Estimation and Inference for High-dimensional Multi-response Growth Curve Model

Xin Zhou[1], Yin Xia[2], and Lexin Li[1]

[1] *University of California at Berkeley, and* [2] *Fudan University*

Supplementary Material

In this Supplementary Material, Section A presents a collection of technical lemmas and some of their proofs. Section B presents the proofs of Theorems 1 to 5 in the main paper. Section C presents some additional simulation and real data analysis results.

A Auxiliary Lemmas

To facilitate the theoretical development, we introduce the oracle estimators, where the covariance matrix $\Sigma^{(r)}$ is treated as known. That is, we plug the true covariance $\Sigma^{(r)}$ in (2.17) to estimate $\beta^{(r)}$, and denote it as $(\hat{\beta}^o)^{(r)}$. We then plug $(\hat{\beta}^o)^{(r)}$ in (2.18) to estimate the test statistics $J, J_{r,j}^2$, and denote them as $J^o, (J_{r,j}^o)^2$. In contrast, we view (2.17) and (2.18) where we use the estimated covariance $\hat{\Sigma}^{(r)}$ given data as the data-driven estimators, and denote them as $(\hat{\beta}^d)^{(r)}$, J^d, $(J_{r,j}^d)^2$, respectively. In our theoretical development, we first study the oracle case, then show that the data-driven case is asymptotically equivalent.

Lemma 1. *(Bonferroni inequality) Let* $B = \bigcup_{t=1}^{p} B_t$. *For any* $k \leq p/2$,

$$\sum_{t=1}^{2k} (-1)^{t-1} F_t \leq \mathbb{P}(B) \leq \sum_{t=1}^{2k-1} (-1)^{t-1} F_t,$$

where $F_t = \sum_{1 \leq i_1 < \ldots i_t \leq p} \mathbb{P}(B_{i_1} \bigcap \ldots B_{i_t})$.

Lemma 2. *(Lemma 6 of Cai et al. (2014)) Let* $(Z_1, \ldots, Z_p)^\intercal$ *be a multivariate normal random vector with mean zero and covariance* Σ *and all diagonal elements* $\Sigma_{i,i} = 1$ *for* $1 \leq i \leq p$. *Suppose* $\max_{i \neq j} |\Sigma_{i,j}| \leq C_1 < 1$, *and* $\max_j \sum_{i=1}^{p} \Sigma_{ij}^2 \leq C_2$,, *for some constants* C_1, C_2. *Then for any* $\phi \in \mathbb{R}$,

$$\mathbb{P}\left(\max_{1 \leq i \leq p} Z_i^2 - 2\log p + \log\log p \leq \phi\right) \to \exp\left\{-\frac{1}{\sqrt{\pi}} \exp\left(-\frac{\phi}{2}\right)\right\}.$$

Lemma 3. *(Lemma 2 of Berman (1962)) Suppose* X *and* Y *follow a bivariate normal distribution with zero mean, unit variance and correlation coefficient* ρ. *Then,*

$$\lim_{c \to \infty} \frac{P(X > c, Y > c)}{\{2\pi(1-\rho)^{1/2} c^2\}^{-1} \exp(-c^2/(1+\rho))(1+\rho)^{1/2}} = 1,$$

uniformly for all ρ *such that* $|\rho| \leq \delta$, $0 < \delta < 1$.

Lemma 4. *Let* Λ *be any subset of* $\{(r, j) : 1 \leq r \leq R, 1 \leq j \leq 2p + 2\}$, *and let* $|\Lambda|$ *denote the cardinality. Then, for some constant* $C > 0$,

$$\mathbb{P}\Big(\max_{(r,j) \in \Lambda} \frac{((\hat{\beta}^o)_j^{(r)} - \beta_j^{(r)})^2}{\{(X^\intercal \{\Sigma^{(r)}\}^{-1} X)^{-1}\}_{j,j}} \geq x^2\Big) \leq C|\Lambda|\{1 - \Phi(x)\} + O(p^{-1}),$$

uniformly for $0 \leq x \leq (8 \log \tilde{p})^{1/2}$ *and* $\Lambda \subseteq \{(r, j) : 1 \leq r \leq R, 1 \leq j \leq 2p + 2\}$. *For* *the sub-Gaussian case, this result holds under Condition (C7).*

Lemma 5. *(Lemma 3 of Chen et al. (2023)) Let* $X = (X_1, \ldots, X_n)^\intercal$ *and* $Y = (Y_1, \ldots, Y_n)^\intercal$ *be two random vectors with independent entries, and* $\mathbb{E}X_i = \mathbb{E}Y_i = 0$, $\max_i(\|X_i\|_{\psi_2}, \|Y_i\|_{\psi_2}) \leq K$*, where* $\|X\|_{\psi_2} = \sup_{p\geq 1} p^{-1}(\mathbb{E}|X|^p)^{1/p}$ *denotes the sub-exponential norm of* X*. Let* A *be an* $n \times n$ *matrix. Then, for some constant* $c > 0$*, and every* $t > 0$*,*

$$P\left(|X^\intercal AY - \mathbb{E}X^\intercal AY| \geq t\right) \leq 2\exp\left\{-c\min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|_2}\right)\right\}.$$

Lemma 6. *Suppose Conditions (C1), (C2) and (C4) hold. Then, for any* $\phi \in \mathbb{R}$*,*

$$\mathbb{P}_{\mathcal{H}_0}\left(J^o - 2\log\tilde{p} + \log\log\tilde{p} \leq \phi\right) \to \exp\left\{-\pi^{-1/2}\exp(-\phi/2)\right\}, \text{ as } NT \text{ and } \tilde{p} \to \infty.$$

Proof: We prove this lemma in three steps. First, we formulate the correlation matrix of the oracle test statistic. Second, we show that it satisfies the correlation conditions of Lemma 2. Finally, we apply Lemma 2 to complete the proof.

Step 1: Since J^o is the maximum of the square of $\tilde{p}$ standardized normal variables, we first consider the correlation matrix $\check{\Sigma}$ of $(J^o_{1,1}, \ldots, J^o_{1,2p+2}, \ldots, J^o_{r,j}, \ldots, J^o_{R,1}, \ldots, J^o_{R,2p+2})$. We note that the diagonals of $\check{\Sigma}$ are all 1. In addition, it can be written in the following block matrix form:

$$\check{\Sigma} = \begin{pmatrix} \check{\Sigma}^{(1,1)} & \ldots & \check{\Sigma}^{(1,R)} \\ \ldots & \ldots & \ldots \\ \check{\Sigma}^{(R,1)} & \ldots & \check{\Sigma}^{(R,R)} \end{pmatrix},$$

where each block is a $(2p+2) \times (2p+2)$ matrix.

Next, for $1 \le j_1 \le 2p+2$, $1 \le j_2 \le 2p+2$, we have,

$$\big[\check{\Sigma}^{(r,r)}\big]_{j_1,j_2} = \big[(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}\big]_{j_1,j_2}\big[(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}\big]_{j_1,j_1}^{-1/2}\big[(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}\big]_{j_2,j_2}^{-1/2},$$

where $\Sigma^{(r)}$ is as defined in (2.6). For $1 \le j_1 \le 2p+2$, $1 \le j_2 \le 2p+2$, $1 \le r_1 \le R$, $1 \le r_2 \le R$, and $r_1 \ne r_2$, we have,

$$\begin{aligned}\big[\check{\Sigma}^{(r_1,r_2)}\big]_{j_1,j_2} &= \big[(X^\intercal\{\Sigma^{(r_1)}\}^{-1}X)^{-1}X^\intercal\{\Sigma^{(r_1)}\}^{-1}\Sigma^{(r_1,r_2)}\{\Sigma^{(r_2)}\}^{-1}X(X^\intercal\{\Sigma^{(r_2)}\}^{-1}X)^{-1}\big]_{j_1,j_2}\\ &\quad\times\big[(X^\intercal\{\Sigma^{(r_1)}\}^{-1}X)^{-1}\big]_{j_1,j_1}^{-1/2}\big[(X^\intercal\{\Sigma^{(r_2)}\}^{-1}X)^{-1}\big]_{j_2,j_2}^{-1/2},\end{aligned}$$

where $\Sigma^{(r_1,r_2)}$ is as defined in Condition (C4).

Step 2: By Condition (C4), the maximum absolute value of the off-diagonal entries of $\check{\Sigma}$ is bounded by a positive constant smaller than 1. We show in this step that, for each column of $\check{\Sigma}$, the sum of squares for all entries is upper bounded by a positive constant c, i.e., $\max_{r,j}\sum_{r_2,j_2}\{[\check{\Sigma}^{(r,r_2)}]_{j,j_2}\}^2 \le c$.

First, we show that $\big[(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}\big]_{j,j}^{-1/2} = O\{(NT)^{1/2}\}$ uniformly for $1 \le j \le 2p+2$, $1 \le r \le R$. Write $(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1} = \tilde{\Sigma}^{(r)}$. Then we need to show that $|[\{\tilde{\Sigma}^{(r)}\}^{-1}]_{j,j}| = O(NT)$. At the same time, we obtain the bounds for $\|\Sigma^{(r)}\|_2$, $\|\{\Sigma^{(r)}\}^{-1}\|_2$, $\|\tilde{\Sigma}^{(r)}\|_2$, and $\|\{\tilde{\Sigma}^{(r)}\}^{-1}\|_2$, which are useful in the proof of both this lemma and Theorem 2.

We begin with the bounds for $\|\Sigma^{(r)}\|_2$ and $\|\{\Sigma^{(r)}\}^{-1}\|_2$. By Conditions (C1), (C2),

$$\|G(I_N\otimes\Sigma_\zeta)G^\intercal\|_2 \le \|G\|_2^2\|I_N\otimes\Sigma_\zeta\|_2 \le \max_i\|G_i\|_2^2\|\Sigma_\zeta\|_2 \le 2T\max_i\|G_i\|_{\max}^2\|\Sigma_\zeta\|_2 = O(1).$$

By Condition (C2), we have

$$\max_r \lambda_{\max}([\Sigma_R]_{r,r}\Sigma_T) = \max_r [\Sigma_R]_{r,r}\lambda_{\max}(\Sigma_T) \leq \lambda_{\max}(\Sigma_T)\lambda_{\max}(\Sigma_R) = O(1).$$

Similarly, we have

$$\min_r \lambda_{\min}([\Sigma_R]_{r,r}\Sigma_T) = \min_r [\Sigma_R]_{r,r}\lambda_{\min}(\Sigma_T) \geq \lambda_{\min}(\Sigma_T)\lambda_{\min}(\Sigma_R) \geq c_2^{-2}.$$

By Weyl's inequality and the definition of $\Sigma^{(r)}$ in (2.6), we have $\lambda_{\max}(\Sigma^{(r)}) = O(1)$, and $\lambda_{\min}(\Sigma^{(r)}) \geq c_2^{-2}$ uniformly for $r = 1, \ldots, R$. This implies that $\|\Sigma^{(r)}\|_2 = O(1)$ and $\|\{\Sigma^{(r)}\}^{-1}\|_2 = O(1)$. Note that

$$\begin{aligned}\{\tilde{\Sigma}^{(r)}\}^{-1} = X^\intercal\{\Sigma^{(r)}\}^{-1}X &= \lambda_{\min}(\{\Sigma^{(r)}\}^{-1})X^\intercal X + X^\intercal\left(\{\Sigma^{(r)}\}^{-1} - \lambda_{\min}(\{\Sigma^{(r)}\}^{-1})I_{NT}\right)X \\ &= \lambda_{\max}^{-1}(\Sigma^{(r)})X^\intercal X + X^\intercal\left(\{\Sigma^{(r)}\}^{-1} - \lambda_{\max}^{-1}(\Sigma^{(r)})I_{NT}\right)X.\end{aligned}$$

By Condition (C1), we have

$$\begin{aligned}\lambda_{\max}\{(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}\} = \lambda_{\min}^{-1}(X^\intercal\{\Sigma^{(r)}\}^{-1}X) &\leq \lambda_{\min}^{-1}\{\lambda_{\max}^{-1}(\Sigma^{(r)})X^\intercal X\} \\ &= \lambda_{\min}^{-1}(X^\intercal X)\lambda_{\max}(\Sigma^{(r)}) = O(N^{-1}T^{-1}),\end{aligned} \tag{A.1}$$

uniformly for $r = 1, \ldots, R$. Similarly, we have

$$\begin{aligned}\lambda_{\min}\{(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}\} = \lambda_{\max}^{-1}(X^\intercal\{\Sigma^{(r)}\}^{-1}X) &\geq \lambda_{\max}^{-1}\{\lambda_{\min}^{-1}(\Sigma^{(r)})X^\intercal X\} \\ &= \lambda_{\max}^{-1}(X^\intercal X)\lambda_{\min}(\Sigma^{(r)}) \geq c_1^{-1}c_2^{-2}N^{-1}T^{-1},\end{aligned} \tag{A.2}$$

uniformly for $r = 1, \ldots, R$. Together, (A.1) and (A.2) imply that

$$\|\tilde{\Sigma}^{(r)}\|_2 = O(N^{-1}T^{-1}), \quad \|\{\tilde{\Sigma}^{(r)}\}^{-1}\|_2 = O(NT), \quad |[\{\tilde{\Sigma}^{(r)}\}^{-1}]_{j,j}| = O(NT),$$

uniformly for $1 \le j \le 2p+2$, $1 \le r \le R$.

Next, we show that $\max_{r,j} \sum_{r_2,j_2} \{[\check{\Sigma}^{(r,r_2)}]_{j,j_2}\}^2 \le c$, for each $1 \le r \le R$, $1 \le j \le 2p+2$. We note that

$$\begin{aligned}
\sum_{r_2,j_2} \{[\check{\Sigma}^{(r,r_2)}]_{j,j_2}\}^2 &= \sum_{j_2=1}^{2p+2} \{[\check{\Sigma}^{(r,r)}]_{j,j_2}\}^2 + \sum_{r_2\neq r} \sum_{j_2=1}^{2p+2} \{[\check{\Sigma}^{(r,r_2)}]_{j,j_2}\}^2 \\
&= \sum_{j_2=1}^{2p+2} \{[\tilde{\Sigma}^{(r)}]_{j,j_2}\}^2 \{[\tilde{\Sigma}^{(r)}]_{j,j}\}^{-1} \{[\tilde{\Sigma}^{(r)}]_{j_2,j_2}\}^{-1} \\
&\quad + \sum_{r_2\neq r} \sum_{j_2=1}^{2p+2} \left[\tilde{\Sigma}^{(r)} X^\intercal \{\Sigma^{(r)}\}^{-1} \Sigma^{(r,r_2)} \{\Sigma^{(r_2)}\}^{-1} X \tilde{\Sigma}^{(r_2)}\right]_{j,j_2}^2 \{[\tilde{\Sigma}^{(r)}]_{j,j}\}^{-1} \{[\tilde{\Sigma}^{(r)}]_{j_2,j_2}\}^{-1}.
\end{aligned}$$

By the fact that $\{[\tilde{\Sigma}^{(r)}]_{j,j}\}^{-1} \le [\{\tilde{\Sigma}^{(r)}\}^{-1}]_{j,j} = O(NT)$ for $1 \le j \le 2p+2$, we have,

$$\sum_{r_2,j_2} \{[\check{\Sigma}^{(r,r_2)}]_{j,j_2}\}^2 = O(N^2T^2) \left(\sum_{j_2=1}^{2p+2} \{[\tilde{\Sigma}^{(r)}]_{j,j_2}\}^2 + \sum_{r_2\neq r} \sum_{j_2=1}^{2p+2} \left[\tilde{\Sigma}^{(r)} X^\intercal \{\Sigma^{(r)}\}^{-1} \Sigma^{(r,r_2)} \{\Sigma^{(r_2)}\}^{-1} X \tilde{\Sigma}^{(r_2)}\right]_{j,j_2}^2 \right).$$

So it suffices to show that

$$\sum_{j_2=1}^{2p+2} \{[\tilde{\Sigma}^{(r)}]_{j,j_2}\}^2 + \sum_{r_2\neq r} \sum_{j_2=1}^{2p+2} \left[\tilde{\Sigma}^{(r)} X^\intercal \{\Sigma^{(r)}\}^{-1} \Sigma^{(r,r_2)} \{\Sigma^{(r_2)}\}^{-1} X \tilde{\Sigma}^{(r_2)}\right]_{j,j_2}^2 = O(N^{-2}T^{-2}), \tag{A.3}$$

uniformly for $1 \le j \le 2p+2$, $1 \le r \le R$.

