Statistica Sinica Preprint No: SS-2019-0319	
Title	Hypothesis Testing for Block-structured Correlation for High Dimensional Variables
Manuscript ID	SS-2019-0319
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0319
Complete List of Authors	Shurong Zheng, Xuming He and Jianhua Guo
Corresponding Author	Shurong Zheng
E-mail	zhengsr@nenu.edu.cn
Notice: Accepted version subject to English editing.	

Hypothesis Testing for Block-structured Correlation for High-dimensional Variables

Shurong Zheng[1], Xuming He[2] and Jianhua Guo[1]

[1]*Northeast Normal University, China and* [2]*University of Michigan, USA*

Abstract: Testing independence or block-independence of high dimensional random vectors is of great importance in multivariate statistical analysis. Recent work on high dimensional block independence tests aim to extend their validity beyond specific distributions (e.g., Gaussian) or restrictive block sizes. In this paper, we propose a new and powerful test on block-structured correlation of high dimensional random vectors for sparse or non-sparse alternatives without strict distributional assumptions. The statistical properties of the proposed test are developed under the asymptotic regime that the dimension grows proportionally with the sample size. Empirically, we find that the proposed test outperforms the existing tests we have considered for a variety of alternatives and works quite well when there are few existing tests at our disposal.

Key words and phrases: Testing block-independence, high-dimension, multivariate statistical analysis, sparse alternatives, non-sparse alternatives.

1. Introduction

Driven by a wide range of scientific applications, testing independence

of random vectors is of great importance in multivariate statistical analysis. In the conventional low-dimensional setting with $p/n \to 0$, where p is the dimension of the random vector and n is the sample size, both complete and block independence tests are well established. For complete independence, Anderson (2003) detailed the *likelihood ratio test* (LRT) for the Gaussian population. For block independence, Wilks (1935) and Sugiura & Fujikoshi (1969) developed effective likelihood ratio tests for the Gaussian population and derived their asymptotic distributions under regularity conditions.

In the high-dimensional setting, the classical LRT is invalid or cannot be defined as the dimension p becomes greater than the sample size n. In recent years, researchers have made great advances on high-dimensional independence tests. For complete independence, Bai *et al.* (2009) proposed the corrected LRT when $p/n \to y \in (0, 1)$. Jiang & Yang (2013) studied the LRT when $p/n \to y \in (0, 1]$. Schott (2005) developed a test based on the Frobenius norm of the sample correlation matrix under the case of $p > n$. Zhou (2007) and Cai & Jiang (2011) extended the results of Jiang (2004) to obtain the extreme distribution of coherence of the sample correlation matrices. Li & Xue (2015) proposed a quadratic type statistic and an extreme-value type statistic. For high dimensional block independence, Jiang, Bai & Zheng (2013) developed a corrected LRT and trace test as

$p/n \to y \in (0,1)$. Jiang & Yang (2013) studied the LRT for the Gaussian population when $p/n \to (0,1]$. Bao, Hu, Pan & Zhou (2016) proposed a Schott type statistic when the dimension of every block of random variables is less than the sample size. Yamada, Hyodo and Nishiyama (2017) allowed a more general setting by using the Frobenius norm of the sample covariance matrix. Paindaveine and Verdebout (2016) proposed a high dimensional sign test for block-structured correlation between random variables of two blocks under appropriate symmetry assumptions.

This paper aims to develop a new and powerful test on block-structured correlation of a high dimensional random vector for sparse or non-sparse alternatives under no strict distributional assumptions under the asymptotic regime of $p/n \to y \in (0,\infty)$. To this end, we propose a two-term test statistic. The first term is $T_{n1} = \mathrm{tr}[\mathbf{S}_n - \mathrm{diag}(\mathbf{S}_{11},\ldots,\mathbf{S}_{KK})]^2$, where the sample covariance matrix $\mathbf{S}_n$ is a natural estimator of the population covariance matrix and the block-diagonal matrix $\mathrm{diag}(\mathbf{S}_{11},\ldots,\mathbf{S}_{KK})$ is a population covariance matrix estimator under block-structured correlation. The statistic T_{n1} does not impose any condition on the dimension because T_{n1} involves no matrix inversion. The statistic T_{n1} is the total sum of the squared entries of $\mathbf{S}_n - \mathrm{diag}(\mathbf{S}_{11},\ldots,\mathbf{S}_{KK})$ to capture the overall difference between $\mathbf{S}_n$ and $\mathrm{diag}(\mathbf{S}_{11},\ldots,\mathbf{S}_{KK})$ even if the individual entries of

$\mathbf{S}_n - \mathrm{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{KK})$ are small. That is, T_{n1}, similar to the test used in Yamada, Hyodo and Nishiyama (2017), will have good power for non-sparse alternatives. The second term is a screening term T_{n0} which is added to T_{n1} for enhancing the power under sparse alternatives. Then the proposed test statistic $T_{n1} + T_{n0}$ is effective not only for non-sparse alternatives but also for sparse alternatives. To examine the performance of the proposed test statistic, the limiting null distribution is derived as $p/n \to y \in (0, \infty)$, allowing y to be greater than 1. Simulation studies show that Type I errors of the proposed test can be well maintained. Moreover, under the alternative hypothesis, the limiting distribution of the proposed test is discussed, and the asymptotic unbiasedness of the proposed test is proved. When the dimension is smaller than the sample size, simulation studies are conducted to compare our proposed test with the existing tests for the Gaussian population. For comparison of empirical powers, our proposed test performs favorably over other tests designed for high dimensions. Even when the population is non-Gaussian and the dimension is greater than the sample size, our proposed test performs well in our studies.

The organization of this paper is as follows. In Section 2, we propose the test statistic, derive its limiting distribution under the null hypothesis and the alternative hypothesis, and present the asymptotic power function

to show that the proposed test is asymptotically unbiased. In Section 3, we conduct simulation studies for comparing the proposed test with several existing tests. A real data set is analyzed in Section 4 for illustration. Section 5 concludes with a discussion. Some proofs are given in the Appendix.

2. Test on block-structured correlation

Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a random sample from the p-dimensional population random vector $\mathbf{x} = (x_1, \ldots, x_p)^\top$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\bar{\mathbf{x}} = n^{-1}\sum_{i=1}^n \mathbf{x}_i$ and $\mathbf{S}_n = (n-1)^{-1}\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$ be the sample mean and sample covariance matrix, respectively. Without loss of generality, the random vector $\mathbf{x} = (x_1, \ldots, x_p)^T$ can be formulated by K random variable blocks: $\{x_1, \ldots, x_{p_1}\}$, $\{x_{p_1+1}, \ldots, x_{p_1+p_2}\}$, $\ldots$, $\{x_{p_1+p_2+\cdots+p_{K-1}+1}, \ldots, x_p\}$, where $p = p_1 + \cdots + p_K$ and K is permitted to increase with n at some rate. Let $\boldsymbol{\Sigma}_{ij}$ be the covariance matrix of the i-th and j-th random variable blocks. The population and sample covariance matrices can be partitioned into $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{ij})_{i,j=1}^K$ and $\mathbf{S}_n = (\mathbf{S}_{ij})_{i,j=1}^K$, respectively. Testing block-structured correlation of $\mathbf{x}$ can be formulated as testing

$$H_0 : \boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\Sigma}_{11}, \ldots, \boldsymbol{\Sigma}_{KK}), \tag{2.1}$$

where $\mathrm{diag}(\mathbf{\Sigma}_{11}, \ldots, \mathbf{\Sigma}_{KK})$ is the block-diagonal matrix from K blocks $\{\mathbf{\Sigma}_{kk}, k = 1, \ldots, K\}$. A natural estimator of $\mathbf{\Sigma}$ is $\mathbf{S}_n$. Under the null hypothesis, a natural estimator of $\mathbf{\Sigma}$ is $\mathrm{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{KK})$. For the Gaussian population, the LRT statistic is (Wilks, 1935)

$$\log |\mathbf{S}_n| - \log |\mathrm{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{KK})|,$$

which is the entropy loss of $\mathbf{S}_n$ and $\mathrm{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{KK})$. The entropy loss for covariance matrix estimation can be found in James & Stein (1961) and Muirhead (1982). Jiang, Bai & Zheng (2013) proposed the following trace test statistic for the case of $K = 2$

$$\mathrm{tr}\left[\left(\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}\right)\left(\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}\right)^{\top}\right]$$

which is the quadratic loss of $\mathbf{S}_n$ and $\mathrm{diag}(\mathbf{S}_{11}, \mathbf{S}_{22})$. The quadratic loss for covariance matrix estimation can be found in Olkin & Selliah (1977), Haff (1980) and Muirhead (1982). For block-structured correlation, regardless of the entropy loss or the quadratic loss for covariance matrix estimation, the inversion of a sample covariance matrix or log-determinant of $\mathbf{S}_{kk}$ is involved; as a consequence, the block dimension cannot be larger than the sample size.

In this paper, we propose a test statistic with two terms where one term is the distance between $\mathbf{S}_n$ and $\mathrm{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{KK})$ and the other term is

a screening term. Motivated by the Frobenius distance between matrices, this paper proposes the following statistic

$$T_{n1} = \mathrm{tr}[\mathbf{S}_n - \mathrm{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{KK})]^2.$$

Note that the statistic T_{n1} as used in Yamada, Hyodo and Nishiyama (2017)is the total sum of the squared entries of $\mathbf{S}_n - \mathrm{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{KK})$, which can capture the overall difference even when the individual entries of $\mathbf{S}_n - \mathrm{diag}(\mathbf{S}_{11}$,..., $\mathbf{S}_{KK})$ are small nonzero numbers. Therefore, the statistic T_{n1} is not only suitable for both low dimensions and high dimensions, but is also expected to have good performance for non-sparse alternatives. Furthermore, to enhance the power of T_{n1} when $\mathbf{\Sigma} - \mathrm{diag}(\mathbf{\Sigma}_{11}, \ldots, \mathbf{\Sigma}_{KK})$ is very sparse, a screening term T_{n0} is added to T_{n1}. A similar idea has been used in Fan, Liao and Yao (2015). Let the screening term be

$$T_{n0} = p^2 \delta_{\{\max_{(\ell_1,\ell_2)\in A_0} n(s_{\ell_1\ell_2})^2(\hat{\theta}_{\ell_1\ell_2})^{-1} > s^*(n,p)\}}$$

where $\delta_{\{\cdot\}}$ is an indicator function, $s^*(n,p)$ is a threshold depending on (n,p), $\mathbf{S}_n = (s_{\ell_1\ell_2})_{\ell_1,\ell_2=1}^p$, $\hat{\theta}_{\ell_1\ell_2} = n^{-1}\sum_{i=1}^n[(x_{\ell_1 i} - \bar{x}_{\ell_1})(x_{\ell_2 i} - \bar{x}_{\ell_2}) - s_{\ell_1\ell_2}]^2$, the set

$$A_0 = \{(\ell_1, \ell_2) : \ell_1 \in \{\tilde{p}_{i-1}+1, \ldots, \tilde{p}_i\}, \ell_2 \in \{\tilde{p}_{j-1}+1, \ldots, \tilde{p}_j\}, 1 \le i < j \le K\}, \tag{2.2}$$

with $\tilde{p}_i = p_1 + \ldots + p_i$, $\mathbf{x}_i = (x_{1i}, \ldots, x_{pi})^\top$, $\bar{x}_{\ell_1} = n^{-1}\sum_{i=1}^n x_{\ell_1 i}$ and $\bar{x}_{\ell_2} = n^{-1}\sum_{i=1}^n x_{\ell_2 i}$. The screening term T_{n0} shows that if some s_{ℓ_1,ℓ_2} is large enough, then T_{n0} is at least in the order of p^2. Thus, the screening term T_{n0} can capture the difference between $\mathbf{S}_n$ and $\mathrm{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{KK})$ even when $\mathbf{\Sigma} - \mathrm{diag}(\mathbf{\Sigma}_{11}, \ldots, \mathbf{\Sigma}_{KK})$ is very sparse. Our proposed test statistic is the sum of the two terms, that is,

$$\begin{aligned} T_n &= T_{n1} + T_{n0} \qquad (2.3) \\ &= \mathrm{tr}[\mathbf{S}_n - \mathrm{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{KK})]^2 + p^2 \delta_{\{\max_{(\ell_1,\ell_2)\in A_0} n(s_{\ell_1\ell_2})^2(\hat{\theta}_{\ell_1\ell_2})^{-1} > s^*(n,p)\}} \end{aligned}$$

which is expected to have good performance not only for non-sparse alternatives but also for sparse alternatives. Conditions needed on the threshold s^* will be given later.

2.1 Limiting null distribution of T_n

To facilitate the formulation, we use the following independent component structure model for the data.

Assumption [A]. Let $\{\mathbf{x}_i\}_{i=1}^n$ satisfy the independent component structure $\mathbf{x}_i = (x_{1i}, \ldots, x_{pi})^T = \boldsymbol{\mu} + \mathbf{\Sigma}^{1/2}\mathbf{w}_i$, where $\mathbf{w}_i = (w_{1i}, \ldots, w_{pi})^\top$, all elements $\{w_{ji} : j = 1, \ldots, p,\ i = 1, \ldots, n\}$ are i.i.d. with $\mathrm{E}(w_{ji}) = 0$, $\mathrm{E}(w_{ji}^2) = 1$, and finite 4th moments.

Remark 1. In fact, by (1.8) of Bai and Silverstein (2004), the existence of the finite 4th moment of w_{ji} implies that there exists a sequence $\{\eta_n\}$ satisfying $\eta_n \to 0$, $\eta_n n^{1/4} \to +\infty$ and $\eta_n^{-4}\mathrm{E}w_{ji}^4\delta_{(|w_{ji}|>\eta_n\sqrt{n})} \to 0$.

Assumption [B]. Assume that the number of blocks satisfies $K\eta_n^2 = o(1)$. Moreover, the spectral norm of $\mathbf{\Sigma}$ is bounded uniformly in p. The convergence regime $p/n \to y \in (0, \infty)$ for some constant y is satisfied.

In Assumption [A], moment conditions are imposed, which is distribution free. For example, the Gaussian distribution and many other distributions readily satisfy the independent component structure. In Assumption [B], $K\eta_n^2 = o(1)$ allows that K increases with n at some rate. Especially, for the Gaussian distribution, we have

$$\begin{aligned}\eta_n^{-4}\mathrm{E}w_{ji}^4\delta_{(|w_{ji}|>\eta_n\sqrt{n})} &\leq \eta_n^{-(4+m)}n^{-m/2}\mathrm{E}w_{ji}^{4+m}\delta_{(|w_{ji}|>\eta_n\sqrt{n})} \\ &= o(\eta_n^{-(4+m)}n^{-m/2}) = o(1),\end{aligned}$$

for any even m, if $\eta_n^{-2} = O(n^{m/(m+4)})$. Then K can have the order $o(n^{1-\epsilon})$ for any $\epsilon > 0$.

Lemma 1. *Under Assumption [A]-[B], and under H_0 specified by (2.1), we have*

$$\frac{T_{n1} - \mu}{\sigma} \to N(0, 1) \quad \textit{and} \quad \frac{T_{n1} - \hat{\mu}}{\sigma_0} \to N(0, 1),$$

where

$$
\begin{aligned}
\mu &= \frac{(n^2-n-1)[(\mathrm{tr}\boldsymbol{\Sigma})^2-\sum_{k=1}^K(\mathrm{tr}\boldsymbol{\Sigma}_{kk})^2]}{n(n-1)^2},\\
\hat{\mu} &= \frac{(n^2-n-1)[(\mathrm{tr}\mathbf{S}_n)^2-\sum_{k=1}^K(\mathrm{tr}\mathbf{S}_{kk})^2]}{n(n-1)^2},\\
\sigma_0^2 &= 4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2)^2-4\sum_{k=1}^K(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{kk}^2)^2,\\
\sigma^2 &= \sigma_0^2+4n^{-3}\sum_{k=1}^K(\mathrm{tr}\boldsymbol{\Sigma}_{kk}-\mathrm{tr}\boldsymbol{\Sigma})^2\left[2\,\mathrm{tr}\boldsymbol{\Sigma}_{kk}^2+\beta_w\sum_{\ell=1}^{p_k}(\mathbf{e}_{\ell k}^\top\boldsymbol{\Sigma}_{kk}\mathbf{e}_{\ell k})^2\right],\\
\beta_w &= \mathrm{E}(w_{ji}^4)-3.
\end{aligned}
\tag{2.4}
$$

Here $\mathbf{e}_\ell$ *is a* p*-dimensional vector with the* ℓ*-th element being one and other elements being zeros and* $\mathbf{e}_{\ell k}$ *is a* p_k*-dimensional vector with the* ℓ*-th element being one and other elements being zeros.*

Note that we have suppressed the subscript n in many of the quantities we use such as μ and σ^2. The proof of Lemma 1 is in the supplementary file 1. The asymptotic variance σ_0^2 depends on the unknown parameters $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ and $\mathrm{tr}(\boldsymbol{\Sigma}_{kk}^2), k=1,\ldots,K$. However,

$$(n-2)^{-1}[\mathrm{tr}(\mathbf{S}_{kk}^2)-(n+2)^{-1}(\mathrm{tr}\mathbf{S}_{kk})^2]-n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{kk}^2)=o_p(1),\quad k=1,\ldots,K$$

which can be used to estimate σ_0^2; see the proof in the supplementary file

1. Moreover, under H_0, we have $\mathrm{tr}(\mathbf{\Sigma}^2) = \sum_{k=1}^K \mathrm{tr}(\mathbf{\Sigma}_{kk}^2)$, and then

$$(n-2)^{-1}\sum_{k=1}^K[\mathrm{tr}(\mathbf{S}_{kk}^2) - (n+2)^{-1}(\mathrm{tr}\mathbf{S}_{kk})^2] - n^{-1}\mathrm{tr}(\mathbf{\Sigma}^2) = o_p(1).$$

Therefore, σ_0^2 can be consistently estimated by

$$\begin{aligned}\hat{\sigma}_0^2 &= 4(n-2)^{-2}\{\sum_{k=1}^K[\mathrm{tr}(\mathbf{S}_{kk}^2) - (n+2)^{-1}(\mathrm{tr}\mathbf{S}_{kk})^2]\}^2 \\ &\quad -4(n-2)^{-2}\sum_{k=1}^K[\mathrm{tr}(\mathbf{S}_{kk}^2) - (n+2)^{-1}(\mathrm{tr}\mathbf{S}_{kk})^2]^2.\end{aligned}$$

Bai and Saranadasa (1996) suggested a uniformly minimum variance unbiased estimator of $\mathrm{tr}(\mathbf{\Sigma}^2)$ under the normality assumption, but we have used an asymptotic approximation with a finite sample correction factor to better control Type I errors. Let

$$p_0^2 = p^2 - p_1^2 - \ldots - p_K^2. \tag{2.5}$$

The following result provides the asymptotic justification to the proposed test.

Theorem 1. *Under Assumptions [A]-[B], and under H_0 specified by (2.1), if* $\liminf_{n\to\infty}\inf_{(i,j)\in A_0}\mathrm{var}[(x_{1i}-\mathrm{E}x_{1i})(x_{1j}-\mathrm{E}x_{1j})][\mathrm{var}(x_{1i})\mathrm{var}(x_{1j})]^{-1/2} > 0$, $s^*(n,p) - 4\log p_0 \to +\infty$, *and* $\sup_{1\le\ell\le p}\mathrm{E}\exp(t_0|x_{\ell 1}|^{m_0}) < \infty$ *for some constants* $t_0 > 0$ *and* $0 < m_0 \le 2$, *we have*

$$\hat{\sigma}_0^{-1}(T_n - \hat{\mu}) \to N(0,1).$$

We note that T_n has the same null distribution as T_{n1} in the asymptotic sense, and the second term T_{n0} plays a role mainly when the alternative hypothesis is true. The one-sided rejection region for H_0 at the nominal level α is

$$\{\mathbf{x}_1, \ldots, \mathbf{x}_n : T_n - \hat{\mu} > \hat{\sigma}_0 q_{1-\alpha}\}, \tag{2.6}$$

where q_α is the α-th quantile of the standard normal distribution.

Remark 2. To apply the proposed test in practice, we need to choose the threshold $s^*(n, p)$. There are many choices for the threshold as long as it satisfies $s^*(n, p) - 4 \log p_0 \to +\infty$. For simplicity, in this paper, the threshold is taken to be

$$s^*(n, p) = [4 + (\log \log n - 1)^2](\log p_0 - 0.25 \log \log p_0) + q \tag{2.7}$$

where q satisfies $\exp[-(8\pi)^{-1/2} \exp(-q/2)] = 0.99$. The threshold ensures that even if n and p_0 are small, the probability of the event $T_{n0} = 0$ is bounded by 0.01 under H_0 because $\max_{(\ell_1,\ell_2)\in A_0} n(s_{\ell_1\ell_2})^2\, \hat{\theta}^{-1}_{\ell_1\ell_2} - 4 \log p_0 + \log \log p_0$ converges to a type I extreme value distribution $\exp[-(8\pi)^{-1/2} \exp(-t/2)]$ under the null hypothesis (see Xiao and Wu, 2013). The probability of the event $T_{n0} = 0$ becomes negligible under H_0 when either n or p_0 is moderately large. For example, if $n = 200$ and $p_0 = 250$, the concerned probability is only 0.002.

