Abstract

Testing for multi-dimensional white noise is an important subject in statistical inference. Such test in the high-dimensional case becomes an open problem waiting to be solved, especially when the dimension of a time series is comparable to or even greater than the sample size. To detect an arbitrary form of departure from high-dimensional white noise, a few tests have been developed. Some of these tests are based on max-type statistics, while others are based on sum-type ones. Despite the progress, an urgent issue awaits to be resolved: none of these tests is robust to the sparsity of the serial correlation structure. Motivated by this, we propose a Fisher’s combination test by combining the max-type and the sum-type statistics, based on the established asymptotically independence between them. This combination test can achieve robustness to the sparsity of the serial correlation structure, and combine the advantages of the two types of tests. We demonstrate the advantages of the proposed test over some existing tests through extensive numerical results and an empirical analysis.

Keywords: Asymptotically independence, Fisher’s combination test, High-dimensional white noise, Hypothesis test, Robustness
1 Introduction

Testing for white noise or serial correlation is an important problem in statistical modeling and inference, especially in diagnostic checking for linear regression and linear time series modeling. In recent years, researchers are increasingly interested in modeling high-dimensional time series data, which are becoming one of the most common data types, and frequently appear in many applications, including meteorology, genomics, chemometrics, biological and environmental research, finance and econometrics, etc. This brings further challenge to diagnostic checking, as we need to perform test for high-dimensional white noise, where the dimension of time series is comparable to or even greater than the sample size, i.e. the observed length of the time series.

For univariate time series, many widely used white noise tests have been proposed in the literature (Li, 2004). Some of these tests have been extended for testing multivariate time series (Hosking, 1980; Li and Mcleod, 1981), which are, however, only suitable for the case that the dimension of the time series is small compared to the sample size. Specifically, for univariate time series, the celebrated Box-Pierce portmanteau test and its variations are considered to be among the most popular omnibus tests for detecting non-specific forms of deviation from white noise. These tests are particularly convenient in practical applications, due to the fact that they are asymptotically distribution-free and χ²-distributed under the null hypothesis (Li, 2004; Lütkepohl, 2005). However, it is widely known that when extended to the multivariate cases, these tests suffer the issue of slow convergence to their asymptotic null distributions (Li et al., 2019).

Recently, multivariate white noise tests have undergone rapid development. Some new omnibus tests, such as the tests proposed by Chang, Yao and Zhou (2017), Li et al. (2019) and Tsay (2020) respectively, can even deal with high-dimensional time series, where the
dimension of the time series is comparable to or even greater than the sample size. Specifically, Chang, Yao and Zhou (2017) proposed a max-type test for high-dimensional white noise, using the maximum absolute auto-correlations and cross-correlations of the component series. Based on an approximation by the \( L_\infty \)-norm of a normal random vector, the critical value of the max-type test can be evaluated by bootstrapping from a multivariate normal distribution. Subsequently, Tsay (2020) proposed a rank-based max-type test for high-dimensional white noise by using Spearman’s rank correlation, and established the limiting null distribution based on the theory of extreme values. On the other hand, Li et al. (2019) proposed a sum-type test for high-dimensional white noise, using the sum of the squared singular values of several lagged sample autocovariance matrices. Using the random matrix theory, the asymptotic normality for the test statistic under the null is established under the Marčenko-Pastur asymptotic regime.

In general, the max-type test performs well in the case of sparse correlations, i.e. there is a small amount of large absolute auto- or cross-correlations at any nonzero lag. In contrast, the sum-type test performs well in the case of non-sparse correlations, which encapsulates the serial correlations within and across all component series. These two types of tests have their own applicability, but neither of them can perform well in both cases. In other words, neither of these two types of tests is applicable in the case of sparse serial correlations. This motivates us to establish a new test, which can combine the advantages of both types and is therefore applicable to sparse and non-sparse serial correlations. To this end, we reconsider the max-type test and the sum-type test, establish their asymptotic independence, then combine them to construct a combination test. It should be noted that the general idea of constructing combination tests by establishing the asymptotic independence of max-type and sum-type statistics has appeared in independence tests and covariance matrix tests for
high-dimensional random vectors, for example, in Li and Xue (2015) and Yu et al. (2020).

To combine the asymptotically independent tests, we employ the framework of combining the p-values of independent tests (Littell and Folks 1971). In a great deal of early literatures, the problem of combining independent tests of hypotheses has been widely considered, such as in Pearson (1938), Fisher (1950), Wilk and Shapiro (1968) and Naik (1969), to name but a few. Among these methods, the well known Fisher’s combination test proposed in Fisher (1950) is usually regarded as one of the best choices, whose advantages were discussed in Littell and Folks (1971). It should be noted that in addition to the way of combining p-values of independent tests, there are other ways for combining independent tests. For example, if all test statistics asymptotically follow Gaussian distributions, a linear combination of the statistics can be used to construct a combined test statistic. However, in situation where the test statistics have different types of asymptotic distributions, such as normal distribution and Gumble distribution, it is usually difficult to directly combine these statistics, hence the way of combining p-values becomes more practical.

In this paper, for testing high-dimensional white noise, we propose a Fisher’s combination test by combining the p-values of the max-type and sum-type tests, which is suitable to detect sparse and non-sparse serial correlations. Using the extreme value theory and the martingale’s central limit theorem, we establish the limiting null distributions of the max-type and sum-type statistics, respectively. Then, we establish the asymptotically independence between the two statistics under the null hypothesis, which enables us to use Fisher’s framework of combining independent tests. Furthermore, we demonstrate the advantages of the proposed Fisher’s combination test over its competitors through extensive numerical results. In the empirical application, we demonstrate the robust performance of
the proposed Fisher’s combination test.

The main contributions of this paper are threefold as follows.

- First, we established the limiting null distribution of the max-type statistic for testing high-dimensional white noise, proved that this max-type test is rate-optimal and investigated its local power function in some special cases.

- Then, we proposed a new sum-type test for testing high-dimensional white noise, where the relationship between the sample size and the dimension is not constrained, while the existing sum-type test needs to impose the Marčenko-Pastur regime, i.e. the ratio of the sample size to the dimension goes to a constant.

- Finally, we proved the asymptotic independence between the above max-type and sum-type test statistics under both Gaussian and non-Gaussian distributions, which is the most important contribution of this paper. Based on this asymptotic independence, we constructed the Fisher’s combination test that is suitable to detect either sparse or non-sparse serial correlations. In particular, we can eliminate the Gaussianity requirement of the error distribution in establishing the asymptotic independence. In contrast, all the existing literatures on establishing such asymptotic independence are limited by the Gaussianity requirement. Indeed, the proof of such asymptotic independence is more challenging than that of the asymptotic independence between the max-type and sum-type test statistics proposed for many other important high-dimensional testing problems, such as the high-dimensional cross-sectional independence testing problem (Feng et al., 2022), the high-dimensional location testing problem (Xu et al., 2016) and the high-dimensional covariance matrix testing problem (Li and Xue, 2015; Yu et al., 2020), which are limited by the Gaussianity requirement.
The rest of this paper is organized as follows. In Section 2, we describe the problem of testing for high-dimensional white noise, reconsider the max-type and sum-type tests, establish their asymptotic independence and then construct the Fisher’s combination test. In Section 3, we present extensive numerical results of the proposed test in comparison with some of its competitors, followed by an empirical application in Section 4. Then, we conclude the paper with some discussions in Section 5 and relegate the technical proofs to the supplementary.

2 Methodology

2.1 Testing problem

Let \( \{\epsilon_t\} \) be a \( p \)-dimensional weakly stationary time series with mean zero, where \( \epsilon_t = (\epsilon_{t1}, \cdots, \epsilon_{tp})^\top \in \mathbb{R}^p \). Let \( \Sigma(k) = \{\sigma_{ij}(k)\}_{1 \leq i, j \leq p} \) denote the autocovariance of \( \epsilon_t \) at lag \( k \), and let \( \Gamma(k) = \{\rho_{ij}(k)\}_{1 \leq i, j \leq p} \) denote the autocorrelation of \( \epsilon_t \) at lag \( k \), where for any matrix \( M \), \( \text{diag}(M) \) denotes the diagonal matrix consisting of the diagonal elements of \( M \) only. Let \( \Sigma = \{\sigma_{ij}\}_{1 \leq i, j \leq p} = \Sigma(0) \) and \( \Gamma = \{\rho_{ij}\}_{1 \leq i, j \leq p} = \Gamma(0) \). Let \( \sigma_i^2 = \sigma_{ii} \), for each \( i \in \{1, \cdots, p\} \). \( \{\epsilon_t\} \) is white noise, if \( \Sigma(k) = 0 \) for all \( k \neq 0 \).

With the observations \( \{\epsilon_1, \cdots, \epsilon_n\} \), let

\[
\hat{\Gamma}(k) = \{\hat{\rho}_{ij}(k)\}_{1 \leq i, j \leq p} \equiv \text{diag}\{\hat{\Sigma}(0)\}^{-1/2}\hat{\Sigma}(k)\text{diag}\{\hat{\Sigma}(0)\}^{-1/2}
\]
denote the sample autocorrelation matrix at lag $k$, where
\[
\hat{\Sigma}(k) = \{\hat{\rho}_{ij}(k)\}_{1 \leq i, j \leq p} = \frac{1}{n} \sum_{t=1}^{n-k} \varepsilon_{t+k} \varepsilon_t^T
\]
denotes the sample autocovariance matrix at lag $k$.

We consider the following testing problem:
\[
H_0 : \{\varepsilon_t\} \text{ is white noise } v.s. \ H_1 : \{\varepsilon_t\} \text{ is not white noise}, \quad (1)
\]
where the dimension of time series $p$ is comparable to or even greater than the sample size $n$.

2.2 The max-type test

Before proposing the Fisher’s combination test for testing high-dimensional white noise, we need to re-examine the max-type and sum-type tests, which will be combined to construct the combination test.

Since $\Gamma(k) = 0$ for any $k \geq 1$ under $H_0$, the max-type test statistic $T_{\text{MAX}}$ is defined as
\[
T_{\text{MAX}} = \max_{1 \leq k \leq K} T_{n,k}, \quad (2)
\]
which was first proposed by Chang, Yao and Zhou [2017], where $T_{n,k} = \max_{1 \leq i, j \leq p} n^{1/2} |\hat{\rho}_{ij}(k)|$ and $K \geq 1$ is an integer. For this max-type test statistic, Chang, Yao and Zhou [2017] evaluated the critical value by bootstrapping from a multivariate normal distribution, which is a widely recognized practice in the case of sparse correlations.

To establish the Fisher’s combination test in this paper, we first derive the limiting null
distribution of the max-type test statistic, which will be presented in the following Theorem 1. Specifically, Theorem 1 states that $T_{\text{MAX}}^2 - 2 \log(Kp^2) + \log \log(Kp^2)$ has an asymptotic extreme-value distribution when both $n$ and $p$ go to infinity. Hence, a level-$\alpha$ test with $\alpha \in (0, 1)$ will be performed by rejecting $H_0$ when $T_{\text{MAX}}^2 - 2 \log(Kp^2) + \log \log(Kp^2)$ is larger than $q_\alpha$, i.e. the $1 - \alpha$ quantile of $G(y) \defeq \exp \{-\pi^{-1/2} \exp(-y/2)\}$.

In deriving Theorem 1, we impose the following three conditions.

(C1) $\varepsilon_{ti}$’s have one of the following two types of tails: (i) sub-gaussian-type tails, i.e. there exist some constant $\eta > 0$ and $M > 0$, such that $\mathbb{E} e^{\eta \varepsilon_{ti}/\sigma_i} \leq M$ for all $i \in \{1, \ldots, p\}$ and $t \in \{1, \ldots, n\}$, where $p$ satisfies $\log p = o(n^{1/5})$; (ii) polynomial-type tails, i.e. for some $\gamma_0 > 0$ and $c_1 > 0$, $p \leq c_1 n^{\gamma_0}$ and for some $\epsilon > 0$ and $M > 0$, $\mathbb{E} |\varepsilon_{ti}/\sigma_i|^{4\gamma_0 + 4 + \epsilon} \leq M$ for all $i \in \{1, \ldots, p\}$ and $t \in \{1, \ldots, n\}$.

(C2) There exists a positive constant $C$ such that $C^{-1} \leq \min_{1 \leq i \leq p} \sigma_i^2 \leq \max_{1 \leq i \leq p} \sigma_i^2 \leq C$.

(C3) There exists $\varrho \in (0, 1)$ s.t. $|\rho_{ij}| \leq \varrho$ for all $1 \leq i < j \leq p$ with $p \geq 2$. $|C_p|/p^2 \to 0$ as $p \to \infty$ if (C1)-(i) holds; and $|C_p|/n^{4/8} \to 0$ if (C1)-(ii) holds. Here $C_p \defeq \{(i, j) : |B_{p,(i,j)}| \geq p^{\kappa_p}\}$ and $B_{p,(i,j)} \defeq \{(s, l) : |\rho_{ij}\rho_{sl}| \geq \delta_p\}$ for $1 \leq i, j \leq p$ with $\delta_p, \kappa_p > 0$, $\delta_p = o(1/\log p)$ and $\kappa_p = o(1)$ as $p \to \infty$.

Remark 1. Condition (C1) requires that the tail of the distributions of $\varepsilon_{ti}$’s is sub-gaussian-type or polynomial-type, which is the same as Condition (C2) or (C2*) used in Cai et al. [2013]. It is a more general moment condition than the normal distribution assumption. Condition (C2) requires that all the variances of $\varepsilon_{ti}$’s are bounded. Condition (C3) requires that the number of variable pairs with strong correlation cannot be too large.
**Theorem 1.** Suppose Conditions (C1)-(C3) hold. Then, under $H_0$, for any $y \in \mathbb{R}$, we have

$$\Pr \left\{ T_{\text{MAX}}^2 - 2 \log(Kp^2) + \log(Kp^2) \leq y \right\} \to G(y)$$

as $n, p \to \infty$, where $G(y) = \exp \left\{ -\pi^{-1/2} \exp(-y/2) \right\}$.

We recall that all technical proofs of the theorems and the proposition are relegated to the supplementary.

Let $\mathcal{U}(c)$ be a set of matrices indexed by a constant $c$, which is given by

$$\left\{ \{\Gamma(1), \ldots, \Gamma(K)\} \in \mathbb{R}^{p \times Kp} : \max_{1 \leq k \leq K, 1 \leq i < j \leq p} |\rho_{ij}(k)| \geq c (\log p/n)^{1/2} \right\}.$$ 

Consider the following sparse alternative

$$H_a^R(c) \triangleq \left[ F(\varepsilon_1, \ldots, \varepsilon_n) : \{\text{cor}_F(\varepsilon_{t+1}, \varepsilon_t), \ldots, \text{cor}_F(\varepsilon_{t+K}, \varepsilon_t)\} \in \mathcal{U}(c) \right], \quad (3)$$

where $F(\varepsilon_1, \ldots, \varepsilon_n)$ denotes the joint distribution of $\{\varepsilon_1, \ldots, \varepsilon_n\}$ and $\text{cor}_F$ denotes the autocorrelation matrix under the joint distribution $F$. Let $\mathcal{T}_\alpha$ denote the set of all measurable size-$\alpha$ tests.

Further, the following theorem characterizes the conditions under which the power of the proposed max-type test $\mathbb{I}\{T_{\text{MAX}} - 2 \log(Kp^2) + \log(Kp^2) \geq q_0\}$ tends to 1 as $n \to \infty$, under the alternative $H_a^R(b_0)$ for some constant $b_0$.

**Theorem 2.** Assume that $\{\varepsilon_t\}$ is a strictly stationary time series and the long-run variance $\gamma_i^L = \lim_{n \to \infty} \text{var} \left( n^{-1/2} \sum_{t=1}^{n} \varepsilon_{it}^2 \right)$ is bounded for all $1 \leq i \leq p$. Then, we have

$$\inf_{F(\varepsilon_1, \ldots, \varepsilon_n) \in H_a^R(b_0)} \Pr \left[ \mathbb{I}\{T_{\text{MAX}} - 2 \log(Kp^2) + \log(Kp^2) \geq q_0\} = 1 \right] = 1 - o(1),$$
for all $b_0 > 3$, where the infimum is taken over the joint distribution family $H^R_a(b_0)$ of \{${\varepsilon}_1, \cdots, {\varepsilon}_n$\} defined in (3).

Theorem 2 indicates that the above max-type test can detect alternatives of order $(\log p/n)^{1/2}$. In Theorem 3, we will show that this test is rate-optimal, i.e. the rate of the signal gap, $(\log p/n)^{1/2}$, cannot be further relaxed.

**Theorem 3.** Suppose $c_0 < 1$ is a positive constant, and let $\beta$ be a positive constant satisfying $\alpha + \beta < 1$. If $\log p/n = o(1)$, we have

$$\inf_{T_{a} \in T_{a}} \sup_{F(\varepsilon_{1}, \cdots, \varepsilon_{n}) \in H^R_a(c_0)} \mathbb{P}(T_{a} = 0) \geq 1 - \alpha - \beta$$

as $n, p \to \infty$, where the supremum is taken over the joint distribution family $H^R_a(c_0)$ of \{${\varepsilon}_1, \cdots, {\varepsilon}_n$\} defined in (3).

Theorem 3 indicates that any measurable size-$\alpha$ test cannot differentiate between the null hypothesis $H_0$ and the sparse alternative, when

$$\max_{1 \leq k \leq K, 1 \leq i < j \leq p} |\rho(k)_{ij}| < c_0(\log p/n)^{1/2}$$

for some constant $c_0 < 1$.

We now 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 $i \in \{2, \cdots, p\}$, $\varepsilon_{t1}$’s are all i.i.d. from $\mathcal{N}(0, 1)$. \{${\varepsilon}_{t1}$\}_{t=1, \cdots, n} are independent of \{${\varepsilon}_{t1}$\}_{t=1, \cdots, n}, $i = 2, \cdots, p$. By the Central Limit Theorem, we have that $\sqrt{n - k} \hat{\rho}_{ij}(k) \overset{d}{\to} \mathcal{N}(0, 1)$ for $k > 1$; $\sqrt{n - 1} \hat{\rho}_{ij}(1) \overset{d}{\to} \mathcal{N}(0, 1)$ for $i \neq j$ and $\sqrt{\frac{n-1}{\frac{3\rho^2 + 2\rho^4 + 3\rho^6}{(1+\rho^2)^2}}} \{\hat{\rho}_{11}(1) - \frac{\rho}{1+\rho^2}\} \overset{d}{\to} \mathcal{N}(0, 1)$ for $k = 1$.  

10
Define \( x_\alpha = 2 \log(Kp^2) - \log \log(Kp^2) + q_\alpha \). Define \( A = \{(i, j, k)|1 \leq i, j \leq p, 1 \leq k \leq K\} \).

