Supplementary material for "Robust estimation for number of factors in high dimensional factor modeling via Spearman correlation matrix"

Contents

This supplementary document contains some auxiliary lemmas and the technical proofs of Lemmas 2.2, 2.3, 6.1, 6.2, S1.6 and S1.7.

S1 Auxiliary lemmas

This section introduces several auxiliary lemmas used in the technical proofs of our theoretical results. Lemmas S1.1 – S1.5 are from existing literature, while Lemmas S1.6 and S1.7 are our original contributions. We provide the proofs of these two new lemmas in Sections S2.5 and S2.6, respectively.

Lemma S1.1 (Weyl's inequality, Corollary 6.3.4 in Horn and Johnson (2012)). *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two $n \times n$ normal matrices, and let $\lambda_1(\boldsymbol{A}) \geqslant \cdots \geqslant \lambda_n(\boldsymbol{A})$ and $\lambda_1(\boldsymbol{B}) \geqslant \cdots \geqslant \lambda_n(\boldsymbol{B})$ be the nonincreasingly ordered eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Then*

$$\max_{1 \leqslant i \leqslant n} |\lambda_i(\boldsymbol{A}) - \lambda_i(\boldsymbol{B})| \leqslant \|\boldsymbol{A} - \boldsymbol{B}\|_2,$$

where $\|\boldsymbol{A} - \boldsymbol{B}\|_2$ denotes the spectral norm of $\boldsymbol{A} - \boldsymbol{B}$.

Lemma S1.2 (Lemma 7 in El Karoui (2009)). *Suppose that the random vector $\boldsymbol{r} \in \mathbb{R}^p$ has the property that for any convex 1-Lipschitz (with respect to the Euclidean norm) function F from $\mathbb{R}^p$ to $\mathbb{R}$, we have*

$$\mathbb{P}\big(|F(\boldsymbol{r}) - m_F| > t\big) \leqslant C \exp\big\{-c(p)t^2\big\},$$

where m_F denotes a median of F, and C and $c(p)$ are independent of F, and C is independent of p. We allow $c(p)$ to be a constant or to go to zero with p like $p^{-\alpha}$, $0 \leqslant \alpha < 1$. Suppose, further, that $\mathbb{E}(\boldsymbol{r}) = \boldsymbol{0}$, $\mathbb{E}(\boldsymbol{r}\boldsymbol{r}^\intercal) = \boldsymbol{\Sigma}$, with $\|\boldsymbol{\Sigma}\|_2 \leqslant \log(p)$. If $\boldsymbol{M}$ is a complex deterministic matrix such that $\|\boldsymbol{M}\|_2 \leqslant \xi$, where ξ is independent of p, then $p^{-1}\boldsymbol{r}^\intercal \boldsymbol{M}\boldsymbol{r}$ is strongly concentrated around its mean, $p^{-1}\mathrm{tr}(\boldsymbol{M}\boldsymbol{\Sigma})$. In particular, if, for $\varepsilon > 0$, $t_p(\varepsilon) = (\log p)^{1+\varepsilon}/\sqrt{pc(p)}$, then

$$\log\left\{\mathbb{P}\left(\left|\frac{1}{p}\boldsymbol{r}^\intercal \boldsymbol{M}\boldsymbol{r} - \frac{1}{p}\mathrm{tr}(\boldsymbol{M}\boldsymbol{\Sigma})\right| > t_p(\varepsilon)\right)\right\} \asymp -(\log p)^{1+2\varepsilon}.$$

Lemma S1.3 (Corollary 4.10 in Ledoux (2001)). *For every product probability $\mathbb{P}$ on $[0,1]^n$,* *every convex* 1*-Lipschitz function* F *on* $\mathbb{R}^n$*, and every* $r \geqslant 0$*, we have*

$$\mathbb{P}\big(|F - m_F| \geqslant r\big) \leqslant 4e^{-r^2/4},$$

where m_F *is a median of* F *for* $\mathbb{P}$*.*

Lemma S1.4 ((3.3.41) in Horn and Johnson (1991))**.** *For any* $n \times n$ *Hermitian* $\boldsymbol{A} = (A_{ij})$ *with eigenvalues* $\lambda_1, \ldots, \lambda_n$*, and convex* f*, we have*

$$\sum_{i=1}^{n} f(A_{ii}) \leqslant \sum_{i=1}^{n} f(\lambda_i).$$

Lemma S1.5 (Grothendieck's identity, Lemma 3.6.6 in Vershynin (2018), Lemma 4.1 in Li et al. (2022))**.** *Consider a bivariate normal vector:*

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \gamma \\ \gamma & \sigma_2^2 \end{pmatrix}\right),$$

we have

$$\mathbb{E}[\mathrm{sign}(Z_1)\mathrm{sign}(Z_2)] = \frac{2}{\pi}\arcsin\Big(\frac{\gamma}{\sigma_1\sigma_2}\Big).$$

Lemma S1.6 is utilized to analyze the structure of $\boldsymbol{\Sigma}_\rho$, as stated in Lemma 2.3.

Lemma S1.6. *Consider a bivariate random vector:*

$$\boldsymbol{Q} := \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \overset{\mathsf{d}}{=} \begin{pmatrix} \sqrt{W_X}X_1 + \sqrt{W_{Y_1}}Y_1 \\ \sqrt{W_X}X_2 + \sqrt{W_{Y_2}}Y_2 \end{pmatrix}, \tag{S1.1}$$

where W_X*,* W_{Y_1}*, and* W_{Y_2} *are three independent scalar-valued random variables with positive* *support, and* $(X_1, X_2, Y_1, Y_2)^\intercal \sim \mathcal{N}_4(\mathbf{0}, \boldsymbol{\Omega}_0)$ *with*

$$\boldsymbol{\Omega}_0 = \begin{pmatrix} \boldsymbol{\Sigma}_X & O \\ O & \boldsymbol{\Sigma}_Y \end{pmatrix}, \qquad \boldsymbol{\Sigma}_X = \begin{pmatrix} \sigma_{X_1}^2 & \gamma \\ \gamma & \sigma_{X_2}^2 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_Y = \begin{pmatrix} \sigma_{Y_1}^2 & 0 \\ 0 & \sigma_{Y_2}^2 \end{pmatrix}.$$

Suppose that $\widetilde{Q}_1$ *and* $\widetilde{Q}_2$ *(they are independent) are independent copies of* Q_1 *and* Q_2*, respectively, then, we have*

$$\mathbb{E}[\mathrm{sign}(Q_1 - \widetilde{Q}_1)\,\mathrm{sign}(Q_2 - \widetilde{Q}_2)] = \frac{2}{\pi}\mathbb{E}\{\arcsin(r)\},$$

where

$$r = \frac{\gamma W_X}{\sqrt{(W_X + \widetilde{W}_{X_1})\sigma_{X_1}^2 + (W_{Y_1} + \widetilde{W}_{Y_1})\sigma_{Y_1}^2}\sqrt{(W_X + \widetilde{W}_{X_2})\sigma_{X_2}^2 + (W_{Y_2} + \widetilde{W}_{Y_2})\sigma_{Y_2}^2}},$$

and $(\widetilde{W}_{X_1}, \widetilde{W}_{Y_1})$ *and* $(\widetilde{W}_{X_2}, \widetilde{W}_{Y_2})$ *are independent copies of* (W_X, W_{Y_1}) *and* (W_X, W_{Y_2})*, respectively.*

Lemma S1.7 is used in the proof of Lemma 6.2.

Lemma S1.7. *Let* $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\intercal$ *and* $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)^\intercal$ *be two* $n \times p$ *independent random matrices satisfying the same assumptions as those in Lemma 6.2. For any* $i \in [n]$*, we denote* $\boldsymbol{X}_i = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_n)^\intercal$*,* $\boldsymbol{X}_{i0} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_n)^\intercal$*, and*

$$\beta_i = 1 - \frac{1}{n}\boldsymbol{x}_i^\intercal \boldsymbol{\Gamma}^{1/2}(\lambda \boldsymbol{I} - n^{-1}\boldsymbol{\Gamma}^{1/2}\boldsymbol{X}_{i0}^\intercal \boldsymbol{X}_{i0}\boldsymbol{\Gamma}^{1/2})^{-1}\boldsymbol{\Gamma}^{1/2}\boldsymbol{x}_i, \tag{S1.2}$$

$$\beta_{i0} = 1 - \frac{1}{n}\mathrm{tr}\big\{\boldsymbol{\Gamma}(\lambda \boldsymbol{I} - n^{-1}\boldsymbol{\Gamma}^{1/2}\boldsymbol{X}_{i0}^\intercal \boldsymbol{X}_{i0}\boldsymbol{\Gamma}^{1/2})^{-1}\big\}, \tag{S1.3}$$

with $\boldsymbol{\Gamma} = \boldsymbol{U}_2\boldsymbol{D}_2\boldsymbol{U}_2^\intercal$ *defined before the equation (18) in the main text. Then, for any* $K \times K$

symmetric matrix $\boldsymbol{W}$, it holds that, as $n \to \infty$,

$$\beta_{i0} \to -\frac{1}{\lambda \underline{m}(\lambda)}, \qquad \varepsilon_i := \beta_i - \beta_{i0} \to 0, \tag{S1.4}$$

$$\mathbb{E}\,\varepsilon_i^2 = o(1), \qquad \mathbb{E}\,\varepsilon_i^4 = o(n^{-1}), \tag{S1.5}$$

$$\mathbb{E}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_i)\}^2 = O(1), \qquad \mathbb{E}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{i0})\}^2 = O(1), \tag{S1.6}$$

where $\underline{m}(\cdot)$ denotes the Stieltjes transform of the LSD of the matrix $n^{-1}\boldsymbol{X}_{i0}\boldsymbol{\Gamma}\boldsymbol{X}_{i0}^{\intercal}$, and $\boldsymbol{\tau}_{i0}$ and $\boldsymbol{\tau}_i$ are defined in (S2.22) *and* (S2.23)*, respectively.*