To bound the first term in the left-hand-side of (A.3), because $\|\tilde{\Sigma}^{(r)}\|_2 = O(N^{-1}T^{-1})$, we have,

$$\sum_{j_2=1}^{2p+2} \{[\tilde{\Sigma}^{(r)}]_{j,j_2}\}^2 \le \|\tilde{\Sigma}^{(r)}\|_2^2 = O(N^{-2}T^{-2}).$$

To bound the second term in the left-hand-side of (A.3), we have,

$$\sum_{r_2\neq r}\sum_{j_2=1}^{2p+2}\Big[\tilde\Sigma^{(r)}X^\intercal\{\Sigma^{(r)}\}^{-1}\Sigma^{(r,r_2)}\{\Sigma^{(r_2)}\}^{-1}X\tilde\Sigma^{(r_2)}\Big]^2_{j,j_2}$$
$$\leq \sum_{r_2\neq r}\Big\|\{\tilde\Sigma^{(r)}X^\intercal\{\Sigma^{(r)}\}^{-1}\Sigma^{(r,r_2)}\{\Sigma^{(r_2)}\}^{-1}X\tilde\Sigma^{(r_2)}\}\Big\|_2^2$$
$$= \sum_{r_2\neq r}\Big\|\{\tilde\Sigma^{(r)}X^\intercal\{\Sigma^{(r)}\}^{-1}\mathrm{diag}(\{[\Sigma_R]_{r,r_2}\Sigma_T\}_{i=1}^N)\{\Sigma^{(r_2)}\}^{-1}X\tilde\Sigma^{(r_2)}\}\Big\|_2^2$$
$$\leq \sum_{r_2\neq r}[\Sigma_R]^2_{r,r_2}\|\tilde\Sigma^{(r)}\|_2^2\ \Big\|X^\intercal\{\Sigma^{(r)}\}^{-1}\mathrm{diag}(\{\Sigma_T\}_{i=1}^N)\{\Sigma^{(r_2)}\}^{-1}X\Big\|_2^2\ \|\tilde\Sigma^{(r_2)}\|_2^2.$$

Recall that $\|\{\Sigma^{(r)}\}^{-1}\|_2 = O(1)$. By Conditions (C1) and (C2), we have that $\|X^\intercal\{\Sigma^{(r)}\}^{-1}$ $\mathrm{diag}(\{\Sigma_T\}_{i=1}^N)\{\Sigma^{(r_2)}\}^{-1}X\|_2^2 = O(N^2T^2)$. Therefore, we have that,

$$\sum_{r_2\neq r}\sum_{j_2=1}^{2p+2}\Big[\tilde\Sigma^{(r)}X^\intercal\{\Sigma^{(r)}\}^{-1}\Sigma^{(r,r_2)}\{\Sigma^{(r_2)}\}^{-1}X\tilde\Sigma^{(r_2)}\Big]^2_{j,j_2}$$
$$\leq \sum_{r_2\neq r}\Sigma^2_{R,r,r_2}c_1(N^{-2}T^{-2})(N^2T^2)(N^{-2}T^{-2}) \leq C_1\|\Sigma_R\|_2^2N^{-2}T^{-2}\leq C_2N^{-2}T^{-2},$$

where C_1 and C_2 are some positive constants. In the first inequality we use the fact that $\|\tilde\Sigma^{(r)}\|_2 = O(N^{-1}T^{-1})$ for $1\leq r\leq R$, and in the last inequality we use Condition (C2). Therefore, we prove (A.3), which yields $\max_{r,j}\sum_{r_2,j_2}\{[\check\Sigma^{(r,r_2)}]_{j,j_2}\}^2\leq c$.

Step 3: Applying Lemma 2 completes the proof of Lemma 6. □

Lemma 7. *Suppose Conditions (C1) - (C5) hold. Suppose $\tilde p_0 = |\mathcal{H}_0| \asymp \tilde p$, and for some $\rho>0$ and $\delta>0$, $|S_\rho|\geq\{1/(\pi^{1/2}\alpha)+\delta\}(\log\tilde p)^{1/2}$. Then,*

$$\lim_{(NT,\tilde p)\to\infty}\frac{\mathrm{FDR}^o}{\alpha\tilde p_0/\tilde p} = 1,\ \lim_{(NT,\tilde p)\to\infty}\frac{\mathrm{FDP}^o(\hat\tau)}{\alpha\tilde p_0/\tilde p} = 1 \textit{ in probability,}$$

where $\mathrm{FDR}^o, \mathrm{FDP}^o$ *are the false discovery rate and proportion under the oracle case by plugging in the oracle test statistic* $(J^o_{r,j})^2$.

Proof: We first show that $P(\hat{\tau}$ exists in $[0, t_{\tilde{p}}]) = 1$. Recall that $S_\rho = \{(r,j) \in \mathcal{H} : \{\beta_j^{(r)}\}^2/\{(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}\}_{j,j} \geq (\log \tilde{p})^{1+\rho}\}$. Then, with probability tending to 1,

$$\sum_{r,j} I\left\{|J^o_{r,j}| > (2\log\tilde{p})^{1/2}\right\} \geq \{1/(\pi^{1/2}\alpha) + \delta\}(\log\tilde{p})^{1/2}.$$

Then with probability tending to 1, we also have that,

$$\frac{\tilde{p}}{\sum_{r,j} I\{|J^o_{r,j}| > (2\log\tilde{p})^{1/2}\} \vee 1} \leq \tilde{p}\{1/(\pi^{1/2}\alpha) + \delta\}^{-1}(\log\tilde{p})^{-1/2}.$$

Denoting by $G(\tau) = 2\{1 - \Phi(\tau)\}$, we have that $G(t_{\tilde{p}}) \sim (2/\pi)^{1/2} t_{\tilde{p}}^{-1} \exp(-t_{\tilde{p}}^2/2)$. Therefore, by the definition of $t_{\tilde{p}}$ in Algorithm 3, we have that $\mathbb{P}(\hat{\tau}$ exists in $[0, t_{\tilde{p}}]) = 1$.

By the definition of $\hat{\tau}$, we have that,

$$\frac{\tilde{p}G(\hat{\tau})}{\sum_{r,j} I(|J^o_{r,j}| \geq \hat{\tau}) \vee 1} = \alpha.$$

Thus, to prove Lemma 7, it suffices to show that

$$\left|\frac{\sum_{(r,j)\in\mathcal{H}_0}\{I(|J^o_{r,j}| \geq \tau) - G(\tau)\}}{\tilde{p}_0 G(\tau)}\right| \to 0 \text{ in probability, uniformly for } 0 \leq \tau \leq t_{\tilde{p}}.$$

Let $0 = \tilde{t}_0 < \tilde{t}_1 < \ldots < \tilde{t}_b = t_{\tilde{p}}$, such that $\tilde{t}_h - \tilde{t}_{h-1} = v_{\tilde{p}}$ for $1 \leq h \leq b-1$ and $\tilde{t}_{\tilde{p}} - \tilde{t}_{\tilde{p}-1} \leq v_{\tilde{p}}$, where $v_{\tilde{p}} = (\log\tilde{p}\log_4\tilde{p})^{-1/2}$. We have that $b \sim t_{\tilde{p}}/v_{\tilde{p}}$. For any τ, such that $\tilde{t}_{h-1} \leq \tau \leq \tilde{t}_h$, we have that,

$$\frac{\sum_{(r,j)\in\mathcal{H}_0} I(|J_{r,j}^o| \geq \tilde{t}_{h-1})}{\tilde{p}_0 G(\tilde{t}_h)} \frac{G(\tilde{t}_h)}{G(\tilde{t}_{h-1})} \leq \frac{\sum_{(r,j)\in\mathcal{H}_0} I(|J_{r,j}^o| \geq \tau)}{\tilde{p}_0 G(\tau)} \leq \frac{\sum_{(r,j)\in\mathcal{H}_0} I(|J_{r,j}^o| \geq \tilde{t}_{h-1})}{\tilde{p}_0 G(\tilde{t}_{h-1})} \frac{G(\tilde{t}_{h-1})}{G(\tilde{t}_h)}.$$

Therefore, it suffices to show that

$$\max_{1\leq h\leq b} \left| \frac{\sum_{(r,j)\in\mathcal{H}_0} \left\{ I\left(|J_{r,j}^o| \geq \tilde{t}_h\right) - G(\tilde{t}_h)\right\}}{\tilde{p}_0 G(\tilde{t}_h)} \right| \to 0 \text{ in probability.} \tag{A.4}$$

We have, for any $\epsilon > 0$,

$$\begin{aligned}
& P\left[\max_{1\leq h\leq b} \left| \frac{\sum_{(r,j)\in\mathcal{H}_0}\{I(|J_{r,j}^o| \geq \tilde{t}_h) - G(\tilde{t}_h)\}}{\tilde{p}_0 G(\tilde{t}_h)} \right| \geq \epsilon \right] \\
\leq & \sum_{h=1}^{b} P\left[\left| \frac{\sum_{(r,j)\in\mathcal{H}_0}\{I(|J_{r,j}^o| \geq \tilde{t}_h) - G(\tilde{t}_h)\}}{\tilde{p}_0 G(\tilde{t}_h)} \right| \geq \epsilon \right] \\
\leq & \frac{1}{v_{\tilde{p}}} \int_0^{t_{\tilde{p}}} P\left[\left| \frac{\sum_{(r,j)\in\mathcal{H}_0}\{I(|J_{r,j}^o| \geq \tau) - G(\tau)\}}{\tilde{p}_0 G(\tau)} \right| \geq \frac{1}{2}\epsilon \right] d\tau \\
& \quad + \sum_{h=b-1}^{b} P\left[\left| \frac{\sum_{(r,j)\in\mathcal{H}_0}\{I(|J_{r,j}^o| \geq \tilde{t}_h) - G(\tilde{t}_h)\}}{\tilde{p}_0 G(\tilde{t}_h)} \right| \geq \epsilon \right].
\end{aligned}$$

Since $\mathbb{P}(|J_{r,j}^o| \geq \tau) = G(\tau)$ for $(r,j) \in \mathcal{H}_0$, we only need to show that, for any $\epsilon > 0$,

$$\int_0^{t_{\tilde{p}}} P\left[\left| \frac{\sum_{(r,j)\in\mathcal{H}_0}\{I(|J_{r,j}^o| \geq \tau) - \mathbb{P}(|J_{r,j}^o| \geq \tau)\}}{\tilde{p}_0 G(\tau)} \right| \geq \epsilon \right] d\tau = o(v_{\tilde{p}}), \tag{A.5}$$

and

$$\max_{0\leq\tau\leq t_{\tilde{p}}} P\left[\left| \frac{\sum_{(r,j)\in\mathcal{H}_0}\{I(|J_{r,j}^o| \geq \tau) - \mathbb{P}(|J_{r,j}^o| \geq \tau)\}}{\tilde{p}_0 G(\tau)} \right| \geq \epsilon \right] = o(1). \tag{A.6}$$

Since the proofs of (A.5) and (A.6) are similar, we only prove (A.5) here.

By Markov's inequality, it suffices to bound

$$E\left|\frac{\sum_{(r,j)\in\mathcal{H}_0}\{I(|J^o_{r,j}|\geq\tau)-\mathbb{P}(|J^o_{r,j}|\geq\tau)\}}{\tilde{p}_0G(\tau)}\right|^2$$
$$=\frac{\sum_{(r_1,j_1),(r_2,j_2)\in\mathcal{H}_0}\{\mathbb{P}(|J^o_{r_1,j_1}|\geq\tau,|J^o_{r_2,j_2}|\geq\tau)-\mathbb{P}(|J^o_{r_1,j_1}|\geq\tau)\mathbb{P}(|J^o_{r_2,j_2}|\geq\tau)\}}{\tilde{p}_0^2G^2(\tau)} \tag{A.7}$$

We next divide the index pairs $(r_1,j_1),(r_2,j_2)\in\mathcal{H}_0$ into several parts depending on the covariance $\mathrm{cov}(J^o_{r_1,j_1},J^o_{r_2,j_2})$. Specifically, for some small enough constant $\gamma>0$, denote by

$$\Gamma_{r,j}(\gamma)=\left\{(r_0,j_0):(r_0,j_0)\neq(r,j),|\mathrm{cov}(J^o_{r,j},J^o_{r_0,j_0})|\geq(\log\tilde{p})^{-2-\gamma}\right\}.$$

In Step 2 of the proof of Lemma 6, we obtain that $|\mathrm{cov}(J^o_{r_1,j_1},J^o_{r_2,j_2})|\leq c_0$ holds uniformly for (r_1,j_1), (r_2,j_2), and some constant $c_0<1$. In addition, $\sum_{r_0,j_0}\mathrm{cov}^2(J^o_{r_0,j_0},J^o_{r,j})=O(1)$ for any (r,j), which implies $\max_{(r,j)\in\mathcal{H}_0}|\Gamma_{r,j}(\gamma)|=o(\tilde{p}^\iota)$ for any constant $\iota>0$.

Then the index pairs $\{(r_1,j_1),(r_2,j_2):(r_1,j_1)\in\mathcal{H}_0,(r_2,j_2)\in\mathcal{H}_0\}$ can be divided into three parts: $H_{01}=\{(r_1,j_1),(r_2,j_2):(r_1,j_1)\in\mathcal{H}_0,(r_2,j_2)\in\mathcal{H}_0,(r_1,j_1)=(r_2,j_2)\}$; $H_{02}=\{(r_1,j_1),(r_2,j_2):(r_1,j_1)\in\mathcal{H}_0,(r_2,j_2)\in\mathcal{H}_0,(r_2,j_2)\in\Gamma_{r_1,j_1}(\gamma)\}$ that contains highly correlated index pairs; and $H_{03}=\{(r_1,j_1),(r_2,j_2):(r_1,j_1)\in\mathcal{H}_0,(r_2,j_2)\in\mathcal{H}_0,(r_2,j_2)\notin\Gamma_{r_1,j_1}(\gamma)\}$ that contains weakly correlated pairs.

