Supplementary Materials to "High-dimensional MANOVA via Bootstrapping and its Application to Functional and Sparse Count Data"

Zhenhua Lin*
Department of Statistics and Applied Probability, National University of Singapore

Miles E. Lopes
Department of Statistics, University of California, Davis

Hans-Georg Müller†
Department of Statistics, University of California, Davis

General remarks and notation. Throughout we refer to the notations introduced in the main text and define $m_\circ = \min\{m_1,\ldots,m_K\}$, $m_{\max} = \max\{m_1,\ldots,m_K\}$, $\ell_\circ = \min\{\ell_1,\ldots,\ell_K\}$, and $\ell_{\max} = \max\{\ell_1,\ldots,\ell_K\}$. Let $\mathcal{J}_{k,l}(m_k,m_l) = \mathcal{J}_k(m_k) \cup \mathcal{J}_l(m_l)$ and $m_\circ \le m_{k,l} := |\mathcal{J}_{k,l}(m_k,m_l)| \le m_k + m_l \le 2m_{\max}$. Define $\ell_{k,l}$ analogously. Define $\lambda_{k,l}^2 = n_l/(n_k+n_l)$ and

$$M_m(k,l) = \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \left(\lambda_{k,l}S_{k,j}/\sigma^\tau_{k,l,j} - \lambda_{l,k}S_{l,j}/\sigma^\tau_{k,l,j}\right),$$

and define $\tilde{M}_m(k,l)$ and $M^\star_m(k,l)$ analogously. Let $N = |\mathcal{P}|$ and suppose we enumerate the pairs in $\mathcal{P}$ by $(k_1,l_1),\ldots,(k_N,l_N)$. Let $\mathbf{m} = (m_{k_1,l_1},\ldots,m_{k_N,l_N})$. Define

$$M_{\mathbf{m}} = \max_{(k,l)\in\mathcal{P}} M_m(k,l),$$

and $\tilde{M}_{\mathbf{m}}$ and $M^\star_{\mathbf{m}}$ analogously. In addition, define

$$\kappa = \alpha(1-\tau).$$

*Research partially supported by NUS start-up grant R-155-001-217-133.
†Research supported by NSF grant DMS-2014626.

Lastly, the constant $c > 0$ used in the proofs below may vary from place to place; however, it does not depend on K, N, p or $n_1, \ldots, n_K$.

Remark. Under Assumption **??**, all ℓ_k and thus all $\ell_{k,l}$ are of the same order as $\ell_\circ$ and $\ell_{\max}$, and similarly, all m_k and $m_{k,l}$ are of the same order as $m_\circ$ and $m_{\max}$.

Remark. It is sufficient to show that the results in the theorems hold for all large values of n. The proofs below implicitly assume $p > \min_{1\le q\le N} m_{k_q,l_q}$ (unless otherwise stated). If $p \le \min_{1\le q\le N} m_{k_q,l_q}$, then $p = m_{k_q,l_q}$ for all $1 \le q \le N$, and consequently, the quantities I and III in the proof of Theorem **??**, $\mathrm{I}'(X)$ and $\mathrm{III}'(X)$ in the proof of Theorem **??**, and I'' and III'' in the proof of Theorem **??** become exactly zero. In this case, the related proofs are simplified to bounding II, $\mathrm{II}'(X)$ and II''.

A Proof of Theorem ??

Proof. Consider the inequality

$$d_{\mathrm{K}}\left(\mathcal{L}(M), \mathcal{L}(\tilde{M})\right) \le \mathrm{I} + \mathrm{II} + \mathrm{III},$$

where we define

$$\begin{aligned}
\mathrm{I} &= d_{\mathrm{K}}\left(\mathcal{L}(M), \mathcal{L}(M_{\mathbf{m}})\right),\\
\mathrm{II} &= d_{\mathrm{K}}\left(\mathcal{L}(M_{\mathbf{m}}), \mathcal{L}(\tilde{M}_{\mathbf{m}})\right),\\
\mathrm{III} &= d_{\mathrm{K}}\left(\mathcal{L}(\tilde{M}_{\mathbf{m}}), \mathcal{L}(\tilde{M})\right).
\end{aligned}$$

Then the conclusion of the theorem follows from Propositions A.1 and A.2 below. ☐

Proposition A.1. *Under the conditions of Theorem* **??**, *we have* $\mathrm{II} \lesssim n^{-\frac{1}{2}+\delta}$.

Proof. Let Π denote the projection onto the coordinates indexed by $\mathcal{J} = \bigcup_{(k,l)\in\mathcal{P}} \mathcal{J}_{k,l}(m_k, m_l)$. Let $J = |\mathcal{J}|$. Define the $J \times J$ diagonal matrix $D_{k,l} = \operatorname{diag}(\sigma_{k,l,j} : j \in \mathcal{J})$. It follows that

$$M_m(k,l) = \max_{j\in\mathcal{I}(k,l)} e_j^\top D_{k,l}^{-\tau} \Pi(\lambda_{k,l} S_k - \lambda_{l,k} S_l),$$

where $e_j \in \mathbb{R}^J$ is the jth standard basis vector, and $\mathcal{I}(k,l)$ denotes the row indices involving $\mathcal{J}_{k,l}(m_k, m_l)$ in the projection Π. Let $\mathfrak{C}_{k,l}^\top = \lambda_{k,l} D_{k,l}^{-\tau} \Pi \Sigma_k^{1/2}$, which is of size $J \times p$.

Consider the QR decomposition $\Sigma_k^{1/2}\Pi^\top = Q_k V_k$ so that

$$\mathfrak{C}_{k,l} = Q_k V_k (\lambda_{k,l} D_{k,l}^{-\tau}) \equiv Q_k R_{k,l},$$

where the columns of $Q_k \in \mathbb{R}^{p\times J}$ are an orthonormal basis for the image of $\mathfrak{C}_{k,l}$ and $R_{k,l} \in \mathbb{R}^{J\times J}$. Define the random vectors

$$\breve{Z}_k = n_k^{-1/2} \sum_{i=1}^{n_k} Q_k^\top Z_{k,i}.$$

Then

$$D_{k,l}^{-\tau} \Pi(\lambda_{k,l} S_k - \lambda_{l,k} S_l) = R_{k,l}^\top \breve{Z}_k - R_{l,k}^\top \breve{Z}_l.$$

Let $R^\top$ be a $JN \times JK$ block matrix with $N \times K$ blocks of size $J \times J$ such that, for $q = 1, \ldots, N$ and $k = 1, \ldots, K$, the (q,k)-block is $R_{k_q,l_q}^\top$ if $k = k_q$, is $-R_{l_q,k_q}^\top$ if $k = l_q$, and is $\mathbf{0}$ otherwise. Then

$$\begin{pmatrix} D_{k_1,l_1}^{-\tau} \Pi(\lambda_{k_1,l_1} S_{k_1} - \lambda_{l_1,k_1} S_{l_1}) \\ D_{k_2,l_2}^{-\tau} \Pi(\lambda_{k_2,l_2} S_{k_2} - \lambda_{l_2,k_2} S_{l_2}) \\ \vdots \\ D_{k_N,l_N}^{-\tau} \Pi(\lambda_{k_N,l_N} S_{k_N} - \lambda_{l_N,k_N} S_{l_N}) \end{pmatrix} = R^\top \breve{Z},$$

where $\breve{Z}$ is the $(JK) \times 1$ column vector obtained by stacking the vectors $\breve{Z}_1, \ldots, \breve{Z}_K$.

It can be checked that for any fixed $t \in \mathbb{R}$, there exists a Borel convex set $\mathcal{A}_t \subset \mathbb{R}^r$, with $r = JK$, such that $\mathbb{P}(M_{\mathbf{m}} \le t) = \mathbb{P}(\breve{Z} \in \mathcal{A}_t)$. By the same reasoning, we also have $\mathbb{P}(\tilde{M}_{\mathbf{m}} \le t) = \gamma_r(\mathcal{A}_t)$, where γ_r is the standard Gaussian distribution on $\mathbb{R}^r$. Thus,

$$\mathrm{II} \le \sup_{\mathcal{A} \in \mathscr{A}} |\mathbb{P}(\breve{Z} \in \mathcal{A}) - \gamma_r(\mathcal{A})|,$$

where $\mathscr{A}$ denotes the collection of all Borel convex subsets of $\mathbb{R}^r$.

Now we apply Theorem 1.2 of (Bentkus, 2005), as follows. Let $n_{1:k} = \sum_{j=1}^k n_j$. Define $Y_i \in \mathbb{R}^r$ in the following way: For $k = 1, \ldots, K$ and $i' = 1, \ldots, n_k$, set $i = n_{1:k} - n_k + i'$ and set all coordinates of Y_i to zero except that $Y_{i,(Jk-J+1):(Jk)} = n_k^{-1/2} Q_k^\top Z_{k,i'}$, i.e., the subvector of Y_i at coordinates $Jk - J + 1, \ldots, Jk$ is equal to the vector $n_k^{-1/2} Q_k^\top Z_{k,i'}$.

Then $\breve{Z} = \sum_{i=1}^{\mathbf{n}} Y_i$, i.e., $\breve{Z}$ is a sum of $\mathbf{n} = \sum_{k=1}^K n_k$ independent random vectors. We also observe that $\mathrm{cov}(\breve{Z}) = I_r$. For $n_{1:k} - n_k + 1 \le i \le n_{1:k}$, $\beta_i := \mathbb{E}\|\{\mathrm{cov}(\breve{Z})\}^{-1} Y_i\|^3 = \mathbb{E}\|Y_i\|^3 = n_k^{-3/2} \mathbb{E}\|Q_k^\top Z_{k,1}\|^3 \le n_k^{-3/2} [\mathbb{E}(Z_{k,1}^\top Q_k Q_k^\top Z_{k,1})^2]^{3/4}$, where the inequality is due to Lyapunov's inequality. Let v_j be the jth column of Q_1. If we put $\zeta_j = Z_{1,1}^\top v_j$, then

$$\mathbb{E}(Z_{1,1}^\top Q_1 Q_1^\top Z_{1,1})^2 = \left\| \sum_{j=1}^J \zeta_j^2 \right\|_2^2 \le \left(\sum_{j=1}^J \|\zeta_j^2\|_2 \right)^2 \lesssim J^2,$$

where we used the fact that $\|Z_{1,1}^\top v_j\|_4^2 \le c$ based on Assumption **??**, where $c > 0$ is a constant depending only on c_0 of Assumption **??**. The same argument applies to the quantity $\mathbb{E}(Z_{k,1}^\top Q_k Q_k^\top Z_{k,1})^2$ for a generic k with the same constant c. This implies that $\beta_i \le c n_k^{-3/2} J^{3/2}$ for all $n_{1:k} - n_k + 1 \le i \le n_{1:k}$, and some constant $c > 0$

not depending on n. Therefore,

$$\mathrm{II} \lesssim J^{1/4} \sum_{i=1}^{n_1+\cdots+n_K} \beta_i \lesssim J^{7/4} \sum_{k=1}^{K} n_k^{-1/2} \lesssim N^{7/4} m_{\max}^{7/4} K n^{-1/2} \lesssim n^{-1/2+\delta},$$

where the third inequality is due to $J \le 2Nm_{\max}$, and the last one follows from $\max\{K, N\} \lesssim e^{\sqrt{\log n}} \lesssim n^{\delta}$ and $m_{\max} \lesssim n_{\max}^{\delta} \asymp n^{\delta}$ for any fixed $\delta > 0$. □

Proposition A.2. *Under the conditions of Theorem ??, we have* $\mathrm{I} \lesssim n^{-1/2+\delta}$ *and* $\mathrm{III} \lesssim n^{-1/2+\delta}$.

Proof. We only establish the bound for I, since the same argument applies to III. For any fixed $t \in \mathbb{R}$,

$$|\mathbb{P}(M \le t) - \mathbb{P}(M_{\mathbf{m}} \le t)| = \mathbb{P}\left(A(t) \cap B(t)\right),$$

where

$$A(t) = \left\{ \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} (\lambda_{k,l} S_{k,j}/\sigma_{k,l,j}^{\tau} - \lambda_{l,k} S_{l,j}/\sigma_{k,l,j}^{\tau}) \le t \right\},$$
$$B(t) = \left\{ \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}^c_{k,l}(m_k,m_l)} (\lambda_{k,l} S_{k,j}/\sigma_{k,l,j}^{\tau} - \lambda_{l,k} S_{l,j}/\sigma_{k,l,j}^{\tau}) > t \right\},$$

and $\mathcal{J}^c_{k,l}(m_k, m_l)$ denotes the complement of $\mathcal{J}_{k,l}(m_k, m_l)$ in $\{1, \ldots, p\}$. Also, if $t_1 \le t_2$, it is seen that

$$A(t) \cap B(t) \subset A(t_2) \cup B(t_1)$$

for all $t \in \mathbb{R}$. By a union bound, we have

$$\mathrm{I} \le \mathbb{P}(A(t_2)) + \mathbb{P}(B(t_1)).$$

Take

$$t_1 = c m_{\circ}^{-\kappa} \log n$$
$$t_2 = c_2 c_{\circ} \ell_{\max}^{-\kappa} \sqrt{\log \ell_{\max}}$$

for a certain constant $c > 0$, where we recall that $c_2 \in (0, 1)$ is defined in Assumption ??. Then, $\mathbb{P}(A(t_2))$ and $\mathbb{P}(B(t_1))$ are at most of order $n^{-1/2+\delta}$, according to Lemma A.3 below. Moreover, the inequality $t_1 \le t_2$ holds for all large n, due to the definitions of $\ell_{\max}$, $m_{\circ}$, and κ, as well as the condition $(1-\tau)\sqrt{\log n} \gtrsim 1$. □

Lemma A.3. *Under the conditions of Theorem ??, there is a positive constant c, not depending on n, that can be selected in the definition of* t_1 *and* t_2*, so that*

$$\mathbb{P}(A(t_2)) \lesssim n^{-\frac{1}{2}+\delta}, \tag{S1}$$

and

$$\mathbb{P}(B(t_1)) \lesssim n^{-1}. \tag{S2}$$

Proof of (S1). Let $\mathcal{I}_{k,l}$ be a subset of $\mathcal{J}_{k,l}(\ell_k,\ell_l)$ constructed in the following way: if $\mathcal{J}_k(\ell_k)\cap\mathcal{J}_l(\ell_l)$ contains at least $\ell_\circ/2$ elements, then $\mathcal{I}_{k,l}=\mathcal{J}_k(\ell_k)\cap\mathcal{J}_l(\ell_l)$, and otherwise, $\mathcal{I}_{k,l}=\mathcal{J}_k(\ell_k)\cap\mathcal{J}_l^c(\ell_l)$ when $\sigma_{k,(\ell_k)}\geq\sigma_{l,(\ell_l)}$ and $\mathcal{I}_{k,l}=\mathcal{J}_k^c(\ell_k)\cap\mathcal{J}_l(\ell_l)$ when $\sigma_{k,(\ell_k)}<\sigma_{l,(\ell_l)}$. According to Proposition A.1 and the fact that $\mathcal{I}_{k,l}\subset\mathcal{J}_{k,l}(m_k,m_l)$, we have

$$\begin{aligned}\mathbb{P}(A(t_2)) &\leq \mathbb{P}\left(\max_{(k,l)\in\mathcal{P}}\max_{j\in\mathcal{J}_{k,l}(m_k,m_l)}(\lambda_{k,l}\tilde{S}_{k,j}/\sigma_{k,l,j}^{\tau}-\lambda_{l,k}\tilde{S}_{l,j}/\sigma_{k,l,j}^{\tau})\leq t_2\right)+\mathrm{II}\\ &\leq \mathbb{P}\left(\max_{(k,l)\in\mathcal{P}}\max_{j\in\mathcal{I}_{k,l}}(\lambda_{k,l}\tilde{S}_{k,j}/\sigma_{k,l,j}^{\tau}-\lambda_{l,k}\tilde{S}_{l,j}/\sigma_{k,l,j}^{\tau})\leq t_2\right)+cn^{-\frac{1}{2}+\delta}.\end{aligned}$$