Remark S1.8. *The conclusions* (S1.4) – (S1.6) *presented in Lemma S1.7 are identical to those in Lemma C.3 and Lemma D.1 of Jiang and Bai (2021). However, the proof provided by Jiang and Bai (2021) is not applicable for our scenario as we do not assume that $\boldsymbol{X}$ (or $\boldsymbol{Y}$) possesses i.i.d. entries, as stated in Assumption (B1). Specifically, we need to establish the concentration properties of certain quadratic forms without the i.i.d. assumption. We address this challenge in Section S2.6 by leveraging concentration inequalities from Ledoux (2001) and El Karoui (2009). See the proof in Section S2.6 for more details.*

S2 Proofs of lemmas

S2.1 Proof of Lemma 2.2

Let $\boldsymbol{\Delta}_n$ denote the difference between $\boldsymbol{\rho}_n$ and $\boldsymbol{W}_n$, that is,

$$\boldsymbol{\Delta}_n := \boldsymbol{\rho}_n - \boldsymbol{W}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{R}_i\boldsymbol{R}_i^{\intercal} - \frac{3}{n}\sum_{i=1}^{n} \boldsymbol{A}_i\boldsymbol{A}_i^{\intercal}, \tag{S2.1}$$

where $\boldsymbol{R}_i^{\intercal}$ is the i-th row of $\boldsymbol{R}$ defined in (2) and $\boldsymbol{A}_i := \mathbb{E}\{\mathrm{sign}(\boldsymbol{y}_i - \boldsymbol{y}_\ell) \mid \boldsymbol{y}_i\}$. From Lemma S1.1, it is sufficient to show that $\|\sqrt{n}\boldsymbol{\Delta}_n\|_2 = O(1)$. To this end, we show that $\sqrt{n}\boldsymbol{\Delta}_n$ is a

Wigner-type matrix by the following moment estimations:

$$\mathbb{E}\{(\boldsymbol{\Delta}_n)_{j\ell}\} = O(n^{-1}), \quad \mathrm{Var}\{(\boldsymbol{\Delta}_n)_{j\ell}\} = O(n^{-1}), \tag{S2.2}$$

$$\mathrm{Corr}\big((\boldsymbol{\Delta}_n)_{j\ell}, (\boldsymbol{\Delta}_n)_{j\ell'}\big) = o(1), \qquad j \neq \ell \neq \ell', \tag{S2.3}$$

$$\mathrm{Corr}\big((\boldsymbol{\Delta}_n)_{jj}, (\boldsymbol{\Delta}_n)_{\ell r}\big) = o(1), \qquad \ell \neq r. \tag{S2.4}$$

Let y_{ij} be the j-th component of $\boldsymbol{y}_i$. For each $j \in [p]$, we define the *empirical cumulative distribution function* (ECDF) as

$$\widehat{F}_j(y) = \frac{1}{n} \sum_{\ell=1}^{n} \mathbb{1}\{y_{\ell j} \leqslant y\}, \tag{S2.5}$$

and let $r_{ij} = n\widehat{F}_j(y_{ij})$ be the rank of y_{ij} among $\{y_{\ell j}\}_{\ell=1}^n$, where and $i \in [n]$ and $j \in [p]$. For notational simplicity, we denote $F_{ij} = F_j(y_{ij})$ and $\widehat{F}_{ij} = \widehat{F}_j(y_{ij})$, where $F_j(\cdot)$ is the CDF of y_{1j}. Using these notations, we write the (i,j)-th entry of the matrix $\boldsymbol{R}$ defined in (2) as

$$R_{ij} = \sqrt{\frac{12}{n^2-1}}\Big(r_{ij} - \frac{n+1}{2}\Big) = \sqrt{\frac{12n^2}{n^2-1}}\Big(\widehat{F}_{ij} - \frac{n+1}{2n}\Big), \tag{S2.6}$$

and write the j-th entry of the vector $\boldsymbol{A}_i$ as

$$A_{ij} = \mathbb{E}\big\{\mathrm{sign}\big(y_{ij} - y_{\ell j}\big) \mid y_{ij}\big\} = 2F_{ij} - 1. \tag{S2.7}$$

First, we prove (S2.2). From (S2.1) and (S2.6) – (S2.7), the j-th diagonal entries of $\boldsymbol{\Delta}_n$ can be written as $(\boldsymbol{\Delta}_n)_{jj} = 1 - (12/n)\sum_{i=1}^n (F_{ij} - 1/2)^2$. By using the fact that $\{F_{ij}\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathsf{Uniform}(0,1)$ for fixed j, we obtain

$$\mathbb{E}\{(\boldsymbol{\Delta}_n)_{jj}\} = 0, \qquad \mathrm{Var}\{(\boldsymbol{\Delta}_n)_{jj}\} = \frac{4}{5n}. \tag{S2.8}$$

Then, we consider the off-diagonal entries. From (S2.1) and (S2.6) – (S2.7), the (j,ℓ)-th entry of $\boldsymbol{\Delta}_n$, where $j \neq \ell$, can be written as

$$
\begin{aligned}
&(\boldsymbol{\Delta}_n)_{j\ell} \\
&= \frac{1}{n}\left\{\sum_{i=1}^{n}\left(\frac{12n^2}{n^2-1}\right)\left(\widehat{F}_{ij}-\frac{n+1}{2n}\right)\left(\widehat{F}_{i\ell}-\frac{n+1}{2n}\right)-\sum_{i=1}^{n}12\left(F_{ij}-\frac{1}{2}\right)\left(F_{i\ell}-\frac{1}{2}\right)\right\} \\
&= \frac{6}{n}\sum_{i=1}^{n}\left\{\left(\frac{2n^2}{n^2-1}\widehat{F}_{ij}\widehat{F}_{i\ell}-2F_{ij}F_{i\ell}\right)-\left(\frac{n}{n-1}\widehat{F}_{ij}-F_{ij}\right)-\left(\frac{n}{n-1}\widehat{F}_{i\ell}-F_{i\ell}\right)+\frac{1}{n-1}\right\} \\
&=: \frac{6}{n}\sum_{i=1}^{n}(\boldsymbol{\Delta}_n)_{j\ell}^{(i)}.
\end{aligned}
$$

From the fact that $\{F_{ij}\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathsf{Uniform}(0,1)$ for fixed j and the definition (S2.5) of ECDF, we obtain

$$
\mathbb{E}F_{ij}=\frac{1}{2}, \qquad \mathbb{E}F_{ij}^2=\frac{1}{3}, \qquad \mathbb{E}\widehat{F}_{ij}=\frac{1}{n}\sum_{\ell=1}^{n}\mathbb{E}\mathbb{1}\{y_{\ell j}\leqslant y_{ij}\}=\frac{n+1}{2n},
$$

and

$$
\mathbb{E}\widehat{F}_{ij}\widehat{F}_{i\ell}=\frac{(n-1)(n-2)}{n^2}\mathbb{E}F_{ij}F_{i\ell}+\frac{n-1}{n^2}\mathbb{E}\mathbb{1}\{y_{sj}\leqslant y_{ij}\}\mathbb{1}\{y_{s\ell}\leqslant y_{i\ell}\}+\frac{1}{n}, \quad (s\neq i).
$$

These identities imply that

$$
\mathbb{E}\big\{(\boldsymbol{\Delta}_n)_{j\ell}^{(1)}\big\}=-\frac{6}{n+1}\mathbb{E}F_{1j}F_{1\ell}+\frac{2}{n+1}\mathbb{E}\mathbb{1}\{y_{2j}\leqslant y_{1j}\}\mathbb{1}\{y_{2\ell}\leqslant y_{1\ell}\}+\frac{1}{n+1}=O(n^{-1}).
$$

As $\{(\boldsymbol{\Delta}_n)_{j\ell}^{(i)}\}_{i=1}^n$ are i.i.d. for fixed j and ℓ, we have

$$
\mathbb{E}\big\{(\boldsymbol{\Delta}_n)_{j\ell}\big\}=6\mathbb{E}\big\{(\boldsymbol{\Delta}_n)_{j\ell}^{(1)}\big\}=O(n^{-1}). \tag{S2.9}
$$

Moreover, a similar calculation gives

$$
\begin{aligned}
&\mathrm{Var}\big\{(\mathbf{\Delta}_n)_{j\ell}^{(1)}\big\}\\
=&\frac{n^3}{(n^2-1)^2}\mathbb{E}\bigg\{\frac{1}{3}+4F_{1j}F_{1\ell}\big(3F_{1j}+3F_{1\ell}-4F_{1j}F_{1\ell}\big)-10F_{1j}F_{1\ell}\\
&\qquad\qquad-4\Big(F_{1j}+F_{1\ell}-2F_{1j}F_{1\ell}-\frac{1}{2}\Big)\mathbb{1}\{y_{2j}\leqslant y_{1j}\}\mathbb{1}\{y_{2\ell}\leqslant y_{1\ell}\}\bigg\}+O(n^{-2})\\
=&\,O(n^{-1}).
\end{aligned}
$$

Thus, we have

$$
\mathrm{Var}\big\{(\mathbf{\Delta}_n)_{j\ell}\big\}=\frac{36}{n}\mathrm{Var}\big\{(\mathbf{\Delta}_n)_{j\ell}^{(1)}\big\}=O(n^{-2}). \tag{S2.10}
$$

This, together with (S2.8) and (S2.9), implies (S2.2).

