Estimating Covariance Matrices at Different Levels in Repeated Measurements

Supplementary Materials

Sunpeng Duan*, Guo Yu*, Juntao Duan, Yuedong Wang

Department of Statistics and Applied Probability

University of California, Santa Barbara, 93106

This supplement includes further optimization implementation details, tuning parameter selection, comparison between the constrained and unconstrained estimators, and proofs of our main theorems.

S1 Further Details on Optimization Algorithm Implementation

The complete algorithm solving the convex optimization problem (3.7) in the main context is summarized in Algorithm 1.

A reasonable stopping criteria suggested by Boyd et al. [2010] is

$$\|\Sigma^{(l+1)} - \Theta^{(l+1)}\|_F \le \epsilon^{\text{pri}} \qquad \text{and} \qquad \|\rho(\Theta^{(l+1)} - \Theta^{(l)})\|_F \le \epsilon^{\text{dual}}. \tag{S1.1}$$

*These authors contributed equally to this work

Algorithm 1 Alternating direction method of multipliers for solving (3.7) in the paper

Require: δ, λ, $\rho^{(0)}$, B, $\Sigma^{(0)}$, $\Theta^{(0)}$, $\Lambda^{(0)}$, and $l = 0$.

1: Repeat

2: $\Sigma^{(l+1)} \leftarrow \frac{1}{1+\rho^{(l)}} \left(B + \rho\Theta^{(l)} - \Lambda^{(l)}, \delta\right)_+$

3: $\Theta^{(l+1)} \leftarrow \mathcal{S}_{\lambda/\rho^{(l)}} \left(\Sigma^{(l+1)} + \frac{1}{\rho^{(l)}}\Lambda^{(l)}\right)$

4: $\Lambda^{(l+1)} \leftarrow \Lambda^{(l)} + \rho^{(l)} \left(\Sigma^{(l+1)} - \Theta^{(l+1)}\right)$

5: Update $\rho^{(l+1)}$ based on equation (3.13) in Boyd et al. [2010]

6: Until convergence

where ϵ^{pri} and ϵ^{dual} are positive feasibility tolerances for the primal and dual feasibility conditions, which are controlled by an absolute criterion ϵ^{abs} and a relative criterion ϵ^{rel}:

$$
\begin{aligned}
\epsilon^{\text{pri}} &= p\epsilon^{\text{abs}} + \epsilon^{\text{rel}} \max\{\|\Sigma^{(l+1)}\|_F, \|\Theta^{(l+1)}\|_F\}, \\
\epsilon^{\text{dual}} &= p\epsilon^{\text{abs}} + \epsilon^{\text{rel}} \|\Lambda^{(l+1)}\|_F, \qquad \text{(S1.2)}
\end{aligned}
$$

where $\epsilon^{\text{abs}} > 0$ and $\epsilon^{\text{rel}} > 0$. In the numerical studies, we choose $\epsilon^{\text{abs}} = \epsilon^{\text{rel}} = 10^{-8}$. Choice of ρ can greatly impact the practical convergence of the alternating direction method procedure. And to improve the convergence, we adopt an adaptive strategy described in Boyd et al. [2010] for varying penalty parameter ρ. In practice, we use the soft-thresholding estimators based on the sample estimates as the initial ($\Sigma^{(0)}$, $\Theta^{(0)}$). And the initial input for $\Lambda^{(0)}$ is a zero matrix. The initial penalty parameter ρ is 0.1. Without the positive semi-definite constraints of Σ_ε and Σ_b in (3.7) in the main context, the unconstrained solutions will be $\mathcal{S}_\lambda(\widehat{\Sigma}_\varepsilon)$ and $\mathcal{S}_\lambda(\widehat{\Sigma}_b)$ with $B = \widehat{\Sigma}_\varepsilon$ and $B = \widehat{\Sigma}_b$, respectively. For efficient computation, we always first check the positive semi-definiteness of $\mathcal{S}_\lambda(\widehat{\Sigma}_\varepsilon)$ and $\mathcal{S}_\lambda(\widehat{\Sigma}_b)$. If $\mathcal{S}_\lambda(\widehat{\Sigma}_\varepsilon)$ and $\mathcal{S}_\lambda(\widehat{\Sigma}_b)$ are positive semi-definite, they are the final solutions to (3.7) in the main context, respectively. Otherwise, we will use Algorithm 1 to solve (3.7) in the main context.

S2 Tuning Parameters Selection using Cross-validation

The main optimization problem (3.7) in the main context defines various estimators that we study in this paper, where λ is the tuning parameter that controls the level of regularization of the sample estimates. We present in this section a cross-validation procedure for selecting the tuning parameter [Bickel and Levina, 2008, Rothman et al., 2009, Cai and Liu, 2011] specifically in the presence of repeated measurements.

For each (of the K) split in a K-fold cross-validation procedure, we randomly partition the m groups into a set of m_1 groups of training set, i.e., $\mathcal{T}_r = \{\boldsymbol{Y}_{ij} : i \in \mathcal{A}\}$ with $|\mathcal{A}| = m_1$ and a set of $m - m_1$ groups of validation set, i.e., $\mathcal{T}_e = \{\boldsymbol{Y}_{ij} : i \in \mathcal{A}^c\}$ with $|\mathcal{A}^c| = m - m_1$.

Let $\widehat{S}^+\{\lambda, \widehat{S}(\mathcal{T})\}$ denote a generic estimator, which is defined as a solution to the optimization problem (3.7) in the main text with the tuning parameter value λ and input sample matrix $\widehat{S}(\mathcal{T})$ evaluated using a dataset $\mathcal{T}$. Specifically, the estimator $\widehat{S}^{+}\{\lambda, \widehat{S}(\mathcal{T})\}$ could refer to $\widehat{\Sigma}_b^{+}$, $\widehat{\Sigma}_{\varepsilon}^{+}$, $\widetilde{\Sigma}_b^{+}$, and $\overline{\Sigma}^{+}$. And $\widehat{S}(\mathcal{T})$ refers to the unbiased estimator $\widehat{\Sigma}_b$, $\widehat{\Sigma}_{\varepsilon}$, $\widetilde{\Sigma}_b$, and the biased estimator $\overline{\Sigma}$. The cross-validation procedure is presented in the following Algorithm 2 to choose the tuning parameter from a path of candidate tuning parameter values $\{\lambda_1 > \lambda_2 > \ldots > \lambda_L\}$.

Algorithm 2 a K-fold Cross-Validation Procedure

Require: $\{\boldsymbol{Y}_{ij} : 1 \le i \le m, 1 \le j \le n_i\}$ and $\{\lambda_1 > \lambda_2 > \ldots > \lambda_L\}$.

```
for ℓ = 1, ..., L do
    for ν = 1, ..., K do
        Divide {Y_ij : 1 ≤ i ≤ m, 1 ≤ j ≤ n_i} into training set T_r^(ν) and validation set T_e^(ν);
        Compute the sample covariance matrix Ŝ(T_e^(ν)) on the validation set T_e^(ν);
        Compute the estimator Ŝ^+{λ_ℓ, Ŝ(T_r^(ν))} on the training set T_r^(ν).
    end for
    Compute CV estimate of error E_ℓ = Σ_{ν=1}^K ||Ŝ^+{λ_ℓ, Ŝ(T_r^(ν))} − Ŝ(T_e^(ν))||_F^2 / K.
end for
Let ℓ̂ = argmin_{ℓ=1,...,L} E_ℓ, and return the selected tuning parameter λ_ℓ̂.
```

S3 Lemmas and Proofs of the Main Theorems

S3.1 Lemmas for the Exponential-tail Condition

We observe $\boldsymbol{Y}_{ij} \in \mathbb{R}^p$, which is the j-th repeated measurement of the i-th subject for $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$, following the model (2.1) in the paper, where

$\boldsymbol{\varepsilon}_{ij}$ and $\boldsymbol{b}_i$ are p-dimensional sub-Gaussian random vectors with the true within-subject and between-subject covariance, i.e., $\mathrm{cov}(\boldsymbol{\varepsilon}_{ij}) = \Sigma_\varepsilon^0$ and $\mathrm{cov}(\boldsymbol{b}_i) = \Sigma_b^0$, respectively, and $\boldsymbol{b}_i$ and $\boldsymbol{\varepsilon}_{ij}$ are mutually independent. Let $N = \sum_{i=1}^m n_i$ be the total number of observations. We begin with several lemmas, which are essential for the proofs of the main results.

Lemma 1. *Consider the true within-subject covariance Σ_ε^0 with $\max_k(\Sigma_\varepsilon^0)_{k,k} \le M_\varepsilon$. Let $\lambda_\varepsilon = C_1\{N\log p\}^{1/2}/(N-m)$ for a sufficiently large constant C_1. If $\log p \le N$, then the unbiased within-subject sample estimate $\widehat{\Sigma}_\varepsilon$ satisfies*

$$pr\left\{\max_{k,l}\left|(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}\right| > \lambda_\varepsilon\right\} \le 4p^{-C_2},$$

where $C_2 > 0$ only depends on C_1 and M_ε.

Proof. We first rewrite $\widehat{\Sigma}_\varepsilon$ as follows,

$$\begin{aligned}\widehat{\Sigma}_\varepsilon &= \frac{1}{N-m}\sum_{i=1}^m\sum_{j=1}^{n_i}(\boldsymbol{Y}_{ij} - \bar{\boldsymbol{Y}}_{i\cdot})(\boldsymbol{Y}_{ij} - \bar{\boldsymbol{Y}}_{i\cdot})^{\mathrm{T}} \\ &= \frac{1}{N-m}\left(\sum_{i=1}^m\sum_{j=1}^{n_i}\boldsymbol{\varepsilon}_{ij}\boldsymbol{\varepsilon}_{ij}^{\mathrm{T}} - \sum_{i=1}^m n_i\bar{\boldsymbol{\varepsilon}}_{i\cdot}\bar{\boldsymbol{\varepsilon}}_{i\cdot}^{\mathrm{T}}\right).\end{aligned}$$

Then,

$$\begin{aligned}(\widehat{\Sigma}_\varepsilon)_{k,l} &= \frac{1}{N-m}\left(\sum_{i=1}^m\sum_{j=1}^{n_i}\varepsilon_{ijk}\varepsilon_{ijl} - \sum_{i=1}^m n_i\bar{\varepsilon}_{i\cdot k}\bar{\varepsilon}_{i\cdot l}\right) \qquad \text{(S3.3)} \\ &= \frac{1}{N-m}\sum_{i=1}^m\sum_{j=1}^{n_i}\varepsilon_{ijk}\varepsilon_{ijl} - \frac{1}{N-m}\sum_{i=1}^m n_i\bar{\varepsilon}_{i\cdot k}\bar{\varepsilon}_{i\cdot l} \\ &= \frac{1}{N-m}\sum_{i=1}^m\sum_{j=1}^{n_i}\left\{\varepsilon_{ijk}\varepsilon_{ijl} - (\Sigma_\varepsilon^0)_{k,l}\right\} - \frac{1}{N-m}\sum_{i=1}^m\left\{\frac{1}{n_i}S_{i\cdot k}S_{i\cdot l} - (\Sigma_\varepsilon^0)_{k,l}\right\} + (\Sigma_\varepsilon^0)_{k,l},\end{aligned}$$

where $S_{i\cdot k} = \sum_{j=1}^{n_i} \varepsilon_{ijk}$.

By (S3.3),

$$\max_{k,l} \left| (\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l} \right| \le \max_{k,l} \frac{1}{N-m} \left| \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left\{ \varepsilon_{ijk}\varepsilon_{ijl} - (\Sigma_\varepsilon^0)_{k,l} \right\} \right| + \frac{1}{N-m} \max_{k,l} \left| \sum_{i=1}^{m} \left\{ \frac{1}{n_i} S_{i\cdot k} S_{i\cdot l} - (\Sigma_\varepsilon^0)_{k,l} \right\} \right|. \tag{S3.4}$$

Now, we assume that $\varepsilon_{ijk} \in \mathcal{SG}(\sigma_{\varepsilon,k}^2)$, i.e., ε_{ijk} is sub-Gaussian with a variance factor $\sigma_{\varepsilon,k}^2$ for $1 \le i \le m, 1 \le j \le n_i, 1 \le k \le p$. It is easy to check that $n_i^{-1/2} S_{i\cdot k} \in \mathcal{SG}(\sigma_{\varepsilon,k}^2)$.

Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a convex function with $\psi(0) = 0$, especially, $\psi_q(v) = \exp(|v|^q) - 1$, for $q \in [1, 2]$. Then for an $\mathbb{R}$-valued random variable X, the Orlicz norm of X is $\|X\|_\psi = \inf\{t \in \mathbb{R}_+ : E\{\psi(|X|/t)\} \le 1\}$. And by the properties of Orlicz norms, for any random variable X and any increasing convex $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ with $\psi(0) = 0$, we have

$$\|X - E(X)\|_\psi \le 2\|X\|_\psi. \tag{S3.5}$$

Moreover, if $X \in \mathcal{SG}(\sigma^2)$, then

$$\|X\|_{\psi_2} \le c_0 \sigma, \tag{S3.6}$$

for some $c_0 \le (8/3)^{1/2}$.