Thus,

\[
P \{ T_{\text{MAX}}^2 - 2 \log(Kp^2) + \log \log(Kp^2) \geq q_\alpha \} \\
= P \left\{ \max_{1 \leq i, j \leq p, 1 \leq k \leq K} (n - k)\hat{\rho}_{ij}^2(k) \geq x_\alpha \right\} \\
\geq P \{ (n - 1)\hat{\rho}_{11}^2(1) \geq x_\alpha \} \\
= P \left\{ \left| \mathcal{N} \left( \frac{\sqrt{n}\rho}{1 + \rho^2}, 1 - \frac{3\rho^2 + 2\rho^4 + 3\rho^6}{(1 + \rho^2)^4} \right) \right| \geq \sqrt{x_\alpha} \right\} + o(1) \\
= P \{ |\mathcal{N}(\sqrt{n}\rho, 1)| \geq \sqrt{x_\alpha} \} + o(1) \\
= \Phi(\sqrt{n}\rho - \sqrt{x_\alpha}) + \Phi(-\sqrt{n}\rho - \sqrt{x_\alpha}) + o(1)
\]

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

Additionally,

\[
P \{ T_{\text{MAX}}^2 - 2 \log(Kp^2) + \log \log(Kp^2) \geq q_\alpha \} \\
= P \left\{ \max_{1 \leq i, j \leq p, 1 \leq k \leq K} (n - k)\hat{\rho}_{ij}^2(k) \geq x_\alpha \right\} \\
\leq P \{ (n - 1)\hat{\rho}_{11}^2(1) \geq x_\alpha \} + P \left\{ \max_{(i, j) \in \mathcal{A} \cap (1, 1, 1)} (n - k)\hat{\rho}_{ij}^2(k) \geq x_\alpha \right\} \\
= \Phi(\sqrt{n}\rho - \sqrt{x_\alpha}) + \Phi(-\sqrt{n}\rho - \sqrt{x_\alpha}) + \alpha + o(1).
\]

In the last equality, we used the following derivation. By the proof of Theorem 1 in [Chang](#).
et al. (2018), we have
\[
P \left\{ \max_{(i,j,k) \in A/(1,1,1)} (n - k) \hat{p}_{ij}^2(k) \geq x_\alpha \right\} - P \left( \max_{1 \leq l \leq Kp^2 - 1} \xi_l^2 \geq x_\alpha \right) \to 0,
\]
where \((\xi_1, \cdots , \xi_{Kp^2 - 1})\) follows the multivariate normal distribution with zero mean and the same correlation matrix of \(\{(n - k) \hat{p}_{ij}^2(k)\}_{(i,j,k) \in A/(1,1,1)}\). After some calculations, we have
\[
\text{cor} \left( \hat{p}_{n1}(k), \hat{p}_{n1}(k - 1) \right) \to 2\rho, \quad \text{cor} \left( (n - k) \hat{p}_{11}(k), (n - k - 2) \hat{p}_{11}(k + 2) \right) \to \frac{\rho^2}{1 + \rho^2}, \quad \text{and the other correlations between } (n - k) \hat{p}_{ij}^2(k) \text{ are all zeros.}
\]
By Theorem 1 in the supplementary and condition \(\rho = O(\sqrt{\log p/n})\), we have \(P \left( \max_{1 \leq l \leq Kp^2 - 1} \xi_l^2 \geq x_\alpha \right) \to \alpha\).

Hence, the power function of the max-type test is
\[
\lim_{n,p \to \infty} \Phi(\sqrt{n}\rho - \sqrt{x_\alpha}) + \lim_{n,p \to \infty} \Phi(-\sqrt{n}\rho - \sqrt{x_\alpha}) \leq \beta_{\text{MAX}}(\rho) \leq \alpha + \lim_{n,p \to \infty} \Phi(\sqrt{n}\rho - \sqrt{x_\alpha}) + \lim_{n,p \to \infty} \Phi(-\sqrt{n}\rho - \sqrt{x_\alpha}).
\]

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

Remark 2. As mentioned in Chang, Zhou and Wang (2017), the max-type tests based on asymptotic Gumble distributions usually have conservative size performance. The proposed MAX test also has such limitation. To solve this problem, some resampling methods can be employed, such as the bootstrap procedure used in the max-type test proposed by Chang, Yao and Zhou (2017). The resampling methods can also relax the conditions imposed on \(\varepsilon_t\)’s
Unfortunately, such methods generally need to pay a much heavier computational cost, especially in high-dimensional situations. Hence, whether to use a test based on asymptotic distribution or based on resampling strongly depends on the requirement of computational efficiency.

2.3 The sum-type test

In this subsection, we reconsider the sum-type test, with the test statistic defined as

$$T_{\text{SUM}} = \frac{1}{n(n - 1)} \sum_{l=1}^{K} \sum_{t \neq s} \epsilon_t^\top \epsilon_{t+l} \epsilon_{s+t}^\top.$$ 

It can be seen from the following Theorem and Proposition that under $H_0$, $T_{\text{SUM}}/\hat{\sigma}_S$ has an asymptotically standard normal distribution when both $n$ and $p$ go to infinity, where

$$\hat{\sigma}_S^2 \doteq \frac{2K}{n(n-1)} \text{tr}(\Sigma^2)^2 \quad \text{and} \quad \text{tr}(\Sigma^2) \doteq \frac{1}{n(n-1)} \sum_{t \neq s} (\epsilon_t^\top \epsilon_s)^2.$$ 

Hence, a level-$\alpha$ test will be performed by rejecting $H_0$ when $T_{\text{SUM}}/\hat{\sigma}_S$ is larger than $z_\alpha$, i.e. the $1-\alpha$ quantile of the standard normal distribution.

Note that the test statistic in (4) is similar to the sum-type test statistic proposed by Li et al. (2019). The differences between them are twofold: first, the test statistic in (4) removes the diagonal elements $\epsilon_t^\top \epsilon_{t+l} \epsilon_{t+l}^\top$ from the summation to reduce the requirement of $\Sigma$; second, we use the martingale’s central limit theorem to establish the limiting null distribution of the test statistic, while Li et al. (2019) used the random matrix theory.

In deriving the asymptotic properties of $T_{\text{SUM}}$, we impose the following two conditions.

(C4) Let $\epsilon_t = \Sigma^{1/2} z_t$ under $H_0$, where $\{z_t\}$ with $z_t = (z_{t1}, \ldots, z_{tp})^\top$ is a sequence of $p$-dimensional independent random vectors with independent components $z_{ti}$’s, satis-
fying $Ez_t = 0, Ez_t^2 = 1$ and $Ez_t^4 < \infty$.

(C5) $\text{tr}(\Sigma^4) = o\{\text{tr}^2(\Sigma^2)\}$. 

**Remark 3.** Condition (C5) is mild and holds automatically if all the eigenvalues of $\Sigma$ are bounded, i.e. Condition (C5) is weaker than the condition of bounded eigenvalues of $\Sigma$ imposed in [Li et al. (2019)](#), which indeed reduces the requirement of $\Sigma$. Note that Condition (C5) is also commonly adopted in the literature of testing high-dimensional covariance matrices, such as in [Chen et al. (2010)](#).

**Theorem 4.** Suppose Conditions (C4)-(C5) hold. Then, under $H_0$, we have $T_{SUM}/\sigma_S \xrightarrow{d} \mathcal{N}(0, 1)$, where $\sigma_S^2 \doteq 2K/n(n-1)\text{tr}^2(\Sigma^2)$.

Following the result in Proposition 1 we use the above $\text{tr}(\Sigma^2)$ to estimate $\text{tr}(\Sigma^2)$.

**Proposition 1.** If $\varepsilon_t = \Sigma^{1/2}z_t$ and $\text{tr}(\Sigma^4) = o\{\text{tr}^2(\Sigma^2)\}$, then under $H_0$, $\text{tr}(\Sigma^2)/\text{tr}(\Sigma^2) \xrightarrow{P} 1$.

Further, we present the asymptotic power function of the sum-type test $I(T_{SUM}/\hat{\sigma}_S \geq z_\alpha)$, when an alternative hypothesis $H_1$ is specified. Here, we assume that under $H_1$, the observations $\{\varepsilon_1, \cdots, \varepsilon_n\}$ follow a $p$-dimensional first-order vector moving average process, abbreviated as VMA(1), of the form

$$H_1 : \varepsilon_t = A_0 z_t + A_1 z_{t-1}, \tag{5}$$

where $A_0, A_1 \in \mathbb{R}^{p \times p}$ are the coefficient matrices. We consider the asymptotic distribution of $T_{SUM}$ in the case of $K = 1$ in the following Theorem 5.

**Theorem 5.** Under $H_1$ in (5) with $K = 1$, we have $(T_{SUM} - \mu_S)/\sigma_{S_1} \xrightarrow{d} \mathcal{N}(0, 1)$, where

$$\mu_S \doteq \text{tr}(\Sigma_0 \Sigma_1) + \frac{2}{T} \text{tr}^2(\Sigma_{01}), \quad \Sigma_0 \doteq A_0^T A_0, \quad \Sigma_1 \doteq A_1^T A_1, \quad \Sigma_{01} \doteq A_0^T A_1.$$

14
\[
\sigma^2_{S_1} = \frac{2}{T^2} \text{tr}^2(\tilde{\Sigma}_0^2 + \tilde{\Sigma}_1^2) + \frac{6}{T^2} \text{tr}^2(\tilde{\Sigma}_0 \tilde{\Sigma}_1) \\
+ \frac{4}{T} \left[ 2 \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_1)^2 + (\nu_4 - 3) \text{tr} \left\{ D^2(\tilde{\Sigma}_0 \tilde{\Sigma}_1) \right\} \right] \\
+ \frac{8}{T^2} \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_1)^2 \text{tr}(\tilde{\Sigma}_0 + \tilde{\Sigma}_1^2) + \frac{16}{T^2} \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_1) \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_0) \\
+ \frac{16}{T^2} \text{tr}(\tilde{\Sigma}_0 + \tilde{\Sigma}_1) \left\{ \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_0) + \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_1) \right\} \\
+ \frac{16}{T^2} \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_0) \left\{ \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_0) + \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_1) + 2 \text{tr}(\tilde{\Sigma}_1 \tilde{\Sigma}_0 \tilde{\Sigma}_0) \right\} \\
+ \frac{4}{T} \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_0 + \tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_1 + 2 \tilde{\Sigma}_0 \tilde{\Sigma}_1 \tilde{\Sigma}_0 \tilde{\Sigma}_0) \\
+ \frac{4}{T} \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_0) + \frac{12}{T^2} \text{tr}^2(\tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_1) + \frac{16}{T^2} \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_1) \\
+ \frac{4}{T^2} \text{tr}^2(\tilde{\Sigma}_0 \tilde{\Sigma}_0) + \frac{4}{T^2} \text{tr}^2(\tilde{\Sigma}_1 \tilde{\Sigma}_0 \tilde{\Sigma}_0) + r_n,
\]

and the remainder \( r_n = o(\sigma^2_{S_1}) \). Here, for each square matrix \( A \), \( D(A) \) denotes the diagonal matrix consisting of the main diagonal elements of \( A \).

Similar to Proposition 1, we have \( \text{tr}(\tilde{\Sigma}^2_0) / \xi_0 \xrightarrow{p} 1 \) under \( H_1 \) in (5), where \( \xi_0 = \text{tr}(\tilde{\Sigma}_0^2 + \tilde{\Sigma}_1^2) + 2 \text{tr}(\tilde{\Sigma}_0 \tilde{\Sigma}_0 \tilde{\Sigma}_1) \). As a result, the asymptotic power function of the proposed sum-type test \( \| T_{\text{SUM}} / \hat{\sigma}_S \| \geq \alpha \) under \( H_1 \) in (5) is approximately equal to

\[
\beta_{\text{SUM}} \doteq \Phi \left( \frac{\mu_S}{\sigma_{S_1}} - \frac{\sqrt{2} n^{-1} \xi_0}{\sigma_{S_1}} \right).
\]

### 2.4 Fisher’s combination test

To combine the proposed max-type and sum-type tests, we propose the Fisher’s combination test, based on the asymptotic independence between \( T_{\text{MAX}} \) and \( T_{\text{SUM}} \), to be provided
in the following Theorem 6. Specifically, let

\[ p_{\text{MAX}} \doteq 1 - G \{ T_{\text{MAX}}^2 - 2 \log(Kp^2) + \log \log (Kp^2) \} \]

and

\[ p_{\text{SUM}} \doteq 1 - \Phi \left( T_{\text{SUM}} / \hat{\sigma}_s \right) \]

denote the p-values with respect to the test statistics \( T_{\text{MAX}} \) and \( T_{\text{SUM}} \) respectively. Based on \( p_{\text{MAX}} \) and \( p_{\text{SUM}} \), the proposed Fisher’s combination test rejects \( H_0 \) at the significance level \( \alpha \), if

\[ T_{\text{FC}} \doteq -2 \log p_{\text{MAX}} - 2 \log p_{\text{SUM}} \]

is larger than \( c_\alpha \), i.e. the \( 1 - \alpha \) quantile of the chi-squared distribution with 4 degrees of freedom \( \text{[Fisher, 1950; Littell and Folks, 1971]} \).

In deriving the asymptotic independence between \( T_{\text{MAX}} \) and \( T_{\text{SUM}} \), we need to impose an additional condition as follows. Let \( \lambda_{\text{min}}(\Sigma) \) and \( \lambda_{\text{max}}(\Sigma) \) denote the minimum and maximum eigenvalues of \( \Sigma \), respectively.

\[ \text{(C6)} \quad \text{tr}^{-1}(\Sigma^2)(\log p)^\gamma \max\{\lambda_{\text{max}}(\Sigma)M_p, M_p^2, M_p^{3/2} \lambda_{\text{max}}^{1/2}(\Sigma)\} \to 0 \quad \text{for some positive constant} \gamma > 1, \text{ where } M_p \doteq \max_{1 \leq i \leq p} \left\{ \sum_{j \neq i} \sigma_{ij}^2 \right\}. \]

**Remark 4.** Condition (C6) requires that the covariance between each pair of variables can not be very large, which holds in many common cases. For example, it automatically holds when all the variables are independent, i.e. \( M_p = 0 \); if all the eigenvalues of \( \Sigma \) are bounded, \( M_p \) is also bounded, hence Condition (C6) holds. In addition, if \( \Sigma \) is a bandable covariance matrices, i.e. \( \sigma_{ij} = 0 \) if \( |i - j| > k \) for some fixed integer \( k \), and all nonzero \( \sigma_{ij} \)’s are bounded by \( c \), then Condition (C6) holds when \( \lambda_{\text{max}}(\Sigma)(\log p)^\gamma \text{tr}^{-1}(\Sigma^2) \to 0 \).
Note that Condition (C6) is significantly different from Assumption 1 (ii) of Yu et al. (2020) for testing high-dimensional covariance matrix. For example, when $M_p = 0$, we do not need to impose conditions on $\lambda_i(\Sigma)$’s. In fact, these two type of conditions cannot contain each other.

**Theorem 6.** Suppose $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$ for $t = 1, \cdots, n$ and Conditions (C2), (C3), (C5) and (C6) hold. Then, we have

$$P \{ T_{\text{MAX}}^2 - 2 \log(Kp^2) + \log(Kp^2) \leq x, T_{\text{SUM}}/\hat{\sigma}_S \leq y \} \rightarrow G(x) \cdot \Phi(y), \quad (7)$$

as $n, p \to \infty$, i.e. $T_{\text{MAX}}$ and $T_{\text{SUM}}$ are asymptotically independent.

Further, we relax the assumption of Gaussian distribution of $\varepsilon_t$ in Theorem 6 to non-Gaussian distributions, with sub-gaussian-type or polynomial-type tails. To establish the theoretical result under non-Gaussian distributions, Condition (C1) is modified as follows.

(C1’) $\varepsilon_t$’s have one of the following two types of tails: (i) sub-gaussian-type tails, i.e. there exist some constant $\eta > 0$ and $M > 0$, such that $E e^{\eta \varepsilon^2_t/\sigma^2_t} \leq M$ for all $i \in \{1, \cdots, p\}$ and $t \in \{1, \cdots, n\}$, where $p$ satisfies $\log p = o(n^{1/6})$; (ii) polynomial-type tails, i.e. for some $\gamma_0$ and $c_1 > 0$, $p \leq c_1 n^{\gamma_0}$ and for some $\epsilon > 0$ and $M > 0$, $E |\varepsilon_t/\sigma_t|^6\gamma_0 + \epsilon \leq M$ for all $i \in \{1, \cdots, p\}$ and $t \in \{1, \cdots, n\}$.

**Theorem 7.** Assume Conditions (C1’) and (C2)-(C6) hold. Then, we have

$$P \{ T_{\text{MAX}}^2 - 2 \log(Kp^2) + \log(Kp^2) \leq x, T_{\text{SUM}}/\hat{\sigma}_S \leq y \} \rightarrow G(x) \cdot \Phi(y), \quad (8)$$

as $n, p \to \infty$, i.e. $T_{\text{MAX}}$ and $T_{\text{SUM}}$ are asymptotically independent.
Remark 5. Note that relaxing the assumption of Gaussian distribution in establishing the asymptotically independence between the max-type and the sum-type statistics is an important contribution of this paper, since all the existing literatures on establishing such asymptotic independence, including [Li and Xue (2015), Xu et al. (2016), Feng et al. (2022) and Yu et al. (2020)], are limited by the assumption of Gaussian distribution. In this paper, we have developed a novel theoretical tool to weaken the Gaussian distribution to non-Gaussian distributions with sub-gaussian-type or polynomial-type tails. Its theoretical framework is enlightening, which can be generalized to analogous studies.

Based on Theorem 6 or Theorem 7, we immediately have the following result for \( T_{FC} \).

**Corollary 1.** Assume the same conditions as in Theorem 6 or Theorem 7, then we have \( T_{FC} \overset{d}{\longrightarrow} \chi^2_4 \) as \( n, p \to \infty \).

Under the alternative hypothesis (3), we have \( p_{MAX} \to 0 \) under the sparse alternatives due to Theorem 2. On the other hand, under the dense alternative hypothesis (5), we have \( p_{SUM} \to 0 \) if \( \mu_S/\sigma_{S1} \to \infty \) due to Theorem 5.

According to the definition of \( T_{FCP} \), if \( p_{MAX} \to 0 \) or \( p_{SUM} \to 0 \), we have \( T_{FC} \to \infty \), hence we reject the null hypothesis.

Remark 6. Note that Conditions (C2), (C3), (C5) and (C6) are all about \( \Sigma \), which hold automatically if all the eigenvalues of \( \Sigma \) are bounded. This indicates that these conditions are compatible and the intersection of these conditions is very conventional, which means that the scope of application of the proposed Fisher’s combination test is relatively broad.

Next, we show that \( T_{SUM} \) is still asymptotically independent of \( T_{MAX} \) under a specific alternative hypothesis. Based on this result, we obtain a low bound of the power function of \( T_{FC} \).
Theorem 8. Assume Conditions (C1') and (C2)-(C5) hold. Assume that all eigenvalues of $\Sigma = \text{cov}(\varepsilon_t)$ are bounded. Let $K = 1$. Then, under the alternative hypothesis (5) with

$$A_0 = \begin{pmatrix} A_{011} & 0 \\ 0 & A_{022} \end{pmatrix}, \quad A_1 = \begin{pmatrix} A_{111} & 0 \\ 0 & 0 \end{pmatrix},$$

$A_{011} \in \mathbb{R}^{d \times d}$, $A_{022} \in \mathbb{R}^{(p-d) \times (p-d)}$, $A_{111} \in \mathbb{R}^{d \times d}$ and $d = o(p)$, we have

$$P \left\{ T_{\text{MAX}}^2 - 2 \log(Kp^2) + \log \log(Kp^2) \leq x, T_{\text{SUM}}/\hat{\sigma}_S \leq y \right\} \rightarrow P \left\{ T_{\text{MAX}}^2 - 2 \log(Kp^2) + \log \log(Kp^2) \leq x \right\} P \left\{ T_{\text{SUM}}/\hat{\sigma}_S \leq y \right\},$$

as $n, p \to \infty$, i.e. $T_{\text{MAX}}$ and $T_{\text{SUM}}$ are still asymptotically independent.