Now, we prove (S2.3). For any $j,\ell,\ell'\in[p]$ with $j\neq\ell\neq\ell'$, direct calculation gives us

$$
\begin{aligned}
&\mathrm{Cov}\big((\mathbf{\Delta}_n)_{j\ell},(\mathbf{\Delta}_n)_{j\ell'}\big)\\
=&\,\frac{36n^2}{(n^2-1)^2}\mathbb{E}\Big\{-16F_{1j}^2F_{1\ell}F_{1\ell'}+8F_{1j}^2F_{1\ell}+4F_{1j}^2\mathbb{1}\{y_{2\ell}\leqslant y_{1\ell}\}\mathbb{1}\{y_{2\ell'}\leqslant y_{1\ell'}\}\\
&+12F_{1j}F_{1\ell}F_{1\ell'}+4F_{1j}\mathbb{1}\{y_{2j}\leqslant y_{1j}\}\big[F_{1\ell}\mathbb{1}\{y_{2\ell'}\leqslant y_{1\ell'}\}+F_{1\ell'}\mathbb{1}\{y_{2\ell}\leqslant y_{1\ell}\}\big]\\
&-3F_{1j}\big(F_{1\ell}+F_{1\ell'}\big)-2F_{1j}\mathbb{1}\{y_{2j}\leqslant y_{1j}\}\big[\mathbb{1}\{y_{2\ell}\leqslant y_{1\ell}\}+\mathbb{1}\{y_{2\ell'}\leqslant y_{1\ell'}\}\big]\\
&-4F_{1j}\mathbb{1}\{y_{2\ell}\leqslant y_{1\ell}\}\mathbb{1}\{y_{2\ell'}\leqslant y_{1\ell'}\}-F_{1\ell}F_{1\ell'}\\
&-2\mathbb{1}\{y_{2j}\leqslant y_{1j}\}\big[F_{1\ell}\mathbb{1}\{y_{2\ell'}\leqslant y_{1\ell'}\}+F_{1\ell'}\mathbb{1}\{y_{2\ell}\leqslant y_{1\ell}\}\big]\\
&+\mathbb{1}\{y_{2j}\leqslant y_{1j}\}\mathbb{1}\{y_{2\ell}\leqslant y_{1\ell}\}+\mathbb{1}\{y_{2j}\leqslant y_{1j}\}\mathbb{1}\{y_{2\ell'}\leqslant y_{1\ell'}\}\\
&+\mathbb{1}\{y_{2\ell}\leqslant y_{1\ell}\}\mathbb{1}\{y_{2\ell'}\leqslant y_{1\ell'}\}+\frac{1}{6}\Big\}+O(n^{-3}).
\end{aligned} \tag{S2.11}
$$

From Assumptions (A3) and (A4), the constant term in the curly brackets of (S2.11) is of order $o(1)$, then we have $\mathrm{Cov}\big((\mathbf{\Delta}_n)_{j\ell},(\mathbf{\Delta}_n)_{j\ell'}\big)=o(n^{-2})$. This, together with (S2.10), implies that

$$\mathrm{Corr}\big((\boldsymbol{\Delta}_n)_{j\ell}, (\boldsymbol{\Delta}_n)_{j\ell'}\big) = \frac{\mathrm{Cov}\big((\boldsymbol{\Delta}_n)_{j\ell}, (\boldsymbol{\Delta}_n)_{j\ell'}\big)}{\sqrt{\mathrm{Var}\big\{(\boldsymbol{\Delta}_n)_{j\ell}\big\} \cdot \mathrm{Var}\big\{(\boldsymbol{\Delta}_n)_{j\ell'}\big\}}} = o(1),$$

which is (S2.3).

Finally, we prove (S2.4). For any $j, \ell, r \in [p]$ with $\ell \neq r$, by using similar argument above, we have

$$\begin{aligned}
&\mathrm{Cov}\big((\boldsymbol{\Delta}_n)_{jj}, (\boldsymbol{\Delta}_n)_{\ell r}\big) \\
=& \frac{72}{n}\frac{1}{n+1}\Bigg\{\Big(2\mathbb{E}\big[F_{1j}\mathbb{1}\{y_{2\ell} \leqslant y_{1\ell}\}\mathbb{1}\{y_{2r} \leqslant y_{1r}\}\big] - 6\mathbb{E}F_{1j}F_{1\ell}F_{1r} + 2\mathbb{E}(F_{1j}F_{1\ell} + F_{1j}F_{1r}) - \frac{1}{2}\Big) \\
&\quad - \Big(2\mathbb{E}\big[F_{1j}^2\mathbb{1}\{y_{2\ell} \leqslant y_{1\ell}\}\mathbb{1}\{y_{2r} \leqslant y_{1r}\}\big] - 6\mathbb{E}F_{1j}^2F_{1\ell}F_{1r} + 2\mathbb{E}(F_{1j}^2F_{1\ell} + F_{1j}^2F_{1r}) - \frac{1}{3}\Big) \\
&\quad - \frac{1}{6}\Big[-6\mathbb{E}F_{1\ell}F_{1r} + 2\mathbb{E}\mathbb{1}\{y_{2\ell} \leqslant y_{1\ell}\}\mathbb{1}\{y_{2r} \leqslant y_{1r}\} + 1\Big]\Bigg\} \\
&= o(n^{-2}).
\end{aligned}$$

This, together with (S2.8) and (S2.10), implies (S2.4) and completes the proof.

S2.2 Proof of Lemma 2.3

Recall that $\boldsymbol{A}_{i\ell} = \mathrm{sign}(\boldsymbol{y}_i - \boldsymbol{y}_\ell)$ and $\boldsymbol{A}_i = \mathbb{E}(\boldsymbol{A}_{i\ell} \mid \boldsymbol{y}_i)$ for any $i, \ell \in [n]$ with $i \neq \ell$. By using the law of iterated expectations repeatedly, we obtain

$$\begin{aligned}
\boldsymbol{\Sigma}_\rho &= 3\mathbb{E}(\boldsymbol{A}_1\boldsymbol{A}_1^\intercal) = 3\mathbb{E}\big\{\mathbb{E}(\boldsymbol{A}_{12} \mid \boldsymbol{y}_1)\boldsymbol{A}_1^\intercal\big\} = 3\mathbb{E}(\boldsymbol{A}_{12}\boldsymbol{A}_1^\intercal) \\
&= 3\mathbb{E}\big\{\boldsymbol{A}_{12}\mathbb{E}(\boldsymbol{A}_{13}^\intercal \mid \boldsymbol{y}_1, \boldsymbol{y}_2)\big\} = 3\mathbb{E}\big\{\mathbb{E}(\boldsymbol{A}_{12}\boldsymbol{A}_{13}^\intercal \mid \boldsymbol{y}_1, \boldsymbol{y}_2)\big\} \\
&= 3\mathbb{E}(\boldsymbol{A}_{12}\boldsymbol{A}_{13}^\intercal) = 3\mathbb{E}\big\{\mathrm{sign}(\boldsymbol{y}_1 - \boldsymbol{y}_2)\,\mathrm{sign}(\boldsymbol{y}_1 - \boldsymbol{y}_3)^\intercal\big\}.
\end{aligned}$$

Let $\boldsymbol{b}_i^\intercal$ denote the i-th row of $\boldsymbol{B}$, and Ψ_{ii} denote the i-th diagonal element of $\boldsymbol{\Psi}$. From Assumption (A2) and Definition 2.1, the i-th component of $\boldsymbol{y}_1 = \boldsymbol{B}\boldsymbol{f}_1 + \boldsymbol{\Psi}\boldsymbol{e}_1$ has a stochastic representation (below, we omit "1" in subscripts to simplify notations)

$$y_i \overset{\mathsf{d}}{=} \sqrt{w^f}\boldsymbol{b}_i^\intercal \boldsymbol{x} + \sqrt{w_i^e}\Psi_{ii} z_i, \quad i \in [p], \tag{S2.12}$$

where $\boldsymbol{x} \sim \mathcal{N}_K(\boldsymbol{0}, \boldsymbol{I}_K)$, $\{z_i\}_{i=1}^p \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, and they are independent. For any $i \in [p]$, let

$$\tilde{\boldsymbol{v}}_i := \frac{\sqrt{w^f}\boldsymbol{b}_i}{\sqrt{(w^f + \widetilde{w}^f)\boldsymbol{b}_i^\intercal \boldsymbol{b}_i + (w_i^e + \widetilde{w}_i^e)\Psi_{ii}^2}}, \qquad \hat{\boldsymbol{v}}_i := \frac{\sqrt{w^f}\boldsymbol{b}_i}{\sqrt{(w^f + \widehat{w}^f)\boldsymbol{b}_i^\intercal \boldsymbol{b}_i + (w_i^e + \widehat{w}_i^e)\Psi_{ii}^2}},$$

where $(\widetilde{w}^f, \widetilde{w}_i^e)$ and $(\widehat{w}^f, \widehat{w}_i^e)$ are two independent copies of (w^f, w_i^e). It follows from Lemma S1.6 that, the (i,j)-th element of $\boldsymbol{\Sigma}_\rho$ can be written as

$$(\boldsymbol{\Sigma}_\rho)_{ij} = \begin{cases} (6/\pi)\mathbb{E}\{\arcsin(\tilde{\boldsymbol{v}}_i^\intercal \hat{\boldsymbol{v}}_j)\}, & i \neq j, \\ 1, & i = j. \end{cases}$$