As $\sigma_{k,l,j}=\sqrt{\lambda_{k,l}^2\sigma_{k,j}^2+\lambda_{l,k}^2\sigma_{l,j}^2}\geq\lambda_{k,l}\sigma_{k,j}$ and $\sigma_{k,(j)}\geq c_\circ j^{-\alpha}$ for $j\in\{1,\ldots,m_k\}$, and due to Assumption **??** with $c_\circ\in(0,1)$ and Assumption **??** with $c_2\in(0,1)$, we have $\sigma_{k,l,j}^{\tau-1}\leq\lambda_{k,l}^{\tau-1}\sigma_{k,j}^{\tau-1}\leq c_2^{\tau-1}\ell_k^{\alpha(1-\tau)}c_\circ^{\tau-1}\leq\ell_{\max}^{\kappa}/(c_2c_\circ)$ for $j\in\mathcal{I}_{k,l}$. With an argument similar to that of Lemma B.1 of Lopes et al. (2020), we can show that

$$\begin{aligned}&\mathbb{P}\left(\max_{(k,l)\in\mathcal{P}}\max_{j\in\mathcal{I}_{k,l}}(\lambda_{k,l}\tilde{S}_{k,j}/\sigma_{k,l,j}^{\tau}-\lambda_{l,k}\tilde{S}_{l,j}/\sigma_{k,l,j}^{\tau})\leq t_2\right)\\ &\leq \mathbb{P}\left(\max_{(k,l)\in\mathcal{P}}\max_{j\in\mathcal{I}_{k,l}}(\lambda_{k,l}\tilde{S}_{k,j}/\sigma_{k,l,j}-\lambda_{l,k}\tilde{S}_{l,j}/\sigma_{k,l,j})\leq\sqrt{\log\ell_{\max}}\right)\\ &\leq \sum_{(k,l)\in\mathcal{P}}\mathbb{P}\left(\max_{j\in\mathcal{I}_{k,l}}(\lambda_{k,l}\tilde{S}_{k,j}/\sigma_{k,l,j}-\lambda_{l,k}\tilde{S}_{l,j}/\sigma_{k,l,j})\leq\sqrt{\log\ell_{\max}}\right).\end{aligned}$$

Note that the cardinality of $\mathcal{I}_{k,l}$ is at least $\ell_\circ/2$. Based on Assumption **??**, for all sufficiently large n, for all $1\leq k<j\leq K$, we have $\log(\ell_{\max})\leq 1.01\log\ell_\circ\leq 1.01^2\log(2|\mathcal{I}_{k,l}|)\leq 1.1^2\log|\mathcal{I}_{k,l}|$. Then,

$$\begin{aligned}&\mathbb{P}\left(\max_{j\in\mathcal{I}_{k,l}}(\lambda_{k,l}\tilde{S}_{k,j}/\sigma_{k,l,j}-\lambda_{l,k}\tilde{S}_{l,j}/\sigma_{k,l,j})\leq\sqrt{\log\ell_{\max}}\right)\\ &\leq \mathbb{P}\left(\max_{j\in\mathcal{I}_{k,l}}(\lambda_{k,l}\tilde{S}_{k,j}/\sigma_{k,l,j}-\lambda_{l,k}\tilde{S}_{l,j}/\sigma_{k,l,j})\leq 1.1\sqrt{\log|\mathcal{I}_{k,l}|}\right).\end{aligned} \tag{S3}$$

To apply Lemma B.2 of Lopes et al. (2020), let Q denote the correlation matrix of the random variables $\{\lambda_{k,l}\tilde{S}_{k,j}/\sigma_{k,l,j}-\lambda_{l,k}\tilde{S}_{l,j}/\sigma_{k,l,j}:1\leq j\leq p\}$. When $\mathcal{I}_{k,l}=\mathcal{J}_k(\ell_k)\cap\mathcal{J}_l(\ell_l)$, for $j,r\in\mathcal{I}_{k,l}$, one has

$$\begin{aligned}Q_{j,r} &= \frac{\lambda_{k,l}^2R_{k,j,r}(p)\sigma_{k,j}\sigma_{k,r}+\lambda_{l,k}^2R_{l,j,r}(p)\sigma_{l,j}\sigma_{l,r}}{\sqrt{\lambda_{k,l}^2\sigma_{k,j}^2+\lambda_{l,k}^2\sigma_{l,j}^2}\sqrt{\lambda_{k,l}^2\sigma_{k,r}^2+\lambda_{l,k}^2\sigma_{l,r}^2}}\\ &\leq (1-\epsilon_0)\frac{\lambda_{k,l}^2\sigma_{k,j}\sigma_{k,r}+\lambda_{l,k}^2\sigma_{l,j}\sigma_{l,r}}{\sqrt{\lambda_{k,l}^2\sigma_{k,j}^2+\lambda_{l,k}^2\sigma_{l,j}^2}\sqrt{\lambda_{k,l}^2\sigma_{k,r}^2+\lambda_{l,k}^2\sigma_{l,k}^2}}\\ &\leq 1-\epsilon_0,\end{aligned}$$

since the construction of $\mathcal{I}_{k,l}$ implies that $\max\{R_{k,j,r},R_{l,j,r}\}\leq 1-\epsilon_0$. When $\mathcal{I}_{k,l}=\mathcal{J}_k(\ell_k)\cap\mathcal{J}_l^c(\ell_l)$ (so that

$\sigma_{k,(\ell_k)} \geq \sigma_{l,(\ell_l)}$), we have

$$
\begin{aligned}
Q_{j,k} &\leq 1 - \frac{\epsilon_0 \lambda_{k,l}^2 \sigma_{k,j}\sigma_{k,r}}{\sqrt{\lambda_{k,l}^2\sigma_{k,j}^2 + \lambda_{l,k}^2\sigma_{l,j}^2}\sqrt{\lambda_{k,l}^2\sigma_{k,r}^2 + \lambda_{l,k}^2\sigma_{l,r}^2}} \\
&\leq 1 - \epsilon_0 \frac{\lambda_{k,l}^2}{\lambda_{k,l}^2 + \lambda_{l,k}^2} \\
&\leq 1 - \epsilon_0,
\end{aligned}
$$

where the first inequality is obtained by using $R_{k,j,r} \leq 1 - \epsilon_0$ for $j, r \in \mathcal{I}_{k,l}$ and the inequality $\lambda_{k,l}^2\sigma_{k,j}\sigma_{k,r} + \lambda_{l,k}^2\sigma_{l,j}\sigma_{l,r} \leq \sqrt{\lambda_{k,l}^2\sigma_{k,j}^2 + \lambda_{l,k}^2\sigma_{l,j}^2}\sqrt{\lambda_{k,l}^2\sigma_{k,r}^2 + \lambda_{l,k}^2\sigma_{l,r}^2}$ and the second is due to $\sigma_{k,j} \geq \sigma_{k,(\ell_k)} \geq \sigma_{l,(\ell_l)} \geq \sigma_{l,j}$ as $j \in \mathcal{J}_k(\ell_k)$ and $j \in \mathcal{J}_l^c(\ell_l)$. A similar argument shows that the inequality $Q_{j,k} \leq 1 - \epsilon_0$ also holds when $\sigma_{k,(\ell_k)} < \sigma_{l,(\ell_l)}$, in which $\mathcal{I}_{k,l} = \mathcal{J}_k^c(\ell_k) \cap \mathcal{J}_l(\ell_l)$.

To apply Lemma B.2 of Lopes et al. (2020), we note that the bound $\sqrt{\log|\mathcal{I}_{k,l}|}$ is required instead of $1.1\sqrt{\log|\mathcal{I}_{k,l}|}$. However, by carefully examining the proof of Lemma B.2 of Lopes et al. (2020), we find that the lemma is still valid for $1.1\sqrt{\log|\mathcal{I}_{k,l}|}$, potentially with constants different from C and $\frac{1}{2}$ in (B.19) of Lopes et al. (2020). This shows that (S3) is bounded by cn^{-1} for some constant c not depending on n. Then $N \lesssim n^{\delta}$ for any $\delta > 0$ implies (S1). □

Proof of (S2). The following argument is similar to the proof for part (b) of Lemma B.1 in Lopes et al. (2020). Define the random variable

$$
V = \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}^c(m_k,m_l)} (\lambda_{k,l}S_{k,j}/\sigma_{k,l,j}^{\tau} - \lambda_{l,k}S_{l,j}/\sigma_{k,l,j}^{\tau})
$$

and let $q = \max\{2\kappa^{-1}, 3, \log n\}$. To bound $\|V\|_q$, we observe that

$$
\begin{aligned}
\|V\|_q^q &= \mathbb{E}\left[\Big|\max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \lambda_{k,l}S_{k,j}/\sigma_{k,l,j}^{\tau} - \lambda_{l,k}S_{l,j}/\sigma_{k,l,j}^{\tau}\Big|^q\right] \\
&\leq \sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \sigma_{k,l,j}^{q(1-\tau)} \mathbb{E}|\lambda_{k,l}S_{k,j}/\sigma_{k,l,j} - \lambda_{l,k}S_{l,j}/\sigma_{k,l,j}|^q.
\end{aligned}
$$

Further, we have

$$
\begin{aligned}
\sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \sigma_{k,l,j}^{q(1-\tau)} &\leq \sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \max\{\sigma_{k,j}, \sigma_{l,j}\}^{q(1-\tau)} \\
&\leq \sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}_{k,l}^c(m_k,m_l)} (\sigma_{k,j}^{q(1-\tau)} + \sigma_{l,j}^{q(1-\tau)}) \\
&\leq c_1^{q(1-\tau)} \sum_{(k,l)\in\mathcal{P}} \left(\sum_{j=m_k+1}^{p} j^{-\alpha q(1-\tau)} + \sum_{j=m_l+1}^{p} j^{-\alpha q(1-\tau)}\right) \\
&\leq c_1^{q(1-\tau)} \sum_{(k,l)\in\mathcal{P}} \left(2\int_{m_\circ}^{p} x^{-q\kappa}\mathrm{d}x\right)
\end{aligned}
$$

$$\leq 2c_1^{q(1-\tau)} N \frac{m_\circ^{-q\kappa+1}}{q\kappa - 1}, \tag{S4}$$

where we recall $\kappa = \alpha(1-\tau)$, and note that $q\kappa \geq 2$. Then, with $\|\lambda_{k,l} S_{k,j}/\sigma_{k,l,j} - \lambda_{l,k} S_{l,j}/\sigma_{k,l,j}\|_q \leq cq$ according to Lemma E.3, we deduce that

$$\|V\|_q^q \leq 2c_1^{q(1-\tau)} (cq)^q N \frac{m_\circ^{-q\kappa+1}}{q\kappa - 1},$$

and with $C = \frac{c}{(q\kappa-1)^{1/q}} m_\circ^{1/q} (2N)^{1/q} \lesssim 1$ that

$$\|V\|_q \leq Cq m_\circ^{-\kappa}.$$

Also, the assumption that $(1-\tau)\sqrt{\log n} \gtrsim 1$ implies that $q \lesssim \log n$. Therefore, with $t = e\|V\|_q$ so that $t \leq cm_\circ^{-\kappa} \log n$ for some constant $c > 0$ not depending on n, by Chebyshev's inequality $\mathbb{P}(V \geq t) \leq t^{-q}\|V\|_q^q$, we obtain that

$$\mathbb{P}\left(V \geq cm_\circ^{-\kappa} \log n\right) \leq \mathbb{P}(V \geq t) \leq e^{-q} \leq n^{-1},$$

completing the proof. □

B Proof of Theorem ??

Proof. Consider the inequality

$$d_{\mathrm{K}}\left(\mathcal{L}(\tilde{M}), \mathcal{L}(M^\star|X)\right) \leq \mathrm{I}' + \mathrm{II}'(X) + \mathrm{III}'(X),$$

where we define

$$\mathrm{I}' = d_{\mathrm{K}}\left(\mathcal{L}(\tilde{M}), \mathcal{L}(\tilde{M}_{\mathbf{m}})\right),$$
$$\mathrm{II}'(X) = d_{\mathrm{K}}\left(\mathcal{L}(\tilde{M}_{\mathbf{m}}), \mathcal{L}(M^\star_{\mathbf{m}}|X)\right),$$
$$\mathrm{III}'(X) = d_{\mathrm{K}}\left(\mathcal{L}(M^\star_{\mathbf{m}}|X), \mathcal{L}(M^\star|X)\right).$$

The first term is equal to III in the proof of Theorem **??** and requires no further treatment. The second term is addressed in Proposition B.2.

To derive the bound for $\mathrm{III}'(X)$, we partially reuse the proof of Proposition A.2. For any real numbers $t_1' \leq t_2'$, the following bound holds

$$\mathrm{III}'(X) \leq \mathbb{P}(A'(t_2')|X) + \mathbb{P}(B'(t_1')|X),$$

where we define the following events for any $t \in \mathbb{R}$,

$$A'(t) = \left\{ \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} (\lambda_{k,l} S^\star_{k,j}/\hat{\sigma}^\tau_{k,l,j} - \lambda_{l,k} S^\star_{l,j}/\hat{\sigma}^\tau_{k,l,j}) \le t \right\},$$
$$B'(t) = \left\{ \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}^c_{k,l}(m_k,m_l)} (\lambda_{k,l} S^\star_{k,j}/\hat{\sigma}^\tau_{k,l,j} - \lambda_{l,k} S^\star_{l,j}/\hat{\sigma}^\tau_{k,l,j}) > t \right\}.$$

Lemma B.1 ensures that t'_1 and t'_2 can be chosen so that the random variables $\mathbb{P}(A'(t'_2)|X)$ and $\mathbb{P}(B'(t'_1)|X)$ are at most $cn^{-\frac{1}{2}+\delta}$ with probability at least $1 - cn^{-1}$. Under Assumption **??**, it can be checked that the choices of t'_1 and t'_2 given in Lemma B.1 satisfy $t'_1 \le t'_2$ when n (and hence all n_k) is sufficiently large. □

Lemma B.1. *Under the conditions of Theorem* **??**, *there are positive constants c'_1, c'_2, and c, not depending on n, for which the following statement is true: If t'_1 and t'_2 are chosen as*

$$t'_1 = c'_1 m_\circ^{-\kappa} \log^{3/2} n$$
$$t'_2 = c'_2 \ell_{\max}^{-\kappa} \sqrt{\log \ell_{\max}},$$

then the events

$$\mathbb{P}(A'(t'_2)|X) \le cn^{-\frac{1}{2}+\delta} \tag{S5}$$

and

$$\mathbb{P}(B'(t'_1)|X) \le n^{-1} \tag{S6}$$

each hold with probability at least $1 - cn^{-1}$.

Proof. By the triangle inequality and the definition of Kolmogorov distance,

$$\mathbb{P}(A'(t'_2)|X) \le \mathbb{P}\left(\max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} (\lambda_{k,l}\tilde{S}_{k,j} - \lambda_{l,k}\tilde{S}_{l,j})/\sigma^\tau_{k,l,j} \le t'_2 \right) + \mathrm{II}'(X).$$

Taking $t'_2 = t_2$ as in the proof of Proposition A.2, the proof of Lemma A.3 shows that the first term is of order $n^{-1/2+\delta}$. Proposition B.2 shows that the second term is bounded by $cn^{-\frac{1}{2}+\delta}$ with probability at least $1 - cn^{-1}$ for some constant $c > 0$ not depending on n. This establishes (S5).

To deal with (S6), we define the random variable

$$V^\star = \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}^c_{k,l}(m_k,m_l)} (\lambda_{k,l} S^\star_{k,j}/\hat{\sigma}^\tau_{k,l,j} - \lambda_{l,k} S^\star_{l,j}/\hat{\sigma}^\tau_{k,l,j}),$$

and let $q = \max\{2\kappa^{-1}, 3, \log n\}$. We shall construct a function $b(\cdot)$ such that the following bound holds for every realization of X,

$$(\mathbb{E}[|V^\star|^q \mid X])^{1/q} \le b(X),$$

and then Chebyshev's inequality gives the following inequality for any number b_n satisfying $b(X) \le b_n$,

$$\mathbb{P}(V^\star \ge e b_n \mid X) \le e^{-q} \le n^{-1}.$$

We will then find b_n so that the event $\{b(X) \le b_n\}$ holds with high probability. Finally, we will see that $t_1' \asymp b_n$.