Define a minimal p-values test $T_{\text{min}} = \min(p_{\text{SUM}}, p_{\text{MAX}})$. According to Theorem 7, we reject the null hypothesis if $p_{\text{SUM}} \leq 1 - \sqrt{1 - \alpha} \approx \alpha/2$ or $p_{\text{MAX}} \leq 1 - \sqrt{1 - \alpha} \approx \alpha/2$. According to the results in [Littell and Folks, 1971], we have that the power of Fisher’s combination test $\beta_{T_{\text{FC}}}$ is comparable to the power of the minimal p-values test $\beta_{T_{\text{min}}}$. Thus, we have

$$\beta_{T_{\text{FC}}} \approx \beta_{T_{\text{min}}} \geq \beta_{T_{\text{SUM}}, \alpha/2} + \beta_{T_{\text{MAX}}, \alpha/2} - \beta_{T_{\text{SUM}}, \alpha/2} \beta_{T_{\text{MAX}}, \alpha/2},$$

where the last inequality is based on the inclusion-exclusion principle and the result of Theorem 8 and $\beta_{T_{\text{SUM}}, \alpha}$, $\beta_{T_{\text{MAX}}, \alpha}$, are respectively the power functions of the sum-type test $T_{\text{SUM}}$ and max-type test $T_{\text{MAX}}$ at significant level $\alpha$. 19
3 Numerical results

We now present some numerical results to demonstrate the performance of the proposed max-type test, sum-type test and Fisher’s combined probability test, abbreviated as MAX, SUM and FC respectively, as well as their comparison with the sum-type test proposed by Li et al. (2019), abbreviated as LY. Note that we exclude Chang, Yao and Zhou (2017) and Tsay (2020)'s max-type tests from comparison because they have almost the same performance as the proposed max-type test in the case of linear correlations.

3.1 Size performance

For the cases under the null hypothesis, we let \( \varepsilon_t = A z_t \) with \( z_t = (z_{t1}, \cdots, z_{tp})^\top \) and \( A = \{a_{ij}\}_{1 \leq i, j \leq p} \). We consider the following two distributions of \( z_t \): (i) \( z_{ti} \overset{i.i.d}{\sim} \mathcal{N}(0, I_p) \); (ii) \( z_{ti} \overset{i.i.d}{\sim} \text{Ga}(4, 0.5) \) \(-2\), and the following three settings of \( A \):

(I) \( A = \Sigma^{1/2} \), \( \Sigma = \{\sigma_{ij}\}_{1 \leq i, j \leq p} \), \( \sigma_{ii} = 1, i = 1, \cdots, p, \sigma_{ij} = 0.5(i-j)^{-2} \) with \( i \neq j \);

(II) \( A = \Sigma^{1/2} \), \( \Sigma = \{\sigma_{ij}\}_{1 \leq i, j \leq p} \), \( \sigma_{ii} = 1, i = 1, \cdots, p, \sigma_{ij} = 0.5\mathbb{1}(|i-j| < 5) \) with \( i \neq j \);

(III) \( a_{ij} \overset{i.i.d}{\sim} \text{U}(-1, 1) \).

Tables 1 and 2 summarize the empirical size performance of MAX, LY, SUM and FC under settings (I)-(III) with the distributions (i) and (ii), respectively. Both of these tables suggest that in terms of size performance, SUM is the best and MAX is the most conservative, while FC and LY are in between. In addition, we find that FC has much better size performance than LY in situations where \( p/n \) is relatively large or \( K > 1 \).
Table 1: Size performance in the case of $\varepsilon_t = A z_t$ with distribution (i): $z_t^{i.i.d} \sim \mathcal{N}(0, I_p)$.

| $n$ | $p$ | $K = 1$ | $K = 2$ | $K = 3$ |
|-----|-----|---------|---------|---------|
|     |     | MAX     | LY      | SUM     | FC      | MAX     | LY      | SUM     | FC      | MAX     | LY      | SUM     | FC      |
| 100 | 30  | 0.016   | 0.039   | 0.046   | 0.043   | 0.010   | 0.025   | 0.043   | 0.025   | 0.016   | 0.026   | 0.068   | 0.050   |
| 100 | 60  | 0.007   | 0.036   | 0.050   | 0.030   | 0.004   | 0.012   | 0.042   | 0.021   | 0.006   | 0.004   | 0.043   | 0.024   |
| 100 | 90  | 0.004   | 0.022   | 0.048   | 0.024   | 0.010   | 0.012   | 0.043   | 0.025   | 0.006   | 0.002   | 0.047   | 0.015   |
| 100 | 120 | 0.008   | 0.016   | 0.043   | 0.021   | 0.006   | 0.004   | 0.040   | 0.016   | 0.005   | 0.000   | 0.048   | 0.025   |
| 200 | 30  | 0.012   | 0.054   | 0.062   | 0.052   | 0.020   | 0.044   | 0.052   | 0.053   | 0.014   | 0.040   | 0.053   | 0.038   |
| 200 | 60  | 0.017   | 0.048   | 0.056   | 0.039   | 0.016   | 0.035   | 0.049   | 0.032   | 0.020   | 0.032   | 0.056   | 0.047   |
| 200 | 90  | 0.016   | 0.032   | 0.049   | 0.027   | 0.011   | 0.030   | 0.054   | 0.037   | 0.017   | 0.021   | 0.056   | 0.050   |
| 200 | 120 | 0.014   | 0.031   | 0.049   | 0.036   | 0.017   | 0.016   | 0.043   | 0.036   | 0.012   | 0.015   | 0.065   | 0.039   |
|     |     | Setting (I) |         |         |         |         | Setting (II) |         |         |         |         | Setting (III) |         |         |
| 100 | 30  | 0.014   | 0.038   | 0.050   | 0.043   | 0.009   | 0.037   | 0.052   | 0.038   | 0.005   | 0.022   | 0.051   | 0.032   |
| 100 | 60  | 0.003   | 0.030   | 0.049   | 0.021   | 0.010   | 0.017   | 0.050   | 0.025   | 0.009   | 0.011   | 0.049   | 0.025   |
| 100 | 90  | 0.008   | 0.027   | 0.045   | 0.031   | 0.011   | 0.011   | 0.056   | 0.040   | 0.006   | 0.007   | 0.046   | 0.021   |
| 100 | 120 | 0.008   | 0.017   | 0.042   | 0.028   | 0.002   | 0.008   | 0.052   | 0.022   | 0.004   | 0.001   | 0.035   | 0.022   |
| 200 | 30  | 0.018   | 0.045   | 0.053   | 0.043   | 0.014   | 0.045   | 0.051   | 0.043   | 0.010   | 0.030   | 0.048   | 0.035   |
| 200 | 60  | 0.016   | 0.041   | 0.052   | 0.033   | 0.016   | 0.031   | 0.056   | 0.039   | 0.013   | 0.017   | 0.047   | 0.027   |
| 200 | 90  | 0.016   | 0.028   | 0.040   | 0.023   | 0.017   | 0.023   | 0.049   | 0.043   | 0.017   | 0.009   | 0.050   | 0.031   |
| 200 | 120 | 0.015   | 0.035   | 0.050   | 0.037   | 0.011   | 0.013   | 0.043   | 0.028   | 0.011   | 0.008   | 0.052   | 0.030   |
Table 2: Size performance in the case of $\boldsymbol{e}_t = \mathbf{A} \boldsymbol{z}_t$ with distribution (ii): $z_{ti} \overset{i.i.d.}{\sim} Ga(4, 0.5) - 2$.

| n   | p   | MAX | LY | SUM | FC | MAX | LY | SUM | FC | MAX | LY | SUM | FC |
|-----|-----|-----|-----|-----|----|-----|-----|-----|----|-----|-----|-----|----|
|     |     | K = 1 |     |     |    | K = 2 |     |     |    | K = 3 |     |     |    |
| 100 | 30  | 0.011 0.050 0.058 0.049 | 0.008 0.038 0.060 0.036 | 0.012 0.018 0.053 0.031 |     |     |     |     |     |     |     |
| 100 | 60  | 0.010 0.023 0.043 0.027 | 0.009 0.014 0.034 0.020 | 0.009 0.007 0.040 0.020 |     |     |     |     |     |     |     |
| 100 | 90  | 0.006 0.026 0.044 0.022 | 0.006 0.013 0.058 0.031 | 0.008 0.002 0.058 0.027 |     |     |     |     |     |     |     |
| 100 | 120 | 0.011 0.017 0.046 0.028 | 0.006 0.012 0.046 0.019 | 0.009 0.001 0.047 0.022 |     |     |     |     |     |     |     |
| 200 | 30  | 0.018 0.046 0.052 0.054 | 0.012 0.047 0.064 0.041 | 0.013 0.027 0.039 0.031 |     |     |     |     |     |     |     |
| 200 | 60  | 0.014 0.047 0.060 0.037 | 0.013 0.038 0.054 0.035 | 0.012 0.029 0.045 0.041 |     |     |     |     |     |     |     |
| 200 | 90  | 0.015 0.034 0.049 0.039 | 0.015 0.026 0.055 0.035 | 0.010 0.009 0.042 0.025 |     |     |     |     |     |     |     |
| 200 | 120 | 0.022 0.040 0.059 0.038 | 0.014 0.010 0.043 0.022 | 0.016 0.010 0.057 0.029 |     |     |     |     |     |     |     |

Setting (II)

| n   | p   | MAX | LY | SUM | FC | MAX | LY | SUM | FC | MAX | LY | SUM | FC |
|-----|-----|-----|-----|-----|----|-----|-----|-----|----|-----|-----|-----|----|
| 100 | 30  | 0.010 0.049 0.061 0.048 | 0.003 0.032 0.053 0.035 | 0.011 0.021 0.044 0.032 |     |     |     |     |     |     |     |
| 100 | 60  | 0.012 0.023 0.042 0.029 | 0.013 0.022 0.053 0.038 | 0.008 0.016 0.063 0.037 |     |     |     |     |     |     |     |
| 100 | 90  | 0.007 0.029 0.053 0.028 | 0.008 0.008 0.051 0.024 | 0.003 0.002 0.055 0.031 |     |     |     |     |     |     |     |
| 100 | 120 | 0.005 0.017 0.052 0.025 | 0.005 0.006 0.040 0.016 | 0.009 0.002 0.051 0.023 |     |     |     |     |     |     |     |
| 200 | 30  | 0.016 0.045 0.057 0.044 | 0.019 0.045 0.059 0.045 | 0.019 0.044 0.060 0.051 |     |     |     |     |     |     |     |
| 200 | 60  | 0.013 0.042 0.055 0.038 | 0.007 0.023 0.044 0.030 | 0.015 0.019 0.050 0.029 |     |     |     |     |     |     |     |
| 200 | 90  | 0.015 0.033 0.046 0.037 | 0.013 0.027 0.070 0.040 | 0.016 0.015 0.049 0.031 |     |     |     |     |     |     |     |
| 200 | 120 | 0.009 0.024 0.048 0.029 | 0.018 0.027 0.050 0.042 | 0.008 0.013 0.053 0.028 |     |     |     |     |     |     |     |

Setting (III)

| n   | p   | MAX | LY | SUM | FC | MAX | LY | SUM | FC | MAX | LY | SUM | FC |
|-----|-----|-----|-----|-----|----|-----|-----|-----|----|-----|-----|-----|----|
| 100 | 30  | 0.010 0.049 0.061 0.048 | 0.003 0.032 0.053 0.035 | 0.011 0.021 0.044 0.032 |     |     |     |     |     |     |     |
| 100 | 60  | 0.012 0.023 0.042 0.029 | 0.013 0.022 0.053 0.038 | 0.008 0.016 0.063 0.037 |     |     |     |     |     |     |     |
| 100 | 90  | 0.007 0.029 0.053 0.028 | 0.008 0.008 0.051 0.024 | 0.003 0.002 0.055 0.031 |     |     |     |     |     |     |     |
| 100 | 120 | 0.005 0.017 0.052 0.025 | 0.005 0.006 0.040 0.016 | 0.009 0.002 0.051 0.023 |     |     |     |     |     |     |     |
| 200 | 30  | 0.016 0.045 0.057 0.044 | 0.019 0.045 0.059 0.045 | 0.019 0.044 0.060 0.051 |     |     |     |     |     |     |     |
| 200 | 60  | 0.013 0.042 0.055 0.038 | 0.007 0.023 0.044 0.030 | 0.015 0.019 0.050 0.029 |     |     |     |     |     |     |     |
| 200 | 90  | 0.015 0.033 0.046 0.037 | 0.013 0.027 0.070 0.040 | 0.016 0.015 0.049 0.031 |     |     |     |     |     |     |     |
| 200 | 120 | 0.009 0.024 0.048 0.029 | 0.018 0.027 0.050 0.042 | 0.008 0.013 0.053 0.028 |     |     |     |     |     |     |     |

22
3.2 Power comparison

In this subsection, we compare the empirical power performance of the above four tests. For the cases under the alternative hypothesis, we only consider the above distribution (i) to avoid redundancy and consider the following three new settings of $\varepsilon_t$:

(IV) VAR(1) model: $\varepsilon_t = A\varepsilon_{t-1} + z_t$;

(V) VMA(1) model: $\varepsilon_t = z_t + A z_{t-1}$;

(VI) VARMA(1) model: $\varepsilon_t = 0.5A\varepsilon_{t-1} + z_t + 0.5A z_{t-1}$.

Here “VAR(1)”, “VMA(1)” and “VARMA(1)” are the abbreviations of 1-order vector autoregressive process, vector moving average process and vector autoregressive moving average process, respectively. Let $A = \{a_{ij}\}_{1 \leq i,j \leq p}$. For the alternative hypothesis, we let the first $a_{ij} \neq 0$ for $1 \leq i, j \leq m$ and $a_{ij} = 0$ otherwise. Note that $m$ controls the signal strength and sparsity of $A$. For the VAR(1) model, if $m = 1$, $a_{ij} \sim U(0.4,0.8)$; if $2 \leq m \leq 10$, $a_{ij} \sim U(-1.4/m,1.4/m)$. For the VMA(1) model, if $m = 1$, $a_{ij} \sim U(0.4,0.9)$; if $2 \leq m \leq 10$, $a_{ij} \sim U(-1.8/m,1.8/m)$. For the VARMA(1) model, if $m = 1$, $a_{ij} \sim U(0.4,0.8)$; if $2 \leq m \leq 10$, $a_{ij} \sim U(-1.6/m,1.6/m)$. Specifically, as $m$ decreases, both the signal strength and sparsity of $A$ increase. Let $n = 200$, $p \in \{60,90\}$ and $K \in \{1,2,3\}$.

Figures 1 and 2 present the empirical power curves of MAX, LY, SUM and FC under settings (IV)-(VI) and distribution (i) for $(n,p) = (200,60)$ and $(200,90)$, respectively. In each panel of these figures, the abscissa $m$ varies between 1 and 10, corresponding to the power performance of the involved tests with different signal strength and sparsity of $A$. Both figures suggest that in terms of empirical power performance, FC is better than its competitors in most cases whether $A$ is sparse or non-sparse, which has robust performance
Figure 1: Power curves of the involved tests with \( m = 1, 2, \cdots, 10 \) and \( (n, p) = (200, 60) \).

due to the combination of the advantages of both MAX and SUM. Although FC cannot outperform its competitors in all cases, it is indeed applicable to both sparse and non-sparse cases of \( A \). In contrast, MAX generally fails when \( A \) is dense enough, while SUM and LY generally fail when \( A \) is very sparse.

Note that in all the above simulation studies, the ways of setting up \( \Sigma \) are very common,
Figure 2: Power curves of the involved tests with $m = 1, 2, \cdots, 10$ and $(n, p) = (200, 90)$.

all of which satisfy the condition of bounded eigenvalues of $\Sigma$. Hence, these simulation settings for $\Sigma$ satisfy all the conditions imposed on $\Sigma$ in the theoretical results established above.
4 Application

In this section, we are interested in testing whether the error series \( \{ \varepsilon_t \} \) under the Fama-French three-factor model (Fama and French, 1993) is white noise, i.e.

\[
H_0 : \{ \varepsilon_t \} \text{ is white noise versus } H_1 : \{ \varepsilon_t \} \text{ is not white noise,} \tag{10}
\]

where \( \varepsilon_t = (\varepsilon_{t1}, \cdots, \varepsilon_{tp})^\top \) and \( p \) is the number of securities. The Fama-French three-factor model is one of the most popular factor pricing models in finance, which has the explicit form

\[
Y_{ti} = r_{ti} - r_{ft} = \alpha_i + \beta_{i1}(r_{mt} - r_{ft}) + \beta_{i2}SMB_t + \beta_{i3}HML_t + \varepsilon_{ti}
\]

for \( t \in \{1, \cdots, n\} \) and \( i \in \{1, \cdots, p\} \), where \( r_{ti} \) is the return of the \( i \)-th security at time \( t \), \( r_{ft} \) is the risk free rate at time \( t \), \( Y_{ti} = r_{ti} - r_{ft} \) is the excess return of the \( i \)-th security at time \( t \) and \( r_{mt} \) is the market return at time \( t \).

We collected the return data of the securities in the S&P 500 index and considered two forms of data compilation. First, we compiled the monthly returns on all the securities that constitute the S&P 500 index each month over the period from January 2005 to November 2018. Because the securities that make up the index change over time, we only consider \( p = 374 \) securities that were included in the S&P 500 index during the entire period. A total of \( T = 165 \) consecutive observations were obtained. The time series data on the safe rate of return, and the market factors are obtained from Ken French’s data library web page. The one-month US treasury bill rate is chosen as the risk-free rate \( (r_{ft}) \). The value-weighted return on all NYSE, AMEX, and NASDAQ stocks from CRSP is used as a proxy for the market return \( (r_{mt}) \). The average return on the three small portfolios minus the average return on the three big portfolios \( (SMB_t) \), and the average return on two value
portfolios minus the average return on two growth portfolios \((HML_t)\) are calculated based on the stocks listed on the NYSE, AMEX and NASDAQ.

Second, we compiled the weekly returns on all the securities that constitute the S&P 500 index over the period from January 2005 to November 2018. The weekly data were calculated using the security prices on Fridays. Similar to the monthly data, we only considered a total of \(p = 381\) stocks that were included in the S&P 500 index during the entire period. We formed a total of \(T = 716\) weekly return rates for each stock during this period after excluding the Fridays that happened to be holidays.

Under these two forms of data compilation, we test the hypotheses in (10) using the proposed Fisher’s combination test as well as its competitors, respectively. Specifically, we let

\[
\hat{\varepsilon}_{ti} = Y_{ti} - \hat{\alpha}_i - \hat{\beta}_{i1}(r_{mt} - r_{ft}) - \hat{\beta}_{i2}SMB_t - \hat{\beta}_{i3}HML_t,
\]

where \(\hat{\alpha}_i, \hat{\beta}_{i1}, \hat{\beta}_{i2}\) and \(\hat{\beta}_{i3}\) are the ordinary least squares (OLS) estimators of \(\alpha_i, \beta_{i1}, \beta_{i2}\) and \(\beta_{i3}\), respectively. To demonstrate the usefulness of the proposed test, we treat \(\hat{\varepsilon}_t = (\hat{\varepsilon}_{t1}, \cdots, \hat{\varepsilon}_{tp})^T\) as the observation of \(\varepsilon_t\), instead of considering the testing problem within the Fama-French three-factor model.

We use the sliding window method for the subsequent application. Given a fixed length \(n\), for each \(\tau \in \{1, \cdots, T-n\}\), we implement each of the involved tests on the data compiled from the period from \(\tau\) to \(\tau + n - 1\), where \(\{\tau, \cdots, \tau + n - 1\}\) is the sliding window of length \(n\). Then, we record the rate of rejecting the null hypothesis in these \(T - n\) testing results corresponding to the \(T - n\) sliding windows.

