Biometrics 000, 000–000 DOI: 000
000 0000

Supporting Information for "Decomposition of Variation of Mixed Variables by a Latent Mixed Gaussian Copula Model" by

Yutong Liu[1], Toni Darville[2], Xiaojing Zheng[1,2,*], and Quefeng Li[1,]**

[1]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.

[2]Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.

**email:* xiaojinz@email.unc.edu

***email:* quefeng@email.unc.edu

This paper has been submitted for consideration for publication in *Biometrics*

Web Appendix A. Proofs

Proof of Proposition 1. Since $(\boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{U})$ and $(\widetilde{\boldsymbol{F}}_1, \widetilde{\boldsymbol{F}}_2, \widetilde{\boldsymbol{U}})$ are both mutually uncorrelated, we have

$$\mathrm{cov}(\boldsymbol{\Lambda}_1\boldsymbol{F}_1, \boldsymbol{\Lambda}_2\boldsymbol{F}_2) = 0, \quad \mathrm{cov}(\boldsymbol{\Lambda}_1\boldsymbol{F}_1, \boldsymbol{U}) = 0, \quad \mathrm{cov}(\boldsymbol{\Lambda}_2\boldsymbol{F}_2, \boldsymbol{U}) = 0,$$
$$\mathrm{cov}(\widetilde{\boldsymbol{\Lambda}}_1\widetilde{\boldsymbol{F}}_1, \widetilde{\boldsymbol{\Lambda}}_2\widetilde{\boldsymbol{F}}_2) = 0, \quad \mathrm{cov}(\widetilde{\boldsymbol{\Lambda}}_1\widetilde{\boldsymbol{F}}_1, \widetilde{\boldsymbol{U}}) = 0, \quad \mathrm{cov}(\widetilde{\boldsymbol{\Lambda}}_2\widetilde{\boldsymbol{F}}_2, \widetilde{\boldsymbol{U}}) = 0.$$

Suppose $\boldsymbol{\Lambda}_g\boldsymbol{F}_g + \boldsymbol{U} = \widetilde{\boldsymbol{\Lambda}}_g\widetilde{\boldsymbol{F}}_g + \widetilde{\boldsymbol{U}}$ for $g \in \{1, 2\}$. Let $\boldsymbol{W} = \widetilde{\boldsymbol{U}} - \boldsymbol{U}$. Then, $\widetilde{\boldsymbol{\Lambda}}_1\widetilde{\boldsymbol{F}}_1 = \boldsymbol{\Lambda}_1\boldsymbol{F}_1 - \boldsymbol{W}$ and $\widetilde{\boldsymbol{\Lambda}}_2\widetilde{\boldsymbol{F}}_2 = \boldsymbol{\Lambda}_2\boldsymbol{F}_2 - \boldsymbol{W}$. We have

$$\begin{aligned}0 = \mathrm{cov}(\widetilde{\boldsymbol{\Lambda}}_1\widetilde{\boldsymbol{F}}_1, \widetilde{\boldsymbol{\Lambda}}_2\widetilde{\boldsymbol{F}}_2) &= \mathrm{cov}(\boldsymbol{\Lambda}_1\boldsymbol{F}_1 - \boldsymbol{W}, \boldsymbol{\Lambda}_2\boldsymbol{F}_2 - \boldsymbol{W}) \\ &= \mathrm{cov}(\boldsymbol{\Lambda}_1\boldsymbol{F}_1, \boldsymbol{\Lambda}_2\boldsymbol{F}_2) + \mathrm{Var}(\boldsymbol{W}) - \mathrm{cov}(\boldsymbol{\Lambda}_1\boldsymbol{F}_1, \boldsymbol{W}) - \mathrm{cov}(\boldsymbol{\Lambda}_2\boldsymbol{F}_2, \boldsymbol{W}).\end{aligned}$$

It implies that

$$\mathrm{Var}(\boldsymbol{W}) = \mathrm{cov}(\boldsymbol{\Lambda}_1\boldsymbol{F}_1, \boldsymbol{W}) + \mathrm{cov}(\boldsymbol{\Lambda}_2\boldsymbol{F}_2, \boldsymbol{W}). \tag{S1}$$

Similarly, we have

$$\begin{aligned}0 = \mathrm{cov}(\widetilde{\boldsymbol{\Lambda}}_g\widetilde{\boldsymbol{F}}_g, \widetilde{\boldsymbol{U}}) &= \mathrm{cov}(\boldsymbol{\Lambda}_g\boldsymbol{F}_g - \boldsymbol{W}, \boldsymbol{U} + \boldsymbol{W}) \\ &= \mathrm{cov}(\boldsymbol{\Lambda}_g\boldsymbol{F}_g, \boldsymbol{U}) - \mathrm{Var}(\boldsymbol{W}) + \mathrm{cov}(\boldsymbol{\Lambda}_g\boldsymbol{F}_g, \boldsymbol{W}) - \mathrm{cov}(\boldsymbol{U}, \boldsymbol{W}).\end{aligned}$$

Then,

$$\mathrm{Var}(\boldsymbol{W}) = -\mathrm{cov}(\boldsymbol{U}, \boldsymbol{W}) + \mathrm{cov}(\boldsymbol{\Lambda}_g\boldsymbol{F}_g, \boldsymbol{W}). \tag{S2}$$

By (S2), we also have

$$\mathrm{cov}(\boldsymbol{\Lambda}_1\boldsymbol{F}_1, \boldsymbol{W}) = \mathrm{cov}(\boldsymbol{\Lambda}_2\boldsymbol{F}_2, \boldsymbol{W}). \tag{S3}$$

By (S1) and (S3), we have $\mathrm{Var}(\boldsymbol{W}) = 2\mathrm{cov}(\boldsymbol{\Lambda}_1\boldsymbol{F}_1, \boldsymbol{W}) = 2\mathrm{cov}(\boldsymbol{\Lambda}_2\boldsymbol{F}_2, \boldsymbol{W})$. By (S1) and (S2), we have $-\mathrm{cov}(\boldsymbol{U}, \boldsymbol{W}) = \mathrm{cov}(\boldsymbol{\Lambda}_1\boldsymbol{F}_1, \boldsymbol{W}) = \mathrm{cov}(\boldsymbol{\Lambda}_2\boldsymbol{F}_2, \boldsymbol{W})$. Then, we have

$$\mathrm{Var}(\boldsymbol{W}) = -2\mathrm{cov}(\boldsymbol{U}, \boldsymbol{W}) = 2\mathrm{cov}(\boldsymbol{\Lambda}_1\boldsymbol{F}_1, \boldsymbol{W}) = 2\mathrm{cov}(\boldsymbol{\Lambda}_2\boldsymbol{F}_2, \boldsymbol{W}).$$

Therefore,

$$\widetilde{\boldsymbol{\Lambda}}_g\widetilde{\boldsymbol{\Lambda}}_g' = \mathrm{Var}(\widetilde{\boldsymbol{\Lambda}}_g\widetilde{\boldsymbol{F}}_g) = \mathrm{Var}(\boldsymbol{\Lambda}_g\boldsymbol{F}_g - \boldsymbol{W}) = \mathrm{Var}(\boldsymbol{\Lambda}_g\boldsymbol{F}_g) + \mathrm{Var}(\boldsymbol{W}) - 2\mathrm{cov}(\boldsymbol{\Lambda}_g\boldsymbol{F}_g, \boldsymbol{W})$$
$$= \mathrm{Var}(\boldsymbol{\Lambda}_g\boldsymbol{F}_g) = \boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_g'.$$

Thus, $\mathrm{Var}(\widetilde{\boldsymbol{U}}) = \mathrm{Var}(\boldsymbol{U}+\boldsymbol{W}) = \mathrm{Var}(\boldsymbol{U}) + \mathrm{Var}(\boldsymbol{W}) + 2\mathrm{cov}(\boldsymbol{U},\boldsymbol{W}) = \mathrm{Var}(\boldsymbol{U})$.