To construct b, we adopt the same argument of the proof of Lemma B.1(b) of Lopes et al. (2020) and show that for any realization of X,

$$\mathbb{E}(|V^\star|^q \mid X) \le \sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \hat{\sigma}_{k,l,j}^{q(1-\tau)} \mathbb{E}(|\lambda_{k,l} S_{k,j}^\star / \hat{\sigma}_{k,l,j}^\tau - \lambda_{l,k} S_{l,j}^\star / \hat{\sigma}_{k,l,j}^\tau|^q \mid X).$$

By Lemma E.3, for every $j \in \{1,\dots,p\}$, the event

$$\mathbb{E}(|\lambda_{k,l} S_{k,j}^\star / \hat{\sigma}_{k,l,j}^\tau - \lambda_{l,k} S_{l,j}^\star / \hat{\sigma}_{k,l,j}^\tau|^q \mid X) \le (cq)^q$$

holds with probability 1. Consequently, if we set $s = q(1-\tau)$ and consider the random variable

$$\hat{\mathfrak{s}} = \left(\sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \hat{\sigma}_{k,l,j}^s \right)^{1/s},$$

as well as

$$b(X) = cq\hat{\mathfrak{s}}^{(1-\tau)},$$

we obtain the bound

$$[\mathbb{E}(|V^\star|^q \mid X)]^{1/q} \le b(X),$$

with probability 1. Now, Lemma E.2 implies that

$$\mathbb{P}\left(b(X) \ge q \frac{(c\sqrt{q})^{1-\tau}}{(q\kappa-1)^{1/q}} m_\circ^{-\kappa+1/q} (2N)^{1/q} \right) \le e^{-q} \le n^{-1}$$

for some constant $c > 0$ not depending on n. By weakening this tail bound slightly, it can be simplified to

$$\mathbb{P}\left(b(X) \ge C' q^{3/2} m_\circ^{-\kappa} \right) \le n^{-1},$$

where $C' = c m_\circ^{1/q} (q\kappa - 1)^{-1/q} (2N)^{1/q}$. Since $C' \lesssim 1$ and $(1-\tau)\sqrt{\log n} \gtrsim 1$ gives $q \asymp \log n$, it follows that there is a constant c_1' not depending on n, such that if $b_n = c_1' m_\circ^{-\kappa} \log^{3/2} n$, then $\mathbb{P}(b(X) \ge b_n) \le n^{-1}$, which completes the proof. □

Proposition B.2. *Under the conditions of Theorem ??, there is a constant $c > 0$, not depending on n, such* *that the event*

$$\mathrm{II}'(X) \le cn^{-\frac{1}{2}+\delta}$$

holds with probability at least $1 - cn^{-1}$.

Proof. Define the random variable

$$\breve{M}^{\star}_{\mathbf{m}} = \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} (\lambda_{k,l} S^{\star}_{k,j} - \lambda_{l,k} S^{\star}_{l,j})/\sigma^{\tau}_{k,l,j} \tag{S7}$$

and consider the triangle inequality

$$\mathrm{II}'(X) \le d_{\mathrm{K}}\left(\mathcal{L}(\tilde{M}_{\mathbf{m}}), \mathcal{L}(\breve{M}^{\star}_{\mathbf{m}}|X)\right) + d_{\mathrm{K}}\left(\mathcal{L}(\breve{M}^{\star}_{\mathbf{m}}|X), \mathcal{L}(M^{\star}_{\mathbf{m}}|X)\right). \tag{S8}$$

Addressing the first term of (S8). Let S be the vector obtained by stacking column vectors $\lambda_{k,l} S^{\star}_{k,j} - \lambda_{l,k} S^{\star}_{l,j}$ for $(k,l) = (k_1,l_1),\ldots,(k_N,l_N)$. As in the proof of Proposition A.1, $\breve{M}^{\star}_{\mathbf{m}}$ can be expressed as coordinate-wise maximum of $\Pi_{\mathbf{m}} R^{\top}\zeta$ with $\zeta \sim N(0,\breve{\mathfrak{S}})$, where $\Pi_{\mathbf{m}}$ denotes the projection matrix onto the superindices $\mathcal{I} = \{(k,l,j) : k,l \in \mathcal{P}, j \in \mathcal{J}_{k,l}(m_k,m_l)\}$, R is a matrix, and

$$\breve{\mathfrak{S}} = \begin{pmatrix} \Pi\hat{\Sigma}_1\Pi^{\top} & & & \\ & \Pi\hat{\Sigma}_2\Pi^{\top} & & \\ & & \ddots & \\ & & & \Pi\hat{\Sigma}_K\Pi^{\top} \end{pmatrix}$$

with Π being defined in the proof of Proposition A.1. Similarly, $\tilde{M}_{\mathbf{m}}$ can be expressed as coordinate-wise maximum of $\Pi_{\mathbf{m}} R^{\top}\xi$, where $\xi \sim N(0,\mathfrak{S})$ with

$$\mathfrak{S} = \begin{pmatrix} \Pi\Sigma_1\Pi^{\top} & & & \\ & \Pi\Sigma_2\Pi^{\top} & & \\ & & \ddots & \\ & & & \Pi\Sigma_K\Pi^{\top} \end{pmatrix}.$$

For $\mathfrak{C}_k^{\top} = \Pi\Sigma_k^{1/2}$ consider the singular value decomposition

$$\mathfrak{C}_k = U_k \Lambda_k V_k^{\top},$$

where $r_k \lesssim J \equiv |\mathcal{I}|$ denotes the rank of $\mathfrak{C}_k$. We may assume that $U_k \in \mathbb{R}^{p\times r_k}$ has orthonormal columns, $\Lambda_k \in \mathbb{R}^{r_k\times r_k}$ to be invertible, and $V_k^{\top}$ to have orthonormal rows. Define

$$W_k = n_k^{-1}\sum_{i=1}^{n_k}(Z_{k,i} - \bar{Z}_k)(Z_{k,i} - \bar{Z}_k)^{\top},$$

where $\bar{Z}_k = n_k^{-1}\sum_{i=1}^{n_k} Z_{k,i}$, and

$$W = \begin{pmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_K \end{pmatrix}.$$

Then $\mathfrak{S} = \mathfrak{C}^\top\mathfrak{C}$ and $\breve{\mathfrak{S}} = \mathfrak{C}^\top W\mathfrak{C}$ with

$$\mathfrak{C} = \begin{pmatrix} \mathfrak{C}_1 & & & \\ & \mathfrak{C}_2 & & \\ & & \ddots & \\ & & & \mathfrak{C}_K \end{pmatrix}.$$

Define r_k-dimensional vectors $\tilde{\xi}_k = V_k^\top\xi_k$ and $\tilde{\zeta}_k = V_k^\top\zeta_k$, where ξ_k and ζ_k are respectively the subvectors of ξ and ζ corresponding to the kth sample. It can be shown that the columns of $\Pi\hat{\Sigma}_k\Pi^\top$ and $\Pi\Sigma_k\Pi^\top$ span the same subspace of $\mathbb{R}^J$ with probability at least $1-cn_k^{-2}$ (due to Lemma D.5 of Lopes et al. (2020) and noting that the probability bound there can be strengthened to $1-cn^{-2}$). Therefore, the event $E = \{$the columns of $\mathfrak{S}$ and $\hat{\mathfrak{S}}$ span the same subspace$\}$ holds with probability at least $1-c\sum_{k=1}^K n_k^{-2} \geq 1-cn^{-1}$, and furthermore, conditionally on E, the random vector ξ lies in the column-span of V, where

$$V = \begin{pmatrix} V_1 & & & \\ & V_2 & & \\ & & \ddots & \\ & & & V_K \end{pmatrix},$$

since $\breve{\mathfrak{S}} = V\Lambda(U^\top WU)\Lambda V^\top$with

$$U = \begin{pmatrix} U_1 & & & \\ & U_2 & & \\ & & \ddots & \\ & & & U_K \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} \Lambda_1 & & & \\ & \Lambda_2 & & \\ & & \ddots & \\ & & & \Lambda_K \end{pmatrix}.$$

The argument below is conditional on the event E.

Given E, the random vector ξ lies in the column-span of V almost surely, which means $V\tilde{\xi} = \xi$ almost surely. The same argument applies to ζ and $\tilde{\zeta}$. It follows that for any $t\in\mathbb{R}$, the events $\{\tilde{M}_{\mathbf{m}} \leq t\}$ and $\{\breve{M}^\star_{\mathbf{m}} \leq t\}$ can be expressed as $\{\tilde{\xi}\in\mathcal{A}_t\}$ and $\{\tilde{\zeta}\in\mathcal{A}_t\}$, respectively, for a convex set $\mathcal{A}_t$. Hence $d_{\mathrm{K}}\left(\mathcal{L}(\tilde{M}_{\mathbf{m}}), \mathcal{L}(\breve{M}^\star_{\mathbf{m}}|X)\right)$ is upper-bounded by the total variation distance between $\mathcal{L}(\tilde{\xi})$ and $\mathcal{L}(\tilde{\zeta})$, and in turn, Pinsker's inequality implies that this is upper-bounded by $c\sqrt{d_{\mathrm{KL}}(\mathcal{L}(\tilde{\zeta}),\mathcal{L}(\tilde{\xi}))}$, where $c>0$ is an absolute constant, and d_{KL} denotes the KL divergence. Since the random vectors $\tilde{\xi}\sim N(0, V^\top\mathfrak{S}V)$ and $\tilde{\zeta}\sim N(0,V^\top\breve{\mathfrak{S}}V)$ are Gaussian (conditional on X), the following exact formula is available if we let $H = (V^\top\mathfrak{S}V)^{1/2}$ (so that $H^\top H = V^\top\mathfrak{S}V$)

and $\tilde{C} = H^{-\top}(V^\top \breve{\mathfrak{S}} V)H^{-1} - I_r$,

$$\begin{aligned} d_{\mathrm{KL}}(\mathcal{L}(\tilde{\zeta}), \mathcal{L}(\tilde{\xi})) &= \frac{1}{2}\{\mathrm{tr}(\tilde{C}) - \log\det(\tilde{C} + I_r)\} \\ &= \frac{1}{2}\sum_{j=1}^{r}\{\theta_j(\tilde{C}) - \log(\theta_j(\tilde{C}) + 1)\}, \end{aligned}$$

where $r = \sum_{k=1}^{K} r_k \le KJ$ and $\theta_j(\tilde{C})$ denotes the eigenvalues of $\tilde{C}$. Note that $\|\tilde{C}\|_{\mathrm{op}} \le cKn^{-1/2}J\log n_{\max}$ by utilizing Lemma D.5 of Lopes et al. (2020) and the diagonal block structure of $\tilde{C}$. Using the inequality $|x - \log(x+1)| \le x^2/(1+x)$ that holds for any $x \in (-1, \infty)$, as well as the condition $|\theta_j(\tilde{C})| \le \|\tilde{C}\|_{\mathrm{op}} \le cKn^{-1/2}J\log n_{\max} \le 1/2$ for sufficiently large n, we have

$$d_{\mathrm{KL}}(\mathcal{L}(\tilde{\zeta}), \mathcal{L}(\tilde{\xi})) \le cr\|\tilde{C}\|_{\mathrm{op}}^2 \le cKJ\left(Kn^{-1/2}J\log n_{\max}\right)^2,$$

for some absolute constant $c > 0$. Thus,

$$d_{\mathrm{K}}\left(\mathcal{L}(\tilde{M}_{\mathbf{m}}), \mathcal{L}(\breve{M}^\star_{\mathbf{m}}|X)\right) \le cJ^{3/2}K^{3/2}n^{-1/2}\log n_{\max}$$

with probability at least $1 - cn^{-1}$. With $J \le Km_{\max}$ and observing

$$cJ^{3/2}K^{3/2}n^{-1/2}\log n_{\max} \lesssim K^3 m_{\max}^{3/2} n^{-1/2}\log n_{\max} \lesssim n^{-\frac{1}{2}+\delta},$$

the first term of (S8) is bounded by $cn^{-\frac{1}{2}+\delta}$ with probability at least $1 - cn^{-1}$.

Addressing the second term of (S8). We proceed by considering the general inequality

$$d_{\mathrm{K}}(\mathcal{L}(\xi), \mathcal{L}(\zeta)) \le \sup_{t\in\mathbb{R}} \mathbb{P}(|\zeta - t| \le \varepsilon) + \mathbb{P}(|\xi - \zeta| > \varepsilon),$$

which holds for any random variables ξ and ζ, and any real number $\varepsilon > 0$. We will let $\mathcal{L}(\breve{M}^\star_{\mathbf{m}}|X)$ play the role of $\mathcal{L}(\xi)$, and $\mathcal{L}(M^\star_{\mathbf{m}}|X)$ play the role of $\mathcal{L}(\zeta)$. Thus we need to establish an anti-concentration inequality for $\mathcal{L}(M^\star_{\mathbf{m}}|X)$, as well as a coupling inequality for $M^\star_{\mathbf{m}}$ and $\breve{M}^\star_{\mathbf{m}}$, conditionally on X.

For the coupling inequality, we put

$$\varepsilon = cn^{-1/2}\log^{5/2} n_{\max}$$

for a suitable constant $c > 0$ not depending on n. Then Lemma E.6 shows that the event

$$\mathbb{P}\left(|\breve{M}^\star_{\mathbf{m}} - M^\star_{\mathbf{m}}| > \varepsilon \mid X\right) \le cn^{-1}$$

holds with probability at least $1 - cn^{-1}$.

For the anti-concentration inequality, we use Nazarov's inequality (Lemma G.2, Lopes et al., 2020). Let

$$\underline{\hat{\sigma}}_{\mathbf{m}} = \min_{(k,l)\in\mathcal{P}} \min_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \hat{\sigma}_{k,l,j}.$$

Then Nazarov's inequality implies that the event

$$\sup_{t\in\mathbb{R}} \mathbb{P}\left(|M^\star_{\mathbf{m}} - t| \le \varepsilon | X\right) \le c\varepsilon \underline{\hat{\sigma}}_{\mathbf{m}}^{\tau-1}\sqrt{\log m} \le c\varepsilon \underline{\hat{\sigma}}_{\mathbf{m}}^{\tau-1}\sqrt{\log(2Nm_{\max})}$$

holds with probability 1, where $m = \sum_{(k,l)\in\mathcal{P}} m_{k,l} \le 2Nm_{\max}$. Meanwhile, we observe that

$$\begin{aligned}\sigma_{k,l,j} &= \sqrt{\lambda_{k,l}^2\sigma_{k,j}^2 + \lambda_{l,k}^2\sigma_{l,j}^2} \ge \max\{\lambda_{k,l}\sigma_{k,j}, \lambda_{l,k}\sigma_{l,j}\} \\ &\ge c_2 \max\{\sigma_{k,j}, \sigma_{l,j}\} \ge c_2 c_\circ \max\{m_k^{-\alpha}, m_l^{-\alpha}\} \ge c m_{\max}^{-\alpha}\end{aligned} \tag{S9}$$

for all $(k,l)\in\mathcal{P}$ and $j\in\mathcal{J}_{k,l}(m_k,m_l)$. Then, Lemma E.4 and Assumption **??** imply that the event

$$\underline{\hat{\sigma}}_{\mathbf{m}}^{\tau-1} \le c m_{\max}^{\kappa}$$

holds with probability at least $1 - Nn^{-2} \ge 1 - cn^{-1}$. Given the above, we conclude that

$$\sup_{t\in\mathbb{R}} \mathbb{P}\left(|M^\star_{\mathbf{m}} - t| \le \varepsilon \mid X\right) \le c m_{\max}^{\kappa}\sqrt{\log(2Nm_{\max})}n^{-1/2}\log^{5/2} n_{\max} \le cn^{-1/2+\delta}$$

holds with probability at least $1 - cn^{-1}$, which completes the proof. □

C Proof of Theorem ??

Define

$$\hat{M}_{\mathbf{m}} = \max_{(k,l)\in\mathcal{P}} \hat{M}_{m_{k,l}}(k,l), \tag{S10}$$

Proof. We first observe that

$$d_{\mathrm{K}}(\mathcal{L}(\hat{M}), \mathcal{L}(M)) \le \mathrm{I}'' + \mathrm{II}'' + \mathrm{III}'',$$

where

$$\begin{aligned}\mathrm{I}'' &= d_{\mathrm{K}}\left(\mathcal{L}(\hat{M}), \mathcal{L}(\hat{M}_{\mathbf{m}})\right), \\ \mathrm{II}'' &= d_{\mathrm{K}}\left(\mathcal{L}(\hat{M}_{\mathbf{m}}), \mathcal{L}(M_{\mathbf{m}})\right), \\ \mathrm{III}'' &= d_{\mathrm{K}}\left(\mathcal{L}(M_{\mathbf{m}}), \mathcal{L}(M)\right).\end{aligned}$$

The last term III″ requires no further consideration, as it is equal to I in the proof of Theorem **??**. The second term is handled in Proposition C.1, while the first term is handled in Proposition C.2. □

Proposition C.1. *Let δ be as in Theorem* **??**. *Under Assumptions* **??**–**??**, *one has* $\mathrm{II}'' \lesssim n^{-\frac{1}{2}+\delta}$.

Proof. We again proceed by considering the general inequality

$$d_{\mathrm{K}}(\mathcal{L}(\xi), \mathcal{L}(\zeta)) \le \sup_{t\in\mathbb{R}} \mathbb{P}(|\zeta - t| \le \varepsilon) + \mathbb{P}(|\xi - \zeta| > \varepsilon),$$

which holds for any random variables ξ and ζ, and any real number $\varepsilon > 0$. We will let $\mathcal{L}(\hat{M}_{\mathbf{m}})$ play the role of $\mathcal{L}(\xi)$, and let $\mathcal{L}(M_{\mathbf{m}})$ play the role of $\mathcal{L}(\zeta)$. As before, we then need to establish an anti-concerntration inequality for $\mathcal{L}(M_{\mathbf{m}})$, as well as a coupling inequality for $\hat{M}_{\mathbf{m}}$ and $M_{\mathbf{m}}$.