Since $\varepsilon_{ijk} \in \mathcal{SG}(\sigma_{\varepsilon,k}^2)$ and $n_i^{-1/2} S_{i\cdot k} \in \mathcal{SG}(\sigma_{\varepsilon,k}^2)$, by Lemma 2.7.7 in Vershyin [2018], $\varepsilon_{ijk}\varepsilon_{ijl}$ and $n_i^{-1} S_{i\cdot k} S_{i\cdot l}$ are sub-Exponential random variables. Let

$\max_k \sigma^2_{\varepsilon,k} = M_\varepsilon$. Combining (S3.5) and (S3.6), Lemma 2.7.7 in Vershyin [2018] implies that

$$\left\|\varepsilon_{ijk}\varepsilon_{ijl} - (\Sigma^0_\varepsilon)_{k,l}\right\|_{\psi_1} \leq 2\left\|\varepsilon_{ijk}\varepsilon_{ijl}\right\|_{\psi_1} \leq 2\left\|\varepsilon_{ijk}\right\|_{\psi_2}\left\|\varepsilon_{ijl}\right\|_{\psi_2} \leq c_1 M_\varepsilon,$$

and

$$\left\|n_i^{-1}S_{i\cdot k}S_{i\cdot l} - (\Sigma^0_\varepsilon)_{k,l}\right\|_{\psi_1} \leq 2\left\|n_i^{-1}S_{i\cdot k}S_{i\cdot l}\right\|_{\psi_1} \leq 2\left\|n_i^{-1/2}S_{i\cdot k}\right\|_{\psi_2}\left\|n_i^{-1/2}S_{i\cdot l}\right\|_{\psi_2} \leq c_1 M_\varepsilon,$$

where $c_1 = 2c_0^2$.

Hence, for the first term in (S3.4), by the union sum inequality and Bernstein's inequality (Theorem 2.8.2 in Vershyin [2018]), we can get

$$\begin{aligned} & \mathrm{pr}\left[\max_{k,l}\frac{1}{N-m}\left|\sum_{i=1}^{m}\sum_{j=1}^{n_i}\{\varepsilon_{ijk}\varepsilon_{ijl} - (\Sigma^0_\varepsilon)_{k,l}\}\right| \geq t\right] \\ \leq\ & 2p^2\exp\left[-c_2\min\left\{\frac{t^2(N-m)^2}{NK_1^2}, \frac{t(N-m)}{K_1}\right\}\right], \end{aligned} \tag{S3.7}$$

where $c_2 > 0$, $K_1 = \max_{i,k,l}\|\varepsilon_{ijk}\varepsilon_{ijl} - (\Sigma^0_\varepsilon)_{k,l}\|_{\psi_1} \leq c_1 M_\varepsilon$.

Similarly,

$$\begin{aligned} & \mathrm{pr}\left[\frac{1}{N-m}\max_{k,l}\left|\sum_{i=1}^{m}\left\{\frac{1}{n_i}S_{i\cdot k}S_{i\cdot l} - (\Sigma^0_\varepsilon)_{k,l}\right\}\right| \geq t\right] \\ \leq\ & 2p^2\exp\left[-c_3\min\left\{\frac{t^2(N-m)^2}{mK_2^2}, \frac{t(N-m)}{K_2}\right\}\right], \end{aligned} \tag{S3.8}$$

where $c_3 > 0$, $K_2 = \max_{i,k,l}\|n_i^{-1}S_{i\cdot k}S_{i\cdot l} - (\Sigma^0_\varepsilon)_{k,l}\|_{\psi_1} \leq c_1 M_\varepsilon$.

By (S3.7) and (S3.8), take $t = C_1(N\log p)^{1/2}/\{2(N-m)\}$ for a sufficiently large constant $C_1 > 0$, with $N > \log p$, we will have

$$\begin{aligned}
& \mathrm{pr}\left[\max_{k,l} \frac{1}{N-m}\left|\sum_{i=1}^{m}\sum_{j=1}^{n_i}\{\varepsilon_{ijk}\varepsilon_{ijl} - (\Sigma_\varepsilon^0)_{k,l}\}\right| \geq t\right] \\
\leq\ & 2\exp\left[\max\left\{\left(2 - \frac{c_2 N C_1^2}{4mK_1^2}\right)\log p, 2\log p - \frac{c_2 C_1}{2K_1}(N\log p)^{1/2}\right\}\right] \\
\leq\ & 2\exp\left\{\max\left(2 - \frac{c_2 C_1^2}{4c_1^2 M_\varepsilon^2}, 2 - \frac{c_2 C_1}{2c_1 M_\varepsilon}\right)\log p\right\}, \qquad \text{(S3.9)}
\end{aligned}$$

and

$$\begin{aligned}
& \mathrm{pr}\left[\frac{1}{N-m}\max_{k,l}\left|\sum_{i=1}^{m}\left\{\frac{1}{n_i}S_{i\cdot k}S_{i\cdot l} - (\Sigma_\varepsilon^0)_{k,l}\right\}\right| \geq t\right] \\
\leq\ & 2\exp\left[\max\left\{\left(2 - \frac{c_3 N C_1^2}{4mK_2^2}\right)\log p, 2\log p - \frac{c_3 C_1}{2K_2}(N\log p)^{1/2}\right\}\right] \\
\leq\ & 2\exp\left\{\max\left(2 - \frac{c_3 C_1^2}{4c_1^2 M_\varepsilon^2}, 2 - \frac{c_3 C_1}{2c_1 M_\varepsilon}\right)\log p\right\}. \qquad \text{(S3.10)}
\end{aligned}$$

Combining (S3.9) and (S3.10), with $\lambda_\varepsilon = C_1(N\log p)^{1/2}/(N-m)$, we have

$$\begin{aligned}
& \mathrm{pr}\left\{\max_{k,l}\left|(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}\right| > \lambda_\varepsilon\right\} \\
\leq\ & \mathrm{pr}\left[\frac{1}{N-m}\max_{k_1}\left|\sum_{i=1}^{m}\left\{\frac{1}{n_i}S_{i\cdot k}S_{i\cdot l} - (\Sigma_\varepsilon^0)_{k,l}\right\}\right| \geq \frac{C_1(N\log p)^{1/2}}{2(N-m)}\right] \\
& \quad + \mathrm{pr}\left[\max_{k,l}\frac{1}{N-m}\left|\sum_{i=1}^{m}\sum_{j=1}^{n_i}\{\varepsilon_{ijk}\varepsilon_{ijl} - (\Sigma_\varepsilon^0)_{k,l}\}\right| \geq \frac{C_1(N\log p)^{1/2}}{2(N-m)}\right] \\
\leq\ & 2\exp\left\{\max\left(2 - \frac{c_3 C_1^2}{4c_1^2 M_\varepsilon^2}, 2 - \frac{c_3 C_1}{2c_1 M_\varepsilon}\right)\log p\right\} \\
& \quad + 2\exp\left\{\max\left(2 - \frac{c_2 C_1^2}{4c_1^2 M_\varepsilon^2}, 2 - \frac{c_2 C_1}{2c_1 M_\varepsilon}\right)\log p\right\} \\
\leq\ & 4p^{-C_2},
\end{aligned}$$

where $C_2 = \min\{c_3 C_1(2c_1 M_\varepsilon)^{-1}, c_3, c_2 C_1(2c_1 M_\varepsilon)^{-1}, c_2\}(2c_1 M_\varepsilon)^{-1}C_1 - 2.$ $\square$

Lemma 2. *Consider the true within-subject covariance* Σ_ε^0 *with* $\max_k(\Sigma_\varepsilon^0)_{k,k} \leq M_\varepsilon$ *and the true between-subject covariance* Σ_b^0 *with* $\max_k(\Sigma_b^0)_{k,k} \leq M_b$*. Let*

$$\lambda_b = C_1 \left(\frac{\log p}{m}\right)^{1/2} + C_2 \frac{(N \log p)^{1/2}}{(N-m)n^*} + \frac{M_b}{m} + \frac{M_\varepsilon}{mn^*}$$

for sufficiently large $C_1, C_2 > 0$*, where* $n^* = m / \sum_{i=1}^m n_i^{-1}$*. If* $\log p \leq m$*, then the unbiased between-subject sample estimate* $\widehat{\Sigma}_b$ *satisfies*

$$pr\left\{\max_{k,l}\left|(\widehat{\Sigma}_b - \Sigma_b^0)_{k,l}\right| > 2\lambda_b\right\} \leq 8p^{-C_3},$$

where $C_3 > 0$ *only depends on* C_1*,* C_2 *and* $\max(M_\varepsilon, M_b)$*.*

Proof. Let $\bar{Y}_{i\cdot k} = b_{ik} + n_i^{-1}\sum_{j=1}^{n_i}\varepsilon_{ijk} = b_{ik} + n_i^{-1}S_{i\cdot k} = W_{ik}$, then by decomposition,

$$\begin{aligned}
(\widehat{\Sigma}_b - \Sigma_b^0)_{k,l} &= \{\overline{\Sigma} - (n^*)^{-1}\widehat{\Sigma}_\varepsilon - \Sigma_b^0\}_{k,l} \\
&= [\overline{\Sigma} - \{\Sigma_b^0 + (n^*)^{-1}\Sigma_\varepsilon^0\}]_{k,l} - (n^*)^{-1}(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l} \\
&= \frac{1}{m-1}\sum_{i=1}^m \left\{W_{ik}W_{il} - \left(\Sigma_b^0 + n_i^{-1}\Sigma_\varepsilon^0\right)_{k,l}\right\} \\
&\quad - \frac{m}{m-1}\left(\frac{1}{m}\sum_{i=1}^m W_{ik}\right)\left(\frac{1}{m}\sum_{i=1}^m W_{il}\right) \\
&\quad + \frac{(\Sigma_b^0)_{k,l}}{m-1} + \frac{(\Sigma_\varepsilon^0)_{k,l}}{(m-1)n^*} - \frac{(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}}{n^*}. \qquad \text{(S3.11)}
\end{aligned}$$

Then, with $|(\Sigma_b^0)_{k,l}| \leq M_b$ and $|(\Sigma_\varepsilon^0)_{k,l}| \leq M_\varepsilon$, we have

$$\begin{aligned}
\max_{k,l}\left|(\widehat{\Sigma}_b - \Sigma_b^0)_{k,l}\right| \;\leq\; & 2\max_{k,l}\left|\frac{1}{m}\sum_{i=1}^{m}\left\{W_{ik}W_{il} - \left(\Sigma_b^0 + n_i^{-1}\Sigma_\varepsilon^0\right)_{k,l}\right\}\right| \\
& +2\max_{k,l}\left|\left(\frac{1}{m}\sum_{i=1}^{m}W_{ik}\right)\left(\frac{1}{m}\sum_{i=1}^{m}W_{il}\right)\right| \\
& +\max_{k,l}(n^*)^{-1}\left|(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}\right| \\
& +\frac{2M_b}{m} + \frac{2M_\varepsilon}{mn^*}. \qquad \text{(S3.12)}
\end{aligned}$$

Assume that $b_{ik} \in \mathcal{SG}(\sigma_{b,k}^2)$, i.e., b_{ik} is sub-Gaussian with a variance factor $\sigma_{b,k}^2$ for $1 \leq i \leq m, 1 \leq k \leq p$. Then $W_{ik} \in \mathcal{SG}(\sigma_{b,k}^2 + n_i^{-1}\sigma_{\varepsilon,k}^2)$. Let $\max_k \sigma_{b,k}^2 = M_b$. Then, by Lemma 2.7.7 in Vershyin [2018], we obtain

$$\begin{aligned}
\left\|W_{ik}W_{il} - \left(\Sigma_b^0 + n_i^{-1}\Sigma_\varepsilon^0\right)_{k,l}\right\|_{\psi_1} \;\leq\; & 2\left\|W_{ik}\right\|_{\psi_2}\left\|W_{il}\right\|_{\psi_2} \\
\leq\; & c_1\left(\Sigma_b^0 + n_i^{-1}\Sigma_\varepsilon^0\right)_{k,l} \\
\leq\; & c_1\left(1 + n_l^{-1}\right)M_* \\
\leq\; & 2c_1M_*,
\end{aligned}$$

where $n_l = \min n_i$ and $M_* = \max(M_\varepsilon, M_b)$. And with the Bernstein's inequality, we have

$$\mathrm{pr}\left[\frac{1}{m}\left|\sum_{i=1}^{m}\left\{W_{ik}W_{il} - (\Sigma_b^0 + n_i^{-1}\Sigma_\varepsilon^0)_{k,l}\right\}\right| \geq t\right] \leq 2\exp\left\{-c_4\min\left(\frac{mt^2}{K_3^2}, \frac{mt}{K_3}\right)\right\},$$

where $c_4 > 0$, $K_3 = \max_{i,k,l}\|W_{ik}W_{il} - (\Sigma_b^0 + n_i^{-1}\Sigma_\varepsilon^0)_{k,l}\|_{\psi_1} \leq 2c_1M_*$.