Proof of Theorem 2.

(a) By definition,

$$\tau_{jk} = \mathbb{E}(\widehat{\tau}_{jk}) = \mathbb{E}\Big[\frac{2}{n(n-1)}\sum_{1\leqslant i<i'\leqslant n}\mathrm{sign}\left\{(X_{ij}-X_{i'j})(X_{ik}-X_{i'k})\right\}\Big].$$

For the simplicity of notation, we omit i and i' from the subscripts and write X_{ij} and $X_{i'j}$ as X_j and X_j', where we treat them as two independent realizations from the same distribution. Since X_j and X_j' are three-level ordinal variables,

$$\begin{aligned}\mathrm{sign}(X_j - X_j') = {} & I(X_j=2,X_j'=0) + I(X_j=2,X_j'=1) + I(X_j=1,X_j'=0)\\
& - I(X_j=1,X_j'=2) - I(X_j=0,X_j'=1) - I(X_j=0,X_j'=2)\\
= {} & I(X_j=2) - I(X_j=2,X_j'=2)\\
& + I(X_j=1,X_j'=0) - I(X_j=0,X_j'=1) - I(X_j'=2) + I(X_j=2,X_j'=2)\\
= {} & I(X_j=2) - I(X_j'=2) + I(X_j=1,X_j'=0) - I(X_j=0,X_j'=1).\end{aligned}$$

Define $\boldsymbol{Z} = \boldsymbol{f}(\boldsymbol{Y})$, where $\boldsymbol{Z}\sim N(\boldsymbol{0},\boldsymbol{\Sigma})$. Since $\mathrm{sign}(x) = 2I(x>0)-1$, we have

$$\begin{aligned}\tau_{jk} &= \mathbb{E}\Big\{\mathrm{sign}(X_j-X_j')\,\mathrm{sign}(X_k-X_k')\Big\} = \mathbb{E}\Big[\mathrm{sign}(X_j-X_j')\{2I(X_k>X_k')-1\}\Big]\\
&= \mathbb{E}\Big\{2\,\mathrm{sign}(X_j-X_j')I(X_k>X_k') - \mathrm{sign}(X_j-X_j')\Big\}\\
&= \mathbb{E}\Big\{2I(X_j=2)I(X_k>X_k') - 2I(X_j'=2)I(X_k>X_k')\\
&\quad + 2I(X_j=1,X_j'=0)I(X_k>X_k') - 2I(X_j=0,X_j'=1)I(X_k>X_k')\end{aligned}$$

$$
\begin{aligned}
&- I(X_j = 2) + I(X_j' = 2) - I(X_j = 1, X_j' = 0) + I(X_j = 0, X_j' = 1)\Big\}\\
&= \mathbb{E}\Big\{2I(Z_j > \Delta_{j2}, Z_k - Z_k' > 0) - 2I(Z_j' > \Delta_{j2}, Z_k - Z_k' > 0)\\
&\quad + 2I(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k - Z_k' > 0)\\
&\quad - 2I(Z_j < \Delta_{j1}, \Delta_{j1} < Z_j' < \Delta_{j2}, Z_k - Z_k' > 0)\\
&\quad - I(Z_j > \Delta_{j2}) + I(Z_j' > \Delta_{j2})\\
&\quad - I(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}) + I(Z_j < \Delta_{j1}, \Delta_{j1} < Z_j' < \Delta_{j2})\Big\}\\
&= 2\mathbb{P}(Z_j > \Delta_{j2}, Z_k - Z_k' > 0) - 2\mathbb{P}(Z_j' > \Delta_{j2}, Z_k - Z_k' > 0)\\
&\quad + 2\mathbb{P}(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k - Z_k' > 0)\\
&\quad - 2\mathbb{P}(Z_j < \Delta_{j1}, \Delta_{j1} < Z_j' < \Delta_{j2}, Z_k - Z_k' > 0)\\
&\quad - \mathbb{P}(Z_j > \Delta_{j2}) + \mathbb{P}(Z_j' > \Delta_{j2})\\
&\quad - \mathbb{P}(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}) + \mathbb{P}(Z_j < \Delta_{j1}, \Delta_{j1} < Z_j' < \Delta_{j2})\\
&= 2\mathbb{P}(Z_k' - Z_k < 0) - 2\mathbb{P}(Z_j < \Delta_{j2}, Z_k' - Z_k < 0)\\
&\quad - 2\mathbb{P}(Z_k' - Z_k < 0) + 2\mathbb{P}(Z_j' < \Delta_{j2}, Z_k' - Z_k < 0)\\
&\quad + 2\mathbb{P}(Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k' - Z_k < 0) - 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k' - Z_k < 0)\\
&\quad - 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j2}, Z_k' - Z_k < 0) + 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k' - Z_k < 0)\\
&= -2\mathbb{P}(Z_j < \Delta_{j2}, Z_k' - Z_k < 0) + 2\mathbb{P}(Z_j' < \Delta_{j2}, Z_k' - Z_k < 0)\\
&\quad + 2\mathbb{P}(Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k' - Z_k < 0) - 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j2}, Z_k' - Z_k < 0)\\
&= 2\Phi_2(\Delta_{j2}, 0; \frac{R_{jk}}{\sqrt{2}}) - 2\Phi_2(\Delta_{j2}, 0; -\frac{R_{jk}}{\sqrt{2}}) - 2\Phi_3(\Delta_{j1}, \Delta_{j2}, 0; \boldsymbol{R}_{3d}) + 2\Phi_3(\Delta_{j2}, \Delta_{j1}, 0; \boldsymbol{R}_{3d}).
\end{aligned}
$$