Remark 3. Our proposed hypothesis test (2.6) is a global test on correlations among different blocks. If the null hypothesis gets rejected, under the sparsity assumption, for identifying the individual nonzero correlations, we may directly use Cai and Liu (2016)'s multiple testing method in two steps. Let

$$T_{ij} = \frac{\sum_{\ell=1}^{n}(x_{i\ell} - \bar{x}_i)(x_{j\ell} - \bar{x}_j)}{\sqrt{n\hat{\theta}_{ij}}} \tag{2.8}$$

where $\hat{\theta}_{ij} = n^{-1}\sum_{\ell=1}^{n}[(x_{i\ell} - \bar{x}_i)(x_{j\ell} - \bar{x}_j) - s_{ij}]^2$.

Step 1: bootstrap procedure. Let $\{x_{j1}^*, ..., x_{jn}^*\}$ be a sample drawn randomly with replacement from $\{x_{j1}, ..., x_{jn}\}$ for every $j \in \{1, ..., p\}$. Let $\mathbf{x}_\ell^* = (x_{1\ell}^*, ..., x_{p\ell}^*)^T$ for $\ell = 1, ..., n$ and the bootstrap test statistic T_{ij}^* is computed from $\mathbf{x}_1^*, ..., \mathbf{x}_n^*$ as in (2.8). When the above bootstrap procedure is repeated N times, then we have N bootstrap test statistics $T_{ij1}^*, ..., T_{ijN}^*$. Let

$$G_{n,N}^*(t) = \frac{2}{Np_0^2}\sum_{\ell=1}^{N}\sum_{(i,j)\in A_0} I\{|T_{ij\ell}^*| \geq t\},$$

where A_0 is in (2.2).

Step 2: Large-scale correlation tests with bootstrap given in Cai and Liu (2016). Let

$$\begin{aligned}\hat{t} \;=\; & \inf\{0 \leq t \leq \sqrt{4\log p_0 - 2\log(\log p_0)} : \\ & \frac{G_{n,N}^*(t)(p_0^2)/2}{\max\{\sum_{(i,j)\in A_0} I\{|T_{ij}| \geq t\}, 1\}} \leq \alpha\}.\end{aligned}$$

If $\hat{t}$ does not exist, then let $\hat{t} = \sqrt{4 \log p_0}$. We reject $H_{0ij} : \sigma_{ij} = 0$ whenever $|T_{ij}| \geq \hat{t}$ for $(i, j) \in A_0$.

Remark 4. On the surface, it seems that we need the eighth moment of $\mathbf{x}_i$ to calculate the variance of T_{n1}. In fact, Yamada, Hyodo and Nishiyama (2017) requires the finite eighth moment condition. However, as we show in this paper, the results of Lemma 1 and Theorem 2 require only the fourth moment of $\mathbf{x}_i$.

2.2 Limiting distribution of T_n under the alternative hypothesis

Next, we study the theoretical property of proposed statistic T_n under the alternative hypothesis. Let the difference between the null hypothesis and alternative hypothesis be $\mathbf{A} = \mathbf{\Sigma}^2 - \mathrm{diag}(\mathbf{\Sigma}_{11}^2, \ldots, \mathbf{\Sigma}_{KK}^2)$.

Theorem 2. *Under Assumptions [A]-[B], we have*

$$\sigma_1^{-1}(T_{n1} - \hat{\mu} - \mu_1) \to N(0, 1)$$

where $\mu_1 = (n^2 - n + 2)\mathrm{tr}\mathbf{A}/(n-1)^2$,

$$\sigma_1^2 \;=\; \sigma_0^2 + 4[2n^{-1}\mathrm{tr}\mathbf{A}^2 + \beta_w n^{-1} \sum_{\ell=1}^{p} (\mathbf{e}_\ell^\top \mathbf{A} \mathbf{e}_\ell)^2],$$

here $\mathbf{e}_\ell$ *is the* p*-dimensional vector with the* ℓ*th element being one and other elements being zeros and* $\beta_w = \mathrm{E}w_{ij}^4 - 3$.

The asymptotic power function of T_n is $\beta_{T_n}(\mathbf{A}) = P(T_n - \hat{\mu} > \hat{\sigma}_0 q_{1-\alpha})$. We have $P(T_n - \hat{\mu} > \hat{\sigma}_0 q_{1-\alpha}) - [1 - \Phi(\sigma_1^{-1}(\sigma_0 q_{1-\alpha} - \mu_1))] = o(1)$. Because $\mathrm{tr}\mathbf{A} = \mathrm{tr}\mathbf{\Sigma}^2 - \sum_{k=1}^{K} \mathrm{tr}\mathbf{\Sigma}_{kk}^2 = \sum_{1\le k_1\neq k_2\le K} \mathrm{tr}\mathbf{\Sigma}_{k_1k_2}\mathbf{\Sigma}_{k_2k_1} \ge 0$, it is easy to see that $\sigma_1^2 \ge \sigma_0^2$ and $\mu_1 \ge 0$. If the population covariance matrix departs from the null hypothesis (in the sense that $\mathrm{tr}\mathbf{A} > \epsilon_0 > 0$ for any positive constant ϵ_0), then $\sigma_1^2 > \sigma_0^2$ and $\mu_1 > 0$. Under such an alternative hypothesis, we have $(\sigma_0 q_{1-\alpha} - \mu_1)/\sigma_1 < q_{1-\alpha}$, that is,

$$\beta_{T_n}(\mathbf{A}) > \alpha.$$

Thus, the proposed test T_n is asymptotically unbiased. In fact, when n is sufficiently large, $\beta_{T_n}(\mathbf{A})$ is an increasing function of $\mathrm{tr}\mathbf{A}$ where $\mathrm{tr}\mathbf{A}$ measures the departure from the null hypothesis.

Theorem 3. *Under Assumptions [A]-[B] and* $\mathbf{\Sigma}^2 = \mathrm{diag}(\mathbf{\Sigma}_{11}^2, \ldots, \mathbf{\Sigma}_{KK}^2) + \mathbf{A}$,

(1). We have $\beta_{T_n}(\mathbf{A}) \ge \alpha$ *when* n *is large enough; Especially, when* $\mathrm{tr}\mathbf{A} > \epsilon_0 > 0$ *for any positive constant* ϵ_0*, we have* $\beta_{T_n}(\mathbf{A}) > \alpha$ *for sufficiently large* n*;*

(2). If $\mathrm{tr}\mathbf{A}$ *tends to infinity or* $P(\max_{(\ell_1,\ell_2)\in A_0} n(s_{\ell_1\ell_2})^2(\hat{\theta}_{\ell_1\ell_2})^{-1} > s^*(n,p))$ *converges to one, then we have* $\beta_{T_n}(\mathbf{A}) \to 1$ *as* $n \to \infty$.

Theorem 3 shows that the proposed test T_n is asymptotically unbi-

ased. If the absolute value of at least one entry of $\mathbf{A}$ is greater than $\sqrt{(\log p_0 \log n)/n}$, then there exists $(\ell_1, \ell_2) \in A_0$ such that $n(s_{\ell_1\ell_2})^2(\hat{\theta}_{\ell_1\ell_2})^{-1}/s^*(n,p) \approx c \log n/\log\log n$ converges to infinity in probability under the conditions of Theorem 1, and thus $P(\max_{(\ell_1,\ell_2)\in A_0} n(s_{\ell_1\ell_2})^2(\hat{\theta}_{\ell_1\ell_2})^{-1} > s^*(n,p)) \to 1$ holds by Remark 2 and then the power converges to one.

Remark 5. Support recovery of $\mathbf{\Sigma}$: Following the proof of Theorem 5 in Cai, Liu and Xia (2013), under the conditions

$$p/n \to y \in (0, +\infty), \quad \min_{(i,j)\in A_0} \theta_{ij}(\sigma_{ii}\sigma_{jj})^{-1/2} > \tau,$$
$$\mathrm{E}|(x_{j1} - \mathrm{E}x_{j1})(\sigma_{jj})^{-1/2}|^{8+\epsilon} \le c_0, \ \forall\, 1 \le j \le p,$$

for some $c_0 > 0$, $\epsilon > 0$, $\tau > 0$ with the set A_0 defined in (2.2), we have

$$\liminf_{\mathbf{\Sigma}\in W_0} P(\hat{\Psi} = \Psi) \to 1,$$

where

$$\Psi = \{(i,j) : \sigma_{ij} \neq 0, (i,j) \in A_0\},$$

$$\hat{\Psi} = \{(i,j) : n(s_{ij} - \sigma_{ij})^2(\hat{\theta}_{ij})^{-1} \ge 4\log p_0, (i,j) \in A_0\},$$

$$W_0 = \{\mathbf{\Sigma} : \min_{(i,j)\in\Psi} n^{1/2}|\sigma_{ij}|(\theta_{ij})^{-1/2} \ge 4\sqrt{\log p_0}, (i,j) \in A_0\},$$

with $\mathbf{\Sigma} = (\sigma_{ij})_{i,j=1}^p$ and $p_0^2 = p^2 - p_1^2 - ... - p_K^2$ given in (2.5).

3. Simulation studies

In this section, we evaluate the finite sample performance of the proposed test in terms of its Type I error rates and powers. Because the proposed test uses the Frobenius distance between covariance matrices, we will denote it as FDS. The test proposed by Paindaveine and Verdebout (2016) was developed for variables with mean zero. When applied to the centered variables (by removing the sample mean) in high dimensions, the test has seriously inflated Type I errors, and therefore we exclude it from the comparisons. The test used by Jiang, Bai and Zheng (2013) is the same as the test of Bao, Hu, Pan and Zhou (arXiv) when $K = 2$ but has slightly poorer performance when $K = 3$, so we will include the latter test only. To be specific, the following three competing tests are used in our comparisons:

- "CLRT": the test of Jiang and Yang (2013);
- "BHPZ": the test of Bao, Hu, Pan and Zhou (arXiv);
- "YHN": the test of Yamada, Hyodo and Nishiyama (2017);

We generate samples of size n from $\mathbf{x}_i = \mathbf{1}_p + \mathbf{\Sigma}^{1/2}\mathbf{w}_i$ for $i = 1, \ldots, n$ where $\mathbf{1}_p$ is a p-dimensional vector with all elements equal to one, $\mathbf{w}_i = (w_{1i}, \ldots, w_{pi})^\top$ and $\{w_{ji}, i = 1, \ldots, n, j = 1, \ldots, p\}$ are independent and identically distributed as $N(0, 1)$. To consider different structures of $\mathbf{\Sigma}$,

we use $\mathbf{\Sigma} = 0.2\mathbf{I}_p + \sum_{i=1}^{3} \theta_i \mathbf{\Sigma}_i$ for some values $(\theta_1, \theta_2, \theta_3)$ where $\mathbf{\Sigma}_1 = (0.5^{|i-j|})_{i,j=1}^{p}$ is approximately sparse in structure, $\mathbf{\Sigma}_2 = \mathbf{I}_p + 0.5(\delta_{\{|i-j|=1\}})_{i,j=1}^{p}$ is sparse, and $\mathbf{\Sigma}_3 = 0.98\mathbf{I}_p + 0.02\mathbf{1}_p\mathbf{1}_p^T$ is a dense structure. For each setting, we conduct 5000 Monte Carlo simulations. For the type I error estimates, the standard errors are approximately 0.006.

At the sample size at $n = 200$, we consider the dimension $p = 60, 120, 180$, and the number of blocks $K = 2, 3$ with the block sizes $p_1 = \ldots = p_K = p/K$. The ROC curves for the competing tests are plotted in Figure 1 under the null hypothesis $\mathbf{\Sigma} = 0.2\mathbf{I}_p$ and the alternative hypotheses $\mathbf{\Sigma} = 0.2\mathbf{I}_p + \mathbf{\Sigma}_i, i = 1, 2, 3$ at $n = 200$ and $p_1 = p_2 = p_3 = 20$. Clearly, the test FDS has the best performance for the non-dense $\mathbf{\Sigma}$. When $\mathbf{\Sigma}$ is dense, FDS and YHN are similar, but YHN is the worst performer for the sparse alterative. Moreover, the empirical test sizes and empirical powers are listed in Table 1 for a variety of settings. All the methods maintain Type I errors well. For comparison of powers, the proposed FDS test outperforms. Especially, when $(p_1, p_2, p_3) = (20, 20, 20)$ and $\mathbf{\Sigma} = 0.2\mathbf{I}_p + \mathbf{\Sigma}_1$, the empirical power of FDS test is about 98% and the empirical powers of other tests are between 36% and 53%. For $(p_1, p_2, p_3) = (60, 60, 60)$ and $\mathbf{\Sigma} = 0.2\mathbf{I}_p + \mathbf{\Sigma}_2$, the empirical power of FDS test is about 88%, but the empirical powers of the other tests range at most from 10% to 14%. Overall, the proposed test

FDS is seen to be much more powerful than its competitors. When $\mathbf{\Sigma}$ is dense, FDS and YHN are indeed similar, and they are both leaders in the comparison.

When the dimension is much greater than the sample size, we only examine the performance of FDS, BHPZ and YHN, because CLRT cannot handle such cases. In the simulation, the null hypothesis is $\mathbf{\Sigma} = 0.2\mathbf{I}_p$ and the alternative hypothesis is $\mathbf{\Sigma} = 0.2\mathbf{I}_p + \theta_1\mathbf{\Sigma}_1 + \theta_2\mathbf{\Sigma}_2^* + \theta_3\mathbf{\Sigma}_3$ where $\mathbf{\Sigma}_2^* = \mathbf{I}_p + \rho_0(\delta_{\{|i-j|=1\}})_{i,j=1}^p$ with $\rho_0 = 0.3 + 0.3\exp(0.009p)/(0.15 + \exp(0.009p))$ and $\theta_i = 0$ or 1 for $i = 1, 2, 3$. The distribution of w_{ji} is taken to be $N(0, 1)$ or Gamma(4, 2)-2. In this study we consider the sample size $n = 150, 300$, the dimension $p = 180, 360, 900$ and the number of blocks $K = 2, 3$ with the block sizes $p_1 = \ldots = p_K = p/K$. The empirical test sizes and powers are listed in Tables 2 and 3. The Type I errors are all close to the nominal level of 0.05. Moreover, as the dimension increases, the empirical powers of the tests increase with n. For example, when $\mathbf{\Sigma} = 0.2\mathbf{I}_p + \mathbf{\Sigma}_2^*$, $p = 180$ and $K = 2$, the power of FDS increases from 71.24% to 99.96% quickly as the sample size increases from $n = 150$ to 300, but the powers of other tests rise much less. To save the space, Table 3 is given in the supplementary file.

We note that the proposed FDS test does not always dominate the others when p is small. We refer to the ROC curve in Figure 1 under the null hypothesis $\mathbf{\Sigma} = 0.2\mathbf{I}_p$ and the alternative hypotheses for $\mathbf{\Sigma} = \mathbf{\Sigma}_4 = 1.2\mathbf{I}_p + 0.18(\delta_{\{|i-j|=1\}})_{i,j=1}^p + 0.1(\delta_{\{|i-j|=3\}})_{i,j=1}^p$ at the sample size $n = 200$, the dimension $p = 6$ with $K = 3$ blocks of equal sizes $p_1 = p_2 = p_3 = 2$. In this case, the population is Gaussian and the likelihood is correctly specified, so it is not surprisingly that CLRT shows slightly better performance than FDS.

To check the sensitivity of the threshold $s^*(n, p)$ and any scaled version of T_{n0}, we consider the rejection region

$$\{\mathbf{x}_1, \ldots, \mathbf{x}_n : T_n(c_1, c_2) - \hat{\mu} > \hat{\sigma}_0 q_{1-\alpha}\}, \tag{3.1}$$

which is similar to (2.6) where $\hat{\mu}$ and $\hat{\sigma}_0$ are in (2.4) and

$$T_n(c_1, c_2) = T_{n1} + c_1 \cdot T_{n0}(c_2),$$

with $T_{n1} = \mathrm{tr}[\mathbf{S}_n - \mathrm{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{KK})]^2$ and

$$\begin{aligned} T_{n0}(c_2) &= p^2 \delta_{\{\max_{(\ell_1,\ell_2)\in A_0} n(s_{\ell_1\ell_2})^2(\hat{\theta}_{\ell_1\ell_2})^{-1} > s^*(n,p,c_2)\}}, \\ s^*(n, p, c_2) &= c_2 \cdot [4 + (\log\log n - 1)^2](\log p_0 - 0.25 \log\log p_0) + q. \end{aligned}$$

We have $s^*(n, p) = s^*(n, p, 1)$, $T_{n0} = T_{n0}(1)$ and $T_n = T_n(1, 1)$. We consider the sample size n=200, the dimension is $p = 60, 120, 180$, and the number

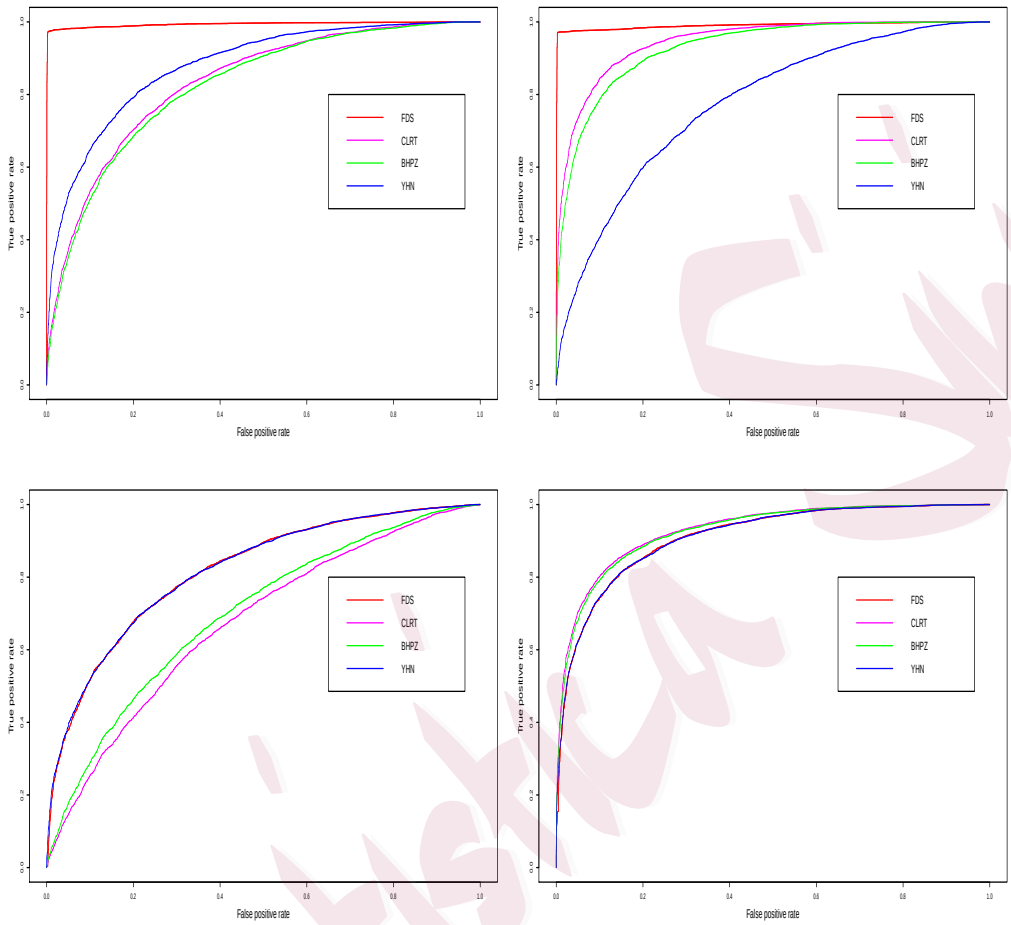

Figure 1: The first three ROC curves are the results from three simulation settings given in Section 3 with three different specifications $\mathbf{\Sigma}_1$ (upper left panel), $\mathbf{\Sigma}_2$ (upper right), $\mathbf{\Sigma}_3$ (lower left) with w_{ij} being i.i.d from $N(0,1)$, $(n,p) = (200,60)$, and $p_1 = p_2 = p_3 = 20$. The ROC curve in the lower right panel refers to the case of $(n,p) = (200,6)$ of $K = 3$ equal block sizes. The curves for FDS and YHN are nearly identical in the lower left panel and lower right panel.

of blocks is $K = 2, 3$ with the block sizes $p_1 = \ldots = p_K = p/K$. The parameters c_1 and c_2 are taken as $c_1 = 0.001, 0.5, 2$ and $c_2 = 0.5, 1, 2$. The empirical test sizes and powers for different values of c_1 and c_2 are listed in Tables 4-5. Simulation results in Table 4 show that when c_1 is small or large, the empirical test sizes and empirical powers are similar for the different values of c_1. Simulation results in Table 5 show that when c_2 is small, the empirical test sizes cannot be controlled; when c_2 is large, although the empirical test sizes can be controlled, the empirical powers will decrease. Then the penalty T_{n0} is somewhat sensitive for the threshold $s^*(n, p)$, but is not sensitive for the scaled version of T_{n0}. Moreover, to show that our test is valid for $p/n \to y = 0$, Table 6 presents some simulation results with $n = 500, 750, 1000$ and $p = 6, 12, 18$. To save the space, Tables 4-5-6 are given in the supplementary file.