For H_{01}, we have,

$$\frac{\sum_{\{(r_1,j_1),(r_2,j_2)\}\in H_{01}}\{\mathbb{P}(|J^o_{r_1,j_1}|\geq\tau,|J^o_{r_2,j_2}|\geq\tau)-\mathbb{P}(|J^o_{r_1,j_1}|\geq\tau)\mathbb{P}(|J^o_{r_2,j_2}|\geq\tau)\}}{\tilde{p}_0^2G^2(\tau)}$$
$$\leq\frac{C\tilde{p}_0G(\tau)}{\tilde{p}_0^2G^2(\tau)}=\frac{C}{\tilde{p}_0G(\tau)}. \tag{A.8}$$

For H_{02}, we have $|H_{02}|\leq\tilde{p}_0\max_{(r,j)\in\mathcal{H}_0}|\Gamma_{r,j}(\gamma)|=o(\tilde{p}^{1+\iota})$. Because for $\tau\leq1$, we have $\mathbb{P}(|J^o_{r_1,j_1}| \geq \tau, |J^o_{r_2,j_2}| \geq \tau)/G^2(\tau) = O(1)$. Therefore, by Lemma 3,

$$\frac{\sum_{\{(r_1,j_1),(r_2,j_2)\}\in H_{02}}\{\mathbb{P}(|J^o_{r_1,j_1}| \geq \tau, |J^o_{r_2,j_2}| \geq \tau) - \mathbb{P}(|J^o_{r_1,j_1}| \geq \tau)\mathbb{P}(|J^o_{r_2,j_2}| \geq \tau)\}}{\tilde{p}_0^2 G^2(\tau)}$$
$$\leq o(\tilde{p}^{-1+\iota}) + o(\tilde{p}^{1+\iota})\frac{(\tau+1)^{-2}\exp\{-\tau^2/(1+c_0)\}}{\tilde{p}_0^2 G^2(\tau)} \leq o[\frac{1}{\tilde{p}^{1-\iota}\{G(\tau)\}^{2c_0/(1+c_0)}}]. \tag{A.9}$$

For H_{03}, by the proof of Lemma 6 in Cai et al. (2014), for any $\{(r_1, j_1), (r_2, j_2)\} \in H_{03}$, we have,

$$\frac{\mathbb{P}(|J^o_{r_1,j_1}| \geq \tau, |J^o_{r_2,j_2}| \geq \tau) - G^2(\tau)}{G^2(\tau)} = o\{(\log \tilde{p})^{-1-\gamma/2}\}.$$

Then we obtain that,

$$\frac{\sum_{\{(r_1,j_1),(r_2,j_2)\}\in H_{03}}\{\mathbb{P}(|J^o_{r_1,j_1}| \geq \tau, |J^o_{r_2,j_2}| \geq \tau) - \mathbb{P}(|J^o_{r_1,j_1}| \geq \tau)\mathbb{P}(|J^o_{r_2,j_2}| \geq \tau)\}}{\tilde{p}_0^2 G^2(\tau)}$$
$$= o\{(\log \tilde{p})^{-1-\gamma/2}\}. \tag{A.10}$$

Combining (A.8), (A.9) and (A.10), we prove (A.5) holds. We can prove (A.6) similarly. Combining (A.5) and (A.6) competes the proof of Lemma 7. □

B Proof of Main Theorems

B.1 Proof of Theorem 1

We prove the result under the sub-Gaussian condition (C7). The result under the normality condition in Theorem 1 follows naturally.

Throughout the proof, we denote $y_{i,r} = y_{i,r,\cdot} = (y_{i,r,1}, \ldots, y_{i,r,T})$, and $y_{i,t} = y_{i,\cdot,t} =$

$(y_{i,1,t},\dots,y_{i,R,t})$. Same rules apply to other notation as well. Let $\delta_{i,r,t}=\epsilon_{i,r,t}+e_t^\intercal G_i\zeta_{i,r}$, where e_t is a vector with the tth entry equal to 1 and all other entries 0. Denote $\check{X}=((X_1-\bar{X})^\intercal,\dots,(X_N-\bar{X}))^\intercal\in\mathbb{R}^{TN\times(2p+q+2)}$, where $\bar{X}=N^{-1}\sum_{i=1}^N X_i\in\mathbb{R}^{T\times(2p+q+2)}$. Denote $\check{X}_i=X_i-\bar{X}$, and $\check{\delta}_{i,r,t}=\delta_{i,r,t}-\bar{\delta}_{r,t}$, where $\bar{\delta}_{r,t}=N^{-1}\sum_{i=1}^N\delta_{i,r,t}$. Let $\check{\mu}_{i,r,t}=e_t^\intercal(X_i-\bar{X})\beta^{(r)}$. We have $\check{y}_{i,r,t}=\check{\mu}_{i,r,t}+\check{\delta}_{i,r,t}$, and $\sum_{i=1}^N\check{\mu}_{i,r,t}=0$. Moreover, we have $\check{X}=PX$, where $(I-P)=N^{-1}(\mathrm{l}_N\mathrm{l}_N^\intercal)\otimes I_T$, and l_N is a vector of ones. Note that P is an orthogonal projection matrix with $P=X_P(X_P^\intercal X_P)^{-1}X_P^\intercal$, where $X_P=\mathrm{l}_N\otimes I_T$. By the Separation Theorem (Takane and Shibayama, 1991), we have $\sigma_{\max}(\check{X})\le\sigma_{\max}(X)$. In addition, Condition (C1) implies that $\|\check{X}\|_2^2=O(NT)$.

We next prove this theorem in six steps.

Step 1: We first show that $\mu_{i,r,t}$'s are negligible, so that our theoretical analysis can be based on the sample covariance matrices as given in (2.9) and (2.13) in the paper. Denote

$$Y_i=\begin{pmatrix}\check{y}_{i,1,1} & \dots & \check{y}_{i,R,1}\\ \dots & \dots & \dots\\ \check{y}_{i,1,T} & \dots & \check{y}_{i,R,T}\end{pmatrix},\quad \Delta_i=\begin{pmatrix}\check{\delta}_{i,1,1} & \dots & \check{\delta}_{i,R,1}\\ \dots & \dots & \dots\\ \check{\delta}_{i,1,T} & \dots & \check{\delta}_{i,R,T}\end{pmatrix}.$$

We next prove the following results:

$$\begin{aligned}
\left\|\frac{1}{NT}\sum_{i=1}^N(Y_i^\intercal Y_i-\Delta_i^\intercal\Delta_i)\right\|_{\max}&=O_{\mathbb{P}}\left\{c_R+\left(\frac{c_R\log R}{NT}\right)^{1/2}\right\},\\
\left\|\frac{1}{NR}\sum_{i=1}^N(Y_iY_i^\intercal-\Delta_i\Delta_i^\intercal)\right\|_{\max}&=O_{\mathbb{P}}\left\{\frac{s_Bc_B^2T}{R}+\left(\frac{s_Bc_B^2T\log T}{NR^2}\right)^{1/2}\right\},\\
\left|\frac{1}{NR}\sum_{(r_1,r_2)\in\mathcal{S}}\sum_{i=1}^N(y_{i,r_1}y_{i,r_2}^\intercal-\delta_{i,r_1}\delta_{i,r_2}^\intercal)\right|&=O_{\mathbb{P}}\left\{\frac{s_Bc_B^2T}{R}+\left(\frac{s_Bc_B^2T\log T}{NR^2}\right)^{1/2}\right\},
\end{aligned}\tag{B.11}$$

where $\mathcal{S}$ is as given in Step 2 of Algorithm 1.

To prove the first result in (B.11), for its entry in the r_1th row and r_2th column,

$$\frac{1}{NT}\sum_{i=1}^{N} e_{r_1}^{\intercal}(Y_i^{\intercal}Y_i-\Delta_i^{\intercal}\Delta_i)e_{r_2} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mu}_{i,r_1,t}\check{\mu}_{i,r_2,t}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\check{\mu}_{i,r_1,t}\check{\delta}_{i,r_2,t}+\check{\mu}_{i,r_2,t}\check{\delta}_{i,r_1,t}).$$

By Condition (C1), we have, for $1 \leq r \leq R$,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mu}_{i,r,t}^2 = \frac{1}{NT}\|\check{X}\beta^{(r)}\|_2^2 = O(c_R).$$

By the inequality of arithmetic and geometric means, we have, for $1 \leq r_1, r_2 \leq R$,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mu}_{i,r_1,t}\check{\mu}_{i,r_2,t} = O(c_R). \tag{B.12}$$

In addition,

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mu}_{i,r_1,t}\check{\delta}_{i,r_2,t} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mu}_{i,r_1,t}(\delta_{i,r_2,t}-\bar{\delta}_{r_2,t}) \\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mu}_{i,r_1,t}\delta_{i,r_2,t} - \frac{1}{NT}\sum_{t=1}^{T}(\sum_{i=1}^{N}\check{\mu}_{i,r_1,t})\bar{\delta}_{r_2,t} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mu}_{i,r_1,t}\delta_{i,r_2,t} \\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mu}_{i,r_1,t}\epsilon_{i,r_2,t} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mu}_{i,r_1,t}e_t^{\intercal}G_i\zeta_{i,r_2}.
\end{aligned} \tag{B.13}$$

We next bound the two terms in the last equation separately.

For the first term, we have that,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mu}_{i,r_1,t}\epsilon_{i,r_2,t} = \frac{1}{NT}\sum_{i=1}^{N}(\check{X}_i\beta^{(r_1)})^{\intercal}\epsilon_{i,r_2} = \sum_{i=1}^{N}\frac{1}{NT}(\check{X}_i\beta^{(r_1)})^{\intercal}\Sigma_T^{1/2}\tilde{\epsilon}_{i,r_2},$$

where $\tilde{\epsilon}_{i,r_2} = \Sigma_T^{-1/2}\epsilon_{i,r_2}$. By the definition of ϵ_{i,r_2}, we have that $\tilde{\epsilon}_{r_2} = (\tilde{\epsilon}_{1,r_2}^\intercal, \ldots, \tilde{\epsilon}_{N,r_2}^\intercal)^\intercal$ has independent entries. Then $(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \check{\mu}_{i,r_1,t}\epsilon_{i,r_2,t}$ is the weighted sum of independent sub-Gaussian variables. By Condition (C1), we have that,

$$\sum_{i=1}^N \left\| \frac{1}{NT}(\check{X}_i\beta^{(r_1)})^\intercal \Sigma_T^{1/2} \right\|_2^2 \leq \frac{1}{(NT)^2}\sum_{i=1}^N \|\check{X}_i\beta^{(r_1)}\|_2^2\|\Sigma_T^{1/2}\|_2^2 = \frac{1}{(NT)^2}\|\check{X}\beta^{(r_1)}\|_2^2\|\Sigma_T^{1/2}\|_2^2$$
$$\leq \frac{C_1}{(NT)^2}\|\check{X}\beta^{(r_1)}\|_2^2 \leq \frac{C_2}{NT}c_R.$$

Therefore, by the tail distribution of the sub-Gaussian random variables, we have that, simultaneously for $1 \leq r_1, r_2 \leq R$,

$$\left| \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \check{\mu}_{i,r_1,t}\epsilon_{i,r_2,t} \right| = O_{\mathbb{P}}\left\{ \left(\frac{c_R \log R}{NT} \right)^{1/2} \right\}. \tag{B.14}$$

For the second term, we have that,

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \check{\mu}_{i,r_1,t}e_t^\intercal G_i\zeta_{i,r_2} = \sum_{i=1}^N \frac{1}{NT}(\check{X}_i\beta^{(r_1)})^\intercal G_i\Sigma_\zeta^{1/2}\tilde{\zeta}_{i,r_2},$$

where $\tilde{\zeta}_{i,r_2} = \Sigma_\zeta^{-1/2}\zeta_{i,r_2}$. Similarly, this term is a weighted sum of independent sub-Gaussian variables. By Conditions (C1) and (C2),

$$\sum_{i=1}^N \left\| \frac{1}{NT}(\check{X}_i\beta^{(r_1)})^\intercal G_i\Sigma_\zeta^{1/2} \right\|_2^2 \leq \frac{1}{(NT)^2}\sum_{i=1}^N \|\check{X}_i\beta^{(r_1)}\|_2^2\|G_i\|_2^2\|\Sigma_\zeta^{1/2}\|_2^2$$
$$\leq \frac{1}{N^2T^2}\|\check{X}\beta^{(r_1)}\|_2^2\|G_i\|_{\max}^2 T\|\Sigma_\zeta^{1/2}\|_2^2 \leq \frac{C_1}{N^2T^2}\|\check{X}\beta^{(r_1)}\|_2^2 \leq \frac{C_2}{NT}c_R.$$

Therefore, we have that, simultaneously for $1 \leq r_1, r_2 \leq R$,

$$\left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \check{\mu}_{i,r_1,t} e_t^\intercal G_i \zeta_{i,r_2} \right| = O_{\mathbb{P}} \left\{ \left(\frac{c_R \log R}{NT} \right)^{1/2} \right\}. \tag{B.15}$$

By (B.12), (B.13), (B.14), (B.15), we prove the first result in (B.11).

To prove the second result in (B.11), for its entry in the t_1th row and t_2th column,

$$\frac{1}{NR} \sum_{i=1}^{N} e_{t_1}^\intercal (Y_i Y_i^\intercal - \Delta_i \Delta_i^\intercal) e_{t_2} = \frac{1}{NR} \sum_{i=1}^{N} \sum_{r=1}^{R} \check{\mu}_{i,r,t_1} \check{\mu}_{i,r,t_2} + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\check{\mu}_{i,r,t_1} \check{\delta}_{i,r,t_2} + \check{\mu}_{i,r,t_2} \check{\delta}_{i,r,t_1}).$$

Recall that $\check{\mu}_{i,r,t} = e_t^\intercal \check{X}_{i,-1} \beta_{-1}^{(r)}$, where $\check{X}_{i,-1} \in \mathbb{R}^{T \times (2p+q+1)}$ is a sub-matrix of $\check{X}_i$ by removing the first column, and $\beta_{-1}^{(r)} \in \mathbb{R}^{2p+q+1}$ is obtained by removing the first entry of $\beta^{(r)}$. Denote $\Upsilon = \sum_{r=1}^{R} \beta_{-1}^{(r)} \{\beta_{-1}^{(r)}\}^\intercal$, and $\tilde{X}_t = (X_{1,-1}^\intercal e_t, \ldots, X_{N,-1}^\intercal e_t)^\intercal \in \mathbb{R}^{N \times (2p+q+1)}$. Then $\|\tilde{X}_t\|_2 \leq \|X\|_2 = O\{(NT)^{1/2}\}$. By Conditions (C1) and (C2), we have that,

$$\begin{aligned}
&\frac{1}{NR} \sum_{r=1}^{N} \sum_{r=1}^{R} \check{\mu}_{i,r,t}^2 = \frac{1}{NR} \sum_{i=1}^{N} \sum_{r=1}^{R} e_t^\intercal \check{X}_{i,-1} (\beta_{-1}^{(r)} \{\beta_{-1}^{(r)}\}^\intercal) \check{X}_{i,-1}^\intercal e_t = \frac{1}{NR} \sum_{i=1}^{N} e_t^\intercal \check{X}_i \Upsilon \check{X}_i^\intercal e_t \\
&= \frac{1}{NR} \mathrm{tr}(\tilde{X}_t \Upsilon \tilde{X}_t^\intercal) = \frac{1}{NR} \|\tilde{X}_t \Upsilon^{1/2}\|_F^2 \leq \frac{1}{NR} \|\Upsilon^{1/2}\|_F^2 \|\tilde{X}_t\|_2^2 = \frac{1}{NR} \mathrm{tr}(\Upsilon) \|\tilde{X}_t\|_2^2 \\
&\leq \frac{1}{NR} \left(\sum_{r=1}^{R} \|\beta_{-1}^{(r)}\|_2^2 \right) \|\tilde{X}_t\|_2^2 = O\left(\frac{s_B c_B^2 T}{R} \right).
\end{aligned} \tag{B.16}$$

Next, we have that,

$$\begin{aligned}
&\frac{1}{NR} \sum_{i=1}^{N} \sum_{r=1}^{R} \check{\mu}_{i,r,t_1} \check{\delta}_{i,r,t_2} = \frac{1}{NR} \sum_{i=1}^{N} \sum_{r=1}^{R} \check{\mu}_{i,r,t_1} (\delta_{i,r,t_2} - \bar{\delta}_{r,t_2}) \\
&= \frac{1}{NR} \sum_{i=1}^{N} \sum_{r=1}^{R} \check{\mu}_{i,r,t_1} \delta_{i,r,t_2} - \frac{1}{NR} \sum_{r=1}^{R} \bar{\delta}_{r,t_2} (\sum_{i=1}^{N} \check{\mu}_{i,r,t_1}) = \frac{1}{NR} \sum_{i=1}^{N} \sum_{r=1}^{R} \check{\mu}_{i,r,t_1} \delta_{i,r,t_2} \qquad \text{(B.17)} \\
&= \frac{1}{NR} \sum_{i=1}^{N} (\check{\mu}_{i,1,t_1}, \ldots, \check{\mu}_{i,R,t_1})^\intercal \Sigma_R^{1/2} \tilde{\epsilon}_{i,t_2} + \frac{1}{NR} \sum_{i=1}^{N} \sum_{r=1}^{R} \check{\mu}_{i,r,t_1} e_{t_2}^\intercal G_i \Sigma_\zeta^{1/2} \tilde{\zeta}_{i,r},
\end{aligned}$$

where $\tilde{\epsilon}_{i,t_2} = \Sigma_R^{-1/2}\epsilon_{i,t_2}$. In addition, by Conditions (C1) and (C2), we have $\|\Sigma_R\|_2 = O(1)$, and $\|e_{t_2}^\intercal G_i \Sigma_\zeta^{1/2}\|_2 = O(1)$ simultaneously for $1 \le t_2 \le T$. Similar to (B.14) and (B.15), (B.17) is the weighted sum of independent sub-Gaussian variables. Therefore,

$$\frac{1}{(NR)^2}\sum_{i=1}^N \|(\check{\mu}_{i,1,t_1},\ldots,\check{\mu}_{i,R,t_1})^\intercal\|_2^2 = \frac{C_1}{(NR)^2}\sum_{i=1}^N\sum_{r=1}^R \check{\mu}_{i,r,t_1}^2 = O\left(\frac{s_B c_B^2 T}{NR^2}\right).$$

Then we have that, simultaneously for $1 \le t_1, t_2 \le T$,

$$\left|\frac{1}{NR}\sum_{i=1}^N\sum_{r=1}^R \check{\mu}_{i,r,t_1}\check{\delta}_{i,r,t_2}\right| = O_{\mathbb{P}}\left(\left(\frac{s_B c_B^2 T\log T}{NR^2}\right)^{1/2}\right). \tag{B.18}$$

By (B.16) and (B.18), we prove the second result in (B.11).