(b) Since X_k and X_k' are binary variables,

$$\text{sign}(X_k - X_k') = I(Y_k > C_k) - I(Y_k' > C_k).$$

We have

$$
\begin{aligned}
\tau_{jk} &= \mathbb{E}\Big[\operatorname{sign}(X_j - X_j')\big\{I(Y_k > C_k) - I(Y_k' > C_k)\big\}\Big] \\
&= \mathbb{E}\Big\{I(X_j = 2, Y_k > C_k) - I(X_j' = 2, Y_k > C_k) + I(X_j = 1, X_j' = 0, Y_k > C_k) \\
&\quad - I(X_j = 0, X_j' = 1, Y_k > C_k) - I(X_j = 2, Y_k' > C_k) + I(X_j' = 2, Y_k' > C_k) \\
&\quad - I(X_j = 1, X_j' = 0, Y_k' > C_k) + I(X_j = 0, X_j' = 1, Y_k' > C_k)\Big\} \\
&= \mathbb{E}\Big\{I(Y_j > C_{j2}, Y_k > C_k) - I(Y_j' > C_{j2}, Y_k > C_k) + I(C_{j1} < Y_j < C_{j2}, Y_j' < C_{j1}, Y_k > C_k) \\
&\quad - I(Y_j < C_{j1}, C_{j1} < Y_j' < C_{j2}, Y_k > C_k) - I(Y_j > C_{j2}, Y_k' > C_k) + I(Y_j' > C_{j2}, Y_k' > C_k) \\
&\quad - I(C_{j1} < Y_j < C_{j2}, Y_j' < C_{j1}, Y_k' > C_k) + I(Y_j < C_{j1}, C_{j1} < Y_j' < C_{j2}, Y_k' > C_k)\Big\} \\
&= \mathbb{P}(Z_j > \Delta_{j2}, Z_k > \Delta_k) - \mathbb{P}(Z_j' > \Delta_{j2}, Z_k > \Delta_k) + \mathbb{P}(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k > \Delta_k) \\
&\quad - \mathbb{P}(Z_j < \Delta_{j1}, \Delta_{j1} < Z_j' < \Delta_{j2}, Z_k > \Delta_k) - \mathbb{P}(Z_j > \Delta_{j2}, Z_k' > \Delta_k) + \mathbb{P}(Z_j' > \Delta_{j2}, Z_k' > \Delta_k) \\
&\quad - \mathbb{P}(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k' > \Delta_k) + \mathbb{P}(Z_j < \Delta_{j1}, \Delta_{j1} < Z_j' < \Delta_{j2}, Z_k' > \Delta_k) \\
&= \mathbb{P}(Z_k > \Delta_k) - \mathbb{P}(Z_j < \Delta_{j2}, Z_k > \Delta_k) - \mathbb{P}(Z_k > \Delta_k) + \mathbb{P}(Z_j' < \Delta_{j2}, Z_k > \Delta_k) \\
&\quad + \mathbb{P}(Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k > \Delta_k) - \mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k > \Delta_k) \\
&\quad - \mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j2}, Z_k > \Delta_k) + \mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k > \Delta_k) \\
&\quad - \mathbb{P}(Z_k' > \Delta_k) + \mathbb{P}(Z_j < \Delta_{j2}, Z_k' > \Delta_k) + \mathbb{P}(Z_k' > \Delta_k) - \mathbb{P}(Z_j' < \Delta_{j2}, Z_k' > \Delta_k) \\
&\quad - \mathbb{P}(Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k' > \Delta_k) + \mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k' > \Delta_k) \\
&\quad + \mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j2}, Z_k' > \Delta_k) - \mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k' > \Delta_k) \\
&= -2\mathbb{P}(Z_j < \Delta_{j2}, Z_k > \Delta_k) + 2\mathbb{P}(Z_j' < \Delta_{j2}, Z_k > \Delta_k) \\
&\quad + 2\mathbb{P}(Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k > \Delta_k) - 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k > \Delta_k) \\
&\quad - 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j2}, Z_k > \Delta_k) + 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k > \Delta_k)
\end{aligned}
$$

$$= -2\mathbb{P}(Z_j < \Delta_{j2}) + 2\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_k) + 2\mathbb{P}(Z_j' < \Delta_{j2}) - 2\mathbb{P}(Z_j' < \Delta_{j2}, Z_k < \Delta_k)$$
$$+ 2\mathbb{P}(Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}) - 2\mathbb{P}(Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k < \Delta_k)$$
$$- 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}) + 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k < \Delta_k)$$
$$- 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j2}) + 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j2}, Z_k < \Delta_k)$$
$$+ 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}) - 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k < \Delta_k)$$
$$= 2\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_k) - 2\mathbb{P}(Z_j' < \Delta_{j2}, Z_k < \Delta_k)$$
$$- 2\mathbb{P}(Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k < \Delta_k) + 2\mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j2}, Z_k < \Delta_k)$$
$$= 2\Big\{\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_k) - \mathbb{P}(Z_j' < \Delta_{j2})\mathbb{P}(Z_k < \Delta_k)$$
$$- \mathbb{P}(Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k < \Delta_k) + \mathbb{P}(Z_j < \Delta_{j1}, Z_j' < \Delta_{j2}, Z_k < \Delta_k)\Big\}$$
$$= 2\Phi_2(\Delta_{j2}, \Delta_k; R_{jk}) - 2\Phi_1(\Delta_{j2})\Phi_1(\Delta_k) - 2\Phi_1(\Delta_{j1})\Phi_2(\Delta_{j2}, \Delta_k; R_{jk}) + 2\Phi_1(\Delta_{j2})\Phi_2(\Delta_{j1}, \Delta_k; R_{jk}.)$$

(c) Since X_j and X_j' are three-level ordinal variables,

$$\text{sign}(X_j - X_j') = I(X_j = 2) - I(X_j' = 2) + I(X_j = 1, X_j' = 0) - I(X_j = 0, X_j' = 1).$$

Since X_k and X_k' are truncated variables,

$$\text{sign}(X_k - X_k') = -I(X_k = 0, X_k' > 0) + I(X_k > 0, X_k' = 0) + I(X_k > 0, X_k' > 0)\,\text{sign}(X_k - X_k')$$
$$= -I(X_k = 0) + I(X_k' = 0) + I(X_k > 0, X_k' > 0)\,\text{sign}(X_k - X_k').$$

Since $\text{sign}(x) = 2I(x > 0) - 1$, we have

$$\tau_{jk} = \mathbb{E}\Big\{\text{sign}(X_j - X_j')\,\text{sign}(X_k - X_k')\Big\}$$
$$= \mathbb{E}\Big\{2I(X_j = 2)I(X_k' = 0) - 2I(X_j = 2)I(X_k = 0) + 2I(X_j = 1, X_j' = 0)I(X_k' = 0)$$
$$- 2I(X_j = 1, X_j' = 0)I(X_k = 0) + 2I(X_k > X_k')I(X_j = 2)I(X_k > 0, X_k' > 0)$$
$$- 2I(X_k > X_k')I(X_j' = 2)I(X_k > 0, X_k' > 0)$$