For the coupling inequality, we put

$$\varepsilon = cn^{-1/2} \log^{5/2} n_{\max}$$

for a suitable constant c not depending on n. Then Lemma E.7 shows that

$$\mathbb{P}\left(|\hat{M}_{\mathbf{m}} - M_{\mathbf{m}}| > \varepsilon\right) \lesssim n^{-1}.$$

For the anti-concentration inequality, we utilize $d_{\mathrm{K}}(\mathcal{L}(M_{\mathbf{m}}), \mathcal{L}(\tilde{M}_{\mathbf{m}})) \lesssim n^{-1/2+\delta}$, which was established in the proof of Theorem **??**, whence

$$\begin{aligned}\sup_{t\in\mathbb{R}} \mathbb{P}(|M_{\mathbf{m}} - t| \le \varepsilon) &= \sup_{t\in\mathbb{R}} \{\mathbb{P}(M_{\mathbf{m}} \le t + \varepsilon) - \mathbb{P}(M_{\mathbf{m}} \le t - \varepsilon)\} \\ &= \sup_{t\in\mathbb{R}} \{\mathbb{P}(\tilde{M}_{\mathbf{m}} \le t + \varepsilon) - \mathbb{P}(\tilde{M}_{\mathbf{m}} \le t - \varepsilon)\} + cn^{-1/2+\delta}.\end{aligned}$$

Let

$$\underline{\sigma}_{\mathbf{m}} = \min_{(k,l)\in\mathcal{P}} \min_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \sigma_{k,l,j}.$$

Then Nazarov's inequality implies that

$$\sup_{t\in\mathbb{R}} \mathbb{P}\left(|\tilde{M}_{\mathbf{m}} - t| \le \varepsilon\right) \lesssim \varepsilon \underline{\sigma}_{\mathbf{m}}^{\tau-1} \sqrt{\log m} \lesssim \varepsilon \underline{\sigma}_{\mathbf{m}}^{\tau-1} \sqrt{\log(2Nm_{\max})} \lesssim \varepsilon m_{\max}^{\alpha(1-\tau)} \sqrt{\log(2Nm_{\max})},$$

where $m = \sum_{(k,l)\in\mathcal{P}} m_{k,l}$, and the last inequality is due to (S9). Given the above, we conclude that

$$\sup_{t\in\mathbb{R}} \mathbb{P}\left(|\tilde{M}_{\mathbf{m}} - t| \le \varepsilon\right) \le cn^{-1/2} (\log^{5/2} n_{\max}) m_{\max}^{\alpha(1-\tau)} \sqrt{\log(2Nm_{\max})} \le cn^{-1/2+\delta}.$$

This completes the proof. □

Proposition C.2. *Under the conditions of Theorem* **??**, *one has* $\mathrm{I}'' \lesssim n^{-\frac{1}{2}+\delta}$.

Proof. Define

$$A''(t) = \left\{ \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} (\lambda_{k,l}S_{k,j}/\hat{\sigma}^{\tau}_{k,l,j} - \lambda_{l,k}S_{l,j}/\hat{\sigma}^{\tau}_{k,l,j}) \le t \right\},$$
$$B''(t) = \left\{ \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}^c_{k,l}(m_k,m_l)} (\lambda_{k,l}S_{k,j}/\hat{\sigma}^{\tau}_{k,l,j} - \lambda_{l,k}S_{l,j}/\hat{\sigma}^{\tau}_{k,l,j}) > t \right\},$$

where $\mathcal{J}^c_{k,l}(m_k,m_l)$ denotes the complement of $\mathcal{J}_{k,l}(m_k,m_l)$ in $\{1,\dots,p\}$. Also, if $t''_1 \le t''_2$, it is seen that

$$A''(t) \cap B''(t) \subset A''(t''_2) \cup B''(t''_1)$$

for all $t\in\mathbb{R}$. By a union bound, we have

$$\mathrm{I}'' \le \mathbb{P}(A''(t''_2)) + \mathbb{P}(B''(t''_1)).$$

Setting

$$t''_1 = cm_\circ^{-\kappa}\log n$$
$$t''_2 = c_\circ \ell_{\max}^{-\kappa}\sqrt{\log \ell_{\max}}$$

for a constant $c>0$, we proceed to show that $\mathbb{P}(A''(t''_2))$ and $\mathbb{P}(B''(t''_1))$ are bounded by $cn^{-1/2+\delta}$. We note the inequality $t''_1 \le t''_2$ holds for all large n, due to the definitions of $\ell_{\max}$, $m_\circ$, and κ, as well as the condition $(1-\tau)\sqrt{\log n} \gtrsim 1$. Specifically we will establish that

$$\mathbb{P}(A''(t''_2)) \lesssim n^{-\frac{1}{2}+\delta}, \tag{S11}$$

and

$$\mathbb{P}(B''(t''_1)) \lesssim n^{-1}. \tag{S12}$$

According to Propositions C.1 and A.1, we have

$$\begin{aligned}
\mathbb{P}(A''(t''_2)) &\le \mathbb{P}\left(\max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} (\lambda_{k,l}S_{k,j}/\sigma^{\tau}_{k,l,j} - \lambda_{l,k}S_{l,j}/\sigma^{\tau}_{k,l,j}) \le t''_2 \right) + \mathrm{II}'' \\
&\le \mathbb{P}\left(\max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} (\lambda_{k,l}\tilde{S}_{k,j}/\sigma^{\tau}_{k,l,j} - \lambda_{l,k}\tilde{S}_{l,j}/\sigma^{\tau}_{k,l,j}) \le t''_2 \right) + \mathrm{II} + \mathrm{II}'' \\
&\le \mathbb{P}\left(\max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} (\lambda_{k,l}\tilde{S}_{k,j}/\sigma^{\tau}_{k,l,j} - \lambda_{l,k}\tilde{S}_{l,j}/\sigma^{\tau}_{k,l,j}) \le t''_2 \right) + cn^{-\frac{1}{2}+\delta}.
\end{aligned}$$

Then (S11) follows from a similar argument as given in the proof of Lemma A.3.

To derive (S12), consider

$$U = \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}^c_{k,l}(m_k,m_l)} \frac{\lambda_{k,l}S_{k,j} - \lambda_{l,k}S_{l,j}}{\hat{\sigma}^{\tau}_{k,l,j}}.$$

For $q = \max\{2\kappa^{-1}, 3, \log n\}$, we first observe that

$$\|U\|_q^q \le \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \mathbb{E}\left|\frac{\lambda_{k,l}S_{k,j} - \lambda_{l,k}S_{l,j}}{\hat{\sigma}_{k,l,j}^\tau}\right|^q \le \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \|V_{k,l,j}\|_{2q}^q \|Y_{k,l,j}\|_{2q}^q$$

with

$$V_{k,l,j} = \left|\frac{\sigma_{k,l,j}^\tau}{\hat{\sigma}_{k,l,j}^\tau}\right|,$$
$$Y_{k,l,j} = \left|\frac{\lambda_{k,l}S_{k,j} - \lambda_{l,k}S_{l,j}}{\sigma_{k,l,j}^\tau}\right|.$$

By Lemma E.10, we further have

$$\begin{aligned}
\|U\|_q^q &\le c^q \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \|Y_{k,l,j}\|_{2q}^q \\
&\le c^q \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \left(\sigma_{k,l,j}^{2q(1-\tau)} \mathbb{E}\left|\frac{\lambda_{k,l}S_{k,j} - \lambda_{l,k}S_{l,j}}{\sigma_{k,l,j}}\right|^{2q}\right)^{1/2} \\
&\le c^q \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \sigma_{k,l,j}^{q(1-\tau)} \left(\mathbb{E}\left|\frac{\lambda_{k,l}S_{k,j} - \lambda_{l,k}S_{l,j}}{\sigma_{k,l,j}}\right|^{2q}\right)^{1/2} \\
&\le (cq)^q \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \sigma_{k,l,j}^{q(1-\tau)} \\
&\le (cq)^q c_1^{q(1-\tau)} N \frac{m_\circ^{-q\kappa+1}}{q\kappa - 1},
\end{aligned}$$

where the last inequality is due to (S4). If we put $C = \frac{c}{(q\kappa-1)^{1/q}} m_\circ^{1/q} N^{1/q} \lesssim 1$, then

$$\|U\|_q \le Cq m_\circ^{-\kappa}.$$

Since $q \asymp \log n$, we have

$$\mathbb{P}\left(U \ge ec m_\circ^{-\kappa} \log n\right) \le e^{-q} \le \frac{1}{n},$$

as needed. ☐

Remark. If Assumption **??** is replaced with the condition $n^{-1/2}\log^3 p \ll 1$, then (S12) can be established in the following way. With the same notations in the proof of Proposition C.2, we first observe that

$$U \le \left(\max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \left|\frac{\sigma_{k,l,j}^\tau}{\hat{\sigma}_{k,l,j}^\tau}\right|\right) V$$

with

$$V = \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \left|\frac{\lambda_{k,l}S_{k,j} - \lambda_{l,k}S_{l,j}}{\sigma_{k,l,j}^\tau}\right|.$$

Under the condition $n^{-1/2}\log^3 p \ll 1$,

$$\max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}^c_{k,l}(m_k,m_l)} \left|\frac{\sigma^{\mathcal{T}}_{k,l,j}}{\hat{\sigma}^{\mathcal{T}}_{k,l,j}}\right| \asymp 1$$

with probability at least $1 - cNn^{-2}$, according to Lemma E.8. With the aid of Lemma E.3, the term V then can be handled by an argument similar to the proof of Lemma A.3.

Remark. The above proofs relied on the condition $p > m_\circ$ and this implies $m_\circ \geq n^{\log^{-a} n}$ and $\ell_\circ \geq \log^3 n$. These conditions are used in the analysis of I and III, as well as I′, III′(X), I″ and III″. If $p \leq m_\circ$, then the definition of $m_\circ$ implies that $p = m_1 = \cdots = m_K$, and the quantities I, III, I′, III′(X), I″ and III″ become exactly 0. In this case, the proofs of Theorems **??**, **??** and **??** reduce to bounding II, II′(X) and II″, and these arguments can be repeated as before.

D Proof of Theorem ??

Proof. Part **??** is handled in Proposition D.1. Below we establish part **??**.

Let q_{M^*} be the quantile of the conditional probability distribution $\mathcal{L}(M^*|X)$. By Theorem **??** and **??**, the event $E = \{d_{\mathrm{K}}(\mathcal{L}(M), \mathcal{L}(M^\star|X)) \leq ca_n\}$ holds with probability at least $1 - cn^{-1}$, where $a_n = n^{-1/2+\delta}$. Below we condition on the event E and observe that $q_M(\varrho - ca_n) \leq q_{M^\star}(\varrho) \leq q_M(\varrho + ca_n)$ conditional on E.

In the derivation of Proposition A.2, with the notation there, we have

$$\begin{aligned}\mathbb{P}(M \leq t) = \mathbb{P}\left(A(t) \cap B^c(t)\right) &= P(A(t)) - \mathbb{P}\left(A(t) \cap B(t)\right)\\ &\geq \mathbb{P}\left(A(t)\right) - \mathbb{P}\left(A(t_2)\right) - \mathbb{P}\left(B(t_1)\right)\\ &\geq \mathbb{P}\left(A(t)\right) - cn^{-1/2+\delta}.\end{aligned}$$

By an argument similar to Lemma A.3, one can show that if

$$V = \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \frac{\lambda_{k,l}S_{k,j} - \lambda_{l,k}S_{l,j}}{\sigma^{\mathcal{T}}_{k,l,j}},$$

then $\|V\|_Q^Q \leq (cQ)^Q N$ if we define $Q = \max\{2\kappa^{-1}, 3, \sqrt{\log n}\}$. Thus,

$$\mathbb{P}\left(A(t)\right) = 1 - \mathbb{P}(V > t) \geq 1 - \frac{\|V\|_Q^Q}{t^Q} \geq 1 - e^{-Q} \to 1$$

if $t \geq e\|V\|_Q$. Therefore, $q_M(\varrho + ca_n) \lesssim \|V\|_Q \lesssim \sqrt{\log n}$, otherwise $\mathbb{P}(M \leq t) \to 1 > \varrho$. Similar arguments show that $q_M(\varrho - ca_n) \lesssim \sqrt{\log n}$.

The above shows that $|q_{M^\star}(\varrho)| \leq c\sqrt{\log n}$ with probability at least $1 - cn^{-1}$. Since conditional on the data X the random variable $M^\star$ is the maximum of a multivariate Gaussian distribution, according to

Nadarajah et al. (2019), the conditional distribution of $M^\star$ has a probability density and thus its cumulative distribution function is continuous. This enables us to apply Theorem 2 of Xia (2019), with an appropriately chosen $t \asymp \sqrt{\log n}$ in that theorem, to conclude that $|\hat{q}_M(\varrho) - q_{M^\star}(\varrho)| \le c\sqrt{\log n}$ with probability at least $1 - e^{-2h^2 B} \ge 1 - cn^{-1}$ for all sufficiently large n when the event E holds, where $h > 0$ is a constant not depending on n. To complete the proof, we should verify that the quantities $|\varphi_F^+(t)|$ and $|\varphi_F^-(t)|$ defined in that theorem, with the chosen $t \asymp \sqrt{\log n}$, are bounded from below by $h > 0$ with probability at least $1 - cn^{-1}$ for all sufficiently large n, where F is the cumulative distribution function of $M^\star$ conditional on the data X. For this, set $h = \min\{(1-\varrho)/4, \varrho/4\} > 0$. Then replacing ϱ with $\varrho + 2h$ in the inequality $|q_{M^\star}(\varrho)| \le c\sqrt{\log n}$, we observe that $|q_{M^\star}(\varrho + 2h)| \le c\sqrt{\log n}$ with probability at least $1 - cn^{-1}$. This means that the event $|q_{M^\star}(\varrho+2h) - q_{M^\star}(\varrho)| \le c\sqrt{\log n}$ occurs with probability at least $1 - cn^{-1}$, and when this event holds we are able to find an appropriate $t \asymp \sqrt{\log n}$ such that $\varphi_F^+(t) = \lfloor (1-\varrho)n \rfloor / n + F(q_{M^\star}(\varrho) + t) - 1 = \lfloor (1-\varrho)n \rfloor / n + \varrho + 2h - 1 \ge h$ for all sufficiently large n. The claim for $|\varphi_F^-(t)|$ can be established in a similar fashion. □

Proposition D.1. *Under Assumptions ??–??, for some constant $c > 0$ not depending on n, one has*

$$\mathbb{P}\left(\max_{(k,l)\in\mathcal{P}} \max_{1\le j\le p} \hat{\sigma}^2_{k,l,j} < 2\sigma^2_{\max} \right) \ge 1 - cn^{-1},$$

where $\sigma_{\max} = \max\{\sigma_{k,j} : 1 \le j \le p, 1 \le k \le K\}$.

Proof. Define

$$A^\circ(t) = \left\{ \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(n,n)} \hat{\sigma}^2_{k,l,j} > t \right\},$$
$$B^\circ(t) = \left\{ \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}^c_{k,l}(n,n)} \hat{\sigma}^2_{k,l,j} > t \right\},$$

where as before $\mathcal{J}^c_{k,l}(n,n)$ denotes the complement of $\mathcal{J}_{k,l}(n,n)$ in $\{1,\ldots,p\}$. With $t^\circ = 2\sigma^2_{\max}$ we will establish that

$$\mathbb{P}(A^\circ(t^\circ)) \lesssim n^{-1}, \tag{S13}$$

and when $\mathcal{J}^c_{k,l}(n,n) \ne \varnothing$ for some (k,l) that

$$\mathbb{P}(B^\circ(t^\circ)) \lesssim n^{-1}. \tag{S14}$$

For (S13), we first observe that

$$\mathbb{P}(\hat{\sigma}^2_{k,l,j} > t^\circ) \le \mathbb{P}(|\hat{\sigma}^2_{k,l,j} - \sigma^2_{k,l,j}| > t^\circ - \sigma^2_{\max}).$$

With the above inequality, by using Lemma E.5 and a union bound, we conclude that

$$\mathbb{P}\left(A^{\circ}(t^{\circ})\right) \le \sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}_{k,l}(n,n)} \mathbb{P}(|\hat{\sigma}^2_{k,l,j} - \sigma^2_{k,l,j}| > t^{\circ} - \sigma^2_{\max})$$
$$\le cNn \cdot n^{-3} \lesssim n^{-1}.$$

To derive (S14), consider

$$U = \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}^c_{k,l}(m_k,m_l)} \hat{\sigma}^2_{k,l,j}.$$

For $q = \max\{\alpha^{-1}, 3, \log n\}$, we first observe that

$$\|U\|_q^q \le \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}^c_{k,l}(n,n)} \mathbb{E}\left|\hat{\sigma}^2_{k,l,j}\right|^q$$

By Lemma E.1, we further have

$$\begin{aligned}
\|U\|_q^q &\le \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}^c_{k,l}(n,n)} \|\hat{\sigma}_{k,l,j}\|_{2q}^{2q} \\
&\le \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}^c_{k,l}(n,n)} (c\sigma_{k,l,j}\sqrt{2q})^{2q} \\
&\le c(cq)^q \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}^c_{k,l}(n,n)} \sigma^{2q}_{k,l,j} \\
&\le c(cq)^q N n^{-2q\alpha+1},
\end{aligned}$$

where the last inequality is derived in analogy to (S4), and this implies

$$\|U\|_q \lesssim q n^{-2\alpha+1/q} N^{1/q} \ll \sigma^2_{\max}.$$

Since $q \asymp \log n$, we have

$$\mathbb{P}(U > 2\sigma^2_{\max}) \le \mathbb{P}\left(U \ge e\|U\|_q\right) \le e^{-q} \le \frac{1}{n}$$

for all sufficiently large n. □

E Technical Lemmas

Lemma E.1. *Suppose the conditions of Theorem* **??** *hold. For any fixed* $b > 0$*, if* $3 \le q \le \max\{2\kappa^{-1}, 3, \log^b n\}$*, there exists a constant* $c > 0$ *not depending on* q*,* K*,* N*,* p *or* $n_1, \dots, n_K$*, such that for any* $k, l \in \{1, \dots, K\}$ *and* $j \in \{1, \dots, p\}$*, we have* $\|\hat{\sigma}_{k,j}\|_q \le c\sigma_{k,j}\sqrt{q}$ *and* $\|\hat{\sigma}_{k,l,j}\|_q \le c\sigma_{k,l,j}\sqrt{q}$.