The proof of the third result in (B.11) is essentially the same as that for the second one.

Combing the three results together, we prove (B.11) .

Step 2: Next, we establish the convergence rate of $[\hat{\Sigma}_R]_{r_1,r_2}$ for $r_1 \neq r_2$.

We first show that $\|\hat{\Sigma}_1 - \Sigma_1\|_{\max} = O_{\mathbb{P}}[c_R + \{\log R/(NT)\}^{1/2}]$, where $\hat{\Sigma}_1$ is as defined in (2.10), and $\Sigma_1 = \Sigma_R + (NT)^{-1}\sum_{i=1}^N \operatorname{tr}(G_i\Sigma_\zeta G_i^\intercal)I_R$. From (B.11), it suffices to show,

$$\left\|\frac{1}{NT}\sum_{i=1}^N \Delta_i^\intercal\Delta_i - \Sigma_1\right\|_{\max} = O_{\mathbb{P}}\left[\{\log R/(NT)\}^{1/2}\right]. \tag{B.19}$$

Note that the definition of Δ_i involves $\check{\delta}_{i,r,t}$, instead of $\delta_{i,r,t}$. We show that, using $\delta_{i,r,t}$ to substitute $\check{\delta}_{i,r,t}$ in Δ_i leads to the same convergence rate. That is, we show that, simultaneously for $1 \le r_1, r_2 \le R$,

$$\left|\frac{1}{NT}\sum_{i=1}^{N}\epsilon_{i,r_1}^{\intercal}\epsilon_{i,r_2}-[\Sigma_R]_{r_1,r_2}\right|=O_{\mathbb{P}}\left(\{\log R/(NT)\}^{1/2}\right),$$
$$\left|\frac{1}{NT}\sum_{i=1}^{N}\zeta_{i,r_1}^{\intercal}G_i^{\intercal}\epsilon_{i,r_2}\right|=O_{\mathbb{P}}\left(\{\log R/(NT^2)\}^{1/2}\right),$$
$$\left|\frac{1}{NT}\sum_{i=1}^{N}\zeta_{i,r_1}^{\intercal}G_i^{\intercal}G_i\zeta_{i,r_2}-\frac{I(r_1=r_2)}{NT}\sum_{i=1}^{N}\mathrm{tr}(G_i\Sigma_\zeta G_i^{\intercal})\right|=O_{\mathbb{P}}\left(\{\log R/(NT^2)\}^{1/2}\right), \tag{B.20}$$

Recall the notations defined in the proofs of (B.11), i.e., $\tilde{\epsilon}_{i,r}=\Sigma_T^{-1/2}\epsilon_{i,r}$ and $\tilde{\zeta}_{i,r}=\Sigma_\zeta^{-1/2}\zeta_{i,r}$. Then (B.20) is equivalent to the following,

$$\left|\frac{1}{NT}\sum_{i=1}^{N}\tilde{\epsilon}_{i,r_1}^{\intercal}\Sigma_T\tilde{\epsilon}_{i,r_2}-[\Sigma_R]_{r_1,r_2}\right|=O_{\mathbb{P}}\left(\{\log R/(NT)\}^{1/2}\right),$$
$$\left|\frac{1}{NT}\sum_{i=1}^{N}\tilde{\zeta}_{i,r_1}^{\intercal}\Sigma_\zeta^{1/2}G_i^{\intercal}\Sigma_T^{1/2}\tilde{\epsilon}_{i,r_2}\right|=O_{\mathbb{P}}\left(\{\log R/(NT^2)\}^{1/2}\right), \tag{B.21}$$
$$\left|\frac{1}{NT}\sum_{i=1}^{N}\tilde{\zeta}_{i,r_1}^{\intercal}\Sigma_\zeta^{1/2}G_i^{\intercal}G_i\Sigma_\zeta^{1/2}\tilde{\zeta}_{i,r_2}-\frac{I(r_1=r_2)}{NT}\sum_{i=1}^{N}\mathrm{tr}(G_i\Sigma_\zeta G_i^{\intercal})\right|=O_{\mathbb{P}}\left(\{\log R/(NT^2)\}^{1/2}\right),$$

By Conditions (C1) and (C2), we have $\|\Sigma_T\|_F^2\leq T\|\Sigma_T\|_2^2=O(T)$, and $\|G_i\Sigma_\zeta^{1/2}\|_F^2\leq\|G_i\|_2^2\|\Sigma_\zeta\|_F\leq 2T\|G_i\|_{\max}^2\|\Sigma_\zeta\|_2=O(1)$. We apply Lemma 5 repeatedly to prove (B.21).

For the first equation in (B.21), let $X=(\tilde{\epsilon}_{1,r_1}^{\intercal},\ldots,\tilde{\epsilon}_{N,r_1}^{\intercal})\in\mathbb{R}^{NT}$, $Y=(\tilde{\epsilon}_{1,r_2}^{\intercal},\ldots,\tilde{\epsilon}_{N,r_2}^{\intercal})\in\mathbb{R}^{NT}$, and $A=\mathrm{diag}(\Sigma_T,\ldots,\Sigma_T)\in\mathbb{R}^{TN\times TN}$. Then we have $\|A\|_F^2=N\|\Sigma_T\|_F^2=O(NT)$. By Lemma 5, the first equation holds.

For the second equation in (B.21), let $X=(\tilde{\zeta}_{1,r_1}^{\intercal},0_{T-2}^{\intercal},\tilde{\zeta}_{1,r_2}^{\intercal},0_{T-2}^{\intercal},\ldots,\tilde{\epsilon}_{N,r_1}^{\intercal},0_{T-2}^{\intercal})\in\mathbb{R}^{NT}$, $Y=(\tilde{\epsilon}_{1,r_2}^{\intercal},\ldots,\tilde{\epsilon}_{N,r_2}^{\intercal})\in\mathbb{R}^{NT}$, and $A=\mathrm{diag}(A_1,\ldots,A_N)\in\mathbb{R}^{TN\times TN}$, where 0_{T-2} denotes the zero vector with length $T-2$ and $A_i\in\mathbb{R}^{T\times T}$, in which the first two rows are $\Sigma_\zeta^{1/2} G_i^\intercal \Sigma_T^{1/2}$, whereas the other rows are 0. Then $\|A\|_F^2 = \sum_{i=1}^N \|\Sigma_\zeta^{1/2} G_i^\intercal \Sigma_T^{1/2}\|_F^2 \leq \sum_{i=1}^N \|\Sigma_\zeta^{1/2} G_i^\intercal\|_F^2 \|\Sigma_T^{1/2}\|_2^2 = O(N)$. Therefore, by Lemma 5, the second equation holds.

For the third equation in (B.21), let $X = (\tilde{\zeta}_{1,r_1}^\intercal, \ldots, \tilde{\epsilon}_{N,r_1}^\intercal) \in \mathbb{R}^{2N}$, $Y = (\tilde{\zeta}_{1,r_2}^\intercal, \ldots, \tilde{\epsilon}_{N,r_2}^\intercal) \in \mathbb{R}^{2N}$, and $A = \mathrm{diag}(\Sigma_\zeta^{1/2} G_1^\intercal G_1 \Sigma_\zeta^{1/2}, \ldots, \Sigma_\zeta^{1/2} G_N^\intercal G_N \Sigma_\zeta^{1/2})$. Then we have $\|A\|_F^2 = \sum_{i=1}^N \|\Sigma_\zeta^{1/2} G_i^\intercal G_i \Sigma_\zeta^{1/2}\|_F^2 \leq C \sum_{i=1}^N \|G_i \Sigma_\zeta^{1/2}\|_F^4 = O(N)$. By Lemma 5, the third equation holds.

Together, we prove (B.21), which leads to (B.20).

Next, we turn to prove (B.19). It suffices to show that, using $\check{\delta}_{i,r,t}$ instead of $\delta_{i,r,t}$ in Δ_i leads to the same result. The proof is similar to that of (B.20), which splits $\delta_{i,r,t} = \epsilon_{i,r,T} + e_i^\intercal G_i \zeta_{i,r}$, and proves the three equations similar as those in (B.20). We only prove the equation below, and the other two follow similarly.

$$
\left| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \epsilon_{i,r_1,t}\epsilon_{i,r_2,t} - \left(\epsilon_{i,r_1,t} - \frac{1}{N} \sum_{j=1}^N \epsilon_{j,r_1,t} \right) \left(\epsilon_{i,r_2,t} - \frac{1}{N} \sum_{j=1}^N \epsilon_{j,r_2,t} \right) \right| = o_{\mathbb{P}}\left(\{\log R/(NT)\}^{1/2} \right). \tag{B.22}
$$

We note that,

$$
\begin{aligned}
&\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \epsilon_{i,r_1,t}\epsilon_{i,r_2,t} - \left(\epsilon_{i,r_1,t} - \frac{1}{N} \sum_{j=1}^N \epsilon_{j,r_1,t} \right) \left(\epsilon_{i,r_2,t} - \frac{1}{N} \sum_{j=1}^N \epsilon_{j,r_2,t} \right) \\
&= \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{N} \sum_{i=1}^N \epsilon_{i,r_1,t} \right) \left(\frac{1}{N} \sum_{j=1}^N \epsilon_{j,r_2,t} \right) \\
&= \frac{1}{N^2 T} (\epsilon_{1,r_1}^\intercal, \ldots, \epsilon_{N,r_1}^\intercal)(I_T \otimes 1_{N\times N})(\epsilon_{1,r_2}^\intercal, \ldots, \epsilon_{N,r_2}^\intercal)^\intercal \\
&= \frac{1}{N^2 T} (\tilde{\epsilon}_{1,r_1}^\intercal, \ldots, \tilde{\epsilon}_{N,r_1}^\intercal) A^{1/2} (I_T \otimes 1_{N\times N}) A^{1/2} (\tilde{\epsilon}_{1,r_2}^\intercal, \ldots, \tilde{\epsilon}_{N,r_2}^\intercal)^\intercal,
\end{aligned}
$$

where $\mathrm{l}_{N\times N} \in \mathbb{R}^{N\times N}$ denotes the matrix with all entries equal to 1, and $A = \mathrm{diag}(\{\Sigma_T\}_{i=1}^N) \in \mathbb{R}^{TN\times TN}$. Moreover, we have that,

$$\mathbb{E}\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \epsilon_{i,r_1,t}\epsilon_{i,r_2,t} - \left(\epsilon_{i,r_1,t} - \frac{1}{N}\sum_{j=1}^N \epsilon_{j,r_1,t}\right)\left(\epsilon_{i,r_2,t} - \frac{1}{N}\sum_{j=1}^N \epsilon_{j,r_2,t}\right)\right\}$$
$$= \frac{1}{N}[\Sigma_R]_{r_1,r_2} = O(N^{-1}).$$

Because $\|A^{1/2}(I_T\otimes \mathrm{l}_{N\times N})A^{1/2}\|_F^2 \le \|I_T\otimes \mathrm{l}_{N\times N}\|_F^2\|A\|_2^2 = O(N^2T)$. By applying Lemma 5 again, we obtain that,

$$\frac{1}{N^2T}(\tilde{\epsilon}_{1,r_1}^\intercal,\ldots,\tilde{\epsilon}_{N,r_1}^\intercal)A^{1/2}(I_T\otimes \mathrm{l}_{N\times N})A^{1/2}(\tilde{\epsilon}_{1,r_2}^\intercal,\ldots,\tilde{\epsilon}_{N,r_2}^\intercal)^\intercal = O(N^{-1}) + O_{\mathbb{P}}\left(\{\log R/(NT)\}^{1/2}\right),$$

which implies (B.22) by the condition that $\theta_{N,T,R,B} = o(1)$. Therefore, we have $\|\hat{\Sigma}_1 - \Sigma_1\|_{\max} = O_{\mathbb{P}}\left(c_R + \{\log R/(NT)\}^{1/2}\right)$. From (2.10), this implies that,

$$\max_{r_1\neq r_2}|[\hat{\Sigma}_R]_{r_1,r_2} - [\Sigma_R]_{r_1,r_2}| = O_{\mathbb{P}}\left(c_R + \{\log R/(NT)\}^{1/2}\right).$$

Step 3: Next, we establish the convergence rate of $\hat{\Sigma}_T$. That is, we aim to show that $\|\hat{\Sigma}_T - \Sigma_T\|_{\max} = O_{\mathbb{P}}[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}]$.

By the definition of $\hat{\Sigma}_T$ in (2.12), we have that,

$$\|\hat{\Sigma}_T - \Sigma_T\|_{\max} \le \left\|\frac{1}{R}\sum_{(r_1,r_2)\in\mathcal{S}}\left(\frac{1}{[\hat{\Sigma}_R]_{r_1,r_2}} - \frac{1}{[\Sigma_R]_{r_1,r_2}}\right)\frac{1}{N}\sum_{i=1}^N(\check{y}_{i,r_1,1},\ldots,\check{y}_{i,r_1,T})^\intercal(\check{y}_{i,r_2,1},\ldots,\check{y}_{i,r_2,T})\right\|_{\max}$$
$$+\left\|\left\{\frac{1}{NR}\sum_{(r_1,r_2)\in\mathcal{S}}\frac{1}{[\Sigma_R]_{r_1,r_2}}\sum_{i=1}^N(\check{y}_{i,r_1,1},\ldots,\check{y}_{i,r_1,T})^\intercal(\check{y}_{i,r_2,1},\ldots,\check{y}_{i,r_2,T})\right\} - \Sigma_T\right\|_{\max} \quad \text{(B.23)}$$

We next bound the two terms in (B.23) separately.

