k) =& \frac{1}{n}\left(\sum_{t=1}^{n-k} \sigma_i^{-1}\sigma_j^{-1}\varepsilon_{ti}\varepsilon_{t+k,j}\right)^2 \\ =& \frac{\left(\sum_{t=1}^{n-k} \sigma_i^{-1}\sigma_j^{-1}\varepsilon_{ti}\varepsilon_{t+k,j}\right)^2}{\sum_{t=1}^{n-k} \sigma_i^{-2}\sigma_j^{-2}\varepsilon_{ti}^2\varepsilon_{t+k,j}^2} \times \frac{1}{n}\sum_{t=1}^{n-k} \sigma_i^{-2}\sigma_j^{-2}\varepsilon_{ti}^2\varepsilon_{t+k,j}^2. \end{aligned}$$

By the self-normalized large deviation theorem for independent random variables (Theorem 1 in Jing et al. (2003)), we can get

$$\max_{1\leq i\leq j\leq p, 1\leq k\leq K} \mathrm{P}\left\{\frac{\left(\sum_{t=1}^{n-k} \sigma_i^{-1}\sigma_j^{-1}\varepsilon_{ti}\varepsilon_{t+k,j}\right)^2}{\sum_{t=1}^{n-k} \sigma_i^{-2}\sigma_j^{-2}\varepsilon_{ti}^2\varepsilon_{t+k,j}^2} \geq x^2\right\} \leq C\{1-\Phi(x)\} \quad \text{(S3.6)}$$

uniformly for $0 \leq x \leq (8\log p)^{1/2}$. By Lemma 3 in Cai et al. (2013), we have

(1) if (C1)-(i) holds, we have

$$\mathrm{P}\left\{\left|\frac{1}{n}\sum_{t=1}^{n-k}\sigma_i^{-2}\sigma_j^{-2}\varepsilon_{ti}^2\varepsilon_{t+k,j}^2-1\right|\geq C\frac{\epsilon_n}{\log p}\right\}=O(p^{-\delta}), \qquad \text{(S3.7)}$$

(2) if (C1)-(ii) holds,

$$\mathrm{P}\left\{\left|\frac{1}{n}\sum_{t=1}^{n-k}\sigma_i^{-2}\sigma_j^{-2}\varepsilon_{ti}^2\varepsilon_{t+k,j}^2-1\right|\geq C\frac{\epsilon_n}{\log p}\right\}=O(n^{-\epsilon/8}), \qquad \text{(S3.8)}$$

for any $\delta > 0$. Thus, together (S3.6), (S3.7) with (S3.8), we can complete the proof of this lemma. □

Now, we are ready to present the proof of Theorem 1.

Proof of Theorem 1 Define $\tilde{T}_n \doteq \max_{1\leqslant k\leqslant K}\tilde{T}_{n,k}$, where

$$\tilde{T}_{n,k} \doteq \max_{1\leqslant i,j\leqslant p} n^{1/2}\left|\tilde{\rho}_{ij}(k)\right|.$$

Conditional on the event $\{\max_{1\leq i\leq p}|\hat{\sigma}_i^2-\sigma_i^2|\geq C\frac{\epsilon_n}{\log p}\}$, we have

$$|T_n^2-\tilde{T}_n^2|\leq C\tilde{T}_n\frac{\epsilon_n}{\log p}.$$

Thus, by Lemma 3, we only need to show that

$$\mathrm{P}\left\{\tilde{T}_n-2\log(Kp^2)+\log\log(Kp^2)\leqslant y\right\}\to\exp\left\{-\pi^{-1/2}\exp(-y/2)\right\}.$$

Restate that $\tilde{\rho}_{ij}(k) = \frac{1}{n}\sum_{t=1}^{n-k}\sigma_i^{-1}\sigma_j^{-1}\varepsilon_{ti}\varepsilon_{t+k,j}$. Without loss of generality, we assume that $\sigma_i = 1$ for all i. After some simply calculation, we have $\mathrm{cov}\left\{n^{1/2}\tilde{\rho}_{ij}(k), n^{1/2}\tilde{\rho}_{sw}(l)\right\} = \rho_{is}\rho_{jw}\mathbb{I}(k=l)$. Thus, we rearrange $\{n^{1/2}\tilde{\rho}_{ij}(k)\}_{1\le i,j\le p, 1\le k\le K}$ as $\{\nu_1, \cdots, \nu_N\}$ with $N = Kp^2$. Let $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_N)^\top$, then

$$\mathrm{cov}(\boldsymbol{\nu}) = \{a_{ij}\}_{1\le i,j\le N} \doteq \mathrm{diag}\{\boldsymbol{\Gamma}(0)\otimes\boldsymbol{\Gamma}(0), \cdots, \boldsymbol{\Gamma}(0)\otimes\boldsymbol{\Gamma}(0)\}, \tag{S3.9}$$

where $\otimes$ denotes the Kronecker product and $\mathrm{diag}\{\boldsymbol{\Gamma}(0)\otimes\boldsymbol{\Gamma}(0), \cdots, \boldsymbol{\Gamma}(0)\otimes\boldsymbol{\Gamma}(0)\}$ denotes the block diagonal matrix composed by $\boldsymbol{\Gamma}(0)\otimes\boldsymbol{\Gamma}(0), \cdots, \boldsymbol{\Gamma}(0)\otimes\boldsymbol{\Gamma}(0)$. For $1 \le i \le N$, define $B_{N,i} = \left\{1 \le j \le N : |a_{ij}| \ge \delta_p\right\}$ and $C_N = \left\{i : |B_{N,i}| \ge p^\kappa\right\}$. By the definition of a_{ij}, we have $|C_N| = |C_p|$.

Define $z = \{2\log(N) - 2\log\log(N) + y\}^{1/2}$. By Lemma 4, if (C1)-(i) holds, we have

$$\mathrm{P}(|\nu_i| \ge z) \le C\{1-\Phi(z)\} + O(p^{-2}) \le C\frac{1}{\sqrt{\pi}}\frac{e^{-y/2}}{N} + O(p^{-2});$$

if (C1)-(ii) holds, we have

$$\mathrm{P}(|\nu_i| \ge z) \le C\{1-\Phi(z)\} + O(n^{-\epsilon/8}) \le C\frac{1}{\sqrt{\pi}}\frac{e^{-y/2}}{N} + O(n^{-\epsilon/8}).$$

Thus, by Condition (C3),

$$\mathrm{P}\left(\max_{i\in C_N}|\nu_i|>z\right)\le |C_N|\cdot \mathrm{P}(|\nu_i|\ge z)\to 0$$

as $n,\ p\to\infty$. Set $D_N=\left\{1\le i\le N:|B_{N,i}|<p^{\kappa}\right\}$. By Condition (C3), $|D_N|/N\to 1$ as $N\to\infty$. Obviously,

$$\mathrm{P}\left(\max_{i\in D_N}|\nu_i|>z\right)\le \mathrm{P}\left(\max_{1\le i\le N}|\nu_i|>z\right)\le \mathrm{P}\left(\max_{i\in D_N}|\nu_i|>z\right)+\mathrm{P}\left(\max_{i\in C_N}|\nu_i|>z\right).$$

Therefore, to prove this theorem, it is enough to show

$$\lim_{N\to\infty}\mathrm{P}\left(\max_{i\in D_N}|\nu_i|>z\right)=1-\exp\left(-\frac{1}{\sqrt{\pi}}e^{-x/2}\right)$$

as $N\to\infty$.

We redefine $\nu_s=n^{1/2}\tilde{\rho}_{ij}(k)=n^{-1/2}\sum_{t=1}^{n-k}\varepsilon_{ti}\varepsilon_{t+k,j}=n^{-1/2}\sum_{t=1}^{n-k}Z_{ts}$. Let $\breve{Z}_{ts}=Z_{ts}\mathbb{I}(Z_{ts}\le\tau_n)-\mathbb{E}\{Z_{ts}I(Z_{ts}\le\tau_n)\}$. Here, $\tau_n=\eta^{-1}8M\log(p+n)$, if (C1)-(i) holds, and $\tau_n=\sqrt{n}/(\log p)^8$, if (C1)-(ii) holds. Define $\breve{\nu}_i=$

$n^{-1/2}\sum_{t=1}^{n-k}\breve{Z}_{ts}$ and $q=|D_N|$. If (C1)-(i) holds, then

$$\begin{aligned}
&\max_{1\le k\le q}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\mathbb{E}\,|Z_{lk}|\,\mathbb{I}\left\{|Z_{lk}|\ge \eta^{-1}8M\log(p+n)\right\}\\
\le& C\sqrt{n}\max_{1\le l\le n}\max_{1\le k\le q}\mathbb{E}\,|Z_{lk}|\,\mathbb{I}\left\{|Z_{lk}|\ge \eta^{-1}8M\log(p+n)\right\}\\
\le& C\sqrt{n}(p+n)^{-4}\max_{1\le l\le n}\max_{1\le k\le q}\mathbb{E}\,|Z_{lk}|\exp\left\{\eta\,|Z_{lk}|\,/\,(2M)\right\}\le C\sqrt{n}(p+n)^{-4}.
\end{aligned}$$

If (C1)-(ii) holds, then

$$\begin{aligned}
&\max_{1\le k\le q}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\mathbb{E}\,|Z_{lk}|\,\mathbb{I}\left\{|Z_{lk}|\ge \sqrt{n}/(\log p)^{8}\right\}\\
\le& C\sqrt{n}\max_{1\le l\le n}\max_{1\le k\le q}\mathbb{E}\,|Z_{lk}|\,\mathbb{I}\left\{|Z_{lk}|\ge \sqrt{n}/(\log p)^{8}\right\}\le Cn^{-\gamma_0-\epsilon/8}.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\mathrm{P}\left\{\max_{1\le k\le q}|\nu_k-\breve{\nu}_k|\ge(\log p)^{-1}\right\}\\
\le&\,\mathrm{P}\left(\max_{1\le k\le q}\max_{1\le l\le n}|Z_{lk}|\ge\tau_n\right)\\
\le& n\,\mathrm{P}\left(\max_{1\le i,j\le p}\max_{1\le s\le K}|\varepsilon_{ti}\varepsilon_{t+s,j}|\ge\tau_n\right)\\
\le& n\max_{1\le i,j\le p}\max_{1\le s\le K}\left\{\mathrm{P}(|\varepsilon_{ti}|\ge\tau_n^{1/2})+\mathrm{P}(|\varepsilon_{t+s,j}|\ge\tau_n^{1/2})\right\}=O(p^{-1}+n^{-\epsilon/8}).
\end{aligned}$$

Note that

$$\left|\max_{1\le k\le q}\nu_k^2-\max_{1\le k\le q}\breve{\nu}_k^2\right|\le 2\max_{1\le k\le q}|\breve{\nu}_k|\max_{1\le k\le q}|\nu_k-\breve{\nu}_k|+\max_{1\le k\le q}|\nu_k-\breve{\nu}_k|^2. \tag{S3.10}$$

Therefore, to prove this theorem, it is enough to show

$$\lim_{N\to\infty}\mathrm{P}\left(\max_{i\in D_N}|\breve{\nu}_i|>z\right)=1-\exp\left(-\frac{1}{\sqrt{\pi}}e^{-x/2}\right)$$

as $N\to\infty$. Then, by Bonferroni inequality,

$$\sum_{t=1}^{2k}(-1)^{t-1}\alpha_t\le \mathrm{P}\left(\max_{i\in D_N}|\breve{\nu}_i|>z\right)\le\sum_{t=1}^{2k+1}(-1)^{t-1}\alpha_t$$

for any $k\ge 1$, where

$$\alpha_t\doteq\sum{}^{*}\,\mathrm{P}\left(|\breve{\nu}_{i_1}|>z,\cdots,|\breve{\nu}_{i_t}|>z\right)$$

for $1\le t\le N$, and the sum runs over all $i_1<\cdots<i_t$ and $i_1,\cdots,i_t\in D_N$. First, we will prove that

$$\lim_{N\to\infty}\alpha_t=\frac{1}{t!}\pi^{-t/2}e^{-ty/2} \tag{S3.11}$$

for each $t\ge 1$. All the assumptions in Theorem 1.1 of Zaitsev (1987) are satisfied. Thus, we have

$$
\begin{aligned}
&\sum^{*} \mathrm{P}\left\{|Z_{i_1}| > z + \zeta_n(\log N)^{-1/2}, \cdots, |Z_{i_t}| > z + \zeta_n(\log N)^{-1/2}\right\} \\
&\quad - \begin{pmatrix} |D_N| \\ t \end{pmatrix} c_1 t^{5/2} \exp\left\{-\frac{n^{1/2}\zeta_n}{c_2 t^3(\log N)^{1/2}}\right\} \\
&\leq \sum^{*} \mathrm{P}\left\{|\breve{\nu}_{i_1}| > z, \cdots, |\breve{\nu}_{i_t}| > z\right\} \\
&\leq \sum^{*} \mathrm{P}\left\{|Z_{i_1}| > z - \zeta_n(\log N)^{-1/2}, \cdots, |Z_{i_t}| > z - \zeta_n(\log N)^{-1/2}\right\} \\
&\quad + \begin{pmatrix} |D_N| \\ t \end{pmatrix} c_1 t^{5/2} \exp\left\{-\frac{n^{1/2}\zeta_n}{c_2 t^3(\log N)^{1/2}}\right\},
\end{aligned}
$$

where $(Z_{i_1}, \cdots, Z_{i_t})^\top$ follows a multivariate normal distribution with mean zero and the same covariance matrix with $(\breve{\nu}_{i_1}, \cdots, \breve{\nu}_{i_t})^\top$. By Lemma 1, we have

$$
\begin{aligned}
&\sum^{*} \mathrm{P}\left\{|Z_{i_1}| > z + \zeta_n(\log N)^{-1/2}, \cdots, |Z_{i_t}| > z + \zeta_n(\log N)^{-1/2}\right\} \to \frac{1}{t!}\pi^{-t/2}e^{-ty/2}, \\
&\sum^{*} \mathrm{P}\left\{|Z_{i_1}| > z - \zeta_n(\log N)^{-1/2}, \cdots, |Z_{i_t}| > z - \zeta_n(\log N)^{-1/2}\right\} \to \frac{1}{t!}\pi^{-t/2}e^{-ty/2},
\end{aligned}
$$

with $\zeta_n \to 0$ and $N \to \infty$. Additionally,

$$
\begin{pmatrix} |D_N| \\ t \end{pmatrix} c_1 t^{5/2} \exp\left\{-\frac{n^{1/2}\zeta_n}{c_2 t^3(\log N)^{1/2}}\right\} \leq C \begin{pmatrix} N \\ t \end{pmatrix} t^{5/2} \exp\left\{-\frac{n^{1/2}\zeta_n}{c_2 t^3(\log N)^{1/2}}\right\} \to 0
$$

for $\zeta_n \to 0$ sufficiently slow. Thus, we have

$$\sum^{*} \mathrm{P}\left(|\breve{\nu}_{i_1}| > z, \cdots, |\breve{\nu}_{i_t}| > z\right) \to \frac{1}{t!}\pi^{-t/2}e^{-ty/2}.$$

