ASYMPTOTIC INDEPENDENCE OF THE SUM AND MAXIMUM OF DEPENDENT RANDOM VARIABLES WITH APPLICATIONS TO HIGH-DIMENSIONAL TESTS

Long Feng[1], Tiefeng Jiang[2], Xiaoyun Li[3], Binghui Liu[4]

[1]*Nankai University,* [2]*University of Minnesota,* [3]*LinkedIn Corporation and* [4]*Northeast Normal University*

Supplementary Material

We propose the combo-type two-sample mean test and regression coefficient test, and present the simulation results in comparison with some of its competitors. We also provide the technical proofs of the theoretical results in Sections 2, 3 of the main text and Section S1-S2 of the Supplementary Material.

S1 Two-sample Mean Test

S1.1 Testing Procedure

Here, we consider the two-sample mean testing problem in the high-dimensional setting. Assume that $\{\boldsymbol{X}_{i1}, \cdots, \boldsymbol{X}_{in_i}\}$ for $i = 1, 2$ are two independent random samples with sizes n_1 and n_2, and from p-variate normal distributions $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, respectively. Consider

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{versus} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \tag{S1}$$

For the case where dimension p is fixed, the classic Hotelling's T^2 test statistic is

$$\frac{n_1 n_2}{n_1 + n_2}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)^T \hat{\mathbf{S}}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2), \tag{S2}$$

where $\bar{\boldsymbol{X}}_i$ is the sample mean vector of the ith sample and $\hat{\mathbf{S}}$ is the pooled sample covariance matrix defined by

$$\hat{\mathbf{S}} = \frac{1}{n_1 + n_2}\Big[\sum_{j=1}^{n_1}(\mathbf{X}_{1j} - \bar{\mathbf{X}}_1)(\mathbf{X}_{1j} - \bar{\mathbf{X}}_1)^T + \sum_{j=1}^{n_2}(\mathbf{X}_{2j} - \bar{\mathbf{X}}_2)(\mathbf{X}_{2j} - \bar{\mathbf{X}}_2)^T\Big]. \tag{S3}$$

Let $n = n_1 + n_2$. In the high-dimensional case with $p > n$, $\hat{\mathbf{S}}$ is not guaranteed to be invertible. Under the assumption that $p/n \to c \in (0, \infty)$, Bai and Saranadasa (1996) proposed a test statistic $(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)^T(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)$ by replacing $\hat{\mathbf{S}}$ in (S2) with the identity matrix. Without the restriction on n and p, Chen and Qin (2010) constructed a different test statistic by excluding the term $\sum_{j=1}^{n_i} \boldsymbol{X}_{ij}^T \boldsymbol{X}_{ij}$ for $i = 1$ and 2 from $(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)^T(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)$. However, the above two tests are not scale-invariant. A statistic $T(\boldsymbol{X}_{11}, \cdots, \boldsymbol{X}_{1n_1}, \boldsymbol{X}_{21}, \cdots, \boldsymbol{X}_{2n_2})$ is said to be location-scale invariant if the corresponding value of T is not changed provided "$\boldsymbol{X}_{ij}$" is replaced by "$a\boldsymbol{X}_{ij} + b$" for all i, j, where a and b are arbitrary constants free of i and j. We say T is scale-invariant if the above holds with $b = 0$. For this reason, many efforts have been devoted to construct location-scale invariant test procedures, including Srivastava and Du (2008); Gregory et al. (2015); Feng et al. (2015), to name a few. In particular, Srivastava and Du (2008) considered the following sum-type test statistic

$$T_{sum}^{(2)} = \frac{\frac{n_1 n_2}{n_1+n_2}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)^T \hat{\mathbf{D}}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) - \frac{(n_1+n_2-2)p}{(n_1+n_2-4)}}{\sqrt{2\left[\mathrm{tr}(\hat{\mathbf{R}}^2) - \frac{p^2}{(n_1+n_2-2)}\right]c_{p,n}}}, \tag{S4}$$

where $\hat{\mathbf{D}}$ is the diagonal matrix of $\hat{\mathbf{S}}$ in (S3) and $\hat{\mathbf{R}} = \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{S}}\hat{\mathbf{D}}^{-1/2}$ is the pooled sample correlation matrix, and $c_{p,n} = 1 + \frac{\mathrm{tr}(\hat{\mathbf{R}}^2)}{p^{3/2}}$. The statistic $T_{sum}^{(2)}$ is location-scale invariant. Similar to the discussion in the paragraph above (3.5), the above sum-type tests usually do not perform well for sparse data. For the sparse alternative, Chen et al. (2019) extended the work of Zhong et al. (2013) by studying the statistic

$$M_{L_n} = \max_{s \in \mathcal{S}_n} \frac{L_n(s) - \hat{\mu}_{L_n(s),0}}{\hat{\sigma}_{L_n(s),0}}. \tag{S5}$$

Note that this formula follows (4.3) from Chen et al. (2019), where the notations "$L_n(s)$, $\mathcal{S}_n$, $\hat{\mu}_{L_n(s),0}$, $\hat{\sigma}_{L_n(s),0}$" are quite involved; interested readers are referred to their paper for more details. Later on, we will compare our proposed test with M_{L_n} in (S5). Again, for sparse data, Cai et al. (2014)

proposed the following max-type test statistic

$$T_{max}^{(2)} = \frac{n_1 n_2}{n_1 + n_2} \max_{1 \leqslant i \leqslant p} \frac{(\bar{\boldsymbol{X}}_{1i} - \bar{\boldsymbol{X}}_{2i})^2}{\hat{\sigma}_{ii}^2},$$

where $\bar{\boldsymbol{X}}_{ji}$ is the ith coordinate of $\bar{\boldsymbol{X}}_j \in \mathbb{R}^p$ for $j = 1, 2$ and $1 \leqslant i \leqslant p$ and $\hat{\sigma}_{ii}^2$ is the ith diagonal element of $\hat{\mathbf{S}}$ in (S3). Similar to the one-sample test case, this test statistic is particularly powerful against sparse alternatives with certain optimality. We will study the asymptotic behavior of the sum-type and max-type tests, and design a new test that takes advantage of both worlds.

Recall that $\boldsymbol{\Sigma}$ is the covariance matrix shared by two populations, and let $\mathbf{D}$ be the diagonal matrix of $\boldsymbol{\Sigma}$ such that $\mathbf{R} := \mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}$ is the population correlation matrix. For soundness, assume that the sample sizes n_1 and n_2 both depend on p. Now we apply the theoretical results in Section 2 to the two-sample mean test as follows.

Theorem S1. *Assume the null hypothesis in* (S1) *holds and* $\lim_{p\to\infty} n_1/n_2 \to \kappa \in (0, \infty)$. *The following are true as* $p \to \infty$*:*

(i) If (3.6) *holds, then* $T_{sum}^{(2)} \to N(0,1)$ *in distribution;*

(ii) If (2.2) *holds with "*$\boldsymbol{\Sigma}$*" being replaced by "*$\mathbf{R}$*" and* $\log p = o(n^{1/3})$, *then* $T_{max}^{(2)} - 2\log p + \log\log p$ *converges weakly to a Gumbel distribution with cdf* $F(x) = \exp\{-\frac{1}{\sqrt{\pi}}\exp(-x/2)\}$;

(iii) Assume (3.6) *holds. If* (2.3) *is true with "$\mathbf{\Sigma}$" replaced by "$\mathbf{R}$", then $T_{sum}^{(2)}$ and $T_{max}^{(2)} - 2\log p + \log\log p$ are asymptotically independent.*

Part (i) of Theorem S1 is from Srivastava and Du (2008). Recently Jiang and Li (2021) obtained a general theory, which also leads to the same conclusion. Same as in the one-sample test, for test $T_{sum}^{(2)}$, a level-α test rejects H_0 when $T_{sum}^{(2)} > z_\alpha = \Phi^{-1}(1-\alpha)$ of $N(0,1)$. For the max-type test, a level-α test will then be carried out through rejecting the null hypothesis when $T_{max}^{(2)} - 2\log p + \log\log p > q_\alpha = -\log\pi - 2\log\log(1-\alpha)^{-1}$ of the distribution function in Theorem S1(ii).

Relying on Theorem S1, we propose the following test statistic which utilizes the max-type and sum-type tests. Define

$$T_{com}^{(2)} = \min\{P_M^{(2)}, P_S^{(2)}\}, \tag{S6}$$

where $P_M^{(2)} = 1 - F(T_{max}^{(2)} - 2\log p + \log\log p)$ with $F(y) = e^{-\pi^{-1/2}e^{-y/2}}$ and $P_S^{(2)} = 1 - \Phi(T_{sum}^{(2)})$. are the p-values of the two tests, respectively. Similar to Corollary 1, we immediately obtain the following result by the asymptotic independence.

Corollary S1. *Assume the condition in Theorem S1(iii) holds. Then $T_{com}^{(2)}$ in* (S6) *converges weakly to a distribution with density $G(w) = 2(1-w)I(0 \leqslant w \leqslant 1)$ as $p \to \infty$.*

According to Corollary S1, the proposed combo-type test leads us to perform a level-α test by rejecting the null hypothesis when $T_{com}^{(2)} < 1 - \sqrt{1-\alpha} \approx \frac{\alpha}{2}$ when α is small. Now we analyze the power of the test $T_{com}^{(2)}$.

Write $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^T$. Similar to (3.8), the power function of our combo-type test $\beta_C^{(2)}(\boldsymbol{\mu}, \alpha)$ is larger than $\max\{\beta_M^{(2)}(\boldsymbol{\mu}, \alpha/2), \beta_S^{(2)}(\boldsymbol{\mu}, \alpha/2)\}$, where $\beta_M^{(2)}(\boldsymbol{\mu}, \alpha)$ and $\beta_S^{(2)}(\boldsymbol{\mu}, \alpha)$ are the power functions of $T_{max}^{(2)}$ and $T_{sum}^{(2)}$ with significant level α, respectively. Following Srivastava and Du (2008), the power function of $T_{sum}^{(2)}$ is given by

$$\beta_S^{(2)}(\boldsymbol{\mu}, \alpha) = \lim_{p\to\infty} \Phi\left(-z_\alpha + \frac{\frac{n_1 n_2}{n_1+n_2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \mathbf{D}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\sqrt{2\mathrm{tr}(\mathbf{R}^2)}}\right).$$

Thus, we have

$$\beta_C^{(2)}(\boldsymbol{\mu}, \alpha) \geqslant \lim_{p\to\infty} \Phi\left(-z_{\alpha/2} + \frac{\frac{n_1 n_2}{n_1+n_2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \mathbf{D}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\sqrt{2\mathrm{tr}(\mathbf{R}^2)}}\right).$$

Further write $\boldsymbol{\mu}_1 = (\mu_{11}, \cdots, \mu_{1p})^T$ and $\boldsymbol{\mu}_2 = (\mu_{21}, \cdots, \mu_{2p})^T$, and define $\delta_i = \mu_{1i} - \mu_{2i}$ for $i = 1, \cdots, p$. We have an analogous set of analysis and claims as for the one-sample test. Firstly, by Theorem 2 from Cai et al. (2014), the asymptotic power of $T_{max}^{(2)}$ converges to one if $\max_{1\leqslant i\leqslant p} |\delta_i/\sigma_{ii}| \geqslant c\sqrt{\log p/n}$ for a certain constant c and if the sparsity level $\gamma < 1/4$ and the locations of the non-zero variables are randomly and uniformly selected from $\{1, \cdots, p\}$, meaning that the power function of our proposed test $T_{com}^{(2)}$ also converges to one under this situation. Secondly, according to Theorem 3 from Cai et al. (2014), the condition $\max_{1\leqslant i\leqslant p}|\delta_i/\sigma_{ii}| \geqslant c\sqrt{\log p/n}$ is minimax rate-optimal for testing against sparse alternatives, and such optimality also holds for our test $T_{max}^{(2)}$.

Similar to the one-sample test problem, we consider a special case with $\mathbf{\Sigma} = \mathbf{I}_p$. There are m nonzeros δ_i and they are all equal to $\delta \neq 0$. Thus,

$$\beta_S^{(2)} = \lim_{p\to\infty} \Phi\left(-z_\alpha + \frac{n_1 n_2 m\delta^2}{n\sqrt{2p}}\right).$$

Take $\xi > 0$ such that $p^{1/2}n^{2\xi-1} \to \infty$. For the non-sparse case: $\delta = O(n^{-\xi})$ and $m = o(p^{1/2}n^{2\xi-1})$, we have $\beta_M^{(2)}(\boldsymbol{\mu},\alpha) \approx \alpha$ and $\beta_C^{(2)}(\boldsymbol{\mu},\alpha) \approx \beta_S^{(2)}(\boldsymbol{\mu},\alpha/2)$. For the sparse case: $\delta = c\sqrt{\log p/n}$ with sufficient large c and $m = o((\log p)^{-1}p^{1/2})$, we have $\beta_S^{(2)}(\boldsymbol{\mu},\alpha) \approx \alpha$ and $\beta_C^{(2)}(\boldsymbol{\mu},\alpha) \approx \beta_M^{(2)}(\boldsymbol{\mu},\alpha/2) \to 1$.

S1.2 Simulation Results

Then, we present the simulation results for the two-sample test problem, where our test $T_{com}^{(2)}$ (abbreviated as COM) from (S6) will be compared with the sum-type test $T_{sum}^{(2)}$ from (S4) proposed by Srivastava and Du (2008) (abbreviated as SUM), the max-type test $T_{max}^{(2)}$ proposed by Cai et al. (2014) (abbreviated as MAX) and the Higher Criticism test proposed by Chen et al. (2019) (abbreviated as HC2).

Recall the three scenarios of covariance matrices appeared in (I), (II) and (III) after Example 1. Since the conclusions from all three scenarios are similar, here we only present the results when the covariance matrix follows Scenario (I), i.e., $\mathbf{\Sigma} = (0.5^{|i-j|})_{1\leqslant i,j\leqslant p}$.

Example 3. We consider $\boldsymbol{X}_{ki} = \boldsymbol{\mu}_k + \mathbf{\Sigma}^{1/2}\boldsymbol{z}_{ki}$ for $k = 1, 2$ and $i = 1 \cdots, n$, and each component of $\boldsymbol{z}_{ki}$ is independently generated from three distributions: (1) $N(0, 1)$; (2) t-distribution, $t(5)/\sqrt{5/3}$; (3) the mixture normal random variable $V/\sqrt{1.8}$, where V is as in Example 1.

We consider two different sample sizes $n = 100, 200$ and three different dimensions $p = 200, 400, 600$. Under the null hypothesis, we set $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$. The significance level is chosen such that $\alpha = 0.05$. Again, for the alternative hypothesis, we only present on $n = 100, p = 200$ and $\boldsymbol{\mu}_2 = \mathbf{0}$ since the observations from different combinations of n and p are similar. Define $\boldsymbol{\mu}_1 = (\mu_{11}, \cdots, \mu_{1p})^T$. For different number of nonzero-mean variables $m = 1, \cdots, 20$, we consider $\mu_{j1} = \delta$ for $0 < j \leqslant m$ and $\mu_{1j} = 0$ for $j > m$. The parameter δ is chosen such that $||\boldsymbol{\mu}_1||^2 = m\delta^2 = 1$.

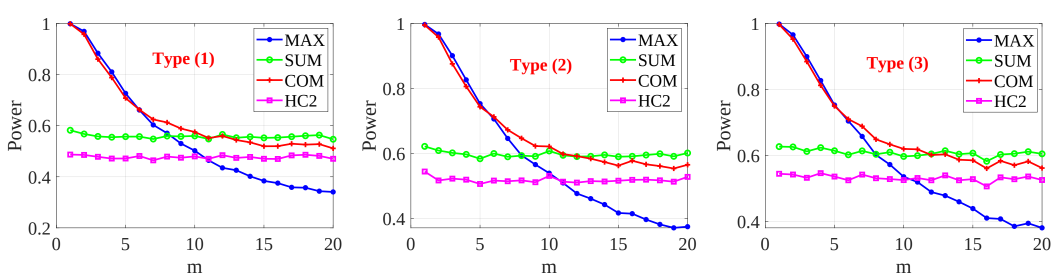
Table S1 reports the empirical sizes of the compared tests. We see that all the tests control the empirical sizes in most cases except that the sizes of HC2 are a little smaller than the nominal level when $n = 200$.

Figure S1 shows the power of each test, where we observe a similar pattern as in Example 1. The power of MAX declines as the number of variables with nonzero means is increasing. The power of SUM and COM

Table S1: Sizes of tests for Example 3, $\alpha = 0.05$.

	Distribution	(1)			(2)			(3)		
	p	200	400	600	200	400	600	200	400	600
$n = 100$	MAX	0.057	0.063	0.058	0.062	0.061	0.060	0.060	0.066	0.064
	SUM	0.055	0.058	0.057	0.060	0.064	0.065	0.055	0.061	0.061
	COM	0.057	0.074	0.060	0.064	0.065	0.068	0.061	0.061	0.062
	HC2	0.043	0.035	0.044	0.047	0.037	0.052	0.042	0.043	0.053
$n = 200$	MAX	0.053	0.053	0.051	0.049	0.062	0.061	0.044	0.065	0.055
	SUM	0.052	0.065	0.059	0.065	0.061	0.064	0.052	0.056	0.053
	COM	0.044	0.049	0.056	0.061	0.063	0.058	0.046	0.061	0.053
	HC2	0.030	0.025	0.025	0.030	0.035	0.036	0.025	0.020	0.033

Figure S1: Power versus number of variables with non-zero means for Example 3. The x-axis m denotes the number of variables with non-zero means; the y-axis is the empirical power.



are always larger than that of HC2 in all cases. The proposed COM matches the power of MAX when the number of variables with nonzero means is small, and almost has same power as SUM when m is large. This justifies the superiority of the proposed combo-type test in the two-sample testing problem, regardless of the sparsity of the data.

S2 Test for Regression Coefficients

S2.1 Testing Procedure

Then, we apply our theory to the testing problem in high-dimensional regression. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n)^T$ be independent and identically distributed p-dimensional covariates, and $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^T$ be the corresponding independent responses. For simplicity, we assume $E\boldsymbol{X}_i = 0$, $\forall i$. We introduce a decomposition of sample point $\boldsymbol{X}_i$ as $\boldsymbol{X}_i = (\boldsymbol{X}_{ia}^T, \boldsymbol{X}_{ib}^T)^T$ with $\boldsymbol{X}_{ia} = (X_{i1}, \cdots, X_{iq})^T \in \mathbb{R}^q$ and $\boldsymbol{X}_{ib} = (X_{i(q+1)}, \cdots, X_{ip})^T \in \mathbb{R}^{p-q}$, where q is smaller than the sample size n, and n is much smaller than p. We consider the following standard linear regression model:

$$Y_i = \boldsymbol{X}_i^T\boldsymbol{\beta} + \varepsilon_i = \boldsymbol{X}_{ia}^T\boldsymbol{\beta}_a + \boldsymbol{X}_{ib}^T\boldsymbol{\beta}_b + \varepsilon_i, \tag{S7}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_a^T, \boldsymbol{\beta}_b^T)^T \in \mathbb{R}^p$, $\boldsymbol{\beta}_a \in \mathbb{R}^q$ and $\boldsymbol{\beta}_b \in \mathbb{R}^{p-q}$ are the regression coefficient vectors. The random noises $\{\varepsilon_i; 1 \leqslant i \leqslant n\}$ are independent with $E\varepsilon_i = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2$ for each i, and are also independent of the data $\boldsymbol{X}$. In this section, we consider the following testing problem:

$$H_0 : \boldsymbol{\beta}_b = 0 \ \text{ vs. } \ H_1 : \boldsymbol{\beta}_b \neq 0 \tag{S8}$$

under the situation that p is much larger than q and the sample size n is small. When $(\boldsymbol{X}_{1a}^T, \cdots, \boldsymbol{X}_{na}^T)^T$ is a null vector, (S8) is equivalent to test $H_0 : \boldsymbol{\beta} = 0$ vs. $H_1 : \boldsymbol{\beta} \neq 0$.

For this problem, Goeman et al. (2006) and Goeman et al. (2011) proposed an empirical Bayes test. It is formulated via a score test on the hyper parameter of a prior distribution on the regression coefficients. By excluding the inverse term in the classical F-statistic, Zhong and Chen (2011) proposed a U-statistic to extend their results in factorial designs.

Denote $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$. To motivate our test procedure, let us consider a set of related but not exactly the same test as follows:

$$H_{0j} : \beta_j = 0 \ \text{ vs. } \ H_{1j} : \beta_j \neq 0 \tag{S9}$$

for each $j = q + 1, \cdots, p$. Some notations are needed before we proceed. For the set of features included in the index set of "a", let $\boldsymbol{X}_a = (\boldsymbol{X}_{1a}, \cdots, \boldsymbol{X}_{na})^T$, $\mathbf{H}_a = \mathbf{X}_a(\mathbf{X}_a^T\mathbf{X}_a)^{-1}\mathbf{X}_a^T$ and $\tilde{\mathbf{X}}_j = (\mathbf{I}_n - \mathbf{H}_a)(X_{1j}, \cdots, X_{nj})^T$, for each j. Notice $\boldsymbol{X}_a$ is $n \times q$, $\mathbf{H}_a$ is $n \times n$ and $\tilde{\mathbf{X}}_j \in \mathbb{R}^n$. For the "$b$" part, we define

$$\boldsymbol{X}_b = (\boldsymbol{X}_{1b}, \cdots, \boldsymbol{X}_{nb})^T \quad \text{and} \quad \tilde{\mathbf{X}}_b = (\tilde{\mathbf{X}}_{q+1}, \cdots, \tilde{\mathbf{X}}_p) = (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_b;\ \hat{\boldsymbol{\Sigma}}_{b|a} = n^{-1}\tilde{\mathbf{X}}_b^T\tilde{\mathbf{X}}_b, \tag{S10}$$

where both $\boldsymbol{X}_b$ and $\tilde{\mathbf{X}}_b$ are $n \times (p-q)$ and $\hat{\boldsymbol{\Sigma}}_{b|a}$ is $(p-q) \times (p-q)$. Regarding the response vector, the residual vector and the sample variance, we denote $\hat{\boldsymbol{\varepsilon}} = (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{Y}$ and $\hat{\sigma}^2 = (n-q)^{-1}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}$. For each test in (S9), the classical partial F-test is given by $F_j = \frac{\boldsymbol{Y}^T\tilde{\mathbf{X}}_j(\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j)^{-1}\tilde{\mathbf{X}}_j^T\boldsymbol{Y}}{\boldsymbol{Y}^T[\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\boldsymbol{Y}/(n-p)}$. However, as $p > n$, the statistic F_j becomes problematic since $\mathbf{X}^T\mathbf{X}$ is not invertible. To overcome this issue, we replace the denominator of F_j by $\boldsymbol{Y}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{Y}/(n-q)$. Under the null hypothesis, $\hat{\sigma}^2$ is an unbiased estimator of σ^2; see (S163) in the supplementary material. Thus, the test statistic for (S9) becomes $\tilde{F}_j = \frac{\boldsymbol{Y}^T\tilde{\mathbf{X}}_j(\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j)^{-1}\tilde{\mathbf{X}}_j^T\boldsymbol{Y}}{\boldsymbol{Y}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{Y}/(n-q)}$. Back to the testing problem (S8) of interest, we will handle it by combining $\tilde{F}_j$ together. There are two classical ways to synthesize them. The first one is the sum-type test statistic: $S_F = \sum_{j=q+1}^p \tilde{F}_j$. Obviously, we have $S_F = T_1/\hat{\sigma}^2$ with $T_1 = \sum_{j=q+1}^p \boldsymbol{Y}^T\tilde{\mathbf{X}}_j(\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j)^{-1}\tilde{\mathbf{X}}_j^T\boldsymbol{Y}$. By normalization, without loss of generality, assume $\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j = n$ for each $j = q+1, \cdots, p$. Then $T_1 = \frac{1}{n}\sum_{j=q+1}^p \boldsymbol{Y}^T\tilde{\mathbf{X}}_j\tilde{\mathbf{X}}_j^T\boldsymbol{Y} = \frac{1}{n}\hat{\boldsymbol{\varepsilon}}^T\boldsymbol{X}_b\boldsymbol{X}_b^T\hat{\boldsymbol{\varepsilon}}$. By standardizing T_1 with estimators of the mean and standard deviation of T_1, we propose the sum-type statistic

$$T_{sum}^{(3)} = \frac{n^{-1}\hat{\boldsymbol{\varepsilon}}^T \boldsymbol{X}_b \boldsymbol{X}_b^T \hat{\boldsymbol{\varepsilon}} - n^{-1}(n-q)(p-q)\hat{\sigma}^2}{\sqrt{2\hat{\sigma}^4 \,\widehat{\mathrm{tr}\left(\boldsymbol{\Sigma}_{b|a}^2\right)}}}, \tag{S11}$$

where the denominator is $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^2)} = \frac{n^2}{(n+1-q)(n-q)}\left\{\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_{b|a}^2\right) - \frac{1}{n-q}\mathrm{tr}^2(\hat{\boldsymbol{\Sigma}}_{b|a})\right\}$. The second statistic we are interested in is of the max-type, defined by

$$T_{max}^{(3)} = \max_{q+1\leqslant j\leqslant p} \tilde{F}_j. \tag{S12}$$

Then, we will study the asymptotic distributions of $T_{sum}^{(3)}$ and $T_{max}^{(3)}$, as well as their asymptotic independence, based on which a combo-type test will be proposed. To begin with, we introduce some additional assumptions that will be used.

Define $\boldsymbol{\Sigma}_{b|a} = E[\mathrm{Cov}(\boldsymbol{X}_{ib}|\boldsymbol{X}_{ia})] = (\sigma_{jk}^*)$ as a $(p-q)\times(p-q)$ matrix. Without loss of generality, we assume that $\boldsymbol{\Sigma}_{b|a}$ is normalized such that its diagonal entries are equal to one, i.e., $\sigma_{jj}^* = 1$ for each j. Next, we introduce moment conditions on a conditional predictor. For each $i = 1, \cdots, n$, by regressing $\boldsymbol{X}_{ib}$ on $\boldsymbol{X}_{ia}$, the residual vector is given by $\boldsymbol{X}_{ib}^* = \boldsymbol{X}_{ib} - \mathbf{B}\boldsymbol{X}_{ia} \in \mathbb{R}^{p-q}$, where $\mathbf{B} := \mathrm{Cov}(\boldsymbol{X}_{ib}, \boldsymbol{X}_{ia})\cdot[\mathrm{Cov}(\boldsymbol{X}_{ia})]^{-1}$ is a $(p-q)\times q$ matrix. By the previous assumption $E\boldsymbol{X}_i = 0$, we immediately have that $E(\boldsymbol{X}_{ib}^*) = 0$ and $\mathrm{Cov}(\boldsymbol{X}_{ib}^*) = \boldsymbol{\Sigma}_{b|a} = (\sigma_{ij}^*)$. Define $\mathbf{X}_b^* = (\boldsymbol{X}_{1b}^*, \cdots, \boldsymbol{X}_{nb}^*)^T$, which is a

$n \times (p-q)$ matrix. The conditions we will use later on are stated below.

$$\boldsymbol{X}_{ib}^{*} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{b|a}) \text{ and the diagonal entries of } \boldsymbol{\Sigma}_{b|a} \text{ are all equal to 1;} \tag{S13}$$

$$\text{There exists a constant } \tau > 1 \text{ such that } \tau^{-1} < \lambda_{min}(\boldsymbol{\Sigma}_{b|a}) \leqslant \lambda_{max}(\boldsymbol{\Sigma}_{b|a}) < \tau. \tag{S14}$$

Assumption (S14) is the same as condition (C1) from Lan et al. (2014), which is also a common assumption in literature for research on high-dimensional data; see, for example, Fan et al. (2008), Rothman et al. (2008), Zhang and Huang (2008) and Wang (2009). We assume that both n and q depend on p and limits will be taken as $p \to \infty$. The random errors $\varepsilon_1, \cdots, \varepsilon_n$ are assumed to be i.i.d. with $E\varepsilon_i = 0$ and $\text{Var}(\varepsilon_i) = \sigma^2$ for each i, and no Gaussian assumption is needed for ε_i. Our main result for high-dimensional regression coefficient test is as follows.

Theorem S2. *Assume* (S13) *and* (S14) *hold and* (2.3) *also holds with "$\boldsymbol{\Sigma}$" replaced by "$\boldsymbol{\Sigma}_{b|a}$". Suppose $p = o(n^3)$, $q = o(p)$, $q \leqslant n^\delta$ for some $\delta \in (0,1)$ and $E(|\epsilon_1|^\ell) < \infty$ with $\ell = 14(1-\delta)^{-1}$. Under H_0 from* (S8)*, as $p \to \infty$ we have:*

(i) $T_{sum}^{(3)} \to N(0,1)$ in distribution;

(ii) $T_{max}^{(3)} - 2\log(p-q) + \log\log(p-q)$ converges weakly to a distribution

with cdf $F(x) = \exp\{-\frac{1}{\sqrt{\pi}}\exp(-\frac{x}{2})\}$, $x \in \mathbb{R}$*;*

(iii) $T_{sum}^{(3)}$ *and* $T_{max}^{(3)} - 2\log(p-q) + \log\log(p-q)$ *are asymptotically independent.*

It is known from Lan et al. (2014) that $\frac{n^{-1}\hat{\boldsymbol{\varepsilon}}^T \boldsymbol{X}_b \boldsymbol{X}_b^T \hat{\boldsymbol{\varepsilon}} - \hat{\sigma}^2 n^{-1}(p-q)\mathrm{tr}(\mathbf{M}(\mathbf{I}_n - \mathbf{H}_a))}{\sqrt{2\sigma^4 \,\mathrm{tr}\left(\boldsymbol{\Sigma}_{b|a}^2\right)}}$ converges to $N(0,1)$ in distribution, with $\mathbf{M} := (p-q)^{-1}\sum_{j=q+1}^{p} \tilde{\mathbf{X}}_j \tilde{\mathbf{X}}_j^T$. Although this is not a statistic (since unknown parameters appear in the denominator), it describes the asymptotic behavior of $\hat{\boldsymbol{\varepsilon}}^T \boldsymbol{X}_b \boldsymbol{X}_b^T \hat{\boldsymbol{\varepsilon}}$. The numerator of our statistic $T_{sum}^{(3)}$ in (S11) is simpler, because no computation of $\mathrm{tr}(\mathbf{M}(\mathbf{I}_n - \mathbf{H}_a))$ is needed.

We now study the implications of Theorem S2 and discuss the rejection rules. For the sum-type test, a level-α test will be performed through rejecting H_0 when $T_{sum}^{(3)}$ is larger than the $(1-\alpha)$-quantile $z_\alpha = \Phi^{-1}(1-\alpha)$ of $N(0,1)$. For the max-type test, a level-α test will then be performed through rejecting H_0 when $T_{max}^{(3)} - 2\log(p-q) + \log\log(p-q)$ is larger than the $(1-\alpha)$-quantile $q_\alpha = -\log\pi - 2\log\log(1-\alpha)^{-1}$ of $F(x)$.

Analogously, a combined test is defined through

$$T_{com}^{(3)} = \min\left\{P_S^{(3)}, P_M^{(3)}\right\}, \tag{S15}$$

where $P_S^{(3)} = 1 - \Phi(T_{sum}^{(3)})$ and $P_M^{(3)} = 1 - F(T_{max}^{(3)} - 2\log(p-q) + \log\log(p-q))$ with $F(x) = \exp\{-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{x}{2}\right)\}$. Similar to Corollary 1, the proposed combo-type test allows us to perform a level-α test by rejecting the null hypothesis when $T_{com}^{(3)} < 1 - \sqrt{1-\alpha} \approx \frac{\alpha}{2}$.