Due to the great complexity and diversity of the financial market, the Fama-French three-factor model is only an approximation and the three included factors may often fail to accurately describe the generating mechanism of the excess returns of a large number
of securities. Nevertheless, it has played an important role in pricing analysis of securities. This certainly motivates the investigation on whether a certain factor pricing model is sufficient and whether more advanced factor pricing models with more explanatory factors are needed. It is not irrational to suspect that the Fama-French three-factor model is not sufficient hence the null hypothesis may not be true, especially in the high-dimensional situations. To this end, of course a testing method with more tendency of rejection may be considered to perform better, as long as the test can control the effect size.

Table 3 summarizes the rejecting rates for each \( n \in \{40, 50, 60, 70\} \) and each form of data compilation, where the prescribed integer \( K \in \{2, 3\} \) is used to establish the proposed test statistics. It suggests that for the weekly data, FC, MAX and SUM are more inclined to reject the null hypothesis than LY, where FC is the most powerful and MAX is the second. This may be due to the stronger dependence of securities on time series of weekly data, compared with the monthly data, which may lead to some correlation matrices with larger signal strength. In such a circumstance, both the max-type and sum-type of tests can perform well, and the Fisher’s combined probability test FC outperform them as a combination of them.

Compared with weekly time series, the time dependence of monthly time series is much weaker, which leads to some correlation matrices with much weaker signal strength. This may be the reason why MAX fails to deal with the monthly data, while SUM, FC and LY have good performance. In particular, SUM outperform all the remaining methods in such circumstance.

Overall, SUM, LY and MAX can only have good performance in their respective suitable situations, while FC can have robust performance in both situations.
Table 3: Rejecting rates for the weekly and monthly data respectively.

| n  | p  | Weekly data |         |         |         | Monthly data |          |          |  |
|----|----|-------------|---------|---------|---------|--------------|---------|---------|---|
|    |    |             | MAX     | LY      | SUM     |             | MAX     | LY      | SUM | FC |
|----|----|-------------|---------|---------|---------|--------------|---------|---------|     |    |
| 40 | 381| 0.74        | 0.19    | 0.83    | 0.93    | 0.72         | 0.24    | 0.63    | 0.86 |    |
| 50 | 381| 0.86        | 0.26    | 0.79    | 0.98    | 0.83         | 0.32    | 0.63    | 0.94 |    |
| 60 | 381| 0.88        | 0.34    | 0.73    | 0.98    | 0.87         | 0.39    | 0.57    | 0.97 |    |
| 70 | 381| 0.89        | 0.46    | 0.58    | 0.98    | 0.90         | 0.41    | 0.64    | 0.99 |    |
| 40 | 374| 0.00        | 0.45    | 0.64    | 0.36    | 0.00         | 0.45    | 0.86    | 0.67 |    |
| 50 | 374| 0.00        | 0.55    | 0.68    | 0.55    | 0.01         | 0.62    | 0.90    | 0.75 |    |
| 60 | 374| 0.00        | 0.50    | 0.70    | 0.49    | 0.01         | 0.59    | 0.95    | 0.84 |    |
| 70 | 374| 0.21        | 0.56    | 0.69    | 0.53    | 0.18         | 0.69    | 0.99    | 0.93 |    |

5 Conclusion

Driven by the task of testing for high-dimensional white noise, we adopt the strategy of combining independent tests of hypotheses. In particular, we employ the well known Fisher’s combination test to combine the max-type and sum-type statistics, which is guaranteed to be valid by the asymptotic independence between the two types of statistics. Through extensive numerical results, we demonstrate that the proposed test has clear advantages in power comparison, due to its robustness to sparsity of the serial correlation structure. Furthermore, via an empirical application, we demonstrate the robust performance of the proposed test in testing white noise of the return data of the S&P 500 securities under the Fama-French three-factor model.
References

Cai, T., Liu, W. and Xia, Y. (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., Qiu, Y., Yao, Q. and Zou, T. (2018), ‘Confidence regions for entries of a large precision matrix’, Journal of Econometrics 206, 57–82.

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

Chang, J., Zhou, W.-X. and Wang, L. (2017), ‘Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering’, Biometrics ,

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

Fama, E. F. and French, K. R. (1993), ‘Common risk factors in the returns on stocks and bonds’, Journal of Financial Economics 33(1), 3–56.

Feng, L., Jiang, T., Liu, B. and Xiong, W. (2022), ‘Max-sum tests for cross-sectional independence of high-dimensional panel data’, The Annals of Statistics 50(2), 1124–1143.

Fisher, R. A. (1950), Statistical methods for research workers, London: Oliver and Boyd, 11th ed.
Hosking, J. R. M. (1980), ‘The multivariate portmanteau statistic’, *Journal of the American Statistical Association* 75(371), 602–608.

Li, D. and Xue, L. (2015), ‘Joint limiting laws for high-dimensional independence tests’, *arXiv:1512.08819*.

Li, W. K. (2004), *Diagnostic checks in time series*, Chapman & Hall/CRC.

Li, W. K. and Mcleod, A. I. (1981), ‘Distribution of the residual autocorrelations in multivariate arma time series models’, *Journal of the Royal Statistical Society: Series B (Methodological)* 43(2), 231–239.

Li, Z., Lam, C., Yao, J. and Yao, Q. (2019), ‘On testing for high-dimensional white noise’, *The Annals of Statistics* 47(6), 3382–3412.

Littell, R. C. and Folks, J. L. (1971), ‘Asymptotic optimality of fisher’s method of combining independent tests’, *Journal of the American Statistical Association* 66(336), 802–806.

Lütkepohl, H. (2005), *New introduction to multiple time series analysis*, Springer, Berlin.

Naik, U. D. (1969), ‘The equal probability test and its applications to some simultaneous inference problems’, *Journal of the American Statistical Association* 64(327), 986–998.

Pearson, E. S. (1938), ‘The probability integral transformation for testing goodness of fit and combining independent tests of significance’, *Biometrika* 30(1-2), 134–148.

Tsay, R. S. (2020), ‘Testing serial correlations in high-dimensional time series via extreme value theory’, *Journal of Econometrics* 216(1), 106–117.
Wilk, M. B. and Shapiro, S. S. (1968), ‘The joint assessment of normality of several independent samples’, *Technometrics* 10(4), 825–839.

Xu, G., Lin, L., Wei, P. and Pan, W. (2016), ‘An adaptive two-sample test for high-dimensional means’, *Biometrika* 103(3), 609–624.

Yu, X., Li, D. and Xue, L. (2020), ‘Fisher’s combined probability test for high-dimensional covariance matrices’, [arXiv:2006.00426](https://arxiv.org/abs/2006.00426).
Supplementary Material of “Testing for high-dimensional white noise”

6 Technical proofs

6.1 Proof of Theorem 1

First, we present some technical results for the proof of Theorem 1. We restate Theorem 2 in Feng et al. (2022) as the following Proposition 2, in which the following condition is imposed.

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

Proposition 2. Suppose $(Z_1, \cdots, Z_p)^\top \sim N(0, \Sigma)$ and Condition (CA1) holds. Then, we have $\max_{1 \leq i \leq p} Z_i^2 - 2 \log p + \log \log p$ converges to a Gumbel distribution with cdf $G(x) = \exp\left(-\frac{1}{\sqrt{\pi}} e^{-x^2/2}\right)$ as $p \to \infty$.

The following lemma is from the proof of Theorem 2 in Feng et al. (2022).

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

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

33
where the sum runs over all \(i_1 < \cdots < i_t\) with \(i_1 \cdots i_t \in D_p = \{1, \cdots, p\} \setminus C_p\). Then,

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

**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 \(\sigma^2 = n^{-1} \sum_{i=1}^{n} \mathbb{E}(X_i^2)\). Then for all \(x > 0\),

\[
P \left( \sum_{i=1}^{n} X_i \geq x \right) \leq \exp \left( -\frac{x^2}{2n\sigma^2 + 2Ax/3} \right).
\]

Define \(\sigma_i^2 = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_{ti}^2\) and \(\sigma_i^2 = \text{var}(\varepsilon_{ti})\).

**Lemma 3.** Suppose Condition (C2) holds. Then, under \(H_0\),

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

\[
P \left( \max_{1 \leq i \leq p} |\hat{\sigma}_i^2 - \sigma_i^2| \geq C \frac{\epsilon_n}{\log p} \right) = O(p^{-1}),
\]

(12)

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

\[
P \left( \max_{1 \leq i \leq p} |\hat{\sigma}_i^2 - \sigma_i^2| \geq C \frac{\epsilon_n}{\log p} \right) = O(n^{-t/8}),
\]

(13)

as \(\epsilon_n = \max \{(\log p)^{1/6}/n^{1/2}, (\log p)^{-1}\} \to 0\).
Proof. We first assume that (C1)-(i) holds. It suffices to show that, for any $\delta > 0$,

$$
P \left\{ \max_{1 \leq i \leq p} \left| \frac{1}{n} \sum_{k=1}^{n} (\varepsilon_{ki}^2 - \mathbb{E}\varepsilon_{ki}^2) \right| \geq C \sqrt{\frac{\log p}{n}} \right\} = O \left( p^{-\delta} \right). \tag{14}$$

Define

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

where $\tau$ is sufficiently large. We have

$$
|\mathbb{E}\varepsilon_{ki}^2 - \mathbb{E}\tilde{\varepsilon}_{ki}^2| \leq C \left( \mathbb{E}\varepsilon_{ki}^4 \mathbb{I} \left\{ |\varepsilon_{ki}| \geq \tau \sqrt{\log(p + n)} \right\} \right)^{1/2}
\leq C(n + p)^{-\tau^2 \eta/2} \left( \mathbb{E}\varepsilon_{ki}^4 \exp \left( 2^{-1} \eta \varepsilon_{ki}^2 \right) \right)^{1/2}
\leq C(n + p)^{-\tau^2 \eta/2}, \tag{15}
$$

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

$$
P \left\{ \max_{1 \leq i \leq p} \left| \frac{1}{n} \sum_{k=1}^{n} (\varepsilon_{ki}^2 - \mathbb{E}\varepsilon_{ki}^2) \right| \geq C \sqrt{\frac{\log p}{n}} \right\}
\leq P \left\{ \max_{1 \leq i \leq p} \left| \frac{1}{n} \sum_{k=1}^{n} (\varepsilon_{ki}^2 - \mathbb{E}\tilde{\varepsilon}_{ki}^2) \right| \geq 2^{-1} C \sqrt{\frac{\log p}{n}} \right\}
+ npP \left\{ |\varepsilon_{ki}| \geq \tau \sqrt{\log(p + n)} \right\},
$$

where

$$
npP \left( |\varepsilon_{ki}| \geq \tau \sqrt{\log(p + n)} \right) \leq np(n + p)^{-\tau^2 \eta} \mathbb{E} \exp \left( \eta \varepsilon_{ki}^2 \right) = O \left( p^{-\delta} \right).
$$
Let $t = \eta (8\tau^2)^{-1} \sqrt{\log p/n}$ and $\tilde{Z}_{ki} = \tilde{\varepsilon}_{ki}^2 - \mathbb{E}\tilde{\varepsilon}_{ki}^2$. Then, we have

$$P \left\{ \frac{1}{n} \sum_{k=1}^{n} (\tilde{\varepsilon}_{ki}^2 - \mathbb{E}\tilde{\varepsilon}_{ki}^2) \geq C \sqrt{\frac{\log p}{n}} \right\} \leq \exp \left( -Ct \sqrt{n \log p} \right) \prod_{k=1}^{n} \mathbb{E} \exp \left( t\tilde{Z}_{ki} \right)$$

$$\leq \exp \left( -Ct \sqrt{n \log p} \right) \prod_{k=1}^{n} \left\{ 1 + \mathbb{E}t^2 \tilde{Z}_{ki}^2 \exp \left( t|\tilde{Z}_{ki}| \right) \right\}$$

$$\leq \exp \left\{ -Ct \sqrt{n \log p} + \sum_{k=1}^{n} \mathbb{E}t^2 \tilde{Z}_{ki}^2 \exp \left( t|\tilde{Z}_{ki}| \right) \right\}$$

$$\leq \exp \left\{ -C\eta (8\tau^2)^{-1} \log p + c_{\tau,\eta} \log p \right\} \leq C p^{-\delta},$$

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

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

which leads to (14).

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

$$\tilde{\varepsilon}_{ki}^2 = \varepsilon_{ki}^2 \mathbb{I} \{ |\varepsilon_{ki}^2| \leq n/(\log p)^8 \}.$$  

Then, as in (15), we can show that $|\mathbb{E}\varepsilon_{ki}^2 - \mathbb{E}\tilde{\varepsilon}_{ki}^2| \leq C n^{-\gamma_0/4}$. It follows that

$$P \left( \max_{1 \leq i \leq p} \left| \sum_{k=1}^{n} (\varepsilon_{ki}^2 - \mathbb{E}\varepsilon_{ki}^2) \right| \geq \frac{n\epsilon_n}{\log p} \right)$$

36
\[
\begin{align*}
\Pr \left( \max_{1 \leq i,j,k \leq p} \left| \frac{1}{n} \sum_{k=1}^{n} (\hat{\varepsilon}^2_{ki} - \mathbb{E}\hat{\varepsilon}^2_{ki}) \right| \geq 2^{-1} \frac{n\varepsilon_n}{\log p} \right) + \Pr \left( \max_{i,k} |\varepsilon^2_{ki}| \geq \frac{n}{(\log p)^8} \right) \\
\leq C p^2 \exp \{-C(\log p)^4\} + Cn^{-\epsilon/8},
\end{align*}
\]

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

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

**Lemma 4.** Under \(H_0\), we have

1. If (C1)-(i) holds, we have

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

2. If (C1)-(ii) holds, we have

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

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

\[
n\tilde{\rho}^2_{ij}(k) = \frac{1}{n} \left( \sum_{l=1}^{n-k} \sigma_i^{-1} \sigma_j^{-1} \varepsilon_{ti+l} \varepsilon_{t+k,j} \right)^2
\]

37
By the self-normalized large deviation theorem for independent random variables (Theorem 1 in Jing et al. (2003)), we can get

$$
\max_{1 \leq i \leq p, 1 \leq k \leq K} \mathbb{P}\left( \frac{\left( \sum_{t=1}^{n-k} \sigma_i^{-1} \sigma_j^{-1} \varepsilon_{ti} \varepsilon_{t+k,j} \right)^2}{\sum_{t=1}^{n-k} \sigma_i^{-2} \sigma_j^{-2} \varepsilon_{ti} \varepsilon_{t+k,j}^2} \geq x^2 \right) \leq C\{1 - \Phi(x)\} \quad (16)
$$

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

$$
\mathbb{P}\left( \left| \frac{1}{n} \sum_{t=1}^{n-k} \sigma_i^{-2} \sigma_j^{-2} \varepsilon_{ti} \varepsilon_{t+k,j}^2 - 1 \right| \geq C \frac{\varepsilon_n}{\log p} \right) = O(p^{-\delta}), \quad (17)
$$

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

$$
\mathbb{P}\left( \left| \frac{1}{n} \sum_{t=1}^{n-k} \sigma_i^{-2} \sigma_j^{-2} \varepsilon_{ti} \varepsilon_{t+k,j}^2 - 1 \right| \geq C \frac{\varepsilon_n}{\log p} \right) = O(n^{-\epsilon/8}), \quad (18)
$$

for any $\delta > 0$. Thus, together (16), (17) with (18), we can complete the proof of this lemma.

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

**Proof of Theorem 1** Define $\hat{T}_n \doteq \max_{1 \leq k \leq K} \hat{T}_{n,k}$, where $\hat{T}_{n,k} \doteq \max_{1 \leq i, j \leq p} n^{1/2} |\tilde{p}_{ij}(k)|$. 

38
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

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

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

\[
cov(\nu) = \{ a_{ij} \}_{1 \leq i,j \leq N} = \text{diag}(\Gamma(0) \otimes \Gamma(0), \cdots, \Gamma(0) \otimes \Gamma(0)), \tag{19}
\]

where \( \otimes \) denotes the Kronecker product and \( \text{diag}(\Gamma(0) \otimes \Gamma(0), \cdots, \Gamma(0) \otimes \Gamma(0)) \) denotes the block diagonal matrix composed by \( \Gamma(0) \otimes \Gamma(0), \cdots, \Gamma(0) \otimes \Gamma(0) \). For \( 1 \leq i \leq N \), define \( B_{N,i} = \{ 1 \leq j \leq N : |a_{ij}| \geq \delta_p \} \) and \( C_N = \{ i : |B_{N,i}| \geq p^k \} \). 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

\[
P(|\nu| \geq z) \leq C \{ 1 - \Phi(z) \} + O(p^{-2}) \leq C \frac{1}{\sqrt{\pi}} \frac{e^{-y/2}}{N} + O(p^{-2});
\]

39
if (C1)-(ii) holds, we have

\[ P(|\nu_i| \geq z) \leq C\{1 - \Phi(z)\} + O(n^{-1/8}) \leq C \frac{1}{\sqrt{\pi}} \frac{e^{-y/2}}{N} + O(n^{-1/8}). \]

Thus, by Condition (C3),

\[ P\left( \max_{i \in C_N} |\nu_i| > z \right) \leq |C_N| \cdot P(|\nu_i| \geq z) \to 0 \]

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

\[ P\left( \max_{i \in D_N} |\nu_i| > z \right) \leq P\left( \max_{1 \leq i \leq N} |\nu_i| > z \right) \leq P\left( \max_{i \in D_N} |\nu_i| > z \right) + P\left( \max_{i \in C_N} |\nu_i| > z \right). \]

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

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

as \( N \to \infty \).