4. Demonstration with a real data example

To further demonstrate the power of the proposed test, we use data from a major supermarket in northern China (Wang, 2009). In the dataset, each record contains the daily sales volume of individual products over a 463-day period. We are interested in understanding the correlation between vegetable sale volumes and dairy sale volumes. We have 26 major vegetables

Table 1: Empirical test sizes and powers (in percentage) for comparison of four methods with $n = 200$, $(p_1, \ldots, p_K) = (p/K, \ldots, p/K)$ and $K = 2, 3$ for Gaussian variables. The vector $(\theta_1, \theta_2, \theta_3)$ specifies the Σ matrix. The rejection region is given in (2.6).

$(\theta_1, \theta_2, \theta_3)$	Methods	$p = 60$	120	180	60	120	180
		$K = 2$			$K = 3$		
		Empirical test sizes					
(0, 0, 0)	FDS	4.50	4.95	4.94	5.10	4.85	4.88
	CLRT	4.74	5.52	4.86	5.02	5.30	5.12
	BHPZ	4.58	5.12	4.52	4.88	5.09	4.68
	YHN	4.64	5.07	5.07	5.18	4.94	4.88
		Empirical powers					
(1, 0, 0)	FDS	87.86	76.52	69.28	98.06	93.20	88.42
	CLRT	19.52	9.40	6.98	38.74	14.28	8.38
	BHPZ	17.46	8.80	6.64	36.08	14.72	9.55
	YHN	27.28	13.22	9.72	52.48	22.78	14.83
(0, 1, 0)	FDS	86.70	75.52	68.62	97.50	92.68	88.02
	CLRT	38.28	13.26	7.86	75.42	24.86	10.92
	BHPZ	30.86	11.82	7.82	66.78	23.62	13.26
	YHN	15.68	92.50	7.60	26.12	14.18	10.02
(0, 0, 1)	FDS	32.46	69.86	90.90	38.48	78.90	95.32
	CLRT	12.82	12.38	8.78	15.62	15.90	11.70
	BHPZ	11.92	11.32	9.00	18.10	20.20	17.62
	YHN	32.62	70.20	91.02	38.96	79.16	95.42

Table 2: Empirical test sizes and powers (in percentage) of comparison of three methods with $(p_1, \ldots, p_K) = (p/K, \ldots, p/K)$ and $K = 2, 3$ for Gaussian variables. The vector $(\theta_1, \theta_2, \theta_3)$ specifies the Σ matrix. The rejection region is given in (2.6). When a test is not applicable, the corresponding entries are marked $-$.

$(\theta_1, \theta_2, \theta_3)$	n	Methods	p=180	360	900	180	360	900
			$K = 2$			$K = 3$		
			Empirical test sizes					
$(0, 0, 0)$	150	FDS	5.11	4.72	4.22	4.86	4.78	4.48
		BHPZ	4.62	—	—	5.08	4.76	—
		YHN	5.50	4.94	5.06	5.26	4.86	5.24
	300	FDS	5.08	4.92	4.93	5.08	5.08	5.02
		BHPZ	5.08	4.70	—	5.26	5.30	—
		YHN	5.04	5.08	5.33	5.42	5.32	5.12
			Empirical powers					
$(1, 0, 0)$	150	FDS	38.22	25.78	14.06	57.02	38.85	21.80
		BHPZ	6.14	—	—	7.84	5.26	—
		YHN	8.74	6.22	5.44	12.41	7.66	5.66
	300	FDS	97.74	94.16	87.52	99.95	99.51	97.74
		BHPZ	8.74	5.92	—	13.76	7.48	—
		YHN	12.42	8.14	6.60	22.86	11.36	7.72
$(0, 1, 0)$	150	FDS	71.24	59.54	41.78	89.52	80.20	61.92
		BHPZ	9.32	—	—	20.72	7.10	—
		YHN	7.68	5.86	5.32	10.22	7.18	5.24
	300	FDS	99.96	99.88	99.74	100	100	100
		BHPZ	32.22	10.50	—	74.24	27.82	—
		YHN	10.42	7.2	6.70	16.02	9.85	7.00
$(0, 0, 1)$	150	FDS	76.18	98.48	100	84.28	99.38	100
		BHPZ	7.24	—	—	11.20	6.48	—
		YHN	76.87	98.52	100	84.56	99.46	100
	300	FDS	99.36	100	100	99.82	100	100
		BHPZ	14.84	9.16	—	34.16	21.02	—
		YHN	99.34	100	100	99.82	100	100

and 58 dairy products in the study, that is, $(p_1, p_2) = (26, 58)$.

To evaluate the power of various tests at small sample sizes, we randomly draw the sale volumes of vegetables and dairies with $p_1 + p_2 + 2$ days, that is, the sample size is $n = p_1 + p_2 + 2$. Based on 10,000 random draws at this sample size, the proposed test, FDS, together with the test YHN, rejects the null hypothesis that sale volumes of the vegetables and dairies are uncorrelated 100% of the time. The tests CLRT and BHPZ reject the null hypothesis 58.71% and 84.22% of the time, respectively. For the sensitivity analysis with $(c_1, c_2) = (0.001, 1), (5, 1), (1, 0.5), (1, 2)$, the proposed test FDS still rejects the null hypothesis 100% of the time.

When we take a small number of days randomly from the data set, autocorrelation is negligible. To use the whole sample to understand or confirm the correlation between the prices of these two products, we use the autoregressive AR(1) model to fit the data first, and then examine the residuals. In this case, all the tests we considered reject the null hypothesis of no correlation at the level 0.001. The fact that the proposed test is able to detect the correlation with high power even when the sample size is slightly above the total dimension indicates that the test is valuable in the analysis of moderately high dimensional problems.

5. Discussion

We propose a test for detecting block-structured correlation in high dimensional variables. The validity of the test is established under a framework where the dimension of the variables grows linearly with the sample size. For the rationale of why the framework of p/n tending to a constant is useful for high dimensional data analysis, we refer to Marčenko and Pastur (1967) and Bai and Silverstein (2010). The test can be used in a wide range of problems for Gaussian or non-Gaussian variables, and attains good power for sparse or non-sparse alternatives. Simulation studies show that the proposed test performs very well in both Type I error rates and powers relative to the existing tests when the latter are applicable. Unlike the other tests, the proposed method does not use the inversion of any covariance matrix and requires only the finite fourth moments of the random variables. More importantly, the proposed test performs quite well even when the dimension exceeds the sample size. When p is small and n is large, and the data are Gaussian, the proposed test will lose some power against the likelihood ratio test, but the loss of power is limited even in those situations in our empirical studies.

Supplementary materials

The first supplementary material consists of the proofs of Lemma 1 and Theorem 1-3. The second supplementary material consists of three lemmas and the detailed proofs of (S.6)-(S.8). These proofs are conducted under Assumptions [A]-[B]. The sample covariance matrix $\mathbf{S}_n$ of 84 major vegetables and 58 dairy products in Section 4 is

$https://math127.nenu.edu.cn/shuxue/HData/webpage/covariancematrix.zip.$

Acknowledgements We thank the associate editor, the referees, and the editor for their many constructive and insightful comments that have led to the significant improvements in the article. Jianhua Guo's researches were supported by NSFC grants 11690012 and 11631003. Xuming He's researches were supported by National Science of Foundation (USA) Awards DMS-1607840 and DMS-1914496.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd Edition). New York: John Wiley & Sons.

Bai, Z. D. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica 6*, pp. 311–329.

Bai, Z. D., Jiang, D. D., Yao, J. F. and Zheng, S. R. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *The Annals of Statistics 37*, pp. 3822–3840.

Bai, Z. D. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices (2nd Edition)*. New York: Springer.

Bao, Z. G., Hu, J., Pan, G. M. and Zhou, W. (2014). Test of independence for high-dimensional random vectors based on block correlation matrices. *ArXiv: 1410.5082*.

Cai, T. T. and Jiang, T. F. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *The Annals of Statistics*, *39*(3), pp. 1496–1525.

Cai, T. T., Liu, W. D. 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), pp. 265–277.

Fan, J. Q., Liao, Y. and Yao, J. W. (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, *83*(4), pp. 1497–1541.

Haff, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *The Annals of Statistics*, *8*(3), pp. 586–597.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent component analysis*. John Wiley & Sons, INC.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp.*

Math. Statist. Prob. 1, pp. 361–379.

Jiang, T. F. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *The Annals of Applied Probability*, *14*(2), pp. 865–880.

Jiang, D. D., Bai, Z. D. and Zheng, S. R. (2013). Testing the independence of sets of large-dimensional variables. *Science China: Mathematics*, *56*(1), pp. 135–147.

Jiang, T. F. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *The Annals of Statistics*, *41*(4), pp. 2029–2074.

Li, D. N. and Xue, L. Z. (2015). Joint limiting laws for high-dimensional independence tests. *ArXiv: 1512.08819v1.*

Marčenko, V. A. and Pastur, L. A. (1967). Distribution for some sets of random matrices. *Math. USSR-Sb. 1*, pp. 457-483.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* New York: John Wiley & Sons.

Olkin, I. and Selliah, J. B. (1977). Estimating covariances in a multivariate normal distribution. In Statistical Decision Theory and Related Topics (S. S. Gupta and D. S. Moore, eds.) **II**, 313–326. New York: Academic Press.

Paindaveine, D. and Verdebout, T. (2016). On high-dimensional sign tests. *Bernoulli*, *22*(3), pp. 1745–1769.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, *92*(4),

pp. 951–956.

Sugiura, N. and Fujikoshi, Y. (1969). Asymptotic expansions of the non-null distributions of the likelihood ratio criteria for multivariate linear hypothesis and independence. *The Annals of Mathematical Statistics*, *40*(3), pp. 942–952.

Wang, H. S. (2009). Forward regression for ultra-high dimensional variable screening. *Journal of the American Statistical Association*, *104*(488), pp. 1512–1524.

Wilks, S. S. (1935). On the indepdendence of k sets of normally distributed statistical variables. *Econometrica*, *3*(3), pp. 309–326.

Xiao, H. and Wu, W. B. (2013). Asymptotic theory for maximum deviations of sample covariance matrix estimates. *Stochastic Processes and their Applications*, *123*, pp. 2899–2920.

Yamada, Y. Hyodo, M. and Nishiyama, T. (2017). Testing block-diagonal covariance structure for hig-dimensional data under non-normality. *Journal of Multivariate Analysis*, *155*, pp. 305–316.

Zheng, S. R., Bai, Z. D. and Yao, J. F. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.*, *43*(2), pp. 546–591.

Zhou, W. (2007). Asymptotic distributions of the largest off-diagonal entry of correlation matrices. *Transactions of the American Mathematical Society*, *359*(11), pp. 5345–5363.

Shurong Zheng, Northeast Normal University and KLAS, China

E-mail: zhengsr@nenu.edu.cn

Xuming He, University of Michigan, USA

E-mail: xmhe@umich.edu

Jianhua Guo, Northeast Normal University and KLAS, China

E-mail: jhguo@nenu.edu.cn

Supplementary material 1

The first supplementary material consists of the proofs of Lemma 1 and Theorem 1-3, and Tables 3-4-5-6. The simulation settings of Tables 3-4-5-6 are in the main paper.

S.1. Tables 3-4-5-6

S.2. Proofs of Lemma 1 and Theorem 1-3

Define $\mathbf{r}_i = n^{-1/2}\mathbf{w}_i$, $\mathbf{w}_i = (w_{1i}, \ldots, w_{pi})^\top$, $\mathbf{r}_{ik} = n^{-1/2}\mathbf{w}_{ik}$, $\mathbf{w}_{ik} = (w_{\tilde{p}_{k-1}+1,i}, \ldots, w_{\tilde{p}_k,i})^\top$ with $\tilde{p}_0 = 0$ and $\tilde{p}_k = p_1 + \ldots + p_k$ for $k = 1, \ldots, K$, $i = 1, \ldots, n$. Then $\mathbf{r}_i = (\mathbf{r}_{i1}^\top, \ldots, \mathbf{r}_{iK}^\top)^\top$ and $\mathbf{w}_i = (\mathbf{w}_{i1}^\top, \ldots, \mathbf{w}_{iK}^\top)^\top$ for $i = 1, \ldots, n$. We have

$$(n-1)^2 n^{-2}\mathrm{tr}(\mathbf{S}_n^2) = \mathrm{tr}[(\sum_{i=1}^n \mathbf{\Sigma}^{1/2}\mathbf{r}_i\mathbf{r}_i^\top\mathbf{\Sigma}^{1/2})^2] + n^2(\bar{\mathbf{r}}^\top\mathbf{\Sigma}\bar{\mathbf{r}})^2 - 2n\bar{\mathbf{r}}^\top\mathbf{\Sigma}\sum_{i=1}^n \mathbf{r}_i\mathbf{r}_i^\top\mathbf{\Sigma}\bar{\mathbf{r}},$$

where $\bar{\mathbf{r}} = n^{-1}\sum_{i=1}^n \mathbf{r}_i$. By Lemma S.2.1 and S.2.2 from the supplementary file 2, letting ϵ be a very small positive number, we have $n^2(\bar{\mathbf{r}}^\top\mathbf{\Sigma}\bar{\mathbf{r}})^2 = (n-1)n^{-3}(\mathrm{tr}\mathbf{\Sigma})^2 + o_p(n^{-(1-\epsilon)})$, and

$$n\bar{\mathbf{r}}^\top\mathbf{\Sigma}\sum_{i=1}^n \mathbf{r}_i\mathbf{r}_i^\top\mathbf{\Sigma}\bar{\mathbf{r}} = (n^{-1}\mathrm{tr}\mathbf{\Sigma})^2 + (n-1)n^{-2}\mathrm{tr}(\mathbf{\Sigma}^2) + o_p(n^{-(1-\epsilon)}).$$

Thus, we have

$$\mathrm{tr}(\mathbf{S}_n^2) = \frac{n^2}{(n-1)^2}\mathrm{tr}[(\sum_{i=1}^n \mathbf{\Sigma}^{1/2}\mathbf{r}_i\mathbf{r}_i^\top\mathbf{\Sigma}^{1/2})^2] - \frac{n+1}{n(n-1)^2}(\mathrm{tr}\mathbf{\Sigma})^2 - \frac{2}{n-1}\mathrm{tr}(\mathbf{\Sigma}^2) + o_p(n^{-(1-\epsilon)}).$$

Table 3: Empirical test sizes and empirical powers (in percentage) of comparison of three methods with with $(p_1, \ldots, p_K) = (p/K, \ldots, p/K)$ and $K = 2, 3$ for Gamma variables. The vector $(\theta_1, \theta_2, \theta_3)$ specifies the Σ matrix. The rejection region is given in (2.6). When a test is not applicable, the corresponding entries are marked $-$.

$(\theta_1, \theta_2, \theta_3)$	n	Methods	p=180	360	900	180	360	900
			$K = 2$			$K = 3$		
			Empirical test sizes					
$(0, 0, 0)$	150	FDS	4.86	4.90	4.44	5.12	4.99	4.54
		BHPZ	4.46	—	—	5.22	4.90	—
		YHN	4.94	5.36	5.48	5.30	5.29	5.06
	300	FDS	4.92	4.82	4.81	4.92	5.02	4.84
		BHPZ	4.76	4.94	—	5.38	5.14	—
		YHN	4.84	4.92	5.10	5.02	5.22	4.90
			Empirical powers					
$(1, 0, 0)$	150	FDS	33.98	22.44	13.36	52.32	33.08	19.24
		BHPZ	5.89	—	—	8.02	5.02	—
		YHN	8.82	7.42	5.88	12.88	7.88	6.38
	300	FDS	95.56	90.78	81.34	99.58	99.02	95.42
		BHPZ	8.34	5.86	—	14.17	7.26	—
		YHN	13.28	8.76	6.12	22.12	11.93	7.16
$(0, 1, 0)$	150	FDS	59.82	47.54	31.44	79.76	67.61	47.70
		BHPZ	9.48	—	—	21.44	7.40	—
		YHN	8.44	6.92	5.76	10.30	7.78	6.10
	300	FDS	99.08	98.46	96.04	99.98	99.96	99.86
		BHPZ	31.06	11.00	—	74.36	27.98	—
		YHN	11.40	8.00	5.84	16.78	10.36	6.62
$(0, 0, 1)$	150	FDS	75.24	98.62	100	83.14	99.28	100
		BHPZ	7.40	—	—	11.42	6.62	—
		YHN	77.30	98.86	100	84.34	99.39	100
	300	FDS	99.38	100	100	99.74	100	100
		BHPZ	14.78	8.02	—	34.00	19.76	—
		YHN	99.50	100	100	99.78	100	100

Table 4: Empirical test sizes and powers (in percentage) for comparison of four methods with $n = 200$, $(p_1, \ldots, p_K) = (p/K, \ldots, p/K)$ and $K = 2, 3$ for Gaussian variables. The vector $(\theta_1, \theta_2, \theta_3)$ specifies the Σ matrix. The rejection region is given in (3.1).

(c_1, c_2)	$(\theta_1, \theta_2, \theta_3)$	Methods	$p = 60$	120	180	60	120	180
			$K = 2$			$K = 3$		
			Empirical test sizes					
(0.001, 1)	(0, 0, 0)	FDS	5.61	5.48	5.65	5.77	5.73	5.20
		CLRT	5.15	5.36	5.38	5.26	5.49	5.29
		BHPZ	5.20	5.08	4.88	4.86	5.29	5.15
		YHN	5.32	5.36	5.54	5.55	5.58	4.87
(5, 1)	(0, 0, 0)	FDS	5.61	5.48	5.65	5.77	5.73	5.20
		CLRT	5.15	5.36	5.38	5.26	5.49	5.29
		BHPZ	5.20	5.08	4.88	4.86	5.29	5.15
		YHN	5.32	5.36	5.54	5.55	5.58	4.87
			Empirical powers					
(0.001, 1)	(1, 0, 0)	FDS	87.63	77.34	70.30	98.20	93.21	88.66
		CLRT	19.54	9.78	7.07	38.47	14.27	8.51
		BHPZ	17.27	9.08	6.69	35.03	14.41	9.75
		YHN	27.55	13.91	9.60	52.16	22.80	14.83
(5, 1)	(1, 0, 0)	FDS	87.77	77.34	70.30	98.21	93.21	88.66
		CLRT	19.54	9.78	7.07	38.47	14.27	8.51
		BHPZ	17.27	9.08	6.69	35.03	14.41	9.75
		YHN	27.55	13.91	9.60	52.16	22.80	14.83

Table 5: Empirical test sizes and powers (in percentage) for comparison of four methods with $n = 200$, $(p_1, \ldots, p_K) = (p/K, \ldots, p/K)$ and $K = 2, 3$ for Gaussian variables. The vector $(\theta_1, \theta_2, \theta_3)$ specifies the Σ matrix. The rejection region is given in (3.1).