Again, using the same argument as (3.8), the power function of our combo-type test $\beta_C^{(3)}(\boldsymbol{\mu}, \alpha)$ is larger than $\max\{\beta_M^{(3)}(\boldsymbol{\mu}, \alpha/2), \beta_S^{(3)}(\boldsymbol{\mu}, \alpha/2)\}$, where $\beta_M^{(3)}(\boldsymbol{\mu}, \alpha)$ and $\beta_S^{(3)}(\boldsymbol{\mu}, \alpha)$ are the power functions of $T_{max}^{(3)}$ and $T_{sum}^{(3)}$ at significant level α, respectively. To demonstrate the power of the tests, assume the simple case where $\boldsymbol{X}_a$ is null vector and $\mathrm{Cov}(\boldsymbol{X}_b) = \mathbf{I}_p$. From Zhong and Chen (2011), the power function of $T_{sum}^{(3)}$ is $\beta_S^{(3)}(\boldsymbol{\beta}, \alpha) = \lim_{p\to\infty} \Phi\left(-z_\alpha + \frac{n\boldsymbol{\beta}^T\boldsymbol{\beta}}{\sqrt{2p}\sigma^2}\right)$. In addition, assuming $\boldsymbol{\beta}$ only contains m non-zeros all equal to $\delta \neq 0$, leading to $\beta_S^{(3)}(\boldsymbol{\beta}, \alpha) = \lim_{p\to\infty} \Phi\left(-z_\alpha + \frac{nm\delta^2}{\sqrt{2p}\sigma^2}\right)$. For the non-sparse case with $\delta = O(n^{-\xi})$ and $m = O(p^{1/2}n^{2\xi-1})$, we have $\beta_M^{(3)}(\boldsymbol{\mu}, \alpha) \approx \alpha$ and $\beta_C^{(3)}(\boldsymbol{\mu}, \alpha) \approx \beta_S^{(3)}(\boldsymbol{\mu}, \alpha/2)$. For the sparse case where $\delta = c\sqrt{\log p/n}$ with sufficient large c and $m = o((\log p)^{-1}p^{1/2})$, we have $\beta_S^{(3)}(\boldsymbol{\mu}, \alpha) \approx \alpha$ and $\beta_C^{(3)}(\boldsymbol{\mu}, \alpha) \approx \beta_M^{(3)}(\boldsymbol{\mu}, \alpha/2) \to 1$. Again, in this testing problem for high-dimensional regression, the combined test statistics also exhibits good performance under both sparse and dense alternative hypotheses.

S2.2 Simulation Results

Then, we present our simulation results on the regression coefficient testing problem. We will compare our combo-type test $T_{com}^{(3)}$ (abbreviated as COM) from (S15) against the test $T^{(3)}_{sum}$ (abbreviated as SUM) from (S11), the test $T^{(3)}_{max}$ (abbreviated as MAX) from (S12) and the empirical Bayes test proposed by Goeman et al. (2006) (abbreviated as EB).

EXAMPLE 2. We generate data from (S7), where the regression coefficients β_j for $j \in \{1, 2, \cdots, q\}$ are simulated from a standard normal distribution, and then we set $\beta_j = 0$ for $j > q$. In addition, the predictor vector is given by $\boldsymbol{X}_i = \boldsymbol{\Sigma}^{1/2}\boldsymbol{z}_i$ for $i = 1 \cdots, n$, and each component of $\boldsymbol{z}_i$ is independently generated from three distributions: (1) the normal distribution $N(0, 1)$; (2) the exponential distribution $\exp(1) - 1$; (3) the mixture normal distribution $V/\sqrt{1.8}$, where V is as in Example 1.

Moreover, the random error ε_i is independently generated from a standard normal distribution. We report the Scenarios (I) result where $\text{Cov}(\boldsymbol{X}_i) = \boldsymbol{\Sigma} = (\sigma_{jk}) \in \mathbb{R}^{p\times p}$ with $\sigma_{jk} = 0.5^{|j-k|}$, while the results in the other two cases are similar. We consider two different sample sizes $n = 100, 200$, three different dimensions $p = 200, 400, 600$ and two dimension of predictors in the reduced model $q = 0$ or 5.

We report the empirical sizes of the tests in Table S2. We observe that the empirical size of MAX tends to be smaller than the nominal level. EB and SUM, as well as the proposed COM, control the empirical sizes well for most of the times.

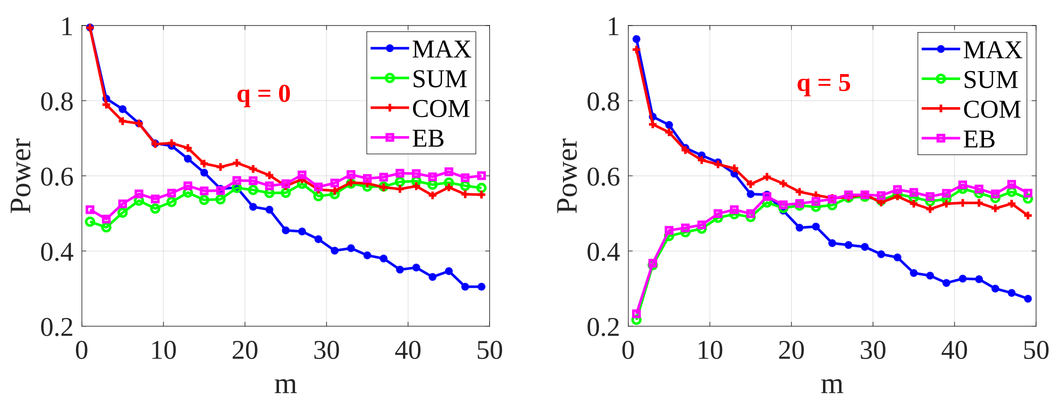
Table S2: Sizes of tests for Example 2, $\alpha = 0.05$.

	Distribution	(1)			(2)			(3)		
	p	200	400	600	200	400	600	200	400	600
$q = 0$										
$n = 100$	MAX	0.032	0.024	0.027	0.026	0.032	0.026	0.027	0.036	0.036
	EB	0.047	0.047	0.047	0.049	0.052	0.050	0.059	0.044	0.044
	SUM	0.061	0.064	0.065	0.060	0.072	0.062	0.079	0.058	0.060
	COM	0.044	0.043	0.047	0.046	0.05	0.043	0.055	0.047	0.044
$n = 200$	MAX	0.032	0.042	0.038	0.030	0.032	0.033	0.045	0.035	0.041
	EB	0.063	0.052	0.049	0.038	0.042	0.036	0.053	0.041	0.055
	SUM	0.069	0.06	0.057	0.054	0.052	0.045	0.063	0.049	0.064
	COM	0.053	0.051	0.048	0.036	0.039	0.044	0.058	0.042	0.050
$q = 5$										
$n = 100$	MAX	0.032	0.029	0.024	0.024	0.030	0.020	0.031	0.037	0.029
	EB	0.048	0.050	0.055	0.056	0.058	0.061	0.061	0.054	0.045
	SUM	0.046	0.049	0.048	0.063	0.059	0.064	0.063	0.052	0.045
	COM	0.038	0.037	0.031	0.043	0.047	0.041	0.048	0.048	0.026
$n = 200$	MAX	0.030	0.031	0.030	0.037	0.033	0.033	0.034	0.032	0.030
	EB	0.068	0.051	0.047	0.045	0.052	0.046	0.058	0.066	0.057
	SUM	0.067	0.051	0.049	0.049	0.054	0.045	0.070	0.068	0.061
	COM	0.048	0.045	0.036	0.040	0.040	0.037	0.051	0.049	0.049

We compare the power of the tests with $n = 100$, $p = 200$. Each entry of $\boldsymbol{z}_i \in \mathbb{R}^p$ is generated from the standard normal distribution (i.e. case (1) in Example 2). Define $\boldsymbol{\beta}_b = \kappa \cdot (\beta_{q+1}, \cdots, \beta_p)^T$. Let m denote the number of nonzero coefficients. For $m = 1, \cdots, 50$, we consider $\beta_j \sim N(0, 1), q < j \leqslant q + m$ and $\beta_j = 0$, $j > q + m$. The parameter κ is chosen so that

$||\boldsymbol{\beta}_b||^2 = 0.5$. As we see from the plots, EB performs similarly to SUM. When the number of nonzero coefficients is small, MAX is more powerful than EB and SUM. In contrast, when the number of nonzero coefficients is large, EB and SUM outperform MAX. Once again, the proposed COM has same power as MAX in the sparse case, and has similar performance to EB and SUM in the non-sparse case. As we mentioned earlier, the results show the benefits of COM, as the true model is usually unknown in practical applications. The proposed COM provides good testing power in all cases.

Figure S2: Power versus the number of nonzero coefficients for Example 2. The x-axis m denotes the number of non-zero coefficients; the y-axis is the empirical power.



S2.3 Search Engine Data

We now use the data from Lan et al. (2014) to make a case study on the regression coefficient test. The data set is obtained from an online mobile phone retailer. It contains a total of $n = 98$ daily records. The response Y is the revenue from the retailer's online sales. The explanatory variable V stands for the advertising spending on each of $p = 164$ different keywords that were bid for Baidu, the leading search engine in China. We sort these explanatory variables by the correlation with the response, from high to low, and denote $V_1, V_2, \cdots$, etc. Since the sales vary with each day of the week, we introduce a 6-dimension indictor variables W to represent Sunday to Friday. So there are 170 explanatory variables $X = (W, V)$ in our model. We will analyze it via the theory established in Section S2.

For different values of k, set $\boldsymbol{X}_a^{(k)} = (W, V_1, \cdots, V_k)^T$ and the rest variables $\boldsymbol{X}_b^{(k)} = (V_{k+1}, \cdots, V_p)^T$. We consider the linear model $\boldsymbol{Y} = (\boldsymbol{X}_a^{(k)})^T \boldsymbol{\beta}_a^{(k)} + (\boldsymbol{X}_b^{(k)})^T \boldsymbol{\beta}_b^{(k)} + \boldsymbol{\varepsilon}$, to test whether advertising spending on the rest of keywords $X_b^{(k)}$ could provide a significant contribution to online sales, conditional on the effect of X_{ak}, i.e. we test $H_0 : \boldsymbol{\beta}_b^{(k)} = \mathbf{0}$. We adopt the tests introduced in Example 2 in Section S2.2, i.e. MAX, SUM, EB and COM.

Table S3 reports the p-values of each tests with different k, which con-

Table S3: p-values of each test in Search Engine Data.

	MAX	SUM	EB	COM
$k = 0$	0.0056	0.1432	0.1100	0.0112
$k = 1$	0.0025	0.2295	0.1776	0.0050
$k = 2$	0.2083	0.1319	0.2464	0.2465
$k = 3$	0.1508	0.1387	0.2877	0.2583

trols the sparsity of the true model. The significant level is set to be $\alpha = 0.05$. From the results, we see that there are two keywords that are significant to the response (revenue) because both MAX and COM reject the null hypothesis as long as $k < 2$ (note that these are two powerful methods for sparse model). When $k \geqslant 2$, all these four tests do not reject the null hypothesis, suggesting that the rest of keywords are not significant to the response. Notice that when $k < 2$, SUM and EB fail to reject the null hypothesis at the significant level 0.05, which shows their poor performance with sparse model, consistent with our theoretical claims and simulation results. On the contrary, COM succeeds in identifying the significant keywords in this problem, illustrating its edge over SUM and EB.

S3 Technical Proofs

S3.1 Proof of Theorem 1

Proof of Theorem 1. Let $\xi_1, \xi_2, \cdots$ be i.i.d. $N(0,1)$-distributed random variables. Let $\mathbf{\Sigma}^{1/2}$ be a non-negative definite matrix such that $\mathbf{\Sigma}^{1/2} \cdot \mathbf{\Sigma}^{1/2} = \mathbf{\Sigma}$. Then $(Z_1, \cdots, Z_p)^T$ and $\mathbf{\Sigma}^{1/2}(\xi_1, \cdots, \xi_p)^T$ have the same distribution. As a consequence, $Z_1^2 + \cdots + Z_p^2$ has the same distribution as that of

$$(\xi_1, \cdots, \xi_p)\mathbf{\Sigma}^{1/2} \cdot \mathbf{\Sigma}^{1/2}(\xi_1, \cdots, \xi_p)^T = (\xi_1, \cdots, \xi_p)\mathbf{\Sigma}(\xi_1, \cdots, \xi_p)^T. \quad \text{(S16)}$$

Let $\lambda_{p,1}, \lambda_{p,2}, \cdots, \lambda_{p,p}$ be the eigenvalues of $\mathbf{\Sigma} = \mathbf{\Sigma}_p$ and $\mathbf{O}$ be a $p \times p$ orthogonal matrix such that $\mathbf{\Sigma}_p = \mathbf{O}^T \operatorname{diag}(\lambda_{p1}, \cdots, \lambda_{pp})\mathbf{O}$. In particular, since all of the diagonal entries of $\mathbf{\Sigma}$ are 1, we have $\lambda_{p,1} + \cdots + \lambda_{p,p} = p$. By the orthogonal invariance of normal distributions, $\mathbf{O}(\xi_1, \cdots, \xi_p)^T$ and $(\xi_1, \cdots, \xi_p)^T$ have the same distribution. By (S16), $Z_1^2 + \cdots + Z_p^2$ is equal to

$$\left[\mathbf{O}(\xi_1, \cdots, \xi_p)^T\right]^T \operatorname{diag}(\lambda_{p,1}, \cdots, \lambda_{p,p}) \left[\mathbf{O}(\xi_1, \cdots, \xi_p)^T\right],$$

and hence has the same distribution as that of $\lambda_{p,1}\xi_1^2 + \cdots + \lambda_{p,p}\xi_p^2$. It is easy to see $E(Z_1^2 + \cdots + Z_p^2) = p$ and

$$\begin{aligned} \operatorname{Var}(Z_1^2 + \cdots + Z_p^2) &= \lambda_{p,1}^2 \operatorname{Var}(\xi_1^2) + \cdots + \lambda_{p,p}^2 \operatorname{Var}(\xi_p^2) \\ &= 2\lambda_{p,1}^2 + \cdots + 2\lambda_{p,p}^2 \end{aligned}$$

$$= \quad 2 \cdot \mathrm{tr}(\mathbf{\Sigma}^2).$$

Easily, $m_k := E(|\xi_1^2 - 1|^k) < \infty$ for any $k \geqslant 1$. Note that $\lambda_{p,1}\xi_1^2 + \cdots + \lambda_{p,p}\xi_p^2$ is a sum of independent random variables. Then,

$$\begin{aligned} \frac{1}{[2 \cdot \mathrm{tr}(\mathbf{\Sigma}^2)]^{(2+\delta)/2}} \sum_{i=1}^{p} E\big|\lambda_{p,i}\xi_i^2 - E(\lambda_{p,i}\xi_i^2)\big|^{2+\delta} &\leqslant \frac{m_{2+\delta}}{[\mathrm{tr}(\mathbf{\Sigma}^2)]^{(2+\delta)/2}} \cdot \sum_{i=1}^{p} \lambda_{p,i}^{2+\delta} \\ &= m_{2+\delta} \cdot \frac{\mathrm{tr}(\mathbf{\Sigma}^{2+\delta})}{[\mathrm{tr}(\mathbf{\Sigma}^2)]^{(2+\delta)/2}}, \end{aligned}$$

which goes to zero by Assumption (2.1). Therefore, by the Lyapunov central limit theorem, $(\lambda_{p,1}\xi_1^2 + \cdots + \lambda_{p,p}\xi_p^2 - p)/\sqrt{2\mathrm{tr}(\mathbf{\Sigma}^2)}$ converges weakly to $N(0,1)$ as $p \to \infty$. This implies that $(Z_1^2 + \cdots + Z_p^2 - p)/\sqrt{2\mathrm{tr}(\mathbf{\Sigma}^2)}$ converges weakly to $N(0,1)$ as $p \to \infty$. □

S3.2 Proof of Theorem 2

For a graph G, we say vertices i and j are neighbors if there is an edge between them. For a set A, we write $|A|$ for its cardinality. We first prove some lemmas.

Lemma S1. *Let $G = (V, E)$ be an undirected graph with $n = |V| \geqslant 4$ vertices. Write $V = \{v_1, \cdots, v_n\}$. Assume each vertex in V has at most q neighbors. Let G_t be the set of subgraphs of G such that each subgraph has t vertices and at least one edge. The following are true.*

(i) $|G_t| \leqslant qn^{t-1}$ for any $2 \leqslant t \leqslant n$.

(ii) Fix integer t with $2 \leqslant t \leqslant n$. Let $G'_t \subset G_t$ such that each member of G'_t is a clique, that is, any two vertices are neighbors. Then $|G'_t| \leqslant nq^{t-1}$.

The following conclusions are true for integer t with $3 \leqslant t \leqslant n$.

(iii) For $j = 2, \cdots, t-1$, let H_j be the subset of $(i_1, \cdots, i_t)$ from G_t satisfying the following: there exists a subgraph S of $\{i_1, \cdots, i_t\}$ with $|S| = j$ and without any edge such that any vertex from $\{i_1, \cdots, i_t\} \backslash S$ has at least two neighbors in S. Then $|H_j| \leqslant (qt)^{t-j+1} n^{j-1}$.

(iv) For $j = 2, \cdots, t-1$, let H'_j be the subset of $(i_1, \cdots, i_t)$ from G_t satisfying the following: for any subgraph S of $\{i_1, \cdots, i_t\}$ with $|S| = j$ and without any edge, we know any vertex from $\{i_1, \cdots, i_t\} \backslash S$ has at least one neighbor in S. Then $|H'_j| \leqslant (qt)^{t-j} n^j$.

Proof of Lemma S1. (i). Choose one vertex from V and choose one of its neighbors. The total number of ways to do this is nq. The total number of ways to fill the rest of $t-2$ vertices arbitrarily is no more than n^{t-2}. Hence, $|G_t| \leqslant nq \cdot n^{t-2} = qn^{t-1}$.

(ii). To form a clique from G_t, we first choose a vertex with n ways. The next vertex has to be one of its q neighbors, the third vertex has to be one of the neighbors of the first two vertices at the same time. Thus the number of choices for the third vertex is no more than q. This has to be true for the picks of the remaining vertices to form a clique. So $|G'_t| \leqslant nq^{t-1}$.

(iii). Now we figure out the ways to get $(i_1, \cdots, i_t) \in H_t$. The number of ways to get i_1 is at most n. Once i_1 is chosen, another vertex b_1 from its neighbors has to be picked to be an element in $\{i_1, \cdots, i_t\}\backslash S$. Once b_1 is taken, since at leat two members from $(i_1, \cdots, i_t)$ are neighbors of b_1, a third vertex i_2 has to be in S. Keep in mind that i_2 has to be a neighbor of b_1. Thus, the total number of ways to pick these three vertices is at most $n \cdot q \cdot q$. The rest of $j-2$ vertices in S have at most n^{j-2} choices to satisfy the requirement; the rest $t-j-1$ vertices from $\{i_1, \cdots, i_t\}\backslash S$ have to be the neighbors of the vertices in S, which amounts to at most $(qt)^{t-j-1}$ ways to fill the $t-j$ vertices. So $|H_t| \leqslant nq^2 \cdot n^{j-2} \cdot (qt)^{t-j-1} = (qt)^{t-j+1}n^{j-1}$.

(iv). The choices of S with $|S| = j$ is no more than n^j. Any of the rest $t-j$ vertices from $(i_1, \cdots, i_t)$ must be the neighbor of a vertex from the chosen j vertices. The total number of neighbors is at most qt. This amounts to no more than $(qt)^{t-j}$ ways to achieve this. Hence $|H_j'| \leqslant (qt)^{t-j}n^j$. □

Lemma S2. *For $p \geqslant 1$, let $\varpi = \varpi_p$ be positive integers with $\lim_{p\to\infty} \varpi_p/p = 1$. Let $Z_{p1}, \cdots, Z_{p\varpi}$ be $N(0,1)$-distributed random variables with covariance matrix $\mathbf{\Sigma} = \mathbf{\Sigma}_p = (\sigma_{ij})_{p\times p}$. Assume $|\sigma_{ij}| \leqslant \delta_p$ for all $1 \leqslant i < j \leqslant \varpi$ and $p \geqslant 1$, where $\{\delta_p;\, p \geqslant 1\}$ are constants satisfying $0 < \delta_p = o(1/\log p)$. Given $x \in \mathbb{R}$, set $z = \big(2\log p - \log\log p + x\big)^{1/2}$. Then, for any fixed $m \geqslant 1$, we* *have*

$$\Big(\frac{\sqrt{\pi}p}{e^{-x/2}}\Big)^m \cdot P(|Z_{pi_1}| > z, \cdots, |Z_{pi_m}| > z) \to 1$$

as $p \to \infty$ *uniformly for all* $1 \leqslant i_1 < \cdots < i_m \leqslant \varpi$.

Proof of Lemma S2. To ease notation, we write "$Z_1, \cdots, Z_\varpi$" for "$Z_{p1}, \cdots, Z_{p\varpi}$" if there is no danger of confusion. First, z is well-defined as p is large. Note that $(Z_1, \cdots, Z_m)^T \sim N(\mathbf{0}, \mathbf{\Sigma}_m)$, where $\mathbf{\Sigma}_m = (\sigma_{ij})_{1\leqslant i,j\leqslant m}$. Recall the density function of $N(\mathbf{0}, \mathbf{\Sigma}_m)$ is given by $\frac{1}{(2\pi)^{m/2}\det(\mathbf{\Sigma}_m)^{1/2}} \exp\big(-\frac{1}{2}x^T\mathbf{\Sigma}_m^{-1}x\big)$ for all $x := (x_1, \cdots, x_m)^T \in \mathbb{R}^m$. It follows that

$$\begin{aligned} & P(Z_1 > z, \cdots, Z_m > z) \\ = & \frac{1}{(2\pi)^{m/2}\det(\mathbf{\Sigma}_m)^{1/2}} \int_z^\infty \cdots \int_z^\infty \exp\Big(-\frac{1}{2}x^T\mathbf{\Sigma}_m^{-1}x\Big)\, dx_1 \cdots dx_m. \end{aligned} \tag{S17}$$

For two non-negative definite matrices $\mathbf{A}$ and $\mathbf{B}$, we write $\mathbf{A} \leqslant \mathbf{B}$ if $\mathbf{B} - \mathbf{A}$ is also non-negative definite. We need to understand $\mathbf{\Sigma}_m^{-1}$ and $\det(\mathbf{\Sigma}_m)$ on the right hand side of (S17). First we claim that

$$(1 - 2m\delta_p)\mathbf{I}_m \leqslant \mathbf{\Sigma}_m^{-1} \leqslant (1 + 2m\delta_p)\mathbf{I}_m \tag{S18}$$

and that

$$1 - m!\delta_p \leqslant \det(\mathbf{\Sigma}_m)^{-1/2} \leqslant 1 + m!\delta_p \tag{S19}$$

as p is sufficiently large.

In fact, by the Gershgorin disc theorem (see, e.g., Horn and Johnson (2012)), all eigenvalues of $\mathbf{\Sigma}_m$ have to be in the set

$$\bigcup_{1 \leqslant i \leqslant m} \Big(\sigma_{ii} - \sum_{j \neq i} \sigma_{ij}, \sigma_{ii} + \sum_{j \neq i} \sigma_{ij} \Big).$$

By assumption, $\sigma_{ii} = 1$ and $|\sigma_{ij}| \leqslant \delta_p$ for all $1 \leqslant i < j \leqslant \varpi$. Thus, all of the eigenvalues of $\mathbf{\Sigma}_m$ are between $1 - m\delta_p$ and $1 + m\delta_p$. The two bounds are positive as p is sufficiently large. On the other hand, $(1 - m\delta_p)^{-1} \leqslant 1 + 2m\delta_p$ and $(1 + m\delta_p)^{-1} \geqslant 1 - 2m\delta_p$ as p is sufficiently large. The assertion (S18) is obtained.

Second, $\det(\mathbf{\Sigma}_m)$ is the sum of $m!$ terms. The term as the product of the diagonal entries of $\mathbf{\Sigma}_m$ is 1; each of the remaining $m! - 1$ terms is the product of m entries from which at least one is an off-diagonal entry. Therefore, $|\det(\mathbf{\Sigma}_m) - 1| \leqslant m!\delta_p$. This implies $1 - m!\delta_p \leqslant \det(\mathbf{\Sigma}_m) \leqslant 1 + m!\delta_p$. Trivially, $(1 + u)^{-1/2} = 1 - \frac{u}{2}(1 + O(u))$ as $u \to 0$. The statement (S19) is confirmed.

The claim (S18) implies that $(1 - 2m\delta_p)|x|^2 \leqslant x^T \mathbf{\Sigma}_m^{-1} x \leqslant (1 + 2m\delta_p)|x|^2$ for each $x \in \mathbb{R}^m$. The tail bound of Gaussian variable gives $P(N(0,1) \geqslant t) \sim \frac{1}{\sqrt{2\pi}\, t} e^{-t^2/2}$ as $t \to \infty$. Since $P(N(0, \sigma^2) \geqslant t) = P(N(0,1) \geqslant t/\sigma)$ for any $\sigma > 0$, for any $\epsilon \in (0, 1/2)$, there exists $t_0 > 0$ such that

$$(1 - \epsilon) \frac{\sigma}{\sqrt{2\pi}\, t} e^{-t^2/(2\sigma^2)} \leqslant P(N(0, \sigma^2) \geqslant t) \leqslant (1 + \epsilon) \frac{\sigma}{\sqrt{2\pi}\, t} e^{-t^2/(2\sigma^2)} \quad \text{(S20)}$$

for all $t \geqslant \sigma t_0$. Thus, from (S17)-(S20), we upper bound the probability of interest by

$$\begin{aligned} & P(Z_1 > z, \cdots, Z_m > z) \\ \leqslant\ & \frac{1+m!\delta_p}{(2\pi)^{m/2}} \int_z^\infty \cdots \int_z^\infty \exp\Big[-\frac{(1-2m\delta_p)}{2}(x_1^2+\cdots+x_m^2)\Big]\, dx_1\cdots dx_m \\ =\ & (1+m!\delta_p)\sigma^m \cdot P\big(N(0,\sigma^2) \geqslant z\big)^m \\ \leqslant\ & (1+m!\delta_p)\sigma^{2m}(1+\epsilon)^m \cdot \Big[\frac{1}{\sqrt{2\pi}\, z} e^{-(1-2m\delta_p)z^2/2}\Big]^m \end{aligned}$$

as $z \geqslant \sigma t_0$, where $\sigma := (1-2m\delta_p)^{-1/2}$. Now

$$\begin{aligned} \frac{1}{z} e^{-(1-2m\delta_p)z^2/2} &= e^{m\delta_p z^2} \cdot \frac{1}{z} e^{-z^2/2} \\ &= e^{o(1)} \cdot \frac{1+o(1)}{\sqrt{2\log p}} \cdot \exp\Big[-\frac{1}{2}(2\log p - \log\log p + x)\Big] \\ &= [1+o(1)] \cdot \frac{e^{-x/2}}{\sqrt{2}p} \end{aligned}$$

as $p \to \infty$ since $\delta_p = o(1/\log p)$, where the last $o(1)$ depends on m and p. Consequently,

$$P(Z_1 > z, \cdots, Z_m > z) \leqslant \Big(\frac{e^{-x/2}}{2\sqrt{\pi}p}\Big)^m C_m$$

where

$$C_m := (1+m!\delta_p)\sigma^{2m}(1+\epsilon)^m[1+o(1)]^m \leqslant (1+2\epsilon)^m$$

as p is sufficiently large because $\delta_p \to 0$. In summary, for fixed $m \geqslant 1$,

$$P(Z_1 > z, \cdots, Z_m > z) \leqslant (1+2\epsilon)^m \Big(\frac{e^{-x/2}}{2\sqrt{\pi}p}\Big)^m \qquad \text{(S21)}$$

as p is sufficiently large. Similarly, from (S17)-(S20), a lower bound can be established as

$$
\begin{aligned}
& P(Z_1 > z, \cdots, Z_m > z) \\
\geqslant\ & \frac{1 - m!\delta_p}{(2\pi)^{m/2}} \int_z^{\infty} \cdots \int_z^{\infty} \exp\Big(- \frac{(1 + 2m\delta_p)}{2} (x_1^2 + \cdots + x_m^2) \Big)\, dx_1 \cdots dx_m \\
=\ & (1 - m!\delta_p)\sigma_1^m \cdot P\big(N(0, \sigma_1^2) \geqslant z\big)^m \\
\geqslant\ & (1 - m!\delta_p)\sigma_1^{2m}(1 - \epsilon)^m \cdot \Big[\frac{1}{\sqrt{2\pi}\, z} e^{-(1+2m\delta_p)z^2/2} \Big]^m
\end{aligned}
$$

where $\sigma_1 := (1 + 2m\delta_p)^{-1/2}$. By taking care of each term above as in the previous arguments, we can get a reverse inequality of (S21) with "$1 + 2\epsilon$" replaced by "$1 - 2\epsilon$". Consequently, we know that

$$
\Big(\frac{e^{-x/2}}{2\sqrt{\pi p}} \Big)^m \cdot (1 - 2\epsilon)^m \leqslant P(Z_1 > z, \cdots, Z_m > z) \leqslant \Big(\frac{e^{-x/2}}{2\sqrt{\pi p}} \Big)^m \cdot (1 + 2\epsilon)^m \tag{S22}
$$

as $p \geqslant p(m, \epsilon)$, where $p(m, \epsilon) \geqslant 1$ is a constant that depends on m and ϵ only. Now we consider the decomposition

$$
P(|Z_1| > z, \cdots, |Z_m| > z) = \sum P(\eta_1 Z_1 > z, \cdots, \eta_m Z_m > z), \tag{S23}
$$

where the summation is over all the 2^m possible cases with $\eta_1 = \pm 1, \cdots, \eta_m = \pm 1$. Notice that $|\mathrm{Cov}(\eta_i Z_i, \eta_j Z_j)| = |\mathrm{Cov}(Z_i, Z_j)| \leqslant \delta_p$, and the derivation of (S22) depends on δ_p rather than the exact values of σ_{ij}'s. By (S23), we have

$$\Big(\frac{e^{-x/2}}{\sqrt{\pi}p}\Big)^m \cdot (1-2\epsilon)^m \leqslant P(|Z_1| > z, \cdots, |Z_m| > z) \leqslant \Big(\frac{e^{-x/2}}{\sqrt{\pi}p}\Big)^m \cdot (1+2\epsilon)^m$$

as $p \geqslant p(m, \epsilon)$. Based on the same reasoning, for any $\epsilon \in (0, 1/2)$, we have

$$\Big(\frac{e^{-x/2}}{\sqrt{\pi}p}\Big)^m \cdot (1-2\epsilon)^m \leqslant P(|Z_{i_1}| > z, \cdots, |Z_{i_m}| > z) \leqslant \Big(\frac{e^{-x/2}}{\sqrt{\pi}p}\Big)^m \cdot (1+2\epsilon)^m$$

as $p \geqslant p(m, \epsilon)$ uniformly for all $1 \leqslant i_1 < \cdots < i_m \leqslant R$. Because the probability above does is independent of ϵ, by letting $p \to \infty$ and $\epsilon \downarrow 0$, we have that

$$\Big(\frac{\sqrt{\pi}p}{e^{-x/2}}\Big)^m \cdot P(|Z_{i_1}| > z, \cdots, |Z_{i_m}| > z) \to 1$$

as $p \to \infty$ uniformly for all $1 \leqslant i_1 < \cdots < i_m \leqslant \varpi$, which completes the proof. □

For any $m \times m$ symmetric matrix $\mathbf{M}$, we use the notation $\|\mathbf{M}\|$ to denote its spectral norm, that is, $\|\mathbf{M}\| = \max\{\lambda_{max}(\mathbf{M}), -\lambda_{min}(\mathbf{M})\}$. Evidently, $\|\mathbf{M}\|^2 \leqslant \operatorname{tr}(\mathbf{M}^T\mathbf{M})$. Also, $\|\mathbf{M}\| = \lambda_{max}(\mathbf{M})$ if $\mathbf{M}$ is non-negative definite. Obviously, for any symmetric matrix $\mathbf{M}$, if $\|\mathbf{M}\| \leqslant a$ then $x^T\mathbf{M}x \leqslant \lambda_{max}(\mathbf{M})x^Tx \leqslant ax^Tx$ for any $x \in \mathbb{R}^m$. The same argument applies to $-\mathbf{M}$. Thus, by symmetry we know that

$$-ax^Tx \leqslant x^T\mathbf{M}x \leqslant ax^Tx \tag{S24}$$

for any $x \in \mathbb{R}^m$. The following lemma would be further needed.

Lemma S3. *Let $m \geqslant 2$ and $Z_1, \cdots, Z_m$ be $N(0,1)$-distributed random variables with positive definite covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})_{m\times m}$. For some $1 > \varrho > \delta > 0$, assume $|\sigma_{12}| \leqslant \varrho$, and $|\sigma_{ij}| \leqslant \delta$ for all $1 \leqslant i < j \leqslant m$ but $(i,j) \neq (1,2)$. Then, if $\delta \leqslant \frac{1}{8m^2}(1-\varrho)^3$, we have*

$$P(|Z_1| > z, \cdots, |Z_m| > z) \leqslant \frac{2^m}{z} \cdot e^{-\alpha z^2/2} \tag{S25}$$

for all $z > 0$, where $\alpha = m - \frac{1}{4}(\varrho + 3)$.