For the first term in (B.23), from Condition (C3) and the convergence rate of $[\hat{\Sigma}_R]_{r_1,r_2}$ for $r_1 \neq r_2$, we have, $\min_{(r_1,r_2)\in\mathcal{S}} |[\Sigma_R]_{r_1,r_2}| \geq c$ and $\min_{(r_1,r_2)\in\mathcal{S}} |[\hat{\Sigma}_R]_{r_1,r_2}| \geq c$ with probability tending to 1 for some constant $c > 0$. In the following, we establish the convergence rate under the event $\{\min_{(r_1,r_2)\in\mathcal{S}} |[\Sigma_R]_{r_1,r_2}| \geq c, \text{ and } \min_{(r_1,r_2)\in\mathcal{S}} |[\hat{\Sigma}_R]_{r_1,r_2}| \geq c\}$. We have, simultaneously for $(r_1, r_2) \in \mathcal{S}$,

$$\left| \frac{1}{[\hat{\Sigma}_R]_{r_1,r_2}} - \frac{1}{[\Sigma_R]_{r_1,r_2}} \right| = O_{\mathbb{P}}[c_R + \{\log R/(NT)\}^{1/2}]. \tag{B.24}$$

Based on Step 1 and following the proof of Step 2, we have,

$$\left\| \frac{1}{N} \sum_{i=1}^{N} (\check{y}_{i,r_1,1}, \ldots, \check{y}_{i,r_1,T})^\intercal (\check{y}_{i,r_2,1}, \ldots, \check{y}_{i,r_2,T}) \right\|_{\max} = O_{\mathbb{P}}(1), \tag{B.25}$$

simultaneously for $(r_1, r_2) \in \mathcal{S}$. By (B.24) and (B.25), we have,

$$\begin{aligned} &\left\| \frac{1}{R} \sum_{(r_1,r_2)\in\mathcal{S}} \left(\frac{1}{[\hat{\Sigma}_R]_{r_1,r_2}} - \frac{1}{[\Sigma_R]_{r_1,r_2}} \right) \frac{1}{N} \sum_{i=1}^{N} (\check{y}_{i,r_1,1}, \ldots, \check{y}_{i,r_1,T})^\intercal (\check{y}_{i,r_2,1}, \ldots, \check{y}_{i,r_2,T}) \right\|_{\max} \\ &= O_{\mathbb{P}} \left(c_R + \{\log R/(NT)\}^{1/2} \right), . \end{aligned} \tag{B.26}$$

For the second term in (B.23), since $\min_{(r_1,r_2)\in\mathcal{S}} |[\Sigma_R]_{r_1,r_2}| \geq c$, similar to the proof of Step 1, it suffices to show that,

$$\begin{aligned} &\left\| \left\{ \frac{1}{NR} \sum_{(r_1,r_2)\in\mathcal{S}} \frac{1}{[\Sigma_R]_{r_1,r_2}} \sum_{i=1}^{N} (\check{\delta}_{i,r_1,1}, \ldots, \check{\delta}_{i,r_1,T})^\intercal (\check{\delta}_{i,r_2,1}, \ldots, \check{\delta}_{i,r_2,T}) \right\} - \Sigma_T \right\|_{\max} \\ &= O_{\mathbb{P}} \left(\{\log T/(NR)\}^{1/2} \right). \end{aligned} \tag{B.27}$$

Similar to Step 2, we only need to prove the version substituting $\delta_{i,r,t}$ by $\check{\delta}_{i,r,t}$, i.e.,

$$\left\|\left\{\frac{1}{NR}\sum_{(r_1,r_2)\in\mathcal{S}}\frac{1}{[\Sigma_R]_{r_1,r_2}}\sum_{i=1}^{N}(\delta_{i,r_1,1},\ldots,\delta_{i,r_1,T})^\intercal(\delta_{i,r_2,1},\ldots,\delta_{i,r_2,T})\right\}-\Sigma_T\right\|_{\max} \tag{B.28}$$
$$= O_{\mathbb{P}}\left(\{\log T/(NR)\}^{1/2}\right).$$

Again, it suffices to show simultaneously, for $1\le t_1,t_2\le T$,

$$\begin{aligned}
&\left|\frac{1}{NR}\sum_{i=1}^{N}\sum_{(r_1,r_2)\in\mathcal{S}}\frac{1}{[\Sigma_R]_{r_1,r_2}}\epsilon_{i,r_1,t_1}\epsilon_{i,r_2,t_2}-[\Sigma_T]_{t_1,t_2}\right| = O_{\mathbb{P}}\left(\{\log T/(NR)\}^{1/2}\right),\\
&\left|\frac{1}{NR}\sum_{i=1}^{N}\sum_{(r_1,r_2)\in\mathcal{S}}\frac{1}{[\Sigma_R]_{r_1,r_2}}\epsilon_{i,r_1,t_1}e_{t_2}^\intercal G_i\zeta_{i,r_2}\right| = O_{\mathbb{P}}\left(\{\log T/(NR)\}^{1/2}\right),\\
&\left|\frac{1}{NR}\sum_{i=1}^{N}\sum_{(r_1,r_2)\in\mathcal{S}}\frac{1}{[\Sigma_R]_{r_1,r_2}}\zeta_{i,r_1}^\intercal G_i^\intercal e_{t_1}e_{t_2}^\intercal G_i\zeta_{i,r_2}\right| = O_{\mathbb{P}}\left(\{\log T/(NR)\}^{1/2}\right).
\end{aligned} \tag{B.29}$$

We first simplify the summation $\sum_{(r_1,r_2)\in\mathcal{S}}$. We put the index pairs in S in any ordering, i.e., $S=\left\{(r_1^{(m)},r_2^{(m)}), m=1,\ldots,K\right\}$. Define two matrices, $U_1,U_2\in\mathbb{R}^{K\times R}$, such that $(U_1)_{m,r_1^{(m)}}=1$, $(U_2)_{m,r_2^{(m)}}=1$, and the other entries of U_1, U_2 are all 0. Recall that $\tilde{\epsilon}_{i,t}=\Sigma_R^{-1/2}\epsilon_{i,t}$, and $\tilde{\zeta}_{i,r}=\Sigma_\zeta^{-1/2}\zeta_{i,r}$. For $1\le t\le T$, $\tilde{\epsilon}_t=(\tilde{\epsilon}_{1,t}^\intercal,\ldots,\tilde{\epsilon}_{N,t}^\intercal)^\intercal$ has independent entries. Then we have, for $d=1,2$, the m-th entry of $U_d\Sigma_R^{1/2}\tilde{\epsilon}_{i,t_d}$ is $\epsilon_{i,r_d^{(m)},t_d}$. Similarly, for $d=1,2$, denote $H_{i,d}=\mathrm{diag}(\{e_{t_d}^\intercal G_i\Sigma_\zeta\}_{m=1}^R)\in\mathbb{R}^{R\times 2R}$, and $\tilde{\zeta}_i=(\tilde{\zeta}_{i,1}^\intercal,\ldots,\tilde{\zeta}_{i,R}^\intercal)^\intercal\in\mathbb{R}^{2R}$. We then have that the mth entry of $U_dH_{i,d}\tilde{\zeta}_i$ is equal to $e_{t_d}^\intercal G_i\zeta_{i,r_d^{(m)}}$. In addition, denote

$$L=\mathrm{diag}\left\{\frac{1}{[\Sigma_R]_{r_1^{(1)},r_2^{(1)}}},\ldots,\frac{1}{[\Sigma_R]_{r_1^{(m)},r_2^{(m)}}},\ldots,\frac{1}{[\Sigma_R]_{r_1^{(K)},r_2^{(K)}}}\right\}.$$

Then (B.29) is equivalent to

$$\begin{aligned}
&\left|\frac{1}{NR}\sum_{i=1}^{N}(U_1\Sigma_R^{1/2}\tilde{\epsilon}_{i,t_1})^\intercal L(U_2\Sigma_R^{1/2}\tilde{\epsilon}_{i,t_2})-[\Sigma_T]_{t_1,t_2}\right| = O_{\mathbb{P}}\left(\{\log T/(NR)\}^{1/2}\right),\\
&\left|\frac{1}{NR}\sum_{i=1}^{N}(U_1\Sigma_R^{1/2}\tilde{\epsilon}_{i,t_1})^\intercal L(U_2 H_{i,2}\tilde{\zeta}_i)\right| = O_{\mathbb{P}}\left(\{\log T/(NR)\}^{1/2}\right),\\
&\left|\frac{1}{NR}\sum_{i=1}^{N}(U_1 H_{i,1}\tilde{\zeta}_i)^\intercal L(U_2 H_{i,2}\tilde{\zeta}_i)\right| = O_{\mathbb{P}}\left(\{\log T/(NR)\}^{1/2}\right).
\end{aligned} \tag{B.30}$$

By the above definitions and Condition (C1), we have that $\|U_1\|_2=\|U_2\|_2=1$, $\|L\|_2=O(1)$, $\|H_{i,d}\|_2=O(1)$, and $\|H_{i,d}\|_F=O(R^{1/2})$. By Condition (C2), we have $\|\Sigma_R^{1/2}\|_F^2\le R\|\Sigma_R^{1/2}\|_2^2=O(R)$. Again, we apply Lemma 5 to prove (B.30).

Specifically, for the first equation in (B.30), let $X=(\tilde{\epsilon}_{1,t_1}^\intercal,\dots,\tilde{\epsilon}_{N,t_1}^\intercal)\in\mathbb{R}^{NR}$, $Y=(\tilde{\epsilon}_{1,t_2}^\intercal,\dots,\tilde{\epsilon}_{N,t_2}^\intercal)\in\mathbb{R}^{NR}$, and $A=\mathrm{diag}(\{\Sigma_R^{1/2}U_1^\intercal LU_2\Sigma_R^{1/2}\}_{i=1}^N)\in\mathbb{R}^{NR\times NR}$, for which

$$\|A\|_F^2=N\|\Sigma_R^{1/2}U_1^\intercal LU_2\Sigma_R^{1/2}\|_F^2\le N\|\Sigma_R^{1/2}\|_F^2\|U_1\|_2^2\|L\|_2^2\|U_2\|_2^2\|\Sigma_R^{1/2}\|_2^2=O(NR).$$

Then the first equation holds by Lemma 5.

For the second equation in (B.30), let $X=(\tilde{\epsilon}_{1,t_1}^\intercal,0_R^\intercal,\tilde{\epsilon}_{2,t_1}^\intercal,0_R^\intercal,\dots,\tilde{\epsilon}_{N,t_1}^\intercal,0_R^\intercal)\in\mathbb{R}^{2NR}$, $Y=(\tilde{\zeta}_1,\dots,\tilde{\zeta}_N)\in\mathbb{R}^{2NR}$, and $A=\mathrm{diag}(\{\tilde{\Sigma}_R^{1/2}U_1^\intercal LU_2H_{i,2}\}_{i=1}^N)\in\mathbb{R}^{2NR\times 2NR}$, where $\tilde{\Sigma}_R^{1/2}\in\mathbb{R}^{2R\times R}$ has the first R rows equal to $\Sigma_R^{1/2}$ and the last R rows equal to 0. Then we also have $\|A\|_F^2=O(NR)$. So the second equation holds by Lemma 5.

For the last equation in (B.30), let $X=Y=(\tilde{\zeta}_1,\dots,\tilde{\zeta}_N)\in\mathbb{R}^{2NR}$, and $A=\mathrm{diag}(\{H_{i,1}^\intercal U_1^\intercal\ LU_2H_{i,2}\}_{i=1}^N)\in\mathbb{R}^{2NR\times 2NR}$. Again we have $\|A\|_F^2=O(NR)$, so the last equation holds by Lemma 5.

Therefore, (B.27) and (B.28) hold. Together with (B.26), we obtain the convergence rate of $\hat{\Sigma}_T$.

Step 4: Next, we establish the convergence rate of $\hat{\kappa}$.

Note that

$$\begin{aligned}
\hat{\kappa} - \kappa &= \frac{\sum_{i=1}^N u_i^\intercal \hat{\Sigma}_{3,i} u_i}{\sum_{i=1}^N u_i^\intercal \hat{\Sigma}_T u_i} - \frac{\sum_{i=1}^N u_i^\intercal \Sigma_{3,i} u_i}{\sum_{i=1}^N u_i^\intercal \Sigma_T u_i} \\
&= \left\{ \frac{\sum_{i=1}^N u_i^\intercal \hat{\Sigma}_{3,i} u_i}{\sum_{i=1}^N u_i^\intercal \hat{\Sigma}_T u_i} - \frac{\sum_{i=1}^N u_i^\intercal \hat{\Sigma}_{3,i} u_i}{\sum_{i=1}^N u_i^\intercal \Sigma_T u_i} \right\} + \left\{ \frac{\sum_{i=1}^N u_i^\intercal \hat{\Sigma}_{3,i} u_i}{\sum_{i=1}^N u_i^\intercal \Sigma_T u_i} - \frac{\sum_{i=1}^N u_i^\intercal \Sigma_{3,i} u_i}{\sum_{i=1}^N u_i^\intercal \Sigma_T u_i} \right\} \\
&= \left(\sum_{i=1}^N u_i^\intercal \hat{\Sigma}_{3,i} u_i \right) \left\{ \frac{1}{\sum_{i=1}^N u_i^\intercal \hat{\Sigma}_T u_i} - \frac{1}{\sum_{i=1}^N u_i^\intercal \Sigma_T u_i} \right\} + \frac{\sum_{i=1}^N u_i^\intercal (\hat{\Sigma}_{3,i} - \Sigma_{3,i}) u_i}{\sum_{i=1}^N u_i^\intercal \Sigma_T u_i}
\end{aligned}$$

By the definition of u_i, we have that,

$$\begin{aligned}
|u_i^\intercal (\hat{\Sigma}_T - \Sigma_T) u_i| &\le \|u_i^\intercal\|_2 \|\hat{\Sigma}_T - \Sigma_T\|_2 \|u_i\|_2 \\
&= O_{\mathbb{P}} \left(T[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}] \right).
\end{aligned}$$

By Condition (C1), we also have $c_1^{-1} \le u_i^\intercal \Sigma_T u_i \le c_1$. Therefore, by the condition that $(T[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}] = o(1)$, we have that,

$$\frac{N}{2c_1} \le \sum_{i=1}^N u_i^\intercal \hat{\Sigma}_T u_i \le 2Nc_1.$$

Similar to Steps 1 and 3, using the definition of $\hat{\Sigma}_{3,i}$, we have that $\mu_{i,r,t}$'s are negligible. By repeatedly applying Lemma 5, we can obtain that

$$\left| \sum_{i=1}^N \{u_i^\intercal (\hat{\Sigma}_{3,i} - \Sigma_{3,i}) u_i\} \right| = O_{\mathbb{P}} \left(T[\{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}] \right),$$

which implies that,

$$\left|\frac{\sum_{i=1}^{N} u_i^\intercal(\hat{\Sigma}_{3,i}-\Sigma_{3,i})u_i}{\sum_{i=1}^{N} u_i^\intercal \Sigma_T u_i}\right| = O_{\mathbb{P}}\left(T[\{\log T/(NR)\}^{1/2}+c_R+s_B c_B^2 TR^{-1}]\right).$$

Therefore, we obtain that,

$$|\hat{\kappa}-\kappa| = O_{\mathbb{P}}\left(T[\{\log R/(NT)\}^{1/2}+\{\log T/(NR)\}^{1/2}+c_R+s_B c_B^2 TR^{-1}]\right).$$

Step 5: Next, we establish the convergence rate of $\hat{\Sigma}_\zeta$.

We first show that $\|v_{i,d}\|_2 = O(1)$ for $d=1,2$. We only prove the result for $d=1$, and the proof for $d=2$ follows similarly. Denote $G_{i,2}=xs_1$, and $G_{i,1}=as_1+bs_2$, where $s_1, s_2 \in \mathbb{R}^T$ are unit vectors and are orthogonal to each other. Then we have $x^2=O(T)$ by Condition (C1). By construction, we have $v_{i,1}=b^{-1}s_2$, so we only need to show that $b^{-1}=O(1)$. Note that

$$G_i^\intercal G_i = \begin{pmatrix} x^2 & ax \\ ax & a^2+b^2 \end{pmatrix}.$$

By Condition (C1), there exists some positive constant c_0, such that

$$x^2+a^2+b^2-\{(x^2+a^2+b^2)^2-4x^2b^2\}^{1/2} = \frac{4x^2b^2}{x^2+a^2+b^2+\{(x^2+a^2+b^2)^2-4x^2b^2\}^{1/2}} \geq c_0,$$

which implies that $b^2 \geq c_0/2$ so $b^{-1}=O(1)$.