Let $N \to \infty$, we have

$$\begin{aligned}
&\sum_{t=1}^{2k}(-1)^{t-1}\frac{1}{t!}\left(\frac{1}{\sqrt{\pi}}e^{-x/2}\right)^{t} \\
\leq &\liminf_{N\to\infty} \mathrm{P}\left(\max_{i\in D_N}|\breve{\nu}_i| > z\right) \\
\leq &\limsup_{N\to\infty} \mathrm{P}\left(\max_{i\in D_N}|\breve{\nu}_i| > z\right) \leq \sum_{t=1}^{2k+1}(-1)^{t-1}\frac{1}{t!}\left(\frac{1}{\sqrt{\pi}}e^{-x/2}\right)^{t}
\end{aligned}$$

for each $k \geq 1$. By letting $k \to \infty$ and using the Taylor expansion of the function $1 - e^{-x}$, we obtain the result. □

S3.2 Proof of Theorem 2

Define $(i_0, j_0, k_0) = \arg\max_{1\leq i<j\leq p, 1\leq k\leq K} |\rho_{ij}(k)|$. Let $\gamma_{il} = \mathbb{E}(\varepsilon_{it}^2\varepsilon_{i,t+l}^2) - \sigma_i^4$. By the condition of Theorem 2, the long-run variance $\gamma_i^L = \lim_{n\to\infty}\{\gamma_{i0} +$

$2\sum_{l=1}^{n}(1-l/n)\gamma_{il}\}$ is bounded. Hence,

$$\begin{aligned}\mathbb{E}\left(\frac{1}{n}\sum_{t=1}^{n-k_0}\varepsilon_{i_0t}^2\right) =&\sigma_{i_0}^2,\\ \mathrm{var}\left(\frac{1}{n}\sum_{t=1}^{n-k_0}\varepsilon_{i_0t}^2\right) =&\frac{1}{n}\{\mathbb{E}(\varepsilon_{i_0t}^4)-\sigma_{i_0}^4\}+\frac{2}{n^2}\sum_{s<t}\{\mathbb{E}(\varepsilon_{i_0t}^2\varepsilon_{i_0s}^2)-\sigma_{i_0}^4\}\\ =&\frac{1}{n}\gamma_{i_0}+\frac{2}{n}\sum_{l=1}^{n-1}(1-l/n)\gamma_{i_0l}\to 0.\end{aligned}$$

Thus, we have $\frac{1}{n}\sum_{t=1}^{n-k_0}\varepsilon_{i_0t}^2\xrightarrow{p}\sigma_{i_0}^2$. Similarly, we have

$$\tfrac{1}{n}\sum_{t=1}^{n-k_0}\varepsilon_{j_0t}^2\xrightarrow{p}\sigma_{j_0}^2 \text{ and } \tfrac{1}{n}\sum_{t=1}^{n-k_0}\varepsilon_{i_0t}\varepsilon_{j_0,t+k_0}\xrightarrow{p}\sigma_{i_0j_0}(k_0).$$

Thus, $\hat{\rho}_{i_0j_0}(k_0)\xrightarrow{p}\rho_{i_0j_0}(k_0)$. As $n,p\to\infty$, we have

$$\begin{aligned}&\mathrm{P}\left\{\max_{1\le k\le K}\max_{1\le i<j\le p} n\hat{\rho}_{ij}^2(k)-2\log(Kp^2)+\log\log(Kp^2)\ge q_\alpha\right\}\\ \ge&\mathrm{P}\left\{n\hat{\rho}_{i_0j_0}^2(k_0)-2\log(Kp^2)+\log\log(Kp^2)\ge q_\alpha\right\}\\ \to&\mathrm{P}\left\{n\rho_{i_0j_0}^2(k_0)-2\log(Kp^2)+\log\log(Kp^2)\ge q_\alpha\right\}=1\end{aligned}$$

by the condition $\rho_{i_0j_0}(k_0)\ge 3\sqrt{\log p/n}>\sqrt{2\log(Kp^2)/n}$. □

S3.3 Proof of Theorem 3

Define $\boldsymbol{X} = (\boldsymbol{\varepsilon}_{\cdot 1}^\top, \cdots, \boldsymbol{\varepsilon}_{\cdot p}^\top)^\top \in \mathbb{R}^d$, $d = np$ and $\boldsymbol{\varepsilon}_{\cdot i} = (\varepsilon_{1i}, \cdots, \varepsilon_{ni})^\top$. Consider the Gaussian setting and a simple alternative set of parameters

$$\mathcal{F}(\rho) \doteq \left\{ \boldsymbol{\Xi} : \boldsymbol{\Xi} = \mathrm{diag}\{\underbrace{\mathbf{I}_n, \cdots, \mathbf{I}_n}_{k-1}, \boldsymbol{\Sigma}(\rho), \mathbf{I}_n, \cdots, \mathbf{I}_n\}, 1 \le k \le p \right\},$$

where

$$\boldsymbol{\Sigma}(\rho) = \begin{pmatrix} 1 & \rho & 0 & \cdots & 0 & 0 \\ \rho & 1 & \rho & \cdots & 0 & 0 \\ 0 & \rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \rho \\ 0 & 0 & 0 & \cdots & \rho & 1 \end{pmatrix}.$$

Let μ_ρ be the uniform measure on $\mathcal{F}(\rho)$ and $\rho = c_0(\log d/n)^{1/2}$ for some small enough constant $c_0 < 1$. Let $\mathrm{pr}_{\boldsymbol{\Xi}}$ denote the probability measure of $N_d(\mathbf{0}, \boldsymbol{\Xi})$ and $\mathrm{pr}_{\mu_\rho} = \int \mathrm{pr}_{\boldsymbol{\Xi}} \mathrm{d}\mu_\rho(\boldsymbol{\Xi})$. Let pr_0 denote the probability measure of $N_d(0, \mathbf{I}_d)$. Note that, for any set A, we have

$$\sup_{\boldsymbol{\Xi} \in \mathcal{F}(\rho)} \mathrm{pr}_{\boldsymbol{\Xi}}\left(A^C\right) \ge \mathrm{pr}_{\mu_\rho}\left(A^C\right), \quad 1 = \mathrm{pr}_{\mu_\rho}\left(A^C\right) + \mathrm{pr}_{\mu_\rho}(A)$$

and

$$\mathrm{pr}_{\mu_\rho}(A) \le \mathrm{pr}_0(A) + \left|\mathrm{pr}_{\mu_\rho}(A) - \mathrm{pr}_0(A)\right|.$$

Letting $A = \{T_\alpha = 1\}$, the above equations yield

$$\begin{aligned}\inf_{T_\alpha \in \mathcal{T}_\alpha, \boldsymbol{\Xi} \in \mathcal{F}(\rho)} \sup_{\mathrm{pr}_{\boldsymbol{\Xi}}} (T_\alpha = 0) \ge & 1 - \alpha - \sup_{A:\mathrm{pr}_0(A)\le\alpha} \left|\mathrm{pr}_{\mu_\rho}(A) - \mathrm{pr}_0(A)\right| \\ \ge & 1 - \alpha - \frac{1}{2}\left\|\mathrm{pr}_{\mu_\rho} - \mathrm{pr}_0\right\|_{TV},\end{aligned}$$

where $\|\cdot\|_{TV}$ denotes the total variation norm. Setting $L_{\mu_\rho}(y) = \mathrm{dpr}_{\mu_\rho}(y)/\mathrm{dpr}_0(y)$, and by Jensen's inequality, we have

$$\begin{aligned}\left\|\mathrm{pr}_{\mu_\rho} - \mathrm{pr}_0\right\|_{TV} = & \int \left|L_{\mu_\rho}(y) - 1\right| \mathrm{dpr}_0(y) \\ = & \mathbb{E}_{\mathrm{pr}_0}\left|L_{\mu_\rho}(Y) - 1\right| \le \left[\mathbb{E}_{\mathrm{pr}_0}\left\{L^2_{\mu_\rho}(Y)\right\} - 1\right]^{1/2}.\end{aligned}$$

Therefore, as long as $\mathbb{E}_{\mathrm{pr}_0}\left\{L^2_{\mu_\rho}(Y)\right\} = 1 + o(1)$, we have

$$\inf_{T_\alpha \in \mathcal{T}_\alpha, \boldsymbol{\Xi} \in \mathcal{F}(\rho)} \sup_{\mathrm{pr}_{\boldsymbol{\Xi}}} (T_\alpha = 0) \ge 1 - \alpha - o(1) > 0.$$

We then prove that $\mathbb{E}_{\mathrm{pr}_0}\left\{L^2_{\mu_\rho}(Y)\right\} = 1 + o(1)$. By construction, we have

$$L_{\mu_\rho} = \frac{1}{p}\sum_{\boldsymbol{\Xi} \in \mathcal{F}(\rho)} \left[\frac{1}{|\boldsymbol{\Xi}|^{1/2}} \exp\left\{-\frac{1}{2} Z_{i,\cdot}^{\mathrm{T}} \left(\boldsymbol{\Omega} - \mathbf{I}_d\right) Z_{i,\cdot}\right\}\right],$$

where $\mathbf{\Omega} = \Xi^{-1}$ and $\{Z_{i,\cdot} : 1 \leq i \leq n\}$ are independent and identically distributed as $\mathcal{N}_d(0, \mathbf{I}_d)$. We have

$$\begin{aligned}&\mathbb{E}_{\mathrm{pr}_0}\left\{L^2_{\mu_\rho}(Y)\right\}\\ =&\frac{1}{p^2}\sum_{\Xi_1,\Xi_2\in\mathcal{F}(\rho)}\mathbb{E}\left[\frac{1}{|\Xi_1|^{1/2}}\frac{1}{|\Xi_2|^{1/2}}\exp\left\{-\frac{1}{2}Z_{i,\cdot}^{\mathrm{T}}\left(\mathbf{\Omega}_1+\mathbf{\Omega}_2-2I_d\right)Z_{i,\cdot}\right\}\right],\end{aligned}$$

where $\mathbf{\Omega}_i = \Xi_i^{-1}$ for $i = 1, 2$. We write

$$\begin{aligned}\mathbb{E}_{\mathrm{pr}_0}\left\{L^2_{\mu_\rho}(Y)\right\} =&\frac{p-1}{p}\mathbb{E}\left[\frac{1}{|\Xi_1|^{1/2}}\frac{1}{|\Xi_2|^{1/2}}\exp\left\{-\frac{1}{2}Z_{i,\cdot}^{\mathrm{T}}\left(\mathbf{\Omega}_1+\mathbf{\Omega}_2-2\mathbf{I}_d\right)Z_{i,\cdot}\right\}\right]\\ &+\frac{1}{p}\mathbb{E}\left[\frac{1}{|\Xi|}\exp\left\{-\frac{1}{2}Z_{i,\cdot}^{\mathrm{T}}\left(2\mathbf{\Omega}-2\mathbf{I}_d\right)Z_{i,\cdot}\right\}\right]\\ \doteq&E_1+E_2,\end{aligned}$$

where E_1 represents the set of (Ξ_1, Ξ_2) with $\Xi_1 \neq \Xi_2$, and E_2 represents the set of (Ξ_1, Ξ_2) with $\Xi_1 = \Xi_2$. By standard argument in moment generating functions of the Gaussian quadratic form, we have

$$\begin{aligned}\mathbb{E}\left\{\exp\left(-\frac{1}{2}\boldsymbol{W}^\top\mathbf{A}\boldsymbol{W}\right)\right\} =&\left[\{1+\lambda_1(\mathbf{A})\}\cdots\{1+\lambda_q(\mathbf{A})\}\right]^{-1/2}\\ =&\{\det(\mathbf{I}_q+\mathbf{A})\}^{-1/2}\end{aligned}$$

if $\boldsymbol{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$ and $\mathbf{A} \in \mathbb{R}^{q\times q}$. Without loss of generality, define $\Xi_1 = \mathrm{diag}\{\mathbf{\Sigma}(\rho), \mathbf{I}_n, \cdots, \mathbf{I}_n\}$ and $\Xi_2 = \mathrm{diag}\{\mathbf{I}_n, \mathbf{\Sigma}(\rho), \mathbf{I}_n, \cdots, \mathbf{I}_n\}$. Thus, $|\Xi_1| =$

$|\boldsymbol{\Xi}_2| = |\boldsymbol{\Sigma}(\rho)|$. Additionally, define $\boldsymbol{\Omega}_1 = \mathrm{diag}\{\boldsymbol{\Sigma}(\rho)^{-1}, \mathbf{I}_n, \cdots, \mathbf{I}_n\}$ and $\boldsymbol{\Omega}_2 = \mathrm{diag}\{\mathbf{I}_n, \boldsymbol{\Sigma}(\rho)^{-1}, \mathbf{I}_n, \cdots, \mathbf{I}_n\}$.

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{|\boldsymbol{\Xi}_1|^{1/2}}\frac{1}{|\boldsymbol{\Xi}_2|^{1/2}}\exp\left\{-\frac{1}{2}Z_{i,\cdot}^{\mathrm{T}}\left(\boldsymbol{\Omega}_1+\boldsymbol{\Omega}_2-2I_d\right)Z_{i,\cdot}\right\}\right]\\
=&\mathbb{E}\left(\frac{1}{|\boldsymbol{\Xi}_1|^{1/2}}\frac{1}{|\boldsymbol{\Xi}_2|^{1/2}}\exp\left[-\frac{1}{2}Z_{1,\cdot}^{\mathrm{T}}\left\{\boldsymbol{\Sigma}(\rho)^{-1}-\mathbf{I}_n\right\}Z_{1,\cdot}-\frac{1}{2}Z_{2,\cdot}^{\mathrm{T}}\left\{\boldsymbol{\Sigma}(\rho)^{-1}-\mathbf{I}_n\right\}Z_{2,\cdot}\right]\right)\\
=&\frac{1}{|\boldsymbol{\Sigma}(\rho)|}|\boldsymbol{\Sigma}(\rho)^{-1}-\mathbf{I}_n+\mathbf{I}_n|^{-1/2}|\boldsymbol{\Sigma}(\rho)^{-1}-\mathbf{I}_n+\mathbf{I}_n|^{-1/2}=1.
\end{aligned}
$$

Thus,

$$
E_1 = \frac{p-1}{p} = 1 + o(1) \tag{S3.12}
$$

as $p \to \infty$. Similarly,

$$
\begin{aligned}
&\frac{1}{p}\mathbb{E}\left[\frac{1}{|\boldsymbol{\Xi}|}\exp\left\{-\frac{1}{2}Z_{i,\cdot}^{\mathrm{T}}\left(2\boldsymbol{\Omega}-2I_d\right)Z_{i,\cdot}\right\}\right]\\
=&\frac{1}{p}\frac{1}{|\boldsymbol{\Sigma}(\rho)|}\mathbb{E}\left[\exp\left\{-\frac{1}{2}Z_{1,\cdot}^{\mathrm{T}}\left(2\boldsymbol{\Sigma}(\rho)^{-1}-2\mathbf{I}_n\right)Z_{1,\cdot}\right\}\right]\\
=&\frac{1}{p}\frac{1}{|\boldsymbol{\Sigma}(\rho)|}|2\boldsymbol{\Sigma}(\rho)^{-1}-\mathbf{I}_n|^{-1/2}.
\end{aligned}
$$

By (10.1) in Daniels (1956), we have $|\boldsymbol{\Sigma}(\rho)| = (1-\rho^2)^{n-1}$. Define

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Then, $\boldsymbol{\Sigma}(\rho) = \mathbf{I}_n + \rho\mathbf{A}$. By the Taylor expansion, we have $(\mathbf{I}_n + \rho\mathbf{A})^{-1} = \sum_{k=0}^{\infty}(-\rho)^k\mathbf{A}^k$ and

$$\begin{aligned} 2\boldsymbol{\Sigma}(\rho)^{-1} - \mathbf{I}_n =& 2(\mathbf{I}_n + \rho\mathbf{A})^{-1} - \mathbf{I}_n = 2(\mathbf{I}_n - \rho\mathbf{A} + \rho^2\mathbf{A}^2 - \rho^3\mathbf{A}^3 + \cdots) - \mathbf{I}_n \\ =& \mathbf{I}_n - 2\rho\mathbf{A} + 2\rho^2\mathbf{A}^2\sum_{k=0}^{\infty}(-\rho)^k\mathbf{A}^k \\ =& \mathbf{I}_n - 2\rho\mathbf{A} + 2\rho^2\mathbf{A}^2(\mathbf{I}_n + \rho\mathbf{A})^{-1}. \end{aligned}$$