(c_1, c_2)	$(\theta_1, \theta_2, \theta_3)$	Methods	$p = 60$	120	180	60	120	180
			$K = 2$			$K = 3$		
			Empirical test sizes					
(1, 0.5)	(0, 0, 0)	FDS	20.53	29.74	39.05	22.40	32.66	41.39
		CLRT	5.15	5.36	5.38	5.26	5.49	5.29
		BHPZ	5.20	5.08	4.88	4.86	5.29	5.15
		YHN	5.32	5.36	5.54	5.55	5.58	4.87
(1, 2)	(0, 0, 0)	FDS	5.43	5.35	5.55	5.60	5.55	5.12
		CLRT	5.15	5.36	5.38	5.26	5.49	5.29
		BHPZ	5.20	5.08	4.88	4.86	5.29	5.15
		YHN	5.32	5.36	5.54	5.55	5.58	4.87
			Empirical powers					
(1, 0.5)	(1, 0, 0)	FDS	99.10	98.47	97.91	99.97	99.93	99.93
		CLRT	19.54	9.78	7.07	38.47	14.27	8.51
		BHPZ	17.27	9.08	6.69	35.03	14.41	9.75
		YHN	27.55	13.91	9.60	52.15	22.80	14.82
(1, 2)	(1, 0, 0)	FDS	40.26	20.03	13.34	64.40	31.24	19.74
		CLRT	19.54	9.78	7.07	38.47	14.27	8.51
		BHPZ	17.27	9.08	6.69	35.03	14.41	9.75
		YHN	27.55	13.91	9.60	52.15	22.80	14.82

Table 6: Empirical test sizes and powers (in percentage) for comparison of four methods with $n = 200$, $(p_1, \ldots, p_K) = (p/K, \ldots, p/K)$ and $K = 2, 3$ for Gaussian variables. The vector $(\theta_1, \theta_2, \theta_3)$ specifies the Σ matrix. The rejection region is given in (2.6).

n	$(\theta_1, \theta_2, \theta_3)$	Methods	$p = 6$	12	18	6	12	18
			$K = 2$			$K = 3$		
			Empirical test sizes					
600	(0, 0, 0)	FDS	6.65	6.28	5.92	6.82	6.12	5.87
		CLRT	6.51	6.13	5.65	6.68	5.90	5.67
		BHPZ	6.46	6.09	5.59	6.69	5.93	5.50
		YHN	6.57	5.97	5.65	6.72	5.92	5.64
750	(0, 0, 0)	FDS	6.36	6.22	5.84	6.46	6.12	6.36
		CLRT	6.48	5.99	5.81	6.49	5.84	6.19
		BHPZ	6.45	5.99	5.72	6.46	5.82	6.23
		YHN	6.35	6.04	5.79	6.39	6.00	6.19
1000	(0, 0, 0)	FDS	6.54	6.07	6.05	6.54	5.87	6.36
		CLRT	6.29	5.86	5.96	6.49	5.69	6.10
		BHPZ	6.26	5.83	5.90	6.39	5.67	6.21
		YHN	6.51	6.01	5.91	6.59	5.87	6.21

Because $\mathrm{tr}\mathbf{S}_n = n(n-1)^{-1}(\sum_{i=1}^n \mathbf{r}_i^\top \mathbf{\Sigma}\mathbf{r}_i - n\bar{\mathbf{r}}^\top \mathbf{\Sigma}\bar{\mathbf{r}})$, we have $\mathrm{tr}\mathbf{S}_n = n(n-1)^{-1}\sum_{i=1}^n \mathbf{r}_i^\top \mathbf{\Sigma}\mathbf{r}_i - (n-1)^{-1}\mathrm{tr}\mathbf{\Sigma} + o_p(n^{-(1-\epsilon)})$ by Lemma S.2.1 from the supplementary file 2. As shown in Bai and Silverstein (2004, p. 559-560),

$$\mathrm{tr}[(\sum_{i=1}^n \mathbf{\Sigma}^{1/2}\mathbf{r}_i\mathbf{r}_i^\top\mathbf{\Sigma}^{1/2})^q] - \mathrm{tr}[(\sum_{i=1}^n \mathbf{\Sigma}^{1/2}\tilde{\mathbf{r}}_i\tilde{\mathbf{r}}_i^\top\mathbf{\Sigma}^{1/2})^q] = o_p(n^{-1/4}),\ \ q = 1, 2$$

where $\tilde{\mathbf{r}}_i = n^{-1/2}\tilde{\mathbf{w}}_i$, $\tilde{\mathbf{w}}_i = (\tilde{w}_{1i}, \ldots, \tilde{w}_{pi})^\top$,

$$\tilde{w}_{\ell i} = [\mathrm{Var}(w_{\ell i}\delta_{\{|w_{\ell i}|\le\sqrt{n}\eta_n\}})]^{-1/2}(w_{\ell i}\delta_{\{|w_{\ell i}|\le\sqrt{n}\eta_n\}} - \mathrm{E}w_{\ell i}\delta_{\{|w_{\ell i}|\le\sqrt{n}\eta_n\}}),$$

$|\tilde{w}_{\ell i}| \le c\sqrt{n}\eta_n$, $\mathrm{E}\tilde{w}_{\ell i} = 0$, $\mathrm{E}(\tilde{w}_{\ell i}^2) = 1$ and $\mathrm{E}(\tilde{w}_{\ell i}^4) < \infty$ for $\ell = 1, \ldots, p$ and $i = 1, \ldots, n$ with $\eta_n \downarrow 0$, $n^{1/4}\eta_n \to \infty$ and c being a positive constant. For simplicity, we shall rename the variables $\tilde{w}_{\ell i}$ simply as $w_{\ell i}$ and proceed by assuming that $|w_{\ell i}| \le \sqrt{n}\eta_n$, $\mathrm{E}w_{\ell i} = 0$, $\mathrm{E}(w_{\ell i}^2) = 1$ and $\mathrm{E}(w_{\ell i}^4) < \infty$ with $\eta_n \downarrow 0$ and $n^{1/4}\eta_n \to \infty$. Let $\mathbf{B}_n = \sum_{i=1}^n \mathbf{\Sigma}^{1/2}\mathbf{r}_i\mathbf{r}_i^\top\mathbf{\Sigma}^{1/2}$, then

$$\mathrm{tr}(\mathbf{S}_n^2) = \frac{n^2}{(n-1)^2}\mathrm{tr}(\mathbf{B}_n^2) - \frac{n+1}{n(n-1)^2}(\mathrm{tr}\mathbf{\Sigma})^2 - \frac{2}{n-1}\mathrm{tr}(\mathbf{\Sigma}^2) + o_p(n^{-1/4}). \quad \text{(S.1)}$$

Similarly, let $\mathbf{B}_n = \sum_{i=1}^n \mathbf{\Sigma}_{kk}^{1/2}\mathbf{r}_{ik}\mathbf{r}_{ik}^\top\mathbf{\Sigma}_{kk}^{1/2}$, then $\mathrm{tr}\mathbf{S}_{kk} = n(n-1)^{-1}\sum_{i=1}^n \mathbf{r}_{ik}^\top\mathbf{\Sigma}_{kk}\mathbf{r}_{ik} - (n-1)^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk} + o_p(1)$ and

$$\mathrm{tr}(\mathbf{S}_{kk}^2) = \frac{n^2}{(n-1)^2}\mathrm{tr}(\mathbf{B}_{kk}^2) - \frac{n+1}{n(n-1)^2}(\mathrm{tr}\mathbf{\Sigma}_{kk})^2 - \frac{2}{n-1}\mathrm{tr}(\mathbf{\Sigma}_{kk}^2) + o_p(n^{-1/4}), \quad \text{(S.2)}$$

where $o_p(n^{-1/4})$ is uniform for $k = 1, \ldots, K$.

S.1 Part I of Lemma 1 and its proof

Lemma S.1. *Under Assumption [A]-[B] and under* $H_0 : \mathbf{\Sigma} = \mathrm{diag}(\mathbf{\Sigma}_{11}, \ldots, \mathbf{\Sigma}_{KK})$, *we have* $\sigma^{-1}(T_{n1} - \mu) \to N(0, 1)$, *where the quantities* μ *and* σ *are given in Lemma 1 in the main paper.*

Proof of Lemma S.1. First note that $T_{n1} = \mathrm{tr}[\mathbf{S}_n - \mathrm{diag}(\mathbf{S}_{11}, \ldots, \mathbf{S}_{KK})]^2 = \mathrm{tr}(\mathbf{S}_n^2) - \sum_{k=1}^K \mathrm{tr}(\mathbf{S}_{kk}^2)$. By (S.1) and (S.2), we have

$$\begin{aligned} T_{n1} &= \frac{n^2}{(n-1)^2}[\mathrm{tr}(\mathbf{B}_n^2) - \sum_{k=1}^K \mathrm{tr}(\mathbf{B}_{kk}^2)] \\ &\quad - \frac{n+1}{n(n-1)^2}[(\mathrm{tr}\mathbf{\Sigma})^2 - \sum_{k=1}^K (\mathrm{tr}\mathbf{\Sigma}_{kk})^2] - \frac{2}{n-1}[\mathrm{tr}(\mathbf{\Sigma}^2) - \sum_{k=1}^K \mathrm{tr}(\mathbf{\Sigma}_{kk}^2)] + o_p(n^{-1/4}). \end{aligned} \tag{S.3}$$

Under H_0, we have

$$T_{n1} = \frac{n^2}{(n-1)^2}[\mathrm{tr}(\mathbf{B}_n^2) - \sum_{k=1}^K \mathrm{tr}(\mathbf{B}_{kk}^2)] - \frac{n+1}{n(n-1)^2}[(\mathrm{tr}\mathbf{\Sigma})^2 - \sum_{k=1}^K (\mathrm{tr}\mathbf{\Sigma}_{kk})^2] + o_p(n^{-1/4}).$$

That is, the central limit theorem for T_{n1} can be obtained by establishing the central limit theorem for $[\mathrm{tr}(\mathbf{B}_n^2) - \sum_{k=1}^K \mathrm{tr}(\mathbf{B}_{kk}^2)]$. We need to compute the mean μ and the variance σ^2 of the statistic T_{n1}. The asymptotic normality is due to the fact that $\{\mathrm{E}_j(\mathrm{tr}\mathbf{B}_n^2) - \mathrm{E}_{j-1}(\mathrm{tr}\mathbf{B}_n^2), j = 1, \ldots, n\}$ and $\{\mathrm{E}_j(\mathrm{tr}\mathbf{B}_{kk}^2) - \mathrm{E}_{j-1}(\mathrm{tr}\mathbf{B}_{kk}^2), j = 1, \ldots, n\}$ for $k = 1, \ldots, K$ are two martingale difference sequences, where we use E_j as the conditional expectation given $\mathbf{x}_1, \ldots, \mathbf{x}_j$. Lemma S.2.3 from the supplementary file 2 shows that these martingale difference

sequences satisfy the Lindeberg's conditions, that is,

$$\sum_{j=1}^{n} \mathrm{E}([\mathrm{E}_j(\mathrm{tr}\mathbf{B}_n^2) - \mathrm{E}_{j-1}(\mathrm{tr}\mathbf{B}_n^2)]^2 \delta_{\{|\mathrm{E}_j(\mathrm{tr}\mathbf{B}_n^2)-\mathrm{E}_{j-1}(\mathrm{tr}\mathbf{B}_n^2)|\geq\epsilon\}}) = O(\eta_n^4), \tag{S.4}$$

$$\sum_{j=1}^{n} \mathrm{E}([\mathrm{E}_j(\mathrm{tr}\mathbf{B}_{kk}^2) - \mathrm{E}_{j-1}(\mathrm{tr}\mathbf{B}_{kk}^2)]^2 \delta_{\{|\mathrm{E}_j(\mathrm{tr}\mathbf{B}_{kk}^2)-\mathrm{E}_{j-1}(\mathrm{tr}\mathbf{B}_{kk}^2)|\geq\epsilon\}}) = O(\eta_n^4), \tag{S.5}$$

for any $\epsilon > 0$ where $O(\eta_n^4)$ is uniform for $k = 1, ..., K$. For simplicity, $\mathrm{E}_j(\mathrm{tr}\mathbf{B}_n^2) - \mathrm{E}_{j-1}(\mathrm{tr}\mathbf{B}_n^2)$ is often written as $(\mathrm{E}_j - \mathrm{E}_{j-1})(\mathrm{tr}\mathbf{B}_n^2)$ in this paper.

To compute the mean and the variance, we take the following two steps.

Step 1 computes the mean

$$\mu = \frac{n^2}{(n-1)^2}\mathrm{E}[\mathrm{tr}(\mathbf{B}_n^2) - \sum_{k=1}^{K}\mathrm{tr}(\mathbf{B}_{kk}^2)] - \frac{n+1}{n(n-1)^2}[(\mathrm{tr}\mathbf{\Sigma})^2 - \sum_{k=1}^{K}(\mathrm{tr}\mathbf{\Sigma}_{kk})^2].$$

We have

$$\mathrm{E}[\mathrm{tr}(\mathbf{B}_n^2)] = n^{-1}[2\mathrm{tr}\mathbf{\Sigma}^2 + \beta_w \sum_{j=1}^{p}(\mathbf{e}_j^\top \mathbf{\Sigma}\mathbf{e}_j)^2] + n^{-1}(\mathrm{tr}\mathbf{\Sigma})^2 + (n-1)n^{-1}\mathrm{tr}(\mathbf{\Sigma}^2),$$

$$\mathrm{E}[\mathrm{tr}(\mathbf{B}_{kk}^2)] = n^{-1}[2\mathrm{tr}\mathbf{\Sigma}_{kk}^2 + \beta_w \sum_{j=1}^{p_k}(\mathbf{e}_{jk}^\top \mathbf{\Sigma}_{kk}\mathbf{e}_{jk})^2] + n^{-1}(\mathrm{tr}\mathbf{\Sigma}_{kk})^2 + (n-1)n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{kk}^2),$$

for $k = 1, \ldots, K$. Then under H_0, we have

$$\mu = \frac{n^2-n-1}{n(n-1)^2}(\mathrm{tr}\mathbf{\Sigma})^2 - \frac{n^2-n-1}{n(n-1)^2}\sum_{k=1}^{K}(\mathrm{tr}\mathbf{\Sigma}_{kk})^2.$$

Step 2 shows that $\sigma^2 = \sigma_{00} + \sum_{k=1}^{K}\sigma_{kk} - 2\sum_{k=1}^{K}\sigma_{0k}$ converges in probability, where $\sigma_{00} = \sum_{j=1}^{n}\mathrm{E}_{j-1}[(\mathrm{E}_j - \mathrm{E}_{j-1})(\mathrm{tr}\mathbf{B}_n^2)]^2$, $\sigma_{kk} = \sum_{j=1}^{n}\mathrm{E}_{j-1}[(\mathrm{E}_j - \mathrm{E}_{j-1})(\mathrm{tr}\mathbf{B}_{kk}^2)]^2$, $\sigma_{0k} = \sum_{j=1}^{n}\mathrm{E}_{j-1}\{[(\mathrm{E}_j - \mathrm{E}_{j-1})(\mathrm{tr}\mathbf{B}_n^2)][(\mathrm{E}_j - \mathrm{E}_{j-1})(\mathrm{tr}\mathbf{B}_{kk}^2)]\}$ for

$k = 1, \cdots, K$. To do so, we have

$$
\begin{aligned}
& (\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_n^2 \\
= \; & 2(n-j)n^{-1}(\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{r}_j^\top \boldsymbol{\Sigma}^2 \mathbf{r}_j + (\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{r}_j^\top \boldsymbol{\Sigma}\mathbf{r}_j \mathbf{r}_j^\top \boldsymbol{\Sigma}\mathbf{r}_j \\
& +2\sum_{\ell \le j-1} (\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{r}_j^\top \boldsymbol{\Sigma}\mathbf{r}_\ell \mathbf{r}_\ell^\top \boldsymbol{\Sigma}\mathbf{r}_j \\
= \; & 2(n-j)n^{-1}(\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{r}_j^\top \boldsymbol{\Sigma}^2 \mathbf{r}_j + (\mathbf{r}_j^\top \boldsymbol{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2 - \mathrm{E}[(\mathbf{r}_j^\top \boldsymbol{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2] \\
& +2(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})(\mathbf{r}_j^\top \boldsymbol{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) + 2\sum_{\ell \le j-1} (\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{r}_j^\top \boldsymbol{\Sigma}\mathbf{r}_\ell \mathbf{r}_\ell^\top \boldsymbol{\Sigma}\mathbf{r}_j,
\end{aligned}
$$

$$
\begin{aligned}
& (\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_{11}^2 \\
= \; & 2(n-j)n^{-1}(\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}^2 \mathbf{r}_{j1} + (\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1}\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1} \\
& +2\sum_{\ell \le j-1}(\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{\ell 1}\mathbf{r}_{\ell 1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1} \\
= \; & 2(n-j)n^{-1}(\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}^2 \mathbf{r}_{j1} \\
& +(\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})^2 - \mathrm{E}[(\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})^2] \\
& +2(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})(\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}) + 2\sum_{\ell \le j-1}(\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{\ell 1}\mathbf{r}_{\ell 1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1},
\end{aligned}
$$

where

$$
\begin{aligned}
(\mathbf{r}_j^\top \mathbf{\Sigma}\mathbf{r}_j)^2 - \mathrm{E}(\mathbf{r}_j^\top \mathbf{\Sigma}\mathbf{r}_j)^2 &= (\mathbf{r}_j^\top \mathbf{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\mathbf{\Sigma})^2 - \mathrm{E}(\mathbf{r}_j^\top \mathbf{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\mathbf{\Sigma})^2 \\
&\quad +2(n^{-1}\mathrm{tr}\mathbf{\Sigma})(\mathbf{r}_j^\top \mathbf{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\mathbf{\Sigma}), \\
(\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1})^2 - \mathrm{E}(\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1})^2 &= (\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})^2 - \mathrm{E}(\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})^2 \\
&\quad +2(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})(\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}).
\end{aligned}
$$

We first compute σ_{01}, and the calculations of $\{\sigma_{0k}, k = 2, \ldots, K\}$ can be similarly obtained.

$$
\begin{aligned}
\sigma_{01} &= \sum_{j=1}^{n} \mathrm{E}_{j-1}[(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_n^2][(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_{11}^2] \\
&= \sum_{j=1}^{n} \mathrm{E}_{j-1}\Bigg\{(\mathrm{E}_j - \mathrm{E}_{j-1})\left[\frac{2(n-j)}{n}\mathbf{r}_j^\top \mathbf{\Sigma}^2\mathbf{r}_j + \mathbf{r}_j^\top \mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top \mathbf{\Sigma}\mathbf{r}_j + 2\sum_{\ell \le j-1}\mathbf{r}_j^\top \mathbf{\Sigma}\mathbf{r}_\ell\mathbf{r}_\ell^\top \mathbf{\Sigma}\mathbf{r}_j\right] \\
&\qquad (\mathrm{E}_j - \mathrm{E}_{j-1})\left[\frac{2(n-j)}{n}\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}^2\mathbf{r}_{j1} + \mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1}\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1} + 2\sum_{\ell \le j-1}\mathbf{r}_{j1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{\ell 1}\mathbf{r}_{\ell 1}^\top \mathbf{\Sigma}_{11}\mathbf{r}_{j1}\right]\Bigg\} \\
&= (S.6) + (S.7) + (S.8),
\end{aligned}
$$

where (S.6)-(S.8) are given as follows.

$$\sum_{j=1}^{n} 2(n-j)n^{-1}\mathrm{E}_{j-1}\big\{(\mathbf{r}_j^\top\boldsymbol{\Sigma}^2\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2)(\mathrm{E}_j-\mathrm{E}_{j-1})\big[2(n-j)n^{-1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1} + \mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1} + 2\sum_{\ell\le j-1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{\ell 1}\mathbf{r}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}\big]\big\}, \tag{S.6}$$

$$\sum_{j=1}^{n} \mathrm{E}_{j-1}\big\{\mathrm{E}_j-\mathrm{E}_{j-1})(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j)(\mathrm{E}_j-\mathrm{E}_{j-1})\big[2(n-j)n^{-1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1} + \mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1} + 2\sum_{\ell\le j-1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{\ell 1}\mathbf{r}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}\big]\big\}, \tag{S.7}$$

$$2\sum_{j=1}^{n} \mathrm{E}_{j-1}\big\{\mathrm{E}_j-\mathrm{E}_{j-1})(\sum_{\ell\le j-1}\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_\ell\mathbf{r}_\ell^\top\boldsymbol{\Sigma}\mathbf{r}_j)(\mathrm{E}_j-\mathrm{E}_{j-1})\big[2(n-j)n^{-1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1} + \mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1} + 2\sum_{\ell\le j-1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{\ell 1}\mathbf{r}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}\big]\big\}. \tag{S.8}$$

As verified in the supplementary file, we have

$$\begin{aligned}
(S.6) &= 2(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^3) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell 1}\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell 1}] \\
&\quad +2[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^4) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell 1})^2] + O_p(\eta_n^2), \\
(S.7) &= 4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^2) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell 1})^2] \\
&\quad +4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^3) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell 1}\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell 1}] + O_p(\eta_n^2), \\
(S.8) &= 2(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^3) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell 1})(\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell 1})] \\
&\quad +2[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^4) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell 1})^2] + 4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}^2)^2 + O_p(\eta_n^2).
\end{aligned}$$

Thus under H_0, we have

$$\begin{aligned}
\sigma_{01} &= (S.6)+(S.7)+(S.8) \\
&= 4[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^4)+\beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_{\ell 1})^2]+4[n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^2)]^2 \\
&\quad +(4n^{-1}\mathrm{tr}\mathbf{\Sigma}+4n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^3)+\beta_w n^{-1}\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell 1}^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_{\ell 1}\mathbf{e}_{\ell 1}^\top\mathbf{\Sigma}_{11}\mathbf{e}_{\ell 1}] \\
&\quad +4(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})(n^{-1}\mathrm{tr}\mathbf{\Sigma})[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^2)+\beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top\mathbf{\Sigma}_{11}\mathbf{e}_{\ell 1})^2]+O_p(\eta_n^2).
\end{aligned}$$

Similarly, for $k=2,\ldots,K$, under H_0, we have

$$\begin{aligned}
\sigma_{0k} &= 4[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{kk}^4)+\beta_w n^{-1}\sum_{\ell=1}^{p_k}(\mathbf{e}_{\ell k}^\top\mathbf{\Sigma}_{kk}^2\mathbf{e}_{\ell k})^2]+4(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk}^2)^2 \\
&\quad +(4n^{-1}\mathrm{tr}\mathbf{\Sigma}+4n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{kk}^3)+\beta_w n^{-1}\sum_{\ell=1}^{p_k}\mathbf{e}_{\ell k}^\top\mathbf{\Sigma}_{kk}^2\mathbf{e}_{\ell k}\mathbf{e}_{\ell k}^\top\mathbf{\Sigma}_{kk}\mathbf{e}_{\ell k}] \\
&\quad +4(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})(n^{-1}\mathrm{tr}\mathbf{\Sigma})[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{kk}^2)+\beta_w n^{-1}\sum_{\ell=1}^{p_k}(\mathbf{e}_{\ell k}^\top\mathbf{\Sigma}_{kk}\mathbf{e}_{\ell k})^2]+O_p(\eta_n^2),
\end{aligned}$$

$$\begin{aligned}
\sigma_{00} &= \sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}[(\mathrm{E}_\ell-\mathrm{E}_{\ell-1})(\mathrm{tr}\mathbf{B}_n^2)]^2 \\
&= 4[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}^4)+\beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}^2\mathbf{e}_\ell)^2] \\
&\quad +4(n^{-1}\mathrm{tr}\mathbf{\Sigma})^2[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}^2)+\beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}\mathbf{e}_\ell)^2] \\
&\quad +4[n^{-1}\mathrm{tr}(\mathbf{\Sigma}^2)]^2+8(n^{-1}\mathrm{tr}\mathbf{\Sigma})[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}^3)+\beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\mathbf{\Sigma}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}\mathbf{e}_\ell]+o_p(1)
\end{aligned}$$

$$\begin{aligned}
\sigma_{kk} &= \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_{kk}^2]^2 \\
&= 4n^{-1}[2\mathrm{tr}(\mathbf{\Sigma}_{kk}^4) + \beta_w \sum_{\ell=1}^{p_k} (\mathbf{e}_{\ell k}^\top \mathbf{\Sigma}_{kk}^2 \mathbf{e}_{\ell k})^2] \\
&\quad +4(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})^2 n^{-1}[2\mathrm{tr}(\mathbf{\Sigma}_{kk}^2) + \beta_w \sum_{\ell=1}^{p_k} (\mathbf{e}_{\ell k}^\top \mathbf{\Sigma}_{kk} \mathbf{e}_{\ell k})^2] + 4[n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{kk}^2)]^2 \\
&\quad +8(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk}) n^{-1}[2\mathrm{tr}(\mathbf{\Sigma}_{kk}^3) + \beta_w \sum_{\ell=1}^{p_k} \mathbf{e}_{\ell k}^\top \mathbf{\Sigma}_{kk}^2 \mathbf{e}_{\ell k} \mathbf{e}_{\ell}^\top \mathbf{\Sigma}_{kk} \mathbf{e}_{\ell}] + O_p(\eta_n^2).
\end{aligned}$$