By using Taylor's expansion, we derive that $(\boldsymbol{\Sigma}_\rho)_{ij} = \frac{6}{\pi}\mathbb{E}(\tilde{\boldsymbol{v}}_i^\intercal \hat{\boldsymbol{v}}_j) + O(p^{-2})$ for $i \neq j$. By definitions of $\tilde{\boldsymbol{v}}_i$ and $\hat{\boldsymbol{v}}_j$ and Assumption (A3), we have

$$\begin{aligned}
&\mathbb{E}(\tilde{\boldsymbol{v}}_i^\intercal \hat{\boldsymbol{v}}_j) \\
&= \mathbb{E}\left\{ \frac{\sqrt{w^f}}{\sqrt{(w^f + \widetilde{w}^f)\boldsymbol{b}_i^\intercal \boldsymbol{b}_i + (w_i^e + \widetilde{w}_i^e)\Psi_{ii}^2}} \cdot \frac{\sqrt{w^f}}{\sqrt{(w^f + \widehat{w}^f)\boldsymbol{b}_j^\intercal \boldsymbol{b}_j + (w_j^e + \widehat{w}_j^e)\Psi_{jj}^2}} \right\} \boldsymbol{b}_i^\intercal \boldsymbol{b}_j \\
&= \mathbb{E}\left\{ \left(\frac{\sqrt{w^f}}{\sqrt{(w_i^e + \widetilde{w}_i^e)\Psi_{ii}^2}} + o(1) \right) \left(\frac{\sqrt{w^f}}{\sqrt{(w_j^e + \widehat{w}_j^e)\Psi_{jj}^2}} + o(1) \right) \right\} \boldsymbol{b}_i^\intercal \boldsymbol{b}_j \\
&= \mathbb{E}\left\{ \frac{w^f}{\sqrt{w_i^e + \widetilde{w}_i^e}\sqrt{w_j^e + \widehat{w}_j^e}} \right\} \frac{\boldsymbol{b}_i^\intercal \boldsymbol{b}_j}{\Psi_{ii}\Psi_{jj}} + o(p^{-1}).
\end{aligned}$$

By using the above estimations, we obtain

$$\boldsymbol{\Sigma}_\rho = \mathrm{diag}(\boldsymbol{I}_p - \gamma \boldsymbol{\Psi}^{-1}\boldsymbol{B}\boldsymbol{B}^\intercal \boldsymbol{\Psi}^{-1}) + \gamma \boldsymbol{\Psi}^{-1}\boldsymbol{B}\boldsymbol{B}^\intercal \boldsymbol{\Psi}^{-1} + \boldsymbol{E}_n, \tag{S2.13}$$

where $\boldsymbol{E}_n := \boldsymbol{\Sigma}_\rho - \mathrm{diag}(\boldsymbol{I}_p - \gamma \boldsymbol{\Psi}^{-1}\boldsymbol{B}\boldsymbol{B}^\intercal \boldsymbol{\Psi}^{-1}) - \gamma \boldsymbol{\Psi}^{-1}\boldsymbol{B}\boldsymbol{B}^\intercal \boldsymbol{\Psi}^{-1}$ with $\|\boldsymbol{E}_n\|_{\max} = o(p^{-1})$ and

$$\gamma := \frac{6}{\pi}\mathbb{E}\big[w^f / \{(w_i^e + \widetilde{w}_i^e)^{1/2}(w_j^e + \widehat{w}_j^e)^{1/2}\}\big]. \tag{S2.14}$$

By the basic norm inequality $\|\boldsymbol{A}\|_2 \leqslant p\|\boldsymbol{A}\|_{\max}$ for any $p \times p$ matrix $\boldsymbol{A}$, we have

$$\|\boldsymbol{E}_n\|_2 \leqslant p \cdot o(p^{-1}) = o(1). \tag{S2.15}$$

Since $(\widetilde{w}_i^e, \widehat{w}_j^e)$ is an independent copy of (w_i^e, w_j^e) for $i \neq j$ and random variables $\{w_i^e\}_{i=1}^p$ are i.i.d., we can replace $w_i^e, w_j^e, \widetilde{w}_i^e, \widehat{w}_j^e$ in (S2.14) by $w_1^e, w_2^e, w_3^e, w_4^e$, respectively. This, together with (S2.13) and (S2.15), completes the proof of Lemma 2.3.

S2.3 Proof of Lemma 6.1

From Lemma 2.2, $\boldsymbol{\rho}_n$ and $\boldsymbol{W}_n$ share the same LSD, thus we only need to derive the LSD of $\boldsymbol{W}_n$. Since the vectors $\{\boldsymbol{A}_i\}_{i=1}^n$ are i.i.d., by Theorem 1.1 of Bai and Zhou (2008), we can prove Lemma 6.1 by verifying that the elements of $\boldsymbol{A}_1$ are *weakly dependent* in the following sense: for any non-random $p \times p$ matrix $\boldsymbol{D}$ with bounded spectral norm,

$$\mathrm{Var}(\boldsymbol{A}_1^\intercal \boldsymbol{D}\boldsymbol{A}_1) = o(p^2). \tag{S2.16}$$

From the Corollary 1.1 in Bai and Zhou (2008), (S2.16) holds true if

$$\sum_{\Lambda}\Big\{\mathbb{E}\big(A_{1i}A_{1j}-\sigma_{ij}\big)\big(A_{1i'}A_{1j'}-\sigma_{i'j'}\big)\Big\}^2 = o(p^2), \tag{S2.17}$$

$$\max_{i\neq j}\mathbb{E}\big|A_{1i}A_{1j}-\sigma_{ij}\big|^2 = o(p), \tag{S2.18}$$

where A_{1i} is the i-th component of $\boldsymbol{A}_1$, $\sigma_{ij} := \mathbb{E}\,A_{1i}A_{1j}$, and

$$\Lambda = \{(i,j,i',j') : i,j,i',j' \in [p]\} \setminus \{(i,j,i',j') : i = i' \neq j = j' \text{ or } i = j' \neq j = i'\}.$$

Now, we prove (S2.17) and (S2.18). From (S2.7), we have $A_{1i} \sim \mathsf{Uniform}(-1,1)$ for any $i \in [p]$, and thus

$$\mathbb{E}\,A_{1i}^4 = \frac{1}{5}, \qquad \sigma_{ii} = \mathbb{E}\,A_{1i}^2 = \frac{1}{3}. \tag{S2.19}$$

Two moment conditions (S2.17) and (S2.18) follow from (S2.19) and the following estimations:

$$\sigma_{ij} = O(p^{-1}), \qquad \mathbb{E}\,A_{1i}A_{1j}A_{1i'}A_{1j'} = O(p^{-2}), \qquad \mathbb{E}\,A_{1i}^2A_{1i'}A_{1j'} = O(p^{-1}), \tag{S2.20}$$

$$\mathbb{E}\,A_{1i}^3A_{1j} = O(p^{-1}), \qquad \mathbb{E}\big(A_{1i}^2-\sigma_{ii}\big)\big(A_{1j}^2-\sigma_{jj}\big) = O(p^{-2}), \tag{S2.21}$$

where $i,j,i',j' \in [p]$ and $i \neq j \neq i' \neq j'$. Proofs of (S2.20) and (S2.21) are provided in Section S2.3.1. From (S2.19) – (S2.21), we obtain

$$\max_{i\neq j}\mathbb{E}\,\big|A_{1i}A_{1j}-\sigma_{ij}\big|^2 = O(1),$$

and

$$
\begin{aligned}
&\sum_{\Lambda}\Big\{\mathbb{E}\big(A_{1i}A_{1j}-\sigma_{ij}\big)\big(A_{1i'}A_{1j'}-\sigma_{i'j'}\big)\Big\}^2\\
&=\sum_{i\neq j\neq i'\neq j'}\big(\mathbb{E}\,A_{1i}A_{1j}A_{1i'}A_{1j'}-\sigma_{ij}\sigma_{i'j'}\big)^2+2\sum_{i\neq i'\neq j'}\big(\mathbb{E}\,A_{1i}^2A_{1i'}A_{1j'}-\sigma_{ii}\sigma_{i'j'}\big)^2\\
&\qquad+2\sum_{i\neq j\neq j'}\big(\mathbb{E}\,A_{1i}^2A_{1j}A_{1j'}-\sigma_{ij}\sigma_{ij'}\big)^2+2\sum_{i\neq j}\big(\mathbb{E}\,A_{1i}^3A_{1j}-\sigma_{ij}\sigma_{ii}\big)^2\\
&\qquad+\sum_{i\neq j}\big(\mathbb{E}\,A_{1i}^2A_{1j}^2-\sigma_{ii}\sigma_{jj}\big)^2+\sum_{i}\big(\mathbb{E}\,A_{1i}^4-\sigma_{ii}^2\big)^2\\
&=O(p^4)\cdot O(p^{-4})+O(p^3)\cdot O(p^{-2})+O(p^3)\cdot O(p^{-2})+O(p^2)\cdot O(p^{-2})\\
&\qquad\quad+O(p^2)\cdot O(p^{-4})+O(p)\cdot O(1)\\
&=O(p).
\end{aligned}
$$

This implies (S2.17) and (S2.18), completing the proof of Lemma 6.1.

S2.3.1 Proofs of (S2.20) and (S2.21)

In this section, we provide the proofs of some moment estimations (S2.20) and (S2.21), which are used in the proof of Lemma 6.1.