We redefine \( \nu_s = n^{1/2} \hat{\rho}_{ij}(k) = n^{-1/2} \sum_{t=1}^{n-k} \varepsilon_{ti} \varepsilon_{t+k,j} = n^{-1/2} \sum_{t=1}^{n-k} Z_{ts} \). Let \( \hat{Z}_{ts} = Z_{ts} I(Z_{ts} \leq \tau_n) - \mathbb{E}\{Z_{ts} I(Z_{ts} \leq \tau_n)\} \). 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 \( \hat{\nu}_i = n^{-1/2} \sum_{t=1}^{n-k} \hat{Z}_{ts} \) and \( q = |D_N| \). If (C1)-(i)
holds, then

\[
\max_{1 \leq k \leq q} \left\{ \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \mathbb{E} \left| Z_{lk} \right| \mathbb{I} \{ \left| Z_{lk} \right| \geq \eta^{-1} 8M \log (p + n) \} \right\}
\]

\[\leq C \sqrt{n} \max_{1 \leq l \leq n} \max_{1 \leq k \leq q} \mathbb{E} \left| Z_{lk} \right| \mathbb{I} \{ \left| Z_{lk} \right| \geq \eta^{-1} 8M \log (p + n) \} \]

\[\leq C \sqrt{n} (p + n)^{-4} \max_{1 \leq l \leq n} \max_{1 \leq k \leq q} \mathbb{E} \left| Z_{lk} \right| \exp \{ \eta \left| Z_{lk} \right| / (2M) \} \]

\[\leq C \sqrt{n} (p + n)^{-4}. \]

If (C1)-(ii) holds, then

\[
\max_{1 \leq k \leq q} \left\{ \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \mathbb{E} \left| Z_{lk} \right| \mathbb{I} \{ \left| Z_{lk} \right| \geq \sqrt{n} / (\log p)^{8} \} \right\}
\]

\[\leq C \sqrt{n} \max_{1 \leq l \leq n} \max_{1 \leq k \leq q} \mathbb{E} \left| Z_{lk} \right| \mathbb{I} \{ \left| Z_{lk} \right| \geq \sqrt{n} / (\log p)^{8} \} \leq C n^{-\gamma_0 - \epsilon / 8}. \]

Thus, we have

\[
P \left\{ \max_{1 \leq k \leq q} | \hat{\nu}_k - \nu_k | \geq (\log p)^{-1} \right\}
\]

\[\leq P \left( \max_{1 \leq k \leq q} \max_{1 \leq l \leq n} \left| Z_{lk} \right| \geq \tau_n \right) \]

\[\leq n P \left( \max_{1 \leq i, j \leq p} \max_{1 \leq s \leq R} | \tilde{\varepsilon}_{i, s, t, j} \varepsilon_{t, \tau_{t+s,j}} | \geq \tau_n \right) \]

\[\leq n \max_{1 \leq i, j \leq p} \max_{1 \leq s \leq R} \left\{ P(| \tilde{\varepsilon}_{i, s, t} | \geq \tau_n^{1/2}) + P(| \varepsilon_{t+s,j} | \geq \tau_n^{1/2}) \right\} \]

\[= O(p^{-1} + n^{-\epsilon / 8}). \]
Note that
\[ \left| \max_{1 \leq k \leq q} \nu_k^2 - \max_{1 \leq k \leq q} \tilde{\nu}_k^2 \right| \leq 2 \max_{1 \leq k \leq q} |\tilde{\nu}_k| \max_{1 \leq k \leq q} |\nu_k - \tilde{\nu}_k| + \max_{1 \leq k \leq q} |\nu_k - \tilde{\nu}_k|^2. \] (20)

Therefore, to prove this theorem, it is enough to show
\[ \lim_{N \to \infty} P \left( \max_{i \in D_N} |\tilde{\nu}_i| > z \right) = 1 - \exp \left( -\frac{1}{\sqrt{\pi}} e^{-\frac{z^2}{2}} \right) \]

as \( N \to \infty \). Then, by Bonferroni inequality,
\[ \sum_{t=1}^{2k} (-1)^{t-1} \alpha_t \leq P \left( \max_{i \in D_N} |\tilde{\nu}_i| > z \right) \leq \sum_{t=1}^{2k+1} (-1)^{t-1} \alpha_t \]

for any \( k \geq 1 \), where
\[ \alpha_t \doteq \sum_{i_1, \cdots, i_t \in D_N} P \left( |\tilde{\nu}_{i_1}| > z, \cdots, |\tilde{\nu}_{i_t}| > z \right) \]

for \( 1 \leq t \leq N \), 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
\[ \lim_{N \to \infty} \alpha_t = \frac{1}{t!} \pi^{-t/2} e^{-t\gamma/2} \] (21)

for each \( t \geq 1 \). All the assumptions in Theorem 1.1 in [Zaitsev (1987)] are satisfied. Thus, we have
\[ \sum_{i_1}^* P \{ |Z_{i_1}| > z + \zeta_n (\log N)^{-1/2}, \cdots, |Z_{i_t}| > z + \zeta_n (\log N)^{-1/2} \} \]
\[- \left( \frac{|D_N|}{t} \right) c_1 t^{5/2} \exp \left\{ - \frac{n^{1/2} \zeta_n}{c_2 t^3 (\log N)^{1/2}} \right\} \]

\[ \leq \sum^* \mathbb{P} \{ |\tilde{Y}_{i_1}| > z, \ldots, |\tilde{Y}_{i_t}| > z \} \]

\[ \leq \sum^* \mathbb{P} \{ |Z_{i_1}| > z - \zeta_n (\log N)^{-1/2}, \ldots, |Z_{i_t}| > z - \zeta_n (\log N)^{-1/2} \} \]

\[ + \left( \frac{|D_N|}{t} \right) c_1 t^{5/2} \exp \left\{ - \frac{n^{1/2} \zeta_n}{c_2 t^3 (\log N)^{1/2}} \right\} , \]

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

\[ \sum^* \mathbb{P} \{ |Z_{i_1}| > z + \zeta_n (\log N)^{-1/2}, \ldots, |Z_{i_t}| > z + \zeta_n (\log N)^{-1/2} \} \to \frac{1}{t!} \pi^{-t/2} e^{-ty/2} , \]

\[ \sum^* \mathbb{P} \{ |Z_{i_1}| > z - \zeta_n (\log N)^{-1/2}, \ldots, |Z_{i_t}| > z - \zeta_n (\log N)^{-1/2} \} \to \frac{1}{t!} \pi^{-t/2} e^{-ty/2} , \]

with \(\zeta_n \to 0\) and \(N \to \infty\). Additionally,

\[ \left( \frac{|D_N|}{t} \right) c_1 t^{5/2} \exp \left\{ - \frac{n^{1/2} \zeta_n}{c_2 t^3 (\log N)^{1/2}} \right\} \]

\[ \leq C \left( \frac{N}{t} \right) t^{5/2} \exp \left\{ - \frac{n^{1/2} \zeta_n}{c_2 t^3 (\log N)^{1/2}} \right\} \to 0 \]

43
for ζn → 0 sufficiently slow. Thus, we have

\[ \sum_{t=1}^{\infty} P(|\tilde{v}_{t1}| > z, \cdots, |\tilde{v}_{t1}| > z) \to \frac{1}{t!} \pi^{-t/2} e^{-t^{2}/2}. \]

Let N → ∞, we have

\[ \leq \liminf_{N \to \infty} P \left( \max_{i \in D_N} |\tilde{v}_i| > z \right) \]
\[ \leq \limsup_{N \to \infty} P \left( \max_{i \in D_N} |\tilde{v}_i| > z \right) \leq \sum_{t=1}^{2k+1} (-1)^{t-1} \frac{1}{t!} \left( \frac{1}{\sqrt{\pi}} e^{-x/2} \right)^t \]

for each k ≥ 1. By letting k → ∞ and using the Taylor expansion of the function 1 − e−x, we obtain the result.

---

6.2 Proof of Theorem 2

Define (i_0, j_0, k_0) = arg max_{1 \leq i < j \leq p, 1 \leq k \leq K} |ρ_{ij}(k)|. Let γ_{i0} = E(ε_{i0}^2 ε_{i,t+l}^2) − σ_i^4. By the condition of Theorem 2, the long-run variance γ_{i0} = \lim_{n \to \infty} \{γ_{i0} + 2 \sum_{l=1}^{n} (1 - l/n)γ_{i0} \} is bounded. Hence,

\[ E \left( \frac{1}{n} \sum_{t=1}^{n-k_0} \varepsilon_{i0t}^2 \right) = \sigma_{i0}^2, \]
\[ \text{var} \left( \frac{1}{n} \sum_{t=1}^{n-k_0} \varepsilon_{i0t}^2 \right) = \frac{1}{n} \{E(\varepsilon_{i0t}^4) - \sigma_{i0}^4 \} + \frac{2}{n^2} \sum_{s<t} \{E(\varepsilon_{i0t}^2 \varepsilon_{i0s}^2) - \sigma_{i0}^4 \} \]
Thus, we have \( \frac{1}{n} \sum_{t=1}^{n-k_0} \varepsilon_{io,t}^2 \xrightarrow{p} \sigma_{io}^2 \). Similarly, we have
\[
\frac{1}{n} \sum_{t=1}^{n-k_0} \varepsilon_{j0,t}^2 \xrightarrow{p} \sigma_{j0}^2 \text{ and } \frac{1}{n} \sum_{t=1}^{n-k_0} \varepsilon_{io,t+k_0} \xrightarrow{p} \sigma_{io,j0}(k_0).
\]
Thus, \( \hat{\rho}_{io,j0}(k_0) \xrightarrow{p} \rho_{io,j0}(k_0) \). As \( n, p \to \infty \), we have
\[
P \left\{ \max_{1 \leq k \leq K} \max_{1 \leq i < j \leq p} n \hat{\rho}_{ij}^2(k) - 2 \log(Kp^2) + \log(\log(Kp^2)) \geq q_\alpha \right\}
\geq P \left\{ n \hat{\rho}_{io,j0}^2(k_0) - 2 \log(Kp^2) + \log(\log(Kp^2)) \geq q_\alpha \right\}
\rightarrow P \left\{ n \hat{\rho}_{io,j0}^2(k_0) - 2 \log(Kp^2) + \log(\log(Kp^2)) \geq q_\alpha \right\} = 1
\]
by the condition \( \rho_{io,j0}(k_0) \geq 3 \sqrt{\log p/n} > \sqrt{2 \log(Kp^2)/n} \). \( \square \)

### 6.3 Proof of Theorem 3

Define \( X = (\varepsilon_1^\top, \ldots, \varepsilon_p^\top) \in \mathbb{R}^d \), \( d = np \) and \( \varepsilon_i = (\varepsilon_{1i}, \ldots, \varepsilon_{ni}) \). Consider the Gaussian setting and a simple alternative set of parameters
\[
\mathcal{F}(\rho) = \left\{ \Xi : \Xi = \text{diag}(I_n, \ldots, I_n, \Sigma(\rho), I_n, \ldots, I_n), 1 \leq k \leq p \right\},
\]

where

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

Let $\mu_\rho$ 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 $\text{pr}_\Xi$ denote the probability measure of $N_d(0, \Xi)$ and $\text{pr}_{\mu_\rho} = \int \text{pr}_\Xi d\mu_\rho(\Xi)$. Let $\text{pr}_0$ denote the probability measure of $N_d(0, I_d)$. Note that, for any set $A$, we have

$$
\sup_{\Xi \in \mathcal{F}(\rho)} \text{pr}_\Xi(A^C) \geq \text{pr}_{\mu_\rho}(A^C), \quad 1 = \text{pr}_{\mu_\rho}(A^C) + \text{pr}_{\mu_\rho}(A)
$$

and

$$
\text{pr}_{\mu_\rho}(A) \leq \text{pr}_{0}(A) + \left| \text{pr}_{\mu_\rho}(A) - \text{pr}_{0}(A) \right|.
$$

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

$$
\inf_{T_\alpha \in T_\alpha, \Xi \in \mathcal{F}(\rho)} \sup_{\text{pr}_\Xi} (T_\alpha = 0) \geq 1 - \alpha - \sup_{A: \text{pr}_{0}(A) \leq \alpha} \left| \text{pr}_{\mu_\rho}(A) - \text{pr}_{0}(A) \right|
$$

$$
\geq 1 - \alpha - \frac{1}{2} \left\| \text{pr}_{\mu_\rho} - \text{pr}_{0} \right\|_{TV},
$$

where $\| \cdot \|_{TV}$ denotes the total variation norm. Setting $L_{\mu_\rho}(y) = \text{dpr}_{\mu_\rho}(y)/\text{dpr}_{0}(y)$, and by
Jensen’s inequality, we have

\[
\| \Pr_{\mu} - \Pr_0 \|_{TV} = \int |L_{\mu}(y) - 1| \, d\Pr_0(y) = \mathbb{E}_{\Pr_0} |L_{\mu}(Y) - 1| \leq \left[ \mathbb{E}_{\Pr_0} \left\{ L_{\mu}^2(Y) \right\} - 1 \right]^{1/2}.
\]

Therefore, as long as \( \mathbb{E}_{\Pr_0} \left\{ L_{\mu}^2(Y) \right\} = 1 + o(1) \), we have

\[
\inf_{T_{\alpha} \in \mathcal{T}, \Xi \in \mathcal{F}(\rho)} \sup_{\Pr \in \Xi} (T_{\alpha} = 0) \geq 1 - \alpha - o(1) > 0.
\]

We then prove that \( \mathbb{E}_{\Pr_0} \left\{ L_{\mu}^2(Y) \right\} = 1 + o(1) \). By construction, we have

\[
L_{\mu} = \frac{1}{p} \sum_{\Xi \in \mathcal{F}(\rho)} \left[ \frac{1}{|\Xi|^{1/2}} \exp \left\{ -\frac{1}{2} Z_{i, \cdot}^T (\Omega - I_d) Z_{i, \cdot} \right\} \right],
\]

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

\[
\mathbb{E}_{\Pr_0} \left\{ L_{\mu}^2(Y) \right\} = \frac{1}{p^2} \sum_{\Xi_1, \Xi_2 \in \mathcal{F}(\rho)} \mathbb{E} \left[ \frac{1}{|\Xi_1|^{1/2}} \frac{1}{|\Xi_2|^{1/2}} \exp \left\{ -\frac{1}{2} Z_{i, \cdot}^T (\Omega_1 + \Omega_2 - 2I_d) Z_{i, \cdot} \right\} \right],
\]

where \( \Omega_i = \Xi_i^{-1} \) for \( i = 1, 2 \). We write

\[
\mathbb{E}_{\Pr_0} \left\{ L_{\mu}^2(Y) \right\} = \frac{p - 1}{p} \mathbb{E} \left[ \frac{1}{|\Xi_1|^{1/2}} \frac{1}{|\Xi_2|^{1/2}} \exp \left\{ -\frac{1}{2} Z_{i, \cdot}^T (\Omega_1 + \Omega_2 - 2I_d) Z_{i, \cdot} \right\} \right].
\]
where $E_1$ represents the set of $(\Xi_1, \Xi_2)$ with $\Xi_1 \neq \Xi_2$, and $E_2$ represents the set of $(\Xi_1, \Xi_2)$ with $\Xi_1 = \Xi_2$. By standard argument in moment generating functions of the Gaussian quadratic form, we have

$$E \left\{ \exp \left( -\frac{1}{2} W^T A W \right) \right\} = \left\{ [1 + \lambda_1(A)] \cdots [1 + \lambda_q(A)] \right\}^{-1/2} = \left\{ \det(\mathbf{I}_q + A) \right\}^{-1/2}$$

if $W \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$ and $A \in \mathbb{R}^{q \times q}$. Without loss of generality, define $\Xi_1 = \text{diag} \{ \Sigma(\rho), \mathbf{I}_n, \cdots, \mathbf{I}_n \}$ and $\Xi_2 = \text{diag} \{ \mathbf{I}_n, \Sigma(\rho), \mathbf{I}_n, \cdots, \mathbf{I}_n \}$. Thus, $|\Xi_1| = |\Xi_2| = |\Sigma(\rho)|$. Additionally, define $\Omega_1 = \text{diag} \{ \Sigma(\rho)^{-1}, \mathbf{I}_n, \cdots, \mathbf{I}_n \}$ and $\Omega_2 = \text{diag} \{ \mathbf{I}_n, \Sigma(\rho)^{-1}, \mathbf{I}_n, \cdots, \mathbf{I}_n \}$.

$$E \left[ \frac{1}{|\Xi_1|^{1/2} |\Xi_2|^{1/2}} \exp \left\{ -\frac{1}{2} Z_i^T (\Omega_1 + \Omega_2 - 2\mathbf{I}_d) Z_i \right\} \right]$$

$$= E \left( \frac{1}{|\Xi_1|^{1/2} |\Xi_2|^{1/2}} \exp \left\{ -\frac{1}{2} Z_i^T \left\{ \Sigma(\rho)^{-1} - \mathbf{I}_n \right\} Z_i, -\frac{1}{2} Z_i^T \left\{ \Sigma(\rho)^{-1} - \mathbf{I}_n \right\} Z_i \right\} \right)$$

$$= \frac{1}{|\Sigma(\rho)|} |\Sigma(\rho)^{-1} - \mathbf{I}_n + \mathbf{I}_n|^{-1/2} |\Sigma(\rho)^{-1} - \mathbf{I}_n + \mathbf{I}_n|^{-1/2} = 1.$$

Thus,

$$E_1 = \frac{p - 1}{p} = 1 + o(1)$$  \hspace{1cm} (22)
as $p \to \infty$. Similarly,

$$
\frac{1}{p} \mathbb{E} \left[ \frac{1}{|\mathcal{E}|} \exp \left\{ -\frac{1}{2} Z_{i,\cdot}^T (2\Omega - 2I_d) Z_{i,\cdot} \right\} \right]
$$

$$
= \frac{1}{p} \frac{1}{|\Sigma(\rho)|} \mathbb{E} \left[ \exp \left\{ -\frac{1}{2} Z_{i,\cdot}^T (2\Sigma(\rho)^{-1} - 2I_n) Z_{i,\cdot} \right\} \right]
$$

$$
= \frac{1}{p} \frac{1}{|\Sigma(\rho)|} |2\Sigma(\rho)^{-1} - I_n|^{-1/2}.
$$

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

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

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

$$
2\Sigma(\rho)^{-1} - I_n = 2(I_n + \rho \mathbf{A})^{-1} - I_n = 2(I_n - \rho \mathbf{A} + \rho^2 \mathbf{A}^2 - \rho^3 \mathbf{A}^3 + \cdots) - I_n
$$

$$
= I_n - 2\rho \mathbf{A} + 2\rho^2 \mathbf{A}^2 \sum_{k=0}^{\infty} (-\rho)^k \mathbf{A}^k
$$

$$
= I_n - 2\rho \mathbf{A} + 2\rho^2 \mathbf{A}^2 (I_n + \rho \mathbf{A})^{-1}.
$$

49
Because $\rho^2 A^2 (I_n + \rho A)^{-1}$ is positive definite, we have

$$|2 \Sigma(\rho)^{-1} - I_n| \geq |I_n - 2 \rho A| = (1 - 4\rho^2)^{-n-1}.$$ 

Thus,

$$E_2 \leq p^{-1}(1 - \rho^2)^{-n+1}(1 - 4\rho^2)^{-(n-1)/2} = p^{-1} \exp(3c_0^2 \log p) \{1 + o(1)\} \to 0 \quad (23)$$

if $\rho = c_0 (\log p/n)^{1/2}$ and $c_0^2 < 1/3$. Combining (22) and (23), we have $\mathbb{E}_{\mathbb{P}_{\rho}} \left\{ L_{\mu_{\rho}}^2(Y) \right\} = 1 + o(1)$. Lastly, we can easily show that for $\rho = c_0 (\log p/n)^{1/2}$,

$$\left[ F(\varepsilon) : \text{cor}_F\{ (\varepsilon_{i,1}, \cdots, \varepsilon_{i,p}^T) \} \in \mathcal{F}(\rho), F(\varepsilon) \text{ is Gaussian} \right] \subset \left[ F(\varepsilon) : R\{ F(\varepsilon) \} \in \mathcal{U}(c) \right],$$

where $\varepsilon = \{ \varepsilon_1, \cdots, \varepsilon_n \}$ and $R\{ F(\varepsilon) \} = \{ \text{cor}_F(\varepsilon_{i,1}, \varepsilon_{i,t}), \cdots, \text{cor}_F(\varepsilon_{i,t+1}, \varepsilon_{i,t}) \}$. Thus,

$$\inf_{T_{\alpha} \in T_{\alpha}} \sup_{R\{ F(\varepsilon) \} \in \mathcal{U}(c)} \mathbb{P}_{\varepsilon}\{ T_{\alpha} = 0 \} \geq \inf_{T_{\alpha} \in T_{\alpha}} \sup_{\varepsilon \in \mathcal{F}(\rho)} \mathbb{P}_{\varepsilon}\{ T_{\alpha} = 0 \} \geq 1 - \alpha - o(1) > 0.$$

This completes the proof.

\section*{6.4 Proof of Theorem 4}

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

\textbf{Lemma 5.} Under Condition (C4), for any $m \times m$ symmetric matrix $A = \{ a_{ij} \}_{1 \leq i, j \leq m}$ and
where \( \Delta = \mathbb{E}(z_{it}^4) - 3 \).