Proof of Lemma S3. Let $\mathbf{1} = (1, \cdots, 1)^T \in \mathbb{R}^m$ and

$$\mathbf{a} = (a_1, \cdots, a_m)^T = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{S26}$$

Then $\mathbf{1}^T\mathbf{a} = 1$. We claim that (which will be prove later)

$$a_i \geqslant 0 \tag{S27}$$

for all $i = 1, 2, \cdots, m$. Assuming this is true, then obviously,

$$P(Z_1 > z, \cdots, Z_m > z) \leqslant P((Z_1, \cdots, Z_m)\mathbf{a} \geqslant z).$$

Let $Y = (Z_1, \cdots, Z_m)\mathbf{a}$. Then $Y \sim N(0, \mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a})$. By (S26), $\mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a} = (\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}$. Therefore,

$$\begin{aligned} P(Z_1 > z, \cdots, Z_m > z) \leqslant & P\Big(N(0,1) \geqslant (\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^{1/2}z\Big) \\ \leqslant & \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^{1/2}z} \cdot e^{-(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})z^2/2} \end{aligned} \tag{S28}$$

for any $z > 0$, where in the last step we use a well-known inequality of the Gaussian tail: $P(N(0,1) \geqslant y) \leqslant \frac{1}{\sqrt{2\pi}\,y} e^{-y^2/2}$ for any $y > 0$. Firstly, if $m = 2$, then

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{1-\sigma_{12}^2} \begin{pmatrix} 1 & -\sigma_{12} \\ -\sigma_{12} & 1 \end{pmatrix}.$$

It is easy to check that $\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = \frac{2}{1+\sigma_{12}} \geqslant \frac{2}{1+\varrho} \geqslant 2 - \frac{1}{4}(\varrho + 3) = \alpha$. Notice $\alpha > 1$. We have from (S28) that

$$P(Z_1 > z, Z_2 > z) \leqslant \frac{1}{z} \cdot e^{-\alpha z^2/2}. \tag{S29}$$

So the conclusion holds for $m = 2$. From now on, we assume $m \geqslant 3$.

Step 1: the proof of (S27). Define

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{A} = \begin{pmatrix} \boldsymbol{\Sigma}_2 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{I}_{m-2} \end{pmatrix}$$

where $\mathbf{0}$ is a $2 \times (m-2)$ matrix whose entries are all equal to zero. Trivially, the eigenvalues of $\boldsymbol{\Sigma}_2$ are $1 + \sigma_{12}$ and $1 - \sigma_{12}$, respectively. Basic algebra gives

$$\boldsymbol{\Sigma}_2^{-1} = \frac{1}{1-\sigma_{12}^2} \begin{pmatrix} 1 & -\sigma_{12} \\ -\sigma_{12} & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{A}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-2} \end{pmatrix}, \tag{S30}$$

and the eigenvalues of $\mathbf{A}^{-1}$ are 1 with $m-2$ folds, $\frac{1}{1+\sigma_{12}}$ and $\frac{1}{1-\sigma_{12}}$, respectively, which by assumption bounds the spectral norm as $\|\mathbf{A}^{-1}\| \leqslant \frac{1}{1-\varrho}$. Also,

$\|\mathbf{A}-\mathbf{\Sigma}\|^2 \leqslant \mathrm{tr}[(\mathbf{A}-\mathbf{\Sigma})^T(\mathbf{A}-\mathbf{\Sigma})] \leqslant m^2\delta^2$ since $|\sigma_{ij}| \leqslant \delta$ for all $1 \leqslant i < j \leqslant m$ but $(i,j) \neq (1,2)$. By the fact that $\mathbf{\Sigma}^{-1}-\mathbf{A}^{-1} = \mathbf{A}^{-1}(\mathbf{A}-\mathbf{\Sigma})\mathbf{\Sigma}^{-1}$, we obtain

$$\|\mathbf{\Sigma}^{-1} - \mathbf{A}^{-1}\| \leqslant \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A} - \mathbf{\Sigma}\| \cdot \|\mathbf{\Sigma}^{-1}\| \leqslant \frac{m\delta}{1-\varrho} \cdot \|\mathbf{\Sigma}^{-1}\|. \tag{S31}$$

In particular, this implies from the triangle inequality that

$$\|\mathbf{\Sigma}^{-1}\| \leqslant \frac{1}{1-\varrho} + \frac{m\delta}{1-\varrho} \cdot \|\mathbf{\Sigma}^{-1}\|.$$

By assumption $\delta \leqslant \frac{1}{8m^2}(1-\varrho)^3$, we know $m\delta + \varrho < (1-\varrho) + \varrho = 1$. By solving the inequality we obtain $\|\mathbf{\Sigma}^{-1}\| \leqslant (1-\varrho-m\delta)^{-1} \leqslant \frac{2}{1-\varrho}$. Substituting this to (S31) we get

$$\|\mathbf{\Sigma}^{-1} - \mathbf{A}^{-1}\| \leqslant \frac{2m\delta}{(1-\varrho)^2}. \tag{S32}$$

From (S26), we know $a_i = \mathbf{e}_i^T\mathbf{\Sigma}^{-1}\mathbf{1}/\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1}$ for $1 \leqslant i \leqslant m$, where $\mathbf{e}_i = (0, \cdots, 1, \cdots, 0)^T$ and the only "1" appears in the ith position. As $\mathbf{\Sigma}$ is positive definite, $\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1} > 0$. To show (S27), it is enough to show $\tilde{a}_i := \mathbf{e}_i^T\mathbf{\Sigma}^{-1}\mathbf{1} > 0$ for each i. Define $h_i = \mathbf{e}_i^T\mathbf{A}^{-1}\mathbf{1}$, which equals the ith row sum of $\mathbf{A}^{-1}$. We have

$$\begin{aligned}
|\tilde{a}_i - h_i| = |\mathbf{e}_i^T(\mathbf{\Sigma}^{-1} - \mathbf{A}^{-1})\mathbf{1}| &\leqslant \|\mathbf{e}_i\| \cdot \|(\mathbf{\Sigma}^{-1} - \mathbf{A}^{-1})\mathbf{1}\| \\
&\leqslant \|\mathbf{e}_i\| \cdot \|\mathbf{\Sigma}^{-1} - \mathbf{A}^{-1}\| \cdot \|\mathbf{1}\| \\
&\leqslant \frac{2m^2\delta}{(1-\varrho)^2}
\end{aligned} \tag{S33}$$

induced by (S32). Therefore, $\tilde{a}_i \geqslant h_i - \frac{2m^2\delta}{(1-\varrho)^2}$. Now observe from (S30) that

$$h_1 = h_2 = \frac{1-\sigma_{12}}{1-\sigma_{12}^2} = \frac{1}{1+\sigma_{12}} \geqslant \frac{1}{1+\varrho},$$

and $h_i = 1$ for $i = 3, \cdots, m$. Thus, (S33) and condition $\delta \leqslant \frac{1}{8m^2}(1-\varrho)^3 \leqslant \frac{1}{2m^2}\frac{(1-\varrho)^3}{1+\varrho}$ conclude that $\tilde{a}_i > 0$ for each i, which implies (S27).

Step 2: the proof of (S25). From (S24) and (S32),

$$\mathbf{1}^T(\mathbf{\Sigma}^{-1} - \mathbf{A}^{-1})\mathbf{1} \geqslant -\frac{(2m\delta)\mathbf{1}^T\mathbf{1}}{(1-\varrho)^2} = -\frac{2m^2\delta}{(1-\varrho)^2}.$$

As a result, we have

$$\begin{aligned} \mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1} &\geqslant \mathbf{1}^T\mathbf{A}^{-1}\mathbf{1} - \frac{2m^2\delta}{(1-\varrho^2)^2} \\ &= \frac{2-2\sigma_{12}}{1-\sigma_{12}^2} + m - 2 - \frac{2m^2\delta}{(1-\varrho)^2} \\ &\geqslant m - 1 + \frac{1-\varrho}{1+\varrho} - \frac{2m^2\delta}{(1-\varrho)^2} \end{aligned}$$

by using the assumption $|\sigma_{12}| \leqslant \varrho$. By assumption $\delta \leqslant \frac{1}{8m^2}(1-\varrho)^3$, we further obtain

$$\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1} \geqslant m - 1 + \frac{1-\varrho}{2(1+\varrho)}.$$

In particular, $\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1} \geqslant m-1 \geqslant \frac{1}{4}m$ for $m \geqslant 3$. Then, we can establish from (S28) that

$$P(Z_1 > z, \cdots, Z_m > z) \leqslant \frac{1}{z\sqrt{m}} \exp\Big\{-\frac{z^2}{2}\Big(m - 1 + \frac{1-\varrho}{2(1+\varrho)}\Big)\Big\}, \quad \text{(S34)}$$

under the assumption $\delta \leqslant \frac{1}{8m^2}(1-\varrho)^3$. In the above derivation, the true values of σ_{ij}'s are not used, instead their bounds ϱ and δ are relevant.

Therefore, (S29) and (S34) still hold if each "Z_i" is replaced by "$\eta_i Z_i$" with $\eta_i = \pm 1$. Trivially, $-1 + \frac{1-\varrho}{2(1+\varrho)} \geqslant -\frac{1}{4}(\varrho + 3)$ for each $\varrho \in (0,1)$. This combining with (S34) and (S23) yields (S25). □

To prove Theorem 2, we need a notation. Let $p \geqslant 2$ and $(\sigma_{ij})_{p\times p}$ be a non-negative definite matrix. For $\delta > 0$ and a set $A \subset \{1, 2, \cdots, m\}$ with $2 \leqslant m \leqslant p$, define

$$\wp(A) = \max\Big\{|S|;\ S \subset A \text{ and } \max_{i\in S, j\in S, i\neq j} |\sigma_{ij}| \leqslant \delta\Big\}. \tag{S35}$$

Specifically, $\wp(A)$ takes possible values $0, 2\cdots, |A|$, where we regard $|\varnothing| = 0$. If $\wp(A) = 0$, then $|\sigma_{ij}| > \delta$ for all $i \in A$ and $j \in A$.

Proof of Theorem 2. For any $x \in \mathbb{R}$, write

$$z = \big(2\log p - \log\log p + x\big)^{1/2}, \tag{S36}$$

which is well defined as p is sufficiently large. We will not mention this matter again since the conclusion is valid as $p \to \infty$. It suffices to show

$$\lim_{p\to\infty} P\Big(\max_{1\leqslant i\leqslant p} |Z_i| > z\Big) = 1 - \exp\Big(-\frac{1}{\sqrt{\pi}} e^{-x/2}\Big) \tag{S37}$$

as $p \to \infty$. The proof will be divided into a few steps.

Step 1: reducing "$\{1 \leqslant i \leqslant p\}$" in (S37) *to a set of friendly indices.* First,

$$P\big(|N(0,1)| \geqslant z\big) \sim \frac{2}{\sqrt{2\pi} z} e^{-z^2/2}$$

$$
\begin{aligned}
&\sim \frac{1}{\sqrt{\pi}} \cdot \frac{1}{\sqrt{\log p}} \exp\Big\{ -\frac{1}{2}\big(2\log p - \log\log p + x\big)\Big\} \\
&\sim \frac{1}{\sqrt{\pi}} \cdot \frac{e^{-x/2}}{p} \qquad \text{(S38)}
\end{aligned}
$$

as $p \to \infty$, where for two sequences of real numbers A_p and H_p, the notion $A_p \sim H_p$ means that $A_p/H_p \to 1$ as $p \to \infty$. Immediately, a union bound implies

$$
P\Big(\max_{i\in C_p} |Z_i| > z\Big) \leqslant |C_p| \cdot P\big(|N(0,1)| \geqslant z\big) \to 0,
$$

as $p \to \infty$, where we recall the definition $C_p := \{1 \leqslant i \leqslant p;\, |B_{p,i}| \geqslant p^{\kappa}\}$ with $B_{p,i} = \big\{1 \leqslant j \leqslant p;\, |\sigma_{ij}| \geqslant \delta_p\big\}$. Further denote $D_p := \{1 \leqslant i \leqslant p;\, |B_{p,i}| < p^{\kappa}\}$. By assumption, $|D_p|/p \to 1$ as $p \to \infty$. It follows that

$$
\begin{aligned}
P\Big(\max_{i\in D_p} |Z_i| > z\Big) &\leqslant P\Big(\max_{1\leqslant i\leqslant p} |Z_i| > z\Big) \\
&\leqslant P\Big(\max_{i\in D_p} |Z_i| > z\Big) + P\Big(\max_{i\in C_p} |Z_i| > z\Big).
\end{aligned}
$$

Therefore, to prove (S37), it is enough to show

$$
\lim_{p\to\infty} P\Big(\max_{i\in D_p} |Z_i| > z\Big) = 1 - \exp\Big(-\frac{1}{\sqrt{\pi}} e^{-x/2}\Big), \qquad \text{(S39)}
$$

as $p \to \infty$ asymptotically.

Step 2: estimation of $P(\max_{i\in D_p} |Z_i| > z)$ via the inclusion-exclusion formula. Set

$$
\alpha_t = \sum P(|Z_{i_1}| > z, \cdots, |Z_{i_t}| > z) \qquad \text{(S40)}
$$

for $1 \leqslant t \leqslant p$, where the sum runs over all $i_1 \in D_p, \cdots, i_t \in D_p$ such that $i_1 < \cdots < i_t$. Then,

$$\sum_{t=1}^{2k} (-1)^{t-1} \alpha_t \leqslant P\Big(\max_{i \in D_p} |Z_i| > z \Big) \leqslant \sum_{t=1}^{2k+1} (-1)^{t-1} \alpha_t \tag{S41}$$

for any $k \geqslant 1$. We will prove next that

$$\lim_{p \to \infty} \alpha_t = \frac{1}{t!} \pi^{-t/2} e^{-tx/2} \tag{S42}$$

for each $t \geqslant 1$. Assuming this is true, let $p \to \infty$ in (S41), we have

$$\begin{aligned} \sum_{t=1}^{2k} (-1)^{t-1} \frac{1}{t!} \Big(\frac{1}{\sqrt{\pi}} e^{-x/2} \Big)^t &\leqslant \liminf_{p \to \infty} P\Big(\max_{i \in D_p} |Z_i| > z \Big) \\ &\leqslant \limsup_{p \to \infty} P\Big(\max_{i \in D_p} |Z_i| > z \Big) \leqslant \sum_{t=1}^{2k+1} (-1)^{t-1} \frac{1}{t!} \Big(\frac{1}{\sqrt{\pi}} e^{-x/2} \Big)^t \end{aligned}$$

for each $k \geqslant 1$. By letting $k \to \infty$ and using the Taylor expansion of the function $f(x) = 1 - e^{-x}$, we obtain (S39). It remains to verify (S42). Evidently, by (S38) and the assumption $|D_p|/p \to 1$, we immediately see (S42) holds as $t = 1$. Now we prove (S42) for any $t \geqslant 2$.

Recalling $D_p := \{1 \leqslant i \leqslant p;\ |B_{p,i}| < p^{\kappa}\}$, we write

$$\Big\{ (i_1, \cdots, i_t) \in (D_p)^t;\ i_1 < \cdots < i_t \Big\} = F_t \cup G_t,$$

where

$$\begin{aligned} F_t :=& \Big\{ (i_1, \cdots, i_t) \in (D_p)^t;\ i_1 < \cdots < i_t \text{ and } |\sigma_{i_r i_s}| \leqslant \delta_p \text{ for all } 1 \leqslant r < s \leqslant t \}; \\ G_t :=& \Big\{ (i_1, \cdots, i_t) \in (D_p)^t;\ i_1 < \cdots < i_t \text{ and } |\sigma_{i_r i_s}| > \delta_p \ \text{ for a pair } (r, s) \text{ with } 1 \leqslant r < s \leqslant t \Big\}. \end{aligned} \tag{S43}$$

Now, think D_p as graph with $|D_p|$ vertices, with $|D_p| \leqslant p$ and $|D_p|/p \to 1$ by assumption. Any two different vertices from D_p, say, i and j are connected if $|\sigma_{ij}| > \delta_p$. In this case we also say there is an edge between them. By the definition D_p, each vertex in the graph has at most p^κ neighbors. Replacing "n", "q" and "t" in Lemma S1(i) with "$|D_p|$", "p^κ" and "t", respectively, we have that $|G_t| \leqslant p^{t+\kappa-1}$ for each $2 \leqslant t \leqslant p$. Therefore, $\binom{|D_p|}{t} \geqslant |F_t| \geqslant \binom{|D_p|}{t} - p^{t+\kappa-1}$. Since $D_p/p \to 1$ and $\kappa = \kappa_p \to 0$ as $p \to \infty$, we know

$$\lim_{p\to\infty} \frac{|F_t|}{p^t} = \frac{1}{t!}. \tag{S44}$$

Decomposing (S40), we see that

$$\alpha_t = \sum_{(i_1,\cdots,i_t)\in F_t} P(|Z_{i_1}| > z, \cdots, |Z_{i_t}| > z) + \sum_{(i_1,\cdots,i_t)\in G_t} P(|Z_{i_1}| > z, \cdots, |Z_{i_t}| > z).$$

From Lemma S2 and (S44) we have

$$\sum_{(i_1,\cdots,i_t)\in F_t} P(|Z_{i_1}| > z, \cdots, |Z_{i_t}| > z) \to \frac{1}{t!}\Big(\frac{e^{-x/2}}{\sqrt{\pi}}\Big)^t = \frac{1}{t!}\pi^{-t/2}e^{-tx/2},$$

as $p \to \infty$. As a consequence, to derive (S42), we only need to show that

$$\sum_{(i_1,\cdots,i_t)\in G_t} P(|Z_{i_1}| > z, \cdots, |Z_{i_t}| > z) \to 0 \tag{S45}$$

as $p \to \infty$ asymptotically for each $t \geqslant 2$.

Step 3: the proof of (S45). If $t = 2$, the sum of probabilities in (S45) is bounded by $|G_2| \cdot \max_{1\leqslant i<j\leqslant p} P(|Z_i| > z, |Z_j| > z)$. By Lemma S1(i),

$|G_2| \leqslant p^{\kappa+1}$. Since $|\sigma_{ij}| \leqslant \varrho$, by Lemma S3,

$$P(|Z_i| > z, |Z_j| > z) \leqslant \exp\big[-(5-\varrho)z^2/8\big] \leqslant \frac{(\log p)^C}{p^{(5-\varrho)/4}} \tag{S46}$$

uniformly for all $1 \leqslant i < j \leqslant p$ as p is sufficiently large, where $C > 0$ is a constant not depending on p. We then know (S45) holds. The remaining job is to prove (S45) for $t \geqslant 3$.

Take $\delta = \delta_p$ in (S35) for the definition of $\wp(A)$ and compare it with G_t from (S43). To proceed, we further classify G_t into the following subsets

$$G_{t,j} = \big\{(i_1, \cdots, i_t) \in G_t;\ \wp(\{i_1, \cdots, i_t\}) = j\big\},$$

for $j = 0, 2, \cdots, t-1$. By the definition of G_t, we know that $G_t = \cup_{j=0,2,\cdots,t-1} G_{t,j}$. Since $t \geqslant 3$ is fixed, to show (S45), it suffices to prove

$$\sum_{(i_1,\cdots,i_t)\in G_{t,j}} P(|Z_{i_1}| > z, \cdots, |Z_{i_t}| > z) \to 0 \tag{S47}$$

for all $j \in \{0, 2, \cdots, t-1\}$.

Assume $(i_1, \cdots, i_t) \in G_{t,0}$, which implies $|\sigma_{i_r i_s}| > \delta_p$ for all $1 \leqslant r < s \leqslant t$ by (S35). Hence, the subgraph $\{i_1, \cdots, i_t\} \subset G_t$ is a clique. Taking $n = |D_p| \leqslant p$, $t = t$ and $q = p^\kappa$ into Lemma S(1)(ii), we get $|G_{t,0}| \leqslant p^{1+\kappa(t-1)} \leqslant p^{1+t\kappa}$. Thus, we obtain

$$(\text{S47}) \leqslant p^{1+t\kappa} \cdot \max_{1\leqslant i<j\leqslant p} P(|Z_i| > z, |Z_j| > z) \leqslant p^{1+t\kappa} \cdot \frac{(\log p)^C}{p^{(5-\varrho)/4}} \to 0 \tag{S48}$$

as $p \to \infty$ by using (S46). So, (S47) holds with $t = 0$.

Now we assume $(i_1, \cdots, i_t) \in G_{t,j}$ with $j \in \{2, \cdots, t-1\}$. By definition, there exits $S \subset \{i_1, \cdots, i_t\}$ such that $\max_{i\in S, j\in S, i\neq j} |\sigma_{ij}| \leqslant \delta_p$ and for each $k \in \{i_1, \cdots, i_t\} \backslash S$, there exists $i \in S$ satisfying $|\sigma_{ik}| > \delta_p$. Looking at the last statement we see two possibilities: (i) for each $k \in \{i_1, \cdots, i_t\} \backslash S$, there exist at leat two indices, say, $i \in S$, $j \in S$ with $i \neq j$ satisfying $|\sigma_{ik}| > \delta_p$ and $|\sigma_{jk}| > \delta_p$; (ii) there exists $k \in \{i_1, \cdots, i_t\} \backslash S$ for which $|\sigma_{ik}| > \delta_p$ for an unique $i \in S$. However, for $(i_1, \cdots, i_t) \in G_{t,j}$, (i) and (ii) could happen at the same time for different S, say, (i) holds for S_1 and (ii) holds for S_2 simultaneously. Thus, to differentiate the two cases, we consider the following two types of sets. Denote

$$\begin{aligned} H_{t,j} = &\Big\{(i_1, \cdots, i_t) \in G_{t,j};\ \text{there exist } S \subset \{i_1, \cdots, i_t\} \text{ with } |S| = j \text{ and} \\ &\max_{i\in S, j\in S, i\neq j} |\sigma_{ij}| \leqslant \delta_p \text{ such that for any } k \in \{i_1, \cdots, i_t\} \backslash S \text{ there exist } r \in S, \\ &s \in S, r \neq s \text{ satisfying } \min\{|\sigma_{kr}|, |\sigma_{ks}|\} > \delta_p\Big\}. \end{aligned} \tag{S49}$$

Replacing "n", "q" and "t" in Lemma S1(iii) with "$|D_p|$", "p^κ" and "t", respectively, we have that $|H_{t,j}| \leqslant t^t \cdot p^{j-1+(t-j+1)\kappa}$ for each $t \geqslant 3$. Analogously, set

$$\begin{aligned} H'_{t,j} = &\Big\{(i_1, \cdots, i_t) \in G_{t,j};\ \text{for any } S \subset \{i_1, \cdots, i_t\} \text{ with } |S| = j \text{ and} \\ &\max_{i\in S, j\in S, i\neq j} |\sigma_{ij}| \leqslant \delta_p \text{ there exists } k \in \{i_1, \cdots, i_t\} \backslash S \text{ such that } |\sigma_{kr}| > \delta_p \\ &\text{for a unique } r \in S\}. \end{aligned} \tag{S50}$$

From Lemma S1(iv) we see $|H'_{t,j}| \leqslant t^t \cdot p^{j+(t-j)\kappa}$. It is easy to see $G_{t,j} = H_{t,j} \cup H'_{t,j}$. Therefore, to show (S47), we only need to prove both

$$\sum_{(i_1,\cdots,i_t)\in H_{t,j}} P(|Z_{i_1}| > z, \cdots, |Z_{i_t}| > z) \to 0 \tag{S51}$$

and

$$\sum_{(i_1,\cdots,i_t)\in H'_{t,j}} P(|Z_{i_1}| > z, \cdots, |Z_{i_t}| > z) \to 0 \tag{S52}$$

as $p \to \infty$ for $j = 2, \cdots, t-1$. In fact, let S be as in (S49), then using Lemma S2, the probability in (S51) is bounded by $P(\bigcap_{l\in S}\{|Z_l| > z\}) \leqslant C \cdot p^{-j}$ uniformly for all S as p is sufficiently large, where C is a constant independent of p. This leads to

$$\sum_{(i_1,\cdots,i_t)\in H_{t,j}} P(|Z_{i_1}| > z, \cdots, |Z_{i_t}| > z) \leqslant t^t \cdot p^{j-1+(t-j+1)\kappa} \cdot \left(C \cdot p^{-j}\right) \leqslant (Ct^t) \cdot p^{-1+t\kappa},$$

as p is sufficiently large. By the assumption that $\kappa = \kappa_p \to 0$, we arrive at (S51).

Finally we validate (S52). Recall the definition of $H'_{t,j}$. For $(i_1, \cdots, i_t) \in H'_{t,j}$, pick $S \subset \{i_1, \cdots, i_t\}$ with $|S| = j$, $\max_{i\in S, j\in S, i\neq j} |\sigma_{ij}| \leqslant \delta_p$ and $k \in \{i_1, \cdots, i_t\}\backslash S$ such that $\delta_p < |\sigma_{kr}| \leqslant \varrho$ for a unique $r \in S$. Then each probability in (S52) is bounded by

$$P\Big(|Z_k| > z, \bigcap_{l\in S}\{|Z_l| > z\}\Big),$$

for $2 \leqslant j \leqslant t-1$. Taking $m = j+1$ and applying Lemma S3, the probability above is dominated by

$$\frac{2^{j+1}}{z} \cdot \exp\Big\{ -\frac{z^2}{2}\Big(j + \frac{1-\varrho}{4}\Big)\Big\} = O\Big(\frac{(\log p)^c}{p^{j+(1-\varrho)/4}}\Big),$$

for some constant c. As stated earlier, $|H'_{t,j}| \leqslant t^t \cdot p^{j+(t-j)\kappa}$. by union bound, since $\kappa = \kappa_p \to 0$, we see the sum from (S52) is of order $O(p^{-(1-\varrho)/8})$. Hence, (S52) holds. We have proved (S47) for any $j \in \{0, 2, \cdots, t-1\}$, which concludes the proof. □

S3.3 Proof of Theorem 3

The proof of Theorem 3 is involved. A preparation with a few of lemmas is given below.

Lemma S4. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be nonnegative definite matrices. Then*

$$tr\,(\mathbf{AB}) \leqslant \lambda_{max}(\mathbf{A})\, tr(\mathbf{B}).$$

Proof of Lemma S4. Assume $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices. There is an orthogonal matrix $\mathbf{O}$ such that $\mathbf{A} = \mathbf{O}^T\mathbf{\Lambda}\mathbf{O}$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$ and $\lambda_1 \geqslant \cdots \geqslant \lambda_n \geqslant 0$. Observe $\mathrm{tr}\,(\mathbf{AB}) = \mathrm{tr}\,[\mathbf{\Lambda}(\mathbf{OBO}^T)]$, $\mathrm{tr}\,(\mathbf{OBO}^T) = \mathrm{tr}(\mathbf{B})$ and $\lambda_{max}(\mathbf{A}) = \lambda_{max}(\mathbf{\Lambda})$. Thus, without loss of generality, we assume $\mathbf{A} = \mathbf{\Lambda}$. Write $\mathbf{B} = (b_{ij})$. Then, $b_{ii} \geqslant 0$ for each i and

$$\mathrm{tr}\,(\mathbf{AB}) = \sum_{i=1}^{n} \lambda_i b_{ii} \leqslant \lambda_1 \cdot \sum_{i=1}^{n} b_{ii} = \lambda_{max}(\mathbf{A})\,\mathrm{tr}(\mathbf{B}).$$

The proof is completed. □.

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

Lemma S5. *Let* $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *with* $\boldsymbol{\Sigma}$ *being invertible. Partition* $\mathbf{X}, \boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *as*

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \tag{S53}$$

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

Lemma S6. *Let* $(Z_1, \cdots, Z_p)^T \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. *Under the notion of Lemma S5, for* $1 \leqslant d < p$, *write* $\boldsymbol{X}_1 = (Z_1, \cdots, Z_d)^T$ *and* $\boldsymbol{X}_2 = (Z_{d+1}, \cdots, Z_p)^T$. *Define* $\mathbf{U}_p = \boldsymbol{X}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1$ *and* $\mathbf{V}_p = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1$. *Assume* $\boldsymbol{\Sigma} = (\sigma_{ij})_{p\times p}$ *with* $\sum_{j=1}^p \sigma_{ij}^2 \leqslant K_p$ *for each* $1 \leqslant i \leqslant p$, *where* K_p *is a constant depending on* p *only. Then there exists a constant* $C > 0$ *free of* d, p, $\boldsymbol{\Sigma}$ *and* K_p, *such that the following holds:*

(i) $Ee^{\theta \mathbf{U}_p^T \mathbf{V}_p} \leqslant \exp(dK_pJ_p\theta^2)$ *for all* $|\theta| \leqslant C/\lambda_{max}(\boldsymbol{\Sigma})$ *where* $J_p = \lambda_{max}(\boldsymbol{\Sigma})/\lambda_{min}(\boldsymbol{\Sigma})$.

(ii) $Ee^{\theta\|\mathbf{V}_p\|^2} \leqslant \exp[2dK_p\theta/\lambda_{min}(\boldsymbol{\Sigma})]$ *for all* $0 \leqslant \theta \leqslant C/\lambda_{max}(\boldsymbol{\Sigma})$.

(iii) $E\exp\left[\theta(Z_1^2 + \cdots + Z_d^2)\right] \leqslant e^{2d\theta}$ *for* $0 \leqslant \theta \leqslant C/d$.

Proof of Lemma S6. Set $k = p - d$, so $\boldsymbol{\Sigma}_{11}$ is $d \times d$, $\boldsymbol{\Sigma}_{12}$ is $d \times k$ and $\boldsymbol{\Sigma}_{22}$

is $k \times k$. Let $\xi = (\xi_1, \cdots, \xi_k)^T$ and $\eta = (\eta_1, \cdots, \eta_k)^T$, where the $2k$ random variables ξ_i's and η_i's are i.i.d. $N(0,1)$-distributed. According to Lemma S5,

$$\mathbf{U}_p \overset{d}{=} (\mathbf{\Sigma}_{22\cdot 1})^{1/2}\xi \quad \text{and} \quad \mathbf{V}_p \overset{d}{=} (\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{1/2}\eta, \tag{S54}$$

and they are independent. From (S54) we see that $\mathbf{U}_p^T\mathbf{V}_p \overset{d}{=} \xi^T\hat{\mathbf{\Sigma}}_p\eta$, where

$$\hat{\mathbf{\Sigma}}_p = (\mathbf{\Sigma}_{22\cdot 1})^{1/2} \cdot (\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{1/2}.$$

By the singular value decomposition theorem, we write $\hat{\mathbf{\Sigma}}_p = \mathbf{O}_1 \,\mathrm{diag}(\lambda_1, \cdots, \lambda_k)\, \mathbf{O}_2$ where $\lambda_1, \cdots, \lambda_k$ are the singular values of $\hat{\mathbf{\Sigma}}_p$, and $\mathbf{O}_1$ and $\mathbf{O}_2$ are orthogonal matrices. Review the well-known facts that

$$\mathrm{tr}\,(\mathbf{AB}) = \mathrm{tr}\,(\mathbf{BA}) \quad \text{and} \quad \mathrm{tr}(\mathbf{CD}) \leqslant \mathrm{tr}(\mathbf{C}) \cdot \mathrm{tr}(\mathbf{D}) \tag{S55}$$

for any matrices $\mathbf{A}$ and $\mathbf{B}$ and any non-negative definite matrices $\mathbf{C}$ and $\mathbf{D}$ (the second fact from (S55) can also be thought as a consequence of Lemma S4). Note that

$$\mathbf{\Sigma}_{22} = \mathbf{\Sigma}_{22\cdot 1} + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} \tag{S56}$$

and all three matrices are non-negative definite. This together with the Weyl interlacing inequality implies $\lambda_{max}(\mathbf{\Sigma}_{22\cdot 1}) \leqslant \lambda_{max}(\mathbf{\Sigma}_{22}) \leqslant \lambda_{max}(\mathbf{\Sigma})$. Consequently, we have from Lemma S4 that

$$\lambda_1^2 + \cdots + \lambda_k^2 = \mathrm{tr}\,(\hat{\mathbf{\Sigma}}_p\hat{\mathbf{\Sigma}}_p^T) = \mathrm{tr}\left[\mathbf{\Sigma}_{22\cdot 1} \cdot (\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})\right]$$

$$\leqslant \lambda_{max}(\mathbf{\Sigma}_{22\cdot 1}) \cdot \text{tr}\,(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})$$
$$\leqslant \lambda_{max}(\mathbf{\Sigma}) \cdot \text{tr}\,(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}).$$

Furthermore, by Lemma S4 again,

$$\begin{aligned}\text{tr}\,(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}) = \text{tr}\,(\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{21}) &\leqslant \lambda_{max}(\mathbf{\Sigma}_{11}^{-1}) \cdot \text{tr}(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{21}) \\ &= \frac{1}{\lambda_{min}(\mathbf{\Sigma}_{11})} \cdot \text{tr}(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{21}) \leqslant \frac{dK_p}{\lambda_{min}(\mathbf{\Sigma})}.\end{aligned} \tag{S57}$$

In fact in the above we use the assertion $\lambda_{min}(\mathbf{\Sigma}) \leqslant \lambda_{min}(\mathbf{\Sigma}_{11})$ by the Weyl interlacing inequality and the fact that

$$\text{tr}(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{21}) = \sum_{i=1}^{d}\sum_{j=d+1}^{p} \sigma_{ij}^2 \leqslant dK_p$$

by assumption. Combing the above, we arrive at

$$\lambda_1^2 + \cdots + \lambda_k^2 \leqslant (dK_p) \cdot \frac{\lambda_{max}(\mathbf{\Sigma})}{\lambda_{min}(\mathbf{\Sigma})}. \tag{S58}$$

Another fact we will use later on is that

$$\Lambda_1 := \max\{\lambda_1, \cdots, \lambda_k\} \leqslant \lambda_{max}(\mathbf{\Sigma}). \tag{S59}$$

In fact, recall that $\|\cdot\|$ denotes the spectral norm of a matrix. By definition,

$$\begin{aligned}\Lambda_1 =& \|(\mathbf{\Sigma}_{22\cdot 1})^{1/2} \cdot (\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{1/2}\| \\ \leqslant& \|\mathbf{\Sigma}_{22\cdot 1}\|^{1/2} \cdot \|\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\|^{1/2}.\end{aligned} \tag{S60}$$

From (S56) we know that both norms in (S60) are bounded by $\lambda_{max}(\mathbf{\Sigma}_{22})^{1/2} \leqslant \lambda_{max}(\mathbf{\Sigma})^{1/2}$. So (S59) is obtained. With these preparation, we are ready to prove (i), (ii) and (iii).