Because $\rho^2\mathbf{A}^2(\mathbf{I}_n + \rho\mathbf{A})^{-1}$ is positive definite, we have

$$|2\boldsymbol{\Sigma}(\rho)^{-1} - \mathbf{I}_n| \geq |\mathbf{I}_n - 2\rho\mathbf{A}| = (1-4\rho^2)^{n-1}.$$

Thus,

$$E_2 \leq p^{-1}(1-\rho^2)^{-n+1}(1-4\rho^2)^{-(n-1)/2} = p^{-1}\exp(3c_0^2\log p)\{1+o(1)\} \to 0 \tag{S3.13}$$

if $\rho = c_0(\log p/n)^{1/2}$ and $c_0^2 < 1/3$. Combining (S3.12) and (S3.13), we have $\mathbb{E}_{\mathrm{pr}_0}\left\{L^2_{\mu_\rho}(Y)\right\} = 1 + o(1)$. Lastly, we can easily show that for $\rho = c_0(\log p/n)^{1/2}$,

$$\left[F(\boldsymbol{\varepsilon}) : \mathrm{cor}_F\{(\boldsymbol{\varepsilon}_{\cdot 1}^\top, \cdots, \boldsymbol{\varepsilon}_{\cdot p}^\top)^\top\} \in \mathcal{F}(\rho), F(\boldsymbol{\varepsilon}) \text{ is Gaussian}\right] \subset \left[F(\boldsymbol{\varepsilon}) : R\{F(\boldsymbol{\varepsilon})\} \in \mathcal{U}(c)\right],$$

where $\boldsymbol{\varepsilon} = \{\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n\}$ and $R\{F(\boldsymbol{\varepsilon})\} = \left\{\mathrm{cor}_F(\boldsymbol{\varepsilon}_{t+1}, \boldsymbol{\varepsilon}_t), \cdots, \mathrm{cor}_F(\boldsymbol{\varepsilon}_{t+K}, \boldsymbol{\varepsilon}_t)\right\}$.

Thus,

$$\inf_{T_\alpha \in \mathcal{T}_\alpha} \sup_{R\{F(\boldsymbol{\varepsilon})\} \in \mathcal{U}(c)} \mathrm{pr}_{\boldsymbol{\Xi}}(T_\alpha = 0) \geq \inf_{T_\alpha \in \mathcal{T}_\alpha} \sup_{\boldsymbol{\Xi} \in \mathcal{F}(\rho)} \mathrm{pr}_{\boldsymbol{\Xi}}(T_\alpha = 0) \geq 1 - \alpha - o(1) > 0.$$

This completes the proof. □

S3.4 Proof of Theorem 4

Firstly, we restate Lemma 2.1 in Srivastava (2009) on the quadratic forms.

Lemma 5. *Under Condition (C4), for any* $m \times m$ *symmetric matrix* $A = \{a_{ij}\}_{1 \leq i,j \leq m}$

and $B = \{b_{ij}\}_{1\le i,j\le m}$ *of constants, we have*

$$\mathbb{E}\{(\boldsymbol{z}_t^T A\boldsymbol{z}_t)^2\} = \Delta\sum_{i=1}^{m} a_{ii}^2 + 2\mathrm{tr}(A^2) + \{\mathrm{tr}(A)\}^2,$$
$$\mathrm{var}(\boldsymbol{z}_t^T A\boldsymbol{z}_t) = \Delta\sum_{i=1}^{p} a_{ii}^2 + 2\mathrm{tr}(A^2),$$
$$\mathbb{E}\{(\boldsymbol{z}_t^T A\boldsymbol{z}_t)(\boldsymbol{z}_t^T B\boldsymbol{z}_t)\} = \Delta\sum_{i=1}^{m} a_{ii}b_{ii} + 2\mathrm{tr}(AB) + \mathrm{tr}(A)\mathrm{tr}(B),$$

where $\Delta = \mathbb{E}(z_{it}^4) - 3$*.*

Now, we are ready to present the proof of Theorem 4.

Proof. Let

$$T_{\mathrm{SUM}} = \sum_{l=1}^{K}\frac{2}{n(n-1)}\sum\sum_{t<s}\boldsymbol{\varepsilon}_t^\top\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_{t+l}^\top\boldsymbol{\varepsilon}_{s+l} \doteq \sum_{l=1}^{K} T_l.$$

We will show that for each $l \in \{1, \cdots, K\}$,

$$\frac{T_l}{\sqrt{\frac{2}{n(n-1)}\mathrm{tr}^2(\boldsymbol{\Sigma}^2)}} \xrightarrow{d} \mathcal{N}(0,1). \tag{S3.14}$$

Define $V_{nj} = n^{-1}(n-1)^{-1}\sum_{i=l+1}^{j-1}\boldsymbol{\varepsilon}_{i-l}^\top\boldsymbol{\varepsilon}_{j-l}\boldsymbol{\varepsilon}_i^T\boldsymbol{\varepsilon}_j$, $j \in \{l+2, \cdots, n\}$ and $W_{nk} = \sum_{i=l+2}^{k} V_{ni}$, $k \in \{l+2, \cdots, n\}$. Let $\mathcal{F}_i \doteq \sigma\{\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_i\}$ be the σ-field generated by $\{\boldsymbol{\varepsilon}_j\}_{j\le i}$. It is easy to show that $\mathbb{E}(V_{ni}|\mathcal{F}_{i-1}) = 0$ and it follows that $\{W_{nk}, \mathcal{F}_k : l+2 \le k \le n\}$ is a zero mean martingale. Let $v_{ni} = \mathbb{E}(V_{ni}^2|\mathcal{F}_{i-1})$,

$l+2 \le i \le n$ and $V_n = \sum_{i=l+2}^{n} v_{ni}$. The central limit theorem (Hall and Hyde, 1980) will hold if we can show

$$\frac{V_n}{\mathrm{var}(W_{nn})} \xrightarrow{p} 1, \tag{S3.15}$$

and for any $\epsilon > 0$,

$$\sum_{i=l+2}^{n} n^2 \mathrm{tr}^{-2}(\boldsymbol{\Sigma}^2)\mathbb{E}\left[V_{ni}^2 \mathbb{I}\left\{|V_{ni}| > \epsilon\sqrt{n^{-2}\mathrm{tr}^2(\boldsymbol{\Sigma}^2)}\right\}|\mathcal{F}_{i-1}\right] \xrightarrow{p} 0. \tag{S3.16}$$

It can be shown that

$$\begin{aligned}
& v_{ni} \\
=& \frac{1}{n^2(n-1)^2}\left\{\sum_{j=l+1}^{i-1}(\boldsymbol{\varepsilon}_{i-l}^{\top}\boldsymbol{\varepsilon}_{j-l})^2\boldsymbol{\varepsilon}_j^T\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_j + 2\sum_{l+1\le j<k<i}\boldsymbol{\varepsilon}_{i-l}^{\top}\boldsymbol{\varepsilon}_{j-l}\boldsymbol{\varepsilon}_{i-l}^{\top}\boldsymbol{\varepsilon}_{k-l}\boldsymbol{\varepsilon}_j^T\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_k\right\}.
\end{aligned}$$

Then,

$$\begin{aligned}
& \frac{V_n}{\mathrm{var}(W_{nn})} \\
=& \frac{2}{n(n-1)\mathrm{tr}^2(\boldsymbol{\Sigma}^2)}\Big\{\sum_{i=l+2}^{n}\sum_{j=l+1}^{i-1}(\boldsymbol{\varepsilon}_{i-l}^{\top}\boldsymbol{\varepsilon}_{j-l})^2\boldsymbol{\varepsilon}_j^T\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_j \\
& \qquad + 2\sum_{i=l+2}^{n}\sum_{l+1\le j<k\le i}\boldsymbol{\varepsilon}_{i-l}^{\top}\boldsymbol{\varepsilon}_{j-l}\boldsymbol{\varepsilon}_{i-l}^{\top}\boldsymbol{\varepsilon}_{k-l}\boldsymbol{\varepsilon}_j^T\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_k\Big\} \\
\doteq& C_{n1} + C_{n2}.
\end{aligned}$$

Simple algebras lead to

$$
\begin{aligned}
&\mathbb{E}(C_{n1}) = \frac{(n-l)(n-l-1)}{n(n-1)}, \\
&\mathrm{var}(C_{n1}) = \frac{4}{n^2(n-1)^2\mathrm{tr}^4(\boldsymbol{\Sigma}^2)}\mathbb{E}\left[\sum_{i=l+2}^{n}\sum_{j=l+1}^{n-1}\left\{(\boldsymbol{\varepsilon}_{i-l}^\top\boldsymbol{\varepsilon}_{j-l})^4(\boldsymbol{\varepsilon}_j^T\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_j)^2 - \mathrm{tr}^4(\boldsymbol{\Sigma}^2)\right\}\right].
\end{aligned}
$$

By Lemma 5, we have $\mathbb{E}\left\{(\boldsymbol{\varepsilon}_j^T\boldsymbol{\Sigma}\boldsymbol{\varepsilon}_j)^2 - \mathrm{tr}^2(\boldsymbol{\Sigma}^2)\right\} = O\{\mathrm{tr}(\boldsymbol{\Sigma}^4)\}$. Next, we will show that $\mathbb{E}\left\{(\boldsymbol{\varepsilon}_{i-l}^\top\boldsymbol{\varepsilon}_{j-l})^4 - \mathrm{tr}^2(\boldsymbol{\Sigma}^2)\right\} = O\{\mathrm{tr}(\boldsymbol{\Sigma}^4)\}$. Define $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{1/2} \doteq \{\omega_{kl}\}_{1\le k,l\le p}$.

$$
\begin{aligned}
&\mathbb{E}\{(\boldsymbol{\varepsilon}_i^\top\boldsymbol{\varepsilon}_s)^4\} = \mathbb{E}\left\{(\boldsymbol{z}_i^T\boldsymbol{\Sigma}\boldsymbol{z}_s)^4\right\} = \mathbb{E}\left\{\left(\sum_{k,l=1}^{m}\sigma_{kl}z_{ik}z_{jl}\right)^4\right\} \\
&= \sum_{k,l=1}^{m}\sigma_{kl}^4\mathbb{E}(z_{ik}^4)\mathbb{E}(z_{jl}^4) + \sum_{k\neq l}^{m}\sum_{s\neq t}^{m}\sigma_{kl}^2\sigma_{st}^2\mathbb{E}(z_{ik}^2)\mathbb{E}(z_{is}^2)\mathbb{E}(z_{jl}^2)\mathbb{E}(z_{jt}^2) \\
&\quad + 2\sum_{k=1}^{m}\sum_{s\neq t}^{m}\sigma_{ks}^2\sigma_{kt}^2\mathbb{E}(z_{ik}^4)\mathbb{E}(z_{js}^2z_{jt}^2) + \sum_{k\neq l}^{m}\sum_{s\neq t}^{m}\sigma_{kl}\sigma_{kt}\sigma_{st}\sigma_{sl}\mathbb{E}(z_{ik}^2)\mathbb{E}(z_{jl}^2)\mathbb{E}(z_{is}^2)\mathbb{E}(z_{jt}^2).
\end{aligned}
$$

Note that $\mathrm{tr}^2(\mathbf{\Sigma}^2) = (\sum_{s,t} \sigma_{st}^2)^2 = \sum_{k,l,s,t} \sigma_{st}^2 \sigma_{kl}^2$ and

$$\sum_{k,l=1}^{m} \sigma_{kl}^4 \leq \left(\sum_{k,l} \sigma_{kl}^2 \right)^2,$$

$$\sum_{k=1}^{m} \sum_{s \neq t}^{m} \sigma_{ks}^2 \sigma_{kt}^2 \leq \left(\sum_{k,l} \sigma_{kl}^2 \right)^2,$$

$$\sum_{k \neq l}^{m} \sum_{s \neq t}^{m} \sigma_{kl}^2 \sigma_{st}^2 \leq \sum_{k,l,s,t} \sigma_{st}^2 \sigma_{kl}^2,$$

$$\sum_{k \neq l}^{m} \sum_{s \neq t}^{m} \sigma_{kl} \sigma_{kt} \sigma_{st} \sigma_{sl} \leq \sum_{k \neq l} \omega_{kl}^2 \leq \sum_{k,l} \omega_{kl}^2 = \mathrm{tr}(\mathbf{\Sigma}^4).$$

Thus, we have

$$\mathbb{E}\{(\boldsymbol{\varepsilon}_i^\top \boldsymbol{\varepsilon}_s)^4\} - \mathrm{tr}^2(\mathbf{\Sigma}^2) = O\{\mathrm{tr}(\mathbf{\Sigma}^4)\}. \tag{S3.17}$$

Hence, $\mathrm{var}(C_{n1}) \to 0$ due to $\mathrm{tr}(\mathbf{\Sigma}^4) = o\{\mathrm{tr}^2(\mathbf{\Sigma}^2)\}$. Then, $C_{n1} \xrightarrow{p} 1$. Similarly, $\mathbb{E}(C_{n2}) = 0$ and

$$\mathrm{var}(C_{n2}) = O(n^{-2}) \frac{\mathrm{tr}^2(\mathbf{\Sigma}^4)}{\mathrm{tr}^4(\mathbf{\Sigma}^2)} \to 0,$$

which implies $C_{n2} \xrightarrow{p} 0$. Thus, (S3.15) holds.