Putting things together, we have under H_0,

$$\begin{aligned}
\sigma^2 &= \sigma_{00} + \sum_{k=1}^{K} \sigma_{kk} - 2\sum_{k=1}^{K} \sigma_{0k} \\
&= 4\sum_{k=1}^{K} (n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk} - n^{-1}\mathrm{tr}\mathbf{\Sigma})^2 [2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{kk}^2) + \beta_w n^{-1} \sum_{\ell=1}^{p_k} (\mathbf{e}_{\ell k}^\top \mathbf{\Sigma}_{kk} \mathbf{e}_{\ell k})^2] \\
&\quad +4[n^{-1}\mathrm{tr}(\mathbf{\Sigma}^2)]^2 - 4\sum_{k=1}^{K} [n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{kk}^2)]^2 + O_p(K\eta_n^2).
\end{aligned}$$

S.2 Proofs of Theorem 1, Part II of Lemma 1 and Theorem 2

S.2.1 Proof of Theorem 2

Under H_0, we have $(T_{n1} - \mu)/\sigma \to N(0,1)$. But

$$\mu = \frac{n^2 - n - 1}{n(n-1)^2} (\mathrm{tr}\mathbf{\Sigma})^2 - \frac{n^2 - n - 1}{n(n-1)^2} \sum_{k=1}^{K} (\mathrm{tr}\mathbf{\Sigma}_{kk})^2$$

is unknown. We now replace $\mathrm{tr}\mathbf{\Sigma}$ and $\mathrm{tr}\mathbf{\Sigma}_{kk}$ by $\mathrm{tr}\mathbf{S}_n$ and $\mathrm{tr}\mathbf{S}_{kk}$ in μ, and establish the asymptotic distribution of

$$T_{n1} - \hat{\mu} = \mathrm{tr}\mathbf{S}_n^2 - \sum_{k=1}^{K} \mathrm{tr}\mathbf{S}_{kk}^2 - \hat{\mu}$$

where $\hat{\mu} = \frac{n^2-n-1}{n(n-1)^2}[(\mathrm{tr}\mathbf{S}_n)^2 - \sum_{k=1}^K (\mathrm{tr}\mathbf{S}_{kk})^2]$. By (S.1) and (S.2), we have

$$\begin{aligned}
& T_{n1} - \hat{\mu} \\
= \; & \frac{n^2}{(n-1)^2}[\mathrm{tr}(\mathbf{B}_n^2) - \sum_{k=1}^K \mathrm{tr}(\mathbf{B}_{kk}^2)] \\
& - \frac{n}{n-1}\frac{n^2-n-1}{(n-1)^3}[(\mathrm{tr}\mathbf{B}_n - n^{-1}\mathrm{tr}\mathbf{\Sigma})^2 - \sum_{k=1}^K (\mathrm{tr}\mathbf{B}_{kk} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})^2] \\
& - \frac{n+1}{n(n-1)^2}[(\mathrm{tr}\mathbf{\Sigma})^2 - \sum_{k=1}^K (\mathrm{tr}\mathbf{\Sigma}_{kk})^2] - \frac{2}{n-1}[\mathrm{tr}(\mathbf{\Sigma}^2) - \sum_{k=1}^K \mathrm{tr}(\mathbf{\Sigma}_{kk}^2)] + o_p(n^{-1/4}).
\end{aligned}$$

That is, the central limit theorem for $T_{n1} - \hat{\mu}$ can be obtained by establishing the central limit theorem for $(\mathrm{tr}\mathbf{B}_n^2 - \sum_{k=1}^K \mathrm{tr}\mathbf{B}_{kk}^2, \mathrm{tr}\mathbf{B}_n, \mathrm{tr}\mathbf{B}_{11}, \ldots, \mathrm{tr}\mathbf{B}_{KK})$. The asymptotic normality is due to the fact that the sequences $\{(\mathrm{E}_j - \mathrm{E}_{j-1})(\mathrm{tr}\mathbf{B}_n^2), j = 1, \ldots, n\}$, $\{(\mathrm{E}_j - \mathrm{E}_{j-1})(\mathrm{tr}\mathbf{B}_n), j = 1, \ldots, n\}$, $\{(\mathrm{E}_j - \mathrm{E}_{j-1})(\mathrm{tr}\mathbf{B}_{kk}), j = 1, \ldots, n\}$ and $\{(\mathrm{E}_j - \mathrm{E}_{j-1})(\mathrm{tr}\mathbf{B}_{kk}^2), j = 1, \ldots, n\}$ for $k = 1, \ldots, K$ are martingale difference sequences and Lindeberg-type conditions are satisfied by Lemma S.2.3 from the supplementary file 2. Then we have

$$\sigma_1^{-1}\{T_{n1} - \hat{\mu} - \mu_1\} \to N(0, 1),$$

where $\mu_1 = n^2(n-1)^{-2}\mathrm{E}[\mathrm{tr}(\mathbf{B}_n^2) - \sum_{k=1}^K \mathrm{tr}(\mathbf{B}_{kk}^2)] - (n+1)n^{-1}(n-1)^{-2}[(\mathrm{tr}\mathbf{\Sigma})^2 - \sum_{k=1}^K (\mathrm{tr}\mathbf{\Sigma}_{kk})^2] - 2(n-1)^{-1}[\mathrm{tr}(\mathbf{\Sigma}^2) - \sum_{k=1}^K \mathrm{tr}(\mathbf{\Sigma}_{kk}^2)] - \mu$ and

$$\begin{aligned}
\sigma_1^2 \;=\; & \sigma_{00A} + \sum_{k=1}^K \sigma_{kkA} - 2\sum_{k=1}^K \sigma_{0kA} - 4(n^{-1}\mathrm{tr}\mathbf{\Sigma})\sigma_{000A} + 4(n^{-1}\mathrm{tr}\mathbf{\Sigma})\sum_{k=1}^K \sigma_{00kA} \\
& + 4\sum_{k=1}^K (n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})\sigma_{0kkA} - 4\sum_{k=1}^K (n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})\sigma_{kkkA} + 4(n^{-1}\mathrm{tr}\mathbf{\Sigma})^2\sigma_{0000A} \\
& + 4\sum_{k=1}^K (n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})^2\sigma_{kkkkA} - 8\sum_{k=1}^K (n^{-1}\mathrm{tr}\mathbf{\Sigma})(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})\sigma_{00kkA},
\end{aligned}$$

if the following terms converge in probability

$$\sigma_{00A} = \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_n^2]^2,$$

$$\sigma_{0kA} = \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_n^2][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_{kk}^2]\},$$

$$\sigma_{kkA} = \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_{kk}^2]^2,$$

$$\sigma_{0000A} = \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_n][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_n]\},$$

$$\sigma_{000A} = \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_n^2][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_n]\},$$

$$\sigma_{kkkkA} = \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_{kk}]^2,$$

$$\sigma_{kkkA} = \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_{kk}^2][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_{kk}]\},$$

$$\sigma_{0kkA} = \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_n^2][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_{kk}]\},$$

$$\sigma_{00kA} = \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_n][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_{kk}^2]\},$$

$$\sigma_{00kkA} = \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_{kk}][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{B}_n].$$

The first step is to compute μ_1. Because $\mathrm{E}[\mathrm{tr}(\mathbf{B}_n^2)] = n^{-1}[2\mathrm{tr}\mathbf{\Sigma}^2 + \beta_w \sum_{j=1}^{p}(\mathbf{e}_j^\top \mathbf{\Sigma}\mathbf{e}_j)^2] + n^{-1}(\mathrm{tr}\mathbf{\Sigma})^2 + (n-1)n^{-1}\mathrm{tr}(\mathbf{\Sigma}^2)$ and $\mathrm{E}[\mathrm{tr}(\mathbf{B}_{kk}^2)] = n^{-1}[2\mathrm{tr}\mathbf{\Sigma}_{kk}^2 + \beta_w \sum_{j=1}^{p_k}(\mathbf{e}_{jk}^\top \mathbf{\Sigma}_{kk}\mathbf{e}_{jk})^2] + n^{-1}(\mathrm{tr}\mathbf{\Sigma}_{kk})^2 + (n-1)n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{kk}^2)$ for $k = 1, \ldots, K$, thus we have

$$\begin{aligned}\mu_1 \;=\; & n^2(n-1)^{-2}\mathrm{E}[\mathrm{tr}(\mathbf{B}_n^2) - \sum_{k=1}^{K}\mathrm{tr}(\mathbf{B}_{kk}^2)] - (n+1)n^{-1}(n-1)^{-2}[(\mathrm{tr}\mathbf{\Sigma})^2 - \sum_{k=1}^{K}(\mathrm{tr}\mathbf{\Sigma}_{kk})^2] \\ & -2(n-1)^{-1}[\mathrm{tr}(\mathbf{\Sigma}^2) - \sum_{k=1}^{K}\mathrm{tr}(\mathbf{\Sigma}_{kk}^2)] - \mu = \frac{n^2-n+2}{(n-1)^2}\mathrm{tr}\mathbf{A}\end{aligned}$$

where $\mathbf{A} = \mathbf{\Sigma}^2 - \mathrm{diag}(\mathbf{\Sigma}_{11}^2, \ldots, \mathbf{\Sigma}_{KK}^2)$.

The second step is to compute σ_1^2. Let $\mathbf{\Sigma}_{(kk)}$ is the $p \times p$ dimensional matrix with the kth diagonal block being $\mathbf{\Sigma}_{kk}$ and other entries being zeros. The detailed proofs of σ_{00A}, σ_{0kA}, σ_{kkA}, σ_{0000A}, σ_{000A}, σ_{kkkkA}, σ_{kkkA}, σ_{0kkA}, σ_{00kA} and σ_{00kkA} are similar for $k = 1, \ldots, K$. Moreover, the proof of σ_{01A} is similar to σ_{01}. Therefore, we do not give the details of the proofs of σ_{01A}. We have

$$\sigma_{01A} = \sum_{j=1}^{n} \mathrm{E}_{j-1}[(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n^2][(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_{11}^2] = (S.6) + (S.7) + (S.8)$$

where under the alternative hypothesis,

$$\begin{aligned}
(S.6) &= 2[2n^{-1}\mathrm{tr}\mathbf{\Sigma}^2\mathbf{\Sigma}_{(11)}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}^2\mathbf{e}_\ell)(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{(11)}^2\mathbf{e}_\ell)] \\
&\quad +2n^{-1}\mathrm{tr}\mathbf{\Sigma}_{(11)}[2n^{-1}\mathrm{tr}\mathbf{\Sigma}^2\mathbf{\Sigma}_{(11)} + \beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\mathbf{\Sigma}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}_{(11)}\mathbf{e}_\ell] + O_p(\eta_n^4), \\
(S.7) &= 4(n^{-1}\mathrm{tr}\mathbf{\Sigma})\big(2n^{-1}\mathrm{tr}\mathbf{\Sigma}\mathbf{\Sigma}_{(11)}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\mathbf{\Sigma}\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}_{(11)}^2\mathbf{e}_\ell\big) \\
&\quad +4(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{(11)})(n^{-1}\mathrm{tr}\mathbf{\Sigma})\big[2n^{-1}\mathrm{tr}\mathbf{\Sigma}\mathbf{\Sigma}_{(11)} + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}\mathbf{e}_\ell)(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{(11)}\mathbf{e}_\ell)\big] + O_p(\eta_n^4), \\
(S.8) &= 2[2n^{-1}\mathrm{tr}\mathbf{\Sigma}^2\mathbf{\Sigma}_{(11)}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}^2\mathbf{e}_\ell)(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{(11)}^2\mathbf{e}_\ell)] \\
&\quad +2(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{(11)})\big[2n^{-1}\mathrm{tr}\mathbf{\Sigma}^2\mathbf{\Sigma}_{(11)} + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}^2\mathbf{e}_\ell)(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{(11)}\mathbf{e}_\ell)\big] + O_p(\eta_n^4).
\end{aligned}$$

Therefore, under the alternative hypothesis, we have

$$\begin{aligned}
\sigma_{01A} &= 4[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2\boldsymbol{\Sigma}_{(11)}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}^2\mathbf{e}_\ell)(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{(11)}^2\mathbf{e}_\ell)] \\
&\quad +4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})(2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{(11)}^3 + \beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\boldsymbol{\Sigma}\mathbf{e}_\ell\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{(11)}^2\mathbf{e}_\ell) \\
&\quad +4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{(11)}(2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2\boldsymbol{\Sigma}_{(11)} + \beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\boldsymbol{\Sigma}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{(11)}\mathbf{e}_\ell) \\
&\quad +4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{(11)})(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{(11)}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}\mathbf{e}_\ell)(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{(11)}\mathbf{e}_\ell)] \\
&\quad +4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{(11)})^2 + O_p(\eta_n^2).
\end{aligned}$$

Similarly, for $k = 1, \ldots, K$, under the alternative hypothesis, we have

$$\begin{aligned}
\sigma_{0kA} &= \sum_{j=1}^{n}\mathrm{E}_{j-1}[(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n^2][(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_{kk}^2] \\
&= 4\big[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2\boldsymbol{\Sigma}_{(kk)}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}^2\mathbf{e}_\ell)(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{(kk)}^2\mathbf{e}_\ell)\big] + 4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{(kk)}^2)^2 \\
&\quad +4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})\big(2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{(kk)}^3 + \beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{(kk)}\mathbf{e}_\ell\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{(kk)}^2\mathbf{e}_\ell\big) \\
&\quad +4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{(kk)})\big(2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2\boldsymbol{\Sigma}_{(kk)} + \beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\boldsymbol{\Sigma}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{(kk)}\mathbf{e}_\ell\big) \\
&\quad +4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{(kk)})(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})\big[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{(kk)}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{(kk)}\mathbf{e}_\ell)^2\big] + O_p(\eta_n^2),
\end{aligned}$$

$$\begin{aligned}
\sigma_{00A} &= \sum_{\ell=1}^{n}\mathrm{E}_{\ell-1}[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n^2]^2 \\
&= 4n^{-1}[2\mathrm{tr}\boldsymbol{\Sigma}^4 + \beta_w\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}^2\mathbf{e}_\ell)^2] + 4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2 n^{-1}\big[2\mathrm{tr}\boldsymbol{\Sigma}^2 + \beta_w\sum_{\ell}^{p}(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}\mathbf{e}_\ell)^2\big] \\
&\quad +4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2)^2 + 8(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})n^{-1}\big[2\mathrm{tr}\boldsymbol{\Sigma}^3 + \beta_w\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\boldsymbol{\Sigma}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\boldsymbol{\Sigma}\mathbf{e}_\ell\big] + O_p(\eta_n^2),
\end{aligned}$$

$$\begin{aligned}
\sigma_{kkA} &= \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_{kk}^2]^2 \\
&= 4n^{-1}[2\mathrm{tr}\mathbf{\Sigma}_{kk}^4 + \beta_w \sum_{\ell=1}^{p_k}(\mathbf{e}_{\ell k}^\top \mathbf{\Sigma}_{kk}^2 \mathbf{e}_{\ell k})^2] + 4(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})^2 n^{-1}\big[2\mathrm{tr}\mathbf{\Sigma}_{kk}^2 + \beta_w \sum_{\ell}^{p_k}(\mathbf{e}_{\ell k}^\top \mathbf{\Sigma}_{kk}\mathbf{e}_{\ell k})^2\big] \\
&\quad +4(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk}^2)^2 + 8(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})n^{-1}\big[2\mathrm{tr}\mathbf{\Sigma}_{kk}^3 + \beta_w \sum_{\ell=1}^{p_k}\mathbf{e}_{\ell k}^\top \mathbf{\Sigma}_{kk}^2 \mathbf{e}_{\ell k}\mathbf{e}_\ell^\top \mathbf{\Sigma}_{kk}\mathbf{e}_\ell\big] + O_p(\eta_n^2),
\end{aligned}$$

$$\begin{aligned}
\sigma_{0000A} &= \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n]\} \\
&= 2n^{-1}\mathrm{tr}(\mathbf{\Sigma}^2) + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top \mathbf{\Sigma}\mathbf{e}_\ell)^2 + O_p(\eta_n^2),
\end{aligned}$$

$$\begin{aligned}
\sigma_{000A} &= \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n^2][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n]\} \\
&= 2\big(2n^{-1}\mathrm{tr}\mathbf{\Sigma}^3 + \beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top \mathbf{\Sigma}\mathbf{e}_\ell \mathbf{e}_\ell^\top \mathbf{\Sigma}^2 \mathbf{e}_\ell\big) \\
&\quad +2(n^{-1}\mathrm{tr}\mathbf{\Sigma})\big(2n^{-1}\mathrm{tr}\mathbf{\Sigma}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top \mathbf{\Sigma}\mathbf{e}_\ell)^2\big) + O_p(\eta_n^2),
\end{aligned}$$

$$\sigma_{kkkkA} = \sum_{j=1}^{n} \mathrm{E}_{j-1}\{[(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_{kk}]^2 = 2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{kk}^2) + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top \mathbf{\Sigma}_{kk}\mathbf{e}_\ell)^2 + O_p(\eta_n^2),$$

$$\begin{aligned}
\sigma_{kkkA} &= \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_{kk}^2][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_{kk}]\} \\
&= 2n^{-1}(2\mathrm{tr}\mathbf{\Sigma}_{kk}^3 + \beta_w \sum_{\ell=1}^{p}\mathbf{e}_{k\ell}^\top \mathbf{\Sigma}_{kk}\mathbf{e}_{k\ell}\mathbf{e}_{k\ell}^\top \mathbf{\Sigma}_{kk}^2 \mathbf{e}_{k\ell}) \\
&\quad +2n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk}[2n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_{k\ell}^\top \mathbf{\Sigma}_{kk}\mathbf{e}_{k\ell})^2] + O_p(\eta_n^2),
\end{aligned}$$

$$\begin{aligned}
\sigma_{0kkA} &= \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n^2][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_{kk}]\} \\
&= 2\big(2n^{-1}\mathrm{tr}\mathbf{\Sigma}^2\mathbf{\Sigma}_{(kk)} + \beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\mathbf{\Sigma}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}_{(kk)}\mathbf{e}_\ell\big) \\
&\quad +2(n^{-1}\mathrm{tr}\mathbf{\Sigma})\big(2n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{kk}\mathbf{e}_\ell)^2\big) + O_p(\eta_n^2),
\end{aligned}$$

$$\begin{aligned}
\sigma_{00kA} &= \sum_{\ell=1}^{n} \mathrm{E}_{\ell-1}\{[(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_n][(\mathrm{E}_\ell - \mathrm{E}_{\ell-1})\mathrm{tr}\mathbf{S}_{kk}^2]\} \\
&= 2\big(2n^{-1}\mathrm{tr}\mathbf{\Sigma}\mathbf{\Sigma}_{(kk)}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\mathbf{\Sigma}\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}_{(kk)}^2\mathbf{e}_\ell\big) \\
&\quad +2(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})\big(2n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{kk}\mathbf{e}_\ell)^2\big) + O_p(\eta_n^2), \\
\sigma_{00kkA} &= \sum_{j=1}^{n} \mathrm{E}_{j-1}[(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_{kk}][(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{S}_n] \\
&= 2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{kk}^2) + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{kk}\mathbf{e}_\ell)^2 + O_p(\eta_n^2).
\end{aligned}$$

Then we have

$$\begin{aligned}
\sigma_1^2 &= \sigma_{00A} + \sum_{k=1}^{K}\sigma_{kkA} - 2\sum_{k=1}^{K}\sigma_{0kA} - 4(n^{-1}\mathrm{tr}\mathbf{\Sigma})\sigma_{000A} + 4(n^{-1}\mathrm{tr}\mathbf{\Sigma})\sum_{k=1}^{K}\sigma_{00kA} \\
&\quad +4\sum_{k=1}^{K}(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})\sigma_{0kkA} - 4\sum_{k=1}^{K}(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})\sigma_{kkkA} + 4(n^{-1}\mathrm{tr}\mathbf{\Sigma})^2\sigma_{0000A} \\
&\quad +4\sum_{k=1}^{K}(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})^2\sigma_{kkkkA} - 8\sum_{k=1}^{K}(n^{-1}\mathrm{tr}\mathbf{\Sigma})(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk})\sigma_{00kkA} \\
&= 4(n^{-1}\mathrm{tr}\mathbf{\Sigma}^2)^2 - 4\sum_{k=1}^{K}(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk}^2)^2 + 4[2n^{-1}\mathrm{tr}\mathbf{A}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{A}\mathbf{e}_\ell)^2] + O_p(K\eta_n^2).
\end{aligned}$$

with $\mathbf{A} = \mathbf{\Sigma}^2 - \mathrm{diag}(\mathbf{\Sigma}_{11}^2, \ldots, \mathbf{\Sigma}_{KK}^2)$. Thus $\sigma_1^{-1}(T_{n1} - \hat{\mu} - \mu_1) \to N(0,1)$.