(i) Since $\mathbf{U}_p^T \mathbf{V}_p = \xi^T \hat{\mathbf{\Sigma}}_p \eta$ and $\hat{\mathbf{\Sigma}}_p = \mathbf{O}_1 \operatorname{diag}(\lambda_1, \cdots, \lambda_k) \mathbf{O}_2$, by the orthogonal invariant property of $N(\mathbf{0}, \mathbf{I}_k)$, we have that

$$\mathbf{U}_p^T \mathbf{V}_p \stackrel{d}{=} \sum_{i=1}^k \lambda_i \xi_i \eta_i. \tag{S61}$$

Review the moment generating functions of Gaussian variables,

$$E e^{\theta \xi_1} = e^{\theta^2/2} \quad \text{and} \quad E e^{\theta \xi_1^2} = (1-2\theta)^{-1/2}, \quad \theta < \frac{1}{2}. \tag{S62}$$

We can write

$$E \exp\left(\theta \mathbf{U}_p^T \mathbf{V}_p\right) = \prod_{i=1}^k E e^{\theta \lambda_i \xi_i \eta_i} = \prod_{i=1}^k E e^{(\theta \lambda_i)^2 \xi_i^2/2} = \exp\Big\{ -\frac{1}{2} \sum_{i=1}^k \log[1-(\theta\lambda_i)^2] \Big\},$$

for all $|\theta \lambda_i| < 1$ with $i = 1, \cdots, k$, that is, $|\theta| \leqslant \frac{1}{\Lambda_1}$. Notice $-\frac{1}{2}\log(1-x) \sim \frac{1}{2}x$ as $x \to 0$, Thus, there exists $x_0 \in (0, \frac{1}{2})$ such that $-\frac{1}{2}\log(1-x) \leqslant x$ for all $x \in (0, x_0^2)$. Then

$$E \exp\left(\theta \mathbf{U}_p^T \mathbf{V}_p\right) \leqslant \exp\left(\theta^2 \sum_{i=1}^k \lambda_i^2\right) \leqslant e^{dK_p J_p \theta^2} \tag{S63}$$

for all θ with $|\theta| \leqslant \frac{1}{\Lambda_1} \wedge \frac{x_0}{\Lambda_1} = \frac{x_0}{\Lambda_1}$ by (S58), with $J_p = \lambda_{max}(\mathbf{\Sigma})/\lambda_{min}(\mathbf{\Sigma})$. The inequality (S63) is particularly true if $|\theta| \leqslant \frac{x_0}{\lambda_{max}(\mathbf{\Sigma})}$ by (S59).

(ii) Let $\rho_1, \cdots, \rho_k$ be the eigenvalues of $\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$. From (S56) we have

$$\Lambda_2 := \max\{\rho_1, \cdots, \rho_k\} \leqslant \lambda_{max}(\mathbf{\Sigma}_{22}) \leqslant \lambda_{max}(\mathbf{\Sigma}). \tag{S64}$$

By (S54) and the orthogonal invariant property of normal distributions similar to (S61),

$$\|\mathbf{V}_p\|^2 \overset{d}{=} \eta^T\big(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\big)\eta \overset{d}{=} \rho_1\eta_1^2 + \cdots + \rho_k\eta_k^2. \tag{S65}$$

As shown in (S57),

$$\rho_1 + \cdots + \rho_k \leqslant \frac{dK_p}{\lambda_{min}(\mathbf{\Sigma})}. \tag{S66}$$

By (S62) and (S65), we have

$$Ee^{\theta\|\mathbf{V}_p\|^2} = \exp\Big[-\frac{1}{2}\sum_{i=1}^{d}\log(1-2\theta\rho_i)\Big]$$

for all $\theta < \frac{1}{2\Lambda_2}$. Recalling x_0 defined earlier, we conclude

$$Ee^{\theta\|\mathbf{V}_p\|^2} \leqslant e^{2\theta(\rho_1+\cdots+\rho_k)} \leqslant \exp\Big(\frac{2dK_p\theta}{\lambda_{min}(\mathbf{\Sigma})}\Big)$$

for all $0 \leqslant \theta < \frac{1}{2\Lambda_2} \wedge \frac{x_0^2}{2\Lambda_2} = \frac{x_0^2}{2\Lambda_2}$. By (S64), the above is particularly true provided $0 \leqslant \theta < \frac{x_0^2}{2\lambda_{max}(\mathbf{\Sigma})}$, which proves the claim.

(iii) By the Hölder inequality and the fact that $Z_i \sim N(0,1)$ for each i,

$$E\exp\big[\theta(Z_1^2+\cdots+Z_d^2)\big] \leqslant \Big[\big(Ee^{d\theta Z_1^2}\big)^{1/d}\Big]^d = (1-2d\theta)^{-1/2}$$

for all $\theta < 1/(2d)$. For x_0 defined earlier, we obtain

$$E\exp\big[\theta(Z_1^2+\cdots+Z_d^2)\big] \leqslant e^{2d\theta}$$

for $0 \leqslant \theta \leqslant \frac{x_0^2}{2d}$ as desired. □

Lemma S7. *Assume the same notations and conditions as in Lemma S6. Denote* $\Theta_p = \|\mathbf{V}_p\|^2 + 2\mathbf{U}_p^T\mathbf{V}_p + \sum_{i=1}^d Z_i^2$ *and* $\upsilon_p = [2tr(\mathbf{\Sigma}^2)]^{1/2}$*. Suppose Assumption* (2.3) *holds and* C *is the constant appearing in* (2.3)*. Let* $d \geqslant 1$ *and* $\epsilon > 0$ *be given. Then*

$$\epsilon_p := \frac{(\log p)^C}{\upsilon_p \lambda_{min}(\mathbf{\Sigma})} \to 0 \quad and \quad t = t_p := \frac{C\epsilon}{8} \cdot \frac{\upsilon_p}{\lambda_{max}(\mathbf{\Sigma}) \log p} \to \infty. \tag{S67}$$

Furthermore,

$$P(|\Theta_p| \geqslant \epsilon \upsilon_p) \leqslant \frac{3}{p^t} + \exp\Big(-\frac{C\epsilon}{8d}\sqrt{p}\Big)$$

for every p *satisfying* $p \geqslant \frac{256d^2}{\epsilon^2}$ *and* $\epsilon_p < \frac{\epsilon}{8(C+2)d}$.

Proof of Lemma S7. First we give an estimate for υ_p. Set $K_p = (\log p)^C$, where $C > 0$ is the constant as given in Assumption (2.3). Evidently, $\mathrm{tr}(\mathbf{\Sigma}^2) = \sum_{1 \leqslant i,j \leqslant p} \sigma_{ij}^2$. Also $\sigma_{ii} = 1$ and $\sum_{j=1}^p \sigma_{ij}^2 \leqslant K_p$ for each $1 \leqslant i \leqslant p$ by Assumption (2.3). It follows that

$$\sqrt{2p} = \Big(2\sum_{i=1}^p \sigma_{ii}^2\Big)^{1/2} \leqslant \upsilon_p \leqslant \sqrt{2p} \cdot (\log p)^C \tag{S68}$$

for $p \geqslant 3$. Now, for any $\epsilon > 0$, union bound gives

$$\begin{aligned} &P(|\Theta_p| \geqslant \epsilon \upsilon_p) \\ \leqslant &P\Big(\|\mathbf{V}_p\|^2 \geqslant \frac{1}{4}\epsilon\upsilon_p\Big) + P\Big(|\mathbf{U}_p^T\mathbf{V}_p| \geqslant \frac{1}{4}\epsilon\upsilon_p\Big) + P\Big(\sum_{i=1}^n Z_i^2 \geqslant \frac{1}{4}\epsilon\upsilon_p\Big). \end{aligned} \tag{S69}$$

Let us bound them one by one. First, by the Markov inequality and Lemma

S6(ii),

$$P\Big(\|\mathbf{V}_p\|^2 \geqslant \frac{1}{4}\epsilon\upsilon_p\Big) \leqslant \exp\Big(-\frac{\theta\epsilon}{4}\upsilon_p\Big)\cdot Ee^{\theta\|\mathbf{V}_p\|^2}$$
$$\leqslant \exp\Big[-\theta\Big(\frac{\epsilon}{4}\upsilon_p - \frac{2dK_p}{\lambda_{min}(\mathbf{\Sigma})}\Big)\Big] \tag{S70}$$

for any $0 \leqslant \theta \leqslant C/\lambda_{max}(\mathbf{\Sigma})$ where $C > 0$ is a constant free of p. By assumption, $K_p \leqslant (\log p)^C$ and $p^{-1/2}(\log p)^C \ll \lambda_{min}(\mathbf{\Sigma}) \leqslant \lambda_{max}(\mathbf{\Sigma}) \ll \sqrt{p}/\log p$. Then,

$$\frac{K_p}{\lambda_{min}(\mathbf{\Sigma})} = o(\upsilon_p) \ \text{ and } \ \frac{\upsilon_p}{\lambda_{max}(\mathbf{\Sigma})} \gg \log p, \tag{S71}$$

due to (S68). Thus, (S67) holds. Choosing $\theta = \frac{C}{\lambda_{max}(\mathbf{\Sigma})}$, we see from (S70) that

$$P\Big(\|\mathbf{V}_p\|^2 \geqslant \frac{1}{4}\epsilon\upsilon_p\Big) \leqslant \exp\Big[-\frac{C\epsilon}{8\lambda_{max}(\mathbf{\Sigma})}\upsilon_p\Big] = e^{-t\log p} = \frac{1}{p^t} \tag{S72}$$

for all p satisfying $\frac{\epsilon}{8}\upsilon_p > \frac{2dK_p}{\lambda_{min}(\mathbf{\Sigma})}$, which is particularly true if

$$\frac{K_p}{\upsilon_p\lambda_{min}(\mathbf{\Sigma})} \leqslant \epsilon_p < \frac{\epsilon}{8(C+2)d}. \tag{S73}$$

Second,

$$P\Big(|\mathbf{U}_p^T\mathbf{V}_p| \geqslant \frac{1}{4}\epsilon\upsilon_p\Big) \leqslant P\Big(\mathbf{U}_p^T\mathbf{V}_p \geqslant \frac{1}{4}\epsilon\upsilon_p\Big) + P\Big(-\mathbf{U}_p^T\mathbf{V}_p \geqslant \frac{1}{4}\epsilon\upsilon_p\Big).$$

By Lemma S6(i) and a similar argument as (S70), we have

$$P\Big(|\mathbf{U}_p^T\mathbf{V}_p| \geqslant \frac{1}{4}\epsilon\upsilon_p\Big) \leqslant 2\cdot\exp\Big[\theta\Big(-\frac{1}{4}\epsilon\upsilon_p + dK_pJ_p\theta\Big)\Big],$$

for all $|\theta| \leqslant C/\lambda_{max}(\mathbf{\Sigma})$, where $J_p = \lambda_{max}(\mathbf{\Sigma})/\lambda_{min}(\mathbf{\Sigma})$. It is trivial to see from the notation $\mathbf{\Sigma} = (\sigma_{ij})_{p\times p}$ that $\lambda_{max}(\mathbf{\Sigma}) \geqslant \sigma_{11} = 1$. Take $\theta = C/\lambda_{max}(\mathbf{\Sigma})$ to obtain

$$\begin{aligned} P\Big(|\mathbf{U}_p^T\mathbf{V}_p| \geqslant \frac{1}{4}\epsilon\upsilon_p\Big) \leqslant & 2\cdot\exp\Big[\frac{C}{\lambda_{max}(\mathbf{\Sigma})}\Big(-\frac{1}{4}\epsilon\upsilon_p + \frac{CdK_p}{\lambda_{min}(\mathbf{\Sigma})}\Big)\Big] \\ \leqslant & 2\cdot\exp\Big(-\frac{C\epsilon}{8\lambda_{max}(\mathbf{\Sigma})}\upsilon_p\Big) = \frac{2}{p^t} \end{aligned} \tag{S74}$$

for all p satisfying (S73). Finally, by the Markov inequality and by taking $\theta = C/d$ from Lemma S6(iii) we obtain

$$P\Big(\sum_{i=1}^d Z_i^2 \geqslant \frac{1}{4}\epsilon\upsilon_p\Big) \leqslant \exp\Big(-\frac{C\epsilon}{4d}\upsilon_p + 2C\Big) \leqslant \exp\Big(-\frac{C\epsilon}{8d}\sqrt{p}\Big)$$

for every p satisfying $2 < \frac{\epsilon}{8d}\sqrt{p}$, or equivalently, $p \geqslant \frac{256d^2}{\epsilon^2}$. In the last step above we use the inequality $\upsilon_p \geqslant \sqrt{p}$ from (S68). Combining this with (S69), (S72) and (S74), we conclude that

$$P(|\Theta_p| \geqslant \epsilon\upsilon_p) \leqslant \frac{3}{p^t} + \exp\Big(-\frac{C\epsilon}{8d}\sqrt{p}\Big)$$

for every p satisfying $\epsilon_p < \frac{\epsilon}{8(C+2)d}$ from (S73) and $p \geqslant \frac{256d^2}{\epsilon^2}$. □

We now introduce more general indexing. Let $(Z_1, \cdots, Z_p)^T \sim N(\mathbf{0}, \mathbf{\Sigma})$. Assume d is an integer with $1 \leqslant d < p$. For any set $\Lambda = \{i_1, \cdots, i_d\}$ with $1 \leqslant i_1 < \cdots < i_d \leqslant p$, write $\boldsymbol{X}_{1,\Lambda} = (Z_{i_1}, \cdots, Z_{i_d})^T$. Let $\boldsymbol{X}_{2,\Lambda}$ be the vector obtained with deleting $Z_{i_1}, \cdots, Z_{i_d}$ from $(Z_1, \cdots, Z_p)^T$, that is, $\boldsymbol{X}_{\Lambda,2} = (Z_{j_1}, \cdots, Z_{j_{p-d}})^T$ where $j_1 < \cdots < j_{p-d}$ and $\{j_1, \cdots, j_{p-d}\} =$

$\{1, 2, \cdots, n\} \backslash \Lambda$. Let $\mathbf{\Sigma}_\Lambda$ be the covariance matrix of

$$\mathbf{X}_\Lambda := \begin{pmatrix} \boldsymbol{X}_{1,\Lambda} \\ \boldsymbol{X}_{2,\Lambda} \end{pmatrix}.$$

Partition $\mathbf{\Sigma}_\Lambda$ similar to (S53) such that

$$\mathbf{\Sigma}_\Lambda = \begin{pmatrix} \mathbf{\Sigma}_{11,\Lambda} & \mathbf{\Sigma}_{12,\Lambda} \\ \mathbf{\Sigma}_{21,\Lambda} & \mathbf{\Sigma}_{22,\Lambda} \end{pmatrix}.$$

In particular, $\boldsymbol{X}_{1,\Lambda} \sim N(\mathbf{0}, \mathbf{\Sigma}_{11,\Lambda})$ and $\boldsymbol{X}_{2,\Lambda} \sim N(\mathbf{0}, \mathbf{\Sigma}_{22,\Lambda})$. We have the following result.

Lemma S8. *Let $\mathbf{X} = (Z_1, \cdots, Z_p)^T \sim N(\mathbf{0}, \mathbf{\Sigma})$. Assume d is an integer with $1 \leqslant d < p$. For any set $\Lambda = \{i_1, \cdots, i_d\}$ with $1 \leqslant i_1 < \cdots < i_d \leqslant p$, we define $\mathbf{U}_{p,\Lambda} = \boldsymbol{X}_{2,\Lambda} - \mathbf{\Sigma}_{21,\Lambda}\mathbf{\Sigma}_{11,\Lambda}^{-1}\boldsymbol{X}_{1,\Lambda}$ and $\mathbf{V}_{p,\Lambda} = \mathbf{\Sigma}_{21,\Lambda}\mathbf{\Sigma}_{11,\Lambda}^{-1}\boldsymbol{X}_{1,\Lambda}$. Set $\Theta_{p,\Lambda} = \|\mathbf{V}_{p,\Lambda}\|^2 + 2\mathbf{U}_{p,\Lambda}^T\mathbf{V}_{p,\Lambda} + \sum_{k=1}^d Z_{i_k}^2$ and $\upsilon_p = [2tr(\mathbf{\Sigma}^2)]^{1/2}$. Then, under Assumption (2.3), for any $\epsilon > 0$ there exists $t = t_p \to \infty$ such that*

$$\max_\Lambda P(|\Theta_{p,\Lambda}| \geqslant \epsilon\upsilon_p) \leqslant \frac{1}{p^t}$$

as p is sufficiently large, where the maximum $\Lambda = \{i_1, \cdots, i_d\}$ runs over all possible indices $i_1, \cdots, i_d$ with $1 \leqslant i_1 < \cdots < i_d \leqslant p$.

Proof of Lemma S8. View $\mathbf{X}_\Lambda$ as the vector after exchanging some rows of $\mathbf{X}$. Then there is a permutation matrix $\mathbf{O}$ such that $\mathbf{X}_\Lambda = \mathbf{O}\mathbf{X}$. Therefore the covariance matrix of $\mathbf{X}_\Lambda$ is $\mathbf{\Sigma}_\Lambda = E(\mathbf{O}\mathbf{X}(\mathbf{O}\mathbf{X})^T) = \mathbf{O}\mathbf{\Sigma}\mathbf{O}^T$.

Set $\upsilon_{p,\Lambda} = [2\mathrm{tr}(\mathbf{\Sigma_{\Lambda}}^2)]^{1/2}$. Then

$$\lambda_{max}(\mathbf{\Sigma}_{\Lambda}) = \lambda_{max}(\mathbf{\Sigma}), \quad \lambda_{min}(\mathbf{\Sigma}_{\Lambda}) = \lambda_{min}(\mathbf{\Sigma}) \quad \text{and} \quad \upsilon_{p,\Lambda} = \upsilon_p. \tag{S75}$$

Second, $\mathbf{O\Sigma O}^T$ is the matrix by exchanging some rows and then exchanging the corresponding columns. So the entries of $\mathbf{\Sigma}_{\Lambda}$ are the same as those of $\mathbf{\Sigma}$; the sum of squares of the entries of a row from $\mathbf{O\Sigma O}^T$ is the same as that of a row from $\mathbf{\Sigma}$, and vice versa. Write $\mathbf{\Sigma}_{\Lambda} = (\sigma_{ij,\Lambda})_{p\times p}$. As a consequence,

$$\max_{1\leqslant i<j\leqslant p} |\sigma_{ij,\Lambda}| = \max_{1\leqslant i<j\leqslant p} |\sigma_{ij}|, \quad \text{and} \quad \max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} \sigma_{ij,\Lambda}^2 = \max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} \sigma_{ij}^2. \tag{S76}$$

Notice that in Assumption (2.3), all conditions are imposed on the four quantities: $\lambda_{max}(\mathbf{\Sigma})$, $\lambda_{min}(\mathbf{\Sigma})$, $\max_{1\leqslant i<j\leqslant p} |\sigma_{ij}|$ and $\max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} \sigma_{ij}^2$. As a result, by (S75) and (S76), we see that (2.3) still holds if "$\mathbf{\Sigma}$" is replaced with "$\mathbf{\Sigma}_{\Lambda}$". Review Lemma S7. Let C be as in (2.3). Let $t = t_p$ be as in (S67). By this display,

$$t' = t'_p := \frac{C\epsilon}{8d} \cdot \min\left\{\frac{\upsilon_p}{\lambda_{max}(\mathbf{\Sigma}) \log p}, \frac{\sqrt{p}}{\log p}\right\} \to \infty$$

as $p \to \infty$. Evidently, $t \geqslant t'$ and $\frac{C\epsilon}{8d}\frac{\sqrt{p}}{\log p} \geqslant t'$. Thus

$$\frac{3}{p^t} + \exp\Big(-\frac{C\epsilon}{8d}\sqrt{p}\Big) \leqslant \frac{4}{p^{t'}} \leqslant \frac{1}{p^{t'/2}}$$

if $p^{t'/2} > 4$. Taking $p_0 \geqslant 3$ such that $p^{t'/2} > 4$, $p \geqslant \frac{256d^2}{\epsilon^2}$ and $\epsilon_p < \frac{\epsilon}{8(C+2)d}$ for all $p \geqslant p_0$ and applying Lemma S7, we know that

$$P(|\Theta_p| \geqslant \epsilon\upsilon_p) \leqslant \frac{1}{p^{t'/2}}, \tag{S77}$$

as $p \geqslant p_0$. Note that in the proof of (S77), although the conclusion is on $\mathbf{\Sigma}$, only five quantities of $\mathbf{\Sigma}$ in (S75) and (S76) are required, and they are the same if "$\mathbf{\Sigma}$" is replaced by "$\mathbf{\Sigma}_\Lambda$" for different Λ. Consequently, we induce from (S77) that

$$P(|\Theta_{p,\Lambda}| \geqslant \epsilon \upsilon_p) \leqslant \frac{1}{p^{t'/2}}.$$

for any $p \geqslant p_0$ and any $\Lambda = \{i_1, \cdots, i_d\}$ with $1 \leqslant i_1 < \cdots < i_d \leqslant p$. The desired conclusion then follows by writing $t'/2$ back to t. □

Lemma S9. *Assume* $(Z_1, \cdots, Z_p)^T \sim N(\mathbf{0}, \mathbf{\Sigma})$ *with* $\mathbf{\Sigma}$ *satisfying* (2.3). *Set* $S_p = Z_1^2 + \cdots + Z_p^2$ *and* $\upsilon_p = [2tr(\mathbf{\Sigma}^2)]^{1/2}$. *For any* $x \in \mathbb{R}$ *and* $y \in \mathbb{R}$, *define* $A_p = \{\frac{S_p - p}{\upsilon_p} \leqslant x\}$ *and* $l_p = (2\log p - \log\log p + y)^{1/2}$ *and* $B_i = \{|Z_i| > l_p\}$. *Then, for each* $d \geqslant 1$,

$$\sum_{1 \leqslant i_1 < \cdots < i_d \leqslant p} \left| P(A_p B_{i_1} \cdots B_{i_d}) - P(A_p) \cdot P(B_{i_1} \cdots B_{i_d}) \right| \to 0$$

as $p \to \infty$.

Proof of Lemma S9. We prove the lemma in two steps.

Step 1: appealing independence from normal distributions. Note that $(Z_1, \cdots, Z_p)^T \sim N(\mathbf{0}, \mathbf{\Sigma})$. Take $\mathbf{X}_1 = (Z_1, \cdots, Z_d)^T$ and $\mathbf{X}_2 = (Z_{d+1}, \cdots, Z_p)^T$. Recall the notation in Lemma S5, which allows us to write

$$\mathbf{X}_2 = \mathbf{U}_p + \mathbf{V}_p,$$

where $\mathbf{U}_p = \mathbf{X}_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{X}_1 \sim N(\mathbf{0}, \mathbf{\Sigma}_{22\cdot 1})$ and $\mathbf{V}_p = \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{X}_1 \sim N(\mathbf{0}, \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})$. Lemma S5 says that

$$\mathbf{U}_p \text{ and } \{Z_1, \cdots, Z_d\} \text{ are independent.} \tag{S78}$$

Further denote

$$S_p = \|\mathbf{X}_1\|^2 + \|\mathbf{X}_2\|^2 = \|\mathbf{U}_p\|^2 + \|\mathbf{V}_p\|^2 + 2\mathbf{U}_p^T\mathbf{V}_p + \sum_{i=1}^d Z_i^2.$$

We will show the last three terms on the right hand side are negligible. Recall

$$\Theta_p = \|\mathbf{V}_p\|^2 + 2\mathbf{U}_p^T\mathbf{V}_p + \sum_{i=1}^d Z_i^2$$

as defined in Lemma S7. By Lemma S8 with $\mathbf{X}_\Lambda = \mathbf{X}$, for any $d \geqslant 1$ and $\epsilon > 0$, there exists $t = t_p > 0$ with $\lim_{p\to\infty} t_p = \infty$ and integer $p_0 \geqslant 1$, such that

$$P(|\Theta_p| \geqslant \epsilon \upsilon_p) \leqslant \frac{1}{p^t} \tag{S79}$$

as $p \geqslant p_0$. Now for clarity we re-write the definition of A_p as

$$A_p(x) = \Big\{\frac{1}{\upsilon_p}(S_p - p) \leqslant x\Big\}, \quad x \in \mathbb{R},$$

for $p \geqslant 1$. Since $S_p = \|\mathbf{U}_p\|^2 + \Theta_p$, we see that

$$\begin{aligned} P(A_p(x)B_1\cdots B_d) \leqslant & P\Big(A_p(x)B_1\cdots B_d,\ \frac{|\Theta_p|}{\upsilon_p} < \epsilon\Big) + \frac{1}{p^t} \\ \leqslant & P\Big(\frac{1}{\upsilon_p}(\|\mathbf{U}_p\|^2 - p) \leqslant x + \epsilon,\ B_1\cdots B_d\Big) + \frac{1}{p^t} \end{aligned}$$

$$=P\Big(\frac{1}{\upsilon_p}(\|\mathbf{U}_p\|^2-p)\leqslant x+\epsilon\Big)\cdot P\big(B_1\cdots B_d\big)+\frac{1}{p^t},$$

by the independence stated in (S78). Regarding the first probability, we have

$$\begin{aligned}P\Big(\frac{1}{\upsilon_p}(\|\mathbf{U}_p\|^2-p)\leqslant x+\epsilon\Big)\leqslant&P\Big(\frac{1}{\upsilon_p}(\|\mathbf{U}_p\|^2-p)\leqslant x+\epsilon,\ \frac{|\Theta_p|}{\upsilon_p}<\epsilon\Big)+\frac{1}{p^t}\\ \leqslant&P\Big(\frac{1}{\upsilon_p}(\|\mathbf{U}_p\|^2+\Theta_p-p)\leqslant x+2\epsilon\Big)+\frac{1}{p^t}\\ \leqslant&P\big(A_p(x+2\epsilon)\big)+\frac{1}{p^t}.\end{aligned}$$

Combine the two inequalities to get

$$P(A_p(x)B_1\cdots B_d)\leqslant P\big(A_p(x+2\epsilon)\big)\cdot P\big(B_1\cdots B_d\big)+\frac{2}{p^t}. \tag{S80}$$

Similarly,

$$\begin{aligned}&P\Big(\frac{1}{\upsilon_p}(\|\mathbf{U}_p\|^2-p)\leqslant x-\epsilon,\ B_1\cdots B_d\Big)\\ \leqslant&P\Big(\frac{1}{\upsilon_p}(\|\mathbf{U}_p\|^2-p)\leqslant x-\epsilon,B_1\cdots B_d,\frac{|\Theta_p|}{\upsilon_p}<\epsilon\Big)+\frac{1}{p^t}\\ \leqslant&P\Big(\frac{1}{\upsilon_p}(S_p-p)\leqslant x,\ B_1\cdots B_d\Big)+\frac{1}{p^t}.\end{aligned}$$

By the independence from (S78),

$$P(A_p(x)B_1\cdots B_d)\geqslant P\Big(\frac{1}{\upsilon_p}(\|\mathbf{U}_p\|^2-p)\leqslant x-\epsilon\Big)\cdot P(B_1\cdots B_d)-\frac{1}{p^t}.$$

Furthermore,

$$P\Big(\frac{1}{\upsilon_p}(S_p-p)\leqslant x-2\epsilon\Big)\leqslant P\Big(\frac{1}{\upsilon_p}(S_p-p)\leqslant x-2\epsilon,\ \frac{|\Theta_p|}{\upsilon_p}<\epsilon\Big)+\frac{1}{p^t}$$

$$\leqslant P\Big(\frac{1}{\upsilon_p}(\|\mathbf{U}_p\|^2 - p) \leqslant x - \epsilon\Big) + \frac{1}{p^t}$$

where the fact $S_p = \|\mathbf{U}_p\|^2 + \Theta_p$ is used again. Combining the above two inequalities we get

$$P(A_p(x)B_1 \cdots B_d) \geqslant P(A_p(x-2\epsilon)) \cdot P(B_1 \cdots B_d) - \frac{2}{p^t}.$$

This together with (S80) implies that

$$\begin{aligned} &\big|P(A_p(x)B_1 \cdots B_d) - P(A_p(x)) \cdot P(B_1 \cdots B_d)\big| \\ \leqslant &\Delta_{p,\epsilon} \cdot P(B_1 \cdots B_d) + \frac{2}{p^t} \end{aligned} \tag{S81}$$

as $p \geqslant p_0$, where

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

since $P(A_p(x))$ is increasing in $x \in \mathbb{R}$. An important observation is that the derivation of (S81) is based on three key facts: inequality (S79), the identity $S_p = \|\mathbf{U}_p\|^2 + \Theta_p$ and the fact $\mathbf{U}_p$ and $\{Z_1, \cdots, Z_d\}$ are independent from (S78).

Recall the notations in Lemma S8. For any $1 \leqslant i_1 < i_2 < \cdots < i_d \leqslant p$, denote $\Lambda = \{i_1, \cdots, i_d\}$. Then, $\mathbf{X}_{2,\Lambda} = \mathbf{U}_{p,\Lambda} + \mathbf{V}_{p,\Lambda}$. By Lemma S5, $\mathbf{U}_{p,\Lambda}$ and $\{Z_{i_1}, \cdots, Z_{i_d}\}$ are independent. In addition,

$$S_p = \|\mathbf{X}_{1,\Lambda}\|^2 + \|\mathbf{X}_{2,\Lambda}\|^2 = \|\mathbf{U}_{p,\Lambda}\|^2 + \|\mathbf{V}_{p,\Lambda}\|^2 + 2\mathbf{U}_{p,\Lambda}^T\mathbf{V}_{p,\Lambda} + \sum_{k=1}^{d} Z_{i_k}^2;$$

$$\Theta_{p,\Lambda} = \|\mathbf{V}_{p,\Lambda}\|^2 + 2\mathbf{U}_{p,\Lambda}^T \mathbf{V}_{p,\Lambda} + \sum_{k=1}^{d} Z_{i_k}^2.$$

Hence, we can write $S_p = \|\mathbf{U}_{p,\Lambda}\|^2 + \Theta_{p,\Lambda}$. Based on Lemma S8,

$$\max_{\Lambda} P(|\Theta_{p,\Lambda}| \geqslant \epsilon \upsilon_p) \leqslant \frac{1}{p^t}$$

when $p \geqslant p_0$. Consequently, the three key facts aforementioned also hold for the corresponding quantities related to Λ. Thus, similar to the derivation of (S81), we have

$$\begin{aligned} &\left|P(A_p(x)B_{i_1}\cdots B_{i_d}) - P(A_p(x)) \cdot P(B_{i_1}\cdots B_{i_d})\right| \\ \leqslant &\Delta_{p,\epsilon} \cdot P(B_{i_1}\cdots B_{i_d}) + \frac{2}{p^t}, \end{aligned}$$

as $p \geqslant p_0$. Taking the summation we get

$$\begin{aligned} \zeta(p,d) := &\sum_{1\leqslant i_1<\cdots<i_d\leqslant p} \left|P(A_p(x)B_{i_1}\cdots B_{i_d}) - P(A_p(x)) \cdot P(B_{i_1}\cdots B_{i_d})\right| \\ \leqslant &\sum_{1\leqslant i_1<\cdots<i_d\leqslant p} \left[\Delta_{p,\epsilon} \cdot P(B_{i_1}\cdots B_{i_d}) + \frac{2}{p^t}\right] \\ \leqslant &\Delta_{p,\epsilon} \cdot H(d,p) + \binom{p}{d} \cdot \frac{2}{p^t}, \end{aligned} \tag{S82}$$

where we denote

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

In the following we will show $\lim_{\epsilon\downarrow 0} \limsup_{p\to\infty} \Delta_{p,\epsilon} = 0$ and $\limsup_{p\to\infty} H(d,p) < \infty$ for each $d \geqslant 1$. Assuming these are true, by using $\binom{p}{d} \leqslant p^d$ and (S82), for fixed $d \geqslant 1$, by sending $p \to \infty$ first and then sending $\epsilon \downarrow 0$, we obtain $\lim_{p\to\infty} \zeta(p,d) = 0$ for each $d \geqslant 1$. The proof is then completed.