By the union sum inequality and taking $t = 2^{-1}C_1(\log p/m)^{1/2}$ for a sufficiently large constant $C_1 > 0$, if $m \geq \log p$, we have

$$
\begin{aligned}
& \operatorname{pr}\left[\max_{k,l} \frac{1}{m}\left|\sum_{i=1}^{m}\left\{W_{ik}W_{il} - (\Sigma_b^0 + n_i^{-1}\Sigma_\varepsilon^0)_{k,l}\right\}\right| \geq t\right] \\
\leq\ & 2p^2 \exp\left[-c_4 \min\left\{\frac{C_1^2 \log p}{4K_3^2}, \frac{C_1(m\log p)^{1/2}}{2K_3}\right\}\right] \\
\leq\ & 2\exp\left[\left\{2 - \min\left(\frac{c_4 C_1^2}{16c_1^2 M_*^2}, \frac{c_4 C_1}{4c_1 M_*}\right)\right\}\log p\right]. \qquad \text{(S3.13)}
\end{aligned}
$$

We use a union bound with the general Hoeffding's inequality (Theorem 2.6.2 by Vershyin [2018]) to bound the second term in (S3.12). Specifically, with $m \geq \log p$ and taking $t = 2^{-1}C_1(\log p/m)^{1/2}$, we have

$$
\begin{aligned}
\operatorname{pr}\left(\max_{k,l}\left|\frac{1}{m}\sum_{i=1}^{m} W_{ik}\right|^2 \geq t\right) &= \operatorname{pr}\left(\max_{k,l}\left|\sum_{i=1}^{m} W_{ik}\right| \geq mt^{1/2}\right) \\
&\leq 2p\exp\left(-\frac{c_5 m^2 t}{\sum_{i=1}^{m}\|W_{ik}\|_{\psi_2}^2}\right) \\
&\leq 2p\exp\left(-\frac{c_5 m t}{c_1 M_*}\right) \\
&= 2p\exp\left\{-\frac{c_5 C_1}{2c_1 M_*}(m\log p)^{1/2}\right\} \\
&\leq 2\exp\left\{\left(1 - \frac{c_5 C_1}{2c_1 M_*}\right)\log p\right\}, \qquad \text{(S3.14)}
\end{aligned}
$$

where $c_5 > 0$.

For the third term in (S3.12), by Lemma 1, for a sufficiently large constant $C_2 > 0$, we have

$$
\operatorname{pr}\left\{\max_{k,l}\frac{\left|(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}\right|}{n^*} \geq 2C_2\frac{(N\log p)^{1/2}}{(N-m)n^*}\right\} \leq 4p^{-C_3'}, \qquad \text{(S3.15)}
$$

where $C_3' > 0$ only depends on C_2 and M_ε.

Collecting (S3.13)-(S3.15), with

$$\lambda_b = C_1\left(\frac{\log p}{m}\right)^{1/2} + C_2\frac{(N\log p)^{1/2}}{(N-m)n_*} + \frac{M_b}{m} + \frac{M_\varepsilon}{mn_*},$$

we have

$$\begin{aligned}
& \mathrm{pr}\left\{\max_{k,l}\left|(\widehat{\Sigma}_b - \Sigma_b^0)_{k,l}\right| \geq 2\lambda_b\right\} \\
\leq\ & \mathrm{pr}\left[\max_{k,l}\frac{2}{m}\left|\sum_{i=1}^m\left\{W_{ik}W_{il} - (\Sigma_b^0 + n_i^{-1}\Sigma_\varepsilon^0)_{k,l}\right\}\right| \geq C_1\left(\frac{\log p}{m}\right)^{1/2}\right] \\
& \quad + \mathrm{pr}\left\{2\max_{k,l}\left|\left(\frac{1}{m}\sum_{i=1}^m W_{ik}\right)\left(\frac{1}{m}\sum_{i=1}^m W_{il}\right)\right| \geq C_1\left(\frac{\log p}{m}\right)^{1/2}\right\} \\
& \quad + \mathrm{pr}\left\{\max_{k,l}\frac{\left|(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}\right|}{n^*} \geq 2C_2\frac{(N\log p)^{1/2}}{(N-m)n^*}\right\} \\
\leq\ & 4p^{-C_3'} + 4p^{-C_3''} \\
\leq\ & 8p^{-C_3},
\end{aligned}$$

where $C_3'' = \min\{c_4C_1^2(16c_1^2M_*^2)^{-1}, c_4C_1(4c_1M_*)^{-1}, c_5C_1(2c_1M_*)^{-1} + 1\} - 2$ and $C_3 = \min(C_3', C_3'')$. □

Lemma 3. *Consider the true within-subject covariance Σ_ε^0 with $\max_k(\Sigma_\varepsilon^0)_{k,k} \leq M_\varepsilon$ and the true between-subject covariance Σ_b^0 with $\max_k(\Sigma_b^0)_{k,k} \leq M_b$. Let*

$$\lambda_0 = C_1\left(\frac{\log p}{m}\right)^{1/2} + \frac{M_b}{m} + \frac{M_\varepsilon}{n^*}$$

for sufficiently large $C_1 > 0$, where $n^* = m/\sum_{i=1}^m n_i^{-1}$. If $\log p \leq m$, then the naive between-subject sample estimate $\overline{\Sigma}$ satisfies

$$pr\left\{\max_{k,l}\left|(\overline{\Sigma} - \Sigma_b^0)_{k,l}\right| > 2\lambda_b\right\} \leq 8p^{-C_2}$$

where $C_2 > 0$ *only depends on* C_1 *and* $\max(M_\varepsilon, M_b)$.

Proof. Now, we will consider the convergence rate of $\max_{k,l} |(\overline{\Sigma} - \Sigma_b^0)_{k,l}|$. By (S3.11), we have

$$\begin{aligned}(\overline{\Sigma} - \Sigma_b^0)_{k,l} &= \frac{1}{m-1} \sum_{i=1}^{m} \left\{ W_{ik} W_{il} - \left(\Sigma_b^0 + n_i^{-1} \Sigma_\varepsilon^0\right)_{k,l} \right\} \\ &\quad - \frac{m}{m-1} \left(\frac{1}{m} \sum_{i=1}^{m} W_{ik} \right) \left(\frac{1}{m} \sum_{i=1}^{m} W_{il} \right) \\ &\quad + \frac{(\Sigma_b^0)_{k,l}}{m-1} + \frac{m(\Sigma_\varepsilon^0)_{k,l}}{(m-1)n^*}. \end{aligned} \tag{S3.16}$$

Then, with $|(\Sigma_b^0)_{k,l}| \leq M_b$ and $|(\Sigma_\varepsilon^0)_{k,l}| \leq M_\varepsilon$, we have

$$\begin{aligned}\max_{k,l} \left|(\overline{\Sigma} - \Sigma_b^0)_{k,l}\right| &\leq 2 \max_{k,l} \left| \frac{1}{m} \sum_{i=1}^{m} \left\{ W_{ik} W_{il} - \left(\Sigma_b^0 + n_i^{-1} \Sigma_\varepsilon^0\right)_{k,l} \right\} \right| \\ &\quad + 2 \max_{k,l} \left| \left(\frac{1}{m} \sum_{i=1}^{m} W_{ik} \right) \left(\frac{1}{m} \sum_{i=1}^{m} W_{il} \right) \right| \\ &\quad + \frac{2M_b}{m} + \frac{2M_\varepsilon}{n^*}. \end{aligned} \tag{S3.17}$$

Following the steps in Lemma 2, with

$$\lambda_0 = C_1 \left(\frac{\log p}{m} \right)^{1/2} + \frac{M_b}{m} + \frac{M_\varepsilon}{n^*}$$

for a sufficiently large constant $C_1 > 0$, we have

$$\mathrm{pr} \left\{ \max_{k,l} \left|(\overline{\Sigma} - \Sigma_b^0)_{k,l}\right| > 2\lambda_0 \right\} \leq 4p^{-C_2},$$

where $C_2 > 0$ only depends on C_1 and $\max(M_\varepsilon, M_b)$. $\square$

Lemma 4. *Consider the true within-subject covariance* Σ_ε^0 *with* $\max_k(\Sigma_\varepsilon^0)_{k,k} \leq M_\varepsilon$ *and the true between-subject covariance* Σ_b^0 *with* $\max_k(\Sigma_b^0)_{k,k} \leq M_b$. *Let*

$$\widetilde{\lambda}_b = C_1 \frac{\max_i n_i}{n_0} \left(\frac{\log p}{m}\right)^{1/2} + C_2 \frac{(N \log p)^{1/2}}{n_0(N-m)} + \frac{(2N - n_0 m)M_b}{2n_0 m} + \frac{M_\varepsilon}{n_0 m}$$

for sufficiently large $C_1, C_2 > 0$. *If* $\log p \leq m$, *then* $\widetilde{\Sigma}_b$ *satisfies*

$$pr\left[\max_{k,l} \left|(\widetilde{\Sigma}_b - \Sigma_b^0)_{k,l}\right| > 2\widetilde{\lambda}_b\right] \leq 8p^{-C_3},$$

where $C_3 > 0$ *only depends on* C_1, C_2 *and* $\max(M_\varepsilon, M_b)$.

Proof. Consider

$$\begin{aligned}\max_{k,l} |(\widetilde{\Sigma}_b - \Sigma_b^0)_{k,l}| = \;& \max_{k,l} \left| \left(\frac{\overline{\Sigma} - \widehat{\Sigma}_\varepsilon}{n_0} - \Sigma_b^0\right)_{k,l} \right| = \max_{k,l} \left| \frac{(\overline{\Sigma} - n_0\Sigma_b^0 - \Sigma_\varepsilon^0)_{k,l}}{n_0} - \frac{(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}}{n_0} \right| \\ & \leq \max_{k,l} \left| \frac{(\overline{\Sigma} - n_0\Sigma_b^0 - \Sigma_\varepsilon^0)_{k,l}}{n_0} \right| + \max_{k,l} \left| \frac{(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}}{n_0} \right|. \quad \text{(S3.18)}\end{aligned}$$

With $\bar{Y}_{i\cdot k} = W_{ik}$, we have

$$\begin{aligned}\bar{Y}_{\cdot\cdot k} &= \frac{1}{N}\sum_{i=1}^m n_i W_{ik}, \\ (\overline{\Sigma})_{k,l} &= \frac{1}{m-1}\sum_{i=1}^m n_i \left(W_{ik} - \frac{1}{N}\sum_{i=1}^m n_i W_{ik}\right)\left(W_{il} - \frac{1}{N}\sum_{i=1}^m n_i W_{il}\right) \\ &= \frac{1}{m-1}\sum_{i=1}^m n_i W_{ik} W_{il} - \frac{1}{(m-1)N}\left(\sum_{i=1}^m n_i W_{ik}\right)\left(\sum_{i=1}^m n_i W_{il}\right).\end{aligned}$$

Thus, we obtain

$$
\begin{aligned}
&\frac{(\overline{\Sigma} - n_0\Sigma_b^0 - \Sigma_\varepsilon^0)_{k,l}}{n_0} \\
= \; & \frac{1}{n_0(m-1)} \sum_{i=1}^m \{n_i W_{ik} W_{il} - (n_i \Sigma_b^0 + \Sigma_\varepsilon^0)_{k,l}\} \\
& \quad - \frac{1}{n_0(m-1)N} \left(\sum_{i=1}^m n_i W_{ik} \right) \left(\sum_{i=1}^m n_i W_{il} \right) \\
& \quad + \left\{ \frac{N}{n_0(m-1)} - 1 \right\} (\Sigma_b^0)_{k,l} + \frac{1}{n_0(m-1)} (\Sigma_\varepsilon^0)_{k,l}.
\end{aligned}
$$

Then, for the first term in (S3.18), with $|(\Sigma_b^0)_{k,l}| \le M_b$ and $|(\Sigma_\varepsilon^0)_{k,l}| \le M_\varepsilon$, we have

$$
\begin{aligned}
&\max_{k,l} \left| \frac{(\overline{\Sigma} - n_0\Sigma_b^0 - \Sigma_\varepsilon^0)_{k,l}}{n_0} \right| \\
\le & \max_{k,l} \frac{2}{n_0 m} \left| \sum_{i=1}^m \{n_i W_{ik} W_{il} - (n_i \Sigma_b^0 + \Sigma_\varepsilon^0)_{k,l}\} \right| \\
& \quad + \max_k \frac{2}{n_0 m N} \left| \sum_{i=1}^m n_i W_{ik} \right|^2 + \left\{ \frac{2N}{n_0 m} - 1 \right\} M_b + \frac{2}{n_0 m} M_\varepsilon. \qquad \text{(S3.19)}
\end{aligned}
$$

Recall that by assumptions $b_{ik} \in \mathcal{SG}(\sigma_{b,k}^2)$, i.e., b_{ik} is sub-Gaussian with a variance factor $\sigma_{b,k}^2$ for $1 \le i \le m, 1 \le k \le p$. Then we have $W_{ik} \in SG(\sigma_{b,k}^2 + n_i^{-1}\sigma_{\varepsilon,k}^2)$. Let $\max_k \sigma_{\varepsilon,k}^2 = M_\varepsilon$ and $\max_k \sigma_{b,k}^2 = M_b$. Together with (S3.5) and (S3.6), we get $\|n_i W_{ik} W_{il} - (n_i \Sigma_b^0 + \Sigma_\varepsilon^0)_{k,l}\|_{\psi_1} \le c_1(n_i M_b + M_\varepsilon)$. Then, $n_i W_{ik} W_{il} - (n_i \Sigma_b^0 + \Sigma_\varepsilon^0)_{k,l}$ is sub-Exponential. With Bernstein's inequality, for any k, l, we have

$$
\mathrm{pr}\left[\left| \frac{1}{n_0 m} \sum_{i=1}^m \left\{ n_i W_{ik} W_{il} - (n_i \Sigma_b^0 + \Sigma_\varepsilon^0)_{k,l} \right\} \right| \ge t \right] \le 2\exp\left\{ -c_6 \min\left(\frac{t^2 n_0^2 m}{K_4^2}, \frac{t n_0 m}{K_4} \right) \right\},
$$

where $c_6 > 0$, $K_4 = \max_{i,k,l} \|n_i W_{ik} W_{il} - (n_i \Sigma_b^0 + \Sigma_\varepsilon^0)_{k,l}\|_{\psi_1} \leq 2c_1 n_u M_*$, $n_u = \max_i n_i$, $M_* = \max(M_\varepsilon, M_b)$.