The proof of Theorem 2 is now complete.

S.2.2 Proof of Part II of Lemma 1

Under H_0, $\mu_1 = 0$ and $\sigma_0^2 = \sigma_1^2 = 4(n^{-1}\sum_{k=1}^K \mathrm{tr}\mathbf{\Sigma}_{kk}^2)^2 - 4\sum_{k=1}^K (n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk}^2)^2$. Then under H_0 and by Theorem 2, we have $\sigma_0^{-1}(T_{n1} - \hat{\mu}) \to N(0,1)$.

The proof of Lemma 1 is complete.

S.2.3 Theorem 1

We have $(n-2)^{-1}[\mathrm{tr}\mathbf{S}_n^2 - (n+2)^{-1}(\mathrm{tr}\mathbf{S}_n)^2] - n^{-1}\mathrm{tr}\mathbf{\Sigma}^2 = o_p(1)$ and $(n-2)^{-1}[\mathrm{tr}\mathbf{S}_{kk}^2 - (n+2)^{-1}(\mathrm{tr}\mathbf{S}_{kk})^2] - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{kk}^2 = o_p(1), \quad k = 1, \ldots, K$. Thus under H_0, we have

$$\hat{\sigma}_0^{-1}(T_{n1} - \hat{\mu}) \to N(0,1) \tag{S.9}$$

where $\hat{\sigma}_0^2 = 4(n-2)^{-2}\{\sum_{k=1}^K [\mathrm{tr}\mathbf{S}_{kk}^2 - (n+2)^{-1}(\mathrm{tr}\mathbf{S}_{kk})^2]\}^2 - 4(n-2)^{-2}\sum_{k=1}^K [\mathrm{tr}\mathbf{S}_{kk}^2 - (n+2)^{-1}(\mathrm{tr}\mathbf{S}_{kk})^2]^2$.

Moreover, Xiao and Wu (2013) presented that $\max_{(\ell_1,\ell_2)\in A_0} n(s_{\ell_1\ell_2})^2\hat{\theta}_{\ell_1\ell_2}^{-1} - 4\log p_0 + \log\log p_0$ converges to a type I extreme value distribution under H_0. Then if the threshold $s^*(n,p)$ is taken to satisfy $s^*(n,p) - 4\log p_0 \to +\infty$, then $P(\max_{(\ell_1,\ell_2)\in A_0} n(s_{\ell_1\ell_2})^2\hat{\theta}_{\ell_1\ell_2}^{-1} > s^*(n,p)) \to 0$ under H_0. That is,

$T_n - T_{n1} = o_p(1)$ under H_0. By (S.9), we have

$$\hat{\sigma}_0^{-1}(T_n - \hat{\mu}) \to N(0,1).$$

The proof of Theorem 1 is complete.

S.3 Proof of Theorem 3

When tr$\mathbf{A}$ tends to infinity, σ_1 also converges and $\mu_1 \to +\infty$. Then

$$\frac{\sigma_0 q_{1-\alpha} - \mu_1}{\sigma_1} \to -\infty.$$

Thus we have $\beta_{T_n}(\mathbf{A}) \to 1$. Moreover, if $P(\max_{(\ell_1,\ell_2)\in A_0} n(s_{\ell_1\ell_2})^2\hat{\theta}^{-1}_{\ell_1\ell_2} > s^*(n,p)) \to 1$, then $T_{n0} \to \infty$ in probability as $n \to \infty$. Then the power function will tend to one.

The proof of Lemma 3 is complete.

Supplementary material 2

This supplementary material consists of three lemmas and the detailed proofs of (S.6)-(S.8). These proofs are conducted under Assumption [A]-[B].

S.2.1. Lemma S.2.1-S.2.4 and their proofs

Let $\mathbf{r}_i = n^{-1/2}\mathbf{w}_i$ and ϵ be a very small positive number.

Lemma S.2.1. *Under Assumptions [A]-[B], we have*

$$n\bar{\mathbf{r}}^\top\mathbf{\Sigma}\bar{\mathbf{r}} = n^{-1}\mathrm{tr}\mathbf{\Sigma} + o_p(n^{-(0.5-\epsilon)}).$$

Proof. We have

$$n\bar{\mathbf{r}}^\top\mathbf{\Sigma}\bar{\mathbf{r}} = 2n^{-1}\sum_{i<j}\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_j + n^{-1}\sum_{i=1}^{n}\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_i.$$

First, we have $\mathrm{E}(n^{-1}\sum_{i<j}\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_j) = 0$ and

$$\begin{aligned}
&\mathrm{E}(n^{-1}\sum_{i<j}\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_j)^2 \\
= \;& (n-1)n^{-1}\mathrm{E}(\mathbf{r}_1^\top\mathbf{\Sigma}\mathbf{r}_2\mathbf{r}_2^\top\mathbf{\Sigma}\mathbf{r}_1) + n^{-2}\sum_{i<j<k<\ell}\mathrm{E}(\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_k^\top\mathbf{\Sigma}\mathbf{r}_\ell) \\
&+2n^{-2}\sum_{i<j<k}\mathrm{E}(\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_k) \\
\leq \;& n^{-2}\mathrm{tr}(\mathbf{\Sigma}^2) = o(n^{-2(0.5-\epsilon)}),
\end{aligned}$$

for any small positive number ϵ. That is,

$$n^{-1}\sum_{i<j}\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_j = o_p(n^{-(0.5-\epsilon)}).$$

Second, we have $\mathrm{E}(n^{-1}\sum_{i=1}^{n}\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_i) = n^{-1}\mathrm{tr}\mathbf{\Sigma}$ and

$$\begin{aligned}
\mathrm{Var}(n^{-1}\sum_{i=1}^{n}\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_i) &= n^{-1}\mathrm{E}[(\mathbf{r}_1^\top\mathbf{\Sigma}\mathbf{r}_1 - n^{-1}\mathrm{tr}\mathbf{\Sigma})^2] \\
&= n^{-2}[2\mathrm{tr}(\mathbf{\Sigma}^2) + \beta_w\sum_{j=1}^{p}(\mathbf{e}_j^\top\mathbf{\Sigma}\mathbf{e}_j)^2] = o(n^{-2(0.5-\epsilon)}),
\end{aligned}$$

where the second equality is from (1.15) of Bai and Silverstein (2004). That is,

$$n^{-1}\sum_{i=1}^{n}\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_i - n^{-1}\mathrm{tr}\mathbf{\Sigma} = o_p(n^{-(0.5-\epsilon)}).$$

Thus we have

$$n\bar{\mathbf{r}}^\top\mathbf{\Sigma}\bar{\mathbf{r}} = n^{-1}\mathrm{tr}\mathbf{\Sigma} + o_p(n^{-(0.5-\epsilon)}).$$

Lemma S.2.2. *Under Assumptions [A]-[B], we have*

$$n\bar{\mathbf{r}}^T\mathbf{\Sigma}\sum_{i=1}^n \mathbf{r}_i\mathbf{r}_i^T\mathbf{\Sigma}\bar{\mathbf{r}} = (n^{-1}\mathrm{tr}\mathbf{\Sigma})^2 + (n-1)n^{-2}\mathrm{tr}(\mathbf{\Sigma}^2) + o_p(1).$$

Proof. We have

$$\begin{aligned} n\bar{\mathbf{r}}^T\mathbf{\Sigma}\sum_{i=1}^n \mathbf{r}_i\mathbf{r}_i^\top\mathbf{\Sigma}\bar{\mathbf{r}} &= n^{-1}\sum_{i,j,\ell\ unequal}\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_\ell + n^{-1}\sum_{i,j\ unequal}\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_i \\ &\quad +2n^{-1}\sum_{i,j\ unequal}\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_i\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j + n^{-1}\sum_{i=1}^n\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_i\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_i. \end{aligned}$$

Step 1. We have

$$\begin{aligned} n^{-1}\sum_{i=1}^n\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_i\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_i &= n^{-1}\sum_{i=1}^n(\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_i - n^{-1}\mathrm{tr}\mathbf{\Sigma})^2 \\ &\quad +2n^{-1}\sum_{i=1}^n(n^{-1}\mathrm{tr}\mathbf{\Sigma})(\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_i - n^{-1}\mathrm{tr}\mathbf{\Sigma}) + (n^{-1}\mathrm{tr}\mathbf{\Sigma})^2. \end{aligned}$$

Because $n^{-1}\sum_{i=1}^n \mathrm{E}[(\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_i - n^{-1}\mathrm{tr}\mathbf{\Sigma})^2] = n^{-2}[2\mathrm{tr}(\mathbf{\Sigma}^2) + \beta_w\sum_{j=1}^p(\mathbf{e}_j^T\mathbf{\Sigma}\mathbf{e}_j)^2] = o(n^{-(1-\epsilon)})$, then we have $n^{-1}\sum_{i=1}^n(\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_i - n^{-1}\mathrm{tr}\mathbf{\Sigma})^2 = o(n^{-(1-\epsilon)})$. Because $n^{-1}\sum_{i=1}^n \mathrm{E}(\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_i - n^{-1}\mathrm{tr}\mathbf{\Sigma}) = 0$ and

$$\mathrm{Var}[n^{-1}\sum_{i=1}^n(\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_i - n^{-1}\mathrm{tr}\mathbf{\Sigma})] = n^{-3}[2\mathrm{tr}(\mathbf{\Sigma}^2) + \beta_w\sum_{j=1}^p(\mathbf{e}_j^T\mathbf{\Sigma}\mathbf{e}_j)^2] = o(n^{-2(1-\epsilon)}),$$

then we have $n^{-1}\sum_{i=1}^n(\mathbf{r}_i^\top\mathbf{\Sigma}\mathbf{r}_i - n^{-1}\mathrm{tr}\mathbf{\Sigma}) = o_p(n^{-(1-\epsilon)})$. Thus,

$$n^{-1}\sum_{i=1}^n\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_i\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_i = (n^{-1}\mathrm{tr}\mathbf{\Sigma})^2 + o_p(n^{-(1-\epsilon)}). \tag{S.2.1}$$

Step 2. We have $n^{-1}\sum_{i,j,\ell\ unequal}\mathrm{E}(\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_\ell)=0$ and

$$
\begin{aligned}
& \mathrm{E}(n^{-1}\sum_{i,j,\ell\ unequal}\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_\ell)^2 \\
= \ & 2n^{-2}\sum_{i,j,\ell,k\ unequal}\mathrm{E}[\mathrm{tr}(\mathbf{\Sigma}^{1/2}\mathbf{r}_i\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_\ell\mathbf{r}_\ell^T\mathbf{\Sigma}\mathbf{r}_k\mathbf{r}_k^T\mathbf{\Sigma}^{1/2})] \\
& +2n^{-2}\sum_{i,j,\ell\ unequal}\mathrm{E}[\mathrm{tr}(\mathbf{\Sigma}^{1/2}\mathbf{r}_i\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_\ell\mathbf{r}_\ell^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}^{1/2})] \\
& +4n^{-2}\sum_{i,j,\ell\ unequal}\mathrm{E}[\mathrm{tr}(\mathbf{\Sigma}^{1/2}\mathbf{r}_i\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_\ell\mathbf{r}_\ell^T\mathbf{\Sigma}\mathbf{r}_i\mathbf{r}_\ell^T\mathbf{\Sigma}^{1/2})] \\
\leq \ & 2n^{-2}\mathrm{tr}(\mathbf{\Sigma}^4)+2n^{-1}\mathrm{E}[(\mathbf{r}_1^T\mathbf{\Sigma}^2\mathbf{r}_1)^2]+4\mathrm{E}[\mathrm{tr}(\mathbf{\Sigma}^{1/2}\mathbf{r}_1\mathbf{r}_1^T\mathbf{\Sigma}^2\mathbf{r}_2\mathbf{r}_2^T\mathbf{\Sigma}\mathbf{r}_1\mathbf{r}_2^T\mathbf{\Sigma}^{1/2})] \\
\leq \ & 2n^{-2}\mathrm{tr}(\mathbf{\Sigma}^4)+2n^{-3}[2\mathrm{tr}(\mathbf{\Sigma}^4)+\beta_w\sum_{j=1}^p(\mathbf{e}_j^T\mathbf{\Sigma}^2\mathbf{e}_j)^2+(\mathrm{tr}\mathbf{\Sigma}^2)^2] \\
& +8\mathrm{E}[(\mathbf{r}_1^T\mathbf{\Sigma}^2\mathbf{r}_2\mathbf{r}_2^T\mathbf{\Sigma}^2\mathbf{r}_1)]+8\mathrm{E}[(\mathbf{r}_1^T\mathbf{\Sigma}\mathbf{r}_2\mathbf{r}_2^T\mathbf{\Sigma}\mathbf{r}_1)^2] \\
= \ & 2n^{-2}\mathrm{tr}(\mathbf{\Sigma}^4)+2n^{-3}[2\mathrm{tr}(\mathbf{\Sigma}^4)+\beta_w\sum_{j=1}^p(\mathbf{e}_j^T\mathbf{\Sigma}^2\mathbf{e}_j)^2+(\mathrm{tr}\mathbf{\Sigma}^2)^2] \\
& +8n^{-2}\mathrm{tr}(\mathbf{\Sigma}^4)+24n^{-2}\mathrm{E}(\mathbf{r}_2^T\mathbf{\Sigma}^2\mathbf{r}_2)^2+8n^{-2}\beta_w\sum_{j=1}^p\mathrm{E}(\mathbf{e}_j^T\mathbf{\Sigma}\mathbf{r}_2)^4 \\
= \ & 10n^{-2}\mathrm{tr}(\mathbf{\Sigma}^4)+(2n^{-3}+24n^{-4})[2\mathrm{tr}(\mathbf{\Sigma}^4)+\beta_w\sum_{j=1}^p(\mathbf{e}_j^T\mathbf{\Sigma}^2\mathbf{e}_j)^2+(\mathrm{tr}\mathbf{\Sigma}^2)^2] \\
& +24n^{-4}\beta_w\sum_{j=1}^p(\mathbf{e}_j^T\mathbf{\Sigma}^2\mathbf{e}_j)^2+8n^{-4}\beta_w^2\sum_{j=1}^p\sum_{\ell=1}^p(\mathbf{e}_j^T\mathbf{\Sigma}\mathbf{e}_\ell)^4 \\
= \ & o_p(n^{-2(0.5-\epsilon)}).
\end{aligned}
$$

Then we have

$$
n^{-1}\sum_{i,j,\ell\ unequal}(\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_\ell)=o_p(n^{-(0.5-\epsilon)}). \tag{S.2.2}
$$

Step 3. We have $n^{-1}\sum_{i,j\ unequal}\mathrm{E}\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_i=(n-1)n^{-2}\mathrm{tr}(\mathbf{\Sigma}^2)$ and

$$
\begin{aligned}
& n^{-2}\mathrm{E}(\sum_{i,j\ unequal}\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_i)^2 \\
= \ & 2n^{-2}\sum_{i,j\ unequal}(\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_i)^2+n^{-2}\sum_{i,j,k,\ell\ unequal}(\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_i)(\mathbf{r}_\ell^T\mathbf{\Sigma}\mathbf{r}_k\mathbf{r}_k^T\mathbf{\Sigma}\mathbf{r}_\ell) \\
& +4n^{-2}\sum_{i,j,\ell\ unequal}(\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_i)(\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_\ell\mathbf{r}_\ell^T\mathbf{\Sigma}\mathbf{r}_i) \\
= \ & 6n^{-3}(n-1)\mathrm{E}[(\mathbf{r}_1^T\mathbf{\Sigma}^2\mathbf{r}_1)^2]+2n^{-3}(n-1)\beta_w\sum_{j=1}^p\mathrm{E}(\mathbf{e}_j^T\mathbf{\Sigma}\mathbf{r}_1)^4 \\
& +n^{-5}(n-1)(n-2)(n-3)[\mathrm{tr}(\mathbf{\Sigma}^2)]^2 \\
& +4n^{-5}(n-1)(n-2)[2\mathrm{tr}(\mathbf{\Sigma}^4)+\beta_w\sum_{j=1}^p(\mathbf{e}_j^T\mathbf{\Sigma}^2\mathbf{e}_j)^2+(\mathrm{tr}\mathbf{\Sigma}^2)^2] \\
\leq \ & 6n^{-5}(n-1)\beta_w\sum_{j=1}^p(\mathbf{e}_j^T\mathbf{\Sigma}^2\mathbf{e}_j)^2+2n^{-5}(n-1)\beta_w^2\sum_{j=1}^p\sum_{\ell=1}^p(\mathbf{e}_j^T\mathbf{\Sigma}\mathbf{e}_\ell)^4 \\
& +n^{-5}(n-1)(n-2)(n-3)[\mathrm{tr}(\mathbf{\Sigma}^2)]^2 \\
& +2n^{-5}(n-1)(2n-1)[2\mathrm{tr}(\mathbf{\Sigma}^4)+\beta_w\sum_{j=1}^p(\mathbf{e}_j^T\mathbf{\Sigma}^2\mathbf{e}_j)^2+(\mathrm{tr}\mathbf{\Sigma}^2)^2].
\end{aligned}
$$

Then we have

$$
\begin{aligned}
& \mathrm{Var}(n^{-1}\sum_{i,j\ unequal}\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_i) \\
= \ & n^{-2}\mathrm{E}\Big[\Big(\sum_{i,j\ unequal}\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_i\Big)^2\Big]-\Big(n^{-1}\sum_{i,j\ unequal}\mathrm{E}\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_i\Big)^2 \\
= \ & o(n^{-2(0.5-\epsilon)}).
\end{aligned}
$$

That is,

$$
n^{-1}\sum_{i,j\ unequal}\mathbf{r}_i^T\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^T\mathbf{\Sigma}\mathbf{r}_i-(n-1)n^{-2}\mathrm{tr}(\mathbf{\Sigma}^2)=o_p(n^{-(0.5-\epsilon)}). \qquad \text{(S.2.3)}
$$

Step 4. By (1.8) of Bai and Silverstein (2004), there exists $\eta_n \downarrow 0$ satisfying $n^{1/4}\eta_n \to \infty$ and $\eta_n^{-4}\mathrm{E}[w_{11}^4\delta(|w_{11}| \geq \eta_n\sqrt{n})] \to 0$. Then let $\hat{\mathbf{r}}_i$ be the truncated version of $\mathbf{r}_i$, that is, $\hat{\mathbf{r}}_i^T = n^{-1/2}\hat{\mathbf{w}}_i$ with $\hat{\mathbf{w}}_i = (\hat{w}_{1i}, \ldots, \hat{w}_{pi})^\top$ and $\hat{w}_{\ell i} = w_{\ell i}\delta_{\{|w_{\ell i}| \leq \sqrt{n}\eta_n\}}$. Then we have $\mathrm{E}\hat{w}_{11} \to 0$, $\mathrm{E}\hat{w}_{11}^2 \to 1$ and $\mathrm{Var}(\hat{w}_{11}) \to 1$ as $n \to \infty$. Let $\hat{\boldsymbol{\mu}} = n^{-1/2}(\mathrm{E}\hat{w}_{11})\mathbf{1}_p$ where $\mathbf{1}_p$ is the p-dimensional vector with all entries being ones. Because $\mathrm{E}w_{11} = 0$, then we have

$$|\mathrm{E}\hat{w}_{11}| = |\mathrm{E}[w_{11}\delta(|w_{11}| > \eta_n\sqrt{n})]| \leq \eta_n^{-3}n^{-3/2}\mathrm{E}[w_{11}^4\delta(|w_{11}| > \eta_n\sqrt{n})] = o(n^{-3/2}).$$

That is

$$\hat{\boldsymbol{\mu}}^\top\hat{\boldsymbol{\mu}} = n^{-1}(\mathrm{E}\hat{w}_{11})^2\mathbf{1}_p^\top\mathbf{1}_p \leq o(n^{-1/2}).$$

Because

$$\begin{aligned}
& P(n^{-1}\sum_{i,j\ unequal}\mathbf{r}_i^T\boldsymbol{\Sigma}\mathbf{r}_i\mathbf{r}_i^T\boldsymbol{\Sigma}\mathbf{r}_j \neq n^{-1}\sum_{i,j\ unequal}\hat{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\mathbf{r}}_i\hat{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\mathbf{r}}_j) \\
\leq\ & P(\text{for some } \ell, \mathrm{i}, \hat{w}_{\ell i} \neq w_{\ell i}) \\
\leq\ & \sum_{\ell=1}^{p}\sum_{i=1}^{n}P(|w_{\ell i}| \geq \eta_n\sqrt{n}) \\
\leq\ & (\eta_n\sqrt{n})^{-4}np\mathrm{E}[w_{11}^4\delta(|w_{11}| \geq \eta_n\sqrt{n})] \\
=\ & (p/n)\eta_n^{-4}\mathrm{E}[w_{11}^4\delta(|w_{11}| \geq \eta_n\sqrt{n})] \to 0
\end{aligned}$$

where the third inequality is from the Chebyshev inequality and the last equality is from (1.8) of Bai and Silverstein (2004), then we have

$$n^{-1}\sum_{i,j\ unequal}\mathbf{r}_i^T\boldsymbol{\Sigma}\mathbf{r}_i\mathbf{r}_i^T\boldsymbol{\Sigma}\mathbf{r}_j = n^{-1}\sum_{i,j\ unequal}\hat{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\mathbf{r}}_i\hat{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\mathbf{r}}_j + o_p(1). \qquad \text{(S.2.4)}$$

Let $\tilde{\mathbf{r}}_i = (\hat{\mathbf{r}}_i - \hat{\boldsymbol{\mu}})/\sqrt{\mathrm{Var}(\hat{w}_{11})}$, then

$$\begin{aligned}
\hat{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\mathbf{r}}_i &= \mathrm{Var}(\hat{w}_{11})\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i + 2\sqrt{\mathrm{Var}(\hat{w}_{11})}\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\mu}}^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}} \\
&= \mathrm{Var}(\hat{w}_{11})(\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) + n^{-1}\mathrm{Var}(\hat{w}_{11})\mathrm{tr}\boldsymbol{\Sigma} \\
&\quad +2\sqrt{\mathrm{Var}(\hat{w}_{11})}\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}} + o(n^{-1/2}),
\end{aligned}$$

and $\hat{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\mathbf{r}}_j = \mathrm{Var}(\hat{w}_{11})\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j + \sqrt{\mathrm{Var}(\hat{w}_{11})}\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}} + \sqrt{\mathrm{Var}(\hat{w}_{11})}\tilde{\mathbf{r}}_j^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}} + o(n^{-1/2})$.