$$
\begin{aligned}
&+ 2I(X_k > X_k')I(X_j = 1, X_j' = 0)I(X_k > 0, X_k' > 0) \\
&- 2I(X_k > X_k')I(X_j = 0, X_j' = 1)I(X_k > 0, X_k' > 0)\Big\} \\
&= \mathbb{E}\Big\{2I(Z_j > \Delta_{j2})I(Z_k' < \Delta_k) - 2I(Z_j > \Delta_{j2})I(Z_k < \Delta_k) \\
&+ 2I(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1})I(Z_k' < \Delta_k) - 2I(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1})I(Z_k < \Delta_k) \\
&+ 2I(Z_k' - Z_k < 0)I(Z_j > \Delta_{j2})I(Z_k > \Delta_k, Z_k' > \Delta_k) \\
&- 2I(Z_k' - Z_k < 0)I(Z_j' > \Delta_{j2})I(Z_k > \Delta_k, Z_k' > \Delta_k) \\
&+ 2I(Z_k' - Z_k < 0)I(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1})I(Z_k > \Delta_k, Z_k' > \Delta_k) \\
&- 2I(Z_k' - Z_k < 0)I(\Delta_{j1} < Z_j' < \Delta_{j2}, Z_j < \Delta_{j1})I(Z_k > \Delta_k, Z_k' > \Delta_k)\Big\} \\
&= 2\Big\{\mathbb{P}(Z_j > \Delta_{j2}, Z_k' < \Delta_k) - \mathbb{P}(Z_j > \Delta_{j2}, Z_k < \Delta_k) \\
&+ \mathbb{P}(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k' < \Delta_k) - \mathbb{P}(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k < \Delta_k) \\
&+ \mathbb{P}(Z_k' - Z_k < 0, Z_j > \Delta_{j2}, Z_k > \Delta_k, Z_k' > \Delta_k) \\
&- \mathbb{P}(Z_k' - Z_k < 0, Z_j' > \Delta_{j2}, Z_k > \Delta_k, Z_k' > \Delta_k) \\
&+ \mathbb{P}(Z_k' - Z_k < 0, \Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k > \Delta_k, Z_k' > \Delta_k) \\
&- \mathbb{P}(Z_k' - Z_k < 0, \Delta_{j1} < Z_j' < \Delta_{j2}, Z_j < \Delta_{j1}, Z_k > \Delta_k, Z_k' > \Delta_k)\Big\} \\
&= 2\Big\{\mathbb{P}(Z_k' < \Delta_k)\mathbb{P}(Z_j > \Delta_{j2}) - \mathbb{P}(Z_k < \Delta_k) - \mathbb{P}(Z_j < \Delta_{j2}) \\
&+ 2\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_k) - \mathbb{P}(Z_j < \Delta_{j2})\mathbb{P}(Z_j' < \Delta_{j1}) \\
&- \mathbb{P}(Z_j' < \Delta_{j2})\mathbb{P}(Z_k < \Delta_k) + 2\mathbb{P}(Z_k' - Z_k < 0, Z_j < \Delta_{j2}) \\
&- 2\mathbb{P}(Z_j' < \Delta_{j1})\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_k) - \mathbb{P}(Z_k' < \Delta_k)\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_k) \\
&+ 2\mathbb{P}(Z_k' - Z_k < 0, Z_j' < \Delta_{j2}, Z_k < \Delta_k) - 2\mathbb{P}(Z_k' - Z_k < 0, Z_j < \Delta_{j2}, Z_k < \Delta_k) \\
&+ 2\mathbb{P}(Z_k' - Z_k < 0, Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}) + 2\mathbb{P}(Z_k' - Z_k < 0, Z_j < \Delta_{j2}, Z_k < \Delta_k, Z_k' < \Delta_k)
\end{aligned}
$$

$$
\begin{aligned}
&+ 2\mathbb{P}(Z_k' - Z_k < 0, Z_j' < \Delta_{j2}, Z_j < \Delta_{j1}, Z_k < \Delta_k) \\
&+ 2\mathbb{P}(Z_k' - Z_k < 0, Z_j' < \Delta_{j2}, Z_j < \Delta_{j1}, Z_k' < \Delta_k) \\
&+ 2\mathbb{P}(Z_k' - Z_k < 0, Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k' < \Delta_k) \\
&- 2\mathbb{P}(Z_k' - Z_k < 0, Z_j < \Delta_{j1}, Z_j' < \Delta_{j1}, Z_k < \Delta_k, Z_k' < \Delta_k) \\
&+ 2\mathbb{P}(Z_k' - Z_k < 0, Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k < \Delta_k, Z_k' < \Delta_k) \\
&- \mathbb{P}(Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k < \Delta_k, Z_k' < \Delta_k)\Big\} \\
= 2\Big\{ &- 2\Phi_1(\Delta_k)\Phi_1(\Delta_{j2}) - \Phi_1(\Delta_{j2}) + 2\Phi_2(\Delta_{j2}, \Delta_k; R_{jk}) - \Phi_1(\Delta_{j1})\Phi_1(\Delta_{j2}) \\
&+ \Phi_2(0, \Delta_{j2}; -\frac{R_{jk}}{\sqrt{2}}) - 2\Phi_1(\Delta_{j1})\Phi_2(\Delta_{j2}, \Delta_k; R_{jk}) - \Phi_1(\Delta_k)\Phi_2(\Delta_{j2}, \Delta_k; R_{jk}) \\
&+ 2\Phi_3(0, \Delta_{j2}, \Delta_k; \boldsymbol{R}_{3e}) - 2\Phi_3(0, \Delta_{j2}, \Delta_k; \boldsymbol{R}_{3f}) + 2\Phi_3(\Delta_{j2}, \Delta_{j1}, 0; \boldsymbol{R}_{3d}) \\
&+ 2\Phi_4(0, \Delta_{j2}, \Delta_k, \Delta_k; \boldsymbol{R}_{4c}) + 2\Phi_4(0, \Delta_{j2}, \Delta_{j1}, \Delta_k; \boldsymbol{R}_{4d}) + 2\Phi_4(0, \Delta_{j2}, \Delta_{j1}, \Delta_k; \boldsymbol{R}_{4e}) \\
&+ 2\Phi_4(0, \Delta_{j1}, \Delta_{j1}, \Delta_k; \boldsymbol{R}_{4e}) - 2\Phi_2(\Delta_{j1}, \Delta_k; R_{jk})\Phi_2(\Delta_{j2}, \Delta_k; R_{jk}) \\
&- 2\Phi_5(0, \Delta_{j1}, \Delta_{j1}, \Delta_k, \Delta_k; \boldsymbol{R}_5) + 2\Phi_5(0, \Delta_{j2}, \Delta_{j1}, \Delta_k, \Delta_k; \boldsymbol{R}_5)\Big\}.
\end{aligned}
$$

(d) We have

$$
\begin{aligned}
\tau_{jk} &= \mathbb{E}\Big\{ \operatorname{sign}(X_j - X_j') \operatorname{sign}(X_k - X_k') \Big\} \\
&= \mathbb{E}\Big[\{I(X_j = 2) - I(X_j' = 2) + I(X_j = 1, X_j' = 0) - I(X_j = 0, X_j' = 1)\} \\
&\qquad \{I(X_k = 2) - I(X_k' = 2) + I(X_k = 1, X_k' = 0) - I(X_k = 0, X_k' = 1)\} \Big] \\
&= \mathbb{E}\Big\{ I(X_j = 2, X_k = 2) - I(X_j = 2, X_k' = 2) \\
&\quad - I(X_j' = 2, X_k = 2) + I(X_j' = 2, X_k' = 2) \\
&\quad + I(X_j = 2, X_k = 1, X_k' = 0) - I(X_j = 2, X_k = 0, X_k' = 1) \\
&\quad + I(X_j' = 2, X_k = 0, X_k' = 1) - I(X_j' = 2, X_k = 1, X_k' = 0) \\
&\quad + I(X_j = 1, X_j' = 0, X_k = 2) - I(X_j = 1, X_j' = 0, X_k' = 2)
\end{aligned}
$$