Take $t = C_1 n_u (2n_0)^{-1} (\log p/m)^{1/2}$ for a sufficiently large constant $C_1 > 0$. With $m \geq \log p$ and the union sum inequality, we obtain

$$\begin{aligned}
& \mathrm{pr}\left[\max_{k,l}\left|\frac{1}{n_0 m}\sum_{i=1}^{m}\left\{n_i W_{ik} W_{il} - (n_i \Sigma_b^0 + \Sigma_\varepsilon^0)_{k,l}\right\}\right| \geq t\right] \\
\leq\ & 2p^2 \exp\left\{-c_6 \min\left(\frac{t^2 n_0^2 m}{4c_1^2 n_u^2 M_*^2}, \frac{t n_0 m}{2c_1 n_u M_*}\right)\right\} \\
=\ & 2\exp\left[2\log p - \min\left\{\frac{c_6 C_1^2}{16 c_1^2 M_*^2}\log p, \frac{c_6 C_1}{4c_1 M_*}(m\log p)^{1/2}\right\}\right] \\
\leq\ & 2\exp\left[\left\{2 - \min\left(\frac{c_6 C_1^2}{16 c_1^2 M_*^2}, \frac{c_6 C_1}{4c_1 M_*}\right)\right\}\log p\right]. \qquad \text{(S3.20)}
\end{aligned}$$

Then we will bound the second term in (S3.19). By the property of sub-Gaussian assumption, $n_i W_{ik} = n_i b_{ik} + \sum_{j=1}^{n_i} \varepsilon_{ijk} \in \mathcal{SG}(n_i^2 M_b + n_i M_\varepsilon)$. Then, according to the general Hoeffding's inequality (Theorem 2.6.2 by Vershyin [2018]), we have

$$\begin{aligned}
\mathrm{pr}\left(\frac{1}{n_0 m N}\left|\sum_{i=1}^{m} n_i W_{ik}\right|^2 \geq t\right) &\leq \mathrm{pr}\left\{\left|\sum_{i=1}^{m} n_i W_{ik}\right| \geq (t n_0 m N)^{1/2}\right\} \\
&\leq 2\exp\left\{-\frac{c_7 t n_0 m N}{\sum_{i=1}^{m} c_0^2 (n_i^2 M_b + n_i M_\varepsilon)}\right\} \\
&\leq 2\exp\left\{-\frac{c_7 t n_0 m N}{\sum_{i=1}^{m} c_0^2 (n_i n_u M_b + n_i M_\varepsilon)}\right\} \\
&\leq 2\exp\left(-\frac{c_7 t n_0 m}{c_1 n_u M_*}\right),
\end{aligned}$$

where $c_7 > 0$.

Then, take $t = C_1 n_u (2n_0)^{-1} (\log p/m)^{1/2}$, with $m \geq \log p$, by the union sum inequality, we have

$$\begin{aligned}
\mathrm{pr}\left(\max_k \frac{1}{n_0 m N}\left|\sum_{i=1}^m n_i W_{ik}\right|^2 \geq t\right) &\leq 2p\exp\left(-\frac{c_7 t n_0 m}{c_1 n_u M_*}\right) \\
&\leq 2p\exp\left\{-\frac{c_7 C_1}{2c_1 M_*}(m\log p)^{1/2}\right\} \\
&\leq 2\exp\left\{\left(1-\frac{c_7 C_1}{2c_1 M_*}\right)\log p\right\} \quad \text{(S3.21)}
\end{aligned}$$

for a sufficiently large constant $C_1 > 0$.

To bound the second term in (S3.18), by Lemma 1, for a sufficiently large constant $C_2 > 0$, we have

$$\mathrm{pr}\left\{\max_{k,l}\left|\frac{(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}}{n_0}\right| > 2C_2\frac{(N\log p)^{1/2}}{n_0(N-m)}\right\} \leq 4p^{-C_3'}, \quad \text{(S3.22)}$$

where $C_3' > 0$ only depends on C_2 and M_ε.

Then, with

$$\widetilde{\lambda}_b = C_1\frac{\max n_i}{n_0}\left(\frac{\log p}{m}\right)^{1/2} + C_2\frac{(N\log p)^{1/2}}{n_0(N-m)} + \frac{(2N-n_0 m)M_b}{2n_0 m} + \frac{M_\varepsilon}{n_0 m}$$

for sufficiently large $C_1, C_2 > 0$, combining (S3.18)-(S3.22), we obtain

$$
\begin{aligned}
& \operatorname{pr}\left\{\max_{k,l}\left|(\widetilde{\Sigma}_b - \Sigma_b^0)_{k,l}\right| > 2\widetilde{\lambda}_b\right\} \\
\leq \; & \operatorname{pr}\left[\max_{k,l}\left|\frac{2}{n_0 m}\sum_{i=1}^{m}\left\{n_i W_{ik} W_{il} - (n_i \Sigma_b^0 + \Sigma_\varepsilon^0)_{k,l}\right\}\right| \geq C_1 \frac{\max_i n_i}{n_0}\left(\frac{\log p}{m}\right)^{1/2}\right] \\
& \quad + \operatorname{pr}\left\{\max_k \frac{2}{n_0 m N}\left|\sum_{i=1}^{m} n_i W_{ik}\right|^2 \geq C_1 \frac{\max_i n_i}{n_0}\left(\frac{\log p}{m}\right)^{1/2}\right\} \\
& \quad + \operatorname{pr}\left\{\max_{k,l}\left|\frac{(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}}{n_0}\right| > 2C_2 \frac{(N\log p)^{1/2}}{n_0(N-m)}\right\} \\
\leq \; & 4p^{-C_3''} + 4p^{-C_3'} \\
\leq \; & 8p^{-C_3},
\end{aligned}
$$

where $C_3'' = \min\{c_6 C_1^2 (16 c_1^2 M_*^2)^{-1}, c_6 C_1 (4 c_1 M_*)^{-1}, c_7 C_1 (2 c_1 M_*)^{-1} + 1\} - 2$ and $C_3 = \min\{C_3', C_3''\}$. □

Lemma 5. *Consider the true within-subject covariance Σ_ε^0 with $\max_k(\Sigma_\varepsilon^0)_{k,k} \leq M_\varepsilon$ and the true between-subject covariance Σ_b^0 with $\max_k(\Sigma_b^0)_{k,k} \leq M_b$. Let*

$$
\lambda_1 = C_1\left(\frac{\log p}{m}\right)^{1/2} + M_b + \frac{(2-n^*)M_\varepsilon}{2n^*}
$$

for sufficiently large $C_1 > 0$, where $n^* = m/\sum_{i=1}^m n_i^{-1}$. If $\log p \leq m$, then $\overline{\Sigma}$ satisfies

$$
pr\left\{\max_{k,l}\left|(\overline{\Sigma} - \Sigma_\varepsilon^0)_{k,l}\right| > 2\lambda_1\right\} \leq 4p^{-C_2}
$$

where $C_2 > 0$ only depends on C_1 and $\max(M_\varepsilon, M_b)$.

Proof. Now, we will consider the convergence rate of $\max_{k,l}|(\overline{\Sigma}-\Sigma_\varepsilon^0)_{k,l}|$. Note that $(\overline{\Sigma}-\Sigma_\varepsilon^0)_{k,l}=(\overline{\Sigma}-\Sigma_b^0)_{k,l}+(\Sigma_b^0)_{k,l}-(\Sigma_\varepsilon^0)_{k,l}$. Then, by (S3.16), with $|(\Sigma_b^0)_{k,l}|\leq M_b$ and $|(\Sigma_\varepsilon^0)_{k,l}|\leq M_\varepsilon$, we have

$$\begin{aligned}\max_{k,l}\left|(\overline{\Sigma}-\Sigma_\varepsilon^0)_{k,l}\right| \;\leq\;& 2\max_{k,l}\left|\frac{1}{m}\sum_{i=1}^m\left\{W_{ik}W_{il}-\left(\Sigma_b^0+n_i^{-1}\Sigma_\varepsilon^0\right)_{k,l}\right\}\right|\\ &+2\max_{k,l}\left|\left(\frac{1}{m}\sum_{i=1}^m W_{ik}\right)\left(\frac{1}{m}\sum_{i=1}^m W_{il}\right)\right|\\ &+2M_b+\frac{(2-n^*)M_\varepsilon}{n^*}. \end{aligned}\tag{S3.23}$$

Following the steps in Lemma 2, with

$$\lambda_1=C_1\left(\frac{\log p}{m}\right)^{1/2}+M_b+\frac{(2-n^*)M_\varepsilon}{2n^*}$$

for a sufficiently large constant $C_1>0$, we have

$$\mathrm{pr}\left\{\max_{k,l}\left|(\overline{\Sigma}-\Sigma_\varepsilon^0)_{k,l}\right|>2\lambda_1\right\}\leq 4p^{-C_2},$$

where $C_2>0$ only depends on C_1 and $\max(M_\varepsilon,M_b)$. □

S3.2 Lemmas for the Polynomial-tail Condition

We observe $\boldsymbol{Y}_{ij}\in\mathbb{R}^p$, which is the j-th repeated measurement of the i-th subject for $j=1,\dots,n_i$ and $i=1,\dots,m$, following the model (2.1) in the paper, where $\boldsymbol{\varepsilon}_{ij}$ and $\boldsymbol{b}_i$ are p-dimensional random vectors with the true within-subject and between-subject covariance, i.e., $\mathrm{cov}(\boldsymbol{\varepsilon}_{ij})=\Sigma_\varepsilon^0$ and $\mathrm{cov}(\boldsymbol{b}_i)=\Sigma_b^0$, respectively, and $\boldsymbol{b}_i$ and $\boldsymbol{\varepsilon}_{ij}$ are mutually independent. Let $N = \sum_{i=1}^{m} n_i$ be the total number of observations.

A generic random variable Z with mean 0 satisfies the polynomial-tail condition with constant K_z if for all $\gamma > 0$ and $\delta > 0$, the following holds

$$\mathrm{E}\left[|Z|^{4(1+\gamma+\delta)}\right] \leq K_z \tag{S3.24}$$

for some $K_z > 0$. We assume, for $i = 1, \ldots, m$ and $k = 1, \ldots, p$, that b_{ik} follows (S3.24) with a constant K_b, and for $j = 1, \ldots, n_i$, that ε_{ijk} follows (S3.24) with a constant K_ε.