From equations (S2.7), (S2.12), and Assumption (A3), we have

$$
\begin{aligned}
\mathbb{E}A_{1i}&=2\mathbb{E}\,F_i\big(\boldsymbol{b}_i^\intercal\boldsymbol{f}_1+\Psi_{ii}e_{1i}\big)-1\\
&=\mathbb{E}\left\{2F_i\big(\Psi_{ii}e_{1i}\big)-1\right\}+O(p^{-1/2})=:\mathbb{E}\tilde{A}_{1i}+O(p^{-1/2}).
\end{aligned}
$$

Since $\mathbb{E}A_{1i}=0$, we obtain that $\mathbb{E}\tilde{A}_{1i}=O(p^{-1/2})$. By using the same argument, we obtain

$$
\mathbb{E}\,\tilde{A}_{1i}^2=\mathbb{E}\,A_{1i}^2+O(p^{-1}),\qquad \mathbb{E}\,\tilde{A}_{1i}^3=O(p^{-1/2}).
$$

Since $\tilde{A}_{1i}$ and $\tilde{A}_{1j}$ are independent for $i \neq j$, we have

$$\sigma_{ij} = \mathbb{E}\, A_{1i}A_{1j} = \mathbb{E}\, \tilde{A}_{1i} \cdot \mathbb{E}\, \tilde{A}_{1j} + O(p^{-1}) = O(p^{-1}).$$

By using the above estimations, for $i \neq j \neq i' \neq j'$, we have

$$\mathbb{E}\, A_{1i}A_{1j}A_{1i'}A_{1j'} = \mathbb{E}\, \tilde{A}_{1i} \cdot \mathbb{E}\, \tilde{A}_{1j} \cdot \mathbb{E}\, \tilde{A}_{1i'} \cdot \mathbb{E}\, \tilde{A}_{1j'} + O(p^{-2}) = O(p^{-2}),$$

$$\begin{aligned}\mathbb{E}\, A_{1i}^2 A_{1i'}A_{1j'} &= \mathbb{E}\Big[A_{1i}^2\big\{\tilde{A}_{1i'} + O(p^{-1/2})\big\}\big\{\tilde{A}_{1j'} + O(p^{-1/2})\big\}\Big] \\ &= \mathbb{E}A_{1i}^2 \cdot \mathbb{E}\tilde{A}_{1i'} \cdot \mathbb{E}\tilde{A}_{1j'} + O(p^{-1}) = O(p^{-1}),\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\, A_{1i}^3 A_{1j} = \mathbb{E}\Big[\big\{\tilde{A}_{1i} + O(p^{-1/2})\big\}^3\big\{\tilde{A}_{1j} + O(p^{-1/2})\big\}\Big] \\ &= \mathbb{E}\Big[\Big\{\tilde{A}_{1i}^3 + O(p^{-1/2}) \cdot \tilde{A}_{1i}^2 + O(p^{-1}) \cdot \tilde{A}_{1i} + O(p^{-3/2})\Big\}\Big\{\tilde{A}_{1j} + O(p^{-1/2})\Big\}\Big] \\ &= O(p^{-1}),\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\big(A_{1i}^2 - \sigma_{ii}\big)\big(A_{1j}^2 - \sigma_{jj}\big) \\ &= \mathbb{E}\Big[\big\{\tilde{A}_{1i}^2 - \sigma_{ii} + O(p^{-1/2}) \cdot \tilde{A}_{1i} + O(p^{-1})\big\}\big\{\tilde{A}_{1j}^2 - \sigma_{jj} + O(p^{-1/2}) \cdot \tilde{A}_{1j} + O(p^{-1})\big\}\Big] \\ &= O(p^{-2}).\end{aligned}$$

These equations complete the proofs of (S2.20) and (S2.21).

S2.4 Proof of Lemma 6.2

Throughout this proof, $\mathbf{\Omega}_K(\lambda, \boldsymbol{X})$ is simply denoted as $\mathbf{\Omega}_K(\boldsymbol{X})$ if no confusion. Recall that $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\intercal$ and $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)^\intercal$. We denote

$$\boldsymbol{X}_i = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_n)^\intercal, \qquad \boldsymbol{X}_{i0} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_n)^\intercal,$$

with convention $\boldsymbol{X} = \boldsymbol{X}_n$ and $\boldsymbol{Y} = \boldsymbol{X}_0$. Using the Sherman-Morrison formula, we obtain

$$\begin{aligned}
&\mathbf{\Omega}_K(\boldsymbol{X}_i) - \mathbf{\Omega}_K(\boldsymbol{X}_{i0}) \\
&= \frac{1}{\beta_i\sqrt{n}}\Big\{(1 + n^{-1}\boldsymbol{x}_i^\intercal \boldsymbol{H}_2 \boldsymbol{x}_i)\boldsymbol{I}_K - \boldsymbol{U}_1^\intercal(\boldsymbol{I}_p + \boldsymbol{H}_1)\boldsymbol{x}_i\boldsymbol{x}_i^\intercal(\boldsymbol{I}_p + \boldsymbol{H}_1)^\intercal \boldsymbol{U}_1\Big\} \\
&= \frac{1}{\beta_{i0}\sqrt{n}}(\boldsymbol{\tau}_{i0} + \boldsymbol{\tau}_i) - \frac{\varepsilon_i}{\beta_{i0}^2\sqrt{n}}(\boldsymbol{\tau}_{i0} + \boldsymbol{\tau}_i) + \frac{\varepsilon_i^2}{\beta_{i0}^2\beta_i\sqrt{n}}(\boldsymbol{\tau}_{i0} + \boldsymbol{\tau}_i),
\end{aligned}$$

where β_i and β_{i0} are defined in (S1.2) and (S1.3), and

$$\begin{aligned}
\boldsymbol{H}_1 &= n^{-1}\boldsymbol{X}_{i0}^\intercal(\lambda \boldsymbol{I}_{n-1} - n^{-1}\boldsymbol{X}_{i0}\mathbf{\Gamma}\boldsymbol{X}_{i0}^\intercal)^{-1}\boldsymbol{X}_{i0}\mathbf{\Gamma}, \\
\boldsymbol{H}_2 &= n^{-1}\mathbf{\Gamma}\boldsymbol{X}_{i0}^\intercal(\lambda \boldsymbol{I}_{n-1} - n^{-1}\boldsymbol{X}_{i0}\mathbf{\Gamma}\boldsymbol{X}_{i0}^\intercal)^{-2}\boldsymbol{X}_{i0}\mathbf{\Gamma}, \\
\boldsymbol{\tau}_{i0} &= (1 + n^{-1}\mathrm{tr}\boldsymbol{H}_2)\boldsymbol{I}_K - \boldsymbol{U}_1^\intercal(\boldsymbol{I}_p + \boldsymbol{H}_1)(\boldsymbol{I}_p + \boldsymbol{H}_1)^\intercal \boldsymbol{U}_1, &\text{(S2.22)} \\
\boldsymbol{\tau}_i &= n^{-1}(\boldsymbol{x}_i^\intercal \boldsymbol{H}_2 \boldsymbol{x}_i - \mathrm{tr}\boldsymbol{H}_2)\boldsymbol{I}_K - \boldsymbol{U}_1^\intercal(\boldsymbol{I}_p + \boldsymbol{H}_1)(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal - \boldsymbol{I}_p)(\boldsymbol{I}_p + \boldsymbol{H}_1)^\intercal \boldsymbol{U}_1. &\text{(S2.23)}
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\mathbf{\Omega}_K(\boldsymbol{X}_{i-1}) - \mathbf{\Omega}_K(\boldsymbol{X}_{i0}) \\
&= \frac{1}{\beta_{i0}\sqrt{n}}(\boldsymbol{\tau}_{i0} + \boldsymbol{\tau}_{iy}) - \frac{\varepsilon_{iy}}{\beta_{i0}^2\sqrt{n}}(\boldsymbol{\tau}_{i0} + \boldsymbol{\tau}_{iy}) + \frac{\varepsilon_{iy}^2}{\beta_{i0}^2\beta_{iy}\sqrt{n}}(\boldsymbol{\tau}_{i0} + \boldsymbol{\tau}_{iy}),
\end{aligned}$$

where $\beta_{iy}, \boldsymbol{\tau}_{iy}$, and ε_{iy} are similarly defined as $\beta_i, \boldsymbol{\tau}_i$, and ε_i with $\boldsymbol{x}_i$ replaced by $\boldsymbol{y}_i$.