It remains to show (S3.16). Since

$$\mathbb{E}\left[V_{ni}^2 \mathbb{I}\left\{ |V_{ni}| > \epsilon \sqrt{n^{-2} \mathrm{tr}^2(\mathbf{\Sigma}^2)} \right\} |\mathcal{F}_{i-1} \right] \leq \mathbb{E}(V_{ni}^4 | \mathcal{F}_{i-1}) / \{\epsilon^2 n^{-2} \mathrm{tr}^2(\mathbf{\Sigma}^2)\},$$

we only need to show that

$$\sum_{i=l+2}^{n} \mathbb{E}(V_{ni}^4) = o\{n^{-4}\mathrm{tr}^4(\mathbf{\Sigma}^2)\}.$$

Note that

$$\sum_{i=l+2}^{n} \mathbb{E}(V_{ni}^4) = O(n^{-4}) \sum_{i=l+2}^{n} \mathbb{E}\left\{\left(\sum_{j=l+1}^{i-1} \boldsymbol{\varepsilon}_{i-l}^\top \boldsymbol{\varepsilon}_{j-l} \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_j\right)^4\right\},$$

which can be decomposed as $3Q + P$ with

$$Q = O(n^{-8}) \sum_{i=l+2}^{n} \sum_{s \neq t}^{i-1} \mathbb{E}\left(\boldsymbol{\varepsilon}_{i-l}^T \boldsymbol{\varepsilon}_{s-l} \boldsymbol{\varepsilon}_{s-l}^T \boldsymbol{\varepsilon}_{i-l} \boldsymbol{\varepsilon}_{i-l}^T \boldsymbol{\varepsilon}_{t-l} \boldsymbol{\varepsilon}_{t-l}^T \boldsymbol{\varepsilon}_{i-l} \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_s \boldsymbol{\varepsilon}_s^T \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T \boldsymbol{\varepsilon}_i\right),$$

$$P = O(n^{-8}) \sum_{i=l+2}^{n} \sum_{s=1}^{i-1} \mathbb{E}\left\{(\boldsymbol{\varepsilon}_{i-l}^\top \boldsymbol{\varepsilon}_{j-l})^4 (\boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_s)^4\right\}.$$

Note that $Q = O(n^{-4})\mathbb{E}^2\{(\boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i)^2\} = o\{n^{-4}\mathrm{tr}^4(\mathbf{\Sigma}^2)\}$ by Lemma 5. By (S3.17), we have $P = O\{n^{-4}\mathrm{tr}^2(\mathbf{\Sigma}^4)\} = o\{n^{-4}\mathrm{tr}^4(\mathbf{\Sigma}^2)\}$. And then (S3.16) follows immediately. This completes the proof of (S3.14). Finally, after some simple algebras, we have $\mathbb{E}(T_l T_k) = 0$ if $l \neq k$. Thus, we have

$$\frac{\sum_{l=1}^{K} T_l}{\sqrt{\frac{2K}{n(n-1)}\mathrm{tr}^2(\mathbf{\Sigma}^2)}} \xrightarrow{d} \mathcal{N}(0,1). \tag{S3.18}$$

Here, we complete the proof. □

S3.5 Proof of Proposition 1

Proof. Under H_0, due to Proposition A.2 in Chen et al. (2010) and the condition $\operatorname{tr}(\mathbf{\Sigma}^4) = o(\operatorname{tr}^2(\mathbf{\Sigma}^2))$, we have

$$\begin{aligned}
\mathbb{E}\{\widehat{\operatorname{tr}(\mathbf{\Sigma}^2)}\} =& \operatorname{tr}(\mathbf{\Sigma}^2), \\
\operatorname{var}\{\widehat{\operatorname{tr}(\mathbf{\Sigma}^2)}\} =& 4n^{-2}\operatorname{tr}^2\left(\mathbf{\Sigma}^2\right) + 8n^{-1}\operatorname{tr}\left(\mathbf{\Sigma}^4\right) + 4\Delta n^{-1}\operatorname{tr}\left(\mathbf{\Sigma}^2 \circ \mathbf{\Sigma}^2\right) \\
&+ O\left\{n^{-3}\operatorname{tr}^2\left(\mathbf{\Sigma}^2\right) + n^{-2}\operatorname{tr}\left(\mathbf{\Sigma}^4\right)\right\} = o\{\operatorname{tr}^2(\mathbf{\Sigma}^2)\},
\end{aligned}$$

where $\mathbf{A} \circ \mathbf{B} = \{a_{ij}b_{ij}\}$ for two matrix $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$. Hence, we complete the proof of this proposition. □

S3.6 Proof of Theorem 5

Proof. Recall that $\boldsymbol{\varepsilon}_t = \mathbf{A}_0\boldsymbol{z}_t + \mathbf{A}_1\boldsymbol{z}_{t-1}$ and $K = 1$. Actually,

$$\begin{aligned}
G_1 =& \frac{1}{n(n-1)}\sum_{s\neq t}^{T}\left(\mathbf{A}_0\boldsymbol{z}_s + \mathbf{A}_1\boldsymbol{z}_{s-1}\right)^\top\left(\mathbf{A}_0\boldsymbol{z}_t + \mathbf{A}_1\boldsymbol{z}_{t-1}\right) \\
&\left(\mathbf{A}_0\boldsymbol{z}_{t-1} + \mathbf{A}_1\boldsymbol{z}_{t-2}\right)^\top\left(\mathbf{A}_0\boldsymbol{z}_{s-1} + \mathbf{A}_1\boldsymbol{z}_{s-2}\right) \\
=& G(I) + G(II) + G(III),
\end{aligned}$$

where

$$
\begin{aligned}
& G(I) \\
\doteq & \frac{1}{n(n-1)} \sum_{s \neq t}^{T} \big(\boldsymbol{z}_s^\top \mathbf{A}_0^\top \mathbf{A}_0 \boldsymbol{z}_t \boldsymbol{z}_{t-1}^\top \mathbf{A}_0^\top \mathbf{A}_0 \boldsymbol{z}_{s-1} + \boldsymbol{z}_{s-1}^\top \mathbf{A}_1^\top \mathbf{A}_1 \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-2}^\top \mathbf{A}_1^\top \mathbf{A}_1 \boldsymbol{z}_{s-2} \\
& + \boldsymbol{z}_s^\top \mathbf{A}_0^\top \mathbf{A}_0 \boldsymbol{z}_t \boldsymbol{z}_{t-2}^\top \mathbf{A}_1^\top \mathbf{A}_1 \boldsymbol{z}_{s-2} + \boldsymbol{z}_{s-1}^\top \mathbf{A}_1^\top \mathbf{A}_1 \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}^\top \mathbf{A}_0^\top \mathbf{A}_0 \boldsymbol{z}_{s-1} \big) \\
& G(II) \\
\doteq & \frac{1}{n(n-1)} \sum_{s \neq t}^{T} \big(\boldsymbol{z}_s^\top \mathbf{A}_0^\top \mathbf{A}_1 \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}^\top \mathbf{A}_0^\top \mathbf{A}_0 \boldsymbol{z}_{s-1} + \boldsymbol{z}_{s-1}^\top \mathbf{A}_1^\top \mathbf{A}_0 \boldsymbol{z}_t \boldsymbol{z}_{t-1}^\top \mathbf{A}_0^\top \mathbf{A}_0 \boldsymbol{z}_{s-1} \\
& + \boldsymbol{z}_{s-1}^\top \mathbf{A}_1^\top \mathbf{A}_1 \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}^\top \mathbf{A}_0^\top \mathbf{A}_1 \boldsymbol{z}_{s-2} + \boldsymbol{z}_{s-1}^\top \mathbf{A}_1^\top \mathbf{A}_1 \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-2}^\top \mathbf{A}_1^\top \mathbf{A}_0 \boldsymbol{z}_{s-1} \\
& + \boldsymbol{z}_s^\top \mathbf{A}_0^\top \mathbf{A}_0 \boldsymbol{z}_t \boldsymbol{z}_{t-2}^\top \mathbf{A}_1^\top \mathbf{A}_0 \boldsymbol{z}_{s-1} + \boldsymbol{z}_s^\top \mathbf{A}_0^\top \mathbf{A}_0 \boldsymbol{z}_t \boldsymbol{z}_{t-1}^\top \mathbf{A}_0^\top \mathbf{A}_1 \boldsymbol{z}_{s-2} \\
& + \boldsymbol{z}_s^\top \mathbf{A}_0^\top \mathbf{A}_1 \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-2}^\top \mathbf{A}_1^\top \mathbf{A}_1 \boldsymbol{z}_{s-2} + \boldsymbol{z}_{s-1}^\top \mathbf{A}_1^\top \mathbf{A}_0 \boldsymbol{z}_t \boldsymbol{z}_{t-2}^\top \mathbf{A}_1^\top \mathbf{A}_1 \boldsymbol{z}_{s-2} \big) \\
& G(III) \\
\doteq & \frac{1}{n(n-1)} \sum_{s \neq t}^{T} \big(\boldsymbol{z}_s^\top \mathbf{A}_0^\top \mathbf{A}_1 \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}^\top \mathbf{A}_0^\top \mathbf{A}_1 \boldsymbol{z}_{s-2} + \boldsymbol{z}_{s-1}^\top \mathbf{A}_1^\top \mathbf{A}_0 \boldsymbol{z}_t \boldsymbol{z}_{t-2}^\top \mathbf{A}_1^\top \mathbf{A}_0 \boldsymbol{z}_{s-1} \\
& + \boldsymbol{z}_s^\top \mathbf{A}_0^\top \mathbf{A}_1 \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-2}^\top \mathbf{A}_1^\top \mathbf{A}_0 \boldsymbol{z}_{s-1} + \boldsymbol{z}_{s-1}^\top \mathbf{A}_1^\top \mathbf{A}_0 \boldsymbol{z}_t \boldsymbol{z}_{t-1}^\top \mathbf{A}_0^\top \mathbf{A}_1 \boldsymbol{z}_{s-2} \big) .
\end{aligned}
$$

After some tedious algebra, we have

$$
\mathbb{E}\{G(I)\} = \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_0 \tilde{\boldsymbol{\Sigma}}_1),\ \mathbb{E}\{G(II)\} = 0,\ \mathbb{E}\{G(III)\} = \frac{2}{n} \mathrm{tr}^2(\tilde{\boldsymbol{\Sigma}}_{01})
$$

and

$$
\begin{aligned}
\operatorname{var}\{G(I)\} =& \frac{2}{n^2}\operatorname{tr}^2(\tilde{\boldsymbol{\Sigma}}_0^2 + \tilde{\boldsymbol{\Sigma}}_1^2) + \frac{6}{n^2}\operatorname{tr}^2(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{\Sigma}}_1) \\
&+ \frac{4}{n}\left[2\operatorname{tr}\left(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{\Sigma}}_1\right)^2 + (\nu_4 - 3)\operatorname{tr}\left\{D^2\left(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{\Sigma}}_1\right)\right\}\right] + r_n,
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{var}\{G(II)\} =& \frac{8}{n^2}\operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_{01}^\top)\operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_0^2 + \tilde{\boldsymbol{\Sigma}}_1^2) + \frac{16}{n^2}\operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_1)\operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_0) \\
&+ \frac{16}{n^2}\operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_0 + \tilde{\boldsymbol{\Sigma}}_1)\left\{\operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_{01}^\top\tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_0) + \operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_{01}^\top\tilde{\boldsymbol{\Sigma}}_1)\right\} \\
&+ \frac{16}{n^2}\operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_{01})\left\{\operatorname{tr}\left(\tilde{\boldsymbol{\Sigma}}_0^2\tilde{\boldsymbol{\Sigma}}_{01}^\top\right) + \operatorname{tr}\left(\tilde{\boldsymbol{\Sigma}}_1^2\tilde{\boldsymbol{\Sigma}}_{01}\right) + 2\operatorname{tr}\left(\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_0\right)\right\} \\
&+ \frac{4}{n}\operatorname{tr}\left(\tilde{\boldsymbol{\Sigma}}_{01}^\top\tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_0^2 + \tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_{01}^\top\tilde{\boldsymbol{\Sigma}}_1^2 + 2\tilde{\boldsymbol{\Sigma}}_{01}^\top\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_0\right) + r_n,
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{var}\{G(III)\} =& \frac{4}{n}\operatorname{tr}\left(\tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_{01}^\top\tilde{\boldsymbol{\Sigma}}_{01}^\top\tilde{\boldsymbol{\Sigma}}_{01}\right) + \frac{12}{n^2}\operatorname{tr}^2\left(\tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_{01}^\top\right) \\
&+ \frac{16}{n^2}\operatorname{tr}\left(\tilde{\boldsymbol{\Sigma}}_{01}\right)\operatorname{tr}\left(\tilde{\boldsymbol{\Sigma}}_{01}\tilde{\boldsymbol{\Sigma}}_{01}^\top\tilde{\boldsymbol{\Sigma}}_{01}^\top\right) + r_n,
\end{aligned}
$$

$$
\operatorname{cov}\{G(I), G(III)\} = \frac{4}{n^2}\operatorname{tr}^2\left(\tilde{\boldsymbol{\Sigma}}_0\tilde{\boldsymbol{\Sigma}}_{01}\right) + \frac{4}{n^2}\operatorname{tr}^2\left(\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{\Sigma}}_{01}\right) + r_n,
$$

$$
\operatorname{cov}\{G(I), G(II)\} = r_n, \quad \operatorname{cov}\{G(II), G(III)\} = r_n.
$$

Similar to the proof of (S3.14), we can show that each element of $G(I)$, $G(II)$ and $G(III)$ is asymptotically normal distributed. Thus, we have

$$\frac{T_{\text{SUM}} - \mu_1}{\sigma_{S1}} \xrightarrow{d} \mathcal{N}(0,1).$$

□

S3.7 Proof of Theorem 6

First, we present some technical results for the proof of Theorem 6.

The following is a well-known formula for conditional distributions of multivariate normal distributions; see, for example, p.12 from Muirhead (1982).

Lemma 6. *Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ being invertible. Partition $\mathbf{X}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as*

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where $\mathbf{X}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. Set $\boldsymbol{\Sigma}_{22\cdot1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. Then $\mathbf{X}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{22\cdot1})$ and is independent of $\mathbf{X}_1$.

Lemma 7. *For any $x \in \mathbb{R}$ and $y \in \mathbb{R}$, define $A_p = \{\frac{T_{\text{SUM}}}{\sigma_S} \leq x\}$ and $l_p = (2\log N - \log\log N + y)^{1/2}$ and $B_i = \{|Z_i| > l_p\}$. Then, for each fixed integer*

h,

$$\sum_{1\le i_1<\cdots<i_h\le N}\left|\boldsymbol{P}(A_pB_{i_1}\cdots B_{i_h})-\boldsymbol{P}(A_p)\cdot\boldsymbol{P}(B_{i_1}\cdots B_{i_h})\right|\to 0$$

as $p\to\infty$.

Proof. The argument is divided into two steps.

Step 1: appealing independence from normal distributions.