$$
\begin{aligned}
&+ I(X_j = 0, X_j' = 1, X_k' = 2) - I(X_j = 0, X_j' = 1, X_k = 2)\\
&+ I(X_j = 1, X_j' = 0, X_k = 1, X_k' = 0) - I(X_j = 1, X_j' = 0, X_k = 0, X_k' = 1)\\
&+ I(X_j = 0, X_j' = 1, X_k = 0, X_k' = 1) - I(X_j = 0, X_j' = 1, X_k = 1, X_k' = 0)\Big\}\\
= \mathbb{E}\Big\{&2I(X_j = 2, X_k = 2) - 2I(X_j' = 2, X_k = 2)\\
&+ 2I(X_j = 2, X_k = 1, X_k' = 0) - 2I(X_j = 2, X_k = 0, X_k' = 1)\\
&+ 2I(X_j = 1, X_j' = 0, X_k = 2) - 2I(X_j = 1, X_j' = 0, X_k' = 2)\\
&+ 2I(X_j = 1, X_j' = 0, X_k = 1, X_k' = 0) - 2I(X_j = 1, X_j' = 0, X_k = 0, X_k' = 1)\Big\}\\
= {}&2\mathbb{P}(X_j = 2, X_k = 2) - 2\mathbb{P}(X_j' = 2, X_k = 2)\\
&+ 2\mathbb{P}(X_j = 2, X_k = 1, X_k' = 0) - 2\mathbb{P}(X_j = 2, X_k = 0, X_k' = 1)\\
&+ 2\mathbb{P}(X_j = 1, X_j' = 0, X_k = 2) - 2\mathbb{P}(X_j = 0, X_j' = 1, X_k = 2)\\
&+ 2\mathbb{P}(X_j = 1, X_j' = 0, X_k = 1, X_k' = 0) - 2\mathbb{P}(X_j = 1, X_j' = 0, X_k = 0, X_k' = 1)\\
= {}&2\mathbb{P}(Z_j > \Delta_{j2}, Z_k > \Delta_{k2}) - 2\mathbb{P}(Z_j' > \Delta_{j2}, Z_k > \Delta_{k2})\\
&+ 2\mathbb{P}(Z_j > \Delta_{j2}, \Delta_{k1} < Z_k < \Delta_{k2}, Z_k' < \Delta_{k1}) - 2\mathbb{P}(Z_j > \Delta_{j2}, Z_k < \Delta_{k1}, \Delta_{k1} < Z_k' < \Delta_{k2})\\
&+ 2\mathbb{P}(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k > \Delta_{k2}) - 2\mathbb{P}(Z_j < \Delta_{j1}, \Delta_{j1} < Z_j' < \Delta_{j2}, Z_k > \Delta_{k2})\\
&+ 2\mathbb{P}(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, \Delta_{k1} < Z_k < \Delta_{k2}, Z_k' < \Delta_{k1})\\
&- 2\mathbb{P}(\Delta_{j1} < Z_j < \Delta_{j2}, Z_j' < \Delta_{j1}, Z_k < \Delta_{k1}, \Delta_{k1} < Z_k' < \Delta_{k2})\\
= {}&2\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_{k2}) - 2\mathbb{P}(Z_j' < \Delta_{j2}, Z_k < \Delta_{k2})\\
&+ 2\mathbb{P}(Z_k < \Delta_{k2}, Z_k' < \Delta_{k1}) - 2\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_{k2}, Z_k' < \Delta_{k1})\\
&- 2\mathbb{P}(Z_k < \Delta_{k1}, Z_k' < \Delta_{k1}) + 2\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_{k1}, Z_k' < \Delta_{k1})\\
&- 2\mathbb{P}(Z_k < \Delta_{k1}, Z_k' < \Delta_{k2}) + 2\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_{k1}, Z_k' < \Delta_{k2})\\
&+ 2\mathbb{P}(Z_k < \Delta_{k1}, Z_k' < \Delta_{k1}) - 2\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_{k1}, Z_k' < \Delta_{k1})
\end{aligned}
$$

$$
\begin{aligned}
&+ 2\mathbb{P}(Z_j < \Delta_{j2}, Z'_j < \Delta_{j1}) - 2\mathbb{P}(Z_j < \Delta_{j2}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k2}) \\
&- 2\mathbb{P}(Z_j < \Delta_{j1}, Z'_j < \Delta_{j1}) + 2\mathbb{P}(Z_j < \Delta_{j1}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k2}) \\
&- 2\mathbb{P}(Z_j < \Delta_{j1}, Z'_j < \Delta_{j2}) + 2\mathbb{P}(Z_j < \Delta_{j1}, Z'_j < \Delta_{j2}, Z_k < \Delta_{k2}) \\
&+ 2\mathbb{P}(Z_j < \Delta_{j1}, Z'_j < \Delta_{j1}) - 2\mathbb{P}(Z_j < \Delta_{j1}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k2}) \\
&+ 2\mathbb{P}(Z_j < \Delta_{j2}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k2}, Z'_k < \Delta_{k1}) \\
&- 2\mathbb{P}(Z_j < \Delta_{j2}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k1}, Z'_k < \Delta_{k1}) \\
&- 2\mathbb{P}(Z_j < \Delta_{j1}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k2}, Z'_k < \Delta_{k1}) \\
&+ 2\mathbb{P}(Z_j < \Delta_{j1}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k1}, Z'_k < \Delta_{k1}) \\
&- 2\mathbb{P}(Z_j < \Delta_{j2}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k1}, Z'_k < \Delta_{k2}) \\
&+ 2\mathbb{P}(Z_j < \Delta_{j2}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k1}, Z'_k < \Delta_{k1}) \\
&+ 2\mathbb{P}(Z_j < \Delta_{j1}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k1}, Z'_k < \Delta_{k2}) \\
&- 2\mathbb{P}(Z_j < \Delta_{j1}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k1}, Z'_k < \Delta_{k1}) \\
= &\ 2\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_{k2}) - 2\mathbb{P}(Z'_j < \Delta_{j2}, Z_k < \Delta_{k2}) - 2\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_{k2}, Z'_k < \Delta_{k1}) \\
&+ 2\mathbb{P}(Z_j < \Delta_{j2}, Z_k < \Delta_{k1}, Z'_k < \Delta_{k2}) - 2\mathbb{P}(Z_j < \Delta_{j2}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k2}) \\
&+ 2\mathbb{P}(Z_j < \Delta_{j1}, Z'_j < \Delta_{j2}, Z_k < \Delta_{k2}) + 2\mathbb{P}(Z_j < \Delta_{j2}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k2}, Z'_k < \Delta_{k1}) \\
&- 2\mathbb{P}(Z_j < \Delta_{j2}, Z'_j < \Delta_{j1}, Z_k < \Delta_{k1}, Z'_k < \Delta_{k2}) \\
= &\ 2\Phi_2(\Delta_{j2}, \Delta_{k2}; R_{jk}) - 2\Phi_1(\Delta_{j2})\Phi_1(\Delta_{k2}) \\
&- 2\Phi_2(\Delta_{j2}, \Delta_{k2}; R_{jk})\Phi_1(\Delta_{k1}) + 2\Phi_2(\Delta_{j2}, \Delta_{k1}; R_{jk})\Phi_1(\Delta_{k2}) \\
&- 2\Phi_2(\Delta_{j2}, \Delta_{k2}; R_{jk})\Phi_1(\Delta_{j1}) + 2\Phi_2(\Delta_{j1}, \Delta_{k2}; R_{jk})\Phi_1(\Delta_{j2}) \\
&+ 2\Phi_2(\Delta_{j2}, \Delta_{k2}; R_{jk})\Phi_2(\Delta_{j1}, \Delta_{k1}; R_{jk}) - 2\Phi_2(\Delta_{j2}, \Delta_{k1}; R_{jk})\Phi_2(\Delta_{j1}, \Delta_{k2}; R_{jk})
\end{aligned}
$$