Proof. According to Lemma D.1 of Lopes et al. (2020) (which still holds when $q = \log^b n \ge 3$), we have $\|\hat{\sigma}_{k,j}\|_q \le c\sigma_{k,j}\sqrt{q}$. Therefore, due to $\hat{\sigma}_{k,l,j} = \sqrt{\lambda^2_{k,l}\hat{\sigma}^2_{k,j} + \lambda^2_{l,k}\hat{\sigma}^2_{l,j}} \le \lambda_{k,l}\hat{\sigma}_{k,j} + \lambda_{l,k}\hat{\sigma}_{l,j}$, and using the fact that

$\|Y\|_q^2 = \|Y^2\|_{q/2}$ for any random variable Y, we deduce that

$$\begin{aligned}
\|\hat{\sigma}_{k,l,j}\|_q^2 &= \|\hat{\sigma}_{k,l,j}^2\|_{q/2} = \|\lambda_{k,l}^2 \hat{\sigma}_{k,j}^2 + \lambda_{l,k}^2 \hat{\sigma}_{l,j}^2\|_{q/2} \\
&\le \lambda_{k,l}^2 \|\hat{\sigma}_{k,j}^2\|_{q/2} + \lambda_{l,k}^2 \|\hat{\sigma}_{l,j}^2\|_{q/2} = \lambda_{k,l}^2 \|\hat{\sigma}_{k,j}\|_q^2 + \lambda_{l,k}^2 \|\hat{\sigma}_{k,j}\|_q^2 \\
&\le c^2 q(\lambda_{k,l}^2 \sigma_{k,j}^2 + \lambda_{l,k}^2 \sigma_{l,j}^2) = c^2 q \sigma_{k,l,j}^2.
\end{aligned}$$

□

Lemma E.2. *Let* $q = \max\{2\kappa^{-1}, 3, \log n\}$ *and* $s = q(1-\tau)$*. Consider the random variables* $\hat{\mathfrak{s}}$ *and* $\hat{\mathfrak{t}}$ *defined by*

$$\hat{\mathfrak{s}} = \left(\sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}_{k,l}^c(m_k,m_l)} \hat{\sigma}_{k,l,j}^s \right)^{1/s}$$

and

$$\hat{\mathfrak{t}} = \left(\sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \hat{\sigma}_{k,l,j}^s \right)^{1/s}.$$

Under the conditions of Theorem **??**, *there is a constant* $c > 0$, *not depending on* q, K, N, p *or* $n_1, \dots, n_K$, *such that*

$$\mathbb{P}\left(\hat{\mathfrak{s}} \ge \frac{c\sqrt{q}}{(q\kappa - 1)^{1/s}} m_\circ^{-\alpha+1/s} (2N)^{1/s} \right) \le e^{-q} \tag{S15}$$

and

$$\mathbb{P}\left(\hat{\mathfrak{t}} \ge \frac{c\sqrt{q}}{(q\kappa - 1)^{1/s}} (2N)^{1/s} \right) \le e^{-q}. \tag{S16}$$

Proof. Using Lemma E.1, this lemma follows from similar arguments as in the proof of Lemma D.2 in Lopes et al. (2020). For further details, consider

$$\begin{aligned}
\|\hat{\mathfrak{s}}\|_q &= \left\| \sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}^c(m_k,m_l)} \hat{\sigma}_{k,l,j}^s \right\|_{q/s}^{1/s} \le \left(\sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}^c(m_k,m_l)} \|\hat{\sigma}_{k,l,j}^s\|_{q/s} \right)^{1/s} \\
&= \left(\sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}^c(m_k,m_l)} \|\hat{\sigma}_{k,l,j}\|_q^s \right)^{1/s} \le c\sqrt{q} \left(\sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}^c(m_k,m_l)} \sigma_{k,l,j}^s \right)^{1/s} \\
&\le c\sqrt{q} \left(\sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}^c(m_k,m_l)} \max\{\sigma_{k,j}, \sigma_{l,j}\}^s \right)^{1/s} \le c\sqrt{q} \left(\sum_{(k,l)\in\mathcal{P}} \sum_{j\in\mathcal{J}^c(m_k,m_l)} (\sigma_{k,j}^s + \sigma_{l,j}^s) \right)^{1/s} \\
&\le c\sqrt{q} \left(\sum_{(k,l)\in\mathcal{P}} \left\{ \int_{m_k}^p x^{-s\alpha} \mathrm{d}x + \int_{m_l}^p x^{-s\alpha} \mathrm{d}x \right\} \right)^{1/s} \le c\sqrt{q} \left(2N \int_{m_\circ}^p x^{-s\alpha} \mathrm{d}x \right)^{1/s} \\
&\le c\sqrt{q} (2N)^{1/s} \frac{m_\circ^{-\alpha+1/s}}{(s\alpha - 1)^{1/s}},
\end{aligned}$$

where for the last step, we use $s\alpha = q\kappa > 1$. The proof for $\hat{\mathfrak{t}}$ can be obtained by the same argument, except that the bound becomes $\sum_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \sigma_{k,j}^s \lesssim 1$. □

Lemma E.3. *Suppose the conditions of Theorem* **??** *hold, and for any fixed $b > 0$, let $q = \max\{2\kappa^{-1}, \log^b n, 3\}$. Then for a constant $c > 0$, not depending on q, K, N, p or $n_1, \ldots, n_K$, such that for any $(k,l) \in \mathcal{P}$ and $j \in \{1, \ldots, p\}$, it holds that*

$$\left\| \frac{\lambda_{k,l} S_{k,j}}{\sigma_{k,l,j}} - \frac{\lambda_{l,k} S_{l,j}}{\sigma_{k,l,j}} \right\|_q \le cq, \tag{S17}$$

and the following event holds with probability 1,

$$\left(\mathbb{E}\left[\left| \frac{\lambda_{k,l} S^\star_{k,j}}{\hat{\sigma}_{k,l,j}} - \frac{\lambda_{l,k} S^\star_{l,j}}{\hat{\sigma}_{k,l,j}} \right|^q \mid X \right] \right)^{1/q} \le cq. \tag{S18}$$

Proof. Without loss of generality, let $(k,l) = (1,2)$, and set $\lambda_1 = \lambda_{k,l}$, $\lambda_2 = \lambda_{l,k}$, and $\sigma_j = \sigma_{k,l,j}$. We reuse the notation k for some index from $\{1,2\}$, i.e., $k \in \{1,2\}$ in what follows.

Since $q > 2$, by Minkowski's inequality and Lemma G.4 of Lopes et al. (2020), we have

$$\begin{aligned}
\|\lambda_1 S_{1,j}/\sigma_j - \lambda_2 S_{2,j}/\sigma_j\|_q \le & \|\lambda_1 S_{1,j}/\sigma_j\|_q + \|\lambda_2 S_{2,j}/\sigma_j\|_q \\
\le & q \max\{\|\lambda_1 S_{1,j}/\sigma_j\|_2, \lambda_1 n_1^{-1/2+1/q} \|(X_{1,1,j} - \mu_{1,j})/\sigma_j\|_q\} \\
& + q \max\{\|\lambda_2 S_{2,j}/\sigma_j\|_2, \lambda_2 n_2^{-1/2+1/q} \|(X_{2,1,j} - \mu_{1,j})/\sigma_j\|_q\},
\end{aligned}$$

and furthermore

$$\|S_{k,j}\|_2^2 = \operatorname{var}(S_{k,j}) = \sigma_{k,j}^2.$$

Thus $\|\lambda_k S_{k,j}/\sigma_j\|_2 = \lambda_k \sigma_{k,j} \sigma_j^{-1} \le 1$, where we note that $\lambda_k^2 \sigma_{k,j}^2 \sigma_j^{-2} = \lambda_k^2 \sigma_{k,j}^2 / (\lambda_1^2 \sigma_{1,j}^2 + \lambda_2^2 \sigma_{2,j}^2) \le 1$. Also, if we define the vector $u_k = \sigma_{k,j}^{-1} \Sigma_k^{1/2} e_j$ in $\mathbb{R}^p$ for standard basis $e_1, \ldots, e_p$ in $\mathbb{R}^p$, which satisfies $\|u\|_2 = 1$, then

$$\lambda_k \|(X_{k,1,j} - \mu_{k,j})/\sigma_j\|_q = \lambda_k \sigma_{k,j} \sigma_j^{-1} \|(X_{k,1,j} - \mu_{k,j})/\sigma_{k,j}\|_q \le \|Z_{k,1}^\top u\|_q \lesssim q,$$

proving (S17). Inequality (S18) follows from the same argument, conditioning on X. □

Define the correlation

$$\rho_{k,l,j,j'} = \frac{\Sigma_{k,l}(j,j')}{\sigma_{k,l,j}\sigma_{k,l,j'}},$$

and its sample version

$$\hat{\rho}_{k,l,j,j'} = \frac{\widehat{\Sigma}_{k,l}(j,j')}{\hat{\sigma}_{k,l,j}\hat{\sigma}_{k,l,j'}},$$

for any $j, j' \in \{1, \ldots, p\}$.

Lemma E.4. *Under Assumption* **??** *and* **??**, *there is a constant $c > 0$, not depending on n, such that the following events*

$$\max_{j \in \mathcal{J}_{k,l}(m_k, m_l)} \left| \frac{\hat{\sigma}_{k,l,j}}{\sigma_{k,l,j}} - 1 \right| \le c a_n,$$

$$\min_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \hat{\sigma}_{k,l,j}^{1-\tau} \geq \left(\min_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \sigma_{k,l,j}^{1-\tau}\right)(1-ca_n),$$

and

$$\max_{j,j'\in j\in\mathcal{J}_{k,l}(m_k,m_l)} \left|\hat{\rho}_{j,j'} - \rho_{j,j'}\right| \leq ca_n$$

each hold with probability at least $1 - cn^{-2}$*, where* $a_n = n^{-1/2}\log n_{\max}$.

Proof. These conclusions are direct consequences of Lemma E.5. ☐

Lemma E.5. *Suppose Assumptions* **??** *and* **??** *hold, and fix any* $1 \leq k < l \leq K$ *and any two (possibly equal) indices* $j, j' \in \{1, \ldots, p\}$*. Then, for any number* $\vartheta \geq 1$*, there are positive constants* c *and* $c_1(\vartheta)$*, not depending on* n*, such that the event*

$$\left|\frac{\widehat{\Sigma}_{k,l}(j,j')}{\sigma_{k,l,j}\sigma_{k,l,j'}} - \rho_{k,l,j,j'}\right| \leq c_1(\vartheta)n^{-1/2}\log n_{\max}$$

holds with probability at least $1 - cn^{-\vartheta}$.

Proof. It is equivalent to showing that

$$\left|\widehat{\Sigma}_{k,l}(j,j') - \Sigma_{k,l}(j,j')\right| \leq c_1(\vartheta)n^{-1/2}(\log n_{\max})\sigma_{k,l,j}\sigma_{k,l,j'}.$$

Furthermore

$$\begin{aligned}
\left|\widehat{\Sigma}_{k,l}(j,j') - \Sigma_{k,l}(j,j')\right| &= \left|\lambda_{k,l}^2\widehat{\Sigma}_k(j,j') - \lambda_{k,l}^2\Sigma_k(j,j') + \lambda_l^2\widehat{\Sigma}_l(j,j') - \lambda_{l,k}^2\Sigma_l(j,j')\right| \\
&\leq \lambda_{k,l}^2\left|\widehat{\Sigma}_k(j,j') - \Sigma_k(j,j')\right| + \lambda_{l,k}^2\left|\widehat{\Sigma}_l(j,j') - \Sigma_l(j,j')\right| \\
&\leq c_1(\vartheta)(n_k^{-1/2}\lambda_{k,l}^2\sigma_{k,j}\sigma_{k,j'}\log n_k + n_l^{-1/2}\lambda_{l,k}^2\sigma_{l,j}\sigma_{l,j'}\log n_l) \\
&\leq c_1(\vartheta)(\log n_{\max})n^{-1/2}(\lambda_{k,l}^2\sigma_{k,j}\sigma_{k,j'} + \lambda_{l,k}^2\sigma_{l,j}\sigma_{l,j'}),
\end{aligned}$$

with probability at least $1 - cn_k^{-\vartheta} - cn_l^{-\vartheta} \geq 1 - 2cn^{-\vartheta}$, where the second inequality is due to Lemma D.7 of Lopes et al. (2020). Now, by the Cauchy–Schwarz inequality,

$$2\sigma_{k,j}\sigma_{k,j'}\sigma_{l,j}\sigma_{l,j'} \leq \sigma_{k,j}^2\sigma_{l,j'}^2 + \sigma_{l,j}^2\sigma_{k,j'}^2,$$

and further

$$\begin{aligned}
\lambda_{k,l}^2\sigma_{k,j}\sigma_{k,j'} + \lambda_{l,k}^2\sigma_{l,j}\sigma_{l,j'} &= \sqrt{(\lambda_{k,l}^2\sigma_{k,j}\sigma_{k,j'} + \lambda_{l,k}^2\sigma_{l,j}\sigma_{l,j'})^2} \\
&\leq \sqrt{(\lambda_{k,l}^2\sigma_{k,j}^2 + \lambda_{l,k}^2\sigma_{l,j}^2)}\sqrt{(\lambda_{k,l}^2\sigma_{k,j'}^2 + \lambda_{l,k}^2\sigma_{l,j'}^2)} \\
&= \sigma_{k,l,j}\sigma_{k,l,j'},
\end{aligned}$$

which completes the proof. ☐

Remark. In the above proof, we note that Lemma D.7 of Lopes et al. (2020) does not depend on Assumption 2 of Lopes et al. (2020).

Lemma E.6. *Under the conditions of Theorem ??, there is a constant $c > 0$, not depending on n, such that*

$$\mathbb{P}\left(|\breve{M}^\star_{\mathbf{m}} - M^\star_{\mathbf{m}}| > r_n|X\right) \le cn^{-1}$$

holds with probability at least $1 - cn^{-1}$, where $\breve{M}^\star_{\mathbf{m}}$ is defined in (S7) and $r_n = cn^{-1/2}\log^{5/2} n_{\max}$.

Proof. Using a similar argument as in the proof of Lemma D.8 of Lopes et al. (2020), we find that

$$|\breve{M}^\star_{\mathbf{m}} - M^\star_{\mathbf{m}}| \le \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \left|\left(\frac{\hat{\sigma}_{k,l,j}}{\sigma_{k,l,j}}\right)^\tau - 1\right| \cdot \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \left|\frac{S^\star_{k,l,j}}{\hat{\sigma}^\tau_{k,l,j}}\right|.$$

It follows from Lemma E.4 that the event

$$\max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \left|\left(\frac{\hat{\sigma}_{k,l,j}}{\sigma_{k,l,j}}\right)^\tau - 1\right| \le cn^{-1/2}\log n_{\max}$$

holds with probability at least $1 - cNn^{-2} \ge 1 - cn^{-1}$. Now consider

$$U^\star = \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \left|\frac{S^\star_{k,l,j}}{\hat{\sigma}^\tau_{k,l,j}}\right|.$$

Showing that

$$\mathbb{P}\left(U^\star \ge c\log^{3/2} n_{\max} \mid X\right) \le cn^{-1}$$

holds with probability at least $1 - cn^{-1}$ will complete the proof.