Note that $(\varepsilon_{t1},\cdots,\varepsilon_{tp})^\top\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$. Take $\mathbf{X}_{t1}=(\varepsilon_{t1},\cdots,\varepsilon_{td})^\top$ and $\mathbf{X}_{t2}=(\varepsilon_{t,d+1},\cdots,\varepsilon_{tp})^\top$. First, we assume that $B_{i_1},\cdots,B_{i_h}$ only dependent on $\boldsymbol{X}_{t1}$ with $d\le 2h$. Recall the notation in Lemma 6. Write $\mathbf{X}_{t2}=\mathbf{U}_t+\mathbf{V}_t$, where $\mathbf{U}_t=\mathbf{X}_{t2}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_{t1}\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{22\cdot 1})$ and $\mathbf{V}_t=\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{X}_{t1}\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$. By Lemma 6,

$$\mathbf{U}_t\text{ and }\{\varepsilon_{t1},\cdots,\varepsilon_{td}\}\text{ are independent.}\qquad\text{(S3.19)}$$

Write

$$
\begin{aligned}
& T_{\mathrm{SUM}} \\
=& \frac{1}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \boldsymbol{\varepsilon}_t^{\top} \boldsymbol{\varepsilon}_s \boldsymbol{\varepsilon}_{t+l}^{\top} \boldsymbol{\varepsilon}_{s+l} \\
=& \frac{1}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} (\boldsymbol{X}_{t1}^{\top} \boldsymbol{X}_{s1} + \boldsymbol{X}_{t2}^{\top} \boldsymbol{X}_{s2})(\boldsymbol{X}_{t+l,1}^{\top} \boldsymbol{X}_{s+l,1} + \boldsymbol{X}_{t+l,2}^{\top} \boldsymbol{X}_{s+l,2}) \\
=& \frac{1}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} (\mathbf{U}_t^{\top} \mathbf{U}_s + \mathbf{U}_t^{\top} \boldsymbol{V}_s + \mathbf{U}_s \boldsymbol{V}_t + \boldsymbol{V}_t \boldsymbol{V}_s + \boldsymbol{X}_{t1}^{\top} \boldsymbol{X}_{s1}) \\
& \times (\mathbf{U}_{t+l}^{\top} \mathbf{U}_{s+l} + \mathbf{U}_{t+l}^{\top} \boldsymbol{V}_{s+l} + \mathbf{U}_{s+l} \boldsymbol{V}_{t+l} + \boldsymbol{V}_{t+l} \boldsymbol{V}_{s+l} + \boldsymbol{X}_{t+l,1}^{\top} \boldsymbol{X}_{s+l,1}) \\
=& \frac{1}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \mathbf{U}_t^{\top} \mathbf{U}_s \mathbf{U}_{t+l}^{\top} \mathbf{U}_{s+l} + \frac{2}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \mathbf{U}_t^{\top} \mathbf{U}_s \mathbf{U}_{t+l}^{\top} \boldsymbol{V}_{s+l} \\
& + \frac{2}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \mathbf{U}_t^{\top} \mathbf{U}_s \boldsymbol{V}_{t+l}^{\top} \boldsymbol{V}_{s+l} + \frac{2}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \mathbf{U}_t^{\top} \mathbf{U}_s \boldsymbol{X}_{t+l,1}^{\top} \boldsymbol{X}_{s+l,1} \\
& + \frac{2}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \mathbf{U}_t^{\top} \boldsymbol{V}_s \mathbf{U}_{t+l}^{\top} \mathbf{U}_{s+l} + \frac{4}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \mathbf{U}_t^{\top} \boldsymbol{V}_s \mathbf{U}_{t+l}^{\top} \boldsymbol{V}_{s+l} \\
& + \frac{2}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \mathbf{U}_t^{\top} \boldsymbol{V}_s \boldsymbol{V}_{t+l}^{\top} \boldsymbol{V}_{s+l} + \frac{2}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \mathbf{U}_t^{\top} \boldsymbol{V}_s \boldsymbol{X}_{t+l,1}^{\top} \boldsymbol{X}_{s+l,1} \\
& + \frac{2}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \boldsymbol{X}_{t1}^{\top} \boldsymbol{X}_{s1} \mathbf{U}_{t+l}^{\top} \mathbf{U}_{s+l} + \frac{4}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \boldsymbol{X}_{t1}^{\top} \boldsymbol{X}_{s1} \mathbf{U}_{t+l}^{\top} \boldsymbol{V}_{s+l} \\
& + \frac{2}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \boldsymbol{X}_{t1}^{\top} \boldsymbol{X}_{s1} \boldsymbol{V}_{t+l}^{\top} \boldsymbol{V}_{s+l} + \frac{2}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \boldsymbol{X}_{t1}^{\top} \boldsymbol{X}_{s1} \boldsymbol{X}_{t+l,1}^{\top} \boldsymbol{X}_{s+l,1} \\
\doteq& \frac{1}{n(n-1)} \sum_{l=1}^{K} \sum\sum_{t \neq s} \mathbf{U}_t^{\top} \mathbf{U}_s \mathbf{U}_{t+l}^{\top} \mathbf{U}_{s+l} + \sum_{q=1}^{11} \Theta_q \\
\doteq& S_p + R_p.
\end{aligned}
$$

Next, we will show that, for any $d \geq 1$ and $\iota > 0$, there exists $t = t_p > 0$ with $\lim_{N\to\infty} t_p = \infty$ and integer $p_0 \geq 1$ such that

$$\mathrm{P}(|\Theta_q| \geq \iota\sigma_S) \leq \frac{1}{p^t} \tag{S3.20}$$

as $p \geq p_0$. Here we only consider Θ_1. The proof of the other parts are similar to Θ_1.

By the decomposition, we know that $\{\boldsymbol{V}_t\}_{t=1}^n$ is independent of $\{\mathbf{U}_t\}_{t=1}^n$. Thus, conditional on $\{\mathbf{U}_t\}_{t=1}^n$, Θ_1 has the normal distribution. Hence, we have

$$\begin{aligned}
&\mathrm{P}(|\Theta_1| \geq \iota\sigma_S) \leq \sum_{l=1}^K \mathrm{P}\left(\left|\frac{2}{n(n-1)}\sum\sum_{t\neq s}\mathbf{U}_t^\top\mathbf{U}_s\mathbf{U}_{t+l}^\top\boldsymbol{V}_{s+l}\right| \geq \iota\sigma_S/K\right)\\
&= \sum_{l=1}^K \mathbb{E}\left(\mathbb{E}\left[\mathbb{I}\left\{\left|\frac{2}{n(n-1)}\sum\sum_{t\neq s}\mathbf{U}_t^\top\mathbf{U}_s\mathbf{U}_{t+l}^\top\boldsymbol{V}_{s+l}\right| \geq \iota\sigma_S/K\right\}|\{\mathbf{U}_t\}_{t=1}^n\right]\right)\\
&= 2\sum_{l=1}^K \mathbb{E}\left\{1-\Phi\left(\frac{\iota\sigma_S}{K\tilde{\sigma}_l}\right)\right\} \simeq \sum_{l=1}^K \mathbb{E}\left\{\frac{2}{\sqrt{2\pi}K^{-1}\iota\tilde{\sigma}_l^{-1}\sigma_S}e^{-(K^{-1}\iota\tilde{\sigma}_l^{-1}\sigma_S)^2/2}\right\},
\end{aligned}$$

where

$$\tilde{\sigma}_l^2 = \frac{4}{n^2(n-1)^2}\sum_{t,m\neq s}\mathbf{U}_t^\top\mathbf{U}_s\mathbf{U}_{t+l}^\top\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{U}_{m+l}\mathbf{U}_m^\top\mathbf{U}_s.$$

Here, $a_n \simeq b_n$ denotes that $a_n/b_n \to 1$. Similar to the proof of Proposition 1, we have

$$\frac{\tilde{\sigma}_l^2}{\frac{4}{n(n-1)}\operatorname{tr}(\boldsymbol{\Sigma}_{22\cdot 1}^2)\operatorname{tr}(\boldsymbol{\Sigma}_{22\cdot 1}\cdot\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})}\xrightarrow{p}1.$$

Thus,

$$K^{-1}\iota\tilde{\sigma}_l^{-1}\sigma_S\xrightarrow{p}\frac{\iota}{K}\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}^{1/2}(\boldsymbol{\Sigma}_{22\cdot 1}^2)\operatorname{tr}^{1/2}(\boldsymbol{\Sigma}_{22\cdot 1}\cdot\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})}$$

and

$$\begin{aligned}
&\operatorname{tr}\left\{\boldsymbol{\Sigma}_{22\cdot 1}\cdot\left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right)\right\}\leq\lambda_{\max}\left(\boldsymbol{\Sigma}_{22\cdot 1}\right)\cdot\operatorname{tr}\left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right)\\
\leq&\lambda_{\max}(\boldsymbol{\Sigma})\cdot\operatorname{tr}\left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right)\leq\lambda_{\max}(\boldsymbol{\Sigma})\lambda_{\max}\left(\boldsymbol{\Sigma}_{11}^{-1}\right)\cdot\operatorname{tr}\left(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{21}\right)\\
=&\lambda_{\max}(\boldsymbol{\Sigma})\frac{1}{\lambda_{\min}\left(\boldsymbol{\Sigma}_{11}\right)}\cdot\operatorname{tr}\left(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{21}\right)\leq dM_p\lambda_{\max}(\boldsymbol{\Sigma}).
\end{aligned}$$

In fact, in the above, we use the assertion $\lambda_{\min}\left(\boldsymbol{\Sigma}_{11}\right)$ is bounded by Condition (C2) and the fact that

$$\operatorname{tr}\left(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{21}\right)=\sum_{i=1}^{d}\sum_{j=d+1}^{p}\sigma_{ij}^2\leq dM_p.$$

Additionally, $\boldsymbol{\Sigma}_{22}=\boldsymbol{\Sigma}_{22\cdot 1}+\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ and all three matrices are non-negative definite. Hence, we have $\mathrm{tr}(\mathbf{\Sigma}^2) \geq \mathrm{tr}(\mathbf{\Sigma}_{22\cdot 1}^2)$. Thus,

$$\frac{\iota}{K}\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}^{1/2}(\mathbf{\Sigma}_{22\cdot 1}^2)\mathrm{tr}^{1/2}(\mathbf{\Sigma}_{22\cdot 1}\cdot \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})} \geq \frac{\iota}{K}\frac{\mathrm{tr}^{1/2}(\mathbf{\Sigma}^2)}{\sqrt{dM_p\lambda_{\max}(\mathbf{\Sigma})}}.$$

Then,

$$\begin{aligned}
&\sum_{l=1}^{K}\mathbb{E}\left\{\frac{2}{\sqrt{2\pi}K^{-1}\iota\tilde{\sigma}_l^{-1}\sigma_S}e^{-(K^{-1}\iota\tilde{\sigma}_l^{-1}\sigma_S)^2/2}\right\}\\
\leq &K\frac{1}{\sqrt{2\pi}}\frac{2}{\frac{\iota}{K}\frac{\mathrm{tr}^{1/2}(\mathbf{\Sigma}^2)}{\sqrt{dM_p\lambda_{\max}(\mathbf{\Sigma})}}}\exp\left\{-\frac{\iota^2}{K^2}\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{dM_p\lambda_{\max}(\mathbf{\Sigma})}\right\} \leq p^{-t_p}
\end{aligned}$$

by Condition (C6) with $t_p = \frac{\iota^2}{2K^2}(\log p)^{\gamma-1}$. Note that $\mathrm{tr}(\mathbf{\Sigma}_{11}^2) \leq C_d$ is bounded.

Similarly, it can be proved that for the other parts $\Theta_2, \cdots, \Theta_{11}$,

$$\begin{aligned}
&\mathrm{P}(|\Theta_2| \geq \iota\sigma_S)\\
\leq &K\frac{1}{\sqrt{2\pi}}\frac{2}{\frac{\iota}{K}\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}^{1/2}(\mathbf{\Sigma}_{22\cdot 1}^2)\mathrm{tr}^{1/2}\left\{(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^2\right\}}}\exp\left[-\frac{\iota^2}{K^2}\frac{\mathrm{tr}^2(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma}_{22\cdot 1}^2)\mathrm{tr}\left\{(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^2\right\}}\right]\\
\leq &K\frac{1}{\sqrt{2\pi}}\frac{2}{\frac{\iota}{K}\frac{\mathrm{tr}^{1/2}(\mathbf{\Sigma}^2)}{dM_p}}\exp\left\{-\frac{\iota^2}{K^2}\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{d^2M_p^2}\right\} \leq p^{-t_p},
\end{aligned}$$

$$\begin{aligned}
&\mathrm{P}(|\Theta_3| \geq \iota\sigma_S) \\
\leq& K\frac{1}{\sqrt{2\pi}}\frac{2}{\frac{\iota}{K}\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}^{1/2}(\mathbf{\Sigma}_{22\cdot 1}^2)\mathrm{tr}^{1/2}(\mathbf{\Sigma}_{11}^2)}}\exp\left\{-\frac{\iota^2}{K^2}\frac{\mathrm{tr}^2(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma}_{22\cdot 1}^2)\mathrm{tr}(\mathbf{\Sigma}_{11}^2)}\right\} \\
\leq& K\frac{1}{\sqrt{2\pi}}\frac{2}{\frac{\iota}{K}\frac{\mathrm{tr}^{1/2}(\mathbf{\Sigma}^2)}{C_d^{1/2}}}\exp\left\{-\frac{\iota^2}{K^2}\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{C_d}\right\} \leq p^{-t_p},
\end{aligned}$$

$$\mathrm{P}(|\Theta_4| \geq \iota\sigma_S) \simeq \mathrm{P}(|\Theta_1| \geq \iota\sigma_S), \;\; \mathrm{P}(|\Theta_5| \geq \iota\sigma_S) \simeq \mathrm{P}(|\Theta_2| \geq \iota\sigma_S),$$

$$\begin{aligned}
&\mathrm{P}(|\Theta_6| \geq \iota\sigma_S) \\
\leq& K\frac{1}{\sqrt{2\pi}}\frac{2}{\frac{\iota}{K}\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}^{1/2}(\mathbf{\Sigma}_{22\cdot 1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})\mathrm{tr}^{1/2}\left\{(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^2\right\}}} \\
&\times \exp\left[-\frac{\iota^2}{K^2}\frac{\mathrm{tr}^2(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma}_{22\cdot 1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})\mathrm{tr}\left\{(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^2\right\}}\right] \\
\leq& K\frac{1}{\sqrt{2\pi}}\frac{2}{\frac{\iota}{K}\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\left\{d^{3/2}K_p^{3/2}\lambda_{\max}(\mathbf{\Sigma})\right\}^{1/2}}}\exp\left\{-\frac{\iota^2}{K^2}\frac{\mathrm{tr}^2(\mathbf{\Sigma}^2)}{d^3M_p^3\lambda_{\max}(\mathbf{\Sigma})}\right\} \leq p^{-t_p},
\end{aligned}$$

$$\begin{aligned}
&\mathrm{P}(|\Theta_7| \geq \iota\sigma_S) \\
\leq& K\frac{1}{\sqrt{2\pi}}\frac{2}{\frac{\iota}{K}\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}^{1/2}(\mathbf{\Sigma}_{22\cdot 1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})\mathrm{tr}^{1/2}(\mathbf{\Sigma}_{11}^2)}}\exp\left\{-\frac{\iota^2}{K^2}\frac{\mathrm{tr}^2(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma}_{22\cdot 1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})\mathrm{tr}(\mathbf{\Sigma}_{11}^2)}\right\} \\
\leq& K\frac{1}{\sqrt{2\pi}}\frac{2}{\frac{\iota}{K}\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\{C_d dM_p\lambda_{\max}(\mathbf{\Sigma})\}^{1/2}}}\exp\left\{-\frac{\iota^2}{K^2}\frac{\mathrm{tr}^2(\mathbf{\Sigma}^2)}{C_d dM_p\lambda_{\max}(\mathbf{\Sigma})}\right\} \leq p^{-t_p},
\end{aligned}$$

$$\mathrm{P}(|\Theta_8| \geq \iota\sigma_S) \simeq \mathrm{P}(|\Theta_3| \geq \iota\sigma_S), \;\; \mathrm{P}(|\Theta_9| \geq \iota\sigma_S) \simeq \mathrm{P}(|\Theta_7| \geq \iota\sigma_S),$$

$$\begin{aligned} &\mathrm{P}(|\Theta_{10}| \geq \iota\sigma_S) \\ \leq &K \frac{1}{\sqrt{2\pi}} \frac{2}{\frac{\iota}{K} \frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}^{1/2}\left\{(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^2\right\}\mathrm{tr}^{1/2}(\mathbf{\Sigma}_{11}^2)}} \exp\left[-\frac{\iota^2}{K^2} \frac{\mathrm{tr}^2(\mathbf{\Sigma}^2)}{\mathrm{tr}\left\{(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^2\right\}\mathrm{tr}(\mathbf{\Sigma}_{11}^2)}\right] \\ \leq &K \frac{1}{\sqrt{2\pi}} \frac{2}{\frac{\iota}{K} \frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{C_d^{1/2} d M_p}} \exp\left\{-\frac{\iota^2}{K^2} \frac{\mathrm{tr}^2(\mathbf{\Sigma}^2)}{C_d d^2 M_p^2}\right\} \leq p^{-t_p}, \end{aligned}$$

$$\begin{aligned} \mathrm{P}(|\Theta_{11}| \geq \iota\sigma_S) \leq &K \frac{1}{\sqrt{2\pi}} \frac{2}{\frac{\iota}{K} \frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma}_{11}^2)}} \exp\left\{-\frac{\iota}{K} \frac{\mathrm{tr}^2(\mathbf{\Sigma}^2)}{\mathrm{tr}^2(\mathbf{\Sigma}_{11}^2)}\right\} \\ \leq &K \frac{1}{\sqrt{2\pi}} \frac{2}{\frac{\iota}{K} \frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{C_d^{1/2}}} \exp\left\{-\frac{\iota^2}{K^2} \frac{\mathrm{tr}^2(\mathbf{\Sigma}^2)}{C_d}\right\} \leq p^{-t_p}. \end{aligned}$$