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

**Proof.** Let

\[
T_{\text{SUM}} = \sum_{l=1}^{K} \frac{2}{n(n-1)} \sum_{t<s} \sum_{i} \epsilon_i^T \epsilon_s \epsilon_{t+l} \epsilon_{s+l} = \sum_{l=1}^{K} T_l.
\]

We will show that for each \( l \in \{1, \ldots, K\} \),

\[
\frac{T_l}{\sqrt{\frac{2}{n(n-1)} \text{tr}^2(\Sigma^2)}} \xrightarrow{d} \mathcal{N}(0, 1).
\]  

Define \( V_{nj} = n^{-1}(n-1)^{-1} \sum_{i=l+1}^{j-1} \epsilon_{i-l}^T \epsilon_j \epsilon_{i} \), \( j \in \{l+2, \ldots, n\} \) and \( W_{nk} = \sum_{i=l+2}^{k} V_{ni}, \) \( k \in \{l+2, \ldots, n\} \). Let \( \mathcal{F}_i = \sigma(\epsilon_1, \ldots, \epsilon_i) \) be the \( \sigma \)-field generated by \( \{\epsilon_j\}_{j \leq i} \). It is easy to show that \( \mathbb{E}(V_{ni} | \mathcal{F}_{i-1}) = 0 \) and it follows that \( \{W_{nk}, \mathcal{F}_k : l+2 \leq k \leq n\} \) is a zero mean martingale. Let \( v_{ni} = \mathbb{E}(V_{ni}^2 | \mathcal{F}_{i-1}) \), \( l+2 \leq i \leq n \) and \( V_n = \sum_{i=l+2}^{n} v_{ni} \). The central limit
Theorem (Hall and Hyde [1980]) will hold if we can show

$$\frac{V_n}{\text{var}(W_{nn})} \overset{p}{\to} 1,$$  \hspace{1cm} (25)

and for any $\epsilon > 0$,

$$\sum_{i=l+2}^{n} n^2 \text{tr}^{-2}(\Sigma^2) \mathbb{E} \left[ V_{ni}^2 \mathbb{I} \left\{ |V_{ni}| > \epsilon \sqrt{n^{-2} \text{tr}^2(\Sigma^2)} \right\} |\mathcal{F}_{i-1} \right] \overset{p}{\to} 0.$$  \hspace{1cm} (26)

It can be shown that

$$v_{ni} = \frac{1}{n^2(n-1)^2} \left\{ \sum_{j=l+1}^{i-1} (\varepsilon_{i-l}^{T} \varepsilon_{j-l})^2 \varepsilon_{j}^{T} \Sigma \varepsilon_{j} + 2 \sum_{l+1 \leq j < k < i} \varepsilon_{i-l}^{T} \varepsilon_{j-l}^{T} \varepsilon_{k-l}^{T} \varepsilon_{j} \Sigma \varepsilon_{k} \right\}.$$  

Then,

$$\frac{V_n}{\text{var}(W_{nn})} = \frac{2}{n(n-1) \text{tr}^2(\Sigma^2)} \left\{ \sum_{i=l+2}^{n} \sum_{j=l+1}^{i-1} (\varepsilon_{i-l}^{T} \varepsilon_{j-l})^2 \varepsilon_{j}^{T} \Sigma \varepsilon_{j} + 2 \sum_{i=l+2}^{n} \sum_{l+1 \leq j < k \leq i} \varepsilon_{i-l}^{T} \varepsilon_{j-l}^{T} \varepsilon_{k-l}^{T} \varepsilon_{j} \Sigma \varepsilon_{k} \right\}$$

$$\approx C_{n1} + C_{n2}.$$
Simple algebras lead to

\[
\mathbb{E}(C_{n1}) = \frac{(n - l)(n - l - 1)}{n(n - 1)},
\]

\[
\text{var}(C_{n1}) = \frac{4}{n^2(n-1)^2} \text{tr}^4(\Sigma^2) \mathbb{E}\left[ \sum_{i=1}^{n} \sum_{j=0}^{n-1} \{(\varepsilon_i^T \Sigma \varepsilon_j)^4 (\Sigma^2)^2 - \text{tr}^4(\Sigma^2)\} \right].
\]

By Lemma 5, we have \( \mathbb{E}\{(\varepsilon_i^T \Sigma \varepsilon_j)^2 - \text{tr}^2(\Sigma^2)\} = O\{\text{tr}(\Sigma^4)\}. \) Next, we will show that \( \mathbb{E}\{(\varepsilon_i^T \Sigma \varepsilon_j)^4 - \text{tr}^2(\Sigma^2)\} = O\{\text{tr}(\Sigma^4)\}. \) Define \( \Sigma^{1/2} \Sigma \Sigma^{1/2} \equiv \left\{ \omega_{kl} \right\}_{1 \leq k,l \leq p}. \)

\[
\mathbb{E}\{(\varepsilon_i^T \varepsilon_s)^4\} = \mathbb{E}\{(z_i^T \Sigma z_s)^4\} = \mathbb{E}\left\{ \left( \sum_{k,l=1}^{m} \sigma_{kl} z_{ik} z_{jl} \right)^4 \right\}
\]

\[
= \sum_{k,l=1}^{m} \sigma_{kl}^4 \mathbb{E}(z_{ik}^4) \mathbb{E}(z_{jl}^4) + \sum_{k=1}^{m} \sum_{s+t} \sigma_{ks}^2 \sigma_{kt}^2 \mathbb{E}(z_{ik}^2) \mathbb{E}(z_{js}^2) \mathbb{E}(z_{jl}^2) \mathbb{E}(z_{jt}^2)
\]

\[
+ 2 \sum_{k=1}^{m} \sum_{s+t} \sigma_{ks}^2 \sigma_{kt}^2 \mathbb{E}(z_{ik}^2) \mathbb{E}(z_{js}^2) \mathbb{E}(z_{jl}^2) \mathbb{E}(z_{jt}^2) + \sum_{k=1}^{m} \sum_{s+t} \sigma_{kl}^2 \sigma_{st}^2 \mathbb{E}(z_{ik}^2) \mathbb{E}(z_{is}^2) \mathbb{E}(z_{jt}^2) \mathbb{E}(z_{jt}^2).
\]

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

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

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

\[
\sum_{k=1}^{m} \sum_{s+t} \sigma_{kl}^2 \sigma_{st}^2 \leq \sum_{k,l,s,t} \sigma_{st}^2 \sigma_{kl}^2.
\]

\[
\sum_{k=1}^{m} \sum_{s+t} \sigma_{kl} \sigma_{st} \sigma_{sl} \sigma_{lt} \leq \sum_{k,l} \omega_{kl}^2 \leq \sum_{k,l} \omega_{kl}^2 = \text{tr}(\Sigma^4).
\]

53
Thus, we have

$$\mathbb{E}\{(\epsilon_i^\top \epsilon_s)^4\} - \text{tr}^2(\Sigma^2) = O\{\text{tr}(\Sigma^4)\}. \quad (27)$$

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

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

which implies $C_{n2} \overset{p}{\to} 0$. Thus, (25) holds.

It remains to show (26). Since

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

we only need to show that

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

Note that

$$\sum_{i=l+2}^{n} \mathbb{E}(V_{ni}^4) = O(n^{-4}) \sum_{i=l+2}^{n} \mathbb{E}\left\{ \left( \sum_{j=l+1}^{i-1} \epsilon_{i-j}^\top \epsilon_{j-i} \epsilon_i^\top \epsilon_j \right)^4 \right\},$$
which can be decomposed as $3Q + P$ with

$$Q = O(n^{-8}) \sum_{i=l+2}^{n} \sum_{s=t}^{i-1} \mathbb{E} \left( \varepsilon_{i-1}^T \varepsilon_{s-i} \varepsilon_{s}^T \varepsilon_{i-1} \varepsilon_{l-i} \varepsilon_{l-1} \varepsilon_{i-1} \varepsilon_{l-1} \varepsilon_{i-1} \varepsilon_{l-1} \varepsilon_{i} \varepsilon_{s} \varepsilon_{s} \varepsilon_{i} \varepsilon_{l} \varepsilon_{l} \varepsilon_{i} \right),$$

$$P = O(n^{-8}) \sum_{i=l+2}^{n} \sum_{s=1}^{i-1} \mathbb{E} \left\{ (\varepsilon_{i-1} \varepsilon_{j-1})^4 (\varepsilon_{i} \varepsilon_{s})^4 \right\}.$$

Note that $Q = O(n^{-4}) \mathbb{E}^2\{(\varepsilon_i^T \varepsilon_i)^2\} = o(n^{-4} \text{tr}^4(\Sigma^2))$ by Lemma 5. By (27), we have $P = O(n^{-4} \text{tr}^2(\Sigma^4)) = o(n^{-4} \text{tr}^4(\Sigma^2))$. And then (26) follows immediately. This completes the proof of (24). Finally, after some simple algebras, we have $\mathbb{E}(T_l T_k) = 0$ if $l \neq k$. Thus, we have

$$\frac{\sum_{l=1}^{K} T_l}{\sqrt{\frac{2K}{n(n-1)} \text{tr}^2(\Sigma^2)}} \xrightarrow{d} N(0, 1).$$

Here, we complete the proof. $\square$

### 6.5 Proof of Proposition 1

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

$$\mathbb{E}\{\text{tr}(\Sigma^2)\} = \text{tr}(\Sigma^2),$$

$$\text{var}\{\text{tr}(\Sigma^2)\} = 4n^{-2} \text{tr}^2(\Sigma^2) + 8n^{-1} \text{tr}(\Sigma^4) + 4\Delta n^{-1} \text{tr}(\Sigma^2 \circ \Sigma^2)$$

$$+ O\{n^{-3} \text{tr}^2(\Sigma^2) + n^{-2} \text{tr}(\Sigma^4)\} = o(\text{tr}^2(\Sigma^2)).$$

55
where $A \circ B = \{a_{ij}b_{ij}\}$ for two matrix $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$. Hence, we complete the proof of this proposition.

6.6 Proof of Theorem 6

Proof. Recall that $\varepsilon_t = A_0z_t + A_1z_{t-1}$ and $K = 1$. Actually,

$$G_1 = \frac{1}{n(n-1)} \sum_{s \neq t}^T (A_0z_s + A_1z_{s-1})^\top (A_0z_t + A_1z_{t-1})$$

$$(A_0z_{t-1} + A_1z_{t-2})^\top (A_0z_{s-1} + A_1z_{s-2})$$

$$= G(I) + G(II) + G(III),$$

where

$G(I)$

$$\doteq \frac{1}{n(n-1)} \sum_{s \neq t}^T (z_s^\top A_0^\top A_0z_t z_{t-1}^\top A_0^\top A_0z_{s-1} + z_{s-1}^\top A_1^\top A_1z_{t-1} z_{t-2}^\top A_1^\top A_1z_{s-2}$$

$$+ z_s^\top A_0^\top A_0z_t z_{t-2}^\top A_1^\top A_1z_{s-2} + z_{s-1}^\top A_1^\top A_1z_{t-1} z_{t-1}^\top A_0^\top A_0z_{s-1})$$

$G(II)$

$$\doteq \frac{1}{n(n-1)} \sum_{s \neq t}^T (z_s^\top A_0^\top A_1z_{t-1} z_{t-1}^\top A_0^\top A_0z_{s-1} + z_{s-1}^\top A_1^\top A_1z_t z_{t-1}^\top A_0^\top A_0z_{s-1}$$

$$+ z_{s-1}^\top A_1^\top A_1z_t z_{t-1}^\top A_0^\top A_0z_{s-1} + z_{s-1}^\top A_1^\top A_1z_{t-1} z_{t-2}^\top A_1^\top A_0^\top A_0z_{s-1}$$

$$+ z_s^\top A_0^\top A_0z_t z_{t-2}^\top A_1^\top A_0^\top A_0z_{s-1} + z_s^\top A_0^\top A_0z_t z_{t-1}^\top A_1^\top A_0^\top A_1z_{s-2}$$

56
After some tedious algebra, we have

$$
\mathbb{E}\{G(I)\} = \text{tr}(\tilde{\Sigma}_0 \hat{\Sigma}_1), \quad \mathbb{E}\{G(II)\} = 0, \quad \mathbb{E}\{G(III)\} = \frac{2}{T} \text{tr}^2(\tilde{\Sigma}_{01})
$$

and

\[
\begin{align*}
\text{var}\{G(I)\} &= \frac{2}{T^2} \text{tr}^2(\tilde{\Sigma}_0^2 + \hat{\Sigma}_1^2) + \frac{6}{T^2} \text{tr}^2(\tilde{\Sigma}_0 \hat{\Sigma}_1) + \frac{4}{T} \left[ 2 \text{tr} \left( \tilde{\Sigma}_0 \hat{\Sigma}_1 \right)^2 + (\nu_4 - 3) \text{tr} \left\{ D^2 \left( \tilde{\Sigma}_0 \hat{\Sigma}_1 \right) \right\} \right] + \tau_n, \\
\text{var}\{G(II)\} &= \frac{8}{T^2} \text{tr}(\tilde{\Sigma}_{01} \tilde{\Sigma}_{01}^\top) \text{tr}(\tilde{\Sigma}_0^2 + \hat{\Sigma}_1^2) + \frac{16}{T^2} \text{tr}(\tilde{\Sigma}_{01} \tilde{\Sigma}_1) \text{tr}(\tilde{\Sigma}_{01} \hat{\Sigma}_0) \\
&\quad + \frac{16}{T^2} \text{tr}(\tilde{\Sigma}_0 + \hat{\Sigma}_1) \left\{ \text{tr} \left( \tilde{\Sigma}_{01}^\top \tilde{\Sigma}_{01} \hat{\Sigma}_0 \right) + \text{tr} \left( \tilde{\Sigma}_{01} \tilde{\Sigma}_{01}^\top \hat{\Sigma}_1 \right) \right\} \\
&\quad + \frac{16}{T^2} \text{tr}(\tilde{\Sigma}_{01}) \left\{ \text{tr} \left( \tilde{\Sigma}_{01}^\top \hat{\Sigma}_{01} \hat{\Sigma}_0 \right) + \text{tr} \left( \tilde{\Sigma}_{01} \tilde{\Sigma}_{01}^\top \hat{\Sigma}_1 \right) + 2 \text{tr} \left( \tilde{\Sigma}_{01} \hat{\Sigma}_1 \hat{\Sigma}_0 \right) \right\} \\
&\quad + \frac{4}{T} \left( \tilde{\Sigma}_{01}^\top \tilde{\Sigma}_{01} \hat{\Sigma}_0 + \tilde{\Sigma}_{01} \tilde{\Sigma}_{01}^\top \hat{\Sigma}_1 + 2 \tilde{\Sigma}_{01}^\top \hat{\Sigma}_1 \tilde{\Sigma}_{01} \hat{\Sigma}_0 \right) + \tau_n, \\
\text{var}\{G(III)\} &= \frac{4}{T} \text{tr} \left( \tilde{\Sigma}_{01} \tilde{\Sigma}_{01}^\top \tilde{\Sigma}_{01} \hat{\Sigma}_0 \right) + \frac{12}{T^2} \text{tr}^2 \left( \tilde{\Sigma}_{01} \hat{\Sigma}_{01} \right) \\
&\quad + \frac{16}{T^2} \text{tr} \left( \tilde{\Sigma}_{01} \tilde{\Sigma}_{01}^\top \tilde{\Sigma}_{01} \hat{\Sigma}_{01} \right) + \tau_n, \\
\text{cov}\{G(I), G(III)\} &= \frac{4}{T^2} \text{tr}^2 \left( \tilde{\Sigma}_0 \tilde{\Sigma}_{01} \right) + \frac{4}{T^2} \text{tr}^2 \left( \tilde{\Sigma}_1 \tilde{\Sigma}_{01} \right) + \tau_n,
\end{align*}
\]
cov\{G(I), G(II)\} = r_n, \ cov\{G(II), G(III)\} = r_n.

Similar to the proof of (24), we have

\[
\frac{T_{\text{SUM}} - \mu_1}{\sigma_{S1}} \overset{d}{\rightarrow} \mathcal{N}(0, 1).
\]

\[\square\]

6.7 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 \(X \sim \mathcal{N}(\mu, \Sigma)\) with \(\Sigma\) being invertible. Partition \(X, \mu\) and \(\Sigma\) as

\[
X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},
\]

where \(X_2 \sim \mathcal{N}(\mu_2, \Sigma_{22})\). Set \(\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\). Then \(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1 \sim \mathcal{N}(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1, \Sigma_{22.1})\) and is independent of \(X_1\).

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 \(l_p = (2\log N - \ldots\)
log log \( N + y \)^{1/2} and \( B_i = \{|Z_i| > l_p\} \). Then, for each \( 1 \leq d \leq p \),

\[
\sum_{1 \leq i_1 < \cdots < i_d \leq p} \left| P(A_pB_{i_1} \cdots B_{i_d}) - P(A_p) \cdot P(B_{i_1} \cdots B_{i_d}) \right| \to 0
\]
as \( p \to \infty \).

**Proof.** The argument is divided into two steps.

**Step 1: appealing independence from normal distributions.**

Note that \((\varepsilon_{t_1}, \cdots, \varepsilon_{t_d})^\top \sim \mathcal{N}(0, \Sigma)\). Take \(X_{t_1} = (\varepsilon_{t_1}, \cdots, \varepsilon_{t_d})^\top\) and \(X_{t_2} = (\varepsilon_{t+d+1}, \cdots, \varepsilon_{t_p})^\top\).

Recall the notation in Lemma 6. Write \(X_{t_2} = U_t + V_t\), where \(U_t = X_{t_2} - \Sigma_{21} \Sigma_{11}^{-1} X_{t_1} \sim \mathcal{N}(0, \Sigma_{22,1})\) and \(V_t = \Sigma_{21} \Sigma_{11}^{-1} X_{t_1} \sim \mathcal{N}(0, \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})\). By Lemma 6,

\[
U_t \text{ and } \{\varepsilon_{t_1}, \cdots, \varepsilon_{t_d}\} \text{ are independent. \hspace{1cm} (29)}
\]

Write

\[
T_{\text{SUM}} = \frac{1}{n(n-1)} \sum_{l=1}^{K} \sum_{t+s} \varepsilon_t^\top \varepsilon_s \varepsilon_{t+l}^\top \varepsilon_{s+l}
\]

\[
= \frac{1}{n(n-1)} \sum_{l=1}^{K} \sum_{t+s} (X_{t_1}^\top X_{s_1} + X_{t_2}^\top X_{s_2})(X_{t+l,1}^\top X_{s+l,1} + X_{t+l,2}^\top X_{s+l,2})
\]

\[
= \frac{1}{n(n-1)} \sum_{l=1}^{K} \sum_{t+s} (U_t^\top U_s + U_t^\top V_s + U_s V_t + V_t V_s + X_{t_1}^\top X_{s_1})
\]

\[
\times (U_{t+l}^\top U_{s+l} + U_{t+l}^\top V_{s+l} + U_{s+l} V_{t+l} + V_{t+l} V_{s+l} + X_{t+l,1}^\top X_{s+l,1})
\]

59
Next, we will show that, for any $d \geq 1$ and $t > 0$, there exists $t = t_p > 0$ with $\lim_{N \to \infty} t_p = \infty$ and integer $p_0 \geq 1$ such that

$$P(|\Theta_q| \geq t_0) \leq \frac{1}{p^t}$$

(30)

as $p \geq p_0$. Here we only consider $\Theta_1$. The proof of the other parts are similar to $\Theta_1$.