$$
\begin{aligned}
&= 2\Phi_2(\Delta_{j2}, \Delta_{k2}; R_{jk}) - 2\Phi_1(\Delta_{j2})\Phi_1(\Delta_{k2}) \\
&\quad - 4\Phi_2(\Delta_{k2}, \Delta_{j2}; R_{jk})\Phi_1(\Delta_{j1}) + 4\Phi_2(\Delta_{j1}, \Delta_{k2}; R_{jk})\Phi_1(\Delta_{j2}) \\
&\quad + 2\Phi_2(\Delta_{j2}, \Delta_{k2}; R_{jk})\Phi_2(\Delta_{j1}, \Delta_{k1}; R_{jk}) - 2\Phi_2(\Delta_{j2}, \Delta_{k1}; R_{jk})\Phi_2(\Delta_{j1}, \Delta_{k2}; R_{jk})
\end{aligned}
$$

Proof of Proposition 2. To prove this proposition, we note that Fan et al. (2017) showed that the bivariate normal distribution function $\Phi_2(\cdot, \cdot; t)$ is strictly increasing with t. Therefore, we have $\partial\Phi_2(\Delta_j, \Delta_k; t)/\partial t > 0$ for fixed constants Δ_j and Δ_k. Let $\Phi_d(a_1, \dots, a_d; \boldsymbol{\Sigma}_d(r))$ be the cumulative distribution function of the d-dimensional multivariate normal distribution whose covariance matrix equals to

$$
\boldsymbol{\Sigma}_d(r) = \begin{pmatrix}
1 & \sigma_{12}(r) & \sigma_{13}(r) & \dots & \sigma_{1d}(r) \\
\sigma_{21}(r) & 1 & \sigma_{23}(r) & \dots & \sigma_{2d}(r) \\
 & & 1 & & \\
\vdots & & & \ddots & \vdots \\
\sigma_{d1}(r) & & \dots & & 1
\end{pmatrix}.
$$

Yoon et al. (2020) proved that for any $a_1, \dots, a_d$, there exists $s_{ij}(r) > 0$ for all $r \in (-1, 1)$ such that

$$
\frac{\partial\Phi_d(a_1, \dots, a_d; \boldsymbol{\Sigma}_d(r))}{\partial r} = \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} s_{ij}(r)\frac{\partial\sigma_{ij}(r)}{\partial r}.
$$

(a) Let

$$
\boldsymbol{R}_{3g} = \begin{pmatrix}
1 & 0 & \frac{R_{jk}}{\sqrt{2}} \\
0 & 1 & -\frac{R_{jk}}{\sqrt{2}} \\
\frac{R_{jk}}{\sqrt{2}} & -\frac{R_{jk}}{\sqrt{2}} & 1
\end{pmatrix} \text{ and } \boldsymbol{R}_{3h} = \begin{pmatrix}
1 & 0 & -\frac{R_{jk}}{\sqrt{2}} \\
0 & 1 & -\frac{R_{jk}}{\sqrt{2}} \\
-\frac{R_{jk}}{\sqrt{2}} & -\frac{R_{jk}}{\sqrt{2}} & 1
\end{pmatrix}.
$$

We have $\Phi_3(\Delta_{j2}, \Delta_{j1}, 0; \boldsymbol{R}_{3d}) = \Phi_3(\Delta_{j1}, \Delta_{j2}, 0; \boldsymbol{R}_{3g})$, $\Phi(\Delta_{j2}) = \Phi_2(\Delta_{j2}, 0; \frac{R_{jk}}{\sqrt{2}}) + \Phi_2(\Delta_{j2}, 0; -\frac{R_{jk}}{\sqrt{2}})$, and $\Phi_2(\Delta_{j1}, \Delta_{j2}; 0) = \Phi_3(\Delta_{j1}, \Delta_{j2}, 0; \boldsymbol{R}_{3d}) + \Phi_3(\Delta_{j1}, \Delta_{j2}, 0; \boldsymbol{R}_{3g})$.

Therefore,

$$\frac{\partial F(t;\Delta_{j1},\Delta_{j2})}{\partial t}$$
$$= 2\partial\left\{\Phi_2(\Delta_{j2},0;\frac{t}{\sqrt{2}}) - \Phi_2(\Delta_{j2},0;-\frac{t}{\sqrt{2}}) - \Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3d}) + \Phi_3(\Delta_{j2},\Delta_{j1},0;\boldsymbol{R}_{3d})\right\}/\partial t$$
$$= 2\partial\Bigg[\Phi_2(\Delta_{j2},0;\frac{t}{\sqrt{2}}) - \left\{\Phi(\Delta_{j2}) - \Phi_2(\Delta_{j2},0;\frac{t}{\sqrt{2}})\right\}$$
$$+ \Phi_3(\Delta_{j2},\Delta_{j1},0;\boldsymbol{R}_{3d}) - \left\{\Phi_2(\Delta_{j1},\Delta_{j2};0) - \Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3g})\right\}\Bigg]/\partial t$$
$$= 2\partial\left\{2\Phi_2(\Delta_{j2},0;\frac{t}{\sqrt{2}}) - \Phi(\Delta_{j2}) - \Phi(\Delta_{j1})\Phi(\Delta_{j2}) + 2\Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3g})\right\}/\partial t$$
$$= 4\frac{\partial\Phi_2(\Delta_{j2},0;\frac{t}{\sqrt{2}})}{\partial t} + 4\frac{\partial\Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3g})}{\partial t}.$$

Then, we only need to prove that $\partial\Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3g})/\partial t \geqslant 0$.

By the chain rule,

$$\frac{\partial\Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3g})}{\partial t} = \sum_{i<k}\left\{\frac{\partial\Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3g})}{\partial\sigma_{ik}(r)}\frac{\partial\sigma_{ik}(r)}{\partial r}\right\}$$
$$= \frac{1}{\sqrt{2}}\left\{\frac{\partial\Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3g})}{\partial\sigma_{13}(r)} - \frac{\partial\Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3g})}{\partial\sigma_{23}(r)}\right\}$$

where σ_{ik} denotes the (i,k)-th element of $\boldsymbol{R}_{3g}$.