Using Chebyshev's inequality with $q = \{2\kappa^{-1}, 3, \log n\}$ gives

$$\mathbb{P}\left(U^\star \ge e[\mathbb{E}(|U^\star|^q \mid X)]^{1/q} \mid X\right) \le e^{-q}.$$

Now it suffices to show that the event

$$[\mathbb{E}(|U^\star|^q \mid X)]^{1/q} \le c\log^{3/2} n_{\max}$$

holds with probability at least $1 - cn^{-1}$. This is done by repeating the argument in Lemma B.1 with the aid of (S16) from Lemma E.2. □

Lemma E.7. *Under the conditions of Theorem ??, for some constant $c > 0$, not depending on n, we have*

$$\mathbb{P}\left(|\hat{M}_{\mathbf{m}} - M_{\mathbf{m}}| > r_n\right) \le cn^{-1},$$

where $\hat{M}_{\mathbf{m}}$ is defined in (S10) and $r_n = cn^{-1/2}\log^{5/2} n_{\max}$.

Proof. A similar argument as in the proof of Lemma D.8 of Lopes et al. (2020) leads to

$$|\hat{M}_{\mathbf{m}} - M_{\mathbf{m}}| \le \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \left| \left(\frac{\sigma_{k,l,j}}{\hat{\sigma}_{k,l,j}}\right)^{\tau} - 1 \right| \cdot \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \left| \frac{\lambda_{k,l}S_{k,j} - \lambda_{l,k}S_{l,j}}{\sigma^{\tau}_{k,l,j}} \right|.$$

It follows from Lemma E.4 that the event

$$\max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \left| \left(\frac{\sigma_{k,l,j}}{\hat{\sigma}_{k,l,j}}\right)^{\tau} - 1 \right| \le cn^{-1/2} \log n_{\max}$$

holds with probability at least $1 - cNn^{-2} \ge 1 - cn^{-1}$. Now consider

$$U = \max_{(k,l)\in\mathcal{P}} \max_{j\in\mathcal{J}_{k,l}(m_k,m_l)} \left| \frac{\lambda_{k,l}S_{k,j} - \lambda_{l,k}S_{l,j}}{\sigma^{\tau}_{k,l,j}} \right|.$$

Then

$$\mathbb{P}\left(U \ge c \log^{3/2} n_{\max}\right) \le cn^{-1}$$

will complete the proof.

Using Chebyshev's inequality with $q = \max\{2\kappa^{-1}, 3, \log n\}$ gives

$$\mathbb{P}\left(U \ge e(\mathbb{E}|U|^q)^{1/q}\right) \le e^{-q}.$$

Now it suffices to show that

$$\|U\|_q = (\mathbb{E}|U|^q)^{1/q} \lesssim \log^{3/2} n_{\max}.$$

Observe that

$$\|U\|_q^q \le \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}_{k,l}(m_k,m_l)} \sigma^{q(1-\tau)}_{k,l,j} \mathbb{E}|\sigma^{-1}_{k,l,j}(\lambda_{k,l}S_{k,j} - \lambda_{l,k}S_{l,j})|^q.$$

By Lemma E.3, and noting that $q\alpha(1-\tau) = q\kappa \ge 2$, we further have

$$\|U\|_q^q \le (cq)^q \sum_{(k,l)\in\mathcal{P}, j\in\mathcal{J}_{k,l}(m_k,m_l)} \sigma^{q(1-\tau)}_{k,l,j} \lesssim N(cq)^q,$$

or equivalently,

$$\|U\|_q \lesssim qN^{1/q} \lesssim \log^{3/2} n_{\max},$$

where we use the fact that $N^{1/q} \lesssim 1$ given the choice of q. □

Define the correlation

$$\rho_{k,j,j'} = \frac{\Sigma_k(j,j')}{\sigma_{k,j}\sigma_{k,j'}},$$

and its sample version

$$\hat{\rho}_{k,j,j'} = \frac{\widehat{\Sigma}_k(j,j')}{\hat{\sigma}_{k,j}\hat{\sigma}_{k,j'}},$$

for any $j, j' \in \{1, \dots, p\}$.

Lemma E.8. *Under Assumption* **??** *and* **??**, *if* $n^{-1/2}\log^3 p \ll 1$, *then for any number* $\theta \geq 2$, *there are positive constants* c *and* c_θ, *not depending on* n, *such that the event*

$$\sup_{1\leq k\leq K}\sup_{1\leq j,j'\leq p}\left|\frac{\widehat{\Sigma}_k(j,j')}{\sigma_{k,j}\sigma_{k,j'}} - \rho_{k,j,j'}\right| \leq c_\theta(\log n_{\max} + \log^3 p)n^{-1/2}$$

holds with probability at least $1 - cKn^{-\theta}$.

Proof. It suffices to show that

$$\sup_{1\leq j,j'\leq p}\left|\frac{\widehat{\Sigma}_k(j,j')}{\sigma_{k,j}\sigma_{k,j'}} - \rho_{k,j,j'}\right| \leq \frac{c_\theta(\log n_k + \log^3 p)}{\sqrt{n_k}}$$

with probability at least $1 - cn_k^{\theta}$. Consider ℓ_2-unit vectors $u = \Sigma_k^{1/2}e_j\sigma_{k,j}^{-1}$ and $v = \Sigma_k^{1/2}e_{j'}\sigma_{k,j'}^{-1}$ in $\mathbb{R}^p$. Define

$$W_k = n_k^{-1}\sum_{i=1}^{n_k}(Z_{k,i} - \bar{Z}_k)(Z_{k,i} - \bar{Z}_k)^\top,$$

where $\bar{Z}_k = \sum_{i=1}^{n_k} Z_{k,i}$. Observe that

$$\frac{\widehat{\Sigma}_k(j,j')}{\sigma_{k,j}\sigma_{k,j'}} - \rho_{k,j,j'} = u^\top(W_k - I_p)v. \tag{S19}$$

For each $1 \leq i \leq n_k$, define the random variable $\zeta_{i,u} = Z_{k,i}^\top u$ and $\zeta_{i,v} = Z_{k,i}^\top v$. In this notation, the relation (S19) becomes

$$\frac{\widehat{\Sigma}_k(j,j')}{\sigma_{k,j}\sigma_{k,j'}} - \rho_{k,j,j'} = \Delta(u,v) + \Delta'(u,v)$$

where

$$\Delta(u,v) = \frac{1}{n_k}\sum_{i=1}^{n_k}\zeta_{i,u}\zeta_{i,v} - u^\top v,$$
$$\Delta'(u,v) = \left(\frac{1}{n_k}\sum_{i=1}^{n_k}\zeta_{i,u}\right)\left(\frac{1}{n_k}\sum_{i=1}^{n_k}\zeta_{i,v}\right).$$

Note that $\mathbb{E}(\zeta_{i,u}\zeta_{i,v}) = u^\top v$. Also, if we let $q = \max\{\theta(\log n_k + \log^3 p), 3\}$, then

$$\|\zeta_{i,u}\zeta_{i,v} - u^\top v\|_q \leq 1 + \|\zeta_{i,u}\zeta_{i,v}\|_q \leq 1 + \|\zeta_{i,u}\|_{2q}\|\zeta_{i,v}\|_{2q} \leq cq^2,$$

where the second inequality is due to the Cauchy–Schwarz inequality, and the third to Assumption **??**. The constant c, although it varies from place to place, does not depend on n_k or p. Then, Lemma G.4 of Lopes et al. (2020) gives the following bound for $q > 2$,

$$\begin{aligned}
\|\Delta(u,v)\|_q &\le cq \max\left\{\|\Delta(u,v)\|_2, n_k^{-1}\left(\sum_{i=1}^{n_k}\|\zeta_{i,u}\zeta_{i,v} - u^\top v\|_q^q\right)^{1/q}\right\} \\
&\le cq \max\{n_k^{-1/2}, n_k^{-1+1/q}q^2\} \\
&\le c(\log n_k + \log^3 p)n_k^{-1/2}.
\end{aligned}$$

By the Chebyshev inequality

$$\mathbb{P}\left(|\Delta(u,v)| \ge e\|\Delta(u,v)\|_q\right) \le e^{-q},$$

whence

$$\mathbb{P}\left(|\Delta(u,v)| \ge \frac{c\theta(\log n_k + \log^3 p)}{\sqrt{n_k}}\right) \le \frac{1}{n^\theta p^\theta}.$$

Similar arguments apply to $\Delta'(u,v)$. Thus,

$$\mathbb{P}\left(\left|\frac{\widehat{\Sigma}_k(j,j')}{\sigma_{k,j}\sigma_{k,j'}} - \rho_{k,j,j'}\right| \ge \frac{c_\theta(\log n_k + \log^3 p)}{\sqrt{n_k}}\right) \le \frac{1}{n^\theta p^\theta}$$

and furthermore by a union bound

$$\begin{aligned}
&\mathbb{P}\left(\sup_{1\le j,j'\le p}\left|\frac{\widehat{\Sigma}_k(j,j')}{\sigma_{k,j}\sigma_{k,j'}} - \rho_{k,j,j'}\right| \ge \frac{c_\theta(\log n_k + \log^3 p)}{\sqrt{n_k}}\right) \\
&\le \sum_{1\le j,j'\le p}\frac{1}{n_k^\theta p^\theta} = \frac{1}{n_k^\theta}\frac{p^2}{p^\theta} \le \frac{1}{n_k^\theta}.
\end{aligned}$$

□

Observing that $\sigma_{k,l,j} = \sqrt{\lambda_{k,l}^2\sigma_{k,j}^2 + \lambda_{l,k}^2\sigma_{l,j}^2}$, one obtains the following corollary.

Corollary E.9. *Under Assumption* **??** *and* **??**, *if* $n^{-1/2}\log^3 p \ll 1$, *for any number* $\theta \ge 2$, *there are positive constants* c *and* c_θ, *not depending on* n, *such that the event*

$$\sup_{(k,l)\in\mathcal{P}}\sup_{1\le j\le p}\left|\frac{\hat{\sigma}_{k,l,j}}{\sigma_{k,l,j}} - 1\right| \le c_\theta(\log n_{\max} + \log^3 p)n^{-1/2}$$

holds with probability at least $1 - cNn^{-\theta}$.

Lemma E.10. *Suppose Assumptions* **??**–**??** *hold. Then, for any fixed* $\theta \in (0,\infty)$ *and* $Q \asymp \log n$, *for some constant* c, *not depending on* n, *one has*

$$\sup_{1\le k\le K, 1\le j\le p}\left\|\frac{\sigma_{k,j}^\theta}{\hat{\sigma}_{k,j}^\theta}\right\|_Q \le c.$$

Proof. Below we suppress the subscripts from $\hat{\sigma}_{k,j}$, μ_k and n_k. Also, the constant c might change its value from place to place and depend on θ. In addition, observing that

$$\frac{\sigma^2}{\hat{\sigma}^2} = \frac{1}{n^{-1}\sum_{i=1}^n[\{(X_i-\mu)-(\overline{X}-\mu)\}/\sigma]^2} = \frac{1}{n^{-1}\sum_{i=1}^n(Y_i-\overline{Y})^2}$$

with $Y_i = (X_i-\mu)/\sigma$ and $\overline{Y} = \sum_{i=1}^n Y_i$, without loss of generality, we assume $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$.

Let $\omega = Q\theta \asymp \log n$ and $c_1 = 1/2$. We first observe that

$$\begin{aligned}\mathbb{E}\hat{\sigma}^{-\omega} &= \int_0^\infty \mathbb{P}(\hat{\sigma}^{-\omega} > t)\mathrm{d}t = \int_0^\infty \mathbb{P}(\hat{\sigma}^2 < t^{-2/\omega})\mathrm{d}t \\ &= \int_0^{c_1^{-\omega/2}} \mathbb{P}(\hat{\sigma}^2 < t^{-2/\omega})\mathrm{d}t + \int_{c_1^{-\omega/2}}^{n^\omega} \mathbb{P}(\hat{\sigma}^2 < t^{-2/\omega})\mathrm{d}t + \int_{n^\omega}^\infty \mathbb{P}(\hat{\sigma}^2 < t^{-2/\omega})\mathrm{d}t.\end{aligned}$$

For the last term, we have

$$\begin{aligned}\mathbb{P}(\hat{\sigma}^2 < t^{-2/\omega}) &= \mathbb{P}\left(n^{-1}\sum_{i=1}^n(X_i-\overline{X})^2 \le t^{-2/\omega}\right) \\ &\le \mathbb{P}\left(\forall\, 1\le i\le n: (X_i-\overline{X})^2 \le nt^{-2/\omega}\right) \\ &\le \mathbb{P}\left(\forall\, 1\le i\le n: |X_i-\overline{X}| \le \sqrt{n}t^{-1/\omega}\right) \\ &\le \mathbb{P}\left(\forall\, 1\le i\le n-1: |X_i-X_n| \le 2\sqrt{n}t^{-1/\omega}\right) \\ &= \mathbb{E}\mathbb{P}\left(\forall\, 1\le i\le n-1: |X_i-X_n| \le 2\sqrt{n}t^{-1/\omega} \mid X_n\right) \\ &= \mathbb{E}\{\mathbb{P}\left(|X_1-X_n| \le 2\sqrt{n}t^{-1/\omega} \mid X_n\right)\}^{n-1} \\ &\le (c\sqrt{n}t^{-1/\omega})^{(n-1)\nu}\end{aligned}$$

for some universal constant $c > 0$ and for all sufficiently large n, where the last inequality is due to Assumption **??**, and the last equality is due to the conditional independence of the random variables $|X_1-X_n|,\ldots,|X_{n-1}-X_n|$ given X_n and that these variables have identical conditional distributions. Therefore,

$$\int_{n^\omega}^\infty \mathbb{P}(\hat{\sigma}^2 < t^{-2/\omega})\mathrm{d}t \le (-(n-1)\nu/\omega+1)^{-1}c^{\nu(n-1)}n^{(n-1)\nu/2}t^{-(n-1)\nu/\omega+1}\Big|_{n^\omega}^\infty \asymp \nu c^{\nu(n-1)}n^{\omega-\frac{(n-1)\nu}{2}} \ll 1.$$

When $t \ge c_1^{-\omega/2}$ or equivalently $t^{-2/\omega} \le 1/2$, noting that $\sigma^2 = 1$ as we have assumed standardized X, one has

$$\begin{aligned}\mathbb{P}(\hat{\sigma}^2 - 1 < t^{-2/\omega} - 1) &\le \mathbb{P}(\hat{\sigma}^2 - 1 < -1/2) \\ &\le \mathbb{P}(|\hat{\sigma}^2 - 1| \ge 1/2) \\ &\le \mathbb{P}(|\hat{\sigma}^2 - 1| \ge 2n^{-1/2}\omega\log n) \\ &\le cn^{-2\omega},\end{aligned}$$

where the last inequality is obtained by an argument identical to that in the proof of Lemma D.7 of Lopes et al. (2020), except that the number $q = \max\{\kappa\log(n), 3\}$ there is replaced by $q = \max\{2\omega\log n, 3\}$. This implies that

$$\int_{c_1^{-\omega/2}}^{n^\omega} \mathbb{P}(\hat{\sigma}^2 < t^{-2/\omega})\mathrm{d}t \le n^\omega \cdot cn^{-2\omega} = cn^{-\omega} \ll 1.$$

Note that when $t \le c_1^{-\omega/2}$, we have the trivial bound $\mathbb{P}(\hat{\sigma}^2 < t) \le 1$. Therefore,

$$\mathbb{E}\hat{\sigma}^{-\omega} \le c_1^{-\omega/2} + cn^{-\omega} + \nu c^{\nu(n-1)} n^{\omega - \frac{(n-1)\nu}{2}} \le cc_1^{-\omega/2} = c2^{\omega/2},$$

or $\|\hat{\sigma}^{-\theta}\|_Q \le c$. □

F Additional Simulation Studies on Functional ANOVA

As observed by Zhang et al. (2019), the level of within-function correlation impacts the power of the test. Following a suggestion of a reviewer, we assessed the effect of the within-function correlation on the proposed method by using the simulation setup of Zhang et al. (2019). Specifically, we set $\mu_k(t) = (1 + 2.3t + 3.4t^2 + 1.5t^3) + \theta(k-1)(1+2t+3t^2+4t^3)/\sqrt{30}$ for $k = 1, 2, 3$ and $t \in [0, 1]$, and $X_k(t) = \mu_k(t) + \sum_{j=1}^{11} \sqrt{1.5}\rho^{j/2}\xi_{kj}\phi_j(t)$, where $\phi_1, \phi_2, \ldots$ are Fourier basis functions defined in Section **??**. Two cases are considered for the random variables ξ_{kj}, namely, the Gaussian case $\xi_{kj} \overset{iid}{\sim} N(0, 1)$ and the non-Gaussian case $\xi_{kj} \overset{iid}{\sim} t_4/\sqrt{2}$, where t_4 denotes Student's t distribution with 4 degrees of freedom. As the number m of design points has little impact on the results (Zhang et al., 2019), we fix $m = 100$ as in our previous simulation setting. As in Zhang et al. (2019), we consider $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$, where small values of ρ correspond to strong within-function correlation and large values signify weak correlation.