By the decomposition, we know that $\{V_i\}_{i=1}^n$ is independent of $\{U_i\}_{i=1}^n$. Thus, condi-
tional on \( \{U_t\}_{t=1}^n \), \( \Theta_1 \) has the normal distribution. Hence, we have

\[
P(|\Theta_1| \geq \nu \sigma_S) \leq \sum_{l=1}^{K} \mathbb{P} \left( \left| \frac{2}{n(n-1)} \sum_{t+s} U_t^T U_s U_{t+l}^T V_{s+l} \right| \geq \nu \sigma_S/K \right)
\]

\[
= \sum_{l=1}^{K} \mathbb{E} \left[ \mathbb{I} \left\{ \left| \frac{2}{n(n-1)} \sum_{t+s} U_t^T U_s U_{t+l}^T V_{s+l} \right| \geq \nu \sigma_S/K \right\} \right]
\]

\[
= \sum_{l=1}^{K} \mathbb{E} \left[ 1 - \Phi \left( \frac{\nu \sigma_S}{K \hat{\sigma}_l} \right) \right] \approx \sum_{l=1}^{K} \mathbb{E} \left\{ \frac{2}{\sqrt{2 \pi K^{-1} l \hat{\sigma}_l} \sigma_S} e^{-\left(K^{-1} l \hat{\sigma}_l \sigma_S\right)^2/2} \right\},
\]

where

\[
\hat{\sigma}_l^2 = \frac{4}{n^2(n-1)^2} \sum_{t,m=s} U_t^T U_s U_{t+l}^T \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} U_{m+l}^T U_m U_s.
\]

Here, \( a_n \approx b_n \) denotes that \( a_n/b_n \to 1 \). Similar to the proof of Proposition \[\text{Proposition} 1\] we have

\[
\frac{\hat{\sigma}_l^2}{\frac{4}{n(n-1)} \text{tr}(\Sigma_{221}^2) \text{tr}(\Sigma_{221} \cdot \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})} \xrightarrow{p} 1.
\]

Thus,

\[
K^{-1} l \hat{\sigma}_l^{-1} \sigma_S \xrightarrow{p} \frac{\text{tr}(\Sigma^2)}{K \text{tr}^{1/2}(\Sigma_{221}^2) \text{tr}^{1/2}(\Sigma_{221} \cdot \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})}
\]

and

\[
\text{tr} \left\{ \Sigma_{221} \cdot \left( \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right) \right\} \leq \lambda_{\max} (\Sigma_{221}) \cdot \text{tr} \left( \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right)
\]

\[
\leq \lambda_{\max}(\Sigma) \cdot \text{tr} \left( \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right) \leq \lambda_{\max}(\Sigma) \lambda_{\max} (\Sigma_{11}^{-1}) \cdot \text{tr} (\Sigma_{12} \Sigma_{21})
\]

61
\[
= \lambda_{\max}(\Sigma) \frac{1}{\lambda_{\min}(\Sigma_{11})} \cdot \text{tr} (\Sigma_{12} \Sigma_{21}) \leq dM_p \lambda_{\max}(\Sigma).
\]

In fact, in the above, we use the assertion \(\lambda_{\min}(\Sigma_{11})\) is bounded by Condition (C2) and the fact that
\[
\text{tr} (\Sigma_{12} \Sigma_{21}) = \sum_{i=1}^{d} \sum_{j=d+1}^{p} \sigma_{ij}^2 \leq dM_p.
\]

Additionally, \(\Sigma_{22} = \Sigma_{22-1} + \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}\) and all three matrices are non-negative definite. Hence, we have \(\text{tr}(\Sigma^2) \geq \text{tr}(\Sigma_{22-1}^2)\). Thus,
\[
\frac{\ell}{K} \frac{\text{tr}(\Sigma^2)}{\text{tr}^{1/2}(\Sigma_{22-1}^2) \text{tr}^{1/2}(\Sigma_{22-1}^2 \Sigma_{11}^{-1} \Sigma_{12})} \geq \frac{\ell}{K} \frac{\text{tr}^{1/2}(\Sigma^2)}{\sqrt{dM_p \lambda_{\max}(\Sigma)}}.
\]

Then,
\[
\sum_{l=1}^{K} \mathbb{E} \left\{ \frac{2}{\sqrt{2\pi} K^{-1} i \hat{\sigma}_l^{-1} \sigma_S} e^{-(K^{-1} i \hat{\sigma}_l^{-1} \sigma_S)^2/2} \right\} \leq K \frac{2}{\sqrt{2\pi} \frac{\ell}{\sqrt{dM_p \lambda_{\max}(\Sigma)}}} \exp \left\{ -\frac{\ell^2}{K^2} \frac{\text{tr}(\Sigma^2)}{dM_p \lambda_{\max}(\Sigma)} \right\} \leq p^{-t_p}
\]

by Condition (C6) with \(t_p = \frac{\pi^2}{2K^2} (\log p)^{\gamma-1}\). Note that \(\text{tr}(\Sigma_{11}^2) \leq C_d\) is bounded.

Similarly, it can be proved that for the other parts \(\Theta_2, \cdots, \Theta_{11}\),
\[
P(|\Theta_2| \geq \nu \sigma_S)
\leq K \frac{1}{\sqrt{2\pi}} \frac{2}{\frac{\ell}{\sqrt{dM_p \lambda_{\max}(\Sigma)}} \frac{\text{tr}(\Sigma^2)}{\text{tr}^{1/2}(\Sigma_{22-1}) \text{tr}^{1/2}(\Sigma_{22-1}^2 \Sigma_{11}^{-1} \Sigma_{12})}} \exp \left[ -\frac{\ell^2}{K^2} \frac{\text{tr}^2(\Sigma^2)}{\text{tr}(\Sigma_{22-1}) \text{tr}(\Sigma_{22-1}^2 \Sigma_{11}^{-1} \Sigma_{12})^2} \right]
\]

62
\[ \leq K \frac{1}{\sqrt{2\pi}} \frac{2}{K \frac{\text{tr}^{1/2}(\Sigma^2)}{dM_p}} \exp \left\{ -\frac{\ell^2}{K^2} \frac{\text{tr}(\Sigma^2)}{d^2 M_p^2} \right\} \leq p^{-t_p}, \]

\[ P(|\Theta_3| \geq \ell \sigma_S) \leq K \frac{1}{\sqrt{2\pi}} \frac{2}{K \frac{\text{tr}(\Sigma^2)}{\text{tr}^{1/2}(\Sigma^2)_{22} \text{tr}^{1/2}(\Sigma^2_{11})}} \exp \left\{ -\frac{\ell^2}{K^2} \frac{\text{tr}^2(\Sigma^2)}{\text{tr}(\Sigma^2)_{22} \text{tr}(\Sigma^2_{11})} \right\} \]

\[ \leq K \frac{1}{\sqrt{2\pi}} \frac{2}{K \frac{\text{tr}(\Sigma^2)}{d^{1/2} C_d}} \exp \left\{ -\frac{\ell^2}{K^2} \frac{\text{tr}(\Sigma^2)}{C_d} \right\} \leq p^{-t_p}, \]

\[ \leq K \frac{1}{\sqrt{2\pi}} \frac{2}{K \frac{\text{tr}(\Sigma^2)}{\text{tr}^{1/2}(\Sigma^2)_{22} \text{tr}^{1/2}(\Sigma^2_{11})}} \exp \left\{ -\frac{\ell^2}{K^2} \frac{\text{tr}(\Sigma^2)}{\text{tr}^{1/2}(\Sigma^2)_{22} \text{tr}^{1/2}(\Sigma^2_{11})} \right\} \leq p^{-t_p}, \]

\[ P(|\Theta_4| \geq \ell \sigma_S) \approx P(|\Theta_1| \geq \ell \sigma_S), \quad P(|\Theta_5| \geq \ell \sigma_S) \approx P(|\Theta_2| \geq \ell \sigma_S), \]

\[ P(|\Theta_6| \geq \ell \sigma_S) \leq K \frac{1}{\sqrt{2\pi}} \frac{2}{K \frac{\text{tr}(\Sigma^2)}{\text{tr}^{1/2}(\Sigma^2)_{22} \text{tr}^{1/2}(\Sigma^2_{11})}} \exp \left\{ -\frac{\ell^2}{K^2} \frac{\text{tr}(\Sigma^2)}{\text{tr}(\Sigma^2)_{22} \text{tr}(\Sigma^2_{11})} \right\} \]

\[ \leq K \frac{1}{\sqrt{2\pi}} \frac{2}{K \frac{\text{tr}(\Sigma^2)}{d^{1/2} K_p^{3/2} \lambda_{\text{max}}(\Sigma)}} \exp \left\{ -\frac{\ell^2}{K^2} \frac{\text{tr}(\Sigma^2)}{d^3 M_p^3 \lambda_{\text{max}}(\Sigma)} \right\} \leq p^{-t_p}, \]

\[ P(|\Theta_7| \geq \ell \sigma_S) \]

63
Thus, we have

\[ P(|R_p| \geq \sigma_S) \leq \frac{1}{p^{t_p}}. \quad (31) \]
Now, for clarity, we will revise the definition of $A_p$ as follows

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

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

$$P\{A_p(x)B_1 \cdots B_d\} \leq P\left\{A_p(x)B_1 \cdots B_d, \frac{|R_p|}{\sigma_S} < \iota \right\} + \frac{1}{p^\iota}$$

$$\leq P\left(\frac{S_p}{\sigma_S} \leq x + \iota, \ B_1 \cdots B_d \right) + \frac{1}{p^\iota}$$

$$= P\left(\frac{S_p}{\sigma_S} \leq x + \iota \right) \cdot P(B_1 \cdots B_d) + \frac{1}{p^\iota}$$

by the independence appeared in (29). Hence,

$$P\left(\frac{S_p}{\sigma_S} \leq x + \iota \right) \leq P\left(\frac{S_p}{\sigma_S} \leq x + \iota, \frac{|R_p|}{\sigma_S} < \iota \right) + \frac{1}{p^\iota}$$

$$\leq P\left\{ \frac{1}{\sigma_S} (S_p + R_p) \leq x + 2\iota \right\} + \frac{1}{p^\iota} \leq P\left\{ A_p(x + 2\iota) \right\} + \frac{1}{p^\iota}.$$

Combine the two inequalities to get

$$P\{A_p(x)B_1 \cdots B_d\} \leq P\left\{ A_p(x + 2\iota) \right\} \cdot P(B_1 \cdots B_d) + \frac{2}{p^\iota}. \quad (32)$$

Similarly,

$$P\left(\frac{S_p}{\sigma_S} \leq x - \iota, \ B_1 \cdots B_d \right)$$

65
\[
\leq \Pr\left(\frac{S_p}{\sigma_S} \leq x - t, B_1 \cdots B_d, \frac{|R_p|}{\sigma_S} < t\right) + \frac{1}{p^t} \leq \Pr\left(\frac{S_p}{\sigma_S} \leq x, B_1 \cdots B_d\right) + \frac{1}{p^t}.
\]

By the independence from (29),

\[
\Pr\{A_p(x)B_1 \cdots B_d\} \geq \Pr\left(\frac{S_p}{\sigma_S} \leq x - t\right) \cdot \Pr(B_1 \cdots B_d) - \frac{1}{p^t}.
\]

Furthermore,

\[
\Pr\left(\frac{T_{SUM}}{\sigma_S} \leq x - 2t\right) \leq \Pr\left(\frac{T_{SUM}}{\sigma_S} \leq x - 2t, \frac{|R_p|}{\sigma_S} < t\right) + \frac{1}{p^t} \leq \Pr\left(\frac{S_p}{\sigma_S} \leq x - t\right) + \frac{1}{p^t},
\]

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

\[
\Pr\{A_p(x)B_1 \cdots B_d\} \geq \Pr\{A_p(x - 2t)\} \cdot \Pr(B_1 \cdots B_d) - \frac{2}{p^t}.
\]

This together with (32) concludes

\[
|\Pr\{A_p(x)B_1 \cdots B_d\} - \Pr\{A_p(x)\} \cdot \Pr(B_1 \cdots B_d)| \leq \Delta_{p,t} \cdot \Pr(B_1 \cdots B_d) + \frac{2}{p^t} \tag{33}
\]

as \(p \geq p_0\), where

\[
\Delta_{p,t} \doteq |\Pr\{A_p(x)\} - \Pr\{A_p(x + 2t)\}| + |\Pr\{A_p(x)\} - \Pr\{A_p(x - 2t)\}|
\]

\[
= \Pr\{A_p(x + 2t)\} - \Pr\{A_p(x - 2t)\}
\]

66
since $P\{A_p(x)\}$ is increasing in $x \in \mathbb{R}$. An important observation is that the derivation of (30) is based on three key facts: inequality (30), the identity $T_{SUM} = S_p + R_p$ and the fact $U_t$ and $\{\varepsilon_{t1}, \cdots, \varepsilon_{td}\}$ are independent from (29).

Thus, the three corresponding key facts aforementioned also hold for the quantities related to $\Lambda = \{i_1, \cdots, i_d\}$. Therefore, similar to the derivation of (33), we have

$$
\left| P\{A_p(x)B_{i_1} \cdots B_{i_d}\} - P\{A_p(x)\} \cdot P(B_{i_1} \cdots B_{i_d}) \right| \\
\leq \Delta_{p,t} \cdot P(B_{i_1} \cdots B_{i_d}) + \frac{2}{p^t}
$$

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

$$
\zeta(p, d) = \sum_{1 \leq i_1 < \cdots < i_d \leq p} \left| P\{A_p(x)B_{i_1} \cdots B_{i_d}\} - P\{A_p(x)\} \cdot P(B_{i_1} \cdots B_{i_d}) \right| \\
\leq \sum_{1 \leq i_1 < \cdots < i_d \leq N} \left\{ \Delta_{p,t} \cdot P(B_{i_1} \cdots B_{i_d}) + \frac{2}{p^t} \right\} \\
\leq \Delta_{p,t} \cdot H(d, p) + \binom{p}{d} \cdot \frac{2}{p^t}
$$

(34)

where

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

In the following, we will show $\lim_{\varepsilon \downarrow 0} \limsup_{p \to \infty} \Delta_{p,t} = 0$ and $\limsup_{p \to \infty} H(d, p) < \infty$ for each $d \geq 1$. Assuming these are true, by using $\binom{p}{d} \leq p^d$ and (34), for fixed $d \geq 1$, sending $p \to \infty$ first, then sending $t \downarrow 0$, we get $\lim_{p \to \infty} \zeta(p, d) = 0$ for each $d \geq 1$. The proof is
then completed.

**Step 2:** the proofs of \(\lim_{\iota \to 0} \lim \sup_{p \to \infty} \Delta_{p,\iota} = 0\) and \(\lim \sup_{p \to \infty} H(d, p) < \infty\) for each \(d \geq 1\).

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

\[
\frac{T_{\text{SUM}}}{\sigma_S} \to \mathcal{N}(0, 1) \text{ weakly} \tag{35}
\]

as \(p \to \infty\) and hence

\[
\Delta_{p,\iota} \to \Phi(x + 2\iota) - \Phi(x - 2\iota) \tag{36}
\]

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

\[
\lim_{\iota \to 0} \lim \sup_{p \to \infty} \Delta_{p,\iota} = 0.
\]

Second, under the normality assumption, by Conditions (C2) and (C3), Theorem 1 holds. Hence, by identifying \(H(t, p)\) here as \(\alpha_t\) in (21) for each \(t \geq 1\), we obtain

\[
\lim_{p \to \infty} H(d, p) = \frac{1}{d!} \pi^{-d/2} e^{-nx/2} \tag{37}
\]

for each \(d \geq 1\). The proof is finished. \(\square\)

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,

\[ P \left( \frac{T_{\text{SUM}}}{\sigma_S} \leq x \right) = \Phi(x) \]  

(38)
as \( p \to \infty \) for any \( x \in \mathbb{R} \), where \( \sigma_S = (2Kn^{-2}\text{tr}(\Sigma^2))^{1/2} \) and \( \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \, dt \).

Define \( N = Kp^2 \) and \( Z_{i+(j-1)p+(k-1)p^2} = n^{1/2}\rho_{ij}(k), i, j = 1, \ldots, p, k = 1, \ldots, K \). From Theorem 1 we have

\[ P \left( \max_{1 \leq i \leq N} Z_i^2 - 2 \log N + \log \log N \leq y \right) \to G(y) = \exp \left( - \frac{1}{\sqrt{\pi}} e^{-y/2} \right) \]  

(39)
as \( N \to \infty \) for any \( y \in \mathbb{R} \). To show asymptotic independence, it is enough to prove

\[ \lim_{N \to \infty} P \left( \frac{T_{\text{SUM}}}{\sigma_S} \leq x, \max_{1 \leq i \leq N} Z_i^2 - 2 \log N + \log \log N \leq y \right) = \Phi(x) \cdot G(y) \]  

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

\[ L_N = \max_{1 \leq i \leq N} |Z_i| \]  

and \( l_N = (2 \log N - \log \log N + y)^{1/2} \),

(40)

where the latter one makes sense for large \( N \). Because of (38), the above is equivalent to that

\[ \lim_{N \to \infty} P \left( \frac{T_{\text{SUM}}}{\sigma_S} \leq x, L_N > l_N \right) = \Phi(x) \cdot \{1 - G(y)\} \]  

(41)
for any $x \in \mathbb{R}$ and $y \in \mathbb{R}$. Recalling the notation in Lemma 7, we have

$$A_p = \left\{ \frac{T_{\text{SUM}}}{\sigma_S} \leq x \right\} \quad \text{and} \quad B_i = \{ |Z_i| > l_N \}$$

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

$$P\left( \frac{T_{\text{SUM}}}{\sigma_S} \leq x, \; L_N > l_N \right) = P\left( \bigcup_{i=1}^{N} A_p B_i \right).$$

(43)

From the inclusion-exclusion principle,

$$P\left( \bigcup_{i=1}^{N} A_p B_i \right) \leq \sum_{1 \leq i_1 \leq N} P(A_{p_i} B_{i_1}) - \sum_{1 \leq i_1 < i_2 \leq N} P(A_{p_i} B_{i_1} B_{i_2}) + \cdots + \sum_{1 \leq i_1 < \cdots < i_{2k+1} \leq N} P(A_{p_i} B_{i_1} \cdots B_{i_{2k+1}})$$

(44)

and

$$P\left( \bigcup_{i=1}^{N} A_p B_i \right) \geq \sum_{1 \leq i_1 \leq N} P(A_{p_i} B_{i_1}) - \sum_{1 \leq i_1 < i_2 \leq N} P(A_{p_i} B_{i_1} B_{i_2}) + \cdots - \sum_{1 \leq i_1 < \cdots < i_{2k} \leq N} P(A_{p_i} B_{i_1} \cdots B_{i_{2k}})$$

(45)

for any integer $k \geq 1$. Define

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

70
for \( d \geq 1 \). From (37) we know

\[
\lim_{d \to \infty} \limsup_{p \to \infty} H(N, d) = 0. \tag{46}
\]

Set

\[
\zeta(N, d) = \sum_{1 \leq i_1 < \cdots < i_d \leq N} \left\{ P(A_{p_i}B_{i_1} \cdots B_{i_d}) - P(A_p) \cdot P(B_{i_1} \cdots B_{i_d}) \right\}
\]

for \( d \geq 1 \). By Lemma 7,

\[
\lim_{p \to \infty} \zeta(N, d) = 0 \tag{47}
\]

for each \( d \geq 1 \). The assertion (44) implies that

\[
P \left( \bigcup_{i=1}^{N} A_pB_i \right) \leq P(A_p) \left\{ \sum_{1 \leq i_1 \leq N} P(B_{i_1}) - \sum_{1 \leq i_1 < i_2 \leq N} P(B_{i_1}B_{i_2}) + \cdots - \sum_{1 \leq i_1 < \cdots < i_{2k} \leq N} P(B_{i_1} \cdots B_{i_{2k}}) \right\} + \sum_{d=1}^{2k} \zeta(N, d) + H(N, 2k + 1)
\]

\[
\leq P(A_p) \cdot P\left( \bigcup_{i=1}^{N} B_i \right) + \sum_{d=1}^{2k} \zeta(N, d) + H(N, 2k + 1), \tag{48}
\]

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

\[
P \left( \bigcup_{i=1}^{N} B_i \right) \geq \sum_{1 \leq i_1 \leq N} P(B_{i_1}) - \sum_{1 \leq i_1 < i_2 \leq N} P(B_{i_1}B_{i_2}) + \cdots - \sum_{1 \leq i_1 < \cdots < i_{2k} \leq N} P(B_{i_1} \cdots B_{i_{2k}})
\]
for all $k \geq 1$. By the definition of $l_N$ and (39),

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

as $p \to \infty$. By (38), $P(A_p) \to \Phi(x)$ as $p \to \infty$. From (43), (47) and (48), by fixing $k$ first and sending $p \to \infty$, we obtain that

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

Now, by letting $k \to \infty$ and using (46) we have

$$
\lim sup_{p \to \infty} P\left(\frac{T_{\text{SUM}}}{\sigma_S} \leq x, L_N > l_N\right) \leq \Phi(x)\{1 - G(y)\}. \quad (49)
$$

By applying the same argument to (45), we see that the counterpart of (48) becomes

$$
P\left(\bigcup_{i=1}^{N} A_p B_i\right) \geq P(A_p)\left\{ \sum_{1 \leq i_1 \leq N} P(B_{i_1} \ldots B_{i_{2k-1}}) \right\} + \sum_{d=1}^{2k-1} \zeta(N, d) - H(N, 2k)
$$

$$
\geq P(A_p) \cdot P\left(\bigcup_{i=1}^{N} B_i\right) + \sum_{d=1}^{2k-1} \zeta(N, d) - H(N, 2k),
$$

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

$$
P\left(\bigcup_{i=1}^{N} B_i\right) \leq \sum_{1 \leq i_1 \leq N} P(B_{i_1}) - \sum_{1 \leq i_1 < i_2 \leq N} P(B_{i_1} B_{i_2})
$$
for all \( k \geq 1 \). Review (43) and repeat the earlier procedure to see

\[
\liminf_{p \to \infty} P\left( \frac{T_{\text{SUM}}}{\sigma_S} \leq x, \; L_N > l_N \right) \geq \Phi(x)\{1 - G(y)\} \quad (50)
\]

by sending \( p \to \infty \) and then sending \( k \to \infty \). Here (50) and (49) yield (41). The proof is then completed.