Because

$$n^{-1}\sum_{i,j\ unequal}\mathrm{E}\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j = 0, \quad \mathrm{E}(n^{-1}\sum_{i,j\ unequal}\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j)^2 \le n^{-2}\mathrm{tr}(\boldsymbol{\Sigma}^2) = o(n^{-2(0.5-\epsilon)}),$$

we have $n^{-1}\sum_{i,j\ unequal}\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j = o_p(n^{-(0.5-\epsilon)})$. Because

$$n^{-1}\sum_{i,j\ unequal}\mathrm{E}\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}} = 0, \quad \mathrm{E}(n^{-1}\sum_{i,j\ unequal}\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}})^2 \le \hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}^2\hat{\boldsymbol{\mu}} = o(n^{-1/2}),$$

we have $n^{-1}\sum_{i,j\ unequal}\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}} = o_p(n^{-1/2})$. Because $n^{-1}\sum_{i,j\ unequal}\mathrm{E}(\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}})^2 = \hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}^2\hat{\boldsymbol{\mu}} = o(n^{-1/2})$, we have $n^{-1}\sum_{i,j\ unequal}(\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}})^2 = o_p(n^{-1/2})$. Because

$$n^{-1}\sum_{i,j\ unequal}\mathrm{E}(\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j) = 0, \mathrm{E}(n^{-1}\sum_{i,j\ unequal}\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j)^2 \le (\hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}^2\hat{\boldsymbol{\mu}})^2 = o(n^{-1}),$$

we have $n^{-1}\sum_{i,j\ unequal}(\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j) = o_p(n^{-1/2})$. Because

$$\begin{aligned}
&\mathrm{E}(n^{-1}\sum_{i,j\ unequal}\tilde{\mathbf{r}}_j^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i\tilde{\mathbf{r}}_i^\top\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}})^2 \\
= \ & n^{-2}\sum_{i,j,\ell\ unequal}\mathrm{E}\tilde{\mathbf{r}}_j^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i\tilde{\mathbf{r}}_i^\top\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_\ell\tilde{\mathbf{r}}_\ell^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j + n^{-2}\sum_{i,j,\ unequal}\mathrm{E}\tilde{\mathbf{r}}_j^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i\tilde{\mathbf{r}}_i^\top\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i\tilde{\mathbf{r}}_i^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j \\
& +n^{-2}\sum_{i,j,\ unequal}\mathrm{E}\tilde{\mathbf{r}}_j^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i\tilde{\mathbf{r}}_i^\top\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j\tilde{\mathbf{r}}_j^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i \\
\le \ & n^{-2}\hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}^4\hat{\boldsymbol{\mu}} + \mathrm{E}\tilde{\mathbf{r}}_1^T\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_1\tilde{\mathbf{r}}_1^\top\boldsymbol{\Sigma}^2\tilde{\mathbf{r}}_1 + \mathrm{E}(\tilde{\mathbf{r}}_1^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_2)^2\tilde{\mathbf{r}}_1^\top\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_2 = o(n^{-1/2}),
\end{aligned}$$

we have $n^{-1}\sum_{i,j\ unequal}\tilde{\mathbf{r}}_j^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i\tilde{\mathbf{r}}_i^\top\boldsymbol{\Sigma}\hat{\boldsymbol{\mu}} = o_p(n^{-1/2})$. Because

$$\begin{aligned}
& \mathrm{E}(n^{-1}\sum_{i,j\ unequal}(\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j)^2 \\
= \ & n^{-2}\sum_{i,j,\ell\ unequal}\mathrm{E}(\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j\tilde{\mathbf{r}}_j^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_\ell(\tilde{\mathbf{r}}_\ell^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_\ell - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) \\
& +n^{-2}\sum_{i,j\ unequal}\mathrm{E}(\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j\tilde{\mathbf{r}}_j^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i(\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) \\
& +n^{-2}\sum_{i,j\ unequal}\mathrm{E}(\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j\tilde{\mathbf{r}}_i^\top\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j(\tilde{\mathbf{r}}_j^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) \\
\leq \ & [\mathrm{E}(\tilde{\mathbf{r}}_1^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2]^2 + \mathrm{E}(\tilde{\mathbf{r}}_1^T\boldsymbol{\Sigma}^2\tilde{\mathbf{r}}_2)^2 + n^{-1}\mathrm{E}(\tilde{\mathbf{r}}_1^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2\tilde{\mathbf{r}}_1^T\boldsymbol{\Sigma}^2\tilde{\mathbf{r}}_1 \\
& +[\mathrm{E}(\tilde{\mathbf{r}}_1^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2]^2 + \mathrm{E}(\tilde{\mathbf{r}}_1^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_2)^4 = o(n^{-2(0.5-\epsilon)}),
\end{aligned}$$

by (1.15) of Bai and Silverstein (2004) and (9.9.6) of Bai and Silverstein (2010), we have

$$n^{-1}\sum_{i,j\ unequal}(\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_i - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})\tilde{\mathbf{r}}_i^T\boldsymbol{\Sigma}\tilde{\mathbf{r}}_j = o_p(n^{-(0.5-\epsilon)}).$$

Thus we have

$$n^{-1}\sum_{i,j\ unequal}\hat{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\mathbf{r}}_i\hat{\mathbf{r}}_i^T\boldsymbol{\Sigma}\hat{\mathbf{r}}_j = o(n^{-(0.5-\epsilon)}). \tag{S.2.5}$$

By (S.2.4) and (S.2.5), we have

$$n^{-1}\sum_{i,j\ unequal}\mathbf{r}_i^T\boldsymbol{\Sigma}\mathbf{r}_i\mathbf{r}_i^T\boldsymbol{\Sigma}\mathbf{r}_j = o_p(1). \tag{S.2.6}$$

By (S.2.1), (S.2.2), (S.2.3) and (S.2.6), we have

$$n\bar{\mathbf{r}}^T\boldsymbol{\Sigma}\sum_{i=1}^n\mathbf{r}_i\mathbf{r}_i^T\boldsymbol{\Sigma}\bar{\mathbf{r}} = (n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2 + (n-1)n^{-2}\mathrm{tr}(\boldsymbol{\Sigma}^2) + o_p(1). \tag{S.2.7}$$

That is, the proof of Lemma S.2.2 is complete.

Lemma S.2.3. *Under Assumptions [A]-[B] with* $|w_{\ell i}| \leq \sqrt{n}\eta_n$, $\mathrm{E}w_{\ell i} = 0$, $\mathrm{E}(w_{\ell i}^2) = 1$ *and* $\mathrm{E}(w_{\ell i}^4) < \infty$ *with* $\eta_n \downarrow 0$ *and* $n^{1/4}\eta_n \to \infty$, *we have*

$$\sum_{j=1}^{n} \mathrm{E}([(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_n]^2 \delta_{\{|(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_n| \geq \epsilon\}}) = O(\eta_n^4),$$

$$\sum_{j=1}^{n} \mathrm{E}([(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_n^2]^2 \delta_{\{|(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_n^2| \geq \epsilon\}}) = O(\eta_n^4),$$

$$\sum_{j=1}^{n} \mathrm{E}([(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_{kk}]^2 \delta_{\{|(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_{kk}| \geq \epsilon\}}) = O(\eta_n^4),$$

and

$$\sum_{j=1}^{n} \mathrm{E}([(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_{kk}^2]^2 \delta_{\{|(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_{kk}^2| \geq \epsilon\}}) = O(\eta_n^4).$$

Proof. We have $\mathrm{tr}\mathbf{B}_n = \sum_{i=1}^n \mathbf{r}_i^T \boldsymbol{\Sigma} \mathbf{r}_i$ and $\mathrm{E}(\mathbf{r}_i^T \boldsymbol{\Sigma} \mathbf{r}_i - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^4 \leq Cn^{-1}\eta_n^4 \|\boldsymbol{\Sigma}\|^4 = O(\eta_n^4 n^{-1})$ by (9.9.6) of Bai and Silverstein (2010) where C is a constant independent of n and p. Then we have

$$\sum_{j=1}^{n} \mathrm{E}([(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_n]^2 \delta_{\{|(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_n| \geq \epsilon\}}) \leq n\mathrm{E}(\mathbf{r}_i^T \boldsymbol{\Sigma} \mathbf{r}_i - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^4/\epsilon^2 = O(\eta_n^4).$$

Similarly, we have

$$\sum_{j=1}^{n} \mathrm{E}([(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_{kk}^2]^2 \delta_{\{|(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_{kk}^2| \geq \epsilon\}}) = O(\eta_n^4).$$

$\mathrm{E}_j(\mathrm{tr}\mathbf{B}_n^2) - \mathrm{E}_{j-1}(\mathrm{tr}\mathbf{B}_n^2)$ can be expressed by

$$\begin{aligned}
\mathrm{E}_j\mathrm{tr}(\mathbf{B}_n^2) - \mathrm{E}_{j-1}\mathrm{tr}(\mathbf{B}_n^2) &= 2(n-j)n^{-1}[\mathbf{r}_j^\top \boldsymbol{\Sigma}^2 \mathbf{r}_j - n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}^2)] \\
&\quad + [\mathbf{r}_j^\top \boldsymbol{\Sigma} \mathbf{r}_j \mathbf{r}_j^\top \boldsymbol{\Sigma} \mathbf{r}_j - \mathrm{E}(\mathbf{r}_j^\top \boldsymbol{\Sigma} \mathbf{r}_j \mathbf{r}_j^\top \boldsymbol{\Sigma} \mathbf{r}_j)] \\
&\quad + 2\sum_{k \leq j-1} [\mathbf{r}_j^\top \boldsymbol{\Sigma} \mathbf{r}_k \mathbf{r}_k^\top \boldsymbol{\Sigma} \mathbf{r}_j - n^{-1}(\mathbf{r}_j^\top \boldsymbol{\Sigma}^2 \mathbf{r}_j)].
\end{aligned}$$

We have

$$\sum_{j=1}^{n}(n-j)^4 n^{-4}\mathrm{E}\left[(\mathbf{r}_1^\top\boldsymbol{\Sigma}^2\mathbf{r}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2)^4\right] \le C\sum_{j=1}^{n}(n-j)^4 n^{-5}\eta_n^4, \quad \text{(S.2.8)}$$

which is from Lemma 9.1 of Bai and Silverstein (2004) and C is a constant not dependent on p or n. Moreover, we have

$$\begin{aligned}
&\sum_{j=1}^{n}\mathrm{E}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j - \mathrm{E}\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j)^4]\\
=\;& n\mathrm{E}[(\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1 - \mathrm{E}\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1)^4]\\
\le\;& Cn\mathrm{E}[(\mathbf{r}_1^T\boldsymbol{\Sigma}\mathbf{r}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^8] + nO(n^{-4}) + Cn(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^4\mathrm{E}[(\mathbf{r}_1^T\boldsymbol{\Sigma}\mathbf{r}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^4]\\
\le\;& O(\eta_n^{12}) + O(n^{-3}) + O(n^{-4}) \qquad \text{(S.2.9)}
\end{aligned}$$

where $(\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2 = n^{-2}[2\mathrm{tr}(\boldsymbol{\Sigma}^2) + \beta_w\sum_{j=1}^{p}(\mathbf{e}_j^T\boldsymbol{\Sigma}\mathbf{e}_j)^2]$, the last inequality is from (9.9.6) of Bai and Silverstein (2010) and

$$\begin{aligned}
&(\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1)^2 - \mathrm{E}[(\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1)^2]\\
=\;& (\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2 - \mathrm{E}[(\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2] + 2(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})(\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})\\
=\;& (\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2 + 2(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})(\mathbf{r}_1^\top\boldsymbol{\Sigma}\mathbf{r}_1 - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}) + O(n^{-1}).
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
&\sum_{j=1}^{n} \mathrm{E}\{[\sum_{k\le j-1}(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\mathbf{\Sigma}\mathbf{r}_j]^4\}\\
=\ &\sum_{j=1}^{n} \mathrm{E}\{[\sum_{k\le j-1}(\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\mathbf{\Sigma}\mathbf{r}_j-n^{-1}\mathbf{r}_k^\top\mathbf{\Sigma}^2\mathbf{r}_k)]^4\}\\
\le\ & C\eta_n^4 n^{-1}\sum_{j=1}^{n}\mathrm{E}(\|\sum_{k\le j-1}\mathbf{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\mathbf{\Sigma}\|^4)\\
\le\ & C\eta_n^4\mathrm{E}(\|\sum_{k=1}^{n}\mathbf{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\mathbf{\Sigma}\|^4)\le C\eta_n^4\|\mathbf{\Sigma}\|^8\mathrm{E}(\|\sum_{k=1}^{n}\mathbf{r}_k\mathbf{r}_k^\top\|^4)\\
\le\ & 2C\eta_n^4\|\mathbf{\Sigma}\|^2(1+\sqrt{y_n})^8=O(\eta_n^4) \qquad \text{(S.2.10)}
\end{aligned}$$

where the second inequality is from (9.9.6) of Bai and Silverstein (2010), $\|\sum_{k\le j-1}\mathbf{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\mathbf{\Sigma}\|$ is the spectral norm of the random matrix $\sum_{k\le j-1}\mathbf{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\mathbf{\Sigma}$, that is, the maximum eigenvalue of $\sum_{k\le j-1}\mathbf{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\mathbf{\Sigma}$ and the last inequality is from (4.2) of Yin, Bai and Krishnaiah (1988). From (S.2.8)-(S.2.9)-(S.2.10), we have

$$\begin{aligned}
&\sum_{j=1}^{n}\mathrm{E}\{[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_n^2]^2\delta_{\{|(\mathrm{E}_j-\mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_n^2|\ge\epsilon\}}\}\\
\le\ & C\sum_{j=1}^{n}\mathrm{E}[2(n-j)n^{-1}(\mathbf{r}_j^\top\mathbf{\Sigma}^2\mathbf{r}_j-n^{-1}\mathrm{tr}\mathbf{\Sigma}^2)]^4+C\sum_{j=1}^{n}\mathrm{E}[\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_j-\mathrm{E}(\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_j)]^4\\
&+C\sum_{j=1}^{n}\mathrm{E}\{[\sum_{k\le j-1}(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\mathbf{\Sigma}\mathbf{r}_j]^4\}=O(\eta_n^4)+O(n^{-3})+O(n^{-4})=O(\eta_n^4).
\end{aligned}$$

Similarly, we have

$$\sum_{j=1}^{n}\mathrm{E}\{[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_{kk}^2]^2\delta_{\{|(\mathrm{E}_j-\mathrm{E}_{j-1})\mathrm{tr}\mathbf{B}_{kk}^2|\ge\epsilon\}}\}=O(\eta_n^4)+O(n^{-3})+O(n^{-4})=O(\eta_n^4).$$

The proof of Lemma S.2.3 is complete.

Lemma S.2.4. *Under Assumptions [A]-[B] with* $|w_{\ell i}| \leq \sqrt{n}\eta_n$, $\mathrm{E}w_{\ell i} = 0$, $\mathrm{E}(w_{\ell i}^2) = 1$ *and* $\mathrm{E}(w_{\ell i}^4) < \infty$ *with* $\eta_n \downarrow 0$ *and* $n^{1/4}\eta_n \to \infty$, *we have*

$$
\begin{aligned}
(S.6) &= 2(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^3) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell 1}\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell 1}] \\
&\quad +2[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^4) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell 1})^2] + O_p(\eta_n^2), \\
(S.7) &= 4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^2) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell 1})^2] \\
&\quad +4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^3) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell 1}\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell 1}] + O_p(\eta_n^2), \\
(S.8) &= 2(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^3) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell 1})(\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell 1})] \\
&\quad +2[2n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^4) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell 1})^2] + 4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}^2)^2 + O_p(\eta_n^2),
\end{aligned}
$$

where $O_p(\eta_n^2)$ *is uniform for* $k = 1, ..., K$.

Proof. We have

$$
\begin{aligned}
(S.6) &= (S.2.11) + (S.2.12) + (S.2.13), \\
(S.7) &= (S.2.14) + (S.2.15) + (S.2.16), \\
(S.8) &= (S.2.17) + (S.2.18) + (S.2.19),
\end{aligned}
$$

where

$$4\sum_{j=1}^{n}\frac{(n-j)^2}{n^2}\mathrm{E}_{j-1}\{[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}^2\mathbf{r}_j][(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1}]\} \tag{S.2.11}$$

$$2\sum_{j=1}^{n}\frac{(n-j)}{n}\mathrm{E}_{j-1}\{[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}^2\mathbf{r}_j][(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}]\} \tag{S.2.12}$$

$$4\sum_{j=1}^{n}\frac{(n-j)}{n}\sum_{k\le j-1}\mathrm{E}_{j-1}\{[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}^2\mathbf{r}_j][(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}]\} \tag{S.2.13}$$

$$2\sum_{j=1}^{n}\frac{(n-j)}{n}\mathrm{E}_{j-1}\{[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j][(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1}]\} \tag{S.2.14}$$

$$\sum_{j=1}^{n}\mathrm{E}_{j-1}\{[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j][(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}]\} \tag{S.2.15}$$

$$2\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}_{j-1}\{[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j][(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}]\} \tag{S.2.16}$$

$$4\sum_{j=1}^{n}\frac{(n-j)}{n}\sum_{k\le j-1}\mathrm{E}_{j-1}\{[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\boldsymbol{\Sigma}\mathbf{r}_j][(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1}]\} \tag{S.2.17}$$

$$2\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}_{j-1}\{[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\boldsymbol{\Sigma}\mathbf{r}_j][(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}]\} \tag{S.2.18}$$

$$4\sum_{j=1}^{n}\sum_{k\le j-1}\sum_{\ell\le j-1}\mathrm{E}_{j-1}\{[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\boldsymbol{\Sigma}\mathbf{r}_j][(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{\ell1}\mathbf{r}_{\ell1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}]\}. \tag{S.2.19}$$

Detailed proof of (S.2.11):

$$\begin{aligned}(S.2.11) &= 4\sum_{j=1}^{n}\frac{(n-j)^2}{n^2}\mathrm{E}_{j-1}[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}^2\mathbf{r}_j(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1}]\\ &= \frac{4n(1+O(n^{-1}))}{3}\mathrm{E}\{[\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1}-n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{11}^2)]^2\}\\ &= \frac{4}{3n}[2\mathrm{tr}(\boldsymbol{\Sigma}_{11}^4)+\beta_w\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell1})^2]+O(n^{-1})\end{aligned}$$

where the last equality is from (1.15) of Bai and Silverstein (2004).