The results in Table S1 and Figure S1 show that our approach outperforms the other methods substantially. This may be partially explained by the fact that the basis functions $\phi_1, \phi_2, \ldots$ used to generate data in Zhang et al. (2019) coincide with the basis functions we use for projection in the proposed method. In light of this, we have provided an additional comparison in a more challenging setting, where the data are generated using a modified version of these basis functions (while our method still uses the original basis functions). Specifically, the modified basis functions, denoted by $\tilde{\phi}_1, \tilde{\phi}_2, \ldots$, are constructed in the following way: We first define $\tilde{\phi}_j(t) = \phi_j(0.8t+0.1)$ for $t \in [0, 1]$ and $j = 1, 2, \ldots, 11$, and then apply the Gram–Schmidt procedure to orthonormalize $\tilde{\phi}_1, \tilde{\phi}_2, \ldots$ within $L^2([0, 1])$. The results for simulation studies with these modified basis functions are shown in Table S2 and Figure S2, where the average value of the selected τ across all settings is 0.843 ± 0.153 and 0.824 ± 0.143 for Gaussian and non-Gaussian cases, respectively. We observe that the empirical sizes of most methods are close to the nominal level and the RP method tends to be conservative when ρ is small. When the within-function correlation is strong, e.g., when $\rho = 0.1, 0.3$, the power of the proposed method is considerably larger than that of MPF and GET, which in turn is much larger than the power of the other methods except RP; the power of RP is close to that of the proposed method when $\rho \ge 0.3$ but substantially less when $\rho = 0.1$. When the within-function correlation becomes weaker, e.g., when $\rho = 0.7, 0.9$, all tests tend to have similar power. This shows that the proposed method is

Table S1: Empirical size of functional ANOVA in the simulation setting of Zhang et al. (2019)

	ρ	n	proposed	L2	F	GPF	MPF	GET	RP
Gaussian	0.1	50,50,50	.044	.044	.044	.044	.038	.039	.011
		30,50,70	.054	.066	.060	.062	.066	.040	.017
	0.3	50,50,50	.042	.060	.050	.058	.060	.042	.021
		30,50,70	.046	.044	.040	.044	.048	.042	.024
	0.5	50,50,50	.058	.060	.056	.062	.052	.038	.021
		30,50,70	.052	.062	.062	.064	.058	.039	.037
	0.7	50,50,50	.034	.040	.038	.040	.036	.047	.041
		30,50,70	.060	.056	.050	.048	.060	.040	.049
	0.9	50,50,50	.042	.050	.048	.050	.054	.025	.044
		30,50,70	.044	.058	.054	.060	.032	.039	.051
non-Gaussian	0.1	50,50,50	.048	.060	.054	.056	.050	.030	.004
		30,50,70	.052	.054	.050	.052	.064	.041	.009
	0.3	50,50,50	.030	.046	.040	.044	.040	.034	.015
		30,50,70	.046	.056	.050	.054	.042	.047	.030
	0.5	50,50,50	.046	.076	.072	.080	.060	.032	.027
		30,50,70	.036	.036	.036	.036	.024	.037	.030
	0.7	50,50,50	.036	.042	.036	.044	.034	.035	.034
		30,50,70	.026	.042	.040	.038	.038	.030	.037
	0.9	50,50,50	.034	.064	.062	.058	.054	.036	.037
		30,50,70	.044	.060	.060	.054	.056	.030	.047

preferred for ANOVA for functional data in which the within-function correlation is typically strong.

For simulation studies in the above and in Section **??**, the tuning parameter τ is selected by the data-driven procedure described in Section **??**. Below we investigate the effectiveness of this procedure by comparing it with using a fixed value of τ. Specifically, in the simulation studies of Section **??**, we also compute the empirical size and power for the proposed method by using each of the values $0, 0.1, \ldots, 0.9, 0.99$ for τ. The results are presented in Table S3 and Figure S3, where the results for $\tau = 0, 0.2, 0.4, 0.6, 0.8, 0.99$ are provided. We observe that values close to 1 such as $\tau = 0.99$ yield inflated sizes for the resulting tests. For the family (M3), all values of τ lead to similar performance. Similar observations emerge for the family (M2) except $\tau = 0.99$ that yields much lower power. In families (M1) and (M4), the power increases as τ. In these families, although $\tau = 0.99$ gives rise to the largest power, its that the empirical size also noticeably deviates from the nominal level. In all cases, the power of the data-driven method is close to that of $\tau = 0.8$. Overall, the data-driven selection procedure selects a value for τ that produces competitive power while keeping the size close to the nominal level.

Table S2: Empirical size of functional ANOVA in the modified simulation setting of Zhang et al. (2019)

	ρ	n	proposed	L2	F	GPF	MPF	GET	RP
Gaussian	0.1	50,50,50	.053	.061	.055	.059	.049	.043	.017
		30,50,70	.048	.052	.050	.050	.038	.042	.010
	0.3	50,50,50	.055	.062	.057	.062	.063	.043	.020
		30,50,70	.061	.056	.054	.055	.060	.034	.021
	0.5	50,50,50	.055	.061	.047	.064	.039	.043	.037
		30,50,70	.056	.047	.044	.048	.049	.040	.033
	0.7	50,50,50	.053	.050	.047	.044	.047	.037	.035
		30,50,70	.052	.059	.058	.056	.050	.026	.051
	0.9	50,50,50	.057	.054	.053	.050	.053	.043	.050
		30,50,70	.051	.053	.052	.045	.044	.038	.057
non-Gaussian	0.1	50,50,50	.043	.048	.045	.046	.051	.036	.009
		30,50,70	.050	.049	.047	.047	.048	.057	.016
	0.3	50,50,50	.052	.050	.048	.047	.043	.041	.017
		30,50,70	.050	.067	.064	.063	.056	.042	.016
	0.5	50,50,50	.048	.048	.043	.042	.033	.031	.018
		30,50,70	.043	.045	.043	.055	.049	.038	.029
	0.7	50,50,50	.042	.053	.054	.046	.053	.032	.033
		30,50,70	.048	.053	.051	.054	.037	.048	.044
	0.9	50,50,50	.053	.054	.050	.048	.047	.030	.040
		30,50,70	.055	.036	.037	.039	.042	.031	.050

G Additional Simulation Studies on MANOVA

We complement the numerical studies in Section **??** by assessing the performance of the proposed method in a more standard MANOVA setting, where data are sampled from continuous distributions. We consider three groups that are represented by random vectors $X_1, X_2, X_3 \in \mathbb{R}^p$, such that each X_k follows an elliptical distribution (Fang et al., 1990) with mean μ_k and covariance Σ_k. Specifically, $X_k = \mu_k + U_k \Sigma_k^{1/2} Z_k$, where U_k is a random variable following the exponential distribution with mean 1, Z_k is a p-dimensional random vector following the standard multivariate normal distribution, and U_k is independent of Z_k. Note that each X_k is not Gaussian due to the random multiplicative factor U_k. We consider $p = 25$ and $p = 100$, and two scenarios of μ_k, namely,

- the sparse case, where $\mu_k(j) = 1 + j \sin(2\pi j/p) \exp(j/p)/p + a_p \theta(k-1)((p-j+1)/p)^4$, so that when $k \neq l$, the difference $\mu_k(j) - \mu_l(j)$ between two mean vectors decays as j increases, and
- the dense case, where $\mu_k(j) = 1 + j \sin(2\pi j/p) \exp(j/p)/p + a_p \theta(k-1)$, so that the difference $\mu_k(j) - \mu_l(j)$ remains constant across different coordinates j,

for $k = 1, 2, 3,\ j = 1, 2, \ldots, p$ and a constant a_p. In these settings, $\theta = 0$ corresponds to the null hypothesis,

Table S3: Empirical size of functional ANOVA with different values of τ

Covariance	M	n	τ						
			0	0.2	0.4	0.6	0.8	0.99	data-driven
common	M1	50,50,50	.053	.056	.058	.060	.061	.076	.051
		30,50,70	.045	.045	.045	.051	.055	.099	.053
	M2	50,50,50	.048	.047	.049	.050	.056	.082	.042
		30,50,70	.043	.044	.045	.046	.050	.094	.057
	M3	50,50,50	.044	.044	.047	.049	.049	.086	.057
		30,50,70	.073	.076	.074	.073	.074	.075	.056
	M4	50,50,50	.045	.047	.046	.046	.052	.073	.046
		30,50,70	.046	.045	.050	.050	.058	.084	.053
group-specific	M1	50,50,50	.056	.056	.058	.057	.060	.098	.055
		30,50,70	.049	.047	.049	.049	.058	.095	.043
	M2	50,50,50	.051	.050	.050	.049	.052	.095	.056
		30,50,70	.061	.061	.060	.062	.068	.108	.052
	M3	50,50,50	.042	.044	.044	.047	.052	.099	.051
		30,50,70	.050	.051	.051	.055	.064	.091	.049
	M4	50,50,50	.060	.058	.059	.056	.054	.095	.052
		30,50,70	.055	.055	.056	.054	.062	.114	.050

under which the mean vectors of all groups are equal. We set $a_p = 0.025$ in the sparse scenario, and in the dense scenario we set $a_p = 0.014$ if $p = 25$ and $a_p = 0.008$ if $p = 100$; these values are chosen in the way that the power is approximately 1 when $\theta = 1$. We set $\Sigma_k(j_1, j_2) = j_1^{-1/4} j_2^{-1/4} \mathcal{C}\big((j_1 - 1)/(p-1), (j_2 - 1)/(p-1)\big)$, where $\mathcal{C}(s,t)$ is the Matérn correlation function defined in (**??**) but with different values of parameters; here we set $\sigma^2 = 1$, $\eta = 5$ and $\nu = 0.1$.

Similar to the study in Section **??**, we compare the proposed method with the classic Lawley–Hotelling trace test (LH), the ridge-regularized Lawley–Hotelling trace test (RRLH) (Li et al., 2020), the procedure (S) of Schott (2007) and the data-adaptive ℓ_p-norm-based test (DALp) (Zhang et al., 2018). As mentioned in Section **??**, the method of Schott (2007) is favored for testing problems with a dense alternative, while the method of Zhang et al. (2018) has been reported to be powerful against different patterns of alternatives. The classic Lawley–Hotelling trace test is included as a baseline procedure. The empirical sizes in Table S4 show that those of the proposed method, Lawley–Hotelling trace test and Schott (2007) are close to the nominal level, while the size of Zhang et al. (2018) is inflated. This result is consistent with the observation made in Section **??**, except that the inflation in size of Zhang et al. (2018) seems more pronounced here, especially in the settings with a dense alternative. The empirical power functions shown in Figure S4 suggest that, in the sparse case, the proposed method consistently outperforms the others. In the dense setting, the proposed test has almost the same power of the test of Schott (2007), and substantially outperforms the Lawley–Hotelling trace test. With regard to the Zhang et al. (2018) test, it turns out that its type I error rate under the null is

Table S4: Empirical size of ANOVA on multivariate Laplace data

	p	(n_1, n_2, n_3)	proposed	S	DALp	LH	RRLH
sparse	25	50,50,50	.056	.048	.067	.049	.032
		30,50,70	.058	.049	.063	.046	.038
	100	50,50,50	.053	.052	.080	.032	.054
		30,50,70	.056	.046	.063	.029	.062
dense	25	50,50,50	.055	.059	.070	.036	.053
		30,50,70	.057	.054	.070	.040	.054
	100	50,50,50	.052	.045	.067	.026	.046
		30,50,70	.054	.045	.080	.039	.044

Table S5: Empirical size of ANOVA on multivariate Laplace data with different values of τ

	p	(n_1, n_2, n_3)	τ						
			0	0.2	0.4	0.6	0.8	0.99	data-driven
sparse	25	50,50,50	.055	.051	.047	.043	.048	.051	.056
		30,50,70	.032	.033	.042	.046	.056	.063	.058
	100	50,50,50	.038	.037	.039	.045	.051	.067	.053
		30,50,70	.048	.046	.049	.055	.061	.075	.056
dense	25	50,50,50	.039	.040	.048	.048	.055	.061	.055
		30,50,70	.048	.044	.042	.046	.049	.057	.057
	100	50,50,50	.047	.045	.045	.052	.058	.076	.052
		30,50,70	.042	.049	.048	.049	.056	.062	.054

substantially higher than the nominal level, which makes it difficult to compare power. We also observe that the classic Lawley–Hotelling trace test, which is not specifically designed for the high-dimensional setting, deteriorates considerably as the dimension becomes larger, e.g., when $p = 100$ in both settings, while the regularized Lawley–Hotelling trace test (Li et al., 2020) improves upon the classic unregularized version when the dimension is relatively large in both settings.

As in the case of functional ANOVA, below we investigate the effectiveness of the data-driven selection procedure for τ by comparing it with using a fixed value of τ. Specifically, in the above simulation studies, we also compute the empirical size and power for the proposed high-dimensional MANOVA by using each of the values $0, 0.1, \ldots, 0.9, 0.99$ for τ. The results are presented in Table S5 and Figure S5, where we report the results for $\tau = 0, 0.2, 0.4, 0.6, 0.8, 0.99$. We observe that values rather close to 1 such as $\tau = 0.99$ result in slightly inflated size. In the sparse setting, all values of τ and the data-driven procedure lead to similar performance, while in the dense setting, large values of τ yield larger power, and the power of the data-driven method is close to that of $\tau = 0.8$. Similar to the observations in the functional ANOVA case, overall the data-driven selection procedure is effective in selecting a near optimal value for τ.

References

Bentkus, V. (2005). A Lyapunov-type bound in R^d. *SIAM Journal on Theory of Probability & Its Applications*, 49(2):311–323.

Fang, K.-T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall.

Li, H., Aue, A., and Paul, D. (2020). High-dimensional general linear hypothesis tests via non-linear spectral shrinkage. *Bernoulli*, 26(4):2541–2571.

Lopes, M. E., Lin, Z., and Müller, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional data analysis. *The Annals of Statistics*, 48(2):1214–1229.

Nadarajah, S., Afuecheta, E., and Chan, S. (2019). On the distribution of maximum of multivariate normal random vectors. *Communications in Statistics - Theory and Methods*, 48(10):2425–2445.

Schott, J. R. (2007). Some high-dimensional tests for a one-way MANOVA. *Journal of Multivariate Analysis*, 98(9):1825–1839.

Xia, D. (2019). Non-asymptotic bounds for percentiles of independent non-identical random variables. *Statistics & Probability Letters*.

Zhang, J.-T., Cheng, M.-Y., Wu, H.-T., and Zhou, B. (2019). A new test for functional one-way ANOVA with applications to ischemic heart screening. *Computational Statistics & Data Analysis*, 132:3–17.

Zhang, M., Zhou, C., He, Y., , and Liu, B. (2018). Data-adaptive test for high-dimensional multivariate analysis of variance problem. *Australian & New Zealand Journal of Statistics*, 60(4):447–470.