Next, we obtain the convergence rate. By Condition (C2) and Step 3, we have that, simultaneously for $1 \leq i \leq N$,

$$|v_{i,j_1}^\intercal \Sigma_T v_{i,j_2}| \leq \|v_{i,j_1}^\intercal\|_2 \|\Sigma_T\|_2 \|v_{i,j_2}\|_2 = O(1),$$

and

$$\begin{aligned}|v_{i,j_1}^\intercal(\hat{\Sigma}_T - \Sigma_T)v_{i,j_2}| &\leq \|v_{i,j_1}^\intercal\|_2 \|\hat{\Sigma}_T - \Sigma_T\|_2 \|v_{i,j_2}\|_2 \\ &= O_{\mathbb{P}}\left(T[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}]\right).\end{aligned}$$

Combining the results above, along with the convergence rate for κ, we obtain that,

$$|\hat{\kappa}(v_{i,j_1}^\intercal \hat{\Sigma}_T v_{i,j_2}) - \kappa(v_{i,j_1}^\intercal \Sigma_T v_{i,j_2})| = O_{\mathbb{P}}(T[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}]).$$

Following similar steps in Steps 1-3, we have that,

$$\frac{1}{N}\sum_{i=1}^{N}\{v_{i,j_1}^\intercal \hat{\Sigma}_{3,i} v_{i,j_2} - v_{i,j_1}^\intercal \Sigma_{3,i} v_{i,j_2}\} = O_{\mathbb{P}}\left(T[\{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}]\right).$$

By the definition of $\hat{\Sigma}_\zeta$ in (2.15), we have that,

$$\|\hat{\Sigma}_\zeta - \Sigma_\zeta\|_{\max} = O_{\mathbb{P}}\left(T[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}]\right).$$

Step 6: Finally, we establish the convergence rate of $[\hat{\Sigma}_R]_{r,r}$ and $\hat{\Sigma}^{(r)}$.

By (2.16), (B.21), and Step 4, we have that,

$$\max_r |[\hat{\Sigma}_R]_{r,r} - [\Sigma_R]_{r,r}| = O_{\mathbb{P}}\left(T[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}]\right).$$

Next, by Condition (C2), we have that,

$$\|G(I_N \otimes \hat{\Sigma}_\zeta)G^\intercal - G(I_N \otimes \Sigma_\zeta)G^\intercal\|_{\max} = \max_{1\le i\le N} \|G_i(\hat{\Sigma}_\zeta - \Sigma_\zeta)G_i^\intercal\|_{\max}$$

$$\le \max_{1\le i\le N} \|G_i\|_{\max}^2 \|\hat{\Sigma}_\zeta - \Sigma_\zeta\|_{1,1} = O_{\mathbb{P}}\left(T[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}]\right).$$

Then by Condition (C2), Steps 3 and 4, for $r = 1, \ldots, R$, we have that,

$$\|[\hat{\Sigma}_R]_{r,r}\hat{\Sigma}_T - [\Sigma_R]_{r,r}\Sigma_T\|_{\max} = |[\hat{\Sigma}_R]_{r,r}| \|\hat{\Sigma}_T - \Sigma_T\|_{\max} + |[\hat{\Sigma}_R]_{r,r} - [\Sigma_R]_{r,r}| \|\Sigma_T\|_{\max}$$

$$= O_{\mathbb{P}}\left(T[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}]\right).$$

Therefore, simultaneously for $r = 1, \ldots, R$, we obtain that,

$$\|\hat{\Sigma}^{(r)} - \Sigma^{(r)}\|_{\max} = O_{\mathbb{P}}\left(T[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}]\right),$$

This completes the proof of Theorem 1. □

B.2 Proof of Theorem 2

By Lemma 6, we have $J^o = O_{\mathbb{P}}(\log \tilde{p})$. Therefore, it suffices to show that

$$\max_{r,j} |J_{r,j}^o - J_{r,j}^d| = o_{\mathbb{P}}\{(\log \tilde{p})^{-1/2}\}. \tag{B.31}$$

Again, we prove (B.31) under the sub-Gaussian assumption (C7), then the result holds under the normality condition.

In Step 2 of the proof of Lemma 6, we have showed that $\|\{\Sigma^{(r)}\}^{-1}\|_2 = O(1)$ uniformly for $r = 1, \ldots, R$. Since $\Sigma^{(r)}$ and $\hat{\Sigma}^{(r)}$ are block diagonal, we have $\|\Sigma^{(r)} - \hat{\Sigma}^{(r)}\|_2 =$

$O_{\mathbb{P}}(T\theta_{N,T,R,B})$, where we recall $\theta_{N,T,R,B} = T[\{\log R/(NT)\}^{1/2} + \{\log T/(NR)\}^{1/2} + c_R + s_B c_B^2 T R^{-1}]$. Therefore, we have that, uniformly for $r = 1, \ldots, R$,

$$\|\{\Sigma^{(r)}\}^{-1} - \{\hat{\Sigma}^{(r)}\}^{-1}\|_2 \le \|\Sigma^{(r)} - \hat{\Sigma}^{(r)}\|_2 \|\{\Sigma^{(r)}\}^{-1}\|_2 \|\{\hat{\Sigma}^{(r)}\}^{-1}\|_2 = O_{\mathbb{P}}(T\theta_{N,T,R,B}).$$

By Condition (C1), we have that,

$$\|X^{\intercal}(\{\Sigma^{(r)}\}^{-1} - \{\hat{\Sigma}^{(r)}\}^{-1})X\|_2 \le \|X\|_2^2 \|\{\Sigma^{(r)}\}^{-1} - \{\hat{\Sigma}^{(r)}\}^{-1}\|_2 = O_{\mathbb{P}}(NT^2\theta_{N,T,R,B}).$$

Again, by Step 2 of the proof of Lemma 6, we have $\|(X^{\intercal}\{\Sigma^{(r)}\}^{-1}X)^{-1}\|_2 = O(N^{-1}T^{-1})$. Similarly, we have that $\|(X^{\intercal}\{\hat{\Sigma}^{(r)}\}^{-1}X)^{-1}\|_2 = O(N^{-1}T^{-1})$. Then,

$$\begin{aligned} &\|(X^{\intercal}\{\Sigma^{(r)}\}^{-1}X)^{-1} - (X^{\intercal}\{\hat{\Sigma}^{(r)}\}^{-1}X)^{-1}\|_2 \\ &\quad \le \|(X^{\intercal}\{\Sigma^{(r)}\}^{-1}X)^{-1}\|_2 \|X^{\intercal}(\{\Sigma^{(r)}\}^{-1} - \{\hat{\Sigma}^{(r)}\}^{-1})X\|_2 \|(X^{\intercal}\{\hat{\Sigma}^{(r)}\}^{-1}X)^{-1}\|_2 \\ &\quad = O_{\mathbb{P}}(N^{-1}\theta_{N,T,R,B}), \end{aligned} \tag{B.32}$$

uniformly for $r = 1, \ldots, R$.

Moreover, for the difference between $(\hat{\beta}^o)^{(r)}$ and $(\hat{\beta}^d)^{(r)}$, we have that,

$$(\hat{\beta}^o)^{(r)} - (\hat{\beta}^d)^{(r)} = \{(X^{\intercal}\{\Sigma^{(r)}\}^{-1}X)^{-1}X^{\intercal}\{\Sigma^{(r)}\}^{-1} - (X^{\intercal}\{\hat{\Sigma}^{(r)}\}^{-1}X)^{-1}X^{\intercal}\{\hat{\Sigma}^{(r)}\}^{-1}\}(G\zeta_r + \epsilon_r).$$

By the fact that $\|\{\Sigma^{(r)}\}^{-1}\|_2 = O(1)$, Condition (C1), and (B.32), we have that,

$$\|(X^{\intercal}\{\Sigma^{(r)}\}^{-1}X)^{-1}X^{\intercal}\{\Sigma^{(r)}\}^{-1} - (X^{\intercal}\{\hat{\Sigma}^{(r)}\}^{-1}X)^{-1}X^{\intercal}\{\hat{\Sigma}^{(r)}\}^{-1}\|_2 = O_{\mathbb{P}}\left(N^{-1/2}T^{1/2}\theta_{N,T,R,B}\right).$$

Denote $G\zeta_r = G(I_N \otimes \Sigma_\zeta)\tilde{\zeta}_r$, and $\epsilon_r = \text{diag}(\{\Sigma_T^{1/2}\}_{i=1}^N)\tilde{\epsilon}_r$, where $\tilde{\zeta}_r$ and $\tilde{\epsilon}_r$ have independent entries. Then we have that,

$$(\hat{\beta}^o)^{(r)} - (\hat{\beta}^d)^{(r)} = \Big\{(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}X^\intercal\{\Sigma^{(r)}\}^{-1} - (X^\intercal\{\hat{\Sigma}^{(r)}\}^{-1}X)^{-1}X^\intercal\{\hat{\Sigma}^{(r)}\}^{-1}\Big\}$$
$$\Big\{G(I_N \otimes \Sigma_\zeta^{1/2})\tilde{\zeta}_r + \text{diag}(\{\Sigma_T^{1/2}\}_{i=1}^N)\tilde{\epsilon}_r\Big\}.$$

By Conditions (C1) and (C2), $\|G(I_N \otimes \Sigma_\zeta^{1/2})\|_2 = O(1)$, and $\|\text{diag}(\{\Sigma_T^{1/2}\}_{i=1}^N)\|_2 = O(1)$. So each entry of $((\hat{\beta}^o)^{(r)} - (\hat{\beta}^d)^{(r)})$ is also sub-Gaussian with the variance of order $O_{\mathbb{P}}(N^{-1}T\theta^2_{N,T,R,B})$. By the maximal inequality for sub-Gaussian variables, we have,

$$\max_{r,j} |(\hat{\beta}^o)_j^{(r)} - (\hat{\beta}^d)_j^{(r)}| = O_{\mathbb{P}}\{N^{-1/2}T^{1/2}(\log \tilde{p})^{1/2}\theta_{N,T,R,B}\}, \tag{B.33}$$

By (B.32), (B.33), and $\min_{r,j} |\{(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}\}_{j,j}| \geq c(NT)^{-1}$ for some constant $c > 0$, we obtain that,

$$\max_{r,j} |J^o_{r,j} - J^d_{r,j}| = O_{\mathbb{P}}\{T(\log \tilde{p})^{1/2}\theta_{N,T,R,B}\}.$$

By Condition (C5), we prove (B.31).

This completes the proof of Theorem 2. □

B.3 Proof of Theorem 3

First, we define

$$J^1 = \max_{r,j} \frac{((\hat{\beta}^d)_j^{(r)} - \beta_j^{(r)})^2}{[(X^\intercal\{\hat{\Sigma}^{(r)}\}^{-1}X)^{-1}]_{j,j}}.$$

By Lemma 4, (B.32), (B.33), and Condition (C5), we have that,

$$\mathbb{P}\left(J^1 \leq 2\log\tilde{p} - \frac{1}{2}\log\log\tilde{p}\right) \to 1.$$

By the fact that $\{\beta_j^{(r)}\} \in \mathcal{U}(2\sqrt{2})$, we have,

$$\max_{r,j} \frac{\{\beta_j^{(r)}\}^2}{\{(X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}\}_{j,j}} \geq 8\log\tilde{p}, \text{ and } \max_{r,j} \frac{\{\beta_j^{(r)}\}^2}{\{(X^\intercal\{\hat{\Sigma}^{(r)}\}^{-1}X)^{-1}\}_{j,j}} \leq 2J^1 + 2J^d.$$

Using (B.32) again, we obtain that,

$$\mathbb{P}(J^d \geq q_\alpha + 2\log\tilde{p} - \log\log\tilde{p}) \to 1,$$

as $(NT, \tilde{p}) \to \infty$. This completes the proof of Theorem 3. □

B.4 Proof of Theorem 4

By (B.31) and the assumption on S_ρ in Theorem 4, we have that, with probability tending to 1,

$$\sum_{r,j} I\{|J_{r,j}^d| > (2\log\tilde{p})^{1/2}\} \geq \{1/(\pi^{1/2}\alpha) + \delta\}(\log\tilde{p})^{1/2}.$$

Similarly as the proof of Lemma 7, we have $\mathbb{P}(\hat{\tau}$ exists in $[0, t_{\tilde{p}}]) = 1$, so we focus on the event $\{\hat{\tau}$ exists in $[0, t_{\tilde{p}}]\}$. Then it suffices to show that, with probability tending to 1,

$$\max_{0\leq\tau\leq t_{\tilde{p}}} \left| \frac{\sum_{(r,j)\in\mathcal{H}_0}\{I(|J_{r,j}^d| \geq \tau) - G(\tau)\}}{\tilde{p}_0 G(\tau)} \right| \to 0.$$

Following the proofs of Lemma 7, (B.31), and the fact that $G[\tau+o\{(\log\tilde{p})^{-1/2}\}]/G(\tau) = 1+o(1)$ uniformly in $0 \le \tau \le \sqrt{2\log\tilde{p}}$, we complete the proof of Theorem 4. $\square$

B.5 Proof of Theorem 5

Recall that, by the proofs of Theorem 1 and Theorem 2, we have the following results under sub-Gaussian Assumption (C7),

$$\max_{r,j} |J^o_{r,j} - J^d_{r,j}| = o_{\mathbb{P}}\{(\log\tilde{p})^{-1/2}\}.$$

Therefore, we focus on the oracle case for $J^o = \max_{r=1,\ldots,R,j=1,\ldots,2p+2}(J^o_{r,j})^2$ under the sub-Gaussian condition. Then, by the proofs of Theorems 2 to 4, the result for the data-driven case follows.

For (i), we note that, under H_0,

$$\begin{aligned}(\hat{\beta}^o)^{(r)} &= \beta^{(r)} + (X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}X^\intercal\{\Sigma^{(r)}\}^{-1}(G\zeta_r + \epsilon_r)\\ &= \beta^{(r)} + (X^\intercal\{\Sigma^{(r)}\}^{-1}X)^{-1}X^\intercal\{\Sigma^{(r)}\}^{-1}\{G(I_N \otimes \Sigma_\zeta^{1/2})\tilde{\zeta}_r + \mathrm{diag}(\{\Sigma_T^{1/2}\}_{i=1}^N)\tilde{\epsilon}_r\},\end{aligned}$$

where $\tilde{\epsilon}_r$ and $\tilde{\zeta}_r$ are as defined in the proof of Theorem 2, and they have independent sub-Gaussian entries with the variance of order $O(1)$ simultaneously for $r = 1,\ldots,R$ by Conditions (C2) and (C7). We arrange the index pairs $\{(r,j) : 1 \le r \le R, 1 \le j \le 2p+2\}$ in any ordering and set them as $\{(r_m, j_m) : m = 1,\ldots,\tilde{p}\}$. Let $\theta_m = (X^\intercal\{\Sigma^{(r_m)}\}^{-1}X)^{-1}_{j_m,j_m}$. In Step 2 of the proof of Lemma 6, we have shown that $|\theta_m|^{-1/2} = O\{(NT)^{1/2}\}$, $\|\{\Sigma^{(r_m)}\}^{-1}\|_2 = O(1)$, and $\|(X^\intercal\{\Sigma^{(r_m)}\}^{-1}X)^{-1}\|_2 = O(N^{-1}T^{-1})$. Then,

$$\|(X^\intercal\{\Sigma^{(r_m)}\}^{-1}X)^{-1}\|_1 = O(\{\max(p,q)\}^{1/2}N^{-1}T^{-1}).$$

Therefore, we have that,

$$\begin{aligned}\|(X^\intercal\{\Sigma^{(r_m)}\}^{-1}X)^{-1}X^\intercal\|_{\max} &\leq \|(X^\intercal\{\Sigma^{(r_m)}\}^{-1}X)^{-1}\|_1\|X^\intercal\|_{\max}\\ &= O\left(\{\max(p,q)\}^{1/2}N^{-1}T^{-1}\right).\end{aligned}$$

Furthermore, by the proofs of Lemma 6, we have that $\|\{\Sigma^{(r_m)}\}^{-1}G(I_N\otimes\Sigma_\zeta^{1/2})\|_2 = O(1)$. Then its block-diagonal form implies that $\|\{\Sigma^{(r_m)}\}^{-1}G(I_N\otimes\Sigma_\zeta^{1/2})\|_1 = O(T^{1/2})$.