Step 2: the proofs of "$\lim_{\epsilon\downarrow 0} \limsup_{p\to\infty} \Delta_{p,\epsilon} = 0$*" and "*$\limsup_{p\to\infty} H(d,p) < \infty$ *for each* $d \geqslant 1$*".* First, as discussed below (2.3), Assumption (2.3) implies Assumption (2.1). Thus, Theorem 1 holds and we have as $p \to \infty$,

$$\frac{S_p - p}{\upsilon_p} \to N(0,1) \text{ weakly,} \tag{S83}$$

and hence

$$\Delta_{p,\epsilon} \to \Phi(x + 2\epsilon) - \Phi(x - 2\epsilon), \tag{S84}$$

as $p \to \infty$, where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$. This implies that $\lim_{\epsilon\downarrow 0} \limsup_{p\to\infty} \Delta_{p,\epsilon} = 0$.

Second, take $\delta_p = 1/(\log p)^2$. Recall $B_{p,i} = \{1 \leqslant j \leqslant p;\, |\sigma_{ij}| \geqslant \delta_p\}$ for $1 \leqslant i \leqslant p$ defined in Theorem 2. By Assumption (2.3), we know $\max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} \sigma_{ij}^2 \leqslant (\log p)^C$ for all $p \geqslant 1$. Then

$$|B_{p,i}| \cdot \frac{1}{(\log p)^2} \leqslant \sum_{j=1}^{p} \sigma_{ij}^2 \leqslant (\log p)^C$$

for each $i = 1, \cdots, p$. This implies that $\max_{1\leqslant i\leqslant p} |B_{p,i}| \leqslant (\log p)^{C+2}$. Take $\kappa = \kappa_p = (C+3)(\log\log p)/\log p$ for $p \geqslant e^e$. Then, $\kappa_p \to 0$ and $(\log p)^{C+2} < p^{\kappa}$, which gives

$$C_p := \{1 \leqslant i \leqslant p;\, |B_{p,i}| \geqslant p^{\kappa}\} = \varnothing. \tag{S85}$$

Hence, $D_p := \{1 \leqslant i \leqslant p;\, |B_{p,i}| < p^{\kappa}\} = \{1, 2, \cdots, p\}$. Recall (S36), (S40) and (S42). By noting that "$H(t,p)$" here is exactly "α_t" there for each $t \geqslant 1$, we know

$$\lim_{p\to\infty} H(d,p) = \frac{1}{d!}\pi^{-d/2}e^{-dx/2}, \tag{S86}$$

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

We are now in the position to prove Theorem 3.

Proof of Theorem 3. Again, since Assumption (2.3) implies Assumption (2.1) and Assumption (2.2), we know that Theorem 1 and Theorem 2 hold. Set $\upsilon_p = [2\mathrm{tr}(\mathbf{\Sigma}^2)]^{1/2}$. By Theorem 1,

$$P\Big(\frac{S_p - p}{\upsilon_p} \leqslant x\Big) = \Phi(x) \tag{S87}$$

as $p \to \infty$ for any $x \in \mathbb{R}$, where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt$. From Theorem 2, we have

$$P\Big(\max_{1\leqslant i\leqslant p}\big\{Z_i^2\big\} - 2\log p + \log\log p \leqslant y\Big) \to F(y) = \exp\Big\{-\frac{1}{\sqrt{\pi}}e^{-y/2}\Big\} \tag{S88}$$

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

$$\lim_{p\to\infty} P\Big(\frac{S_p - p}{\upsilon_p} \leqslant x,\ \max_{1\leqslant i\leqslant p} Z_i^2 - 2\log p + \log\log p \leqslant y\Big) = \Phi(x)\cdot F(y)$$

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

$$L_p = \max_{1\leqslant i\leqslant p}|Z_i| \quad \text{and} \quad l_p = (2\log p - \log\log p + y)^{1/2}, \tag{S89}$$

where the latter one makes sense for sufficiently large p. Due to (S87), the above condition we want to prove is equivalent to

$$\lim_{p\to\infty} P\Big(\frac{S_p - p}{\upsilon_p} \leqslant x,\ L_p > l_p\Big) = \Phi(x)\cdot[1-F(y)], \tag{S90}$$

for any $x \in \mathbb{R}$ and $y \in \mathbb{R}$. Recalling the notation in Lemma S9, we have

$$A_p = \Big\{\frac{S_p - p}{\upsilon_p} \leqslant x\Big\} \quad \text{and} \quad B_i = \big\{|Z_i| > l_p\big\} \tag{S91}$$

for $1 \leqslant i \leqslant p$. We can then write

$$P\Big(\frac{1}{\upsilon_p}(S_p - p) \leqslant x,\ L_p > l_p\Big) = P\Big(\bigcup_{i=1}^{p} A_p B_i\Big). \tag{S92}$$

Here the notation A_pB_i stands for $A_p \cap B_i$. From the inclusion-exclusion principle,

$$\begin{aligned} P\Big(\bigcup_{i=1}^{p} A_p B_i\Big) \leqslant & \sum_{1\leqslant i_1\leqslant p} P(A_pB_{i_1}) - \sum_{1\leqslant i_1<i_2\leqslant p} P(A_pB_{i_1}B_{i_2}) \\ & + \cdots + \sum_{1\leqslant i_1<\cdots<i_{2k+1}\leqslant p} P(A_pB_{i_1}\cdots B_{i_{2k+1}}), \end{aligned} \tag{S93}$$

and

$$\begin{aligned} P\Big(\bigcup_{i=1}^{p} A_p B_i\Big) \geqslant & \sum_{1\leqslant i_1\leqslant p} P(A_pB_{i_1}) - \sum_{1\leqslant i_1<i_2\leqslant p} P(A_pB_{i_1}B_{i_2}) \\ & + \cdots - \sum_{1\leqslant i_1<\cdots<i_{2k}\leqslant p} P(A_pB_{i_1}\cdots B_{i_{2k}}) \end{aligned} \tag{S94}$$

for any integer $k \geqslant 1$. As in the proof of Lemma S9, define

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

for $d \geqslant 1$. From (S86) we know

$$\lim_{d\to\infty} \limsup_{p\to\infty} H(p,d) = 0. \tag{S95}$$

Denote

$$\zeta(p,d) = \sum_{1\leqslant i_1<\cdots<i_d\leqslant p} \big[P(A_p B_{i_1}\cdots B_{i_d}) - P(A_p)\cdot P(B_{i_1}\cdots B_{i_d})\big]$$

By Lemma S9, we have

$$\lim_{p\to\infty} \zeta(p,d) = 0 \tag{S96}$$

for each $d \geqslant 1$. The assertion (S93) implies that

$$\begin{aligned} P\Big(\bigcup_{i=1}^{p} A_p B_i\Big) \leqslant & P(A_p)\Big[\sum_{1\leqslant i_1\leqslant p} P(B_{i_1}) - \sum_{1\leqslant i_1<i_2\leqslant p} P(B_{i_1}B_{i_2}) + \cdots - \\ & \sum_{1\leqslant i_1<\cdots<i_{2k}\leqslant p} P(B_{i_1}\cdots B_{i_{2k}})\Big] + \Big[\sum_{d=1}^{2k}\zeta(p,d)\Big] + H(p,2k+1) \\ \leqslant & P(A_p)\cdot P\Big(\bigcup_{i=1}^{p} B_i\Big) + \Big[\sum_{d=1}^{2k}\zeta(p,d)\Big] + H(p,2k+1), \end{aligned} \tag{S97}$$

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

$$P\Big(\bigcup_{i=1}^{p} B_i\Big) \geqslant \sum_{1\leqslant i_1\leqslant p} P(B_{i_1}) - \sum_{1\leqslant i_1<i_2\leqslant p} P(B_{i_1}B_{i_2}) + \cdots - \sum_{1\leqslant i_1<\cdots<i_{2k}\leqslant p} P(B_{i_1}\cdots B_{i_{2k}}),$$

for all $k \geqslant 1$. By the definition of l_p and (S88),

$$P\Big(\bigcup_{i=1}^{p} B_i\Big) = P\big(L_p > l_p\big) = P\big(L_p^2 - 2\log p + \log\log p > y\big) \to 1 - F(y)$$

as $p \to \infty$. By (S87), $P(A_p) \to \Phi(x)$ as $p \to \infty$. From (S92), (S96) and (S97), by fixing k first and sending $p \to \infty$ we obtain that

$$\limsup_{p\to\infty} P\Big(\frac{1}{\upsilon_p}(S_p - p) \leqslant x,\ L_p > l_p\Big) \leqslant \Phi(x) \cdot [1 - F(y)] + \lim_{p\to\infty} H(p, 2k+1).$$

Now, by letting $k \to \infty$ and using (S95), we have

$$\limsup_{p\to\infty} P\Big(\frac{1}{\upsilon_p}(S_p - p) \leqslant x,\ L_p > l_p\Big) \leqslant \Phi(x) \cdot [1 - F(y)]. \qquad \text{(S98)}$$

We next prove the lower bound in a similar way. By applying the same argument to (S94), we see that the counterpart of (S97) becomes

$$\begin{aligned} P\Big(\bigcup_{i=1}^{p} A_p B_i\Big) \geqslant & P(A_p)\Big[\sum_{1\leqslant i_1\leqslant p} P(B_{i_1}) - \sum_{1\leqslant i_1<i_2\leqslant p} P(B_{i_1}B_{i_2}) + \cdots + \\ & \sum_{1\leqslant i_1<\cdots<i_{2k-1}\leqslant p} P(B_{i_1}\cdots B_{i_{2k-1}})\Big] + \Big[\sum_{d=1}^{2k-1} \zeta(p,d)\Big] - H(p,2k) \\ \geqslant & P(A_p)\cdot P\Big(\bigcup_{i=1}^{p} B_i\Big) + \Big[\sum_{d=1}^{2k-1} \zeta(p,d)\Big] - H(p,2k), \end{aligned}$$

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

$$P\Big(\bigcup_{i=1}^{p} B_i\Big) \leqslant \sum_{1\leqslant i_1\leqslant p} P(B_{i_1}) - \sum_{1\leqslant i_1<i_2\leqslant p} P(B_{i_1}B_{i_2}) + \cdots + \sum_{1\leqslant i_1<\cdots<i_{2k-1}\leqslant p} P(B_{i_1}\cdots B_{i_{2k-1}})$$

for all $k \geqslant 1$. Review (S92) and repeat the earlier procedure to see

$$\liminf_{p\to\infty} P\Big(\frac{1}{\upsilon_p}(S_p - p) \leqslant x,\ L_p > l_p\Big) \geqslant \Phi(x) \cdot [1 - F(y)]$$

with $p \to \infty$ and $k \to \infty$, which, together with (S98), yields (S90). The proof is now complete. □

S3.4 Proof of Theorem 4 and Theorem S1

Lemma S10. *Let* $\{(U, U_p, \tilde{U}_p) \in \mathbb{R}^3;\, p \geqslant 1\}$ *and* $\{(V, V_p, \tilde{V}_p) \in \mathbb{R}^3;\, p \geqslant 1\}$ *be two sequences of random variables with* $U_p \to U$ *and* $V_p \to V$ *in distribution as* $p \to \infty$. *Assume* U *and* V *are continuous random variables and*

$$\tilde{U}_p = U_p + o_p(1) \quad \text{and} \quad \tilde{V}_p = V_p + o_p(1). \tag{S99}$$

If U_p *and* V_p *are asymptotically independent, then* $\tilde{U}_p$ *and* $\tilde{V}_p$ *are also asymptotically independent.*

Proof of Lemma S10. Define

$$\Omega_{p,\epsilon} = \left\{ \left| U_p - \tilde{U}_p \right| \leqslant \epsilon,\ \left| V_p - \tilde{V}_p \right| \leqslant \epsilon \right\}$$

for any $p \geqslant 1$ and $\epsilon > 0$. By (S99),

$$\lim_{p\to\infty} P(\Omega_{p,\epsilon}) = 1 \tag{S100}$$

for any $\epsilon > 0$. Fix $x \in \mathbb{R}$ and $y \in \mathbb{R}$. We note that

$$\begin{aligned} P\Big(\tilde{U}_p \leqslant x, \tilde{V}_p \leqslant y\Big) \leqslant & P\Big(\tilde{U}_p \leqslant x, \tilde{V}_p \leqslant y,\, \Omega_{p,\epsilon}\Big) + P(\Omega_{p,\epsilon}^c) \\ \leqslant & P\Big(U_p \leqslant x + \epsilon, V_p \leqslant y + \epsilon\Big) + P(\Omega_{p,\epsilon}^c). \end{aligned} \tag{S101}$$

By the assumption on the asymptotic independence,

$$\lim_{p\to\infty} P(U_p \leqslant s, V_p \leqslant t) = P(U \leqslant s) \cdot P(V \leqslant t) \tag{S102}$$

for any $s \in \mathbb{R}$ and $t \in \mathbb{R}$. By letting $p \to \infty$ and then $\epsilon \downarrow 0$ in (S101), since U and V are continuous, we deduce from (S100) and (S102) that

$$\limsup_{p\to\infty} P\Big(\tilde{U}_p \leqslant x, \tilde{V}_p \leqslant y\Big) \leqslant P(U \leqslant x)\cdot P(V \leqslant y). \tag{S103}$$

By switching the roles of "$U_p \to U$ and $V_p \to V$" and "$\tilde{U}_p$ and $\tilde{V}_p$"in (S101), we have

$$P\Big(U_p \leqslant x, V_p \leqslant y\Big) \leqslant P\Big(\tilde{U}_p \leqslant x+\epsilon, \tilde{V}_p \leqslant y+\epsilon\Big) + P(\Omega_{p,\epsilon}^c)$$

for any $x \in \mathbb{R}$, $y \in \mathbb{R}$ and $\epsilon > 0$. Or, equivalently,

$$P\Big(\tilde{U}_p \leqslant x, \tilde{V}_p \leqslant y\Big) \geqslant P\Big(U_p \leqslant x-\epsilon, V_p \leqslant y-\epsilon\Big) - P(\Omega_{p,\epsilon}^c).$$

Similar to the derivation of (S103), we get

$$\liminf_{p\to\infty} P\Big(\tilde{U}_p \leqslant x, \tilde{V}_p \leqslant y\Big) \geqslant P(U \leqslant x)\cdot P(V \leqslant y).$$

This and (S103) lead to

$$\lim_{p\to\infty} P(\tilde{U}_p \leqslant x, \tilde{V}_p \leqslant y) = P(U \leqslant x)\cdot P(V \leqslant y),$$

which shows the asymptotic independence between $\tilde{U}_p$ and $\tilde{V}_p$ as claimed.

□

Proof of Theorem 4. By Theorem 3.1 in Srivastava (2009), we get claim (i). We next prove claim (ii).

Recall $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are i.i.d. $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distributed random vectors and $\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i = (\bar{\boldsymbol{X}}_1, \cdots, \bar{\boldsymbol{X}}_p)^T$. Note that $\sqrt{n}\bar{\boldsymbol{X}} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ under the null hypothesis in (3.1). Write $\mathbf{\Sigma} = (\sigma_{ij})_{p\times p}$, and let $\mathbf{R} = \mathbf{D}^{-1/2}\mathbf{\Sigma}\mathbf{D}^{-1/2} = (\rho_{ij})_{1\leqslant i,j\leqslant p}$ denote the population correlation matrix, where $\mathbf{D}$ is the diagonal matrix of $\mathbf{\Sigma}$. Then $Z_i := \sqrt{n}\bar{\boldsymbol{X}}_i/\sqrt{\sigma_{ii}} \sim N(0,1)$ for each $1 \leqslant i \leqslant p$ and $\mathrm{Cov}(Z_i, Z_j) = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}} = \rho_{ij}$ for $i \neq j$. In other words, $(Z_1, \cdots, Z_p)^T \sim N(\mathbf{0}, \mathbf{R})$. By assumption, (2.2) holds with "$\mathbf{\Sigma}$" replaced by "$\mathbf{R}$". Set $\tilde{T}^{(1)}_{max} = \max_{1\leqslant i\leqslant p} Z_i^2$. Since Assumption (2.3) is stronger than Assumption (2.2), by Theorem 2 and Assumption (2.2), it holds that

$$\tilde{T}^{(1)}_{max} - 2\log p + \log\log p = \max_{1\leqslant i\leqslant p} Z_i^2 - 2\log p + \log\log p \ \text{ converges}$$
$$\text{weakly to a distribution with cdf } F(x) = \exp\big\{ - e^{-x/2}/\sqrt{\pi}\big\},\ x \in \mathbb{R}. \tag{S104}$$

Observe that the distribution of $(Z_1, \cdots, Z_p)^T \sim N(\mathbf{0}, \mathbf{R})$ is free of n, hence the above limit holds for any $n = n_p$. Now, to prove (ii), we only need to show that $T^{(1)}_{max} - \tilde{T}^{(1)}_{max} = o_p(1)$. Indeed, we have

$$\begin{aligned} |T^{(1)}_{max} - \tilde{T}^{(1)}_{max}| =& \Big| n \max_{1\leqslant i\leqslant p} \hat{\sigma}_{ii}^{-2}\bar{\boldsymbol{X}}_i^2 - n \max_{1\leqslant i\leqslant p} \sigma_{ii}^{-1}\bar{\boldsymbol{X}}_i^2 \Big| \\ \leqslant& \Big| n \max_{1\leqslant i\leqslant p} \sigma_{ii}^{-1}\bar{\boldsymbol{X}}_i^2 \Big| \cdot \max_{1\leqslant i\leqslant p} |\sigma_{ii}\hat{\sigma}_{ii}^{-2} - 1| \\ =& \Big(\max_{1\leqslant i\leqslant p} |Z_i| \Big)^2 \cdot \max_{1\leqslant i\leqslant p} |\sigma_{ii}\hat{\sigma}_{ii}^{-2} - 1|. \end{aligned} \tag{S105}$$

First, use the inequality $P(N(0,1) \geqslant x) \leqslant e^{-x^2/2}$ for $x > 0$ to see

$$P\Big(\max_{1\leqslant i\leqslant p} |Z_i| \geqslant 2\sqrt{\log p}\Big) \leqslant p \cdot P(|N(0,1)| \geqslant 2\sqrt{\log p}) \leqslant \frac{2}{p}.$$

Thus,

$$(\max_{1\leqslant i\leqslant p}|Z_i|)^2 = O_p(\log p). \tag{S106}$$

Based on the explanation below (3.5), we have

$$\sigma_{ii}^{-1}\hat{\sigma}_{ii}^2 \sim \frac{1}{n}\chi^2(n-1) \tag{S107}$$

for each i. Set $a_p = \alpha\sqrt{n^{-1}\log p}$ with the constant α to be determined. Then

$$\begin{aligned} P\Big(\max_{1\leqslant i\leqslant p}|\sigma_{ii}\hat{\sigma}_{ii}^{-2}-1| \geqslant a_p\Big) \leqslant & p\cdot P\Big(\Big|\frac{n}{\chi^2(n-1)}-1\Big| \geqslant a_p\Big) \\ \leqslant & p\cdot P\Big(|\chi^2(n-1)-n| \geqslant \frac{na_p}{2}\Big) + p\cdot P\Big(\chi^2(n-1) < \frac{1}{2}n\Big) \end{aligned}$$

by considering $\chi^2(n-1) < \frac{1}{2}n$ or not. Recall the Chernoff bound and the moderate deviation for sum of i.i.d. random variables (see, for example, p.31 and p.109 from Dembo and Zeitouni (1998)). There exists a constant $C > 0$ such that $P(\chi^2(n-1) < \frac{1}{2}n) \leqslant e^{-Cn}$ for all $n \geqslant 1$ and

$$P\Big(|\chi^2(n-1)-n| \geqslant \frac{na_p}{2}\Big) \leqslant P\Big(\Big|\frac{\chi^2(m)-m}{\sqrt{m\log p}}\Big| \geqslant \frac{\alpha}{3}\Big) \leqslant \exp\Big(-\frac{1}{3}\cdot\frac{\alpha^2}{9}\log p\Big) \tag{S108}$$

as p is sufficiently large, where $m := n-1$ and the fact $\log p = o(n)$ is used in the last step. Choose $\alpha = 8$ to bound the above probability by $O(p^{-2})$. It follows that

$$P\Big(\max_{1\leqslant i\leqslant p}|\sigma_{ii}\hat{\sigma}_{ii}^{-2}-1| \geqslant a_p\Big) = p\cdot O(p^{-2}) + e^{\log p - Cn} \to 0,$$

as $p \to \infty$. This says

$$\max_{1 \leqslant i \leqslant p} |\sigma_{ii} \hat{\sigma}_{ii}^{-2} - 1| = o_p(\sqrt{n^{-1} \log p}). \tag{S109}$$

This together with (S105) and (S106) implies $T_{max}^{(1)} - \tilde{T}_{max}^{(1)} = o_p((n^{-1} \log^3 p)^{1/2}) \to 0$ as long as $\log p = o(n^{1/3})$. This confirms $T_{max}^{(1)} - \tilde{T}_{max}^{(1)} = o_p(1)$, and the proof of part (ii) is completed by using (S104).

Finally we prove part (iii). According to the proof of Theorem 3.1 in Srivastava (2009) or the proof of Theorem 1 in Jiang and Li (2021), we have

$$T_{sum}^{(1)} = \tilde{T}_{sum}^{(1)} + o_p(1) = \frac{n \bar{\boldsymbol{X}}^T \mathbf{D}^{-1} \bar{\boldsymbol{X}} - p}{\sqrt{2\mathrm{tr}(\mathbf{R}^2)}} + o_p(1). \tag{S110}$$

Also, using an conclusion from the proof of (ii) above,

$$T_{max}^{(1)} = \tilde{T}_{max}^{(1)} + o_p(1) = \max_{1 \leqslant i \leqslant p} \frac{\sqrt{n} \bar{\boldsymbol{X}}_i}{\sqrt{\sigma_{ii}}} + o_p(1). \tag{S111}$$

Since $\sqrt{n}(\bar{\boldsymbol{X}}_1, \cdots, \bar{\boldsymbol{X}}_p)^T = \sqrt{n} \bar{\boldsymbol{X}}^T \sim N(\mathbf{0}, \mathbf{\Sigma})$, then $\sqrt{n} \mathbf{D}^{-1/2} \bar{\boldsymbol{X}} \sim N(\mathbf{0}, \mathbf{R})$ by using the notation $\mathbf{R} = \mathbf{D}^{-1/2} \mathbf{\Sigma} \mathbf{D}^{-1/2}$. Recall an earlier notation $Z_i = \sqrt{n} \bar{\boldsymbol{X}}_i / \sqrt{\sigma_{ii}} \sim N(0,1)$ for each $1 \leqslant i \leqslant p$. Obviously, $(Z_1, \cdots, Z_p)^T = \sqrt{n} \mathbf{D}^{-1/2} \bar{\boldsymbol{X}} \sim N(\mathbf{0}, \mathbf{R})$. We are able to rewrite (S110) and (S111) in terms of Z's as

$$T_{sum}^{(1)} = \frac{Z_1^2 + \cdots + Z_p^2 - p}{\sqrt{2\mathrm{tr}(\mathbf{R}^2)}} + o_p(1) \quad \text{and} \quad T_{max}^{(1)} = \max_{1 \leqslant i \leqslant p} Z_i^2 + o_p(1).$$

As aforementioned, Assumption (2.3) is stronger than Assumption (2.2). We then conclude (iii) by Theorem 3, Lemma S10 and the fact $(Z_1, \cdots, Z_p)^T \sim N(\mathbf{0}, \mathbf{R})$. □

Verification of $\beta_M^{(1)}(\boldsymbol{\mu},\alpha)\approx\alpha$. Recall the simplified assumption that $\boldsymbol{\Sigma}=(\sigma_{ij})_{p\times p}=\mathbf{I}_p$, $\xi\in(1/2,5/6]$ and $\delta=O(n^{-\xi})$. In this case, $n^{1-2\xi}\to 0$. We also assume that $\log p=o(n^{\xi-\frac{1}{2}})$. Because of the condition $\xi\in(1/2,5/6]$, we know $\log p=o(n^{1/3})$. As a consequence, the requirement on p vs n imposed in Theorem 4(ii) is satisfied. Notice $\beta_M^{(1)}(\boldsymbol{\mu},\alpha)$ is equal to

$$
\begin{aligned}
& P\left(T_{max}^{(1)}-2\log p+2\log\log p>q_\alpha\right)\\
=& P\left(n\max_{1\leqslant i\leqslant p}\frac{\bar{\boldsymbol{X}}_i^2}{\hat{\sigma}_{ii}^2}-2\log p+2\log\log p>q_\alpha\right)\\
=& P\left(n\max_{1\leqslant i\leqslant p}\frac{(\bar{\boldsymbol{X}}_i-\delta)^2+\delta^2+2\delta(\bar{\boldsymbol{X}}_i-\delta)}{\hat{\sigma}_{ii}^2}-2\log p+2\log\log p>q_\alpha\right)\\
\leqslant& P\left(n\max_{1\leqslant i\leqslant p}\frac{(\bar{\boldsymbol{X}}_i-\delta)^2}{\hat{\sigma}_{ii}^2}+n\max_{1\leqslant i\leqslant p}\frac{\delta^2}{\hat{\sigma}_{ii}^2}+n\max_{1\leqslant i\leqslant p}\frac{2\delta|\bar{\boldsymbol{X}}_i-\delta|}{\hat{\sigma}_{ii}^2}-2\log p+2\log\log p>q_\alpha\right).
\end{aligned}
$$

Since $\sigma_{ii}=1$ by assumption, we have from (S109) that $\max_{1\leqslant i\leqslant p}|\hat{\sigma}_{ii}^{-2}-\sigma_{ii}^{-1}|=O_p(\sqrt{n^{-1}\log p})$. In particular, we have from the triangle inequality that

$$
\max_{1\leqslant i\leqslant p}\hat{\sigma}_{ii}^{-1}\leqslant 1+\max_{1\leqslant i\leqslant p}\frac{|\hat{\sigma}_{ii}^{-2}-1|}{\hat{\sigma}_{ii}^{-1}+1}\leqslant 1+\max_{1\leqslant i\leqslant p}|\hat{\sigma}_{ii}^{-2}-1|=1+O_p(\sqrt{n^{-1}\log p}). \tag{S112}
$$

From the fact $\max_{1\leqslant i\leqslant p}|\hat{\sigma}_{ii}^{-2}-\sigma_{ii}^{-1}|=O_p(\sqrt{n^{-1}\log p})$, we see

$$
\begin{aligned}
n\max_{1\leqslant i\leqslant p}\frac{\delta^2}{\hat{\sigma}_{ii}^2}&\leqslant n\max_{1\leqslant i\leqslant p}\frac{\delta^2}{\sigma_{ii}}+n\max_{1\leqslant i\leqslant p}\delta^2\left|\hat{\sigma}_{ii}^{-2}-\sigma_{ii}^{-1}\right|\\
&=O_p(n^{1-2\xi})+O_p(n^{1-2\xi}\sqrt{n^{-1}\log p})=O_p(n^{1-2\xi}).
\end{aligned}
$$

According to Theorem 4(ii), we have $\max_{1\leqslant i\leqslant p}\frac{|\bar{\boldsymbol{X}}_i-\delta|}{\hat{\sigma}_{ii}}=O_p(\sqrt{(\log p)/n})$.

This and (S112) conclude that

$$n \max_{1\leqslant i\leqslant p} \frac{2\delta|\bar{\boldsymbol{X}}_i - \delta|}{\hat{\sigma}_{ii}^2} = O_p\big(n^{\frac{1}{2}-\xi}\sqrt{\log p}\big).$$

Thus,

$$\begin{aligned}
&\beta_M^{(1)}(\boldsymbol{\mu}, \alpha) \\
\leqslant &P\left(n \max_{1\leqslant i\leqslant p} \frac{(\bar{\boldsymbol{X}}_i - \delta)^2}{\hat{\sigma}_{ii}^2} + O_p(n^{1-2\xi}) + O_p(n^{\frac{1}{2}-\xi}\log p) - 2\log p + 2\log\log(p) > q_\alpha\right) \\
\leqslant &P\left(n \max_{1\leqslant i\leqslant p} \frac{(\bar{\boldsymbol{X}}_i - \delta)^2}{\hat{\sigma}_{ii}^2} + o_p(1) - 2\log p + 2\log\log(p) > q_\alpha\right)
\end{aligned}$$

which goes to α. The verification is completed. □

Proof of Theorem S1. The proof shares same spirit as Theorem 4. Denote $n = n_1 + n_2$. According to Section 5 in Srivastava and Du (2008) or Theorem 2 in Jiang and Li (2021), (i) holds. We prove (ii) next.

Under the normality assumption and the null hypothesis in (S1), we have $\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2 \sim N(\mathbf{0}, \frac{n_1+n_2}{n_1 n_2}\boldsymbol{\Sigma})$. Let $\mathbf{D} = \mathrm{diag}(\sigma_{11}, \cdots, \sigma_{pp})$ be the diagonal matrix of $\boldsymbol{\Sigma}$. Recall $\mathbf{R} = \mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}$. Then

$$(Z_1, \cdots, Z_p)^T := \left(\frac{n_1 n_2}{n_1 + n_2}\right)^{1/2} \mathbf{D}^{-1/2}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) \sim N(\mathbf{0}, \mathbf{R}). \tag{S113}$$

According to Section 5 and the proof of Theorem 2.1 in Srivastava and Du (2008) or the proof of Theorem 2 in Jiang and Li (2021), we have

$$\begin{aligned}
T_{sum}^{(2)} =& \frac{\frac{n_1 n_2}{n_1+n_2}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)^T\mathbf{D}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) - p}{\sqrt{2\mathrm{tr}(\mathbf{R}^2)}} + o_p(1) \\
=& \frac{Z_1^2 + \cdots + Z_p^2}{\sqrt{2\mathrm{tr}(\mathbf{R}^2)}} + o_p(1).
\end{aligned} \tag{S114}$$

Let $\bar{\boldsymbol{X}}_{ji}$ be the ith coordinate of $\bar{\boldsymbol{X}}_j \in \mathbb{R}^p$ for $j = 1, 2$ and $1 \leqslant i \leqslant p$. Then

$$\Big(\frac{n_1 n_2}{n_1 + n_2}\Big)^{1/2}\Big(\frac{\bar{\boldsymbol{X}}_{11} - \bar{\boldsymbol{X}}_{21}}{\sqrt{\sigma_{11}}}, \cdots, \frac{\bar{\boldsymbol{X}}_{1p} - \bar{\boldsymbol{X}}_{2p}}{\sqrt{\sigma_{pp}}}\Big)^T = (Z_1, \cdots, Z_p)^T.$$

Recall

$$T_{max}^{(2)} = \frac{n_1 n_2}{n_1 + n_2} \max_{1 \leqslant i \leqslant p} \frac{(\bar{\boldsymbol{X}}_{1i} - \bar{\boldsymbol{X}}_{2i})^2}{\hat{\sigma}_{ii}^2},$$

where $\hat{\sigma}_{ii}^2$ is the ith diagonal element of $\hat{\mathbf{S}}$ in (S3). Set

$$\tilde{T}_{max}^{(2)} = \frac{n_1 n_2}{n_1 + n_2} \max_{1 \leqslant i \leqslant p} \frac{(\bar{\boldsymbol{X}}_{1i} - \bar{\boldsymbol{X}}_{2i})^2}{\sigma_{ii}}.$$

Then $\tilde{T}_{max}^{(2)} = \max_{1 \leqslant i \leqslant p} Z_i^2$. Recalling the discussion below (2.3), Assumption (2.3) is stronger than Assumption (2.2). By Theorem 2 and Assumption (2.2), we obtain

$$\tilde{T}_{max}^{(2)} - 2\log p + \log\log p = \max_{1 \leqslant i \leqslant p} Z_i^2 - 2\log p + \log\log p \ \text{ converges}$$
$$\text{weakly to a distribution with cdf } F(x) = \exp\big\{-e^{-x/2}/\sqrt{\pi}\big\},\ x \in \mathbb{R}. \tag{S115}$$

Thus, to prove (ii), it suffices to show that $T_{max}^{(2)} - \tilde{T}_{max}^{(2)} = o_p(1)$. By (S3), $(n_1 + n_2)\hat{\mathbf{S}}$ follows a Wishart distribution with parameter $n_1 + n_2 - 2$ and covariance matrix $\boldsymbol{\Sigma}$. Since $\hat{\sigma}_{ii}^2$ is the ith diagonal element of $\hat{\mathbf{S}}$, we know $(n_1 + n_2)\hat{\sigma}_{ii}^2 \sim \sigma_{ii}\chi^2(n_1 + n_2 - 2)$, or equivalently,

$$\hat{\sigma}_{ii}^2 \sigma_{ii}^{-1} \sim \frac{\chi^2(n_1 + n_2 - 2)}{n_1 + n_2}$$

for each $1 \leqslant i \leqslant p$. By the same argument as between (S107) and (S109), we have from the above assertion that $\max_{1\leqslant i\leqslant p} |\hat{\sigma}_{ii}^{-2}\sigma_{ii} - 1| = O_p(\sqrt{n^{-1}\log p})$. Notice (S115) implies $\tilde{T}^{(2)}_{max} = O(\log p)$. By the triangle inequality of the maximum and the trivial inequality $\max_{1\leqslant i\leqslant p} |a_i b_i| \leqslant \max_{1\leqslant i\leqslant p} |a_i| \cdot \max_{1\leqslant i\leqslant p} |b_i|$ for any $\{a_i\}$ and $\{b_i\}$, we get

$$\begin{aligned} |T^{(2)}_{max} - \tilde{T}^{(2)}_{max}| =& \left| \frac{n_1 n_2}{n_1+n_2} \max_{1\leqslant i\leqslant p} \hat{\sigma}_{ii}^{-2}(\bar{\boldsymbol{X}}_{1i} - \bar{\boldsymbol{X}}_{2i})^2 - \frac{n_1 n_2}{n_1+n_2} \max_{1\leqslant i\leqslant p} \sigma_{ii}^{-1}(\bar{\boldsymbol{X}}_{1i} - \bar{\boldsymbol{X}}_{2i})^2 \right| \\ \leqslant& |\tilde{T}^{(2)}_{max}| \cdot \max_{1\leqslant i\leqslant p} |\hat{\sigma}_{ii}^{-2}\sigma_{ii} - 1| \\ =& O\big(n^{-1/2}(\log p)^{3/2}\big) \to 0. \end{aligned} \tag{S116}$$

Consequently, (ii) follows from (S115) under the assumption $\log p = o(n^{1/3})$.