Thus, we have

$$\mathrm{P}(|R_p| \geq \iota\sigma_S) \leq \frac{1}{p^{t_p}}. \tag{S3.21}$$

Now, for clarity, we will revise the definition of A_p as follows

$$A_p(x) = \left\{\frac{T_{\mathrm{SUM}}}{\sigma_S} \leq x\right\}, \;\; x \in \mathbb{R},$$

for $p \geq 1$. Due to the fact that $T_{\text{SUM}} = S_p + R_p$, we see that

$$\begin{aligned} \mathrm{P}\{A_p(x)B_{i_1}\cdots B_{i_h}\} &\leq \mathrm{P}\Big\{A_p(x)B_{i_1}\cdots B_{i_h},\ \frac{|R_p|}{\sigma_S} < \iota\Big\} + \frac{1}{p^t} \\ &\leq \mathrm{P}\Big(\frac{S_p}{\sigma_S} \leq x + \iota,\ B_{i_1}\cdots B_{i_h}\Big) + \frac{1}{p^t} \\ &= \mathrm{P}\Big(\frac{S_p}{\sigma_S} \leq x + \iota\Big) \cdot P\big(B_{i_1}\cdots B_{i_h}\big) + \frac{1}{p^t} \end{aligned}$$

by the independence appeared in (S3.19). Hence,

$$\begin{aligned} \mathrm{P}\Big(\frac{S_p}{\sigma_S} \leq x + \iota\Big) &\leq \mathrm{P}\Big(\frac{S_p}{\sigma_S} \leq x + \iota,\ \frac{|R_p|}{\sigma_S} < \iota\Big) + \frac{1}{p^t} \\ &\leq \mathrm{P}\Big\{\frac{1}{\sigma_S}(S_p + R_p) \leq x + 2\iota\Big\} + \frac{1}{p^t} \leq \mathrm{P}\big\{A_p(x + 2\iota)\big\} + \frac{1}{p^t}. \end{aligned}$$

Combine the two inequalities to get

$$\mathrm{P}\{A_p(x)B_{i_1}\cdots B_{i_h}\} \leq \mathrm{P}\left\{A_p(x + 2\iota)\right\} \cdot \mathrm{P}\left(B_{i_1}\cdots B_{i_h}\right) + \frac{2}{p^t}. \qquad \text{(S3.22)}$$

Similarly,

$$\begin{aligned} &\mathrm{P}\Big(\frac{S_p}{\sigma_S} \leq x - \iota,\ B_{i_1}\cdots B_{i_h}\Big) \\ \leq\ &\mathrm{P}\Big(\frac{S_p}{\sigma_S} \leq x - \iota, B_{i_1}\cdots B_{i_h}, \frac{|R_p|}{\sigma_S} < \iota\Big) + \frac{1}{p^t} \leq \mathrm{P}\Big(\frac{S_p}{\sigma_S} \leq x,\ B_{i_1}\cdots B_{i_h}\Big) + \frac{1}{p^t}. \end{aligned}$$

By the independence from (S3.19),

$$\mathrm{P}\{A_p(x)B_{i_1}\cdots B_{i_h}\} \geq \mathrm{P}\Big(\frac{S_p}{\sigma_S} \leq x - \iota\Big) \cdot \mathrm{P}(B_{i_1}\cdots B_{i_h}) - \frac{1}{p^t}.$$

Furthermore,

$$\mathrm{P}\Big(\frac{T_{\mathrm{SUM}}}{\sigma_S} \leq x - 2\iota\Big) \leq \mathrm{P}\Big(\frac{T_{\mathrm{SUM}}}{\sigma_S} \leq x - 2\iota,\ \frac{|R_p|}{\sigma_S} < \iota\Big) + \frac{1}{p^t} \leq \mathrm{P}\Big(\frac{S_p}{\sigma_S} \leq x - \iota\Big) + \frac{1}{p^t},$$

where the fact $T_{\mathrm{SUM}} = S_p + R_p$ is used again. Combining the above two inequalities, we get

$$\mathrm{P}\{A_p(x)B_{i_1}\cdots B_{i_h}\} \geq \mathrm{P}\{A_p(x - 2\iota)\} \cdot \mathrm{P}(B_{i_1}\cdots B_{i_h}) - \frac{2}{p^t}.$$

This together with (S3.22) concludes

$$|\mathrm{P}\{A_p(x)B_{i_1}\cdots B_{i_h}\} - \mathrm{P}\{A_p(x)\} \cdot \mathrm{P}(B_{i_1}\cdots B_{i_h})| \leq \Delta_{p,\iota} \cdot \mathrm{P}(B_{i_1}\cdots B_{i_h}) + \frac{2}{p^t} \tag{S3.23}$$

as $p \geq p_0$, where

$$\begin{aligned} \Delta_{p,\iota} &\doteq |\mathrm{P}\{A_p(x)\} - \mathrm{P}\{A_p(x + 2\iota)\}| + |\mathrm{P}\{A_p(x)\} - \mathrm{P}\{A_p(x - 2\iota)\}| \\ &= \mathrm{P}\{A_p(x + 2\iota)\} - \mathrm{P}\{A_p(x - 2\iota)\} \end{aligned}$$

since $\mathrm{P}\{A_p(x)\}$ is increasing in $x \in \mathbb{R}$. An important observation is that the derivation of (S3.20) is based on three key facts: inequality (S3.20), the identity $T_{\mathrm{SUM}} = S_p + R_p$ and the fact $\mathbf{U}_t$ and $\{\varepsilon_{t1}, \cdots, \varepsilon_{td}\}$ are independent from (S3.19).

Thus, the three corresponding key facts aforementioned also hold for the quantities related to $\Lambda = \{i_1, \cdots, i_h\}$. Therefore, similar to the derivation of (S3.23), we have

$$\begin{aligned} & \left|\mathrm{P}\{A_p(x)B_{i_1}\cdots B_{i_h}\} - \mathrm{P}\{A_p(x)\}\cdot \mathrm{P}(B_{i_1}\cdots B_{i_h})\right| \\ \leq \; & \Delta_{p,\iota}\cdot \mathrm{P}(B_{i_1}\cdots B_{i_h}) + \frac{2}{p^t} \end{aligned}$$

as $p \geq p_0$. As a result,

$$\begin{aligned} \zeta(N,h) \doteq & \sum_{1\leq i_1<\cdots<i_h\leq N} \left|\mathrm{P}\{A_p(x)B_{i_1}\cdots B_{i_h}\} - \mathrm{P}\{A_p(x)\}\cdot \mathrm{P}(B_{i_1}\cdots B_{i_h})\right| \\ \leq & \sum_{1\leq i_1<\cdots<i_h\leq N} \left\{\Delta_{p,\iota}\cdot \mathrm{P}(B_{i_1}\cdots B_{i_h}) + \frac{2}{p^t}\right\} \\ \leq & \Delta_{p,\iota}\cdot H(h,N) + \binom{N}{2h}\cdot\frac{2}{p^t}, \end{aligned} \tag{S3.24}$$

where

$$H(h,N) \doteq \sum_{1\leq i_1<\cdots<i_h\leq N} \mathrm{P}(B_{i_1}\cdots B_{i_h}).$$

In the following, we will show $\lim_{\iota\downarrow 0}\limsup_{p\to\infty}\Delta_{p,\iota}=0$ and $\limsup_{p\to\infty}H(h,N)<\infty$ for each $d\geq 1$. Assuming these are true, by using $\binom{N}{2h}\leq p^{4h}$ and (S3.24), for fixed $h\geq 1$, sending $p\to\infty$ first, then sending $\iota\downarrow 0$, we get $\lim_{p\to\infty}\zeta(N,h)=0$ for each $d\geq 1$. The proof is then completed.

Step 2: the proofs of "$\lim_{\iota\downarrow 0}\limsup_{p\to\infty}\Delta_{p,\iota}=0$" *and* "$\limsup_{p\to\infty}H(h,N)<\infty$ *for each* $h\geq 1$".

Under Condition (C5), Theorem 4 holds and we have

$$\frac{T_{\rm SUM}}{\sigma_S}\to\mathcal{N}(0,1)\text{ weakly} \tag{S3.25}$$

as $p\to\infty$ and hence

$$\Delta_{p,\iota}\to\Phi(x+2\iota)-\Phi(x-2\iota) \tag{S3.26}$$

as $p\to\infty$, where $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}\,dt$. This implies that

$$\lim_{\iota\downarrow 0}\limsup_{p\to\infty}\Delta_{p,\iota}=0.$$

Second, under the normality assumption, by Conditions (C2) and (C3), Theorem 1 holds. Hence, by identifying "$H(t,p)$" here as "α_t" in (S3.11) for each

$t \geq 1$, we obtain

$$\lim_{p \to \infty} H(h, N) = \frac{1}{h!} \pi^{-h/2} e^{-hx/2} \tag{S3.27}$$

for each $d \geq 1$. The proof is finished. □

Now, we are ready to present the proof of Theorem 6.

Proof. By Conditions (C2), (C3) and (C5), Theorems 1 and 4 hold. By Theorem 4,

$$\mathrm{P}\Big(\frac{T_{\mathrm{SUM}}}{\sigma_S} \leq x\Big) = \Phi(x) \tag{S3.28}$$

as $p \to \infty$ for any $x \in \mathbb{R}$, where $\sigma_S = \{2Kn^{-2}\mathrm{tr}^2(\boldsymbol{\Sigma}^2)\}^{1/2}$ and $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$. Define $N = Kp^2$ and $Z_{i+(j-1)p+(k-1)p^2} = n^{1/2}\tilde{\rho}_{ij}(k)$, $i, j = 1, \cdots, p,\ k = 1, \cdots, K$. From Theorem 1, we have

$$\mathrm{P}\,\big(\max_{1 \leq i \leq N} Z_i^2 - 2\log N + \log\log N \leq y\big) \to G(y) = \exp\Big(-\frac{1}{\sqrt{\pi}} e^{-y/2}\Big) \tag{S3.29}$$

as $N \to \infty$ for any $y \in \mathbb{R}$. To show asymptotic independence, it is enough to prove

$$\lim_{N \to \infty} \mathrm{P}\Big(\frac{T_{\mathrm{SUM}}}{\sigma_S} \leq x,\ \max_{1 \leq i \leq N} Z_i^2 - 2\log N + \log\log N \leq y\Big) = \Phi(x) \cdot G(y)$$

for any $x \in \mathbb{R}$ and $y \in \mathbb{R}$. Set

$$L_N = \max_{1 \le i \le N} |Z_i| \text{ and } l_N = (2 \log N - \log \log N + y)^{1/2}, \tag{S3.30}$$

where the latter one makes sense for large N. Because of (S3.28), the above is equivalent to that

$$\lim_{N \to \infty} \mathrm{P}\Big(\frac{T_{\mathrm{SUM}}}{\sigma_S} \le x,\ L_N > l_N\Big) = \Phi(x) \cdot \{1 - G(y)\} \tag{S3.31}$$

for any $x \in \mathbb{R}$ and $y \in \mathbb{R}$. Recalling the notation in Lemma 7, we have

$$A_p = \Big\{\frac{T_{\mathrm{SUM}}}{\sigma_S} \le x\Big\} \quad \text{and} \quad B_i = \big\{|Z_i| > l_N\big\} \tag{S3.32}$$

for $1 \le i \le N$. Therefore,

$$\mathrm{P}\Big(\frac{T_{\mathrm{SUM}}}{\sigma_S} \le x,\ L_N > l_N\Big) = \mathrm{P}\Big(\bigcup_{i=1}^{N} A_p B_i\Big). \tag{S3.33}$$

From the inclusion-exclusion principle,

$$\mathrm{P}\Big(\bigcup_{i=1}^{N} A_p B_i\Big) \le \sum_{1 \le i_1 \le N} \mathrm{P}(A_p B_{i_1}) - \sum_{1 \le i_1 < i_2 \le N} \mathrm{P}(A_p B_{i_1} B_{i_2}) + \cdots + \sum_{1 \le i_1 < \cdots < i_{2k+1} \le N} \mathrm{P}(A_p B_{i_1} \cdots B_{i_{2k+1}}) \tag{S3.34}$$

and

$$\mathrm{P}\Big(\bigcup_{i=1}^{N} A_p B_i\Big) \geq \sum_{1\leq i_1\leq N} \mathrm{P}(A_p B_{i_1}) - \sum_{1\leq i_1<i_2\leq N} \mathrm{P}(A_p B_{i_1} B_{i_2}) + \cdots - \sum_{1\leq i_1<\cdots<i_{2k}\leq N} \mathrm{P}(A_p B_{i_1}\cdots B_{i_{2k}}) \tag{S3.35}$$

for any integer $k \geq 1$. Define

$$H(N,d) = \sum_{1\leq i_1<\cdots<i_d\leq N} \mathrm{P}(B_{i_1}\cdots B_{i_d})$$

for $d \geq 1$. From (S3.27) we know

$$\lim_{d\to\infty} \limsup_{p\to\infty} H(N,d) = 0. \tag{S3.36}$$

Set

$$\zeta(N,d) = \sum_{1\leq i_1<\cdots<i_d\leq N} \Big\{\mathrm{P}(A_p B_{i_1}\cdots B_{i_d}) - \mathrm{P}(A_p)\cdot \mathrm{P}(B_{i_1}\cdots B_{i_d})\Big\}$$

for $d \geq 1$. By Lemma 7,

$$\lim_{p\to\infty} \zeta(N,d) = 0 \tag{S3.37}$$

for each $d \geq 1$. The assertion (S3.34) implies that

$$\begin{aligned}\mathrm{P}\Big(\bigcup_{i=1}^{N} A_p B_i\Big) \leq & \mathrm{P}(A_p)\Big\{\sum_{1\leq i_1\leq N}\mathrm{P}(B_{i_1}) - \sum_{1\leq i_1<i_2\leq N}\mathrm{P}(B_{i_1}B_{i_2}) + \cdots - \\ & \sum_{1\leq i_1<\cdots<i_{2k}\leq N}\mathrm{P}(B_{i_1}\cdots B_{i_{2k}})\Big\} + \sum_{d=1}^{2k}\zeta(N,d) + H(N,2k+1) \\ \leq & \mathrm{P}(A_p)\cdot P\Big(\bigcup_{i=1}^{N} B_i\Big) + \sum_{d=1}^{2k}\zeta(N,d) + H(N,2k+1), \quad \text{(S3.38)}\end{aligned}$$

where the inclusion-exclusion formula is used again in the last inequality, that is,

$$\begin{aligned}\mathrm{P}\Big(\bigcup_{i=1}^{N} B_i\Big) \;\; \geq \;\; & \sum_{1\leq i_1\leq N}\mathrm{P}(B_{i_1}) - \sum_{1\leq i_1<i_2\leq N}\mathrm{P}(B_{i_1}B_{i_2}) \\ & + \cdots - \sum_{1\leq i_1<\cdots<i_{2k}\leq N}\mathrm{P}(B_{i_1}\cdots B_{i_{2k}})\end{aligned}$$

for all $k \geq 1$. By the definition of l_N and (S3.29),

$$\mathrm{P}\Big(\bigcup_{i=1}^{N} B_i\Big) = \mathrm{P}\big(L_N > l_N\big) = \mathrm{P}\big(L_N^2 - 2\log N + \log\log N > y\big) \to 1 - G(y)$$

as $p \to \infty$. By (S3.28), $\mathrm{P}(A_p) \to \Phi(x)$ as $p \to \infty$. From (S3.33), (S3.37) and (S3.38), by fixing k first and sending $p \to \infty$, we obtain that

$$\limsup_{p\to\infty}\mathrm{P}\Big(\frac{T_{\mathrm{SUM}}}{\sigma_S} \leq x,\ L_N > l_N\Big) \leq \Phi(x)\{1 - G(y)\} + \lim_{p\to\infty} H(N,2k+1).$$