By Conditions (C1) and (C2), we have that,

$$\begin{aligned}&\|(X^\intercal\{\Sigma^{(r_m)}\}^{-1}X)^{-1}X^\intercal\{\Sigma^{(r_m)}\}^{-1}G(I_N\otimes\Sigma_\zeta^{1/2})\|_{\max} = O\left(\{\max(p,q)\}^{1/2}N^{-1}T^{-1/2}\right),\\ &\|(X^\intercal\{\Sigma^{(r_m)}\}^{-1}X)^{-1}X^\intercal\{\Sigma^{(r_m)}\}^{-1}\mathrm{diag}(\{\Sigma_T^{1/2}\}_{i=1}^N)\|_{\max} = O\left(\{\max(p,q)\}^{1/2}N^{-1}T^{-1/2}\right).\end{aligned}$$

For $d=1,2$, $i=1,\ldots,N$ and $t=1,\ldots,T$, define

$$V_{i,d,m,1} = NT\{e_{l_m}^\intercal(X^\intercal\{\Sigma^{(r_m)}\}^{-1}X)^{-1}X^\intercal\{\Sigma^{(r_m)}\}^{-1}G(I_N\otimes\Sigma_\zeta^{1/2})e_{2i+d-2}\}\tilde{\zeta}_{r_m,i,d},$$

$$V_{i,t,m,2} = NT[e_{l_m}^\intercal(X^\intercal\{\Sigma^{(r_m)}\}^{-1}X)^{-1}X^\intercal\{\Sigma^{(r_m)}\}^{-1}\mathrm{diag}(\{\Sigma_T^{1/2}\}_{i=1}^N)e_{iT+t-T}]\tilde{\epsilon}_{r_m,i,t}.$$

Note that $\beta_{j_m}^{(r_m)} = 0$ under H_0. Then we can express J_m^o by

$$J_m^o = (NT)^{-1}|\theta_m|^{-1/2}\left(\sum_{i=1}^N\sum_{d=1}^2 V_{i,d,m,1} + \sum_{i=1}^N\sum_{t=1}^T V_{i,t,m,2}\right).$$

Denote $\check{V}_{i,d,m,1} = V_{i,d,m,1}/(\check{C}\{\max(p,q)T\}^{1/2})$, and $\check{V}_{i,t,m,2} = V_{i,t,m,2}/(\check{C}\{\max(p,q)T\}^{1/2})$, where $\check{C}$ is sufficiently large, so that $\mathbb{E}\exp(\nu\check{V}_{i,d,m,1}^2) \le C$ and $\mathbb{E}\exp(\nu\check{V}_{i,t,m,2}^2) \le C$ uniformly for all $d = 1, 2$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where ν and C are as defined in (C7). Let $\hat{V}_{i,d,m,1} = \check{V}_{i,d,m,1} I(|\check{V}_{i,d,m,1}| \le \varrho) - \mathbb{E}\{\check{V}_{i,d,m,1} I(|\check{V}_{i,d,m,1}| \le \varrho)\}$, and $\hat{V}_{i,t,m,2} = \check{V}_{i,t,m,2}\, I(|\check{V}_{i,t,m,2}| \le \varrho) - \mathbb{E}\{\check{V}_{i,t,m,2} I(|\check{V}_{i,t,m,2}| \le \varrho)\}$, with $\varrho = 10^{1/2}\nu^{-1/2}\{\log(TN + \tilde{p})\}^{1/2}$. Denote

$$\hat{J}_m^o = (NT)^{-1}|\theta_m|^{-1/2}(\check{C}\{\max(p,q)T\}^{1/2}) \left(\sum_{i=1}^N \sum_{d=1}^2 \hat{V}_{i,d,m,1} + \sum_{i=1}^N \sum_{t=1}^T \hat{V}_{i,t,m,2} \right).$$

We next show $\mathbb{E}\{\check{V}_{i,d,m,1} I(|\check{V}_{i,d,m,1}| \le \varrho)\}$ and $\mathbb{E}\{\check{V}_{i,t,m,2} I(|\check{V}_{i,t,m,2}| \le \varrho)\}$ are negligible.

Note that, by the facts that $\mathbb{E}\exp(\nu\check{V}_{i,d,m,1}^2) \le C$, and $\mathbb{E}\exp(\nu\check{V}_{i,t,m,2}^2) \le C$,

$$\begin{aligned}
\max\nolimits_{1\le m\le \tilde{p}} \quad & (NT)^{-1}|\theta_m|^{-1/2}(\check{C}\{\max(p,q)T\}^{1/2}) \sum_{i=1}^N \sum_{d=1}^2 \mathbb{E}\{|\check{V}_{i,d,m,1}| I(|\check{V}_{i,d,m,1}| \ge \varrho)\} \\
&= O(N^{1/2}\{\max(p,q)\}^{1/2}) \max_{1\le m\le\tilde{p}} \max_{1\le i\le N} \max_{1\le d\le 2} \mathbb{E}\{|\check{V}_{i,d,m,1}| I(|\check{V}_{i,d,m,1}| \ge \varrho)\} \\
&= O(N^{1/2}\{\max(p,q)\}^{1/2})(NT + \tilde{p})^{-5}, \\
\max\nolimits_{1\le m\le \tilde{p}} \quad & (NT)^{-1}|\theta_m|^{-1/2}(\check{C}\{\max(p,q)T\}^{1/2}) \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}\{|\check{V}_{i,t,m,2}| I(|\check{V}_{i,t,m,2}| \ge \varrho)\} \\
&= O(N^{1/2}T\{\max(p,q)\}^{1/2}) \max_{1\le m\le\tilde{p}} \max_{1\le i\le N} \max_{1\le t\le T} \mathbb{E}\{|\check{V}_{i,t,m,2}| I(|\check{V}_{i,t,m,2}| \ge \varrho)\} \\
&= O(N^{1/2}T\{\max(p,q)\}^{1/2})(NT + \tilde{p})^{-5}.
\end{aligned}$$

Therefore, we have that,

$$\begin{aligned}
&\mathbb{P}\left\{\max_m |J_m^o - \hat{J}_m^o| \ge (\log\tilde{p})^{-1}\right\} \\
&\le \mathbb{P}\left(\max_i \max_d \max_m |\check{V}_{i,d,m,1}| \ge \varrho\right) + \mathbb{P}\left(\max_i \max_t \max_m |\check{V}_{i,t,m,2}| \ge \varrho\right) = o(\tilde{p}^{-1}).
\end{aligned} \tag{B.34}$$

Given the fact that $|\max_m (J_m^o)^2 - \max_m (\hat{J}_m^o)^2| \le 2\max_m |J_m^o| \max_m |J_m^o - \hat{J}_m^o| + \max_m |J_m^o - \hat{J}_m^o|^2$, it suffices to show that, under H_0, for any fixed ϕ,

$$\mathbb{P}\left(\max_m (\hat{J}_m^o)^2 - 2\log\tilde{p} + \log\log\tilde{p} \le \phi\right) \to \exp\left\{-\frac{1}{\sqrt{\pi}}\exp(-\frac{\phi}{2})\right\}. \tag{B.35}$$

By Lemma 1, we have that, for any fixed $k \le \tilde{p}/2$,

$$\sum_{b=1}^{2k}(-1)^{b-1}\sum_{1\le m_1<\ldots<m_b\le\tilde{p}}\mathbb{P}\left\{(\hat{J}_{m_1}^o)^2 \ge \phi_{\tilde{p}},\ldots,(\hat{J}_{m_b}^o)^2 \ge \phi_{\tilde{p}}\right\} \le \mathbb{P}\left\{\max_m (\hat{J}_m^o)^2 \ge \phi_{\tilde{p}}\right\}$$
$$\le \sum_{b=1}^{2k-1}(-1)^{b-1}\sum_{1\le m_1<\ldots<m_b\le\tilde{p}}\mathbb{P}\left\{(\hat{J}_{m_1}^o)^2 \ge \phi_{\tilde{p}},\ldots,(\hat{J}_{m_b}^o)^2 \ge \phi_{\tilde{p}}\right\}, \tag{B.36}$$

where $\phi_{\tilde{p}} = 2\log\tilde{p} - \log\log\tilde{p} + \phi$. Define $|a|_{\min} = \min_{1\le s\le b}|a_s|$ for any vector $a \in \mathbb{R}^b$, and let $\tilde{V}_{i,m} = (NT|\theta_m|)^{-1/2}\left(\check{C}\{\max(p,q)T\}^{1/2}\right)\left(\sum_{d=1}^{2}\hat{V}_{i,d,m,1} + \sum_{t=1}^{T}\hat{V}_{i,t,m,2}\right)$, and $W_{i,\{m_1,\ldots,m_b\}} = (\tilde{V}_{i,m_1},\ldots,\tilde{V}_{i,m_b})$. Then we have that,

$$\mathbb{P}\left\{(\hat{J}_{m_1}^o)^2 \ge \phi_{\tilde{p}},\ldots,(\hat{J}_{m_b}^o)^2 \ge \phi_{\tilde{p}}\right\} = \mathbb{P}\left\{\left|(NT)^{-1/2}\sum_{i=1}^{N}W_{i,\{m_1,\ldots,m_b\}}\right|_{\min} \ge \phi_{\tilde{p}}^{1/2}\right\}.$$

By Theorem 1.1 in Zaïtsev (1987), we have that,

$$\mathbb{P}\left\{\left|(NT)^{-1/2}\sum_{i=1}^{N}W_{i,\{m_1,\ldots,m_b\}}\right|_{\min} \ge \phi_{\tilde{p}}^{1/2}\right\}$$
$$\le \mathbb{P}\left\{|Z_b|_{\min} \ge \phi_{\tilde{p}}^{1/2} - \epsilon_{NT,\tilde{p}}(\log\tilde{p})^{-1/2}\right\} + c_1 b^{5/2}\exp\left\{-\frac{(NT)^{1/2}\epsilon_{NT,\tilde{p}}}{c_2 b^3\{\max(p,q)T\}^{1/2}\varrho(\log\tilde{p})^{1/2}}\right\}, \tag{B.37}$$

where c_1 and c_2 are some positive constants, $\epsilon_{NT,\tilde{p}} \to 0$ that is to be specified later, and $Z_b = (Z_{m_1},\ldots,Z_{m_b})^\intercal$ is a normal vector with mean zero and the covariance matrix

$\text{cov}(W_{i,\{m_1,\ldots,m_b\}})/T$. Because b is fixed, $\theta_{N,T,R,B} = o(1)$, and $\log \tilde{p} = o\{(N/\max(p,q))^{1/4}\}$, we let $\epsilon_{NT,\tilde{p}} \to 0$ sufficiently slowly, so that for any large constant $\tilde{C} > 0$,

$$c_1 b^{5/2} \exp\left\{-\frac{(NT)^{1/2}\epsilon_{NT,\tilde{p}}}{c_2 b^3\{\max(p,q)T\}^{1/2}\varrho \log \tilde{p}}\right\} = O\left(\tilde{p}^{-\tilde{C}}\right).$$

By (B.36), we have that,

$$\mathbb{P}\left\{\max_m (\hat{J}_m^o)^2 \geq \phi_{\tilde{p}}\right\} \leq \sum_{b=1}^{2k-1}(-1)^{b-1} \sum_{1\leq m_1<\ldots<m_b\leq\tilde{p}} \mathbb{P}\left\{|Z_b|_{\min} \geq \phi_{\tilde{p}}^{1/2} - \epsilon_{NT,\tilde{p}}(\log \tilde{p})^{-1/2}\right\} + o(1). \tag{B.38}$$

Similarly, we also obtain that,

$$\mathbb{P}\left\{\max_m (\hat{J}_m^o)^2 \geq \phi_{\tilde{p}}\right\} \geq \sum_{b=1}^{2k}(-1)^{b-1} \sum_{1\leq m_1<\ldots<m_b\leq\tilde{p}} \mathbb{P}\left\{|Z_b|_{\min} \geq \phi_{\tilde{p}}^{1/2} + \epsilon_{NT,\tilde{p}}(\log \tilde{p})^{-1/2}\right\} + o(1). \tag{B.39}$$

Then it suffices to show that, for any fixed integer b and any ϕ,

$$\sum_{1\leq m_1<\ldots<m_b\leq\tilde{p}} \mathbb{P}\left\{|Z_b|_{\min} \geq \phi_{\tilde{p}}^{1/2} \pm \epsilon_{NT,\tilde{p}}(\log \tilde{p})^{-1/2}\right\} = \frac{1}{b!}\pi^{-b/2}\exp(-b\phi/2)\{1+o(1)\}. \tag{B.40}$$

Then by Lemma 2, the result in (i) follows.

For (ii), the result follows based on the proofs above and the proof in Theorem 3.

For (iii), similarly as the proof of Lemma 7, it suffices to show that

$$\int_0^{t_{\tilde{p}}} P\left[\left|\frac{\sum_{(r,j)\in\mathcal{H}_0}\{I(|J_{r,j}^o| \geq \tau) - \mathbb{P}(|J_{r,j}^o| \geq \tau)\}}{\tilde{p}_0 G(\tau)}\right| \geq \epsilon\right] d\tau = o(v_{\tilde{p}}).$$

Using the Markov inequality, we only need to bound

$$\frac{\sum_{(r_1,j_1),(r_2,j_2)\in\mathcal{H}_0}\{\mathbb{P}(|J^o_{r_1,j_1}|\geq\tau,|J^o_{r_2,j_2}|\geq\tau)-\mathbb{P}(|J^o_{r_1,j_1}|\geq\tau)\mathbb{P}(|J^o_{r_2,j_2}|\geq\tau)\}}{\tilde{p}_0^2G^2(\tau)}.$$

Following the proof of Lemma 7, we divide the index pairs $(r_1, j_1), (r_2, j_2) \in \mathcal{H}_0$ into three subsets based on the magnitude of the covariance $\text{cov}(J^o_{r_1,j_1}, J^o_{r_2,j_2})$. For each subset, by employing the truncations from (i) and the proof of Lemma 7 in the supplement of Sun et al. (2023), the result in (iii) follows under $(\log\tilde{p})^{7+\epsilon} = O(N/\max(p, q))$.

This completes the proof of Theorem 5. □

C Additional Numerical Results

C.1 Experiment results when $T = 8$

We present the additional estimation and multiple testing results for $T = 8$ in Tables S1 and S2, respectively.

C.2 Experiment results under non-Gaussian error distribution

We carry out a simulation in which the error terms in (2.3) follow a heavy-tailed distribution, with the entries of $\Sigma_T^{-1/2}\epsilon_i\Sigma_R^{-1/2}$ and $\Sigma_\zeta^{-1/2}\zeta_{i,r}$ drawn from a t-distribution with 6 degrees of freedom. We consider $T = 4$, while all other settings are the same as in Section 4.4.

Table S3 presents the multiple-testing results. It shows that our proposed method continues to control the FDR at the target level $\alpha = 0.1$, whereas the REML fails to do

Table S1: Parameter estimation: the bias and standard error based on 200 data replications for the autoregressive and moving average temporal structures with $T = 8$.