Now we prove (iii). Recall (S114). By (S116), we see

$$\tilde{T}^{(2)}_{max} - 2\log p + \log\log p = \max_{1\leqslant i\leqslant p} Z_i^2 - 2\log p + \log\log p + o_p(1).$$

As discussed earlier, Assumption (2.3) is stronger than Assumption (2.2). Then, under Assumption (2.3) with "$\boldsymbol{\Sigma}$" replaced by "$\mathbf{R}$", we conclude (iii) from Theorem 3, (S113) and Lemma S10. □

S3.5 Proof of Theorem S2

To prove Theorem S2, we need a preparation. In fact, an asymptotic ratio-consistent estimator of $\operatorname{tr}\big(\boldsymbol{\Sigma}^2_{b|a}\big)$ will be derived (the notation of "$\boldsymbol{\Sigma}_{b|a}$" is given in (S118)). It is stated in Proposition S1. We will develop a series of auxiliary results for this purpose.

Review the setting in Section S2. In what follows, we assume the integers n, p and q satisfy $1 \leqslant q < p$ and $q < n$.

Lemma S11. *Let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ be i.i.d. from the p-dimensional distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$. Write*

$$\boldsymbol{X}_i = \begin{pmatrix} \mathbf{X}_{ia} \\ \mathbf{X}_{ib} \end{pmatrix} \quad and \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}$$

for each $i = 1, \cdots, n$, where $\mathbf{X}_{ia}$ is a q-dimensional vector with distribution $N(\mathbf{0}, \boldsymbol{\Sigma}_{aa})$. Then $\boldsymbol{\Sigma}_{bb\cdot a} := \boldsymbol{\Sigma}_{bb} - \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab}$ is a $(p-q) \times (p-q)$ matrix. Recall $\hat{\boldsymbol{\Sigma}}_{b|a}$ in (S10). *We then have*

$$\hat{\boldsymbol{\Sigma}}_{b|a} \overset{d}{=} \frac{1}{n}\boldsymbol{\Sigma}_{bb\cdot a}^{1/2} \cdot \mathbf{W}\mathbf{W}^T\boldsymbol{\Sigma}_{bb\cdot a}^{1/2}$$

where $\mathbf{W}$ is a $(p-q) \times (n-q)$ matrix and the entries are i.i.d. $N(0,1)$.

Proof of Lemma S11. Recall the notation between (S9) and (S10),

$$\boldsymbol{X}_a = (\boldsymbol{X}_{1a}, \cdots, \boldsymbol{X}_{na})^T, \quad \mathbf{H}_a = \mathbf{X}_a(\mathbf{X}_a^T\mathbf{X}_a)^{-1}\mathbf{X}_a^T, \quad \boldsymbol{X}_b = (\boldsymbol{X}_{1b}, \cdots, \boldsymbol{X}_{nb})^T,$$
$$\tilde{\mathbf{X}}_b = (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_b, \quad \hat{\boldsymbol{\Sigma}}_{b|a} = n^{-1}\tilde{\mathbf{X}}_b^T\tilde{\mathbf{X}}_b,$$

where $\boldsymbol{X}_a$ is $n \times q$, $\mathbf{H}_a$ is $n \times n$, both $\boldsymbol{X}_b$ and $\tilde{\mathbf{X}}_b$ are $n \times (p-q)$ and $\hat{\boldsymbol{\Sigma}}_{b|a}$ is $(p-q) \times (p-q)$. $\hat{\boldsymbol{\Sigma}}_{b|a}$ is defined as

$$\hat{\boldsymbol{\Sigma}}_{b|a} = \frac{1}{n}\tilde{\mathbf{X}}_b^T\tilde{\mathbf{X}}_b = \frac{1}{n}\boldsymbol{X}_b^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_b. \tag{S117}$$

Then, by Lemma S5, the $(p-q)$-dimensional random vector $\mathbf{X}_{ib}-\mathbf{\Sigma}_{ba}\mathbf{\Sigma}_{aa}^{-1}\mathbf{X}_{ia} \sim N(\mathbf{0}, \mathbf{\Sigma}_{bb\cdot a})$ and is independent of $\mathbf{X}_{ia}$.

It follows that the conditional distribution of $\mathbf{X}_{ib}$ given $\mathbf{X}_{ia}$ is characterized by

$$\mathcal{L}(\mathbf{X}_{ib}|\mathbf{X}_{ia}) = N(\mathbf{\Sigma}_{ba}\mathbf{\Sigma}_{aa}^{-1}\mathbf{X}_{ia}, \mathbf{\Sigma}_{bb\cdot a}).$$

Therefore, we have from the definition $\mathbf{\Sigma}_{b|a} = E\mathrm{Cov}(\mathbf{X}_{1b}|\mathbf{X}_{1a})$ that

$$\mathbf{\Sigma}_{b|a} = \mathbf{\Sigma}_{bb\cdot a}. \tag{S118}$$

Write

$$\boldsymbol{X}_b^T = \big(\mathbf{X}_{1b} - \mathbf{\Sigma}_{ba}\mathbf{\Sigma}_{aa}^{-1}\mathbf{X}_{1a}, \cdots, \mathbf{X}_{nb} - \mathbf{\Sigma}_{ba}\mathbf{\Sigma}_{aa}^{-1}\mathbf{X}_{na}\big) + \big(\mathbf{\Sigma}_{ba}\mathbf{\Sigma}_{aa}^{-1}\mathbf{X}_{1a}, \cdots, \mathbf{\Sigma}_{ba}\mathbf{\Sigma}_{aa}^{-1}\mathbf{X}_{na}\big).$$

Notice the last two vectors are both normal and they are independent since that $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are i.i.d. from the p-dimensional population $N(\mathbf{0}, \mathbf{\Sigma})$. Moreover, we can also write

$$\boldsymbol{X}_b^T = (\mathbf{V}_1, \cdots, \mathbf{V}_n) + \mathbf{\Sigma}_{ba}\mathbf{\Sigma}_{aa}^{-1}\mathbf{X}_a^T \tag{S119}$$

where $\mathbf{V}_1, \cdots, \mathbf{V}_n$ are i.i.d. $(p-q)$-dimensional random vectors with distribution $N(\mathbf{0}, \mathbf{\Sigma}_{bb\cdot a})$ also independent of $\mathbf{X}_a$. By definition (S117),

$$\begin{aligned}\hat{\mathbf{\Sigma}}_{b|a} =& \frac{1}{n}\boldsymbol{X}_b^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_b \\ =& \frac{1}{n}\big[(\mathbf{V}_1, \cdots, \mathbf{V}_n) + \mathbf{\Sigma}_{ba}\mathbf{\Sigma}_{aa}^{-1}\mathbf{X}_a^T\big](\mathbf{I}_n - \mathbf{H}_a)\big[(\mathbf{V}_1, \cdots, \mathbf{V}_n) + \mathbf{\Sigma}_{ba}\mathbf{\Sigma}_{aa}^{-1}\mathbf{X}_a^T\big]^T\end{aligned}$$

$$=\frac{1}{n}(\mathbf{V}_1,\cdots,\mathbf{V}_n)(\mathbf{I}_n-\mathbf{H}_a)(\mathbf{V}_1,\cdots,\mathbf{V}_n)^T, \tag{S120}$$

since $\mathbf{X}_a^T(\mathbf{I}_n-\mathbf{H}_a)=\mathbf{0}$. Let $\{\xi_{ij};\, 1\leqslant i\leqslant p-q, 1\leqslant j\leqslant n\}$ be i.i.d. $N(0,1)$-distributed random variables independent of $\mathbf{X}_a$. Without loss of generality, assume $\mathbf{V}_j=\mathbf{\Sigma}_{bb\cdot a}^{1/2}\cdot(\xi_{ij})_{(p-q)\times 1}$ for each j. Therefore,

$$(\mathbf{V}_1,\cdots,\mathbf{V}_n)=\mathbf{\Sigma}_{bb\cdot a}^{1/2}\cdot(\xi_{ij})_{(p-q)\times n}. \tag{S121}$$

Since the $n\times n$ idempotent matrix $\mathbf{H}_a$ has rank q, we know $\mathbf{I}_n-\mathbf{H}_a$ has rank $n-q$. As a function of $\boldsymbol{X}_a$, $\mathbf{I}_n-\mathbf{H}_a$ is independent of $(\mathbf{V}_1,\cdots,\mathbf{V}_n)$. As a result, there exists an $n\times n$ random orthogonal matrix $\mathbf{\Gamma}$ independent of $\{\xi_{ij};\, 1\leqslant i\leqslant p-q, 1\leqslant j\leqslant n\}$ such that

$$\mathbf{I}_n-\mathbf{H}_a=\mathbf{\Gamma}\begin{pmatrix}\mathbf{I}_{(n-q)\times(n-q)} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}_{q\times q}\end{pmatrix}\mathbf{\Gamma}^T,$$

where the three $\mathbf{0}$ above are matrices of entries 0 with proper size. By the orthogonal invariance of i.i.d. standard normals, $(\xi_{ij})_{(p-q)\times n}\mathbf{\Gamma}$ has the same distribution as that of $(\xi_{ij})_{(p-q)\times n}$. From (S120) and (S121), we have

$$\begin{aligned}\hat{\mathbf{\Sigma}}_{b|a}&\overset{d}{=}\frac{1}{n}\mathbf{\Sigma}_{bb\cdot a}^{1/2}\cdot(\xi_{ij})_{(p-q)\times n}\begin{pmatrix}\mathbf{I}_{(n-q)\times(n-q)} & \mathbf{0}\\ \mathbf{0}\mathbf{0}_{q\times q} & \end{pmatrix}(\xi_{ij})_{(p-q)\times n}^T\mathbf{\Sigma}_{bb\cdot a}^{1/2}\\ &=\frac{1}{n}\mathbf{\Sigma}_{bb\cdot a}^{1/2}\cdot\mathbf{W}\mathbf{W}^T\mathbf{\Sigma}_{bb\cdot a}^{1/2}\end{aligned}$$

where $\mathbf{W}=(\xi_{ij})_{(p-q)\times(n-q)}$. $\square$

Lemma S12. *Recall the notations in Lemma S11 and* (S118)*. Let* $\lambda_1, \cdots \lambda_{p-q}$ *be the eigenvalues of* $\boldsymbol{\Sigma}_{b|a} = \boldsymbol{\Sigma}_{bb\cdot a}$*. Let* $\{\mathbf{w}_i;\, 1 \leqslant i \leqslant p-q\}$ *be i.i.d.* $(n-q)$*-dimensional vectors whose entries are i.i.d.* $N(0,1)$*. Then* $tr\big(\hat{\boldsymbol{\Sigma}}_{b|a}\big) \overset{d}{=} \frac{1}{n}\sum_{i=1}^{p-q} \lambda_i \|\mathbf{w}_i\|^2$ *and*

$$tr\big(\hat{\boldsymbol{\Sigma}}^2_{b|a}\big) \overset{d}{=} \frac{1}{n^2}\sum_{i=1}^{p-q} \lambda_i^2 (\mathbf{w}_i^T\mathbf{w}_i)^2 + \frac{2}{n^2}\sum_{1\leqslant i<j\leqslant p-q} \lambda_i\lambda_j (\mathbf{w}_i^T\mathbf{w}_j)^2.$$

In particular,

$$Etr\big(\hat{\boldsymbol{\Sigma}}^2_{b|a}\big) = \frac{n-q}{n^2}\cdot\big[tr(\boldsymbol{\Sigma}_{b|a})\big]^2 + \frac{(n-q)(n-q+1)}{n^2} tr\big(\boldsymbol{\Sigma}^2_{b|a}\big). \tag{S122}$$

Proof of Lemma S12. Let $\mathbf{W}$ be a $(p-q)\times(n-q)$ matrix whose entries are i.i.d. $N(0,1)$. Immediately, $E\hat{\boldsymbol{\Sigma}}_{b|a} = \frac{1}{n}\boldsymbol{\Sigma}^{1/2}_{bb\cdot a}\cdot E(\mathbf{W}\mathbf{W}^T)\boldsymbol{\Sigma}^{1/2}_{bb\cdot a} = \frac{n-q}{n}\cdot\boldsymbol{\Sigma}_{bb\cdot a}$. Using the identity $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ for any matrices $\mathbf{A}$ and $\mathbf{B}$, we know that

$$\begin{aligned}\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}\big) &\overset{d}{=} \frac{1}{n}\mathrm{tr}\big(\boldsymbol{\Sigma}^{1/2}_{bb\cdot a}\cdot\mathbf{W}\mathbf{W}^T\boldsymbol{\Sigma}^{1/2}_{bb\cdot a}\big)\\ &= \frac{1}{n}\mathrm{tr}\big(\mathbf{W}^T\boldsymbol{\Sigma}_{bb\cdot a}\mathbf{W}\big)\end{aligned} \tag{S123}$$

and

$$\begin{aligned}\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}^2_{b|a}\big) &\overset{d}{=} \frac{1}{n^2}\mathrm{tr}\big[\boldsymbol{\Sigma}^{1/2}_{bb\cdot a}\cdot\mathbf{W}\mathbf{W}^T\boldsymbol{\Sigma}_{bb\cdot a}\mathbf{W}\mathbf{W}^T\boldsymbol{\Sigma}^{1/2}_{bb\cdot a}\big]\\ &= \frac{1}{n^2}\mathrm{tr}\big[(\mathbf{W}^T\boldsymbol{\Sigma}_{bb\cdot a}\mathbf{W})^2\big].\end{aligned} \tag{S124}$$

Since $\lambda_1, \cdots, \lambda_{p-q}$ are the eigenvalues of $\boldsymbol{\Sigma}_{bb\cdot a}$, we are able to decompose $\boldsymbol{\Sigma}_{bb\cdot a} = \boldsymbol{\Gamma}_1^T\boldsymbol{\Lambda}\boldsymbol{\Gamma}_1$, where $\boldsymbol{\Gamma}_1$ is an orthogonal matrix and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_{p-q})$.

By the orthogonal invariance of i.i.d. $N(0,1)$-entries, we get

$$\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}\big) \overset{d}{=} \frac{1}{n}\mathrm{tr}\big[(\mathbf{W}^T\boldsymbol{\Lambda}\mathbf{W})\big] \;\text{ and }\; \mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}^2_{b|a}\big) \overset{d}{=} \frac{1}{n^2}\mathrm{tr}\big[(\mathbf{W}^T\boldsymbol{\Lambda}\mathbf{W})^2\big].$$

Furthermore write $\mathbf{W}^T = (\mathbf{w}_1, \cdots, \mathbf{w}_{p-q})$. Then $\mathbf{w}_1, \cdots, \mathbf{w}_{p-q}$ are i.i.d. $(n-q)$-dimensional vectors of distribution $N(\mathbf{0}, \mathbf{I}_{n-q})$. Hence,

$$\mathbf{W}^T\boldsymbol{\Lambda}\mathbf{W} = \sum_{i=1}^{p-q} \lambda_i \mathbf{w}_i\mathbf{w}_i^T,$$

which gives

$$\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}\big) \overset{d}{=} \frac{1}{n}\sum_{i=1}^{p-q} \lambda_i \|\mathbf{w}_i\|^2. \tag{S125}$$

Additionally, we have

$$(\mathbf{W}^T\boldsymbol{\Lambda}\mathbf{W})^2 = \sum_{i=1}^{p-q} \lambda_i^2(\mathbf{w}_i\mathbf{w}_i^T)^2 + 2\sum_{1\leqslant i<j\leqslant p-q} \lambda_i\lambda_j(\mathbf{w}_i\mathbf{w}_i^T)(\mathbf{w}_j\mathbf{w}_j^T).$$

Observe

$$\mathrm{tr}[(\mathbf{w}_i\mathbf{w}_i^T)^2] = \mathrm{tr}[(\mathbf{w}_i\mathbf{w}_i^T)\mathbf{w}_i\mathbf{w}_i^T] = \mathrm{tr}[(\mathbf{w}_i^T\mathbf{w}_i\mathbf{w}_i^T\mathbf{w}_i] = (\mathbf{w}_i^T\mathbf{w}_i)^2.$$

Similarly, $\mathrm{tr}[(\mathbf{w}_i\mathbf{w}_i^T)(\mathbf{w}_j\mathbf{w}_j^T)] = (\mathbf{w}_i^T\mathbf{w}_j)^2$. Then we end up with

$$\mathrm{tr}\big[(\mathbf{W}^T\boldsymbol{\Lambda}\mathbf{W})^2\big] = \sum_{i=1}^{p-q} \lambda_i^2(\mathbf{w}_i^T\mathbf{w}_i)^2 + 2\sum_{1\leqslant i<j\leqslant p-q} \lambda_i\lambda_j(\mathbf{w}_i^T\mathbf{w}_j)^2.$$

It follows from (S124) that

$$\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}^2_{b|a}\big) \overset{d}{=} \frac{1}{n^2}\sum_{i=1}^{p-q} \lambda_i^2(\mathbf{w}_i^T\mathbf{w}_i)^2 + \frac{2}{n^2}\sum_{1\leqslant i<j\leqslant p-q} \lambda_i\lambda_j(\mathbf{w}_i^T\mathbf{w}_j)^2. \tag{S126}$$

Since $\mathbf{w}_1, \cdots, \mathbf{w}_{p-q}$ are i.i.d. $N(\mathbf{0}, \mathbf{I}_{n-q})$, we know $\mathbf{w}_i^T \mathbf{w}_i \sim \chi^2(n-q)$ and $\mathbf{w}_i^T \mathbf{w}_j \overset{d}{=} \|\mathbf{w}_i\| \cdot \eta$, where $\eta \sim N(0,1)$ and is independent of $\|\mathbf{w}_i\|$. Recall $E\chi^2(m) = m$ and $\mathrm{Var}(\chi^2(m)) = 2m$ for any integer $m \geqslant 1$. Thus, by independence we have $E(\|\mathbf{w}_i\|^2 \cdot \eta^2) = n - q$. From (S126), we obtain

$$E\mathrm{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}^2\big) = \frac{1}{n^2} \cdot \big[2(n-q) + (n-q)^2\big] \sum_{i=1}^{p-q} \lambda_i^2 + \frac{2}{n^2} \cdot (n-q) \sum_{1 \leqslant i < j \leqslant p-q} \lambda_i \lambda_j.$$

Using the identity

$$2 \sum_{1 \leqslant i < j \leqslant p-q} \lambda_i \lambda_j = \Big(\sum_{i=1}^{p-q} \lambda_i\Big)^2 - \sum_{i=1}^{p-q} \lambda_i^2,$$

we arrive at

$$E\mathrm{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}^2\big) = \frac{n-q}{n^2} \cdot \big[\mathrm{tr}(\mathbf{\Sigma}_{b|a})\big]^2 + \frac{(n-q)(n-q+1)}{n^2} \mathrm{tr}\big(\mathbf{\Sigma}_{b|a}^2\big)$$

by using (S118). This completes the proof of the lemma. □

Lemma S13. *Let $r \geqslant 1$ and $\mathbf{w}_1$ and $\mathbf{w}_2$ be i.i.d. r-dimensional random vectors with distribution $N(\mathbf{0}, \mathbf{I}_r)$. Then the following are true.*

(i) Var$((\mathbf{w}_1^T \mathbf{w}_1)^2) = 8r(r+2)(r+3)$.

(ii) Var$((\mathbf{w}_1^T \mathbf{w}_2)^2) = 2r(r+3)$.

(iii) Cov$((\mathbf{w}_1^T \mathbf{w}_2)^2, (\mathbf{w}_1^T \mathbf{w}_3)^2) = 2r$.

(iv) Cov$((\mathbf{w}_1^T \mathbf{w}_1)^2, (\mathbf{w}_1^T \mathbf{w}_2)^2) = 4r(r+2)$.

Proof of Lemma S13. It is well-known that

$$E[\chi^2(r)^m] = r(r+2) \cdots (r+2m-2) \tag{S127}$$

for all integers $m \geqslant 1$. We will use this formula to prove the results.

(i) Since $\mathbf{w}_1^T\mathbf{w}_1 \sim \chi^2(r)$, we have

$$E(\|\mathbf{w}_1\|^4) = E(\mathbf{w}_1^T\mathbf{w}_1)^2 = r(r+2) \tag{S128}$$

and $E(\mathbf{w}_1^T\mathbf{w}_1)^4 = r(r+2)(r+4)(r+6)$. This leads to

$$\begin{aligned}\mathrm{Var}((\mathbf{w}_1^T\mathbf{w}_1)^2) =& r(r+2)(r+4)(r+6) - [r(r+2)]^2 \\ =& 8r(r+2)(r+3).\end{aligned}$$

(ii) By Proposition 7.3 from Eaton (1983) or Theorem 1.5.6 from Muirhead (1982), it holds that

$$\|\mathbf{w}_i\| \ \text{ and } \ \mathbf{e}_i := \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|} \ \text{ are independent} \tag{S129}$$

for $i = 1, 2$. Therefore, $(\mathbf{w}_1^T\mathbf{w}_2)^2 = \|\mathbf{w}_1\|^2 \cdot (\mathbf{e}_1^T\mathbf{w}_2)^2 \stackrel{d}{=} \|\mathbf{w}_1\|^2 \cdot \eta^2$, where $\eta \sim N(0,1)$ and η is independent of $\|\mathbf{w}_1\|$. Consequently, by (S128),

$$\begin{aligned}\mathrm{Var}((\mathbf{w}_1^T\mathbf{w}_2)^2) =& E(\|\mathbf{w}_1\|^4 \cdot \eta^4) - \left[E(\|\mathbf{w}_1\|^2 \cdot \eta^2)\right]^2 \\ =& r(r+2) \cdot 3 - r^2 \\ =& 2r(r+3).\end{aligned}$$

(iii) By (S129), we have

$$(\mathbf{w}_1^T\mathbf{w}_2)^2(\mathbf{w}_1^T\mathbf{w}_3)^2 = \|\mathbf{w}_1\|^4 \cdot (\mathbf{e}_1^T\mathbf{w}_2)^2 \cdot (\mathbf{e}_1^T\mathbf{w}_3)^2.$$

Easily, be checking their covariance, we know $\mathbf{e}_1^T\mathbf{w}_2$ and $\mathbf{e}_1^T\mathbf{w}_3$ are i.i.d. $N(0,1)$. This implies that $\|\mathbf{w}_1\|^4$, $(\mathbf{e}_1^T\mathbf{w}_2)^2$ and $(\mathbf{e}_1^T\mathbf{w}_3)^2$ are independent.

Thus, from (S128) we obtain

$$\begin{aligned}\mathrm{Cov}((\mathbf{w}_1^T\mathbf{w}_2)^2,(\mathbf{w}_1^T\mathbf{w}_3)^2) =& E\|\mathbf{w}_1\|^4\cdot 1\cdot 1 - E(\mathbf{w}_1^T\mathbf{w}_2)^2\cdot E(\mathbf{w}_1^T\mathbf{w}_3)^2\\ =& r(r+2)-r^2\\ =& 2r.\end{aligned}$$

(iv) Write $(\mathbf{w}_1^T\mathbf{w}_1)^2(\mathbf{w}_1^T\mathbf{w}_2)^2 = \|\mathbf{w}_1\|^6\cdot(\mathbf{e}_1^T\mathbf{w}_2)^2$. From the independence between $\|\mathbf{w}_1\|$ and $\mathbf{e}_1^T\mathbf{w}_2$, we conclude that

$$\begin{aligned}&\mathrm{Cov}((\mathbf{w}_1^T\mathbf{w}_1)^2,(\mathbf{w}_1^T\mathbf{w}_2)^2)\\ =& E\|\mathbf{w}_1\|^6\cdot E(\mathbf{e}_1^T\mathbf{w}_2)^2 - E(\mathbf{w}_1^T\mathbf{w}_1)^2\cdot E(\mathbf{w}_1^T\mathbf{w}_2)^2\\ =& r(r+2)(r+4)-r(r+2)\cdot r\\ =& 4r(r+2).\end{aligned}$$

Thus, the verification is completed. □

Lemma S14. *Let $\hat{\mathbf{\Sigma}}_{b|a}$ and $\mathbf{\Sigma}_{b|a}$ be as in* (S10) *and* (S118)*, respectively. Then*

$$Var\big(tr\big(\hat{\mathbf{\Sigma}}_{b|a}^2\big)\big) \leqslant \frac{1024(n-q)(n+p-2q)^2}{n^4}\cdot tr(\mathbf{\Sigma}_{b|a}^4).$$

Proof of Lemma S14. Set $r = n-q$ and $s = p-q$. Let $\lambda_1,\cdots\lambda_s$ be the eigenvalues of $\mathbf{\Sigma}_{b|a} = \mathbf{\Sigma}_{bb\cdot a}$. Standard computation gives

$$\mathrm{Var}(\gamma_1+\cdots+\gamma_s) = \sum_{i=1}^s \mathrm{Var}(\gamma_i) + 2\sum_{1\leqslant i<j\leqslant s}\mathrm{Cov}(\gamma_i,\gamma_j), \qquad \text{(S130)}$$

for any random variables $\gamma_1, \cdots, \gamma_s$. Then, from Lemma S12 we see

$$
\begin{aligned}
n^4 \cdot \operatorname{Var}\big(\operatorname{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}^2\big)\big) =& \sum_{i=1}^{s} \lambda_i^4 \operatorname{Var}\big((\mathbf{w}_i^T \mathbf{w}_i)^2\big) + 4 \sum_{1 \leqslant i < j \leqslant s} \lambda_i^2 \lambda_j^2 \operatorname{Var}\big((\mathbf{w}_i^T \mathbf{w}_j)^2\big) \\
&+ 8 \sum_{1 \leqslant i < j \leqslant s} \sum_{(k,l) \in A_{i,j}} \lambda_i \lambda_j \lambda_k \lambda_l \operatorname{Cov}\big((\mathbf{w}_i^T \mathbf{w}_j)^2, (\mathbf{w}_k^T \mathbf{w}_l)^2\big) \\
&+ 4 \sum_{1 \leqslant i < j \leqslant s} \sum_{k \in \{i,j\}} \lambda_i \lambda_j \lambda_k^2 \operatorname{Cov}\big((\mathbf{w}_i^T \mathbf{w}_j)^2, (\mathbf{w}_k^T \mathbf{w}_k)^2\big) \\
:=& C_1 + C_2 + C_3 + C_4, \qquad \text{(S131)}
\end{aligned}
$$

where $A_{i,j} = \{(k,l) \neq (i,j);\, 1 \leqslant k < l \leqslant s,\ \{k,l\} \cap \{i,j\} \neq \varnothing\}$. By Lemma S13(i),

$$
C_1 = 8r(r+2)(r+3) \sum_{i=1}^{s} \lambda_i^4 \leqslant 8(r+3)^3 \cdot \operatorname{tr}(\mathbf{\Sigma}_{b|a}^4). \qquad \text{(S132)}
$$

By Lemma S13(ii),

$$
\begin{aligned}
C_2 =& 4r(r+3) \cdot 2 \sum_{1 \leqslant i < j \leqslant s} \lambda_i^2 \lambda_j^2 \\
\leqslant& 4(r+3)^2 \Big(\sum_{1 \leqslant i \leqslant s} \lambda_i^2 \Big)^2 \\
=& 4(r+3)^2 \left[\operatorname{tr}(\mathbf{\Sigma}_{b|a}^2)\right]^2. \qquad \text{(S133)}
\end{aligned}
$$

By Lemma S13(iii),

$$
\begin{aligned}
C_3 =& 2r \cdot 8 \sum_{1 \leqslant i < j \leqslant s} \sum_{(k,l) \in A_{i,j}} \lambda_i \lambda_j \lambda_k \lambda_l \\
\leqslant& 16r \sum_{1 \leqslant i,j,k \leqslant s} \lambda_i \lambda_j \lambda_k^2
\end{aligned}
$$

$$=16r \cdot \left[\mathrm{tr}(\boldsymbol{\Sigma}_{b|a})\right]^2 \cdot \mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^2). \tag{S134}$$

By Lemma S13(iv),

$$\begin{aligned} C_4 \leqslant & 4r(r+2) \cdot 4 \sum_{1 \leqslant i < j \leqslant s} \sum_{k \in \{i,j\}} \lambda_i \lambda_j \lambda_k^2 \\ \leqslant & 16(r+2)^2 \sum_{1 \leqslant i,j \leqslant s} \lambda_i \lambda_j^3 \\ = & 16(r+2)^2 \left[\mathrm{tr}(\boldsymbol{\Sigma}_{b|a})\right] \cdot \mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^3). \end{aligned} \tag{S135}$$

Let I be uniformly distributed over $\{1, \cdots, s\}$. By Hölder's inequality, $(E\lambda_I^\alpha)^{1/\alpha} \leqslant (E\lambda_I^\beta)^{1/\beta}$ for any $0 < \alpha < \beta$. This says that

$$\left(\frac{\lambda_1^\alpha + \cdots + \lambda_s^\alpha}{s}\right)^{1/\alpha} \leqslant \left(\frac{\lambda_1^\beta + \cdots + \lambda_s^\beta}{s}\right)^{1/\beta}.$$

By taking $\alpha = 1, 2, 3$, respectively, and $\beta = 4$, we have

$$\mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^i) \leqslant s^{1-(i/4)} \cdot [\mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^4)]^{i/4} \tag{S136}$$

for $i = 1, 2, 3$. Consequently,

$$\begin{aligned} & C_2 \leqslant 4(r+3)^2 s \cdot \mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^4), \quad C_3 \leqslant 16rs^2 \cdot \mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^4), \\ & C_4 \leqslant 16(r+2)^2 s \cdot \mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^4). \end{aligned}$$

Combing the above with (S131), we get

$$\begin{aligned} \mathrm{Var}\big(\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}^2\big)\big) \leqslant & \frac{8(r+3)^3 + 4(r+3)^2 s + 16rs^2 + 16(r+2)^2 s}{n^4} \cdot \mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^4) \\ \leqslant & \frac{16(r+3)(r+s+3)^2}{n^4} \cdot \mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^4). \end{aligned}$$

Using the fact $3 \leqslant 3r$ to see $16(r+3)(r+s+3)^2 \leqslant 4^5 r(r+s)^2$ and plugging in $r = n - q$ and $s = p - q$, we obtain

$$\operatorname{Var}\big(\operatorname{tr}\big(\hat{\mathbf{\Sigma}}^2_{b|a}\big)\big) \leqslant \frac{4^5(n-q)(n+p-2q)^2}{n^4} \cdot \operatorname{tr}(\mathbf{\Sigma}^4_{b|a}),$$

which concludes the proof. □

Lemma S15. *Let $\hat{\mathbf{\Sigma}}_{b|a}$ and $\mathbf{\Sigma}_{b|a}$ be as in* (S10) *and* (S118)*, respectively. Then, there exists a constant $K > 0$ not depending on p, q, n or $\mathbf{\Sigma}_{b|a}$ such that*

$$Var\Big(\big(tr\big(\hat{\mathbf{\Sigma}}_{b|a}\big)\big)^2\Big) \leqslant \frac{K(n-q)^2(p-q)^2}{n^4} tr\big(\mathbf{\Sigma}^4_{b|a}\big).$$