Detailed proof of (S.2.12):

$$
\begin{aligned}
&(S.2.12)\\
=\ & 2\sum_{j=1}^{n}\frac{(n-j)}{n}\mathrm{E}_{j-1}[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}^2\mathbf{r}_j(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}]\\
=\ & n(1+O(n^{-1}))\mathrm{E}[\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1}-\mathrm{E}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1}][(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1})^2-\mathrm{E}(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1})^2]\\
=\ & n(1+O(n^{-1}))\mathrm{E}[(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1}-\mathrm{E}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1})(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})^2]\\
&+2n(1+O(n^{-1}))(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})\mathrm{E}[(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1}-\mathrm{E}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1})(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})] \quad \text{(S.2.20)}
\end{aligned}
$$

where the last equality is from the following equality

$$
\begin{aligned}
&(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1})^2-\mathrm{E}[(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1})^2]\\
=\ & (\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})^2-\mathrm{E}[(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})^2]\\
&+2(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}).
\end{aligned}
$$

By (9.9.6) of Bai and Silverstein (2010), we have

$$
n\mathrm{E}[(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1}-\mathrm{E}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1})(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})^2]\le C_0\|\boldsymbol{\Sigma}_{11}\|^2\|\boldsymbol{\Sigma}_{11}^2\|\eta_n^2=O(\eta_n^2) \quad \text{(S.2.21)}
$$

where C_0 is a constant. By (1.15) of Bai and Silverstein (2004), we have

$$
\begin{aligned}
&\mathrm{E}[(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1}-\mathrm{E}\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{j1})(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})]\\
=\ & n^{-2}[2\mathrm{tr}(\boldsymbol{\Sigma}_{11}^3)+\beta_w\sum_{\ell=1}^{p}\mathbf{e}_{\ell1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell1}\mathbf{e}_{\ell1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell1}]=O(n^{-1}). \quad \text{(S.2.22)}
\end{aligned}
$$

By (S.2.20)-(S.2.21)-(S.2.22), we have

$$(S.2.12) = 2(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^3)+\beta_w n^{-1}\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell 1}^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_{\ell 1}\mathbf{e}_{\ell 1}^\top\mathbf{\Sigma}_{11}\mathbf{e}_{\ell 1}]+O(\eta_n^2).$$

Moreover, the detailed proof of (S.2.14) is similar to the proof of (S.2.12).

Detailed proof of (S.2.13):

$$\begin{aligned}
&(S.2.13)\\
=\ & 4\sum_{j=1}^n\sum_{k\le j-1}\frac{(n-j)}{n}\mathrm{E}_{j-1}[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\mathbf{\Sigma}^2\mathbf{r}_j(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{j1}]\\
=\ & 4\sum_{j=1}^n\sum_{k\le j-1}\frac{(n-j)}{n}\mathrm{E}_{j-1}\{[\mathbf{r}_j^\top\mathbf{\Sigma}^2\mathbf{r}_j-n^{-1}\mathrm{tr}(\mathbf{\Sigma}^2)](\mathbf{r}_{j1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^2\mathbf{r}_{k1})\}\\
=\ & 4\sum_{j=1}^n\sum_{k\le j-1}\frac{(n-j)}{n}\mathrm{E}_{j-1}\{[\mathbf{r}_{j1}^\top\mathbf{\Sigma}_{11}^2\mathbf{r}_{j1}-n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^2)](\mathbf{r}_{j1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^2\mathbf{r}_{k1})\}\\
=\ & 4\sum_{j=1}^n\sum_{k\le j-1}\frac{(n-j)}{n^3}(2\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^4\mathbf{r}_{k1}+\beta_w\sum_{\ell=1}^{p_1}\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}\mathbf{e}_\ell) \qquad (S.2.23)
\end{aligned}$$

where the last equality is from (1.15) of Bai and Silverstein (2004). It is clear that $\sum_{j=1}^n\sum_{k\le j-1}(n-j)n^{-3}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^4\mathbf{r}_{k1}$ is the weighted sum of independent random variables $\{\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^4\mathbf{r}_{k1}, k=1,...,n\}$ with $\mathrm{E}\left[\sum_{j=1}^n\sum_{k\le j-1}\frac{n-j}{n^3}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^4\mathbf{r}_{k1}\right]=(3n)^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^4+O(n^{-1})$ and $\mathrm{var}\left[\sum_{j=1}^n\sum_{k\le j-1}\frac{n-j}{n^3}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^4\mathbf{r}_{k1}\right]=O(n^{-1})$. That is

$$\sum_{j=1}^n\sum_{k\le j-1}(n-j)n^{-3}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^4\mathbf{r}_{k1}=(3n)^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^4+O_p(n^{-1/2}). \qquad (S.2.24)$$

It is clear that $\sum_{j=1}^n(n-j)n^{-3}\sum_{k\le j-1}\sum_{\ell=1}^p\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}\mathbf{e}_\ell$ is the weighted sum of the independent random variables $\{\sum_{\ell=1}^p\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}\mathbf{e}_\ell, k=$

$1, ..., n\}$ with

$$\mathrm{E}\Big[\sum_{j=1}^{n}(n-j)n^{-3}\sum_{k\le j-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}\mathbf{e}_\ell\Big] = (6n)^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_\ell)^2 + O(n^{-1})$$

$$\mathrm{var}\Big[\sum_{j=1}^{n}(n-j)n^{-3}\sum_{k\le j-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}\mathbf{e}_\ell\Big] = O(n^{-1}).$$

That is,

$$\sum_{j=1}^{n}(n-j)n^{-3}\sum_{k\le j-1}\sum_{\ell=1}^{p}\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}\mathbf{e}_\ell = (6n)^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell1}^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_{\ell1})^2 + O_p(n^{-1/2}). \tag{S.2.25}$$

By (S.2.23)-(S.2.24)-(S.2.25), we have

$$(S.2.13) = \frac{2}{3n}[2\mathrm{tr}\mathbf{\Sigma}_{11}^4 + \beta_w\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_\ell)^2] + O_p(n^{-1/2}).$$

Moreover, the detailed proof of (S.2.17) is similar to the proof of (S.2.13).

Detailed proof of (S.2.15):

$$\begin{aligned}
(S.2.15) &= \sum_{j=1}^{n}\mathrm{E}_{j-1}[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_j(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{j1}\mathbf{r}_{j1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{j1}] \\
&= n\mathrm{E}[(\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\mathbf{\Sigma})^2(\mathbf{r}_{j1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})^2] \\
&\quad -n\mathrm{E}[(\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\mathbf{\Sigma})^2]\mathrm{E}[(\mathbf{r}_{j1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})^2] \\
&\quad +2n(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})\mathrm{E}[(\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\mathbf{\Sigma})^2(\mathbf{r}_{j1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})] \\
&\quad +2n(n^{-1}\mathrm{tr}\mathbf{\Sigma})\mathrm{E}[(\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\mathbf{\Sigma})(\mathbf{r}_{j1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})^2] \\
&\quad +4n(n^{-1}\mathrm{tr}\mathbf{\Sigma})(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})\mathrm{E}[(\mathbf{r}_j^\top\mathbf{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\mathbf{\Sigma})(\mathbf{r}_{j1}^\top\mathbf{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})] \qquad (S.2.26)
\end{aligned}$$

By (1.15) of Bai and Silverstein (2004), we have

$$\begin{aligned} & n\mathrm{E}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2]\mathrm{E}[(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})^2] \\ = \; & n^{-1}[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}\mathbf{e}_\ell)^2][2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell1}^\top\boldsymbol{\Sigma}\mathbf{e}_{\ell1})^2] = O(n^{-1}), \end{aligned} \tag{S.2.27}$$

and

$$\mathrm{E}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})] = n^{-2}[2\mathrm{tr}\boldsymbol{\Sigma}_{11}^2 + \beta_w\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell1})^2]. \tag{S.2.28}$$

By (9.9.6) of Bai and Silverstein (2010), we have

$$\begin{aligned} & n\mathrm{E}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})^2] \le \eta_n^4\cdot C_0\|\boldsymbol{\Sigma}\|^2\|\boldsymbol{\Sigma}_{11}\|^2 = O(\eta_n^4), \\ & n\mathrm{E}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})] \le \eta_n^2\cdot C_0\|\boldsymbol{\Sigma}\|^2\|\boldsymbol{\Sigma}_{11}\| = O(\eta_n^2), \\ & n\mathrm{E}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})^2] \le \eta_n^2\cdot C_0\|\boldsymbol{\Sigma}\|^2\|\boldsymbol{\Sigma}_{11}\| = O(\eta_n^2). \end{aligned} \tag{S.2.29}$$

By (1.15) of Bai and Silverstein (2004), we have

$$n\mathrm{E}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j - n^{-1}\mathrm{tr}\boldsymbol{\Sigma})(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1} - n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})] = 2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell1})^2. \tag{S.2.30}$$

Then by (S.2.26)-(S.2.27)-(S.2.28)-(S.2.29)-(S.2.30), we have

$$(S.2.15) = 4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11})(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})[2n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}^2 + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell1}^\top\boldsymbol{\Sigma}_{11}\mathbf{e}_{\ell1})^2] + O(\eta_n^2).$$

Detailed proof of (S.2.16):

$$
\begin{aligned}
&(S.2.16)\\
=\ & 2\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}_{j-1}[(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}]\\
=\ & 2\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}_{j-1}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j-n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{k1})]\\
&+4(n^{-1}\mathrm{tr}\boldsymbol{\Sigma})\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}_{j-1}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j-n^{-1}\mathrm{tr}\boldsymbol{\Sigma})(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{k1})] \qquad \text{(S.2.31)}
\end{aligned}
$$

By (9.9.6) of Bai and Silverstein (2010), we have

$$
\begin{aligned}
&\left|\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}_{j-1}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j-n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{k1})]\right|\\
\le\ & \sum_{j=1}^{n}\left|\mathrm{E}_{j-1}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_j-n^{-1}\mathrm{tr}\boldsymbol{\Sigma})^2(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\sum_{k\le j-1}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\sum_{k\le j-1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{k1})]\right|\\
\le\ & \sum_{j=1}^{n}n^{-1}\eta_n^2\cdot C_0\|\boldsymbol{\Sigma}\|^2\|\boldsymbol{\Sigma}_{11}\sum_{k\le j-1}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}\|\\
\le\ & \eta_n^2\cdot C_0\|\boldsymbol{\Sigma}\|^2\|\boldsymbol{\Sigma}_{11}\sum_{k=1}^{n}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}\|\\
\le\ & \eta_n^2\cdot C_0\|\boldsymbol{\Sigma}\|^2\|\boldsymbol{\Sigma}_{11}\|^2\|\sum_{k=1}^{n}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top\|\\
=\ & \eta_n^2\cdot C_0\|\boldsymbol{\Sigma}\|^2\|\boldsymbol{\Sigma}_{11}\|^2\lambda_{\max}(\sum_{k=1}^{n}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top)\\
=\ & \eta_n^2\cdot C_0\|\boldsymbol{\Sigma}\|^2\|\boldsymbol{\Sigma}_{11}\|^2[(1+\sqrt{y_n})^2+o_{a.s.}(1)]=O_{a.s.}(\eta_n^2) \qquad \text{(S.2.32)}
\end{aligned}
$$

where $\lambda_{\max}(\sum_{k=1}^{n}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top)=(1+\sqrt{y_n})^2+o_{a.s.}(1)$ is the maximum eigenvalue of the random matrix $\sum_{k=1}^{n}\mathbf{r}_{k1}\mathbf{r}_{k1}^\top$ by Yin, Bai and Krishnaiah (1988).

Similar to the proofs of (S.2.24) and (S.2.25), we have

$$2n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}\mathbf{r}_{k1}^{\top}\mathbf{\Sigma}_{11}^{3}\mathbf{r}_{k1}=n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^{3}+O_p(n^{-1/2})$$

and

$$n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell1}^{\top}\mathbf{\Sigma}_{11}\mathbf{e}_{\ell1}\mathbf{r}_{k1}^{\top}\mathbf{\Sigma}_{11}\mathbf{e}_{\ell1}\mathbf{e}_{\ell1}^{\top}\mathbf{\Sigma}_{11}\mathbf{r}_{k1}=0.5n^{-1}\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell1}^{\top}\mathbf{\Sigma}_{11}\mathbf{e}_{\ell1}\mathbf{e}_{\ell1}^{\top}\mathbf{\Sigma}_{11}^{2}\mathbf{e}_{\ell1}+O_p(n^{-1/2}).$$

Thus we have

$$\begin{aligned}
&\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}_{j-1}[(\mathbf{r}_j^{\top}\mathbf{\Sigma}\mathbf{r}_j-n^{-1}\mathrm{tr}\mathbf{\Sigma})(\mathbf{r}_{j1}^{\top}\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^{\top}\mathbf{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathbf{r}_{k1}^{\top}\mathbf{\Sigma}_{11}^{2}\mathbf{r}_{k1})]\\
=&\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}_{j-1}[(\mathbf{r}_{j1}^{\top}\mathbf{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})(\mathbf{r}_{j1}^{\top}\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\mathbf{r}_{k1}^{\top}\mathbf{\Sigma}_{11}\mathbf{r}_{j1}-n^{-1}\mathbf{r}_{k1}^{\top}\mathbf{\Sigma}_{11}^{2}\mathbf{r}_{k1})]\\
=&\ n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}\left[2\mathbf{r}_{k1}^{\top}\mathbf{\Sigma}_{11}^{3}\mathbf{r}_{k1}+\beta_w\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell1}^{\top}\mathbf{\Sigma}_{11}\mathbf{r}_{\ell1}\mathbf{r}_{k1}^{\top}\mathbf{\Sigma}_{11}\mathbf{e}_{\ell1}\mathbf{e}_{\ell1}^{\top}\mathbf{\Sigma}_{11}\mathbf{r}_{k1}\right]\\
=&\ 0.5n^{-1}(2\mathrm{tr}\mathbf{\Sigma}_{11}^{3}+\beta_w\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell1}^{\top}\mathbf{\Sigma}_{11}\mathbf{e}_{\ell1}\mathbf{e}_{\ell1}^{\top}\mathbf{\Sigma}_{11}^{2}\mathbf{e}_{\ell1})+O_p(n^{-1}). \qquad \text{(S.2.33)}
\end{aligned}$$

By (S.2.31)-(S.2.32)-(S.2.33), we have

$$(S.2.16)=2(n^{-1}\mathrm{tr}\mathbf{\Sigma})(2n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^{3}+\beta_w n^{-1}\sum_{\ell=1}^{p}\mathbf{e}_{\ell1}^{\top}\mathbf{\Sigma}_{11}\mathbf{e}_{\ell1}\mathbf{e}_{\ell1}^{\top}\mathbf{\Sigma}_{11}^{2}\mathbf{e}_{\ell1})+O_p(\eta_n^2).$$

The detailed proofs of (S.2.18) is similar to the proofs of (S.2.16).

Detailed proofs of (S.2.19):

$$
\begin{aligned}
(S.2.19) &= 4\sum_{j=1}^{n}\mathrm{E}_{j-1}\sum_{k\le j-1}\sum_{i\le j-1}(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_k\mathbf{r}_k^\top\boldsymbol{\Sigma}\mathbf{r}_j(\mathrm{E}_j-\mathrm{E}_{j-1})\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{i1}\mathbf{r}_{i1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{j1}\\
&= 4\sum_{j=1}^{n}\mathrm{E}_{j-1}\sum_{k\le j-1}\sum_{i\le j-1}[(\mathbf{r}_j^\top\boldsymbol{\Sigma}\mathbf{r}_k)^2-n^{-1}\mathbf{r}_k^\top\boldsymbol{\Sigma}^2\mathbf{r}_k][(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{i1})^2-n^{-1}\mathbf{r}_{i1}^\top\boldsymbol{\Sigma}^2\mathbf{r}_{i1}]\\
&= 4\sum_{j=1}^{n}\mathrm{E}_{j-1}\sum_{k\le j-1}\sum_{i\le j-1}[(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{k1})^2-n^{-1}\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{k1}][(\mathbf{r}_{j1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{i1})^2-n^{-1}\mathbf{r}_{i1}^\top\boldsymbol{\Sigma}^2\mathbf{r}_{i1}]\\
&= 4n^{-2}\sum_{j=1}^{n}\mathrm{E}_{j-1}\sum_{k\le j-1}\sum_{i\le j-1}\left[2(\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{i1})^2+\beta_w\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{k1})^2(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{i1})^2\right]\\
&= 4n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}\left[2(\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{k1})^2+\beta_w\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{k1})^4\right]\\
&\quad +4n^{-2}\sum_{j=1}^{n}\sum_{1\le k\ne i\le j-1}\left[2(\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{i1})^2+\beta_w\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{k1})^2(\mathbf{e}_{\ell1}^\top\boldsymbol{\Sigma}_{11}\mathbf{r}_{i1})^2\right] \quad \text{(S.2.34)}
\end{aligned}
$$

where the fourth equality is from (1.15) of Bai and Silverstein (2004). Because

$$
\begin{cases}
n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}(\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{k1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}^2)^2=0.5n^{-2}[2\mathrm{tr}\boldsymbol{\Sigma}_{11}^4+\beta_w\sum_{\ell=1}^{p}(\mathbf{e}_{\ell1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{e}_{\ell1})^2]=O(n^{-1}),\\
n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}|\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{k1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}^2)|\le 0.5\{\mathrm{E}[(\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{k1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}^2)^2]\}^{1/2}=O(n^{-1/2}),
\end{cases}
$$

leads to

$$
\begin{cases}
n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}\mathrm{E}(\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{k1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}^2)^2=O_p(n^{-1}),\\
n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}(n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}^2)(\mathbf{r}_{k1}^\top\boldsymbol{\Sigma}_{11}^2\mathbf{r}_{k1}-n^{-1}\mathrm{tr}\boldsymbol{\Sigma}_{11}^2)=O_p(n^{-1/2}),
\end{cases}
$$

then we have

$$\begin{aligned} n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}(\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^2\mathbf{r}_{k1})^2 &= n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}(\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^2\mathbf{r}_{k1}-n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^2)^2+0.5(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^2)^2 \\ &\quad +2n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^2)(\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^2\mathbf{r}_{k1}-n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^2) \\ &= 0.5(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^2)^2+O_p(n^{-1/2}). \end{aligned} \tag{S.2.35}$$

Because $n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1})^4 = n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}\sum_{\ell=1}^{p}(\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}\mathbf{e}_\ell\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1})^2$, similar to the proof of $(S.2.35)$, we have

$$n^{-2}\sum_{j=1}^{n}\sum_{k\le j-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1})^4 = O_p(n^{-1/2}). \tag{S.2.36}$$

Because

$$\begin{aligned} &4n^{-2}\sum_{j=1}^{n}\sum_{1\le k\ne i\le j-1}\mathrm{E}\left[2(\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^2\mathbf{r}_{i1})^2+\beta_w\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1})^2(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{i1})^2\right] \\ =\ &\frac{4}{3}[2n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^4+\beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell1}^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_{\ell1})^2]+O(n^{-1}), \end{aligned}$$

and

$$\begin{cases} n^{-4}\mathrm{var}[\sum_{j=1}^{n}\sum_{1\le k\ne i\le j-1}(\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^2\mathbf{r}_{i1})^2] = O(n^{-1}), \\ n^{-4}\mathrm{var}[\sum_{j=1}^{n}\sum_{1\le k\ne i\le j-1}\sum_{\ell=1}^{p}(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1})^2(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{i1})^2] = O(n^{-1}), \end{cases}$$

then we have

$$\begin{aligned} &4n^{-2}\sum_{j=1}^{n}\sum_{1\le k\ne i\le j-1}\left[2(\mathbf{r}_{k1}^\top\mathbf{\Sigma}_{11}^2\mathbf{r}_{i1})^2+\beta_w\sum_{\ell=1}^{p_1}(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{k1})^2(\mathbf{e}_\ell^\top\mathbf{\Sigma}_{11}\mathbf{r}_{i1})^2\right] \\ =\ &\frac{4}{3}[2n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^4+\beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell1}^\top\mathbf{\Sigma}_{11}^2\mathbf{e}_{\ell1})^2]+4(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^2)^2+O_p(n^{-1/2}) \end{aligned} \tag{S.2.37}$$

By (S.2.34)-(S.2.35)-(S.2.36)-(S.2.37), we have

$$(S.2.19) = \frac{4}{3}[2n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^4 + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top \mathbf{\Sigma}_{11}^2 \mathbf{e}_{\ell 1})^2] + 4(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^2)^2 + O_p(n^{-1/2}).$$

Thus, we have

$$\begin{aligned}
(S.6) &= 2(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^3) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell 1}^\top \mathbf{\Sigma}_{11}^2 \mathbf{e}_{\ell 1}\mathbf{e}_{\ell 1}^\top \mathbf{\Sigma}_{11}\mathbf{e}_{\ell 1}] \\
&\quad + 2[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^4) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top \mathbf{\Sigma}_{11}^2 \mathbf{e}_{\ell 1})^2] + O_p(\eta_n^2), \\
(S.7) &= 4(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})(n^{-1}\mathrm{tr}\mathbf{\Sigma})[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^2) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top \mathbf{\Sigma}_{11}\mathbf{e}_{\ell 1})^2] \\
&\quad + 4(n^{-1}\mathrm{tr}\mathbf{\Sigma})[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^3) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}\mathbf{e}_{\ell 1}^\top \mathbf{\Sigma}_{11}\mathbf{e}_{\ell 1}\mathbf{e}_{\ell 1}^\top \mathbf{\Sigma}_{11}^2\mathbf{e}_{\ell 1}] + O_p(\eta_n^2), \\
(S.8) &= 2(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11})[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^3) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top \mathbf{\Sigma}_{11}\mathbf{e}_{\ell 1})(\mathbf{e}_{\ell 1}^\top \mathbf{\Sigma}_{11}^2\mathbf{e}_{\ell 1})] \\
&\quad + 2[2n^{-1}\mathrm{tr}(\mathbf{\Sigma}_{11}^4) + \beta_w n^{-1}\sum_{\ell=1}^{p_1}(\mathbf{e}_{\ell 1}^\top \mathbf{\Sigma}_{11}^2 \mathbf{e}_{\ell 1})^2] + 4(n^{-1}\mathrm{tr}\mathbf{\Sigma}_{11}^2)^2 + O_p(\eta_n^2).
\end{aligned}$$

The proof of Lemma S.2.4 is complete.