6.8 Proof of Theorem 7

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

\[
P\left\{ \max_{1 \leq k \leq K} \max_{1 \leq i,j \leq p} n \hat{p}_{ij}^2(k) - 2 \log(K \hat{p}^2) + \log \log(K \hat{p}^2) \leq x, T_{\text{SUM}}/\sigma_S \leq y \right\} \to G(x) \cdot \Phi(y),
\]

(51)

Define

\[
W(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{1}{n(n-1)\sigma_S} \sum_{t=1}^{K} \sum_{t \neq s} \mathbf{x}_t^\top \mathbf{x}_s \mathbf{x}_{t+s}^\top \mathbf{x}_{s+t}.
\]

Hence, \( T_{\text{SUM}}/\sigma_S = W(\epsilon_1, \ldots, \epsilon_n) \).

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

\[
F_{\beta}(z) = \beta^{-1} \log \left( \sum_{j=1}^{p} \exp (\beta z_j) \right),
\]
where $\beta > 0$ is the smoothing parameter that controls the level of approximation. An elementary calculation shows that for all $z \in \mathbb{R}^p$,

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

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

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

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

$$P \left\{ V^2(\bm{\varepsilon}_1, \cdots, \bm{\varepsilon}_n) - 2 \log (Kp^2) + \log \log (Kp^2) \leq x, W(\bm{\varepsilon}_1, \cdots, \bm{\varepsilon}_n) \leq y \right\} \to G(x) \cdot \Phi(y).$$

(52)

Suppose $\bm{\xi}_1, \cdots, \bm{\xi}_n$ are independent and identical distributed as $\mathcal{N}(\bm{0}, \Sigma)$ and independent of $(\bm{\varepsilon}_1, \cdots, \bm{\varepsilon}_n)$. Next, we will show that $(W(\bm{\varepsilon}_1, \cdots, \bm{\varepsilon}_n), V(\bm{\varepsilon}_1, \cdots, \bm{\varepsilon}_n))$ has the same limited distribution as $(W(\bm{\xi}_1, \cdots, \bm{\xi}_n), V(\bm{\xi}_1, \cdots, \bm{\xi}_n))$. 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 $\xi$ if and only if for every $f \in \mathcal{C}_b^2(\mathbb{R}^2)$, $E(f(\xi_n)) \to E(f(\xi))$; see, e.g., Pollard (1984), Chapter III, Theorem 12. We use this property to give a metrization of the weak convergence in $\mathbb{R}^2$. 

74
Thus, we only need to show that

$$\mathbb{E}[f(W(\varepsilon_1, \ldots, \varepsilon_n), V(\varepsilon_1, \ldots, \varepsilon_n))] - \mathbb{E}[f(W(\xi_1, \ldots, \xi_n), V(\xi_1, \ldots, \xi_n))] \to 0$$

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

$$W_d = W(\varepsilon_1, \ldots, \varepsilon_{d-1}, \xi_d, \ldots, \varepsilon_n), \quad V_d = V(\varepsilon_1, \ldots, \varepsilon_{d-1}, \xi_d, \ldots, \varepsilon_n).$$

We have

$$|\mathbb{E}[f(W(\varepsilon_1, \ldots, \varepsilon_n), V(\varepsilon_1, \ldots, \varepsilon_n))] - \mathbb{E}[f(W(\xi_1, \ldots, \xi_n), V(\xi_1, \ldots, \xi_n))]|$$

$$\leq \sum_{d=1}^{n} |\mathbb{E}(f(W_d, V_d)) - \mathbb{E}(f(W_{d+1}, V_{d+1}))|.$$

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

Define

$$W_{d,0} = \frac{1}{n(n-1)\sigma_S} \sum_{1 \leq t \neq s \leq d-2} x_t^\top x_s x_{t+1}^\top x_{s+1} + \frac{1}{n(n-1)\sigma_S} \sum_{d+1 \leq t \neq s \leq n} x_t^\top x_s x_{t+1}^\top x_{s+1}$$

$$+ \frac{2}{n(n-1)\sigma_S} \sum_{t=1}^{d-2} \sum_{s=d+1}^{n} x_t^\top x_s x_{t+1}^\top x_{s+1},$$

75
which only relies on $F_d = \sigma\{\varepsilon_1, \ldots, \varepsilon_{d-2}, \xi_{d+1}, \ldots, \xi_n\}$. Hence,

$$
W_d - W_{d,0} = \frac{1}{n(n-1)\sigma_S} \sum_{t=1}^{d-2} \varepsilon_{d-1}^\top \varepsilon_t \varepsilon_{d+1}^\top + \frac{1}{n(n-1)\sigma_S} \sum_{t=d}^n \varepsilon_{d-1}^\top \xi_t \xi_{d+1}^\top \\
+ \frac{1}{n(n-1)\sigma_S} \sum_{t=1}^{d-2} \xi_{d+1}^\top \varepsilon_t \xi_{d+1}^\top \\
+ \frac{1}{n(n-1)\sigma_S} \sum_{t=d}^n \xi_{d+1}^\top \xi_t \xi_{d+1}^\top + \frac{1}{n(n-1)\sigma_S} \varepsilon_{d-1}^\top \varepsilon_{d+1}^\top.
$$

$$
W_{d+1} - W_{d,0} = \frac{1}{n(n-1)\sigma_S} \sum_{t=1}^{d-2} \varepsilon_{d-1}^\top \varepsilon_t \varepsilon_{d+1}^\top + \frac{1}{n(n-1)\sigma_S} \sum_{t=d}^n \varepsilon_{d-1}^\top \xi_t \xi_{d+1}^\top \\
+ \frac{1}{n(n-1)\sigma_S} \sum_{t=1}^{d-2} \xi_{d+1}^\top \varepsilon_t \xi_{d+1}^\top \\
+ \frac{1}{n(n-1)\sigma_S} \sum_{t=d}^n \xi_{d+1}^\top \xi_t \xi_{d+1}^\top + \frac{1}{n(n-1)\sigma_S} \varepsilon_{d-1}^\top \varepsilon_{d+1}^\top.
$$

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

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

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

$$
V_{d,0} = \beta^{-1} \log \left( \sum_{l=1}^{p^2} \exp \left( \beta n^{1/2} \hat{\rho}_l^{(d,0)} \right) \right).
$$
Similarly, we define $l = i + (j - 1)p$ and

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

for all pairs $(i,j)$. Then,

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

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

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

and

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

Because \( \mathbb{E}(\varepsilon_t) = \mathbb{E}(\xi_t) = 0 \) and \( \mathbb{E}(\varepsilon_t\varepsilon_t^\top) = \mathbb{E}(\xi_t\xi_t^\top) \), we can verify that

\[
\mathbb{E}(W_d - W_{d,0}|F_d) = \mathbb{E}(W_{d+1} - W_{d,0}|F_d),
\]

\[
\mathbb{E}((W_d - W_{d,0})^2|F_d) = \mathbb{E}((W_{d+1} - W_{d,0})^2|F_d).
\]

Thus,

\[
\mathbb{E}(f_1(W_{d,0}, V_{d,0})(W_d - W_{d,0})) = \mathbb{E}(f_1(W_{d,0}, V_{d,0})(W_{d+1} - W_{d,0})),
\]

\[
\mathbb{E}(f_{11}(W_{d,0}, V_{d,0})(W_d - W_{d,0})^2) = \mathbb{E}(f_{11}(W_{d,0}, V_{d,0})(W_{d+1} - W_{d,0})^2).
\]

Next, we consider \( V_d - V_{d,0} \). Define \( z_{d,0} = (n^{1/2}\hat{\rho}_1^{(d,0)}, \ldots, n^{1/2}\hat{\rho}_{p^2}^{(d,0)})^\top \)
and \( z_d = (n^{1/2}\hat{\rho}_1^{(d)}, \ldots, n^{1/2}\hat{\rho}_{p^2}^{(d)})^\top \). By Taylor’s expansion, we have

\[
V_d - V_{d,0} = n^{1/2}\sum_{l=1}^{p^2} \hat{\partial}_l F_\beta(z_{d,0})(\hat{\rho}_l^{d} - \hat{\rho}_l^{d,0}) + \frac{n}{2}\sum_{l=1}^{p^2}\sum_{k=1}^{p^2} \hat{\partial}_k\hat{\partial}_l F_\beta(z_{d,0})(\hat{\rho}_l^{d} - \hat{\rho}_l^{d,0})(\hat{\rho}_k^{d} - \hat{\rho}_k^{d,0})
\]

\[
+ \frac{1}{6} n^{3/2}\sum_{l=1}^{p^2}\sum_{k=1}^{p^2}\sum_{q=1}^{p^2} \hat{\partial}_q\hat{\partial}_k\hat{\partial}_l F_\beta(z_{d,0} + \delta(z_d - z_{d,0}))(\hat{\rho}_l^{d} - \hat{\rho}_l^{d,0})(\hat{\rho}_k^{d} - \hat{\rho}_k^{d,0})(\hat{\rho}_q^{d} - \hat{\rho}_q^{d,0}).
\]

(53)

By \( \mathbb{E}(\varepsilon_t) = \mathbb{E}(\xi_t) = 0 \) and \( \mathbb{E}(\varepsilon_t\varepsilon_t^\top) = \mathbb{E}(\xi_t\xi_t^\top) \), we can also verify that

\[
\mathbb{E}((\hat{\rho}_l^{d} - \hat{\rho}_l^{d,0})|F_d) = \mathbb{E}((\hat{\rho}_l^{d+1} - \hat{\rho}_l^{d,0})|F_d),
\]

\[
\mathbb{E}((\hat{\rho}_l^{d} - \hat{\rho}_l^{d,0})^2|F_d) = \mathbb{E}((\hat{\rho}_l^{d+1} - \hat{\rho}_l^{d,0})^2|F_d),
\]

78
By Lemma A.2 in Chernozhukov et al. (2013), we have
\[
\left| \sum_{l=1}^{p^2} \sum_{k=1}^{p^2} \sum_{q=1}^{p^2} \hat{c}_q \hat{c}_k \hat{c}_l F_\beta(z_{d,0} + \delta(z_d - z_{d,0})) \right| \leq C \beta^2
\]
for some positive constant \(C\). By Condition (C1'), if \(\varepsilon_t\) has polynomial-type tails, we have
\[
P\left( \max_{1 \leq t \leq n, 1 \leq i \leq p} |\varepsilon_{it}| > C n^{\frac{1}{2}-\delta} \right) \leq np(C n^{\frac{1}{2}-\delta})^6 \gamma_0 + \epsilon \mathbb{E}(\frac{|\varepsilon_t|}{\sigma_t})^6 \gamma_0 + \epsilon \to 0,
\]
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
\[
P \left( \max_{1 \leq t \leq n, 1 \leq i \leq p} |\xi_{it}| > C \log(np) \right) \to 0.
\]
Thus, we have
\[
\frac{1}{6} n^{3/2} \sum_{l=1}^{p^2} \sum_{k=1}^{p^2} \sum_{q=1}^{p^2} \hat{c}_q \hat{c}_k \hat{c}_l F_\beta(z_{d,0} + \delta(z_d - z_{d,0}))(\hat{p}_d^l - \hat{p}_d^{d,0})(\hat{p}_d^k - \hat{p}_d^{d,0})(\hat{p}_d^q - \hat{p}_d^{d,0})
\leq C \beta^2 n^{-7/6 - 2\delta}
\]
as probability tending to one. Hence, we have
\[
|\mathbb{E}(f_2(W_{d,0}, V_{d,0})(V_d - V_{d,0})) - \mathbb{E}(f_2(W_{d,0}, V_{d,0})(V_{d+1} - V_{d,0}))| \leq \beta^2 n^{-7/6 - 2\delta}.
\]
Similarly, we can show that

\[
\begin{align*}
|\mathbb{E}(f_{22}(W_{d,0}, V_{d,0})(V_d - V_{d,0})^2) - \mathbb{E}(f_{22}(W_{d,0}, V_{d,0})(V_{d+1} - V_{d,0})^2)| & \leq \beta^2 n^{-7/6 - 2\delta}, \\
|\mathbb{E}(f_{12}(W_{d,0}, V_{d,0})(W_d - W_{d,0})(V_d - V_{d,0})) - \mathbb{E}(f_{12}(W_{d,0}, V_{d,0})(W_{d+1} - W_{d,0})(V_{d+1} - V_{d,0}))| & \leq \beta^2 n^{-7/6 - 2\delta}.
\end{align*}
\]

Hence, we can have

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

By (53), 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

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

as \(n \to \infty\). Then, we obtain the result. If \(\varepsilon_t\) has sub-gaussian-type tails, we can also prove the result by the similar arguments.
6.9 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

\[
A_0 = \begin{pmatrix} A_{011} & 0 \\ 0 & A_{022} \end{pmatrix}, \quad A_1 = \begin{pmatrix} A_{111} & 0 \\ 0 & 0 \end{pmatrix}.
\]

Under the alternative hypothesis, we have \(X_{t1} = A_{011}y_t + A_{111}y_{t-1}\), where \(y_t = (z_{t1}, \ldots, z_{td})^\top\) is independent of \(X_{t2} = A_{022}(z_{td+1}, \ldots, z_{tp})^\top\). As \(\varepsilon_t = (X_{t1}^\top, X_{t2}^\top)^\top\), we can decompose \(T_{\text{SUM}}\) as follows

\[
T_{\text{SUM}} = \frac{1}{n(n - 1)} \sum_{t \neq s} \sum_{t+1} \varepsilon_t^\top \varepsilon_s \varepsilon_{t+1}^\top \varepsilon_{s+1}
\]

\[
= \frac{1}{n(n - 1)} \sum_{t \neq s} (X_{t1}^\top X_{s1} + X_{t2}^\top X_{s2})(X_{t+1,1}^\top X_{s+1,1} + X_{t+1,2}^\top X_{s+1,2})
\]

\[
= \frac{1}{n(n - 1)} \sum_{t \neq s} X_{t1}^\top X_{s1} X_{t+1,1}^\top X_{s+1,1} + \frac{1}{n(n - 1)} \sum_{t \neq s} X_{t1}^\top X_{s1} X_{t+1,2}^\top X_{s+1,2}
\]

\[
+ \frac{1}{n(n - 1)} \sum_{t \neq s} X_{t2}^\top X_{s2} X_{t+1,1}^\top X_{s+1,1} + \frac{1}{n(n - 1)} \sum_{t \neq s} X_{t2}^\top X_{s2} X_{t+1,2}^\top X_{s+1,2}.
\]

Similar to the proof of Theorem 5, we have

\[
\sigma_{S1}^{-2} \text{var} \left( \frac{1}{n(n - 1)} \sum_{t \neq s} X_{t1}^\top X_{s1} X_{t+1,1}^\top X_{s+1,1} \right) = O \left( \frac{d^2}{p^2} \right),
\]

81
\[
\sigma_{S1}^{-2} \text{var} \left( \frac{1}{n(n-1)} \sum_{t\neq s} \sum_{t+s} X_{t1} X_{s1} X_{s1}^\top X_{s1} \right) = O \left( \frac{d}{p} \right),
\]

\[
\sigma_{S1}^{-2} \text{var} \left( \frac{1}{n(n-1)} \sum_{t\neq s} \sum_{t+s} X_{t2} X_{s2} X_{s2}^\top X_{s2} \right) = O \left( \frac{d}{p} \right),
\]

by the condition that the eigenvalues of \( \Sigma \) are all bounded. Thus, we have

\[
T_{\text{SUM}} = \frac{1}{n(n-1)} \sum_{t\neq s} \sum_{t+s} X_{t2} X_{s2} X_{s2}^\top X_{s2} + E(T_{\text{SUM}}) + o_p(\sigma_{S1})
\]

\[
= T_{\text{SUM}}^{(2)} + E(T_{\text{SUM}}) + o_p(\sigma_{S1}).
\]

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

\[
T_{\text{MAX}} = \max_{1\leq i,j \leq p} |n^{1/2} \tilde{\rho}_{ij}(1)| + o_p(1)
\]

\[
= \max \left\{ \max_{1\leq i,j \leq d} |n^{1/2} \tilde{\rho}_{ij}(1)|, \max_{1\leq i\leq d, d+1\leq j \leq p} |n^{1/2} \tilde{\rho}_{ij}(1)|, \max_{d+1\leq i, j \leq p} |n^{1/2} \tilde{\rho}_{ij}(1)| \right\} + o_p(1).
\]

By the independence between \( X_{t1} \) and \( X_{t2} \), we know that \( \max_{1\leq i,j \leq d} |n^{1/2} \tilde{\rho}_{ij}(1)| \) is independent of \( T_{\text{SUM}}^{(2)} \). 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 \( T_{\text{SUM}}^{(2)} \). Because \( X_{t1} \) is independent of \( X_{t2} \), we also can prove that \( \max_{d+1\leq i,j \leq p} |n^{1/2} \tilde{\rho}_{ij}(1)| \) is asymptotically independent of \( T_{\text{SUM}}^{(2)} \) by taking the same procedure as Theorem 6. Thus, we can prove that \( T_{\text{MAX}} \) is asymptotically independent of \( T_{\text{SUM}}^{(2)} \). By Lemma 7.10 in [Feng et al., 2022], we can complete the proof.
References

Cai, T., Liu, W. and Xia, Y. (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.

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

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013), ‘Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors this seminar will be held at ucl - drayton room 321’, *Annals of Statistics* **41**, 2786–2819.

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

Feng, L., Jiang, T., Li, P. and Liu, B. (2022), ‘Asymptotic independence of the sum and maximum of dependent random variables with applications to high-dimensional tests’, *arXiv:2205.01638*.

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

Jing, B., Shao, Q. and Wang, Q. (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*, New York: 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.