Now, we begin to prove that $\boldsymbol{\Omega}_K(\boldsymbol{X})$ and $\boldsymbol{\Omega}_K(\boldsymbol{Y})$ have the same limiting distribution. To this end, we show that the difference between the characteristic functions tends to zero, that is, for any $K \times K$ deterministic symmetric matrix $\boldsymbol{W}$,

$$\mathbb{E}\Big(\exp\big[\mathrm{i}\,\mathrm{tr}\{\boldsymbol{W}\boldsymbol{\Omega}_K(\boldsymbol{X})\}\big]\Big) - \mathbb{E}\Big(\exp\big[\mathrm{i}\,\mathrm{tr}\{\boldsymbol{W}\boldsymbol{\Omega}_K(\boldsymbol{Y})\}\big]\Big) \to 0,$$

where $\mathrm{i} = \sqrt{-1}$. Using the notations we introduced above, we can write

$$\begin{aligned}
&\mathbb{E}\Big(\exp\big[\mathrm{i}\,\mathrm{tr}\{\boldsymbol{W}\boldsymbol{\Omega}_K(\boldsymbol{X})\}\big]\Big) - \mathbb{E}\Big(\exp\big[\mathrm{i}\,\mathrm{tr}\{\boldsymbol{W}\boldsymbol{\Omega}_K(\boldsymbol{Y})\}\big]\Big)\\
=&\sum_{i=1}^{n}\mathbb{E}\exp\bigg(\mathrm{i}\,\mathrm{tr}\Big[\boldsymbol{W}\Big\{\boldsymbol{\Omega}_K(\boldsymbol{X}_{i0}) + \frac{\boldsymbol{\tau}_{i0}}{\beta_{i0}\sqrt{n}}\Big\}\Big]\bigg)\bigg\{\mathbb{E}_i\exp\bigg(\mathrm{i}\,\mathrm{tr}\Big[\boldsymbol{W}\Big\{\frac{\boldsymbol{\tau}_i}{\beta_{i0}\sqrt{n}} - \frac{(\boldsymbol{\tau}_{i0}+\boldsymbol{\tau}_i)\varepsilon_i}{\beta_{i0}^2\sqrt{n}}\Big\}\Big]\bigg)\\
&\qquad - \mathbb{E}_i\exp\bigg(\mathrm{i}\,\mathrm{tr}\Big[\boldsymbol{W}\Big\{\frac{\boldsymbol{\tau}_{iy}}{\beta_{i0}\sqrt{n}} - \frac{(\boldsymbol{\tau}_{i0}+\boldsymbol{\tau}_{iy})\varepsilon_{iy}}{\beta_{i0}^2\sqrt{n}}\Big\}\Big]\bigg)\bigg\}\\
&+\sum_{i=1}^{n}\mathbb{E}\exp\bigg(\mathrm{i}\,\mathrm{tr}\Big[\boldsymbol{W}\Big\{\boldsymbol{\Omega}_K(\boldsymbol{X}_{i0}) + \frac{\boldsymbol{\tau}_{i0}}{\beta_{i0}\sqrt{n}}\Big\}\Big]\bigg)\\
&\quad\bigg\{\mathbb{E}_i\exp\bigg(\mathrm{i}\,\mathrm{tr}\Big[\boldsymbol{W}\Big\{\frac{\boldsymbol{\tau}_i}{\beta_{i0}\sqrt{n}} - \frac{(\boldsymbol{\tau}_{i0}+\boldsymbol{\tau}_i)\varepsilon_i}{\beta_{i0}^2\sqrt{n}}\Big\}\Big]\bigg)\bigg(\exp\Big[\mathrm{i}\,\mathrm{tr}\Big\{\frac{\boldsymbol{W}(\boldsymbol{\tau}_{i0}+\boldsymbol{\tau}_i)\varepsilon_i^2}{\beta_{i0}^2\beta_i\sqrt{n}}\Big\}\Big]-1\bigg)\\
&\quad - \mathbb{E}_i\exp\bigg(\mathrm{i}\,\mathrm{tr}\Big[\boldsymbol{W}\Big\{\frac{\boldsymbol{\tau}_{iy}}{\beta_{i0}\sqrt{n}} - \frac{(\boldsymbol{\tau}_{i0}+\boldsymbol{\tau}_{iy})\varepsilon_{iy}}{\beta_{i0}^2\sqrt{n}}\Big\}\Big]\bigg)\bigg(\exp\Big[\mathrm{i}\,\mathrm{tr}\Big\{\frac{\boldsymbol{W}(\boldsymbol{\tau}_{i0}+\boldsymbol{\tau}_{iy})\varepsilon_{iy}^2}{\beta_{i0}^2\beta_{iy}\sqrt{n}}\Big\}\Big]-1\bigg)\bigg\},
\end{aligned}$$

where $\mathbb{E}_i(\cdot) := \mathbb{E}(\cdot \mid \boldsymbol{X}_{i0})$ denotes the conditional expectation given $\boldsymbol{X}_{i0}$. The second term on the RHS of the above identity can be shown to be negligible by using the same argument as in (Jiang and Bai, 2021, p. 16) and our Lemma S1.7. Hence, we have

$$\begin{aligned}
&\mathbb{E}\Big(\exp\big[\mathrm{i}\,\mathrm{tr}\{\boldsymbol{W}\boldsymbol{\Omega}_K(\boldsymbol{X})\}\big]\Big) - \mathbb{E}\Big(\exp\big[\mathrm{i}\,\mathrm{tr}\{\boldsymbol{W}\boldsymbol{\Omega}_K(\boldsymbol{Y})\}\big]\Big)\\
=&\sum_{i=1}^{n}\mathbb{E}\exp\bigg(\mathrm{i}\,\mathrm{tr}\Big[\boldsymbol{W}\Big\{\boldsymbol{\Omega}_K(\boldsymbol{X}_{i0}) + \frac{\boldsymbol{\tau}_{i0}}{\beta_{i0}\sqrt{n}}\Big\}\Big]\bigg)\cdot\bigg\{\mathbb{E}_i\exp\bigg(\mathrm{i}\,\mathrm{tr}\Big[\boldsymbol{W}\Big\{\frac{\boldsymbol{\tau}_i}{\beta_{i0}\sqrt{n}} - \frac{(\boldsymbol{\tau}_{i0}+\boldsymbol{\tau}_i)\varepsilon_i}{\beta_{i0}^2\sqrt{n}}\Big\}\Big]\bigg)\\
&\quad - \mathbb{E}_i\exp\bigg(\mathrm{i}\,\mathrm{tr}\Big[\boldsymbol{W}\Big\{\frac{\boldsymbol{\tau}_{iy}}{\beta_{i0}\sqrt{n}} - \frac{(\boldsymbol{\tau}_{i0}+\boldsymbol{\tau}_{iy})\varepsilon_{iy}}{\beta_{i0}^2\sqrt{n}}\Big\}\Big]\bigg)\bigg\} + o(1).
\end{aligned}$$

Moreover, by using the same argument, we obtain

$$
\begin{aligned}
&\mathbb{E}\Big(\exp\big[\mathrm{i}\,\mathrm{tr}\{\boldsymbol{W}\boldsymbol{\Omega}_K(\boldsymbol{X})\}\big]\Big)-\mathbb{E}\Big(\exp\big[\mathrm{i}\,\mathrm{tr}\{\boldsymbol{W}\boldsymbol{\Omega}_K(\boldsymbol{Y})\}\big]\Big)\\
&=\sum_{i=1}^{n}\mathbb{E}\exp\Big(\mathrm{i}\,\mathrm{tr}\Big[\boldsymbol{W}\Big\{\boldsymbol{\Omega}_K(\boldsymbol{X}_{i0})+\frac{\boldsymbol{\tau}_{i0}}{\beta_{i0}\sqrt{n}}\Big\}\Big]\Big)\cdot\Big(\mathbb{E}_i\exp\Big[\mathrm{i}\,\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_i}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_i}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}\Big]\\
&\qquad-\mathbb{E}_i\exp\Big[\mathrm{i}\,\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_{iy}}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_{iy}}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}\Big]\Big)+o(1).
\end{aligned}
$$

This, together with Taylor's expansion, gives us

$$
\begin{aligned}
&\mathbb{E}\Big(\exp\big[\mathrm{i}\,\mathrm{tr}\{\boldsymbol{W}\boldsymbol{\Omega}_K(\boldsymbol{X})\}\big]\Big)-\mathbb{E}\Big(\exp\big[\mathrm{i}\,\mathrm{tr}\{\boldsymbol{W}\boldsymbol{\Omega}_K(\boldsymbol{Y})\}\big]\Big)\\
&\leqslant\Bigg|\sum_{i=1}^{n}\mathbb{E}\exp\Big(\mathrm{i}\,\mathrm{tr}\Big[\boldsymbol{W}\Big\{\boldsymbol{\Omega}_K(\boldsymbol{X}_{i0})+\frac{\boldsymbol{\tau}_{i0}}{\beta_{i0}\sqrt{n}}\Big\}\Big]\Big)\times\\
&\qquad C\Bigg\{\mathbb{E}_i\Big(1+\mathrm{i}\,\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_i}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_i}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}-\frac{1}{2}\Big[\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_i}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_i}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}\Big]^2+o(n^{-1})\Big)\\
&\qquad-\mathbb{E}_i\Big(1+\mathrm{i}\,\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_{iy}}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_{iy}}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}-\frac{1}{2}\Big[\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_{iy}}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_{iy}}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}\Big]^2+o(n^{-1})\Big)\Bigg\}\Bigg|\\
&=o(1),
\end{aligned}
$$

where we use some facts as follows:

$$
\begin{aligned}
&\mathbb{E}_i\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_i}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_i}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}=\mathbb{E}_i\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_{iy}}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_{iy}}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}=0,\\
&\mathbb{E}_i\Big[\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_i}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_i}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}\Big]^2-\mathbb{E}_i\Big[\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_{iy}}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_{iy}}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}\Big]^2=o(n^{-1}).
\end{aligned}
\tag{S2.24}
$$

Finally, it remains to prove (S2.24). Note that

$$
\begin{aligned}
&\Big[\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_i}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_i}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}\Big]^2\\
&=\frac{1}{n}\Big[\frac{1}{\beta_{i0}^2}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_i)\}^2-\frac{2\varepsilon_i}{\beta_{i0}^3}\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_i)\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{i0})+\frac{\varepsilon_i^2}{\beta_{i0}^4}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{i0})\}^2\Big],
\end{aligned}
$$

and from Lemma S1.7, we have

$$\frac{1}{\beta_{i0}^3}\mathbb{E}\big\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_i)\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{i0}\varepsilon_i)\big\} \asymp \mathbb{E}\,\varepsilon_i = o(1),$$
$$\frac{1}{\beta_{i0}^4}\mathbb{E}\big[\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{i0})\}^2\varepsilon_i^2\big] \asymp \mathbb{E}\,\varepsilon_i^2 = o(1).$$

Since $\boldsymbol{X}$ and $\boldsymbol{Y}$ satisfy (B1) – (B4), by using above estimates and Lemma S1.7, we obtain

$$\begin{aligned}&\mathbb{E}_i\Big[\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_i}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_i}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}\Big]^2-\mathbb{E}_i\Big[\mathrm{tr}\Big\{\boldsymbol{W}\Big(\frac{\boldsymbol{\tau}_{iy}}{\beta_{i0}\sqrt{n}}-\frac{\boldsymbol{\tau}_{i0}\varepsilon_{iy}}{\beta_{i0}^2\sqrt{n}}\Big)\Big\}\Big]^2\\&=\frac{1}{n}\mathbb{E}_i\bigg(\frac{1}{\beta_{i0}^2}\Big[\big\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_i)\big\}^2-\big\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{iy})\big\}^2\Big]+o(1)\bigg)=o(n^{-1}),\end{aligned}$$

which is (S2.24), completing the proof of Lemma 6.2.