Lemma 6. *Suppose Z_i with mean 0 are independent across $i = 1, \ldots, n$ and satisfy the polynomial-tail condition with constant K_z. Then $n^{-1/2} \sum_{i=1}^{n} Z_{ik}$ also satisfies the polynomial-tail condition* (S3.24)*.*

Proof. Note that

$$\begin{aligned}
\mathrm{E}\left[\left|\sum_{i=1}^{n} Z_i\right|^{4(1+\gamma+\delta)}\right] &\leq B_{\gamma,\delta}\mathrm{E}\left[\left(\sum_{i=1}^{n}|Z_i|^2\right)^{2(1+\gamma+\delta)}\right] \\
&= B_{\gamma,\delta} n^{2(1+\gamma+\delta)}\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}|Z_i|^2\right)^{2(1+\gamma+\delta)}\right] \\
&\leq B_{\gamma,\delta} n^{2(1+\gamma+\delta)}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[|Z_i|^{4(1+\gamma+\delta)}\right] \\
&\leq B_{\gamma,\delta} n^{2(1+\gamma+\delta)} K_z,
\end{aligned}$$

where the first inequality is the Marcinkiewicz–Zygmund inequality, the second inequality above is the Jensen's inequality, and the last inequality holds from the definition in (S3.24). Therefore,

$$\mathrm{E}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i\right|^{4(1+\gamma+\delta)}\right] \leq B_{\gamma,\delta}K_z.$$

As a result, $n^{-1/2}\sum_{i=1}^{n} Z_i$ also satisfies the polynomial-tail condition with constant $B_{\gamma,\delta}K_z$. □

Lemma 7. *Suppose Z_{ik} with mean 0 are independent across $i = 1, \ldots, n$ and satisfy the polynomial-tail condition with constant K_z for $k = 1, \ldots, p$. For any $M > 0$, let*

$$\epsilon = 8(K_z + 1)(M + 1)\left(\frac{\log p}{n}\right)^{1/2},$$

if $p \leq cn^{\gamma}$ for some c, then the following concentration inequality hold:

$$pr\left[\max_k \left|\sum_{i=1}^{n} Z_{ik}\right| > n\epsilon\right] \leq O(p^{-M-1}) + K_z p(\log n)^{2(1+\gamma+\delta)} n^{-(\gamma+\delta)}. \qquad \text{(S3.25)}$$

Proof. This result is proved in the proof of Theorem 2 in Xue et al. [2012]. □

Lemma 8. *Suppose Z_{ik} with mean 0 are independent across $i = 1, \ldots, n$ and satisfy the polynomial-tail condition with constant K_z for $k = 1, \ldots, p$. For any $M > 0$, let*

$$\epsilon = 8(K_z + 1)(M + 2)\left(\frac{\log p}{n}\right)^{1/2},$$

if $p \leq cn^{\gamma}$ for some c, then the following concentration inequality hold:

$$pr\left[\max_{k,\ell} \left|\sum_{i=1}^{n} (Z_{ik}Z_{i\ell} - \mathrm{E}[Z_{ik}Z_{i\ell}])\right| > n\epsilon\right] \leq O(p^{-M}) + K_z p(\log n)^{2(1+\gamma+\delta)} n^{-(\gamma+\delta)}.$$

Proof. This result is proved in the proof of Theorem 2 in Xue et al. [2012]. □

Lemma 9. *Suppose the random errors $\boldsymbol{\varepsilon}_{ij}$'s are i.i.d. across i and j, and has zero-mean entries ε_{ijk} that satisfies the polynomial-tail condition* (S3.24) *with constant K_ε, and consider the true within-subject covariance Σ_ε^0. For any $M > 0$, let $\lambda_\epsilon = 16(K_\varepsilon + 1)(M + 2)\frac{(N \log p)^{1/2}}{N-m}$. If $p \le cN^\gamma$ for some c, then the unbiased within-subject sample estimate $\widehat{\Sigma}_\varepsilon$ satisfies*

$$pr\left\{\max_{k,l}\left|(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}\right| > \lambda_\varepsilon\right\}$$
$$\le O(p^{-M}) + K_\varepsilon p(\log N)^{2(1+\gamma+\delta)}N^{-(\gamma+\delta)} + K_\varepsilon p(\log m)^{2(1+\gamma+\delta)}m^{-(\gamma+\delta)}. \quad \text{(S3.26)}$$

Proof. As in the proof of Lemma 1, we first rewrite $\widehat{\Sigma}_\varepsilon$ as follows:

$$\begin{aligned}
\widehat{\Sigma}_\varepsilon &= \frac{1}{N-m}\sum_{i=1}^m\sum_{j=1}^{n_i}(\boldsymbol{Y}_{ij} - \bar{\boldsymbol{Y}}_{i\cdot})(\boldsymbol{Y}_{ij} - \bar{\boldsymbol{Y}}_{i\cdot})^{\mathrm{T}} \\
&= \frac{1}{N-m}\left(\sum_{i=1}^m\sum_{j=1}^{n_i}\boldsymbol{\varepsilon}_{ij}\boldsymbol{\varepsilon}_{ij}^{\mathrm{T}} - \sum_{i=1}^m n_i\bar{\boldsymbol{\varepsilon}}_{i\cdot}\bar{\boldsymbol{\varepsilon}}_{i\cdot}^{\mathrm{T}}\right),
\end{aligned}$$

which implies that

$$\begin{aligned}
(\widehat{\Sigma}_\varepsilon)_{k,l} &= \frac{1}{N-m}\left(\sum_{i=1}^m\sum_{j=1}^{n_i}\varepsilon_{ijk}\varepsilon_{ijl} - \sum_{i=1}^m n_i\bar{\varepsilon}_{i\cdot k}\bar{\varepsilon}_{i\cdot l}\right) \\
&= \frac{1}{N-m}\sum_{i=1}^m\sum_{j=1}^{n_i}\varepsilon_{ijk}\varepsilon_{ijl} - \frac{1}{N-m}\sum_{i=1}^m n_i\bar{\varepsilon}_{i\cdot k}\bar{\varepsilon}_{i\cdot l} \\
&= \frac{1}{N-m}\sum_{i=1}^m\sum_{j=1}^{n_i}\left\{\varepsilon_{ijk}\varepsilon_{ijl} - (\Sigma_\varepsilon^0)_{k,l}\right\} - \frac{1}{N-m}\sum_{i=1}^m\left\{\frac{1}{n_i}S_{i\cdot k}S_{i\cdot l} - (\Sigma_\varepsilon^0)_{k,l}\right\} + (\Sigma_\varepsilon^0)_{k,l},
\end{aligned}$$

where $S_{i\cdot k} = \sum_{j=1}^{n_i} \varepsilon_{ijk}$. Then we have

$$\max_{k,l} \left| (\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l} \right| \le \max_{k,l} \frac{1}{N-m} \left| \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left\{ \varepsilon_{ijk}\varepsilon_{ijl} - (\Sigma_\varepsilon^0)_{k,l} \right\} \right| + \frac{1}{N-m} \max_{k,l} \left| \sum_{i=1}^{m} \left\{ \frac{1}{n_i} S_{i\cdot k} S_{i\cdot l} - (\Sigma_\varepsilon^0)_{k,l} \right\} \right|. \tag{S3.27}$$

Hence, for the first term in (S3.27), with

$$\epsilon_1 = 8(K_\varepsilon + 1)(M+2) \frac{N}{N-m} \left(\frac{\log p}{N} \right)^{1/2}, \tag{S3.28}$$

Lemma 8 leads to

$$\begin{aligned}
&\mathrm{pr}\left[\frac{1}{N-m} \max_{k,\ell} \left| \sum_{i=1}^{m} \sum_{j=1}^{n_i} \{ \varepsilon_{ijk}\varepsilon_{ij\ell} - (\Sigma_\varepsilon^0)_{k,\ell} \} \right| \ge \epsilon_1 \right] \\
=&\mathrm{pr}\left[\frac{1}{N} \max_{k,\ell} \left| \sum_{i=1}^{m} \sum_{j=1}^{n_i} \{ \varepsilon_{ijk}\varepsilon_{ij\ell} - (\Sigma_\varepsilon^0)_{k,\ell} \} \right| \ge 8(K_\varepsilon+1)(M+2) \left(\frac{\log p}{N} \right)^{1/2} \right] \\
\le& O(p^{-M}) + K_\varepsilon p (\log N)^{2(1+\gamma+\delta)} N^{-(\gamma+\delta)}.
\end{aligned} \tag{S3.29}$$

Similarly, by Lemma 6, $n_i^{-1/2} S_{i\cdot k}$ satisfies polynomial-tail conditions, with

$$\epsilon_2 = 8(K_\varepsilon + 1)(M+2) \frac{m}{N-m} \left(\frac{\log p}{m} \right)^{1/2}, \tag{S3.30}$$

Lemma 8 implies that

$$\begin{aligned}
&\mathrm{pr}\left[\frac{1}{N-m} \max_{k,\ell} \left| \sum_{i=1}^{m} \left\{ \frac{1}{n_i} S_{i\cdot k} S_{i\cdot \ell} - (\Sigma_\varepsilon^0)_{k,\ell} \right\} \right| \ge \epsilon_2 \right] \\
=&\mathrm{pr}\left[\frac{1}{m} \max_{k,\ell} \left| \sum_{i=1}^{m} \left\{ \frac{1}{n_i} S_{i\cdot k} S_{i\cdot \ell} - (\Sigma_\varepsilon^0)_{k,\ell} \right\} \right| \ge 8(K_\varepsilon+1)(M+2) \left(\frac{\log p}{m} \right)^{1/2} \right] \\
\le& O(p^{-M}) + K_\varepsilon p (\log m)^{2(1+\gamma+\delta)} m^{-(\gamma+\delta)}.
\end{aligned}$$

Putting everything together, let

$$\lambda_\epsilon = 16(K_\varepsilon + 1)(M + 2)\frac{(N \log p)^{1/2}}{N - m}, \tag{S3.31}$$

and observe that $\epsilon_1 \geq \epsilon_2$, we have

$$\begin{aligned}
&\mathrm{pr}\left\{\max_{k,l}\left|(\widehat{\Sigma}_\varepsilon - \Sigma^0_\varepsilon)_{k,l}\right| > \lambda_\varepsilon\right\} \\
\leq& \mathrm{pr}\left[\frac{1}{N-m}\max_{k_1}\left|\sum_{i=1}^{m}\left\{\frac{1}{n_i}S_{i\cdot k}S_{i\cdot l} - (\Sigma^0_\varepsilon)_{k,l}\right\}\right| \geq \epsilon_1\right] \\
&\quad + \mathrm{pr}\left[\max_{k,l}\frac{1}{N-m}\left|\sum_{i=1}^{m}\sum_{j=1}^{n_i}\{\varepsilon_{ijk}\varepsilon_{ijl} - (\Sigma^0_\varepsilon)_{k,l}\}\right| \geq \epsilon_1\right] \\
\leq& \mathrm{pr}\left[\frac{1}{N-m}\max_{k_1}\left|\sum_{i=1}^{m}\left\{\frac{1}{n_i}S_{i\cdot k}S_{i\cdot l} - (\Sigma^0_\varepsilon)_{k,l}\right\}\right| \geq \epsilon_2\right] \\
&\quad + \mathrm{pr}\left[\max_{k,l}\frac{1}{N-m}\left|\sum_{i=1}^{m}\sum_{j=1}^{n_i}\{\varepsilon_{ijk}\varepsilon_{ijl} - (\Sigma^0_\varepsilon)_{k,l}\}\right| \geq \epsilon_1\right] \\
\leq& O(p^{-M}) + K_\varepsilon p(\log N)^{2(1+\gamma+\delta)}N^{-(\gamma+\delta)} + K_\varepsilon p(\log m)^{2(1+\gamma+\delta)}m^{-(\gamma+\delta)}.
\end{aligned}$$

$\square$

Lemma 10. *Suppose the random errors $\boldsymbol{\varepsilon}_{ij}$'s are i.i.d. across i and j, and has zero-mean entries ε_{ijk} that satisfies the polynomial-tail condition* (S3.24) *with constant K_ε, and consider the true within-subject covariance Σ^0_ε. Similarly, suppose the random effects $\boldsymbol{b}_i$'s are i.i.d across $i = 1, \ldots, m$, and have zero-mean entries b_{ik} that satisfy the polynomial-tail condition* (S3.24) *with constant K_b,* *and consider the true within-subject covariance Σ_b^0. For any $M > 0$, let*

$$\lambda_b = 16(M+2)\left[(K+1)\left(\frac{\log p}{m}\right)^{1/2} + (K_\varepsilon + 1)\frac{(N\log p)^{1/2}}{(N-m)n^*}\right] + \frac{M_b}{m} + \frac{M_\varepsilon}{mn_*},$$

where $n^* = m/\sum_{i=1}^m n_i^{-1}$ and K is a constant that only depends on K_b and K_ε. If $p \le cm^\gamma$ for some c, then the unbiased between-subject sample estimate $\widehat{\Sigma}_b$ satisfies

$$\begin{aligned} & pr\left\{\max_{k,l}\left|(\widehat{\Sigma}_b - \Sigma_b^0)_{k,l}\right| > \lambda_b\right\} \\ & \le O(p^{-M}) + K_\varepsilon p(\log N)^{2(1+\gamma+\delta)}N^{-(\gamma+\delta)} + Cp(\log m)^{2(1+\gamma+\delta)}m^{-(\gamma+\delta)}. \end{aligned} \tag{S3.32}$$