Temporal structure			Autoregressive				Moving average			
			$\omega = 0.03$		$\omega = 0.05$		$\omega = 0.03$		$\omega = 0.05$	
R	N	method	hub	small	hub	small	hub	small	hub	small
Bias and SE of covariance estimation										
50	100	Bias	0.0541	0.0555	0.0991	0.0994	0.0535	0.0547	0.0991	0.0985
		SE	0.1759	0.1405	0.1765	0.1934	0.1862	0.1516	0.1890	0.2051
	200	Bias	0.0550	0.0550	0.1195	0.1201	0.0542	0.0544	0.1192	0.1201
		SE	0.1342	0.1315	0.1565	0.1523	0.1427	0.1367	0.1638	0.1607
100	100	Bias	0.0785	0.0768	0.1262	0.1273	0.0780	0.0759	0.1252	0.1270
		SE	0.1687	0.1912	0.2275	0.1752	0.1811	0.2037	0.2344	0.1868
	200	Bias	0.0626	0.0617	0.1022	0.1034	0.0623	0.0616	0.1018	0.1032
		SE	0.1193	0.1192	0.1517	0.1284	0.1268	0.1269	0.1590	0.1366
Bias and SE of coefficient estimation										
50	100	Bias	-0.0002	-0.0002	0.0006	0.0002	-0.0002	-0.0002	0.0006	0.0001
		SE	0.1131	0.1129	0.1130	0.1140	0.1124	0.1121	0.1124	0.1134
	200	Bias	0.0000	0.0001	-0.0002	0.0004	0.0000	0.0001	-0.0002	0.0004
		SE	0.0755	0.0755	0.0755	0.0753	0.0749	0.0750	0.0750	0.0748
100	100	Bias	-0.0001	0.0001	-0.0004	-0.0002	-0.0001	0.0002	-0.0004	-0.0002
		SE	0.1131	0.1131	0.1134	0.1133	0.1124	0.1125	0.1127	0.1126
	200	Bias	-0.0001	-0.0001	-0.0002	0.0001	-0.0001	-0.0001	-0.0002	0.0001
		SE	0.0754	0.0755	0.0752	0.0755	0.0748	0.0749	0.0747	0.0749

so in certain cases. In terms of empirical power, our method is generally more powerful than REML, and this advantage becomes increasingly pronounced as N or T grows.

C.3 Experiment results under misspecification of Kronecker structure

We carry out a simulation in which the covariance of random error $\epsilon_{i,r,t}$ does *not* follow a Kronecker structure as in (2.3), but instead,

$$(\epsilon_{i,1,1}, \ldots, \epsilon_{i,1,T}, \ldots, \epsilon_{i,R,1}, \ldots, \epsilon_{i,R,T})^\intercal \sim \text{Normal}(0, \Sigma_{\text{non-sep}}), \quad \text{where}$$

$$\Sigma_{\text{non-sep}}(t_1, r_1, t_2, r_2) = (1 - \lambda_k)[\Sigma_T]_{t_1,t_2}[\Sigma_R]_{r_1,r_2} + \lambda_k \frac{1}{(t_1 - t_2)^2 + 1} \exp\left\{-\frac{(r_1 - r_2)^2}{(t_1 - t_2)^2 + 1}\right\},$$

Table S2: Multiple testing: the empirical FDR and power in percentage based on 200 data replication for the autoregressive and moving average temporal structures with $T = 8$ and the FDR level $\alpha = 0.1$.

Temporal structure			Autoregressive				Moving average			
			$\omega = 0.03$		$\omega = 0.05$		$\omega = 0.03$		$\omega = 0.05$	
R	N	method	hub	small	hub	small	hub	small	hub	small
Empirical FDR										
50	100	Proposed	8.52	7.89	6.46	7.83	9.21	8.30	6.87	8.77
		REML	12.74	12.81	11.55	11.75	12.74	12.43	11.40	11.82
	200	Proposed	7.09	7.47	5.96	6.15	7.03	7.64	5.99	6.10
		REML	10.50	10.27	10.17	10.43	10.50	10.42	10.15	10.20
100	100	Proposed	7.34	7.95	6.84	5.94	7.77	8.89	7.53	6.21
		REML	12.03	12.32	11.61	11.58	11.84	12.27	11.61	11.57
	200	Proposed	6.38	7.10	5.83	5.79	6.57	7.04	5.83	5.76
		REML	9.60	10.23	10.13	9.88	9.65	10.22	10.07	9.98
Empirical Power										
50	100	Proposed	87.08	86.09	86.95	88.27	87.59	86.95	87.74	88.78
		REML	83.29	81.88	84.92	85.84	83.39	81.83	84.87	85.38
	200	Proposed	99.80	99.70	99.88	99.91	99.88	99.77	99.91	99.92
		REML	99.58	98.59	99.79	99.64	99.67	98.67	99.79	99.67
100	100	Proposed	89.44	88.52	90.60	89.77	89.83	89.30	91.22	90.32
		REML	85.08	85.15	87.84	86.78	84.69	84.36	87.67	86.14
	200	Proposed	99.87	99.92	99.90	99.87	99.88	99.94	99.93	99.89
		REML	99.48	99.74	99.74	99.59	99.57	99.75	99.76	99.62

for $1 \leq t_1, t_2 \leq T$ and $1 \leq r_1, r_2 \leq R$, and the hyper-parameter λ_k controls the degree of deviation from the Kronecker structure. We consider $\lambda_k = \{0, 0.2, 0.4, 0.6\}$, with $R = 100, N = 100, T = 4$. All other settings are the same as in Section 4.4.

We report the multiple testing results in Table S4. We see that our proposed method consistently maintains the empirical FDR under the pre-specified level $\alpha = 0.1$, while the REML-based testing method shows noticeable inflation. Moreover, as λ_k increases, reflecting a greater deviation from the Kronecker structure, the empirical power of our method only shows a modest decline, which demonstrates the robustness of our method to departure from the Kronecker condition.

Table S3: Multiple testing for t-distribution error with 6 degrees of freedom: the empirical FDR and power in percentage basyed on 200 data replication for the autoregressive and moving average temporal structures with $T = 4$ and the FDR level $\alpha = 0.1$.

Temporal structure			Autoregressive				Moving average			
			$\omega = 0.03$		$\omega = 0.05$		$\omega = 0.03$		$\omega = 0.05$	
R	N	method	hub	small	hub	small	hub	small	hub	small
Empirical FDR										
50	100	Proposed	8.24	10.05	7.97	5.20	7.93	9.97	7.59	6.08
		REML	10.11	8.44	9.51	9.18	10.73	8.74	10.00	9.27
	200	Proposed	7.99	9.13	7.94	7.70	8.20	8.96	7.91	8.02
		REML	10.27	11.31	10.06	10.40	10.30	11.28	10.03	10.30
100	100	Proposed	6.29	6.07	8.25	8.47	6.49	6.52	8.04	8.41
		REML	9.12	10.22	13.06	11.85	8.69	9.03	13.28	11.92
	200	Proposed	8.72	8.37	7.46	7.56	8.98	8.22	7.48	7.56
		REML	11.19	10.14	10.02	9.89	11.11	10.20	9.93	9.94
Empirical Power										
50	100	Proposed	12.95	15.59	23.44	17.54	12.85	15.64	23.55	17.93
		REML	12.44	13.48	23.34	18.88	11.73	12.88	22.97	18.17
	200	Proposed	74.38	74.89	75.75	76.49	75.08	75.33	76.43	77.11
		REML	69.85	69.24	72.35	72.54	70.30	69.71	73.01	73.02
100	100	Proposed	13.25	13.08	23.14	23.64	13.03	13.48	23.24	23.90
		REML	14.36	15.17	26.02	25.69	13.66	15.28	25.75	25.23
	200	Proposed	72.40	72.82	76.26	77.86	73.45	73.77	76.80	78.53
		REML	68.83	67.61	73.75	74.62	69.11	68.08	73.92	75.00

C.4 Experiment results with $R > N$

We carry out a simulation in which the number of response variables R is much larger than the sample size N. We consider $R = \{500, 1000\}, N = \{100, 200\}, T = 4$. All other settings are the same as in Section 4.4.

We report the multiple testing results in Table S5. We see that our proposed method continues to control the FDR at the target level $\alpha = 0.1$, while REML fails to do so. Moreover, as N increases from 100 to 200, both our method and REML achieve similar levels of empirical power.

Table S4: Multiple testing under misspecification of the Kronecker structure: the empirical FDR and power in percentage based on 200 data replications for the autoregressive and moving average temporal structures with $R = 100$, $N = 100$, $T = 4$ and the FDR level $\alpha = 0.1$.

Temporal structure		Autoregressive				Moving average			
		$\omega = 0.03$		$\omega = 0.05$		$\omega = 0.03$		$\omega = 0.05$	
λ_k	Method	hub	small	hub	small	hub	small	hub	small
Empirical FDR									
0	Proposed	7.05	7.85	6.78	6.85	7.02	7.60	6.68	6.73
	REML	12.48	13.18	11.74	11.96	12.53	13.31	11.78	11.89
0.2	Proposed	8.13	9.99	7.53	7.96	8.61	10.36	7.78	8.11
	REML	12.52	13.76	11.60	12.06	12.27	13.64	11.56	11.96
0.4	Proposed	7.61	9.72	6.85	7.95	7.27	10.10	6.95	7.95
	REML	12.76	13.68	11.59	12.11	12.72	13.79	11.68	12.02
0.6	Proposed	5.13	5.87	4.84	5.67	4.81	5.99	4.97	5.47
	REML	12.79	13.64	11.78	12.06	12.80	13.65	11.80	12.09
Empirical Power									
0	Proposed	39.76	40.09	47.05	46.87	40.29	40.54	47.49	47.36
	REML	41.23	41.44	49.55	49.33	41.61	41.38	49.61	49.59
0.2	Proposed	38.02	40.27	45.81	45.25	38.67	40.77	46.15	45.76
	REML	40.59	40.42	49.11	47.81	40.64	40.49	49.03	47.85
0.4	Proposed	34.54	37.30	42.61	42.72	34.62	37.24	42.87	42.71
	REML	40.24	40.18	48.79	47.65	40.17	40.01	48.87	47.75
0.6	Proposed	26.14	27.77	36.66	35.36	25.77	27.45	36.60	34.94
	REML	39.97	39.80	48.60	47.52	39.95	39.98	48.69	47.42

C.5 Model diagnosis for longitudinal neuroimaging data

We perform model diagnosis for the OASIS-2 longitudinal AD data. We focus on the normality and linearity conditions. Specifically, for each region, we first estimate the covariance matrices, then apply the least squares method to estimate the coefficients. Based on the fitted model, we then compute the residuals. Figure S1 shows the QQ-plots of the standardized residuals for three randomly selected regions. We see that the points lie close to the reference line. Moreover, the Shapiro-Wilk normality test shows that, for 59 out of 68 regions, the p-value exceeds 0.05. Both results suggest that the

Table S5: Multiple testing with R being much larger than N: the empirical FDR and power in percentage based on 200 data replication for the autoregressive and moving average temporal structures with $T = 4$ and the FDR level $\alpha = 0.1$.

Temporal structure			Autoregressive				Moving average			
			$\omega = 0.03$		$\omega = 0.05$		$\omega = 0.03$		$\omega = 0.05$	
R	N	method	hub	small	hub	small	hub	small	hub	small
Empirical FDR										
500	100	Proposed	7.32	6.47	5.93	5.28	7.63	6.47	6.09	5.61
		REML	13.24	12.94	12.06	12.00	13.31	13.03	12.15	12.09
	200	Proposed	7.38	7.16	6.64	6.35	7.48	7.19	6.71	6.27
		REML	10.50	10.39	10.24	10.32	10.55	10.38	10.16	10.28
1000	100	Proposed	6.94	6.40	5.58	5.17	7.23	6.55	6.02	5.54
		REML	12.95	13.05	11.69	11.86	13.01	13.05	11.75	11.87
	200	Proposed	7.16	6.70	6.78	6.32	7.23	6.77	6.80	6.34
		REML	10.68	10.87	10.35	10.29	10.73	10.97	10.31	10.23
Empirical Power										
500	100	Proposed	41.78	39.31	46.39	43.43	42.45	39.72	46.88	44.45
		REML	42.90	41.81	49.93	49.66	42.31	41.24	49.47	48.99
	200	Proposed	90.03	89.79	91.70	91.53	90.71	90.39	92.29	92.10
		REML	88.61	88.03	90.86	91.00	89.27	88.72	91.36	91.53
1000	100	Proposed	40.33	39.06	42.97	42.82	40.96	39.43	44.25	43.92
		REML	42.17	42.04	48.22	48.40	41.78	41.64	48.01	48.25
	200	Proposed	89.44	89.09	92.25	91.97	90.13	89.76	92.8	92.58
		REML	88.32	88.27	91.54	91.55	88.99	88.83	91.99	92.06

normality holds reasonably well. Figure S2 shows the fitted values versus the Pearson residuals for three randomly selected regions. We see that the residuals center around the horizontal line at zero with no discernible pattern, indicating no clear evidence of nonlinearity.

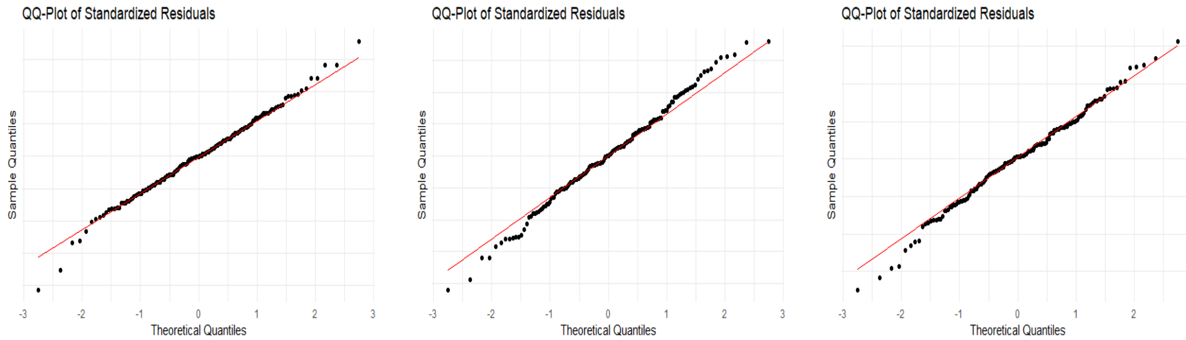

Figure S1: Model diagnosis for longitudinal AD data: QQ-plots for the standardized residuals for three randomly selected regions.

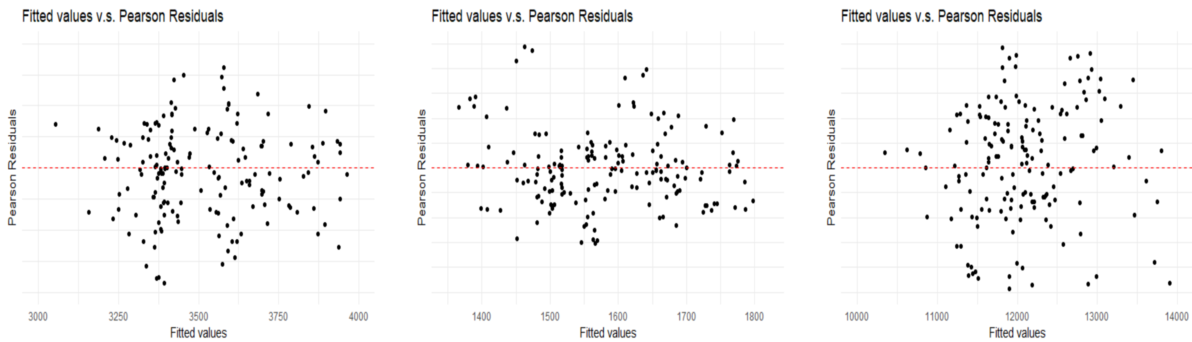

Figure S2: Model diagnosis for longitudinal AD data: the fitted values versus Pearson residuals for three randomly selected regions.

References

Berman, S. M. (1962). A law of large numbers for the maximum in a stationary gaussian sequence. *Ann. Math. Statist. 33*, 93–97.

Cai, T. T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependency. *J. R. Statist. Soc. B 76*, 349–372.

Chen, X., D. Yang, Y. Xu, Y. Xia, D. Wang, and H. Shen (2023). Testing and support recovery of correlation structures for matrix-valued observations with an application to stock market data. *J. Econom. 232*(2), 544–564.

Sun, Y., L. Ma, and Y. Xia (2023). A decorrelating and debiasing approach to simultaneous inference for high-dimensional confounded models. *Journal of the American Statistical Association accepted*, 1–24.

Takane, Y. and T. Shibayama (1991). Principal component analysis with external information on both subjects and variables. *Psychometrika 56*(1), 97–120.

Zaïtsev, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of S.N. Bernstein's inequality conditions. *Probab. Theory Rel. 74*(4), 535–566.