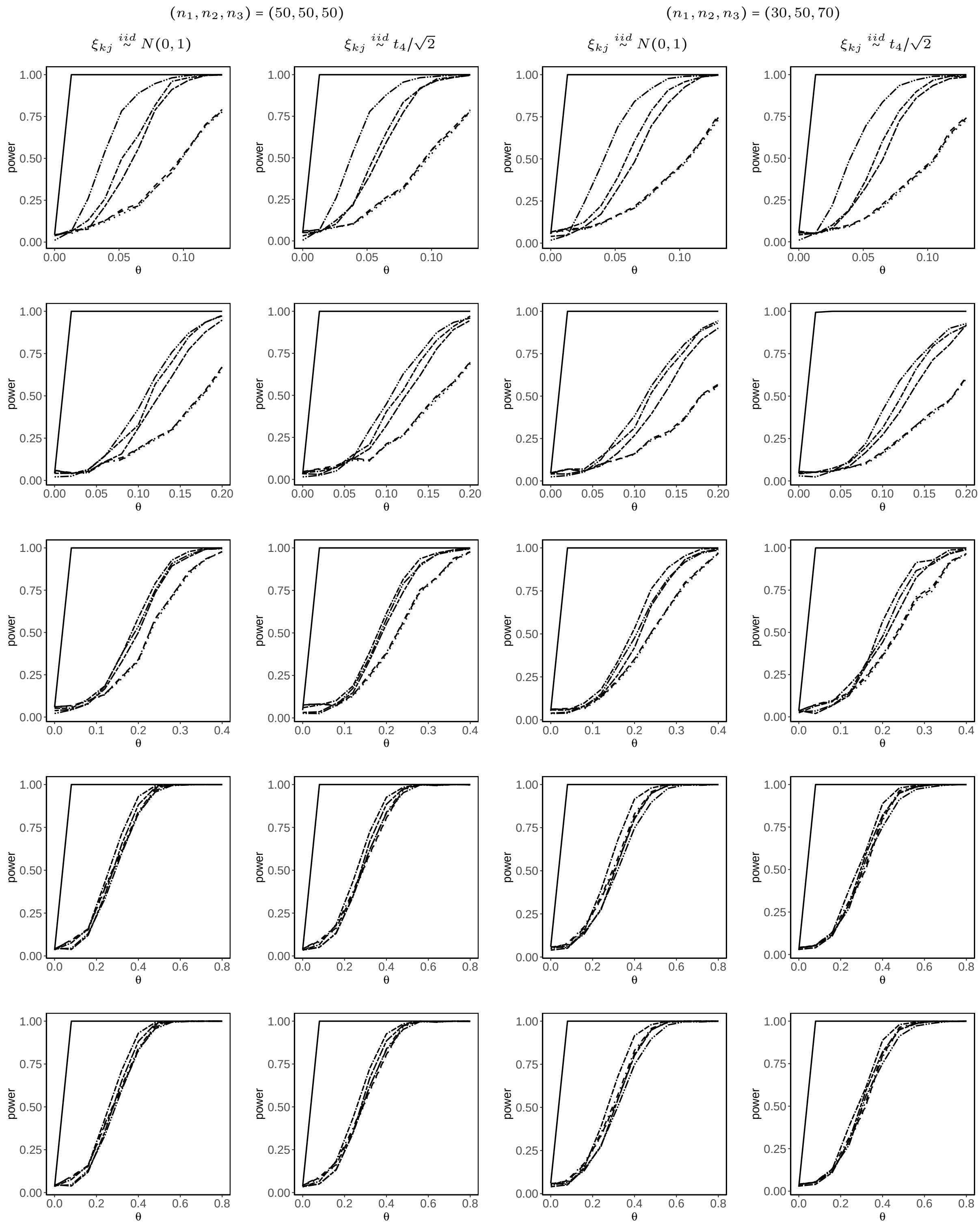

Figure S1: Empirical power of the proposed functional ANOVA (solid), L2 (dashed), F (dotted), GPF (dot-dashed), MPF (dot-dash-dashed), GET (short-long-dashed) and RP (dot-dot-dashed) in the simulation setting of Zhang et al. (2019). Rows 1 to 5 correspond to $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$, respectively. The power functions of L2, F and GPF are nearly indistinguishable.

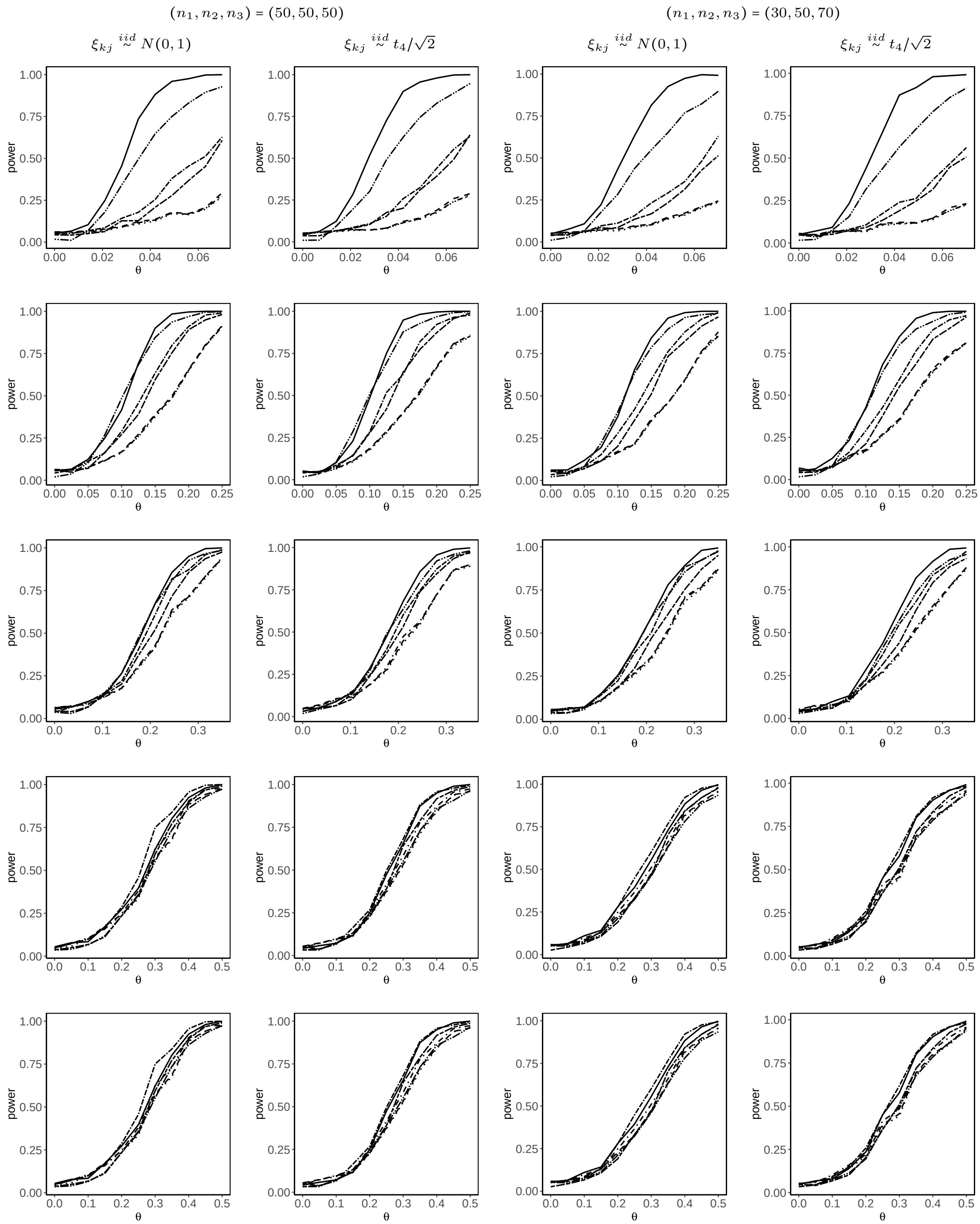

Figure S2: Empirical power of the proposed functional ANOVA (solid), L2 (dashed), F (dotted), GPF (dot-dashed), MPF (dot-dash-dashed), GET (short-long-dashed) and RP (dot-dot-dashed) in the modified simulation setting of Zhang et al. (2019). Rows 1 to 5 correspond to $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$, respectively. The power functions of L2, F and GPF are nearly indistinguishable.

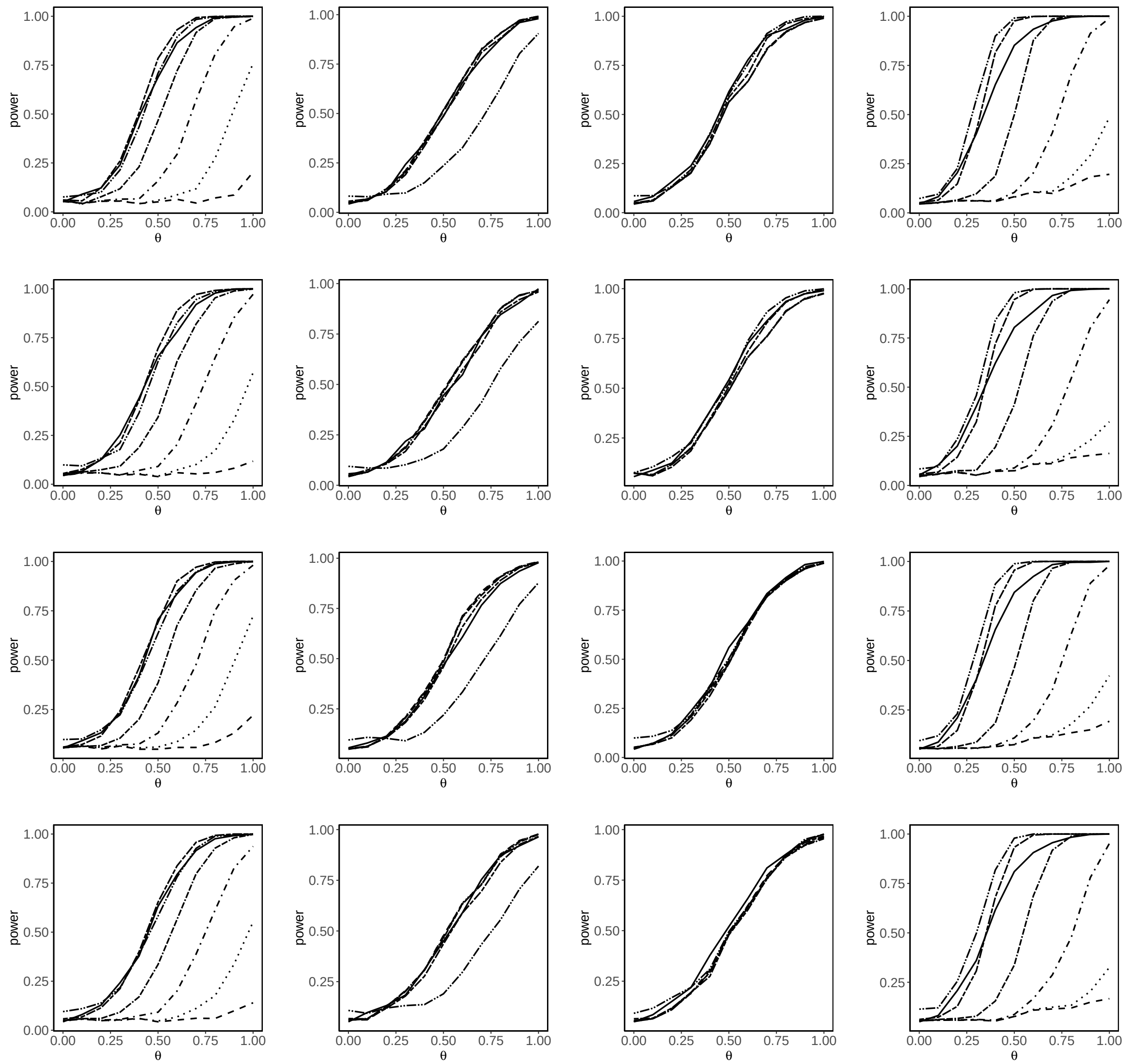

Figure S3: Empirical power of the proposed functional ANOVA with adaptively selected τ (solid), and fixed τ at 0 (dashed), 0.2 (dotted), 0.4 (dot-dashed), 0.6 (dot-dash-dashed), 0.8 (short-long-dashed) and 0.99 (dot-dot-dashed). From left to right the panels display the empirical power functions for families (M1), (M2), (M3) and (M4). The first two rows correspond to the "common covariance" setting and the last two correspond to the "group-specific covariance" setting. The sample sizes are $(n_1, n_2, n_3) = (50, 50, 50)$ for the first and third rows , and $(n_1, n_2, n_3) = (30, 50, 70)$ for the second and fourth rows.

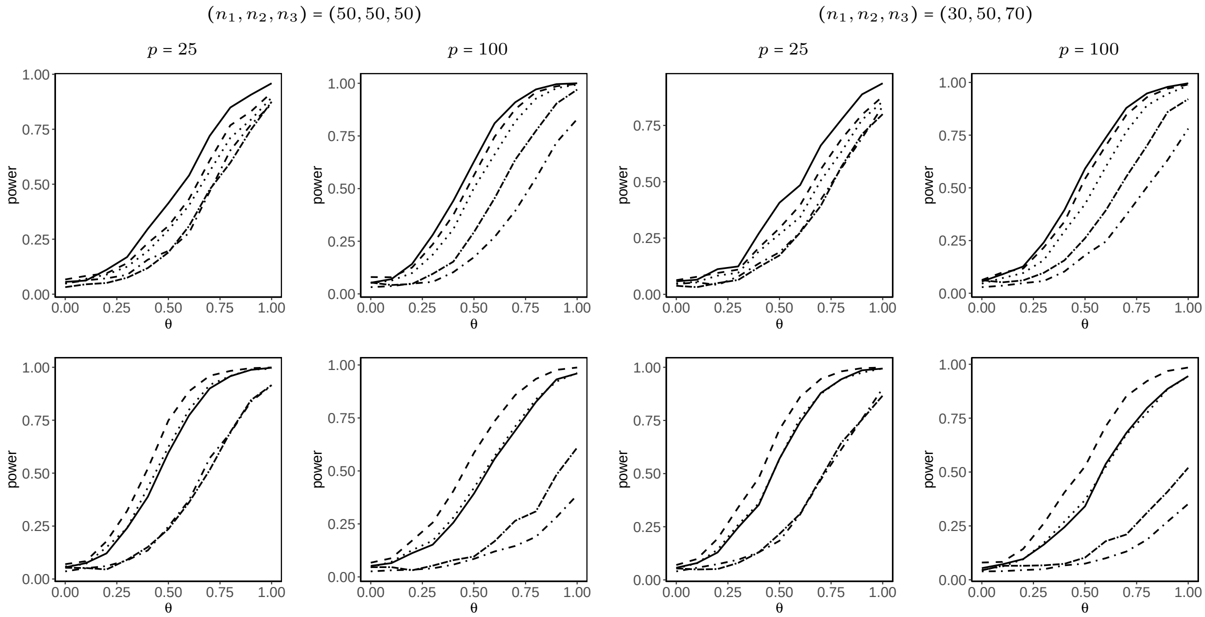

Figure S4: Empirical power of the proposed high-dimensional ANOVA (solid), DALp (dashed), S (dotted), LH (dot-dashed) and RRLH (dot-dash-dashed). Top, sparse setting; bottom, dense setting.

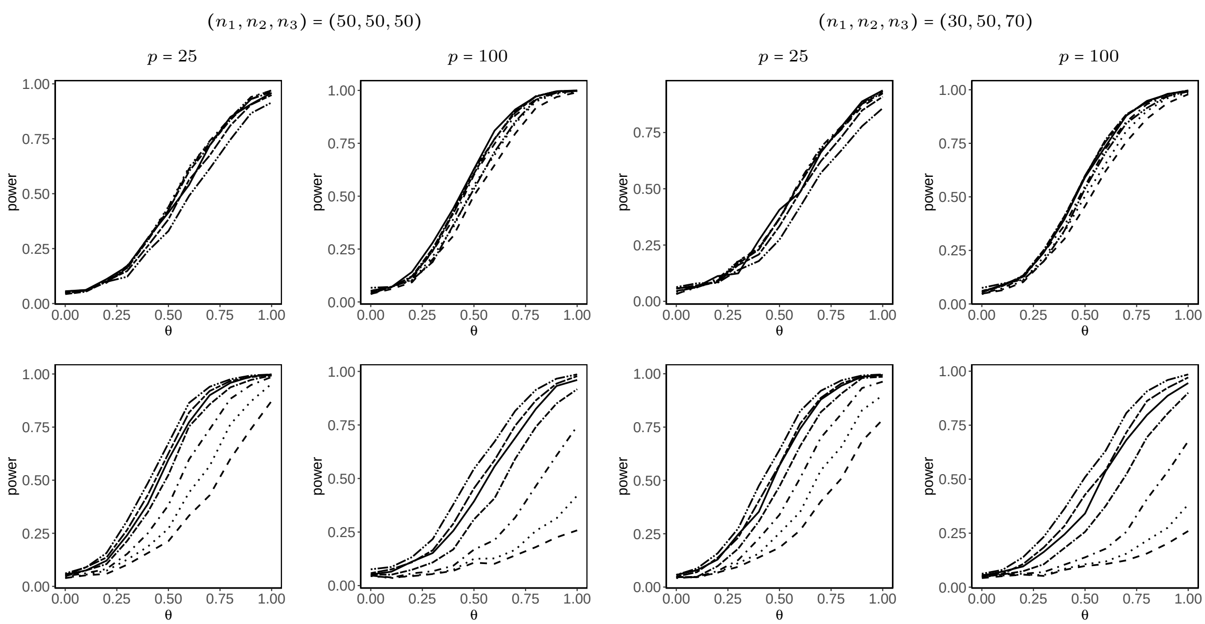

Figure S5: Empirical power of the proposed high-dimensional ANOVA with adaptively selected τ (solid), and fixed τ at 0 (dashed), 0.2 (dotted), 0.4 (dot-dashed), 0.6 (dot-dash-dashed), 0.8 (short-long-dashed) and 0.99 (dot-dot-dashed).