Proof. Let $\bar{Y}_{i\cdot k} = b_{ik} + n_i^{-1}\sum_{j=1}^{n_i}\varepsilon_{ijk} = b_{ik} + n_i^{-1}S_{i\cdot k} = W_{ik}$, then by decomposition,

$$\begin{aligned} (\widehat{\Sigma}_b - \Sigma_b^0)_{k,l} &= \{\overline{\Sigma} - (n^*)^{-1}\widehat{\Sigma}_\varepsilon - \Sigma_b^0\}_{k,l} \\ &= [\overline{\Sigma} - \{\Sigma_b^0 + (n^*)^{-1}\Sigma_\varepsilon^0\}]_{k,l} - (n^*)^{-1}(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l} \\ &= \frac{1}{m-1}\sum_{i=1}^m\left\{W_{ik}W_{il} - \left(\Sigma_b^0 + n_i^{-1}\Sigma_\varepsilon^0\right)_{k,l}\right\} \\ &\quad -\frac{m}{m-1}\left(\frac{1}{m}\sum_{i=1}^m W_{ik}\right)\left(\frac{1}{m}\sum_{i=1}^m W_{il}\right) \\ &\quad +\frac{(\Sigma_b^0)_{k,l}}{m-1} + \frac{(\Sigma_\varepsilon^0)_{k,l}}{(m-1)n^*} - \frac{(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}}{n^*}. \end{aligned} \tag{S3.33}$$

For each k, by assumptions b_{ik} satisfies the sub-exponential tail condition (S3.24) with constant K_b, and is independent of $n_i^{-1}S_{i\cdot k}$, which itself satisfies the sub-exponential tail condition with K_b. By Lemma 6, we have that W_{ik}'s are zero-mean random variables that satisfy the polynomial-tail condition (S3.24) with constant K_w, which only depends on K_b and K_ε. Furthermore, we note that $\mathrm{E}[W_{ik}W_{il}] = (\Sigma_b^0 + n_i^{-1}\Sigma_\varepsilon^0)_{k,l}$. With Lemma 8 and

$$\epsilon_1 = 8(K_w + 1)(M + 2)\left(\frac{\log p}{m}\right)^{1/2},$$

we have

$$\mathrm{pr}\left[\frac{1}{m}\max_{k,l}\left|\sum_{i=1}^{m}\left\{W_{ik}W_{il} - (\Sigma_b^0 + n_i^{-1}\Sigma_\varepsilon^0)_{k,l}\right\}\right| \geq \epsilon_1\right] \leq O(p^{-M}) + K_w p(\log m)^{2(1+\gamma+\delta)}m^{-(\gamma+\delta)}.$$

Furthermore, with Lemma 7 and

$$\epsilon_2 = 8(K_w + 1)(M + 1)\left(\frac{\log p}{m}\right)^{1/2},$$

we have

$$\mathrm{pr}\left(\max_k\left|\frac{1}{m}\sum_{i=1}^{m}W_{ik}\right|^2 \geq \epsilon_2^2\right) = \mathrm{pr}\left(\max_k\left|\frac{1}{m}\sum_{i=1}^{m}W_{ik}\right| \geq \epsilon_2\right)$$
$$\leq O(p^{-M-1}) + K_w p(\log n)^{2(1+\gamma+\delta)}n^{-(\gamma+\delta)}.$$

For the last term in (S3.33), by Lemma 9, we have

$$\mathrm{pr}\left\{\frac{1}{n^*}\max_{k,l}\left|(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}\right| > \frac{\lambda_\varepsilon}{n^*}\right\}$$
$$\leq O(p^{-M}) + K_\varepsilon p(\log N)^{2(1+\gamma+\delta)}N^{-(\gamma+\delta)} + K_\varepsilon p(\log m)^{2(1+\gamma+\delta)}m^{-(\gamma+\delta)}.$$

Collecting the results and using a union bound, with

$$\lambda_b = 16(K_w + 1)(M + 2)\left(\frac{\log p}{m}\right)^{1/2} + \frac{\lambda_\varepsilon}{n^*} + \frac{M_b}{m} + \frac{M_\varepsilon}{mn_*},$$

we have

$$
\begin{aligned}
& \operatorname{pr}\left\{\max_{k,l}\left|(\widehat{\Sigma}_b-\Sigma_b^0)_{k,l}\right|\geq\lambda_b\right\} \\
\leq\ & \operatorname{pr}\left[\max_{k,l}\frac{1}{m}\left|\sum_{i=1}^{m}\left\{W_{ik}W_{il}-(\Sigma_b^0+n_i^{-1}\Sigma_\varepsilon^0)_{k,l}\right\}\right|\geq\epsilon_1\right] \\
& \quad +\operatorname{pr}\left\{\max_{k,l}\left|\left(\frac{1}{m}\sum_{i=1}^{m}W_{ik}\right)\left(\frac{1}{m}\sum_{i=1}^{m}W_{il}\right)\right|\geq\epsilon_2\right\} \\
& \quad +\operatorname{pr}\left\{\max_{k,l}\frac{\left|(\widehat{\Sigma}_\varepsilon-\Sigma_\varepsilon^0)_{k,l}\right|}{n^*}\geq\frac{\lambda_\varepsilon}{n^*}\right\} \\
\leq\ & O(p^{-M})+K_\varepsilon p(\log N)^{2(1+\gamma+\delta)}N^{-(\gamma+\delta)}+Cp(\log m)^{2(1+\gamma+\delta)}m^{-(\gamma+\delta)}.
\end{aligned}
$$

□

S3.3 Proof of Theorem 1

Proof. Now, we will consider the convergence rate of $\max_{k,l}|(\overline{\Sigma}-\Sigma_b^0)_{k,l}|$. By (S3.16), we have

$$
\begin{aligned}
(\overline{\Sigma}-\Sigma_b^0)_{k,l} &= \frac{1}{m-1}\sum_{i=1}^{m}\left\{W_{ik}W_{il}-\left(\Sigma_b^0+n_i^{-1}\Sigma_\varepsilon^0\right)_{k,l}\right\} \\
&\quad -\frac{m}{m-1}\left(\frac{1}{m}\sum_{i=1}^{m}W_{ik}\right)\left(\frac{1}{m}\sum_{i=1}^{m}W_{il}\right) \\
&\quad +\frac{(\Sigma_b^0)_{k,l}}{m-1}+\frac{m(\Sigma_\varepsilon^0)_{k,l}}{(m-1)n^*}.
\end{aligned}
$$

By the triangle inequality,

$$\max_{k,l}\left|(\overline{\Sigma}-\Sigma_b^0)_{k,l}\right| \geq \max_{k,l}\left|\frac{m(\Sigma_\varepsilon^0)_{k,l}}{(m-1)n^*}+\frac{(\Sigma_b^0)_{k,l}}{m-1}\right| - \frac{2}{m}\max_{k,l}\left|\sum_{i=1}^m\left\{W_{ik}W_{il}-\left(\Sigma_b^0+n_i^{-1}\Sigma_\varepsilon^0\right)_{k,l}\right\}\right|$$
$$-2\max_{k,l}\left|\left(\frac{1}{m}\sum_{i=1}^m W_{ik}\right)\left(\frac{1}{m}\sum_{i=1}^m W_{il}\right)\right|.$$

Following the steps in Lemma 2, i.e., using (S3.13) and (S3.14), with

$$\lambda_{0,b} = \max_{k,l}\left|\frac{m(\Sigma_\varepsilon^0)_{k,l}}{(m-1)n^*}+\frac{(\Sigma_b^0)_{k,l}}{m-1}\right| - C_1\left(\frac{\log p}{m}\right)^{1/2}$$

for a sufficiently large constant $C_1 > 0$, we have

$$\mathrm{pr}\left\{\max_{k,l}\left|(\overline{\Sigma}-\Sigma_b^0)_{k,l}\right| \geq \lambda_{0,b}\right\} \geq 1-4p^{-C_2},$$

where $C_2 > 0$ only depends on C_1 and $\max(M_\varepsilon, M_b)$.

Similarly, we have

$$(\overline{\Sigma}-\Sigma_\varepsilon^0)_{k,l} = \frac{1}{m-1}\sum_{i=1}^m\left\{W_{ik}W_{il}-\left(\Sigma_b^0+n_i^{-1}\Sigma_\varepsilon^0\right)_{k,l}\right\} - \frac{m}{m-1}\left(\frac{1}{m}\sum_{i=1}^m W_{ik}\right)\left(\frac{1}{m}\sum_{i=1}^m W_{il}\right)$$
$$+\frac{m}{m-1}(\Sigma_b^0)_{k,l} - \left(1-\frac{\sum_{i=1}^m n_i^{-1}}{(m-1)}\right)(\Sigma_\varepsilon^0)_{k,l}.$$

By the triangle inequality,

$$\max_{k,l}\left|(\overline{\Sigma}-\Sigma_\epsilon^0)_{k,l}\right| \geq \max_{k,l}\left|\frac{m}{m-1}(\Sigma_b^0)_{k,l} - \left(1-\frac{\sum_{i=1}^m n_i^{-1}}{(m-1)}\right)(\Sigma_\varepsilon^0)_{k,l}\right|$$
$$-\frac{2}{m}\max_{k,l}\left|\sum_{i=1}^m\left\{W_{ik}W_{il}-\left(\Sigma_b^0+n_i^{-1}\Sigma_\varepsilon^0\right)_{k,l}\right\}\right|$$
$$-2\max_{k,l}\left|\left(\frac{1}{m}\sum_{i=1}^m W_{ik}\right)\left(\frac{1}{m}\sum_{i=1}^m W_{il}\right)\right|.$$

Using the same probabilistic argument about the last two terms in the inequality above, and with

$$\lambda_{0,\varepsilon} = \max_{k,l} \left| \frac{m}{m-1} (\Sigma_b^0)_{k,l} - \left(1 - \frac{\sum_{i=1}^m n_i^{-1}}{(m-1)}\right) (\Sigma_\varepsilon^0)_{k,l} \right| - C_1 \left(\frac{\log p}{m}\right)^{1/2},$$

we have, for the same C_1 and C_2, that

$$\mathrm{pr}\left\{\max_{k,l} \left|(\overline{\Sigma} - \Sigma_\epsilon^0)_{k,l}\right| \geq \lambda_{0,\epsilon}\right\} \geq 1 - 4p^{-C_2}.$$

$\square$

S3.4 Proof of Theorem 2

Proof. Define $\Delta_\varepsilon = \Sigma_\varepsilon - \Sigma_\varepsilon^0$ and $F_\varepsilon(\Delta_\varepsilon) = \|\Delta_\varepsilon + \Sigma_\varepsilon^0 - \widehat{\Sigma}_\varepsilon\|_F^2/2 + \lambda_\varepsilon|\Delta_\varepsilon + \Sigma_\varepsilon^0|_1$, then the objective function (3.7) in the main context is equivalent to

$$\min_{\Delta_\varepsilon : \Delta_\varepsilon = \Delta_\varepsilon^{\mathrm{T}}, \Delta_\varepsilon + \Sigma_\varepsilon^0 \succeq \delta I} F_\varepsilon(\Delta_\varepsilon).$$

Consider the set

$$\{\Delta_\varepsilon : \Delta_\varepsilon = \Delta_\varepsilon^{\mathrm{T}}, \Delta_\varepsilon + \Sigma_\varepsilon^0 \succeq \delta I, \|\Delta_\varepsilon\|_F = 5\lambda_\varepsilon (ps_\varepsilon)^{1/2}\}. \tag{S3.34}$$

According to Xue et al. [2012], under the probability event $\{|(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{,k,l}| \leq \lambda_\varepsilon, \forall(i,j)\}$, we have

$$\begin{aligned}
F_\varepsilon(\Delta_\varepsilon) - F_\varepsilon(\mathbf{0}) &\geq \frac{1}{2}\|\Delta_\varepsilon\|_F^2 - 2\lambda_\varepsilon \left[\sum_{k,l=1}^{p} 1\{(\Sigma_\varepsilon^0)_{k,l} \neq 0\}\right]^{1/2} \|\Delta_\varepsilon\|_F \\
&\geq \frac{1}{2}\|\Delta_\varepsilon\|_F^2 - 2\lambda_\varepsilon(ps_\varepsilon)^{1/2}\|\Delta_\varepsilon\|_F \\
&= \frac{5}{2}\lambda_\varepsilon^2 ps_\varepsilon \\
&> 0.
\end{aligned}$$

Note that $F_\varepsilon(\Delta_\varepsilon)$ is a convex function and $F_\varepsilon(\widehat{\Delta}_\varepsilon) \leq F_\varepsilon(\mathbf{0}) = 0$. Then, the minimizer $\widehat{\Delta}_\varepsilon$ must be inside the sphere (S3.34). Hence, we have

$$\begin{aligned}
& \operatorname{pr}\left\{\left\|\widehat{\Sigma}_\varepsilon^+ - \Sigma_\varepsilon^0\right\|_F \leq 5\lambda_\varepsilon(ps_\varepsilon)^{1/2}\right\} \\
\geq\ & 1 - \operatorname{pr}\left\{\max_{k,l}\left|(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}\right| > \lambda_\varepsilon\right\} \\
\geq\ & 1 - 4p^{-C_2}.
\end{aligned}$$

□

The proof of Theorem 3, 4, and 6 in the manuscript and Theorem 7 in this document follow straightforwardly from Theorem 2, combined with Lemma 9, Lemma 2, Lemma 4, and Lemma 10 respectively.