S2.5 Proof of Lemma S1.6

From the definition of $\boldsymbol{Q}$ in (S1.1), we obtain $\boldsymbol{Q}|(W_X,W_{Y_1},W_{Y_2})\sim\mathcal{N}_2(\boldsymbol{0},\boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma}=\begin{pmatrix}W_X\sigma_{X_1}^2+W_{Y_1}\sigma_{Y_1}^2 & \gamma W_X\\ \gamma W_X & W_X\sigma_{X_2}^2+W_{Y_2}\sigma_{Y_2}^2\end{pmatrix}.$$

For $\widetilde{Q}_1$ and $\widetilde{Q}_2$, they have the following stochastic representations:

$$\widetilde{Q}_1\overset{\mathsf{d}}{=}\sqrt{\widetilde{W}_{X_1}}\widetilde{X}_1+\sqrt{\widetilde{W}_{Y_1}}\widetilde{Y}_1,\qquad \widetilde{Q}_2\overset{\mathsf{d}}{=}\sqrt{\widetilde{W}_{X_2}}\widetilde{X}_2+\sqrt{\widetilde{W}_{Y_2}}\widetilde{Y}_2,$$

where $(\widetilde{W}_{X_1},\widetilde{W}_{Y_1})$, $(\widetilde{W}_{X_2},\widetilde{W}_{Y_2})$, $(\widetilde{X}_1,\widetilde{Y}_1)$, and $(\widetilde{X}_2,\widetilde{Y}_2)$ are independent copies of (W_X,W_{Y_1}), (W_X,W_{Y_2}), (X_1,Y_1), and (X_2,Y_2), respectively. Note that $\widetilde{X}_1$ and $\widetilde{X}_2$ are independent. Two random vectors $\boldsymbol{Q}$ and $\widetilde{\boldsymbol{Q}}=(\widetilde{Q}_1,\widetilde{Q}_2)^\intercal$ are independent, and $\widetilde{\boldsymbol{Q}}|(\widetilde{W}_{X_1},\widetilde{W}_{X_2},\widetilde{W}_{Y_1},\widetilde{W}_{Y_2})\sim$

$\mathcal{N}_2(\mathbf{0}, \widetilde{\mathbf{\Sigma}})$, where

$$\widetilde{\mathbf{\Sigma}} = \begin{pmatrix} \widetilde{W}_{X_1}\sigma_{X_1}^2 + \widetilde{W}_{Y_1}\sigma_{Y_1}^2 & 0 \\ 0 & \widetilde{W}_{X_2}\sigma_{X_2}^2 + \widetilde{W}_{Y_2}\sigma_{Y_2}^2 \end{pmatrix}.$$

Comnbining $\boldsymbol{Q}$ and $\widetilde{\boldsymbol{Q}}$ into a vector $\boldsymbol{\mathcal{Q}} := \binom{\boldsymbol{Q}}{\tilde{\boldsymbol{Q}}}$, then we have $\boldsymbol{\mathcal{Q}}|\boldsymbol{W} \sim \mathcal{N}_4(\mathbf{0}, \mathbf{\Omega})$, where $\boldsymbol{W} := (W_X, W_{Y_1}, W_{Y_2}, \widetilde{W}_{X_1}, \widetilde{W}_{X_2}, \widetilde{W}_{Y_1}, \widetilde{W}_{Y_2})$, and $\mathbf{\Omega} = \left(\begin{smallmatrix} \mathbf{\Sigma} & O \\ O & \widetilde{\mathbf{\Sigma}} \end{smallmatrix}\right)$. From

$$\left(\begin{smallmatrix} Q_1 - \widetilde{Q}_1 \\ Q_2 - \widetilde{Q}_2 \end{smallmatrix}\right) = (\boldsymbol{I}_2, -\boldsymbol{I}_2)\boldsymbol{\mathcal{Q}} =: \boldsymbol{B}_0\boldsymbol{\mathcal{Q}},$$

we obtain $\left(\begin{smallmatrix} Q_1 - \widetilde{Q}_1 \\ Q_2 - \widetilde{Q}_2 \end{smallmatrix}\right) \Big| \boldsymbol{W} \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{B}_0\mathbf{\Omega}\boldsymbol{B}_0^\intercal)$, where

$$\boldsymbol{B}_0\mathbf{\Omega}\boldsymbol{B}_0^\intercal = \begin{pmatrix} (W_X + \widetilde{W}_{X_1})\sigma_{X_1}^2 + (W_{Y_1} + \widetilde{W}_{Y_1})\sigma_{Y_1}^2 & \gamma W_X \\ \gamma W_X & (W_X + \widetilde{W}_{X_2})\sigma_{X_2}^2 + (W_{Y_2} + \widetilde{W}_{Y_2})\sigma_{Y_2}^2 \end{pmatrix}.$$

Given $\boldsymbol{W}$, the Pearson correlation coefficient between $Q_1 - \widetilde{Q}_1$ and $Q_2 - \widetilde{Q}_2$ is

$$r = \frac{\gamma W_X}{\sqrt{(W_X + \widetilde{W}_{X_1})\sigma_{X_1}^2 + (W_{Y_1} + \widetilde{W}_{Y_1})\sigma_{Y_1}^2}\sqrt{(W_X + \widetilde{W}_{X_2})\sigma_{X_2}^2 + (W_{Y_2} + \widetilde{W}_{Y_2})\sigma_{Y_2}^2}}.$$

By Lemma S1.5, we conclude that

$$\mathbb{E}\{\mathrm{sign}(Q_1 - \widetilde{Q}_1)\mathrm{sign}(Q_2 - \widetilde{Q}_2)\} = \mathbb{E}\Big[\mathbb{E}\Big\{\mathrm{sign}(Q_1 - \widetilde{Q}_1)\mathrm{sign}(Q_2 - \widetilde{Q}_2)\Big|\boldsymbol{W}\Big\}\Big] = \frac{2}{\pi}\mathbb{E}\big\{\arcsin(r)\big\}.$$

This completes the proof of Lemma S1.6.

S2.6 Proof of Lemma S1.7

The convergence of β_{i0} can be proven by using the same argument in the proof of Lemma C.3 of Jiang and Bai (2021). The conclusion $\varepsilon_i \to 0$ follows from (S1.5). First, we prove that $\mathbb{E}\,\varepsilon_i^2 = o(1)$. Since $\|\mathbf{\Gamma}\|_2$ is bounded and the variable λ (spiked eigenvalue) stays away from the bulk eigenvalues of $n^{-1}\mathbf{\Gamma}^{1/2}\boldsymbol{X}_{i0}^{\intercal}\boldsymbol{X}_{i0}\mathbf{\Gamma}^{1/2}$, the spectral norm of $\boldsymbol{B} := \mathbf{\Gamma}^{1/2}(\lambda\boldsymbol{I} - n^{-1}\mathbf{\Gamma}^{1/2}\boldsymbol{X}_{i0}^{\intercal}\boldsymbol{X}_{i0}\mathbf{\Gamma}^{1/2})^{-1}\mathbf{\Gamma}^{1/2}$ is bounded. By Assumptions (A1) and (B3), we have

$$\mathbb{E}\,\varepsilon_i^2 = \frac{1}{n^2}\mathbb{E}\big|\boldsymbol{x}_i^{\intercal}\boldsymbol{B}\boldsymbol{x}_i - \mathrm{tr}\boldsymbol{B}\big|^2 = o(1).$$

Second, we prove $\mathbb{E}\varepsilon_i^4 = o(n^{-1})$. From Assumption (B4), we can apply Lemma S1.2 to estimating moments of ε_i. For $\delta > 0$, let $t_n(\delta) = (p/n)(\log p)^{1+\delta}/\sqrt{pc(p)}$, where $c(p)$ is defined in Assumption (B4). Define an event

$$\mathcal{F}_n(\delta) = \{\boldsymbol{x}_i : n^{-1}(\boldsymbol{x}_i^{\intercal}\boldsymbol{B}\boldsymbol{x}_i - \mathrm{tr}\boldsymbol{B}) < t_n(\delta)\}.$$