Proof of Lemma S15. Set $r = n - q$ and $s = p - q$. Let $\lambda_1, \cdots \lambda_s$ be the eigenvalues of the $s \times s$ matrix $\mathbf{\Sigma}_{b|a}$. Let $\xi_1, \cdots, \xi_s$ be i.i.d. random variables with distribution $\chi^2(r)$, with $E\xi_1 = r$ and $\operatorname{Var}(\xi_1) = 2r$. From (S125),

$$n^4 \cdot \operatorname{Var}\big(\big(\operatorname{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}\big)\big)^2\big) = \operatorname{Var}\Big(\Big(\sum_{i=1}^{s} \lambda_i \xi_i\Big)^2\Big). \tag{S137}$$

Let U be a random variable with mean μ. Write $U^2 = (U-\mu)^2 + 2\mu U - \mu^2$. Trivially, $\operatorname{Var}(W_1 + W_2) \leqslant 2\operatorname{Var}(W_1) + 2\operatorname{Var}(W_2)$ for any random variables W_1 and W_2. Then

$$\begin{aligned}\operatorname{Var}(U^2) \leqslant & 2\operatorname{Var}((U-\mu)^2) + 8\mu^2 \operatorname{Var}(U) \\ \leqslant & 2E(U-\mu)^4 + 8\mu^2 \operatorname{Var}(U).\end{aligned}$$

Now consider U to be $U = \sum_{i=1}^{s} \lambda_i \xi_i$, with $\mu = EU = r\,\mathrm{tr}(\mathbf{\Sigma}_{b|a})$. Use (S136) to see

$$\mu^2 \cdot \mathrm{Var}(U) = 2\mu^2 \sum_{i=1}^{s} \lambda_i^2 = 2r^2 \left[\mathrm{tr}(\mathbf{\Sigma}_{b|a})\right]^2 \cdot \mathrm{tr}\big(\mathbf{\Sigma}_{b|a}^2\big) \leqslant 2r^2 s^2 \cdot \mathrm{tr}(\mathbf{\Sigma}_{b|a}^4). \tag{S138}$$

By the Marcinkiewtz-Zygmund inequality [see, e.g., Theorem 2 on p. 386 from Chow and Teicher (1997)] and the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned} E(U-\mu)^4 =& E\Big[\sum_{i=1}^{s} \lambda_i(\xi_i - r)\Big]^4 \\ \leqslant& K_1 \cdot E\Big[\sum_{i=1}^{s} \lambda_i^2(\xi_i - r)^2\Big]^2 \\ \leqslant& K_1 \cdot \Big(\sum_{i=1}^{s} \lambda_i^4\Big) \cdot E \sum_{i=1}^{s} (\xi_i - r)^4, \end{aligned}$$

where K_1 is a numerical constant. Write $\xi_1 = \sum_{j=1}^{r} \eta_i$, where $\eta_1, \cdots, \eta_r$ are i.i.d. $\chi^2(1)$-distributed random variables. By Corollary 2 on p. 387 from Chow and Teicher (1997), $E[\sum_{j=1}^{r}(\eta_i - 1)]^4 \leqslant K_2 r^2$, where K_2 is also a numerical constant. That is, $E(\xi_1 - r)^4 \leqslant K_2 r^2$. In summary, we get $E(U-\mu)^4 \leqslant (K_1K_2) r^2 s \cdot \mathrm{tr}\big(\mathbf{\Sigma}_{b|a}^4\big)$ and hence

$$n^4 \cdot \mathrm{Var}\Big(\big(\mathrm{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}\big)\big)^2\Big) = \mathrm{Var}(U^2) \leqslant (2K_1K_2) r^2 s \cdot \mathrm{tr}\big(\mathbf{\Sigma}_{b|a}^4\big) + 16 r^2 s^2 \cdot \mathrm{tr}(\mathbf{\Sigma}_{b|a}^4)$$

from (S137) and (S138). The conclusion then follows by using the fact $r^2 s \leqslant r^2 s^2$. □

Proposition S1. *Let* $\hat{\mathbf{\Sigma}}_{b|a}$ *and* $\mathbf{\Sigma}_{b|a}$ *be as in* (S10) *and* (S118). *Set* $r = n - q$ *and*

$$\widehat{tr\big(\mathbf{\Sigma}_{b|a}^2\big)} = \frac{n^2}{r(r+1)} \cdot \Big[tr\big(\hat{\mathbf{\Sigma}}_{b|a}^2\big) - \frac{1}{r}\big(tr\big(\hat{\mathbf{\Sigma}}_{b|a}\big)\big)^2 \Big].$$

Assume (S14) *holds. If* $q = o(n)$, $q = o(p)$ *and* $p = o(n^3)$ *then* $\widehat{tr\big(\mathbf{\Sigma}_{b|a}^2\big)}/tr\big(\mathbf{\Sigma}_{b|a}^2\big) \to 1$ *in probability as* $p \to \infty$. *Hence,* $\widehat{tr\big(\mathbf{\Sigma}_{b|a}^2\big)}$ *is an asymptotic ratio-consistent estimator of* $tr\big(\mathbf{\Sigma}_{b|a}^2\big)$.

Proof of Proposition S1. Since $\mathrm{Var}(W_1 + W_2) \leqslant 2\,\mathrm{Var}(W_1) + 2\,\mathrm{Var}(W_2)$ for any random variables W_1 and W_2, by Lemma S14 and Lemma S15, there exists a constant $K > 0$ independent of n, p, q or $\mathbf{\Sigma}_{b|a}$ such that

$$\begin{aligned}
&\mathrm{Var}\Big(\mathrm{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}^2\big) - \frac{1}{r}\big(\mathrm{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}\big)\big)^2\Big) \\
\leqslant &K \cdot \Big[\frac{(n-q)(n+p-2q)^2}{n^4} + \frac{(n-q)^2(p-q)^2}{n^4 r^2}\Big] \cdot \mathrm{tr}\big(\mathbf{\Sigma}_{b|a}^4\big) \\
\leqslant &K \cdot \Big[\frac{(n+p)^2}{n^3} + \frac{p^2}{n^4}\Big] \cdot \mathrm{tr}(\mathbf{\Sigma}_{b|a}^4).
\end{aligned} \tag{S139}$$

On the other hand, by (S122),

$$\begin{aligned}
&E\Big[\mathrm{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}^2\big) - \frac{1}{r}\big(\mathrm{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}\big)\big)^2\Big] \\
=&\frac{r}{n^2} \cdot \big[\mathrm{tr}(\mathbf{\Sigma}_{b|a})\big]^2 + \frac{r(r+1)}{n^2}\mathrm{tr}\big(\mathbf{\Sigma}_{b|a}^2\big) - \frac{1}{r}E\big[\big(\mathrm{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}\big)\big)^2\big].
\end{aligned} \tag{S140}$$

Define $s = p - q$. Let $\lambda_1, \cdots \lambda_s$ be the eigenvalues of the $s \times s$ matrix $\mathbf{\Sigma}_{b|a}$.

By (S125),

$$n^2 \cdot E\Big[\big(\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}\big)\big)^2\Big] = E\Big(\sum_{i=1}^{s} \lambda_i \|\mathbf{w}_i\|^2\Big)^2$$
$$= \sum_{i=1}^{s} \lambda_i^2 E\big(\|\mathbf{w}_i\|^4\big) + 2 \sum_{1 \leqslant i < j \leqslant s} \lambda_i \lambda_j E\big(\|\mathbf{w}_i\|^2 \|\mathbf{w}_j\|^2\big),$$

where $\|\mathbf{w}_1\|^2, \cdots, \|\mathbf{w}_s\|^2$ are i.i.d. $\chi^2(r)$-distributed random variables. From (S128), $E(\|\mathbf{w}_1\|^4) = r(r+2)$. Thus, we have from independence that

$$n^2 \cdot E\Big[\big(\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}\big)\big)^2\Big] = r(r+2) \cdot \mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}^2\big) + r^2 \cdot 2 \sum_{1 \leqslant i < j \leqslant s} \lambda_i \lambda_j$$
$$= r(r+2) \cdot \mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}^2\big) + r^2 \Big\{\big[\mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}\big)\big]^2 - \mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}^2\big)\Big\}$$
$$= 2r \cdot \mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}^2\big) + r^2 \cdot \big[\mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}\big)\big]^2,$$

or equivalently,

$$\frac{1}{r} E\Big[\big(\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}\big)\big)^2\Big] = \frac{2}{n^2} \cdot \mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}^2\big) + \frac{r}{n^2} \cdot \big[\mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}\big)\big]^2.$$

Combine this with (S140) to get

$$EF_p = E\Big[\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}^2\big) - \frac{1}{r}\big(\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}\big)\big)^2\Big] = \frac{r^2 + r - 2}{n^2} \cdot \mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}^2\big), \qquad \text{(S141)}$$

where $F_p = \mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}^2\big) - \frac{1}{r}\big[\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}\big)\big]^2$. Under Assumption (S14), $\tau^{-1} < \lambda_{min}(\boldsymbol{\Sigma}_{b|a}) \leqslant \lambda_{max}(\boldsymbol{\Sigma}_{b|a}) < \tau$ for some constant $\tau > 1$. This then implies

$$\mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^4) = \lambda_1^4 + \cdots + \lambda_s^4 \leqslant s\tau^4 \quad \text{and} \quad \big[\mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}^2\big)\big]^2 = (\lambda_1^2 + \cdots + \lambda_s^2)^2 \geqslant s^2\tau^{-4}.$$

We deduce from (S139) and (S141) that

$$\begin{aligned}\mathrm{Var}\Big(\frac{F_p}{EF_p}\Big) \leqslant& K \cdot \frac{n(n+p)^2+p^2}{(r^2+r-2)^2} \cdot \frac{\mathrm{tr}(\mathbf{\Sigma}_{b|a}^4)}{\big[\mathrm{tr}\big(\mathbf{\Sigma}_{b|a}^2\big)\big]^2} \\ \leqslant& (3K)\tau^8 \cdot \frac{n^3+np^2}{(r^2+r-2)^2} \cdot \frac{1}{s},\end{aligned}$$

where the fact $n(n+p)^2+p^2 \leqslant 3(n^3+np^2)$ is used in the second inequality. Under the conditions $q/n \to 0$, $q/p \to 0$ and $p = o(n^3)$, we know $r/n \to 1$ and $s/p \to 1$, hence the last term is of order $\frac{1}{np} + \frac{p}{n^3} \to 0$. This leads to $\frac{F_p}{EF_p} \to 1$ in probability and we conclude from (S141) that

$$\Big[\mathrm{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}^2\big) - \frac{1}{r}\big(\mathrm{tr}\big(\hat{\mathbf{\Sigma}}_{b|a}\big)\big)^2\Big] \cdot \Big[\frac{r(r+1)}{n^2} \cdot \mathrm{tr}\big(\mathbf{\Sigma}_{b|a}^2\big)\Big]^{-1} \to 1$$

in probability. The proof is complete. □

Lemma S 16. *Let $\mathbf{y}_1, \cdots, \mathbf{y}_m$ be i.i.d. p-dimensional random vectors with distribution $N(\mathbf{0}, \mathbf{B})$, where $\mathbf{B}$ is a $p \times p$ non-negative definite matrix. Let $\mathbf{x} \in \mathbb{R}^m$ be a non-zero random vector independent of $\mathbf{y}_1, \cdots, \mathbf{y}_m$. Write $\mathbf{x}_1 = \mathbf{x}/\|\mathbf{x}\|$ and the $m \times p$ matrix $(\mathbf{y}_1 \cdots \mathbf{y}_m)^T = (\mathbf{z}_1 \cdots \mathbf{z}_p)$. Then $(\mathbf{x}_1^T\mathbf{z}_1 \cdots \mathbf{x}_1^T\mathbf{z}_p)^T \sim N(\mathbf{0}, \mathbf{B})$ and is independent of $\|\mathbf{x}\|$.*

Proof of Lemma S16. Write $\mathbf{B} = (b_{ij})_{p\times p}$ and $(\mathbf{y}_1 \cdots \mathbf{y}_m)^T = (y_{ij})_{m\times p}$. Use the fact that the rows of the matrix are i.i.d. to see $E(y_{ki}y_{lj}) = 0$ if $k \neq l$ and $E(y_{ki}y_{lj}) = b_{ij}$ if $k = l$. Thus,

$$E(\mathbf{z}_i\mathbf{z}_j^T) = \big(E(y_{ki}y_{lj})\big)_{1\leqslant k,l\leqslant p} = b_{ij}\mathbf{I}_p. \tag{S142}$$

Obviously, $\mathbf{z}_i \sim N(0, b_{ii}\mathbf{I}_m)$ for each i. Since $\mathbf{x}_1$ is a unit random vector, then conditional on $\mathbf{x}$, we know $\mathbf{x}_1^T\mathbf{z}_1, \cdots, \mathbf{x}_1^T\mathbf{z}_p$ are jointly Gaussian random variables with $\mathbf{x}_1^T\mathbf{z}_i \sim N(0, b_{ii})$ for each $1 \leqslant i \leqslant p$. Let us check their covariance matrix. In fact, conditional on $\mathbf{x}$,

$$E\big[(\mathbf{x}_1^T\mathbf{z}_i)(\mathbf{x}_1^T\mathbf{z}_j)\big] = E\big(\mathbf{x}_1^T\mathbf{z}_i\mathbf{z}_j^T\mathbf{x}_1\big) = \mathbf{x}_1^T E\big(\mathbf{z}_i\mathbf{z}_j^T\big)\mathbf{x}_1 = b_{ij},$$

by (S142) and the fact $\mathbf{x}_1^T\mathbf{x}_1 = 1$. In summary, conditional on $\mathbf{x}$, the random vector $(\mathbf{x}_1^T\mathbf{z}_1 \cdots \mathbf{x}_1^T\mathbf{z}_p)^T \sim N(\mathbf{0}, \mathbf{B})$. Since $N(\mathbf{0}, \mathbf{B})$ is free of $\mathbf{x}_1$, this implies, unconditionally, it is also true that $(\mathbf{x}_1^T\mathbf{z}_1 \cdots \mathbf{x}_1^T\mathbf{z}_p)^T \sim N(\mathbf{0}, \mathbf{B})$. Finally, for any set $F \subset \mathbb{R}^p$ and $G \subset [0, \infty)$,

$$\begin{aligned} P\big((\mathbf{x}_1^T\mathbf{z}_1 \cdots \mathbf{x}_1^T\mathbf{z}_p)^T \in F,\ \|\mathbf{x}\| \in G\big) =& E\big[P\big((\mathbf{x}_1^T\mathbf{z}_1 \cdots \mathbf{x}_1^T\mathbf{z}_p)^T \in F\big|\mathbf{x}\big) \cdot I(\|\mathbf{x}\| \in G)\big] \\ =& E\big[P\big(N(\mathbf{0}, \mathbf{B}) \in F\big) \cdot I(\|\mathbf{x}\| \in G)\big] \\ =& P\big(N(\mathbf{0}, \mathbf{B}) \in F\big) \cdot P(\|\mathbf{x}\| \in G). \end{aligned}$$

This shows that $(\mathbf{x}_1^T\mathbf{z}_1 \cdots \mathbf{x}_1^T\mathbf{z}_p)^T$ and $\|\mathbf{x}\|$ are independent. □

Lemma S17 (Bai and Silverstein (2010))**.** *Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ non-random matrix and $\mathbf{X} = (x_1, \cdots, x_n)^T$ be a random vector of independent entries. Assume that $\mathrm{E}x_i = 0$, $E\left|x_i\right|^2 = 1$ and $E\left|x_j\right|^\ell \leqslant \nu_\ell$. Then, for any $p \geqslant 1$,*

$$E\left|\mathbf{X}^T\mathbf{A}\mathbf{X} - \operatorname{tr}\mathbf{A}\right|^p \leqslant C_p\left(\left(\nu_4 \operatorname{tr}\left(\mathbf{A}\mathbf{A}^T\right)\right)^{p/2} + \nu_{2p}\operatorname{tr}\left(\mathbf{A}\mathbf{A}^T\right)^{p/2}\right),$$

where C_p is a constant depending on p only.

Lemma S18. *Assume $\boldsymbol{\varepsilon} = (\epsilon_1, \cdots, \epsilon_n)^T \in \mathbb{R}^n$ satisfies that $\epsilon_1, \cdots, \epsilon_n$ are i.i.d. with $E\epsilon_1 = 0$, $E\epsilon_1^2 = \sigma^2$ and $E|\epsilon_1|^{2k} < \infty$ for an integer $k \geqslant 2$. Let $\mathbf{H}$ be an $n \times n$ symmetric, random matrix satisfying $\mathbf{H}^2 = \mathbf{H}$ and rank$(\mathbf{H}) = q$. Assume $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I}_n)$ and that $\mathbf{u}, \mathbf{H}$ and $\boldsymbol{\varepsilon}$ are independent. Then*

(i) $\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon} = (n-q)\sigma^2 + O_p(\sqrt{n})$;

(ii) $E(\mathbf{u}^T\mathbf{H}\boldsymbol{\varepsilon})^{2k} \leqslant C(k,\sigma) \cdot q^k$, where $C(k,\sigma)$ is a constant depending on k, σ only.

Proof of Lemma S18. It is easy to see $E(\boldsymbol{\varepsilon}^T\mathbf{A}\boldsymbol{\varepsilon}) = \sigma^2\,\mathrm{tr}(A)$ for any matrix $\mathbf{A}$.

(i) Obviously, the conditional mean $E[\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon}|\mathbf{H}] = (n-q)\sigma^2$. Take another expectation to see $E[\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon})] = (n-q)\sigma^2$. By Lemma S17,

$$
\begin{aligned}
E\big[\big(\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon} - (n-q)\sigma^2\big)^2\big|\mathbf{H}\big] \leqslant & C \cdot E\epsilon_1^4 \cdot \mathrm{tr}(\mathbf{I}_n - \mathbf{H}) \\
= & C \cdot E\epsilon_1^4 \cdot (n-q),
\end{aligned}
$$

where $C > 0$ is a constant free of n, q, k, σ. By taking another expectation, we get $\mathrm{Var}(\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon}) \leqslant Cn$. Thus, by the Chebyshev inequality,

$$
P\big(|\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon} - (n-q)\sigma^2| \geqslant A\sqrt{n}\big) \leqslant \frac{\mathrm{Var}(\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon})}{A^2} \leqslant \frac{C}{A^2}.
$$

This implies $\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon} = (n-q)\sigma^2 + O_p(\sqrt{n})$.

(ii) First, $E(\boldsymbol{\varepsilon}^T\mathbf{H}\boldsymbol{\varepsilon}) = \sigma^2 \cdot \mathrm{tr}(\mathbf{H}) = \sigma^2 q$. Trivially,

$$E\big(\boldsymbol{\varepsilon}^T\mathbf{H}\boldsymbol{\varepsilon}\big)^k \leqslant 2^k \cdot \big[E\big|\boldsymbol{\varepsilon}^T\mathbf{H}\boldsymbol{\varepsilon} - \sigma^2 q\big|^k + (\sigma^2 q)^k\big].$$

From Lemma S17 and the fact $\mathbf{H}^l = \mathbf{H}$ for any $l = 1, 2, \cdots$, we see

$$\begin{aligned} E\big|\boldsymbol{\varepsilon}^T\mathbf{H}\boldsymbol{\varepsilon} - \sigma^2 q\big|^k \leqslant & C_k \cdot \big[(E\epsilon_1^4)^{k/2} + E\epsilon_1^{2k}\big] \cdot \big[(\mathrm{tr}(\mathbf{H}))^{k/2} + \mathrm{tr}(\mathbf{H})\big] \\ \leqslant & C_k q^{k/2}, \end{aligned}$$

where C_k is a constant depending on k only. Therefore,

$$E\big(\boldsymbol{\varepsilon}^T\mathbf{H}\boldsymbol{\varepsilon}\big)^k \leqslant C(k,\sigma) \cdot q^k, \tag{S143}$$

where $C(k,\sigma)$ is a constant depending on k, σ only. On the other hand, noting that the n entries of $\mathbf{u}$ are i.i.d. $N(0,1)$, by conditioning on $\mathbf{H}$ and $\boldsymbol{\varepsilon}$, the random variable $\mathbf{u}^T\mathbf{H}\boldsymbol{\varepsilon}$ has the distribution of $G \cdot \|\mathbf{H}\boldsymbol{\varepsilon}\|$, where $G \sim N(0,1)$. Or, equivalently,

$$\text{given } \{\mathbf{H}, \boldsymbol{\varepsilon}\}, \text{ random variable } \mathbf{u}^T\mathbf{H}\boldsymbol{\varepsilon} \text{ has distribution } G \cdot \big(\boldsymbol{\varepsilon}^T\mathbf{H}\boldsymbol{\varepsilon}\big)^{1/2}.$$

In particular, this implies that

$$E(\mathbf{u}^T\mathbf{H}\boldsymbol{\varepsilon})^{2k} = E\big(G^{2k}\big) \cdot E\big(\boldsymbol{\varepsilon}^T\mathbf{H}\boldsymbol{\varepsilon}\big)^k.$$

By combining this, (S143) and the fact $E(G^{2k}) = (2k-1)!!$, we obtain

$$E(\mathbf{u}^T\mathbf{H}\boldsymbol{\varepsilon})^{2k} \leqslant (2k - 1!! \cdot C(k,\sigma) \cdot q^k.$$

The proof is completed. □

Recall random error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^T \in \mathbb{R}^n$ in the linear regression model from Section S2. The components $\{\varepsilon_i;\, 1 \leqslant i \leqslant n\}$ are assumed to be i.i.d. random variables.

Proposition S2. *Let $T_{max}^{(3)}$ be defined as in* (S12). *Assume* (S13) *and* (S14) *are true and* (2.3) *holds with "$\boldsymbol{\Sigma}$" replaced by "$\boldsymbol{\Sigma}_{b|a}$". Suppose $p = o(n^3)$, $q = o(p)$, $q \leqslant n^\delta$ for some $\delta \in (0,1)$ and $E(|\epsilon_1|^\ell) < \infty$ with $\ell = 14(1-\delta)^{-1}$. Then, under H_0 from* (S8), *$T_{max}^{(3)} - 2\log(p-q) + \log\log(p-q)$ converges weakly to a distribution with cdf $F(x) = \exp\{-\frac{1}{\sqrt{\pi}}\exp(-\frac{x}{2})\}$.*

Proof of Proposition S2. Recall the notation between (S9) and (S10). In particular, $\mathbf{H}_a = \mathbf{X}_a(\mathbf{X}_a^T\mathbf{X}_a)^{-1}\mathbf{X}_a^T$. Under the null hypothesis in (S8), $\mathbf{Y} = \boldsymbol{X}_a\boldsymbol{\beta}_a + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^T \in \mathbb{R}^n$ and the random errors $\{\varepsilon_i;\, 1 \leqslant i \leqslant n\}$ are i.i.d. with $E\varepsilon_i = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2$ for each i. Also, $\boldsymbol{\varepsilon}$ is assumed to be independent of $\{\boldsymbol{X}_i;\, 1 \leqslant i \leqslant n\}$. Use $\boldsymbol{X}_a^T(\mathbf{I}_n - \mathbf{H}_a) = \mathbf{0}$ and $(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_a = \mathbf{0}$ to see

$$
\begin{aligned}
\boldsymbol{Y}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{Y} =& (\boldsymbol{X}_a\boldsymbol{\beta}_a + \boldsymbol{\varepsilon})^T(\mathbf{I}_n - \mathbf{H}_a)(\boldsymbol{X}_a\boldsymbol{\beta}_a + \boldsymbol{\varepsilon}) \\
=& \boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}. \qquad \text{(S144)}
\end{aligned}
$$

Recalling (S10), we denote

$$\boldsymbol{X}_b = (\boldsymbol{X}_{1b}, \cdots, \boldsymbol{X}_{nb})^T \quad \text{and} \quad \tilde{\mathbf{X}}_b = (\tilde{\mathbf{X}}_{q+1}, \cdots, \tilde{\mathbf{X}}_p) = (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_b, \tag{S145}$$

where $\boldsymbol{X}_b$ and $\tilde{\mathbf{X}}_b$ are $n \times (p-q)$ matrices. Write $\boldsymbol{X}_b = (\mathbf{w}_{q+1}, \cdots, \mathbf{w}_p)$. Then the last assertion from (S145) says that $\tilde{\mathbf{X}}_j = (\mathbf{I}_n - \mathbf{H}_a)\mathbf{w}_j$ for each $q+1 \leqslant j \leqslant p$. This leads to

$$\begin{aligned}
\boldsymbol{Y}^T\tilde{\mathbf{X}}_j\tilde{\mathbf{X}}_j^T\boldsymbol{Y} =& (\boldsymbol{X}_a\boldsymbol{\beta}_a + \boldsymbol{\varepsilon})^T(\mathbf{I}_n - \mathbf{H}_a)\mathbf{w}_j\mathbf{w}_j^T(\mathbf{I}_n - \mathbf{H}_a)(\boldsymbol{X}_a\boldsymbol{\beta}_a + \boldsymbol{\varepsilon}) \\
=& \boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H}_a)\mathbf{w}_j\mathbf{w}_j^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon} \\
=& \boldsymbol{\varepsilon}^T\tilde{\mathbf{X}}_j\tilde{\mathbf{X}}_j^T\boldsymbol{\varepsilon}
\end{aligned}$$

via the facts $\boldsymbol{X}_a^T(\mathbf{I}_n - \mathbf{H}_a) = \mathbf{0}$ and $(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_a = \mathbf{0}$ again. Note that $(\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j)^{-1}$ is a scalar. By the definition of $T_{max}^{(3)}$ in (S12), we derive

$$\begin{aligned}
T_{max}^{(3)} =& \max_{q+1 \leqslant j \leqslant p} \frac{\boldsymbol{Y}^T\tilde{\mathbf{X}}_j(\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j)^{-1}\tilde{\mathbf{X}}_j^T\boldsymbol{Y}}{\boldsymbol{Y}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{Y}/(n-q)} \\
=& \frac{(n-q)\sigma^2}{\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}} \cdot \max_{q+1 \leqslant j \leqslant p} \left\{n^{-1}\sigma^{-2}\boldsymbol{\varepsilon}^T\tilde{\mathbf{X}}_j(n^{-1}\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j)^{-1}\tilde{\mathbf{X}}_j^T\boldsymbol{\varepsilon}\right\}.
\end{aligned}$$

Set

$$W_p = \frac{(n-q)\sigma^2}{\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}}.$$

By the triangle inequality,

$$\left|T_{max}^{(3)} - W_p \cdot \max_{q+1 \leqslant j \leqslant p} \left\{n^{-1}\sigma^{-2}\boldsymbol{\varepsilon}^T\tilde{\mathbf{X}}_j\tilde{\mathbf{X}}_j^T\boldsymbol{\varepsilon}\right\}\right|$$

$$\leqslant W_p \cdot \max_{q+1\leqslant j\leqslant p} \Big\{ n^{-1}\sigma^{-2}\boldsymbol{\varepsilon}^T\tilde{\mathbf{X}}_j\tilde{\mathbf{X}}_j^T\boldsymbol{\varepsilon} \cdot \big|1-(n^{-1}\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j)^{-1}\big| \Big\}. \tag{S146}$$

Define

$$H_{31} = \max_{q+1\leqslant j\leqslant p} \Big\{ n^{-1}\sigma^{-2}\boldsymbol{\varepsilon}^T\tilde{\mathbf{X}}_j\tilde{\mathbf{X}}_j^T\boldsymbol{\varepsilon} \Big\};$$

$$H_{32} = \max_{q+1\leqslant j\leqslant p} \Big\{ n^{-1}\sigma^{-2}\boldsymbol{\varepsilon}^T\tilde{\mathbf{X}}_j\tilde{\mathbf{X}}_j^T\boldsymbol{\varepsilon} \cdot \big|1-(n^{-1}\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j)^{-1}\big| \Big\}.$$

By Lemma S18(i), we know $1/W_p = 1 + O_p(n^{-1/2})$, which implies

$$W_p = 1 + O_p\Big(\frac{1}{\sqrt{n}}\Big). \tag{S147}$$

We will show that, as $p \to \infty$,

$$H_{31}' := H_{31} - 2\log(p-q) + \log\log(p-q) \to \text{a distribution with cdf } e^{-e^{-x/2}/\sqrt{\pi}}; \tag{S148}$$

$$H_{32} \to 0 \text{ in probability.} \tag{S149}$$

Assuming they are true, using assumption $p = o(n^3)$ and (S146) we have

$$\begin{aligned} T_{max}^{(3)} =& \Big[1 + O_p\Big(\frac{1}{\sqrt{n}}\Big)\Big] \cdot \big[H_{31}' + 2\log(p-q) - \log\log(p-q) + o_p(1)\big] \\ =& H_{31}' + 2\log(p-q) - \log\log(p-q) + o_p(1). \end{aligned} \tag{S150}$$

This and (S148) show that $T_{max}^{(3)} - 2\log(p-q) + \log\log(p-q)$ converges weakly to a distribution with cdf $F(x) = \exp\{-\frac{1}{\sqrt{\pi}}\exp(-\frac{x}{2})\}$, $x \in \mathbb{R}$. It remains to prove (S148) and (S149), which will be done in two steps as follows.