S3.5 Upper Bounds on Estimation Error Rate for Regularized Aggregated Estimator

Theorem 5. *Consider the true between-subject covariance matrix* $\Sigma_b^0 \in \mathcal{U}(M_b, s_b)$ *and the true within-subject covariance matrix* $\Sigma_\varepsilon^0 \in \mathcal{U}(M_\varepsilon, s_\varepsilon)$*. Let*

$$\lambda_1 = C_1 \left(\frac{\log p}{m}\right)^{1/2} + M_b + \frac{(2-n^*)M_\varepsilon}{2n^*}$$

be the value of the tuning parameter λ *in (3.7) in the main context for sufficiently large* $C_1 > 0$*, and the same* n^* *defined in Theorem 3. If* $\log p \leq m$*, then the naive estimator* $\overline{\Sigma}^+$ *satisfies*

$$\left\|\overline{\Sigma}^+ - \Sigma_\varepsilon^0\right\|_F \leq 10\lambda_1 (ps_\varepsilon)^{1/2}$$

with probability at least $1 - 4p^{-C_2}$*, where* $C_2 > 0$ *only depends on* C_1 *and* $\max(M_\varepsilon, M_b)$.

Theorem 6. *Assume that the true between-subject covariance matrix* $\Sigma_b^0 \in \mathcal{U}(M_b, s_b)$ *and the true within-subject covariance matrix* $\Sigma_\varepsilon^0 \in \mathcal{U}(M_\varepsilon, s_\varepsilon)$*. Let*

$$\lambda_0 = C_1 \left(\frac{\log p}{m}\right)^{1/2} + \frac{M_b}{m} + \frac{M_\varepsilon}{n^*}$$

be the value of the tuning parameter λ *in (3.7) for sufficiently large* $C_1 > 0$*, and the same* n^* *defined in Theorem 3. If* $\log p \leq m$*, then the aggregated between-subject estimator* $\overline{\Sigma}^+$ *satisfies*

$$\left\|\overline{\Sigma}^+ - \Sigma_b^0\right\|_F \leq 10\lambda_0 (ps_b)^{1/2}$$

with probability at least $1 - 4p^{-C_2}$*, where* $C_2 > 0$ *only depends on* C_1 *and* $\max(M_\varepsilon, M_b)$*.*

S3.6 Estimation Error for Correlation Matrices

In certain scenarios, estimating between-subject and within-subject correlation matrices (rather than covariance matrices) is of interest. And the sparse and positive-definite estimate of R_ε, denoted as $\widehat{R}_\varepsilon^+$, and of Σ_b, denoted as $\widehat{R}_b^+$ are defined as solution of (3.7) in the main context with $B = \widehat{R}_\varepsilon = D_\varepsilon^{-1/2}\widehat{\Sigma}_\varepsilon D_\varepsilon^{-1/2}$ and $B = \widehat{R}_b = D_b^{-1/2}\widehat{\Sigma}_b D_b^{-1/2}$, where $D_\varepsilon = \operatorname{diag}\{(\widehat{\Sigma}_\varepsilon)_{1,1}, \ldots, (\widehat{\Sigma}_\varepsilon)_{p,p}\}$ and $D_b = \operatorname{diag}\{(\widehat{\Sigma}_b)_{1,1}, \ldots, (\widehat{\Sigma}_b)_{p,p}\}$.

Corollary 1. *Under conditions of Theorem 1, if* $\min_k(\Sigma_\varepsilon^0)_{k,k}$ *is bounded from below, then*

$$\left\|\widehat{R}_\varepsilon^+ - R_\varepsilon^0\right\|_F = O_P\left\{\frac{(ps_\varepsilon N\log p)^{1/2}}{N-m}\right\},$$

uniformly on $\Sigma_\varepsilon^0 \in \mathcal{U}(M_\varepsilon, s_\varepsilon)$*, as* $N, m \to \infty$*.*

Proof. By Lemma 1, we have

$$\operatorname{pr}\left\{\max_{k,l}\left|(\widehat{\Sigma}_\varepsilon - \Sigma_\varepsilon^0)_{k,l}\right| > C_1\frac{(N\log p)^{1/2}}{N-m}\right\} = o(1). \tag{S3.35}$$

By Lemma 2 in Cui et al. [2016], with (S3.35) and the fact $(\widehat{R}_\varepsilon)_{k,l} = (\widehat{\Sigma}_\varepsilon)_{k,l}/\{(\widehat{\Sigma}_\varepsilon)_{k,k}(\widehat{\Sigma}_\varepsilon)_{l,l}\}^{1/2}$, for a sufficiently large constant $C_1' > 0$, we have

$$\operatorname{pr}\left\{\max_{k,l}\left|(\widehat{R}_\varepsilon - R_\varepsilon^0)_{k,l}\right| > C_1'\frac{(N\log p)^{1/2}}{N-m}\right\} = o(1).$$

Following the steps in the proof of Theorem 1, it is easily shown that

$$\left\| \widehat{R}_\varepsilon^+ - R_\varepsilon^0 \right\|_F = O_P \left\{ \frac{(ps_\varepsilon N \log p)^{1/2}}{N-m} \right\}.$$

□

Corollary 2. *Under conditions of Theorem 2, if* $\min_k(\Sigma_\varepsilon^0)_{k,k}$ *and* $\min_k(\Sigma_b^0)_{k,k}$ *are bounded from below, then*

$$\left\| \widehat{R}_b^+ - R_b^0 \right\|_F = O_P \left[(ps_b)^{1/2} \left\{ C_1' \left(\frac{\log p}{m} \right)^{1/2} + C_2' \frac{(N \log p)^{1/2}}{(N-m)n_0} \right\} \right],$$

uniformly on $\Sigma_\varepsilon^0 \in \mathcal{U}(M_\varepsilon, s_\varepsilon)$ *and* $\Sigma_b^0 \in \mathcal{U}(M_b, s_b)$*, for some large* $C_1', C_2' > 0$*, as* $m, n \to \infty$*.*

S3.7 Estimation Error for $\widehat{\Sigma}_b^+$ under the Polynomial-tail Condition

Theorem 7. *In addition to the assumptions in Theorem 3 in the manuscript, suppose that the random effects* $\boldsymbol{b}_i \in \mathbb{R}^p$ *are i.i.d. random vectors (for* $i = 1, \ldots, m$*) with mean zero and between-subject covariance matrix* $\Sigma_b \in \mathcal{U}(M_b, s_b)$*, and that the entries* b_{ik} *satisfy the polynomial-tail condition* (S3.24) *with constant* K_b *for* $k = 1, \ldots, p$*. For any constant* $M > 0$*, let*

$$\lambda_b = 16(M+2) \left[(K+1) \left(\frac{\log p}{m} \right)^{1/2} + (K_\varepsilon + 1) \frac{(N \log p)^{1/2}}{(N-m)n^*} \right] + \frac{M_b}{m} + \frac{M_\varepsilon}{mn_*},$$

where $n^* = m / \sum_{i=1}^m n_i^{-1}$ *and* K *is a constant that only depends on* K_b *and* K_ε*. If* $p \leq cm^\gamma$ *for some* c*, then the regularized between-subject covariance estimate*

$\widehat{\Sigma}_b^+$ *satisfies*

$$\begin{aligned} & \mathrm{pr}\left\{\left\|\widehat{\Sigma}_b^+ - \Sigma_b^0\right\|_F \leq 10\lambda_b(ps_b)^{1/2}\right\} \\ \geq \;& 1 - O(p^{-M}) - K_\varepsilon p(\log N)^{2(1+\gamma+\delta)}N^{-(\gamma+\delta)} - 4Kp(\log m)^{2(1+\gamma+\delta)}m^{-(\gamma+\delta)}. \end{aligned}$$

S4 More Results from Numerical Studies

S4.1 Cross-validation Curve and ROC

In Fig. 1 we present the cross-validation curves and the receiver operating characteristic (ROC) of the sparsity recovery of these estimators in Model 1 in the manuscript with $p = 100$ and under three different levels of data imbalance. The discussions of this plot is in the manuscript.

S4.2 Understanding the Effects of the Bias in Sample Estimates

As seen in Fig 1 in the manuscript, the estimator $\overline{\Sigma}^+$ based on the biased sample estimate $\overline{\Sigma}$ surprisingly has relatively acceptable numerical performance. This subsection investigates this observation by comparing our proposed between-subject estimator $\widehat{\Sigma}_b^+$ with $\overline{\Sigma}^+$. We consider two models as follows:

Model 1. For any given $a > 0$, we set $(\Sigma_b^0)_{j,k} = (1 - |j-k|/10)_+$ and $(\Sigma_\varepsilon^0)_{j,k} = a(1 - |j-k|/10)_+$.

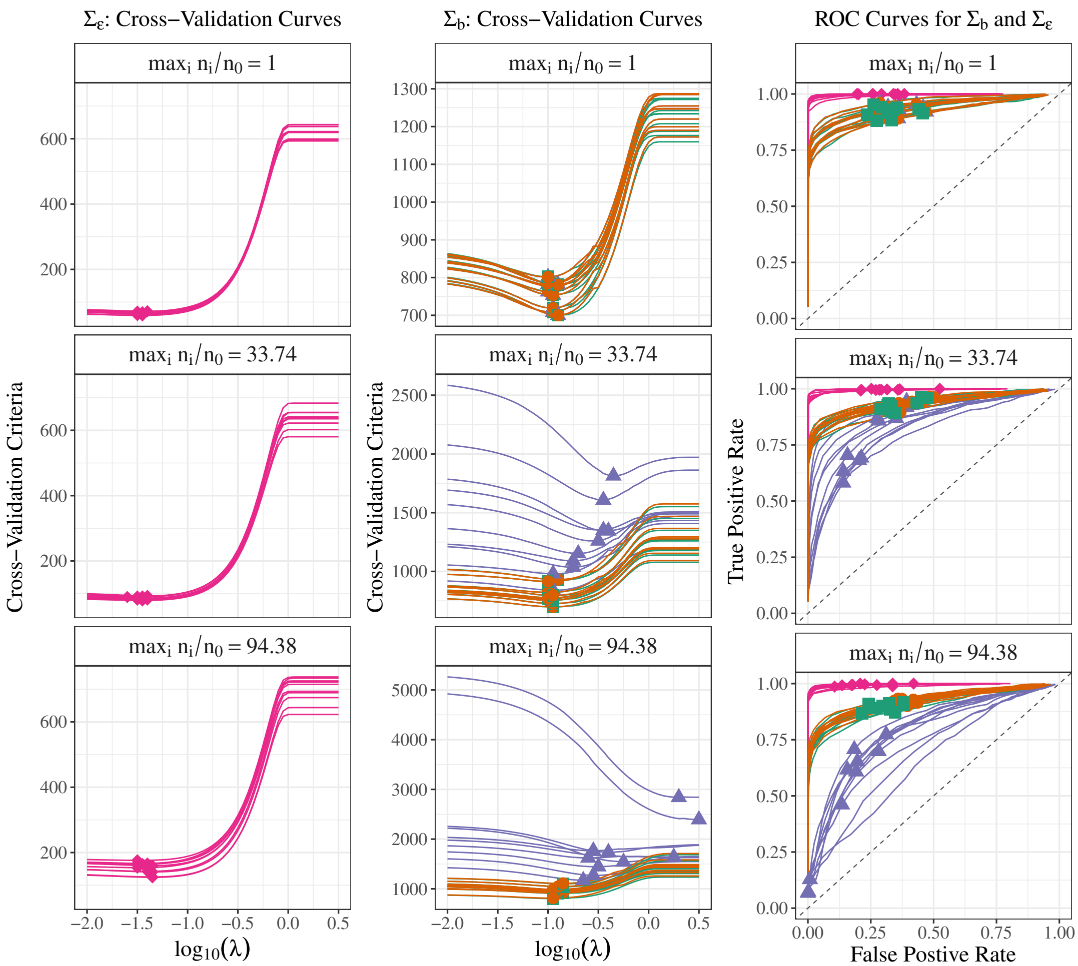

Figure 1: Cross-validation curves and receiver operating characteristic (ROC) curves between-subject and within-subject covariance sparsity recovery in Model 2 in the manuscript with $p = 100$ and different values of $\max_i n_i/n_0$. The top, middle, and bottom rows correspond to different levels of data imbalance (with $a = 10$, 7, and 4, respectively). For simplicity of presentation, we randomly select 10 out of the 100 replicates. The left and middle panels exhibit 5-fold cross-validation curves of $\widehat{\Sigma}_\varepsilon^+$ (pink) for within-subject covariance, $\widehat{\Sigma}_b^+$ (orange), $\widetilde{\Sigma}_b^+$ (violet), and $\overline{\Sigma}^+$ (green) for between-subject covariance. Diamonds ($\widehat{\Sigma}_\varepsilon^+$), circles ($\widehat{\Sigma}_b^+$), triangles ($\widetilde{\Sigma}_b^+$), and squares ($\overline{\Sigma}^+$) in these two panels mark the minimum points on these curves. The right panels present the ROC curves. The diamonds ($\widehat{\Sigma}_\varepsilon^+$), circles ($\widehat{\Sigma}_b^+$), triangles ($\widetilde{\Sigma}_b^+$), and squares ($\overline{\Sigma}^+$) represent the true positive rate and false positive rate with λ values selected by the 5-fold cross-validation.