Now, by letting $k \to \infty$ and using (S3.36) we have

$$\limsup_{p\to\infty} \mathrm{P}\Big(\frac{T_{\mathrm{SUM}}}{\sigma_S} \le x,\ L_N > l_N\Big) \le \Phi(x)\{1 - G(y)\}. \tag{S3.39}$$

By applying the same argument to (S3.35), we see that the counterpart of (S3.38) becomes

$$\begin{aligned}
\mathrm{P}\Big(\bigcup_{i=1}^{N} A_p B_i\Big) \ &\ge\ \mathrm{P}(A_p)\Big\{\sum_{1\le i_1\le N} \mathrm{P}(B_{i_1}) - \sum_{1\le i_1<i_2\le N} \mathrm{P}(B_{i_1}B_{i_2}) + \cdots + \\
&\qquad \sum_{1\le i_1<\cdots<i_{2k-1}\le N} \mathrm{P}(B_{i_1}\cdots B_{i_{2k-1}})\Big\} + \sum_{d=1}^{2k-1}\zeta(N,d) - H(N,2k) \\
&\ge\ \mathrm{P}(A_p)\cdot \mathrm{P}\Big(\bigcup_{i=1}^{N} B_i\Big) + \sum_{d=1}^{2k-1}\zeta(N,d) - H(N,2k),
\end{aligned}$$

where in the last step we use the inclusion-exclusion principle such that

$$\begin{aligned}
\mathrm{P}\Big(\bigcup_{i=1}^{N} B_i\Big) \ &\le\ \sum_{1\le i_1\le N} \mathrm{P}(B_{i_1}) - \sum_{1\le i_1<i_2\le N} \mathrm{P}(B_{i_1}B_{i_2}) \\
&\qquad + \cdots + \sum_{1\le i_1<\cdots<i_{2k-1}\le N} \mathrm{P}(B_{i_1}\cdots B_{i_{2k-1}})
\end{aligned}$$

for all $k \ge 1$. Review (S3.33) and repeat the earlier procedure to see

$$\liminf_{p\to\infty} \mathrm{P}\Big(\frac{T_{\mathrm{SUM}}}{\sigma_S} \le x,\ L_N > l_N\Big) \ge \Phi(x)\{1 - G(y)\} \tag{S3.40}$$

by sending $p \to \infty$ and then sending $k \to \infty$. Here (S3.40) and (S3.39) yield

(S3.31). The proof is then completed. □

S3.8 Proof of Theorem 7

Proof. By the Slutsky's Theorem, we only need to show that

$$\mathrm{P}\left\{\max_{1\le k\le K}\max_{1\le i,j\le p} n\tilde{\rho}_{ij}^2(k) - 2\log(Kp^2) + \log\log(Kp^2) \le x, T_{\mathrm{SUM}}/\sigma_S \le y\right\}$$
$$\to G(x)\cdot\Phi(y), \tag{S3.41}$$

Define

$$W(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n) = \frac{1}{n(n-1)\sigma_S}\sum_{l=1}^{K}\sum\sum_{t\ne s}\boldsymbol{x}_t^\top\boldsymbol{x}_s\boldsymbol{x}_{t+l}^\top\boldsymbol{x}_{s+l}.$$

Hence, $T_{\mathrm{SUM}}/\sigma_S = W(\boldsymbol{\varepsilon}_1,\cdots,\boldsymbol{\varepsilon}_n)$.

For $z = (z_1,\ldots,z_p)' \in \mathbb{R}^p$, consider the function

$$F_\beta(z) \doteq \beta^{-1}\log\left(\sum_{j=1}^{p}\exp(\beta z_j)\right),$$

where $\beta > 0$ is the smoothing parameter that controls the level of approximation. An elementary calculation shows that for all $z \in \mathbb{R}^p$,

$$0 \le F_\beta(z) - \max_{1\le j\le p} z_j \le \beta^{-1}\log p.$$

In the following, we define $\beta = n^{1/12}\log p$. Define

$$V(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) = \beta^{-1}\log\left[\sum_{k=1}^{K}\sum_{1\le i,j\le p}\exp\left\{\beta n^{1/2}\sigma_i^{-1}\sigma_j^{-1}\left(n^{-1}\sum_{t=1}^{n-k}x_{t+k,i}x_{tj}\right)\right\}\right].$$

Then, $V(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n) = \beta^{-1}\log\left[\sum_{k=1}^{K}\sum_{1\le i,j\le p}\exp\left\{\beta n^{1/2}\tilde{\rho}_{ij}(k)\right\}\right]$. Because $\beta^{-1}\log p = n^{-1/12} \to 0$, we only need to show that

$$\begin{aligned}&\mathrm{P}\left\{V^2(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n) - 2\log(Kp^2) + \log\log(Kp^2) \le x, W(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n) \le y\right\}\\ \to& G(x)\cdot\Phi(y).\end{aligned} \tag{S3.42}$$

Suppose $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n$ are independent and identical distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and independent of $(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n)$. Next, we show that $\{W(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n), V(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n)\}$ has the same limited distribution as $(\{W(\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n), V(\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n)\}$. Then, according to Theorem 6, we will obtain the result.

It is known that a sequence of random variables $\{\xi_n\}_{n=1}^{\infty}$ converges weakly to a random variable ξ if and only if for every $f \in \mathscr{C}_b^3(\mathbb{R}^2)$, $\mathbb{E}f(\xi_n) \to \mathrm{E}f(\xi)$; see, e.g., Pollard (1984), Chapter III, Theorem 12. We use this property to give a metrization of the weak convergence in $\mathbb{R}^2$.

Thus, we only need to show that

$$\mathbb{E}[f\{W(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n), V(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n)\}] - \mathbb{E}[f\{W(\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n), V(\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n)\}] \to 0$$

for every $f \in \mathscr{C}_b^3(\mathbb{R}^2)$ as $n, p \to \infty$. Define

$$W_d = W(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_{d-1}, \boldsymbol{\xi}_d, \cdots, \boldsymbol{\xi}_n),\ V_d = V(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_{d-1}, \boldsymbol{\xi}_d, \cdots, \boldsymbol{\xi}_n).$$

We have

$$\begin{aligned}
&|\mathbb{E}[f\{W(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n), V(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n)\}] - \mathbb{E}[f\{W(\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n), V(\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n)\}]| \\
\le &\sum_{d=1}^{n} \left|\mathbb{E}\{f(W_d, V_d)\} - \mathbb{E}\{f(W_{d+1}, V_{d+1})\}\right|.
\end{aligned}$$

In the following, we only proof the result with $K = 1$. For the other fixed integer K, the proof are very similar.

Define

$$\begin{aligned}
W_{d,0} =& \frac{1}{n(n-1)\sigma_S} \sum\sum_{1 \le t \ne s \le d-2} \boldsymbol{x}_t^\top \boldsymbol{x}_s \boldsymbol{x}_{t+1}^\top \boldsymbol{x}_{s+1} + \frac{1}{n(n-1)\sigma_S} \sum\sum_{d+1 \le t \ne s \le n} \boldsymbol{x}_t^\top \boldsymbol{x}_s \boldsymbol{x}_{t+1}^\top \boldsymbol{x}_{s+1} \\
&+ \frac{2}{n(n-1)\sigma_S} \sum_{t=1}^{d-2} \sum_{s=d+1}^{n} \boldsymbol{x}_t^\top \boldsymbol{x}_s \boldsymbol{x}_{t+1}^\top \boldsymbol{x}_{s+1},
\end{aligned}$$

which only relies on $\mathcal{F}_d = \sigma\{\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_{d-2}, \boldsymbol{\xi}_{d+1}, \cdot, \boldsymbol{\xi}_n\}$. Hence,

$$\begin{aligned}
W_d - W_{d,0} =& \frac{1}{n(n-1)\sigma_S}\sum_{t=1}^{d-2}\boldsymbol{\varepsilon}_{d-1}^\top\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_d^\top\boldsymbol{\varepsilon}_{t+1} + \frac{1}{n(n-1)\sigma_S}\sum_{t=d}^{n}\boldsymbol{\varepsilon}_{d-1}^\top\boldsymbol{\xi}_t\boldsymbol{\varepsilon}_d^\top\boldsymbol{\xi}_{t+1} \\
&+ \frac{1}{n(n-1)\sigma_S}\sum_{t=1}^{d-2}\boldsymbol{\xi}_d^\top\boldsymbol{\varepsilon}_t\boldsymbol{\xi}_{d+1}^\top\boldsymbol{\varepsilon}_{t+1} \\
&+ \frac{1}{n(n-1)\sigma_S}\sum_{t=d}^{n}\boldsymbol{\xi}_d^\top\boldsymbol{\xi}_t\boldsymbol{\xi}_{d+1}^\top\boldsymbol{\xi}_{t+1} + \frac{1}{n(n-1)\sigma_S}\boldsymbol{\varepsilon}_{d-1}^\top\boldsymbol{\xi}_d\boldsymbol{\varepsilon}_d^\top\boldsymbol{\xi}_{d+1}, \\
W_{d+1} - W_{d,0} =& \frac{1}{n(n-1)\sigma_S}\sum_{t=1}^{d-2}\boldsymbol{\varepsilon}_{d-1}^\top\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_d^\top\boldsymbol{\varepsilon}_{t+1} + \frac{1}{n(n-1)\sigma_S}\sum_{t=d}^{n}\boldsymbol{\varepsilon}_{d-1}^\top\boldsymbol{\xi}_t\boldsymbol{\varepsilon}_d^\top\boldsymbol{\xi}_{t+1} \\
&+ \frac{1}{n(n-1)\sigma_S}\sum_{t=1}^{d-2}\boldsymbol{\varepsilon}_d^\top\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{d+1}^\top\boldsymbol{\varepsilon}_{t+1} + \frac{1}{n(n-1)\sigma_S}\sum_{t=d}^{n}\boldsymbol{\varepsilon}_d^\top\boldsymbol{\xi}_t\boldsymbol{\varepsilon}_{d+1}^\top\boldsymbol{\xi}_{t+1} \\
&+ \frac{1}{n(n-1)\sigma_S}\boldsymbol{\varepsilon}_{d-1}^\top\boldsymbol{\varepsilon}_d\boldsymbol{\varepsilon}_d^\top\boldsymbol{\varepsilon}_{d+1}.
\end{aligned}$$

Without loss of generality, we assume that $\sigma_i = 1$, $i = 1, \cdots, p$. Define

$$V_{d,0} = \beta^{-1}\log\left[\sum_{1\le i,j\le p}\exp\left\{\beta\left(n^{-1/2}\sum_{t=1}^{d-2}\varepsilon_{t+1,i}\varepsilon_{tj} + n^{-1/2}\sum_{t=d+1}^{n-1}\xi_{t+1,i}\xi_{tj}\right)\right\}\right],$$

which also only relies on $\mathcal{F}_d$. For simplicity, we define $l = i + (j-1)p$ and $\breve{\rho}_l^{(d,0)} = n^{-1}\sum_{t=1}^{d-2}\varepsilon_{t+1,i}\varepsilon_{tj} + n^{-1}\sum_{t=d+1}^{n-1}\xi_{t+1,i}\xi_{tj}$ for all pairs (i, j). Then,

$$V_{d,0} = \beta^{-1}\log\left\{\sum_{l=1}^{p^2}\exp\left(\beta n^{1/2}\breve{\rho}_l^{(d,0)}\right)\right\}.$$

Similarly, we define define $l = i + (j-1)p$ and

$$\breve{\rho}_l^{(d)} = \breve{\rho}_l^{(d,0)} + n^{-1}\varepsilon_{d-1,i}\varepsilon_{d-2,j} + n^{-1}\varepsilon_d\xi_{d-1} + n^{-1}\xi_{d+1}\xi_d$$

for all pairs (i,j). Then,

$$V_d = \beta^{-1}\log\left\{\sum_{l=1}^{p^2}\exp\left(\beta n^{1/2}\breve{\rho}_l^{(d)}\right)\right\}.$$

Define $f = f(x,y)$ and $\frac{\partial f}{\partial x} = f_1(x,y)$, $\frac{\partial f}{\partial y} = f_2(x,y)$, $\frac{\partial f^2}{\partial^2 x} = f_{11}(x,y)$, $\frac{\partial f^2}{\partial^2 y} = f_{22}(x,y)$, $\frac{\partial f^2}{\partial x\partial y} = f_{12}(x,y)$. By Taylor's expansion, we have

$$\begin{aligned}
&f(W_d,V_d) - f(W_{d,0},V_{d,0})\\
=&f_1(W_{d,0},V_{d,0})(W_d - W_{d,0}) + f_2(W_{d,0},V_{d,0})(V_d - V_{d,0})\\
&+\frac{1}{2}f_{11}(W_{d,0},V_{d,0})(W_d - W_{d,0})^2 + \frac{1}{2}f_{22}(W_{d,0},V_{d,0})(V_d - V_{d,0})^2\\
&+\frac{1}{2}f_{12}(W_{d,0},V_{d,0})(W_d - W_{d,0})(V_d - V_{d,0})\\
&+O\{|(V_d - V_{d,0})|^3\} + O\{|(W_d - W_{d,0})|^3\}
\end{aligned}$$

and

$$
\begin{aligned}
& f(W_{d+1}, V_{d+1}) - f(W_{d,0}, V_{d,0}) \\
=& f_1(W_{d,0}, V_{d,0})(W_{d+1} - W_{d,0}) + f_2(W_{d,0}, V_{d,0})(V_{d+1} - V_{d,0}) \\
& + \frac{1}{2} f_{11}(W_{d,0}, V_{d,0})(W_{d+1} - W_{d,0})^2 + \frac{1}{2} f_{22}(W_{d,0}, V_{d,0})(V_{d+1} - V_{d,0})^2 \\
& + \frac{1}{2} f_{12}(W_{d,0}, V_{d,0})(W_{d+1} - W_{d,0})(V_{d+1} - V_{d,0}) \\
& + O\{|(V_{d+1} - V_{d,0})|^3\} + O\{|(W_{d+1} - W_{d,0})|^3\}.
\end{aligned}
$$

Because $\mathbb{E}(\boldsymbol{\varepsilon}_t) = \mathbb{E}(\boldsymbol{\xi}_t) = \mathbf{0}$ and $\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) = \mathbb{E}(\boldsymbol{\xi}_t \boldsymbol{\xi}_t^\top)$, we can verify that

$$
\begin{aligned}
\mathbb{E}(W_d - W_{d,0} | \mathcal{F}_d) &= \mathbb{E}(W_{d+1} - W_{d,0} | \mathcal{F}_d), \\
\mathbb{E}\{(W_d - W_{d,0})^2 | \mathcal{F}_d\} &= \mathbb{E}\{(W_{d+1} - W_{d,0})^2 | \mathcal{F}_d\}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& \mathbb{E}\{f_1(W_{d,0}, V_{d,0})(W_d - W_{d,0})\} = \mathbb{E}\{f_1(W_{d,0}, V_{d,0})(W_{d+1} - W_{d,0})\}, \\
& \mathbb{E}\{f_{11}(W_{d,0}, V_{d,0})(W_d - W_{d,0})^2\} = \mathbb{E}\{f_{11}(W_{d,0}, V_{d,0})(W_{d+1} - W_{d,0})^2\}.
\end{aligned}
$$