Step 1: the proof of (S148). By (S119) and (S145), $\boldsymbol{X}_b^T = (\mathbf{V}_1, \cdots, \mathbf{V}_n) + \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}\mathbf{X}_a^T$, where $\mathbf{V}_1, \cdots, \mathbf{V}_n$ are i.i.d. $(p-q)$-dimensional random vectors with distribution $N(\mathbf{0}, \boldsymbol{\Sigma}_{bb\cdot a})$ and they are also independent of $\mathbf{X}_a$. In particular, $\mathbf{V}_1, \cdots, \mathbf{V}_n$ are independent of $\mathbf{H}_a$, a function of $\mathbf{X}_a$. As $(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_a = \mathbf{0}$, we see that

$$
\begin{aligned}
(\tilde{\mathbf{X}}_{q+1}, \cdots, \tilde{\mathbf{X}}_p) =& (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_b \\
=& (\mathbf{I}_n - \mathbf{H}_a)\big[\mathbf{X}_a\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ba}^T + (\mathbf{V}_1, \cdots, \mathbf{V}_n)^T\big] \\
=& (\mathbf{I}_n - \mathbf{H}_a)(\mathbf{V}_1, \cdots, \mathbf{V}_n)^T.
\end{aligned}
$$

Write $(\mathbf{V}_1, \cdots, \mathbf{V}_n)^T = (\mathbf{u}_{q+1}, \cdots, \mathbf{u}_p)$. In other words, each $\mathbf{u}_{q+1}, \cdots, \mathbf{u}_p \in \mathbb{R}^n$ is a column of $(\mathbf{V}_1, \cdots, \mathbf{V}_n)^T$. Immediately we have

$$
\tilde{\mathbf{X}}_j = (\mathbf{I}_n - \mathbf{H}_a)\mathbf{u}_j \quad \text{and} \quad \mathbf{u}_j \sim N(\mathbf{0}, \mathbf{I}_n) \tag{S151}
$$

for each $j = q+1, \cdots, p$. This means $\tilde{\mathbf{X}}_j^T\boldsymbol{\varepsilon} = \mathbf{u}_j^T\boldsymbol{\varepsilon} - \mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon}$, which then leads to

$$
(\tilde{\mathbf{X}}_j^T\boldsymbol{\varepsilon})^2 = \big(\mathbf{u}_j^T\boldsymbol{\varepsilon}\big)^2 + (\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon})^2 - 2(\boldsymbol{\varepsilon}^T\mathbf{u}_j)\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon} \tag{S152}
$$

by using the fact $\mathbf{u}_j^T\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^T\mathbf{u}_j \in \mathbb{R}$. Obviously, it holds that

$$
\begin{aligned}
&\Big|\max_{q+1\leqslant j\leqslant p}\big\{n^{-1}(\tilde{\mathbf{X}}_j^T\boldsymbol{\varepsilon})^2\big\} - \max_{q+1\leqslant j\leqslant p}\big\{n^{-1}\big(\mathbf{u}_j^T\boldsymbol{\varepsilon}\big)^2\big\}\Big| \\
\leqslant& \max_{q+1\leqslant j\leqslant p}\big\{n^{-1}(\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon})^2\big\} + 2\cdot\max_{q+1\leqslant j\leqslant p}\big\{n^{-1}\big|(\boldsymbol{\varepsilon}^T\mathbf{u}_j)\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon}\big|\big\}.
\end{aligned} \tag{S153}
$$

A key observation is that the three random quantities $\mathbf{u}_j$, $\mathbf{H}_a$ and $\boldsymbol{\varepsilon}$ are independent. By the last assertion from (S151) and Lemma S18(ii), we have

$$E(\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon})^{2k} \leqslant C(k,\sigma)\cdot q^k \tag{S154}$$

for any $q+1 \leqslant j \leqslant n$, where $k = [\frac{6}{1-\delta}]+1$ and $C(k,\sigma)$ is a constant depending on k, σ only. We next show

$$\frac{1}{n}\cdot \max_{q+1\leqslant j\leqslant p}(\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon})^2 \to 0 \quad \text{and} \quad \frac{1}{n}\cdot \max_{q+1\leqslant j\leqslant p}\left|(\boldsymbol{\varepsilon}^T\mathbf{u}_j)\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon}\right| \to 0 \tag{S155}$$

in probability as $p\to\infty$. In fact, for any $\beta>0$,

$$\begin{aligned} P\Big(\frac{1}{n}\cdot \max_{q+1\leqslant j\leqslant p}(\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon})^2 \geqslant \beta\Big) &\leqslant p\cdot \max_{q+1\leqslant j\leqslant p} P\Big((\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon})^2 \geqslant n\beta\Big)\\ &\leqslant p\cdot \max_{q+1\leqslant j\leqslant p}\frac{E(\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon})^{2k}}{(n\beta)^k}. \end{aligned}$$

Therefore, from assumption $q\leqslant n^\delta$ for some $\delta\in(0,1)$, we have

$$P\Big(\frac{1}{n}\cdot \max_{q+1\leqslant j\leqslant p}(\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon})^2 \geqslant \beta\Big) \leqslant C(\beta,k,\sigma)\cdot\frac{pq^k}{n^k} \leqslant C(\beta,k,\sigma)\cdot\frac{p}{n^{k(1-\delta)}},$$

where $C(\beta,k,\sigma)$ is a constant depending on β, k and σ. Since $k(1-\delta)>6$, we get the first limit of (S155) by using the assumption $p=o(n^3)$. For the second limit, by setting $\boldsymbol{\varepsilon}_1=\boldsymbol{\varepsilon}/\|\boldsymbol{\varepsilon}\|$ we have

$$\frac{1}{n}\cdot \max_{q+1\leqslant j\leqslant p}\left|(\boldsymbol{\varepsilon}^T\mathbf{u}_j)\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon}\right| = \frac{\|\boldsymbol{\varepsilon}\|}{\sqrt{n}}\cdot\frac{1}{\sqrt{n}}\cdot \max_{q+1\leqslant j\leqslant p}\left|(\boldsymbol{\varepsilon}_1^T\mathbf{u}_j)\mathbf{u}_j^T\mathbf{H}_a\boldsymbol{\varepsilon}\right|.$$

By the law of large numbers, $\|\boldsymbol{\varepsilon}\|/\sqrt{n} \to \sigma$ in probability. Thus, to get the second limit of (S155), it suffices to prove

$$\frac{1}{\sqrt{n}} \cdot \max_{q+1 \leqslant j \leqslant p} \left|(\boldsymbol{\varepsilon}_1^T \mathbf{u}_j)\mathbf{u}_j^T \mathbf{H}_a \boldsymbol{\varepsilon}\right| \to 0 \tag{S156}$$

in probability. Similar to an earlier argument, we have from the fact $\boldsymbol{\varepsilon}_1^T \mathbf{u}_j \sim N(0,1)$ that

$$\begin{aligned} P\Big(\frac{1}{\sqrt{n}} \cdot \max_{q+1 \leqslant j \leqslant p} \left|(\boldsymbol{\varepsilon}_1^T \mathbf{u}_j)\mathbf{u}_j^T \mathbf{H}_a \boldsymbol{\varepsilon}\right| \geqslant \beta\Big) \leqslant & p \cdot \max_{q+1 \leqslant j \leqslant p} P\Big(|\boldsymbol{\varepsilon}_1^T \mathbf{u}_j| \cdot |\mathbf{u}_j^T \mathbf{H}_a \boldsymbol{\varepsilon}| \geqslant \sqrt{n}\beta\Big) \\ \leqslant & \frac{p}{(\sqrt{n}\beta)^k} \cdot \max_{q+1 \leqslant j \leqslant p} E\big[|\boldsymbol{\varepsilon}_1^T \mathbf{u}_j|^k \cdot |\mathbf{u}_j^T \mathbf{H}_a \boldsymbol{\varepsilon}|^k\big] \\ \leqslant & C(k,\beta) \cdot \frac{p}{n^{k/2}} \cdot \big[E(\mathbf{u}_1^T \mathbf{H}_a \boldsymbol{\varepsilon})^{2k}\big]^{1/2} \end{aligned}$$

by the Cauchy-Schwartz inequality and $E(\mathbf{u}_j^T \mathbf{H}_a \boldsymbol{\varepsilon})^{2k} = E(\mathbf{u}_1^T \mathbf{H}_a \boldsymbol{\varepsilon})^{2k}$, where $C(k,\beta)$ is a constant depending on k and β only. Use (S154) to see

$$P\Big(\frac{1}{\sqrt{n}} \cdot \max_{q+1 \leqslant j \leqslant p} \left|(\boldsymbol{\varepsilon}_1^T \mathbf{u}_j)\mathbf{u}_j^T \mathbf{H}_a \boldsymbol{\varepsilon}\right| \geqslant \beta\Big) \leqslant C(k,\beta,\sigma) \cdot \frac{pq^{k/2}}{n^{k/2}} = O\Big(\frac{p}{n^{k(1-\delta)/2}}\Big),$$

by the assumption $q \leqslant n^{\delta}$ again. Since $p = o(n^3)$ and $k(1-\delta) > 6$, we get (S156) and then the second limit of (S155).

Now we study $\max_{q+1 \leqslant j \leqslant p} \big\{n^{-1}\big(\mathbf{u}_j^T \boldsymbol{\varepsilon}\big)^2\big\}$ in (S153). Since $\epsilon_1, \cdots, \epsilon_n$ are i.i.d. with mean zero and variance σ^2. By assumption, $E(|\epsilon_1|^{\ell}) < \infty$ with $\ell = 14(1-\delta)^{-1}$. This concludes $E(\epsilon_1^{14}) < \infty$. Write $\|\boldsymbol{\varepsilon}\|^2 = \epsilon_1^2 + \cdots + \epsilon_n^2$. By the central limit theorem, $n^{-1}\sigma^{-2}\|\boldsymbol{\varepsilon}\|^2 = 1 + O_p(n^{-1/2})$. Use $\boldsymbol{\varepsilon}_1 = \boldsymbol{\varepsilon}/\|\boldsymbol{\varepsilon}\|$ to see

$$\begin{aligned}\max_{q+1\leqslant j\leqslant p}\left\{n^{-1}\sigma^{-2}\left(\mathbf{u}_j^T\boldsymbol{\varepsilon}\right)^2\right\} =&\left(n^{-1}\sigma^{-2}\|\boldsymbol{\varepsilon}\|^2\right)\cdot\max_{q+1\leqslant j\leqslant p}\left(\mathbf{u}_j^T\boldsymbol{\varepsilon}_1\right)^2\\ =&\left[1+O_p(n^{-1/2})\right]\cdot\max_{q+1\leqslant j\leqslant p}\left(\mathbf{u}_j^T\boldsymbol{\varepsilon}_1\right)^2. \qquad \text{(S157)}\end{aligned}$$

Review $\mathbf{V}_1,\cdots,\mathbf{V}_n$ are i.i.d. $(p-q)$-dimensional random vectors with distribution $N(\mathbf{0},\boldsymbol{\Sigma}_{bb\cdot a})$ and $(\mathbf{V}_1,\cdots,\mathbf{V}_n)^T=(\mathbf{u}_{q+1},\cdots,\mathbf{u}_p)$. By Lemma S16,

$$(\mathbf{u}_{q+1}^T\boldsymbol{\varepsilon}_1,\cdots,\mathbf{u}_p^T\boldsymbol{\varepsilon}_1)^T\sim N(\mathbf{0},\boldsymbol{\Sigma}_{bb\cdot a}). \qquad \text{(S158)}$$

Based on assumption, (2.3) holds with "$\boldsymbol{\Sigma}$" replaced by "$\boldsymbol{\Sigma}_{b|a}$", we know Assumption (2.2) also holds by the discussion below (2.3). We then have from Theorem 2 that

$$P\Big(\max_{q+1\leqslant j\leqslant p}(\mathbf{u}_j^T\boldsymbol{\varepsilon}_1)^2-2\log(p-q)+\log\log(p-q)\leqslant x\Big)\to\exp\Big\{-\frac{1}{\sqrt{\pi}}e^{-x/2}\Big\}$$

as $p\to\infty$. Denote $U_p=\max_{q+1\leqslant j\leqslant p}(\mathbf{u}_j^T\boldsymbol{\varepsilon}_1)^2-2\log(p-q)+\log\log(p-q)$. Then the above says U_p converges weakly to the Gumbel distribution with cdf $F(x)=\exp\{-\frac{1}{\sqrt{\pi}}e^{-x/2}\}$. It then follows from (S157) that

$$\begin{aligned}\max_{q+1\leqslant j\leqslant p}\left\{n^{-1}\sigma^{-2}\left(\mathbf{u}_j^T\boldsymbol{\varepsilon}\right)^2\right\} =&\left[1+O_p(n^{-1/2})\right]\cdot\left[U_p+2\log(p-q)-\log\log(p-q)\right]\\ =&U_p+2\log(p-q)-\log\log(p-q)+O_p(n^{-1/2}\log p).\end{aligned}$$

Obviously, the last term goes to zero since $p=o(n^3)$ by assumption. Use the Slutsky lemma to see that

$$\max_{q+1\leqslant j\leqslant p}\left\{n^{-1}\left(\mathbf{u}_j^T\boldsymbol{\varepsilon}\right)^2\right\}-2\log(p-q)+\log\log(p-q)$$

converges weakly to the Gumbel distribution with cdf $F(x) = \exp\{-\frac{1}{\sqrt{\pi}}e^{-x/2}\}$. We then obtain (S148) by (S153) and (S155).

Step 2: the proof of (S149). Easily, by definition,

$$|H_{32}| \leqslant |H_{31}| \cdot \max_{q+1\leqslant j\leqslant p} \left|1 - (n^{-1}\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j)^{-1}\right|.$$

By (S148), $H_{31} = O(\log p)$. Thus, to prove (S149), it suffices to show

$$H'_{32} := \max_{q+1\leqslant j\leqslant p} \left|1 - (n^{-1}\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j)^{-1}\right| = o_p\Big(\frac{1}{\log p}\Big). \tag{S159}$$

Now we prove this assertion. From assumption $q \leqslant n^{\delta}$ for some $\delta \in (0,1)$, we choose $\delta' \in (0, \min\{1/2, 1-\delta\})$. This indicates that

$$\max\left\{\frac{1}{2}, \delta\right\} < 1 - \delta' < 1. \tag{S160}$$

Note that, for $x > 0$ and $s \in (0,1)$ satisfying $|1 - x^{-1}| \geqslant s$, one of the two inequalities $x^{-1} \geqslant 1 + s$ and $x^{-1} \leqslant 1 - s$ must hold. Equivalently, $x \leqslant \frac{1}{1+s}$ or $x \geqslant \frac{1}{1-s}$. Both of them imply that $|x - 1| \geqslant \frac{s}{1+s} \geqslant \frac{1}{2}s$. Consequently,

$$\begin{aligned} P\Big(H'_{32} \geqslant n^{-\delta'}\Big) \leqslant & p \cdot \max_{q+1\leqslant j\leqslant p} P\Big(\left|1 - (n^{-1}\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j)^{-1}\right| \geqslant n^{-\delta'}\Big) \\ \leqslant & p \cdot \max_{q+1\leqslant j\leqslant p} P\Big(\left|\frac{1}{n}\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j - 1\right| \geqslant \frac{1}{2}n^{-\delta'}\Big). \end{aligned}$$

Recalling (S151) and that the matrix $(\mathbf{I}_n - \mathbf{H}_a)^2 = \mathbf{I}_n - \mathbf{H}_a$ has rank $n-q$, we know $\tilde{\mathbf{X}}_j^T\tilde{\mathbf{X}}_j \sim \chi^2(n-q)$. Denote $n' = n-q$. Observe that $|\frac{y}{n} - 1| \geqslant (1/2)n^{-\delta'}$ implies $|y - n| \geqslant (1/2)n^{1-\delta'}$, which implies $|y - n'| \geqslant (1/2)n^{1-\delta'} - q$. Use (S160) and the assumption $q \leqslant n^{\delta}$ to see $|y - n'|/\sqrt{n'} \geqslant [(1/2)n^{1-\delta'} -$

$q]/\sqrt{n'} \sim (1/2)n^{(1/2)-\delta'}$. By definition, $0 < (1/2) - \delta' < 1/2$. Similar to (S108), we have

$$
\begin{aligned}
P\big(H'_{32} \geqslant n^{-\delta'}\big) \leqslant & p \cdot P\Big(\Big|\frac{1}{\sqrt{n'}}(\chi^2(n') - n')\Big| \geqslant \frac{1}{3}n^{(1/2)-\delta'}\Big) \\
\leqslant & p \cdot \exp\big(C \cdot n^{1-2\delta'}\big),
\end{aligned}
$$

which is equal to $o(1)$ by assumption $p = o(n^3)$. This says $H'_{32} = o_p(n^{-\delta'})$, which implies (S159), and hence (S149) holds as aforementioned. ▫

Recall the error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^T \in \mathbb{R}^n$ in the linear regression from Section S2. We assume $\{\varepsilon_i; 1 \leqslant i \leqslant n\}$ are i.i.d. random variables.

Proposition S3. *Let $T_{sum}^{(3)}$ be defined as in* (S11). *Assume* (S13) *and* (S14) *hold and* (2.3) *also holds with "$\boldsymbol{\Sigma}$" replaced by "$\boldsymbol{\Sigma}_{b|a}$". Suppose $p = o(n^3)$, $q = o(p)$, $q \leqslant n^\delta$ for some $\delta \in (0,1)$ and $E(|\epsilon_1|^\ell) < \infty$ with $\ell = 14(1-\delta)^{-1}$. Under H_0 from* (S8) *we have $T_{sum}^{(3)}$ converges to $N(0,1)$ in distribution as $p \to \infty$.*

Proof of Proposition S3. Recall the notation from Section S2. In particular,

$$
\begin{aligned}
&\hat{\boldsymbol{\Sigma}}_{b|a} = n^{-1}\tilde{\mathbf{X}}_b^T\tilde{\mathbf{X}}_b, \qquad \hat{\sigma}^2 = (n-q)^{-1}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}, \\
&\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^2)} = \frac{n^2}{(n+1-q)(n-q)}\Big\{\mathrm{tr}\big(\hat{\boldsymbol{\Sigma}}_{b|a}^2\big) - \frac{1}{n-q}\mathrm{tr}^2(\hat{\boldsymbol{\Sigma}}_{b|a})\Big\}
\end{aligned}
$$

and

$$T_{sum}^{(3)} = \frac{n^{-1}\hat{\boldsymbol{\varepsilon}}^T \boldsymbol{X}_b \boldsymbol{X}_b^T \hat{\boldsymbol{\varepsilon}} - n^{-1}(n-q)(p-q)\hat{\sigma}^2}{\sqrt{2\hat{\sigma}^4 \operatorname{tr}\widehat{\left(\boldsymbol{\Sigma}_{b|a}^2\right)}}}. \tag{S161}$$

Next we will first derive a workable form for the main ingredient $\hat{\boldsymbol{\varepsilon}}^T \boldsymbol{X}_b \boldsymbol{X}_b^T \hat{\boldsymbol{\varepsilon}}$ above.

Recall $\hat{\boldsymbol{\varepsilon}} = (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{Y}$. Under the null hypothesis in (S8), $\mathbf{Y} = \boldsymbol{X}_a\boldsymbol{\beta}_a + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^T \in \mathbb{R}^n$, and $\{\varepsilon_i;\ 1 \leqslant i \leqslant n\}$ are i.i.d. random variables with $E\varepsilon_1 = 0$ and $\operatorname{Var}(\varepsilon_1) = \sigma^2$. By assumption, $\{\varepsilon_i;\ 1 \leqslant i \leqslant n\}$ are also independent of $\{\boldsymbol{X}_i;\ 1 \leqslant i \leqslant n\}$. Recalling (S145), we denote

$$\boldsymbol{X}_b = (\boldsymbol{X}_{1b}, \cdots, \boldsymbol{X}_{nb})^T \quad \text{and} \quad \tilde{\mathbf{X}}_b = (\tilde{\mathbf{X}}_{q+1}, \cdots, \tilde{\mathbf{X}}_p) = (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_b.$$

Since $\mathbf{H}_a = \mathbf{X}_a(\mathbf{X}_a^T\mathbf{X}_a)^{-1}\mathbf{X}_a^T$, both $\boldsymbol{X}_a^T(\mathbf{I}_n - \mathbf{H}_a) = \mathbf{0}$ and $(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_a = \mathbf{0}$. Thus, $\hat{\boldsymbol{\varepsilon}} = (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}$. By definition,

$$\hat{\sigma}^2 = \frac{1}{n-q}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} = \frac{1}{n-q}\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon} \tag{S162}$$

and

$$\begin{aligned}
\hat{\boldsymbol{\varepsilon}}^T \boldsymbol{X}_b \boldsymbol{X}_b^T \hat{\boldsymbol{\varepsilon}} &= \boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{X}_b\boldsymbol{X}_b^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon} \\
&= \|(\tilde{\mathbf{X}}_{q+1}, \cdots, \tilde{\mathbf{X}}_p)^T\boldsymbol{\varepsilon}\|^2 \\
&= \sum_{j=q+1}^{p} (\tilde{\mathbf{X}}_j^T\boldsymbol{\varepsilon})^2.
\end{aligned}$$

It is easy to verify that $E(\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}) = \sigma^2\operatorname{tr}(\mathbf{I}_n - \mathbf{H}_a) = (n-q)\sigma^2$.

Consequently,

$$E\hat{\sigma}^2 = \sigma^2. \tag{S163}$$

According to (S151), $\tilde{\mathbf{X}}_j = (\mathbf{I}_n - \mathbf{H}_a)\mathbf{u}_j$ and $\mathbf{u}_j \sim N(\mathbf{0}, \mathbf{I}_n)$. Hence,

$$(\tilde{\mathbf{X}}_j^T \boldsymbol{\varepsilon})^2 = [\mathbf{u}_j^T (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}]^2 = \|(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}\|^2 \cdot (\mathbf{u}_j^T \mathbf{e}_2)^2, \tag{S164}$$

where $\mathbf{e}_2 := (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}/\|(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}\|$. Eventually, we arrive at an ideal form to work with, that is,

$$\frac{1}{n}\hat{\boldsymbol{\varepsilon}}^T \boldsymbol{X}_b \boldsymbol{X}_b^T \hat{\boldsymbol{\varepsilon}} = \frac{1}{n}\boldsymbol{\varepsilon}^T (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon} \cdot \sum_{j=q+1}^{p} (\mathbf{u}_j^T \mathbf{e}_2)^2. \tag{S165}$$

Now we start to prove the central limit theorem. The assumption $E(|\epsilon_1|^\ell) < \infty$ with $\ell = 14(1-\delta)^{-1}$ implies that $E(|\epsilon_1|^{2k}) < \infty$ with $k = [\frac{6}{1-\delta}] + 1$. It follows from Lemma S18(i) that

$$\boldsymbol{\varepsilon}^T (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon} = (n-q)\sigma^2 + O_p(\sqrt{n}). \tag{S166}$$

This and (S162) imply that

$$\hat{\sigma}^2 = \sigma^2 + O_p(n^{-1/2}). \tag{S167}$$

By assumption $q \leqslant n^\delta$ for some $\delta \in (0,1)$, we see

$$\frac{1}{n}\boldsymbol{\varepsilon}^T (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon} = \sigma^2 + O_p\big(n^{-\delta'}\big), \tag{S168}$$

with $\delta' = \min\{1-\delta, 1/2\}$.

In lieu of the explanation before (S151), $\{\mathbf{u}_j;\, q+1 \leqslant j \leqslant p\}$ are independent of $\boldsymbol{X}_a$ and $\boldsymbol{\varepsilon}$, and hence are independent of the unit random vector $\mathbf{e}_2$. We then have from Lemma S16 that $(\mathbf{u}_{q+1}^T\mathbf{e}_2, \cdots, \mathbf{u}_p^T\mathbf{e}_2)^T$ has distribution $N(\mathbf{0}, \boldsymbol{\Sigma}_{bb\cdot a})$ and is also independent of $\|(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}\|$. By assumption, (2.3) holds with "$\boldsymbol{\Sigma}$" replaced by "$\boldsymbol{\Sigma}_{b|a}$". From Theorem 1 and (S118) we have

$$\frac{\sum_{j=q+1}^{p}(\mathbf{u}_j^T\mathbf{e}_2)^2 - (p-q)}{\sqrt{2\,\mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^2)}} \to N(0,1) \tag{S169}$$

in distribution as $p \to \infty$. From (S162), we know $\frac{1}{n}\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon} = \frac{n-q}{n}\hat{\sigma}^2$. By (S165) and (S169) we have

$$\frac{\frac{n\hat{\sigma}^{-2}}{n-q}\cdot\left(n^{-1}\hat{\boldsymbol{\varepsilon}}^T\boldsymbol{X}_b\boldsymbol{X}_b^T\hat{\boldsymbol{\varepsilon}}\right) - (p-q)}{\sqrt{2\,\mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^2)}} \to N(0,1).$$

Use the fact $n/(n-q) \to 1$ to see

$$\frac{n^{-1}\hat{\boldsymbol{\varepsilon}}^T\boldsymbol{X}_b\boldsymbol{X}_b^T\hat{\boldsymbol{\varepsilon}} - n^{-1}(n-q)(p-q)\hat{\sigma}^2}{\sqrt{2\hat{\sigma}^4\,\mathrm{tr}(\boldsymbol{\Sigma}_{b|a}^2)}} \to N(0,1). \tag{S170}$$

By Proposition S1, $\widehat{\mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}^2\big)}/\mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}^2\big) \to 1$ in probability. We then arrive at

$$T_{sum}^{(3)} = \frac{n^{-1}\hat{\boldsymbol{\varepsilon}}^T\boldsymbol{X}_b\boldsymbol{X}_b^T\hat{\boldsymbol{\varepsilon}} - n^{-1}(n-q)(p-q)\hat{\sigma}^2}{\sqrt{2\hat{\sigma}^4\,\widehat{\mathrm{tr}\big(\boldsymbol{\Sigma}_{b|a}^2\big)}}} \to N(0,1)$$

through an application of the Slutsky lemma. The proof is completed. □

Proof of Theorem S2. Parts (i) and (ii) follow from Propositions S3 and S2, respectively. The setting here is the same as those in the two propositions, so we will continue to use the same notation in the two propositions to prove (iii) for the asymptotic independence. Recall the definition

$$T_{sum}^{(3)} = \frac{n^{-1}\hat{\boldsymbol{\varepsilon}}^T \boldsymbol{X}_b \boldsymbol{X}_b^T \hat{\boldsymbol{\varepsilon}} - n^{-1}(n-q)(p-q)\hat{\sigma}^2}{\sqrt{2\hat{\sigma}^4 \operatorname{tr}\widehat{(\boldsymbol{\Sigma}_{b|a}^2)}}}.$$

By (S162) and (S165), we have

$$\frac{1}{n}\hat{\boldsymbol{\varepsilon}}^T \boldsymbol{X}_b \boldsymbol{X}_b^T \hat{\boldsymbol{\varepsilon}} = \frac{n-q}{n}\hat{\sigma}^2 \sum_{j=q+1}^{p} (\mathbf{u}_j^T \mathbf{e}_2)^2.$$

Thus,

$$\begin{aligned} T_{sum}^{(3)} =& \frac{n-q}{n} \cdot \frac{\sum_{j=q+1}^{p} (\mathbf{u}_j^T \mathbf{e}_2)^2 - (p-q)}{\sqrt{2\operatorname{tr}\widehat{(\boldsymbol{\Sigma}_{b|a}^2)}}} \\ =& \tilde{T}_{sum}^{(3)} + (\omega_p - 1)\cdot \tilde{T}_{sum}^{(3)} - \frac{q}{n}\cdot \omega_p \tilde{T}_{sum}^{(3)} \end{aligned}$$

where

$$\tilde{T}_{sum}^{(3)} =: \frac{\sum_{j=q+1}^{p} (\mathbf{u}_j^T \mathbf{e}_2)^2 - (p-q)}{\sqrt{2\operatorname{tr}(\boldsymbol{\Sigma}_{b|a}^2)}} \quad \text{and} \quad \omega_p := \left(\frac{\operatorname{tr}(\boldsymbol{\Sigma}_{b|a}^2)}{\operatorname{tr}\widehat{(\boldsymbol{\Sigma}_{b|a}^2)}}\right)^{1/2}.$$

By Proposition S1, $\operatorname{tr}\widehat{(\boldsymbol{\Sigma}_{b|a}^2)}/\operatorname{tr}(\boldsymbol{\Sigma}_{b|a}^2) \to 1$ in probability. Hence $\omega_p \to 1$ in probability. Also, $\tilde{T}_{sum}^{(3)} \to N(0,1)$ by (S169). This together with the assumption $q = O(n^\delta)$ for some $\delta \in (0,1)$ implies that

$$T_{sum}^{(3)} = \frac{\sum_{j=q+1}^{p} (\mathbf{u}_j^T \mathbf{e}_2)^2 - (p-q)}{\sqrt{2\operatorname{tr}(\boldsymbol{\Sigma}_{b|a}^2)}} + o_p(1). \tag{S171}$$

On the other hand, based on (S148) and (S150),

$$\begin{aligned} & T_{max}^{(3)} - 2\log(p-q) + \log\log(p-q) \\ =& H'_{31} + o_p(1) \end{aligned}$$

$$= \max_{q+1\leqslant j\leqslant p} \left\{ n^{-1}\sigma^{-2}\boldsymbol{\varepsilon}^T\tilde{\mathbf{X}}_j\tilde{\mathbf{X}}_j^T\boldsymbol{\varepsilon} \right\} - 2\log(p-q) + \log\log(p-q) + o_p(1)$$

$$= \max_{q+1\leqslant j\leqslant p} \left\{ \frac{\|(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}\|^2}{n\sigma^2} \cdot \left(\mathbf{u}_j^T \boldsymbol{e}_2\right)^2 \right\} - 2\log(p-q) + \log\log(p-q) + o_p(1) \tag{S172}$$

by(S164). Remember that $\mathbf{e}_2 = (\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}/\|(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}\|$ is independent of $\mathbf{u}_j$ because $\mathbf{u}_j, \mathbf{H}_a, \boldsymbol{\varepsilon}$ are independent. It follows that

$$\begin{aligned} &\left| \max_{q+1\leqslant j\leqslant p} \left\{ \frac{\|(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}\|^2}{n\sigma^2} \cdot \left(\mathbf{u}_j^T \boldsymbol{e}_2\right)^2 \right\} - \max_{q+1\leqslant j\leqslant p} \left(\mathbf{u}_j^T \boldsymbol{e}_2\right)^2 \right| \\ =&\left| \frac{\|(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}\|^2}{n\sigma^2} - 1 \right| \cdot \max_{q+1\leqslant j\leqslant p} \left(\mathbf{u}_j^T \boldsymbol{e}_2\right)^2. \end{aligned} \tag{S173}$$

According to (S168), we know that

$$\frac{\|(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}\|^2}{n\sigma^2} = \frac{1}{n\sigma^2}\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon} = 1 + O_p\left(n^{-\delta'}\right)$$

with $\delta' = \min\{1-\delta, 1/2\}$. As a consequence,

$$\left| \frac{\|(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}\|^2}{n\sigma^2} - 1 \right| = O_p\left(\frac{1}{n^{\delta'}}\right). \tag{S174}$$

From the discussion between (S168) and (S169), it holds that

$$(\mathbf{u}_{q+1}^T\mathbf{e}_2, \cdots, \mathbf{u}_p^T\mathbf{e}_2)^T \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{bb\cdot a}). \tag{S175}$$

By assumption, (2.3) holds with "$\boldsymbol{\Sigma}$" replaced by "$\boldsymbol{\Sigma}_{b|a}$", we know Assumption (2.2) also holds by the discussion below (2.3). Due to Theorem 2, we get

$$\max_{q+1\leqslant j\leqslant p}(\mathbf{u}_j^T\mathbf{e}_2)^2 - 2\log(p-q) + \log\log(p-q) \text{ converges weakly to a distribution}$$

with cdf $\exp\Big\{-\dfrac{1}{\sqrt{\pi}}e^{-x/2}\Big\}$

for any $x \in \mathbb{R}$ as $p \to \infty$. By assumption, $p = o(n^3)$, the above particularly implies $\max_{q+1 \leqslant j \leqslant p}(\mathbf{u}_j^T \mathbf{e}_2)^2 = O(\log n)$. This together with (S173) and (S174) concludes that

$$\max_{q+1\leqslant j\leqslant p}\Big\{\frac{\|(\mathbf{I}_n - \mathbf{H}_a)\boldsymbol{\varepsilon}\|^2}{n\sigma^2}\cdot\big(\mathbf{u}_j^T\boldsymbol{e}_2\big)^2\Big\} = \max_{q+1\leqslant j\leqslant p}\big(\mathbf{u}_j^T\boldsymbol{e}_2\big)^2 + o_p(1).$$

According to (S172), we have

$$\begin{aligned}&T_{max}^{(3)} - 2\log(p-q) + \log\log(p-q)\\ =&\max_{q+1\leqslant j\leqslant p}\big(\mathbf{u}_j^T\boldsymbol{e}_2\big)^2 - 2\log(p-q) + \log\log(p-q) + o_p(1).\end{aligned}$$

Joining this with (S171) and (S175), by Theorem 3 and Lemma S10, we obtain $T_{sum}^{(3)}$ and $T_{max}^{(3)} - 2\log(p-q) + \log\log(p-q)$ are asymptotically independent. □

Bibliography

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

Bai, Z. and J. W. Silverstein (2010). *Spectral analysis of large dimensional random matrices*, Volume 20. Springer.

Cai, T. T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 349–372.

Chen, S. X., J. Li, and P.-S. Zhong (2019). Two-sample and anova tests for high dimensional means. *The Annals of Statistics 47*(3), 1443–1474.

Chen, S. X. and Y.-L. Qin (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics 38*(2), 808–835.

Chow, Y. S. and H. Teicher (1997). Probability theory: Independence, interchangeability, martingales.

Dembo, A. and O. Zeitouni (1998). *Large Deviations Techniques and Applications*. Springer, 2nd ed.

Eaton, M. L. (1983). Multivariate statistics: a vector space approach. *JOHN WILEY & SONS, INC., 605 THIRD AVE., NEW YORK, NY 10158, USA, 1983, 512*.

Fan, J., Y. Fan, and J. Lv (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics 147*(1), 186–197.

Feng, L., C. Zou, Z. Wang, and L. Zhu (2015). Two-sample behrens-fisher problem for high-dimensional data. *Statistica Sinica*, 1297–1312.

Goeman, J. J., S. A. Van De Geer, and H. C. Van Houwelingen (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(3), 477–493.

Goeman, J. J., H. C. Van Houwelingen, and L. Finos (2011). Testing against a high-dimensional alternative in the generalized linear model: asymptotic type i error control. *Biometrika*, 381–390.

Gregory, K. B., R. J. Carroll, V. Baladandayuthapani, and S. N. Lahiri (2015). A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association 110*(510), 837–849.

Horn, R. A. and C. R. Johnson (2012). *Matrix analysis.* Cambridge university press.

Jiang, T. and P. Li (2021). Mean test with fewer observation than dimension and ratio unbiased estimator for correlation matrix. *arXiv preprint arXiv:2108.06892*.

Lan, W., H. Wang, and C.-L. Tsai (2014). Testing covariates in high-dimensional regression. *Annals of the Institute of Statistical Mathematics 66*(2), 279–301.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory.* John Wiley & Sons.

Rothman, A. J., P. J. Bickel, E. Levina, and J. Zhu (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics 2*, 494–515.

Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis 100*(3), 518–532.

Srivastava, M. S. and M. Du (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis 99*(3), 386–402.

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

Zhang, C.-H. and J. Huang (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. *The Annals of Statistics 36*(4), 1567–1594.

Zhong, P.-S. and S. X. Chen (2011). Tests for high-dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association 106*(493), 260–274.

Zhong, P.-S., S. X. Chen, and M. Xu (2013). Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *The Annals of Statistics 41*(6), 2820–2851.