Model 2. For any given $a > 0$, we set $(\Sigma_b^0)_{j,k} = (1 - |j-k|/10)_+$ and $(\Sigma_\varepsilon^0)_{j,k} = a(-1)^{|j-k|}(1 - |j-k|/10)_+$.

From (2.3) in the manuscript, the matrix of Σ_ε can be considered as the additive noise for the task of estimating Σ_b. We thus define the inverse signal-to-noise ratio as $|\Sigma_\varepsilon^0|_\infty/|\Sigma_b^0|_\infty$. By varying $|\Sigma_\varepsilon^0|_\infty/|\Sigma_b^0|_\infty = a \in \{1, 2, \ldots, 10\}$ in Model 1 and Model 2, we construct settings where the relative signal strength from Σ_ε and Σ_b is different. In comparison with Model 1, we alternate the signs of sub-diagonal elements in Σ_ε^0 in Model 2. In both models, we generate balanced data with $n_i = 5$ for $i = 1, \ldots, m = 100$ and $p = 50$. Estimation errors in Frobenius norm are summarized (over 100 replications) in Fig. 2.

In general, our proposed between-subject sample estimate $\widehat{\Sigma}_b$ significantly outperforms $\overline{\Sigma}$ in both examples. This demonstrates the effect of the bias correction as in (2.3) in the manuscript. Moreover, for both sample estimators, their regularized versions (dashed lines) achieve lower estimation errors, indicating the benefit of regularization.

Surprisingly, as $|\Sigma_\varepsilon^0|_\infty/|\Sigma_b^0|_\infty$ gets relatively small, $\overline{\Sigma}^+$ achieves an even smaller estimation error than $\widehat{\Sigma}_b^+$. This is an interesting cancellation of two biases with opposite signs: the estimation bias in the sample estimate $\overline{\Sigma}$ and the shrinkage bias in the ℓ_1-penalty. Specifically, for any index pair (j, k), (2.3) in the manuscript indicates that the bias of $\overline{\Sigma}_{j,k}$ in estimating $(\Sigma_b^0)_{j,k}$ is

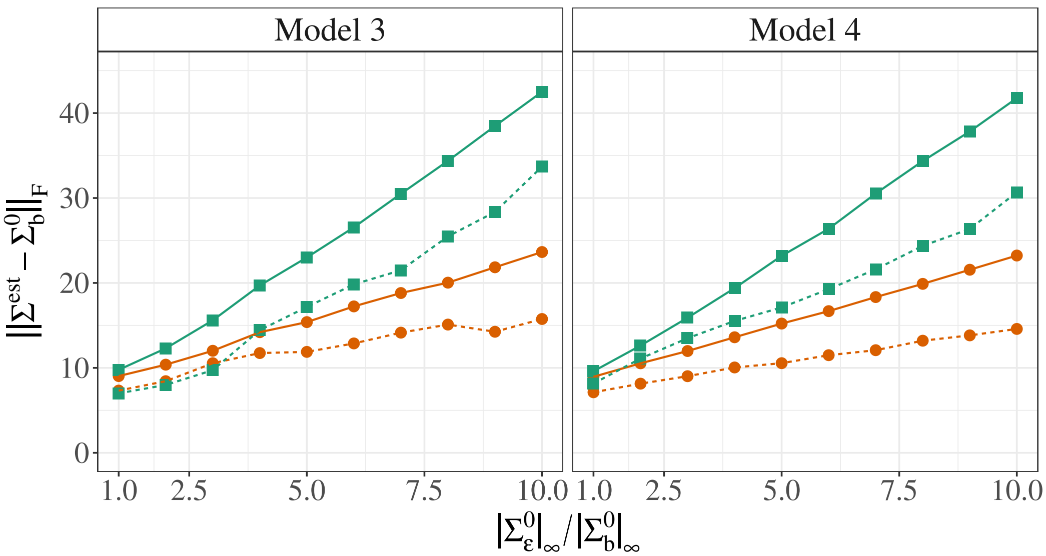

Figure 2: Estimation error (in Frobenius norm, averaged over 100 replicates) of the two between-subject sample covariance (solid) estimators ($\overline{\Sigma}$ and $\widehat{\Sigma}_b$) and their corresponding sparse and positive definite (dash) covariance estimators ($\overline{\Sigma}^+$ and $\widehat{\Sigma}_b^+$). The horizontal axis is the inverse signal-to-noise ratio, i.e., $|\Sigma_\varepsilon^0|_\infty/|\Sigma_b^0|_\infty$. The estimation errors of $\overline{\Sigma}$ and $\overline{\Sigma}^+$ are marked in green, and the estimation errors of $\widehat{\Sigma}_b$ and $\widehat{\Sigma}_b^+$ are marked in orange.

$\sum_i (mn_i)^{-1} (\Sigma_\varepsilon^0)_{j,k}$. In cases where $(\Sigma_\varepsilon^0)_{j,k}$ and $(\Sigma_b^0)_{j,k}$ have the same signs (as in Model 1), this sample estimation bias has the opposite effect from the shrinkage bias from the ℓ_1 penalty. Consequently, these two biases could cancel each other when they have similar magnitudes, which is achieved when $(\Sigma_\varepsilon^0)_{j,k}$ is on a similar scale as λ, and thus resulting in the surprisingly better performance of $\widehat{\Sigma}_b$ than $\widehat{\Sigma}_b^+$. Notably, when the estimation bias (as characterized by $|\Sigma_\varepsilon^0|_\infty / |\Sigma_b^0|_\infty$) is too large to be canceled by the shrinkage bias, or when both biases have the same signs (as in Model 2), the performance of $\widehat{\Sigma}_b^+$ is dominating that of $\overline{\Sigma}^+$.

S4.3 Simulation Studies in Heavy-tailed Settings

This additional simulation study examines the numerical performance of the proposed estimators when the Gaussian assumption is violated, e.g., in heavy-tailed data. Specifically, we consider simulation settings where random effects and random errors are both generated from a t_5 distribution, and all other specifications remain the same as in Model 1 with $p = 100$. The following figure suggests that our proposed methods still perform favorably in heavy-tailed settings.

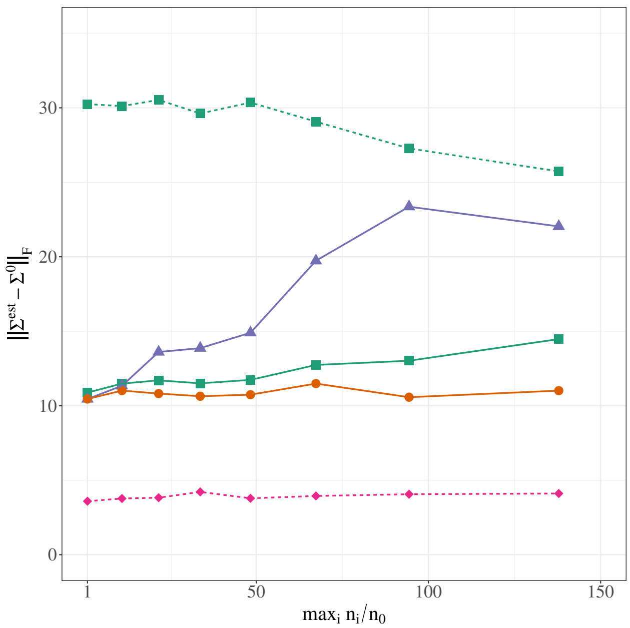

Figure 3: Estimation error (in Frobenius norm, medium over 100 replicates) for two between-subject (solid) and one within-subject (dash) covariance matrix estimator when random effects and random errors are generated from a t_5 distribution: $\widetilde{\Sigma}_b^+$ (violet triangle), $\widehat{\Sigma}_b^+$ (orange circle), and $\widehat{\Sigma}_\varepsilon^+$ (pink diamond). The estimation error of the aggregated estimator ($\overline{\Sigma}^+$, green square) is evaluated in estimating the within-subject (dash) and the between-subject (solid) covariance matrices. The x-axis is $\max_i n_i/n_0$, which characterizes the imbalance of the data.

S4.4 Unconstrained Estimators Versus Constrained Estimators

We generate 100 independent data sets for both balanced Model 1 and Model 2 with $n_i = 2$ and $m = 100$. We compare the performance of the unconstrained estimators, $\mathcal{S}_\lambda(\widehat{\Sigma}_\varepsilon)$ and $\mathcal{S}_\lambda(\widehat{\Sigma}_b)$, and the constrained estimators, $\widehat{\Sigma}_\varepsilon^+$ and $\widehat{\Sigma}_b^+$, in terms of estimation errors and the percentage of positive definite estimators, where $\mathcal{S}_\lambda()$ is the soft-thresholding operator defined in Section 2 in the main context. The simulation results are summarized in Table 1. Generally, constrained estimators exhibit slightly better performance in terms of estimation errors. In addition, we demonstrate that the positive definite constraint is crucial by observing that in most of the cases, the unconstrained estimators are not guaranteed to be positive definite, making them less qualified for interpretation or downstream statistical tasks.

Table 1: Comparison of the unconstrained and constrained estimators under the balanced setting. Each metric is averaged over 100 replicates, with the standard error shown in the parentheses. Comparisons are in terms of the estimation errors (F-error and L_2-error) and the percentage of positive definite estimators.

		Model 1		Model 2	
	p	100	200	100	200
		Within-Subject			
F-error	$\mathcal{S}_\lambda(\widehat{\Sigma}_\varepsilon)$	7.1804 (0.0562)	11.4040 (0.0490)	5.3956 (0.0202)	8.3116 (0.0159)
	$\widehat{\Sigma}_\varepsilon^+$	7.0548 (0.0552)	11.1804 (0.0490)	5.3956 (0.0202)	8.3116 (0.0159)
L_2-error	$\mathcal{S}_\lambda(\widehat{\Sigma}_\varepsilon)$	3.6179 (0.0451)	4.2217 (0.0282)	2.7131 (0.0115)	2.1257 (0.0083)
	$\widehat{\Sigma}_\varepsilon^+$	3.5553 (0.0438)	4.1564 (0.0286)	2.7131 (0.0115)	2.1257 (0.0083)
PD%	$\mathcal{S}_\lambda(\widehat{\Sigma}_\varepsilon)$	18%	3%	100%	100%
	$\widehat{\Sigma}_\varepsilon^+$	100%	100%	100%	100%
		Between-Subject			
F-error	$\mathcal{S}_\lambda(\widehat{\Sigma}_b)$	10.8195 (0.0611)	17.0538 (0.0416)	7.6064 (0.0212)	11.6116 (0.0187)
	$\widehat{\Sigma}_b^+$	10.1304 (0.0635)	16.1446 (0.0436)	7.5382 (0.0222)	11.6005 (0.0139)
L_2-error	$\mathcal{S}_\lambda(\widehat{\Sigma}_b)$	4.5419 (0.0447)	5.3739 (0.0258)	2.3508 (0.0104)	2.5681 (0.0051)
	$\widehat{\Sigma}_b^+$	4.2857 (0.0467)	5.0994 (0.0257)	2.3143 (0.0104)	2.5358 (0.0046)
PD%	$\mathcal{S}_\lambda(\widehat{\Sigma}_b)$	0%	0%	7%	12%
	$\widehat{\Sigma}_b^+$	100%	100%	100%	100%

$\mathcal{S}_\lambda(\widehat{\Sigma}_\varepsilon)$ and $\widehat{\Sigma}_\varepsilon^+$: unconstrained and constrained estimators for within-subject covariance; $\mathcal{S}_\lambda(\widehat{\Sigma}_b)$ and $\widehat{\Sigma}_b^+$: unconstrained and constrained estimators for between-subject covariance; PD%, percentage of positive definite estimators.

Bibliography

P.J. Bickel and E. Levina. Covariance regularization by thresholding. *Ann. Statist.*, 36:2577–2604, 2008.

S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3:1–122, 2010.

T. Cai and W. Liu. Adaptive thresholding for sparse covariance matrix estimation. *J. Am. Statist. Ass.*, 106:672–684, 2011.

Y. Cui, C. Leng, and D. Sun. Sparse estimation of high-dimensional correlation matrices. *Computational Statistics and Data Analysis*, 93:390–403, 2016.

A.J. Rothman, E. Levina, and J. Zhu. Generalized thresholding of large covariance matrices. *J. Am. Statist. Ass.*, 104:177–186, 2009.

R. Vershyin. *High-dimensional probability: an introduction with applications in data science*. Cambridge University Press, Cambridge, 2018.

L. Xue, S. Ma, and H. Zou. Positive-definite ℓ_1-penalized estimation of large covariance matrices. *J. Am. Statist. Ass.*, 107:1480–1491, 2012.