From Lemma S1.2 and the fact $\mathbb{E}(\boldsymbol{x}_1\boldsymbol{x}_1^{\intercal}) = \boldsymbol{I}_p$ (see Assumption (B2)), we have $\mathbb{P}\big(\mathcal{F}_n^c(\delta)\big) = o(n^{-1})$. By the elementary inequality $|a+b|^r \leqslant 2^{r-1}(|a|^r + |b|^r)$ for $r > 1$, we have

$$\mathbb{E}\,\varepsilon_i^4 = \mathbb{E}\big|\varepsilon_i\mathbb{1}_{\{\mathcal{F}_n(\delta)\}} + \varepsilon_i\mathbb{1}_{\{\mathcal{F}_n^c(\delta)\}}\big|^4 \leqslant 8\big[\mathbb{E}\varepsilon_i^4\mathbb{1}_{\{\mathcal{F}_n(\delta)\}} + \mathbb{E}\varepsilon_i^4\mathbb{1}_{\{\mathcal{F}_n^c(\delta)\}}\big]. \tag{S2.25}$$

For the first term, we have

$$\mathbb{E}\,\varepsilon_i^4\mathbb{1}_{\{\mathcal{F}_n(\delta)\}} = \int_0^{\infty} 4t^3\mathbb{P}\big(\big|\varepsilon_i\mathbb{1}_{\{\mathcal{F}_n(\delta)\}}\big| > t\big)\,\mathrm{d}t \leqslant \int_0^{t_n(\delta)} 4t^3\,\mathrm{d}t = o(n^{-1}). \tag{S2.26}$$

For the second term, by the fact that $|\varepsilon_i| = |n^{-1}\big(\boldsymbol{x}_i^{\intercal}\boldsymbol{B}\boldsymbol{x}_i - \mathrm{tr}\boldsymbol{B}\big)| \leqslant C\|\boldsymbol{B}\|_2$ (using Assumption (B2) and the Courant–Fischer principle) and $\|\boldsymbol{B}\|_2 = O(1)$, we obtain

$$\mathbb{E}\,\varepsilon_i^4\mathbb{1}_{\{\mathcal{F}_n^c(\delta)\}} \leqslant C^4\|\boldsymbol{B}\|_2^4\mathbb{P}\big(\mathcal{F}_n^c(\delta)\big) = o(n^{-1}).$$

This, together with (S2.25) and (S2.26), implies that $\mathbb{E}\,\varepsilon_i^4 = o(n^{-1})$.

Finally, we prove (S1.6). The proofs of $\mathbb{E}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_i)\}^2 = O(1)$ and $\mathbb{E}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{i0})\}^2 = O(1)$ are similar, and thus we only prove the first conclusion. Let

$$\boldsymbol{\tau}_{i1} = n^{-1}(\boldsymbol{x}_i^\intercal \boldsymbol{H}_2 \boldsymbol{x}_i - \mathrm{tr}\boldsymbol{H}_2)\boldsymbol{I}_K,$$

$$\boldsymbol{\tau}_{i2} = \boldsymbol{U}_1^\intercal(\boldsymbol{I}_p + \boldsymbol{H}_1)(\boldsymbol{x}_i\boldsymbol{x}_i^\intercal - \boldsymbol{I}_p)(\boldsymbol{I}_p + \boldsymbol{H}_1)^\intercal \boldsymbol{U}_1,$$

then we can write $\boldsymbol{\tau}_i = \boldsymbol{\tau}_{i1} - \boldsymbol{\tau}_{i2}$. To prove $\mathbb{E}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_i)\}^2 = O(1)$, it suffices to show that $\mathbb{E}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{i2})\}^2 = O(1)$, since

$$\mathbb{E}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_i)\}^2 = \mathbb{E}\{\mathrm{tr}\boldsymbol{W}(\boldsymbol{\tau}_{i1} - \boldsymbol{\tau}_{i2})\}^2 \leqslant 2\mathbb{E}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{i1})\}^2 + 2\mathbb{E}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{i2})\}^2,$$

and

$$\mathbb{E}\{\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{i1})\}^2 = n^{-2}\mathbb{E}(\boldsymbol{x}_i^\intercal \boldsymbol{H}_2 \boldsymbol{x}_i - \mathrm{tr}\boldsymbol{H}_2)^2 \cdot (\mathrm{tr}\boldsymbol{W})^2 = o(1),$$

which follows from Assumption (B3) and the fact $\|\boldsymbol{H}_2\|_2 = O(1)$. Let

$$\boldsymbol{\Delta} = (\boldsymbol{I}_p + \boldsymbol{H}_1^\intercal)\boldsymbol{U}_1\boldsymbol{W}\boldsymbol{U}_1^\intercal(\boldsymbol{I}_p + \boldsymbol{H}_1),$$

then we have

$$\mathbb{E}[\mathrm{tr}(\boldsymbol{W}\boldsymbol{\tau}_{i2})]^2 = \mathbb{E}|\boldsymbol{x}_i^\intercal \boldsymbol{\Delta}\boldsymbol{x}_i - \mathrm{tr}\boldsymbol{\Delta}|^2 \leqslant 2\mathbb{E}|\boldsymbol{x}_i^\intercal \boldsymbol{\Delta}\boldsymbol{x}_i|^2 + 2\mathbb{E}|\mathrm{tr}\boldsymbol{\Delta}|^2.$$

Note that

$$\mathrm{tr}(\boldsymbol{\Delta}) = \mathrm{tr}\big\{\boldsymbol{W}\boldsymbol{U}_1^\intercal(\boldsymbol{I}_p + \boldsymbol{H}_1)(\boldsymbol{I}_p + \boldsymbol{H}_1^\intercal)\boldsymbol{U}_1\big\}$$

$$\leqslant K\|\boldsymbol{W}\boldsymbol{U}_1^\intercal(\boldsymbol{I}_p+\boldsymbol{H}_1)(\boldsymbol{I}_p+\boldsymbol{H}_1^\intercal)\boldsymbol{U}_1\|_2 = O(1),$$

and similarly, $(\mathrm{tr}\boldsymbol{\Delta})^2 = O(1)$. Thus, we need only to show $\mathbb{E}|\boldsymbol{x}_i^\intercal\boldsymbol{\Delta}\boldsymbol{x}_i|^2 = O(1)$. Let $\Delta_{k\ell}$ denote the (k,ℓ)-th entry of $\boldsymbol{\Delta}$. From Lemma S1.4, we obtain $\sum_{k=1}^p \Delta_{kk}^2 \leqslant \mathrm{tr}(\boldsymbol{\Delta}^2) = O(1)$. By using Cauchy-Schwarz inequality, we obtain

$$\sum_{k\neq\ell}\Delta_{kk}\Delta_{\ell\ell} \leqslant \Big(\sum_k \Delta_{kk}\Big)\Big(\sum_\ell \Delta_{\ell\ell}\Big) = (\mathrm{tr}\boldsymbol{\Delta})^2 = O(1),$$

$$\sum_{k\neq\ell\neq s}\Delta_{k\ell}\Delta_{ks} \asymp \sum_{k\neq\ell\neq s}\Delta_{kk}\Delta_{\ell s} \leqslant \Big(\sum_k \Delta_{kk}\Big)\cdot\Big(\sum_{\ell\neq s}\Delta_{\ell s}\Big) \leqslant (\mathrm{tr}\boldsymbol{\Delta})\cdot p(\mathrm{tr}\boldsymbol{\Delta}^2)^{1/2} = O(p),$$

$$\sum_{k\neq\ell\neq s\neq t}\Delta_{k\ell}\Delta_{st} \leqslant \Big(\sum_{k\neq\ell}\Delta_{k\ell}\Big)\Big(\sum_{s\neq t}\Delta_{st}\Big) \leqslant p^2\mathrm{tr}(\boldsymbol{\Delta}^2) = O(p^2).$$

It follows from the above inequalities and Assumption (B2) that

$$\begin{aligned}
\mathbb{E}|\boldsymbol{x}_i^\intercal\boldsymbol{\Delta}\boldsymbol{x}_i|^2 &= \sum_{k=1}^p \mathbb{E}\Delta_{kk}^2\cdot\mathbb{E}X_{ik}^4 + \sum_{k\neq\ell}\mathbb{E}(\Delta_{kk}\Delta_{\ell\ell}+2\Delta_{k\ell}^2)\cdot\mathbb{E}X_{ik}^2X_{i\ell}^2 \\
&\quad + \sum_{k\neq\ell\neq s}\mathbb{E}(2\Delta_{kk}\Delta_{\ell s}+4\Delta_{k\ell}\Delta_{ks})\cdot\mathbb{E}X_{ik}^2X_{i\ell}X_{is} \\
&\quad + \sum_{k\neq\ell\neq s\neq t}\mathbb{E}\Delta_{k\ell}\Delta_{st}\cdot\mathbb{E}X_{ik}X_{i\ell}X_{is}X_{it} \\
&= O(1).
\end{aligned}$$

This completes the proof of Lemma S1.7.

References

Bai, Z. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statistica Sinica*, 18(2):425–442.

El Karoui, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *The Annals of Applied Probability*, 19(6):2362 – 2405.

Horn, R. A. and Johnson, C. R. (1991). *Topic in matrix analysis.* Cambridge University Press.

Horn, R. A. and Johnson, C. R. (2012). *Matrix analysis.* Cambridge University Press.

Jiang, D. and Bai, Z. (2021). Supplement to "Generalized four moment theorem and an application to CLT for spiked eigenvalues of high-dimensional covariance matrices". *Bernoulli.*

Ledoux, M. (2001). *The concentration of measure phenomenon*, volume 89. American Mathematical Society.

Li, Z., Wang, C., and Wang, Q. (2022). On eigenvalues of a high dimensional Kendall's rank correlation matrix with dependence. *Science China Mathematics*, to appear.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press.