By Equation (3) in Plackett (1954), we have $\partial\phi_d/\partial\sigma_{ik} = \partial^2\phi_d/(\partial x_i\partial x_k)$. Hence,

$$\frac{\partial\Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3g})}{\partial t} = \frac{1}{\sqrt{2}}\left\{\frac{\partial\Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3g})}{\partial\sigma_{13}(r)} - \frac{\partial\Phi_3(\Delta_{j1},\Delta_{j2},0;\boldsymbol{R}_{3g})}{\partial\sigma_{23}(r)}\right\}$$
$$= \frac{1}{\sqrt{2}}\left[\int_{-\infty}^{\Delta_{j1}}\int_{-\infty}^{\Delta_{j2}}\int_{-\infty}^{0}\left\{\frac{\partial\phi_3(x_1,x_2,x_3;\boldsymbol{R}_{3g})}{\partial\sigma_{13}(r)} - \frac{\partial\phi_3(x_1,x_2,x_3;\boldsymbol{R}_{3g})}{\partial\sigma_{23}(r)}\right\}dx_1dx_2dx_3\right]$$
$$= \frac{1}{\sqrt{2}}\left[\int_{-\infty}^{\Delta_{j1}}\int_{-\infty}^{\Delta_{j2}}\int_{-\infty}^{0}\left\{\frac{\partial^2\phi_3(x_1,x_2,x_3;\boldsymbol{R}_{3g})}{\partial x_1\partial x_3} - \frac{\partial^2\phi_3(x_1,x_2,x_3;\boldsymbol{R}_{3g})}{\partial x_2\partial x_3}\right\}dx_1dx_2dx_3\right]$$

$$= \frac{1}{\sqrt{2}} \left\{ \int_{-\infty}^{\Delta_{j2}} \phi_3(\Delta_{j1}, x_2, 0; \boldsymbol{R}_{3g}) dx_2 - \int_{-\infty}^{\Delta_{j1}} \phi_3(x_1, \Delta_{j2}, 0; \boldsymbol{R}_{3g}) dx_1 \right\}$$
$$= \frac{1}{\sqrt{2}} \left\{ \int_{-\infty}^{\Delta_{j2}} \phi_3(\Delta_{j1}, x, 0; \boldsymbol{R}_{3g}) dx - \int_{-\infty}^{\Delta_{j1}} \phi_3(\Delta_{j2}, x, 0; \boldsymbol{R}_{3g}) dx \right\}$$
$$= \frac{1}{\sqrt{2}} \left\{ \phi_2(\Delta_{j1}, 0; \frac{R_{jk}}{\sqrt{2}}) \int_{-\infty}^{\Delta_{j2}} \phi(x) dx - \phi_2(\Delta_{j2}, 0; \frac{R_{jk}}{\sqrt{2}}) \int_{-\infty}^{\Delta_{j1}} \phi(x) dx \right\}$$
$$= \frac{1}{\sqrt{2}} \left\{ \phi_2(\Delta_{j1}, 0; \frac{R_{jk}}{\sqrt{2}}) \Phi(\Delta_{j2}) - \phi_2(\Delta_{j2}, 0; \frac{R_{jk}}{\sqrt{2}}) \Phi(\Delta_{j1}) \right\}.$$

We have

$$\phi_2(x, y; \frac{R_{jk}}{\sqrt{2}}) = f_{X,Y}(x, y) = f_{X|Y}(x|y) f_Y(y) = \phi\Big(\frac{x}{\sqrt{1 - R_{jk}^2/2}}\Big)\phi(y).$$

Then,

$$\frac{\partial \Phi_3(\Delta_{j1}, \Delta_{j2}, 0; \boldsymbol{R}_{3g})}{\partial t} = \frac{1}{\sqrt{2}} \left\{ \phi_2(\Delta_{j1}, 0; \frac{R_{jk}}{\sqrt{2}}) \Phi(\Delta_{j2}) - \phi_2(\Delta_{j2}, 0; \frac{R_{jk}}{\sqrt{2}}) \Phi(\Delta_{j1}) \right\}$$
$$= \frac{1}{\sqrt{2}} \phi(0) \left\{ \phi\Big(\frac{\Delta_{j1}}{\sqrt{1 - R_{jk}^2/2}}\Big) \Phi(\Delta_{j2}) - \phi\Big(\frac{\Delta_{j2}}{\sqrt{1 - R_{jk}^2/2}}\Big) \Phi(\Delta_{j1}) \right\}.$$

Therefore, we need to show that

$$\phi\left(\frac{\Delta_{j1}}{\sqrt{1 - R_{jk}^2/2}}\right) \Phi(\Delta_{j2}) > \phi\left(\frac{\Delta_{j2}}{\sqrt{1 - R_{jk}^2/2}}\right) \Phi(\Delta_{j1}),$$

which is equivalent to

$$\frac{\Phi(\Delta_{j2})}{\phi\Big(\frac{\Delta_{j2}}{s}\Big)} > \frac{\Phi(\Delta_{j1})}{\phi\Big(\frac{\Delta_{j1}}{s}\Big)},$$

where $s = \sqrt{1 - R_{jk}^2/2}$. Let $h(x) = \Phi(x)/\phi(x/s)$. Since $\Delta_{j2} > \Delta_{j1}$, we just need to show that $h(x)$ is an increasing function. We have

$$\frac{dh(x)}{dx} = \frac{\phi(x)\phi(x/s) + \frac{x}{s^2}\Phi(x)\phi(x/s)}{\phi^2(x/s)} = \frac{\phi(x)}{\phi(x/s)} + \frac{x}{s^2}\frac{\Phi(x)}{\phi(x/s)} = h(x)\left\{\frac{\phi(x)}{\Phi(x)} + \frac{x}{s^2}\right\}.$$

When $x \geqslant 0$, it's obvious that $\phi(x)/\Phi(x) + x/s^2 \geqslant 0$. Since $x \sim N(0, s^2)$, from the property of Mill's ratio, we also know that for $x \geqslant 0$, $\phi(-x)/\Phi(-x) - x/s^2 \geqslant 0$, which means that when $x < 0$, it also holds that $\phi(x)/\Phi(x) + x/s^2 \geqslant 0$. Hence $h(x)$ is an increasing function.

(b)

$$\begin{aligned}\frac{\partial F(t;\Delta_{j1},\Delta_{j2},\Delta_k)}{\partial t} &= 2\left[\frac{\partial \Phi_2(\Delta_{j2},\Delta_k;t)\{1-\Phi_1(\Delta_{j1})\}}{\partial t}+\frac{\partial \Phi_1(\Delta_{j2})\Phi_2(\Delta_{j1},\Delta_k;t)}{\partial t}\right]\\ &= 2\left[\{1-\Phi_1(\Delta_{j1})\}\frac{\partial \Phi_2(\Delta_{j2},\Delta_k;t)}{\partial t}+\Phi_1(\Delta_{j2})\frac{\partial \Phi_2(\Delta_{j1},\Delta_k;t)}{\partial t}\right]\\ &> 0.\end{aligned}$$