Next, we consider $V_d - V_{d,0}$. Define $\boldsymbol{z}_{d,0} = (n^{1/2} \breve{\rho}_1^{(d,0)}, \cdots, n^{1/2} \breve{\rho}_{p^2}^{(d,0)})^\top$

and $\boldsymbol{z}_d = (n^{1/2}\breve{\rho}_1^{(d)}, \cdots, n^{1/2}\breve{\rho}_{p^2}^{(d)})^\top$. By Taylor's expansion, we have

$$\begin{aligned}
& V_d - V_{d,0} \\
=& n^{1/2}\sum_{l=1}^{p^2}\partial_l F_\beta(\boldsymbol{z}_{d,0})(\breve{\rho}_l^d - \breve{\rho}_l^{d,0}) + \frac{n}{2}\sum_{l=1}^{p^2}\sum_{k=1}^{p^2}\partial_k\partial_l F_\beta(\boldsymbol{z}_{d,0})(\breve{\rho}_l^d - \breve{\rho}_l^{d,0})(\breve{\rho}_k^d - \breve{\rho}_k^{d,0}) \\
& + \frac{1}{6}n^{3/2}\sum_{l=1}^{p^2}\sum_{k=1}^{p^2}\sum_{q=1}^{p^2}\partial_q\partial_k\partial_l F_\beta\{\boldsymbol{z}_{d,0} + \delta(\boldsymbol{z}_d - \boldsymbol{z}_{d,0})\}(\breve{\rho}_l^d - \breve{\rho}_l^{d,0})(\breve{\rho}_k^d - \breve{\rho}_k^{d,0})(\breve{\rho}_q^d - \breve{\rho}_q^{d,0}).
\end{aligned} \tag{S3.43}$$

By $\mathbb{E}(\boldsymbol{\varepsilon}_t) = \mathbb{E}(\boldsymbol{\xi}_t) = \mathbf{0}$ and $\mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top) = \mathbb{E}(\boldsymbol{\xi}_t\boldsymbol{\xi}_t^\top)$, we can also verify that

$$\mathbb{E}\{(\breve{\rho}_l^d - \breve{\rho}_l^{d,0})|\mathcal{F}_d\} = \mathbb{E}\{(\breve{\rho}_l^{d+1} - \breve{\rho}_l^{d,0})|\mathcal{F}_d\},$$

$$\mathbb{E}\{(\breve{\rho}_l^d - \breve{\rho}_l^{d,0})^2|\mathcal{F}_d\} = \mathbb{E}\{(\breve{\rho}_l^{d+1} - \breve{\rho}_l^{d,0})^2|\mathcal{F}_d\},$$

By Lemma A.2 in Chernozhukov et al. (2013), we have

$$\left|\sum_{l=1}^{p^2}\sum_{k=1}^{p^2}\sum_{q=1}^{p^2}\partial_q\partial_k\partial_l F_\beta(\boldsymbol{z}_{d,0} + \delta(\boldsymbol{z}_d - \boldsymbol{z}_{d,0}))\right| \le C\beta^2$$

for some positive constant C. By Condition (C1$'$), if $\boldsymbol{\varepsilon}_t$ has polynomial-type tails, we have

$$\mathrm{P}(\max_{1\le t\le n, 1\le i\le p}|\varepsilon_{it}| > Cn^{\frac{1}{6}-\delta}) \le np(Cn^{\frac{1}{6}-\delta})^{6\gamma_0+6+\epsilon}\mathbb{E}(|\varepsilon_{it}/\sigma_i|^{6\gamma_0+6+\epsilon}) \to 0,$$

where $0 < \delta < \frac{\epsilon}{6(6\gamma_0+6+\epsilon)}$. And for random variables $\xi_{it} \sim \mathcal{N}(0,1)$, we also have

$$\mathrm{P}\left\{\max_{1\le t\le n, 1\le i\le p} |\xi_{it}| > C\log(np)\right\} \to 0.$$

Thus, we have

$$\left|\frac{1}{6}n^{3/2}\sum_{l=1}^{p^2}\sum_{k=1}^{p^2}\sum_{q=1}^{p^2}\partial_q\partial_k\partial_l F_\beta\{\boldsymbol{z}_{d,0}+\delta(\boldsymbol{z}_d-\boldsymbol{z}_{d,0})\}(\breve{\rho}_l^d-\breve{\rho}_l^{d,0})(\breve{\rho}_k^d-\breve{\rho}_k^{d,0})(\breve{\rho}_q^d-\breve{\rho}_q^{d,0})\right|$$
$$\le C\beta^2 n^{-7/6-2\delta}$$

as probability tending to one. Hence, we have

$$|\mathbb{E}\{f_2(W_{d,0},V_{d,0})(V_d-V_{d,0})\} - \mathbb{E}\{f_2(W_{d,0},V_{d,0})(V_{d+1}-V_{d,0})\}| \le \beta^2 n^{-7/6-2\delta}.$$

Similarly, we can show that

$$\left|\mathbb{E}\{f_{22}(W_{d,0},V_{d,0})(V_d-V_{d,0})^2\} - \mathbb{E}\{f_{22}(W_{d,0},V_{d,0})(V_{d+1}-V_{d,0})^2\}\right| \le \beta^2 n^{-7/6-2\delta},$$

$$\begin{aligned}&\Big|\mathbb{E}\{f_{12}(W_{d,0},V_{d,0})(W_d-W_{d,0})(V_d-V_{d,0})\}\\ -&\mathbb{E}\{f_{12}(W_{d,0},V_{d,0})(W_{d+1}-W_{d,0})(V_{d+1}-V_{d,0})\}\Big| \le \beta^2 n^{-7/6-2\delta}.\end{aligned}$$

Hence, we can have

$$\begin{aligned}&\sum_{d=1}^{n}|\mathbb{E}\{f(W_d,V_d)\}-\mathbb{E}\{f(W_{d+1},V_{d+1})\}|\\ \leq & C\beta^2 n^{-1/6-2\delta}+2\sum_{d=1}^{n}[\mathbb{E}\{|(V_d-V_{d,0})|^3\}+\mathbb{E}\{|(W_d-W_{d,0})|^3\}].\end{aligned}$$

By (S3.43), we have $\mathbb{E}\{|(V_d-V_{d,0})|^3\}=O(n^{-1-3\delta})$ and

$$\sum_{d=1}^{n}\mathbb{E}|(W_d-W_{d,0})|^3\leq\sum_{d=1}^{n}[\mathbb{E}\{(W_d-W_{d,0})^4\}]^{3/4}.$$

Similar to the proof of Theorem 4, we have $\mathbb{E}\{(W_d-W_{d,0})^4\}=O(n^{-2})$. So we have

$$\sum_{d=1}^{n}|\mathbb{E}\{f(W_d,V_d)\}-\mathbb{E}\{f(W_{d+1},V_{d+1})\}|\leq C\beta^2 n^{-1/6-2\delta}+Cn^{-3\delta}+Cn^{-1/2}\to 0$$

as $n\to\infty$. Then, we obtain the result. If ε_t has sub-gaussian-type tails, we can also prove the result by the similar arguments.

S3.9 Proof of Theorem 8

Similar to the proof of Theorem 7, we only need to prove the result under the normality assumption. Without loss of generality, under the assumption of Theorem 8, we assume that

$$\mathbf{A}_0 = \begin{pmatrix} \mathbf{A}_{011} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{022} \end{pmatrix}, \ \ \mathbf{A}_1 = \begin{pmatrix} \mathbf{A}_{111} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Under the alternative hypothesis, we have $\mathbf{X}_{t1} = \mathbf{A}_{011}\boldsymbol{y}_t + \mathbf{A}_{111}\boldsymbol{y}_{t-1}$, where $\boldsymbol{y}_t = (z_{t1}, \cdots, z_{td})^\top$ is independent of $\boldsymbol{X}_{t2} = \mathbf{A}_{022}(z_{td+1}, \cdots, z_{tp})^\top$. As $\boldsymbol{\varepsilon}_t = (\boldsymbol{X}_{t1}^\top, \boldsymbol{X}_{t2}^\top)^\top$, we can decompose T_{SUM} as follows

$$\begin{aligned}
&T_{\text{SUM}} \\
=&\frac{1}{n(n-1)}\sum\sum_{t\neq s}\boldsymbol{\varepsilon}_t^\top\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_{t+1}^\top\boldsymbol{\varepsilon}_{s+1} \\
=&\frac{1}{n(n-1)}\sum\sum_{t\neq s}(\boldsymbol{X}_{t1}^\top\boldsymbol{X}_{s1} + \boldsymbol{X}_{t2}^\top\boldsymbol{X}_{s2})(\boldsymbol{X}_{t+1,1}^\top\boldsymbol{X}_{s+1,1} + \boldsymbol{X}_{t+1,2}^\top\boldsymbol{X}_{s+1,2}) \\
=&\frac{1}{n(n-1)}\sum\sum_{t\neq s}\boldsymbol{X}_{t1}^\top\boldsymbol{X}_{s1}\boldsymbol{X}_{t+1,1}^\top\boldsymbol{X}_{s+1,1} + \frac{1}{n(n-1)}\sum\sum_{t\neq s}\boldsymbol{X}_{t1}^\top\boldsymbol{X}_{s1}\boldsymbol{X}_{t+1,2}^\top\boldsymbol{X}_{s+1,2} \\
&+ \frac{1}{n(n-1)}\sum\sum_{t\neq s}\boldsymbol{X}_{t2}^\top\boldsymbol{X}_{s2}\boldsymbol{X}_{t+1,1}^\top\boldsymbol{X}_{s+1,1} + \frac{1}{n(n-1)}\sum\sum_{t\neq s}\boldsymbol{X}_{t2}^\top\boldsymbol{X}_{s2}\boldsymbol{X}_{t+1,2}^\top\boldsymbol{X}_{s+1,2}.
\end{aligned}$$

Similar to the proof of Theorem 5, we have

$$\sigma_{S1}^{-2}\mathrm{var}\left(\frac{1}{n(n-1)}\sum\sum_{t\neq s}\boldsymbol{X}_{t1}^{\top}\boldsymbol{X}_{s1}\boldsymbol{X}_{t+1,1}^{\top}\boldsymbol{X}_{s+1,1}\right) = O\left(\frac{d^2}{p^2}\right),$$
$$\sigma_{S1}^{-2}\mathrm{var}\left(\frac{1}{n(n-1)}\sum\sum_{t\neq s}\boldsymbol{X}_{t1}^{\top}\boldsymbol{X}_{s1}\boldsymbol{X}_{t+1,2}^{\top}\boldsymbol{X}_{s+1,2}\right) = O\left(\frac{d}{p}\right),$$
$$\sigma_{S1}^{-2}\mathrm{var}\left(\frac{1}{n(n-1)}\sum\sum_{t\neq s}\boldsymbol{X}_{t2}^{\top}\boldsymbol{X}_{s2}\boldsymbol{X}_{t+1,1}^{\top}\boldsymbol{X}_{s+1,1}\right) = O\left(\frac{d}{p}\right),$$

by the condition that the eigenvalues of $\boldsymbol{\Sigma}$ are all bounded. Thus, we have

$$\begin{aligned}T_{\mathrm{SUM}} =& \frac{1}{n(n-1)}\sum\sum_{t\neq s}\boldsymbol{X}_{t2}^{\top}\boldsymbol{X}_{s2}\boldsymbol{X}_{t+1,2}^{\top}\boldsymbol{X}_{s+1,2} + E(T_{\mathrm{SUM}}) + o_p(\sigma_{S1}) \\ \doteq& T_{\mathrm{SUM}}^{(2)} + E(T_{\mathrm{SUM}}) + o_p(\sigma_{S1}).\end{aligned}$$

Furthermore, taking the same procedure as the proof of Theorem 1, we have

$$\begin{aligned}&T_{\mathrm{MAX}} \\ =& \max_{1\leq i,j\leq p}|n^{1/2}\tilde{\rho}_{ij}(1)| + o_p(1) \\ =& \max\{\max_{1\leq i,j\leq d}|n^{1/2}\tilde{\rho}_{ij}(1)|, \max_{1\leq i\leq d, d+1\leq j\leq p}|n^{1/2}\tilde{\rho}_{ij}(1)|, \max_{d+1\leq i,j\leq p}|n^{1/2}\tilde{\rho}_{ij}(1)|\} + o_p(1).\end{aligned}$$

By the independence between $\boldsymbol{X}_{t1}$ and $\boldsymbol{X}_{t2}$, we know that $\max_{1\leq i,j\leq d}|n^{1/2}\tilde{\rho}_{ij}(1)|$ is independent of $T_{\mathrm{SUM}}^{(2)}$. Due to Theorem 6, we have $\max_{d+1\leq i,j\leq p}|n^{1/2}\tilde{\rho}_{ij}(1)|$ is asymptotically independent of $T_{\mathrm{SUM}}^{(2)}$. Because $\boldsymbol{X}_{t1}$ is independent of $\boldsymbol{X}_{t2}$, we also can prove that $\max_{1\le i\le d, d+1\le j\le p} |n^{1/2}\tilde{\rho}_{ij}(1)|$ is asymptotically independent of $T_{\mathrm{SUM}}^{(2)}$ by taking the same procedure as Theorem 6. Thus, we can prove that T_{MAX} is asymptotically independent of $T_{\mathrm{SUM}}^{(2)}$. By Lemma 7.10 in Feng et al. (2024), we can complete the proof.

Bibliography

Cai, T., W. Liu, and Y. Xia (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association 108*(501), 265–277.

Chang, J., Q. Jiang, and X. Shao (2023). Testing the martingale difference hypothesis in high dimension. *Journal of Econometrics 235*(2), 972–1000.

Chang, J., Y. Qiu, Q. Yao, and T. Zou (2018). Confidence regions for entries of a large precision matrix. *Journal of Econometrics 206*, 57–82.

Chang, J., Q. Yao, and W. Zhou (2017). Testing for high-dimensional white noise using maximum cross-correlations. *Biometrika 104*(1), 111–127.

Chen, S. X., L. X. Zhang, and P. S. Zhong (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association 105*, 810–819.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics 41*, 2786–2819.

Daniels, H. E. (1956). The approximate distribution of serial correlation coefficients. *Biometrika 43*(1), 169–185.

Feng, L., T. Jiang, X. Li, and B. Liu (2024). Asymptotic independence of the sum and maximum of dependent random variables with applications to high-dimensional tests. *Statistica Sinica*, In press.

Hall, P. G. and C. C. Hyde (1980). *Martingale Central Limit Theory and its Applications*. New York: Academic Press.

Jing, B., Q. Shao, and Q. Wang (2003). Self-normalized cramer-type large deviations for independent random variables. *The Annals of probability 31*(4), 2167–2215.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* NewYork: Wiley.

Pollard, D. (1984). *Convergence of stochastic processes*. New York: Springer, 1st ed.

Srivastava, M. (2009). A test of the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis 100*, 518–532.

Zaitsev, A. Y. (1987). On the gaussian approximation of convolutions under multidimensional analogues of S.N. Bernstein's inequality conditions. *Probability Theory and Related Fields 74*, 535–566.