(c) Let

$$\boldsymbol{R}_{3i}=\begin{pmatrix}1 & \frac{R_{jk}}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{R_{jk}}{\sqrt{2}} & 1 & R_{jk}\\ \frac{1}{\sqrt{2}} & R_{jk} & 1\end{pmatrix},\quad \boldsymbol{R}_{4f}=\begin{pmatrix}1 & \frac{R_{jk}}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{R_{jk}}{\sqrt{2}} & 1 & R_{jk} & 0\\ \frac{1}{\sqrt{2}} & R_{jk} & 1 & 0\\ \frac{1}{\sqrt{2}} & 0 & 0 & 1\end{pmatrix},\quad \boldsymbol{R}_{4g}=\begin{pmatrix}1 & 0 & 0 & \frac{R_{jk}}{\sqrt{2}}\\ 0 & 1 & R_{jk} & \frac{R_{jk}}{\sqrt{2}}\\ 0 & R_{jk} & 1 & \frac{1}{\sqrt{2}}\\ \frac{R_{jk}}{\sqrt{2}} & \frac{R_{jk}}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1\end{pmatrix}.$$

$$\boldsymbol{R}_{4h}=\begin{pmatrix}1 & \frac{R_{jk}}{\sqrt{2}} & \frac{1}{\sqrt{2}} & -\frac{R_{jk}}{\sqrt{2}}\\ \frac{R_{jk}}{\sqrt{2}} & 1 & R_{jk} & 0\\ \frac{1}{\sqrt{2}} & R_{jk} & 1 & 0\\ -\frac{R_{jk}}{\sqrt{2}} & 0 & 0 & 1\end{pmatrix},\quad \boldsymbol{R}_{4i}=\begin{pmatrix}1 & \frac{R_{jk}}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{R_{jk}}{\sqrt{2}}\\ \frac{R_{jk}}{\sqrt{2}} & 1 & R_{jk} & 0\\ \frac{1}{\sqrt{2}} & R_{jk} & 1 & 0\\ \frac{R_{jk}}{\sqrt{2}} & 0 & 0 & 1\end{pmatrix}.$$

We have

$$\Phi_3(0,\Delta_{j2},\Delta_k;\boldsymbol{R}_{3f})=\Phi_2(\Delta_{j2},\Delta_k;R_{jk})-\Phi_3(0,\Delta_{j2},\Delta_k;\boldsymbol{R}_{3i}).$$

$$\Phi_4(0,\Delta_{j2},\Delta_k,\Delta_k;\boldsymbol{R}_{4c})=\Phi_1(\Delta_k)\Phi_2(\Delta_{j2},\Delta_k;R_{jk})-\Phi_3(0,\Delta_{j2},\Delta_k;\boldsymbol{R}_{3i})+\Phi_4(0,\Delta_{j2},\Delta_k,-\Delta_k;\boldsymbol{R}_{4f}).$$

$$\Phi_4(0,\Delta_{j2},\Delta_{j1}\Delta_k;\boldsymbol{R}_{4d})=\Phi_1(\Delta_{j2})\Phi_2(\Delta_{j1},\Delta_k;R_{jk})-\Phi_3(0,\Delta_{j1},\Delta_k;\boldsymbol{R}_{3i})+\Phi_4(\Delta_{j2},\Delta_{j1},\Delta_k,0;\boldsymbol{R}_{4g}).$$

$$\Phi_4(0,\Delta_{j2},\Delta_{j1},\Delta_k;\boldsymbol{R}_{4e})=\Phi_4(0,\Delta_{j2},\Delta_k,\Delta_{j1};\boldsymbol{R}_{4h})=\Phi_3(0,\Delta_{j2},\Delta_k;\boldsymbol{R}_{3i})-\Phi_4(0,\Delta_{j2},\Delta_k,-\Delta_{j1};\boldsymbol{R}_{4i}).$$

$$\Phi_4(0,\Delta_{j1},\Delta_{j1},\Delta_k;\boldsymbol{R}_{4e})=\Phi_4(0,\Delta_{j1},\Delta_k,\Delta_{j1};\boldsymbol{R}_{4h})=\Phi_3(0,\Delta_{j1},\Delta_k;\boldsymbol{R}_{3i})-\Phi_4(0,\Delta_{j1},\Delta_k,-\Delta_{j1};\boldsymbol{R}_{4i}).$$

$$
\begin{aligned}
&\partial F(t;\Delta_{j1},\Delta_{j2},\Delta_k)/\partial t\\
&=2\partial\Bigg[2\Big\{1-\Phi_1(\Delta_{j1})\Big\}\Phi_2(\Delta_{j2},\Delta_k;R_{jk})+\Phi_2(0,\Delta_{j2};-\frac{R_{jk}}{\sqrt{2}})-\Phi_1(\Delta_k)\Phi_2(\Delta_{j2},\Delta_k;R_{jk})\\
&-2\Phi_2(\Delta_{j1},\Delta_k;R_{jk})\Phi_2(\Delta_{j2},\Delta_k;R_{jk})+2\Phi_3(0,\Delta_{j2},\Delta_k;\boldsymbol{R}_{3e})-2\Phi_3(0,\Delta_{j2},\Delta_k;\boldsymbol{R}_{3f})\\
&+2\Phi_3(\Delta_{j2},\Delta_{j1},0;\boldsymbol{R}_{3d})+2\Phi_4(0,\Delta_{j2},\Delta_k,\Delta_k;\boldsymbol{R}_{4c})+2\Phi_4(0,\Delta_{j2},\Delta_{j1}\Delta_k;\boldsymbol{R}_{4d})\\
&+2\Phi_4(0,\Delta_{j2},\Delta_{j1},\Delta_k;\boldsymbol{R}_{4e})+2\Phi_4(0,\Delta_{j1},\Delta_{j1},\Delta_k;\boldsymbol{R}_{4e})\\
&-2\Phi_5(0,\Delta_{j1},\Delta_{j1},\Delta_k,\Delta_k;\boldsymbol{R}_5)+2\Phi_5(0,\Delta_{j2},\Delta_{j1},\Delta_k,\Delta_k;\boldsymbol{R}_5)\Bigg]/\partial t\\
&=2\partial\Bigg[2\Big\{1-\Phi_1(\Delta_{j1})\Big\}\Phi_2(\Delta_{j2},\Delta_k;R_{jk})+\Phi_1(\Delta_{j2})-\Phi_2(0,\Delta_{j2};R_{jk}/\sqrt{2})\\
&-2\Phi_2(\Delta_{j1},\Delta_k;R_{jk})\Phi_2(\Delta_{j2},\Delta_k;R_{jk})+2\Phi_3(0,\Delta_{j2},\Delta_k;\boldsymbol{R}_{3e})-2\Phi_2(\Delta_{j2},\Delta_k;R_{jk})\\
&+2\Phi_3(\Delta_{j2},\Delta_{j1},0;\boldsymbol{R}_{3d})+\Phi_1(\Delta_k)\Phi_2(\Delta_{j2},\Delta_k;R_{jk})+2\Phi_1(\Delta_{j2})\Phi_2(\Delta_{j1},\Delta_k;R_{jk})\\
&+2\Phi_4(0,\Delta_{j2},\Delta_k,-\Delta_k;\boldsymbol{R}_{4f})+2\Phi_4(\Delta_{j2},\Delta_{j1},\Delta_k,0;\boldsymbol{R}_{4g})\\
&+2\Phi_3(0,\Delta_{j2},\Delta_k;\boldsymbol{R}_{3i})-2\Phi_4(0,\Delta_{j2},\Delta_k,-\Delta_{j1};\boldsymbol{R}_{4i})\\
&-2\Phi_4(0,\Delta_{j1},\Delta_k,-\Delta_{j1};\boldsymbol{R}_{4i})\\
&-2\Phi_5(0,\Delta_{j1},\Delta_{j1},\Delta_k,\Delta_k;\boldsymbol{R}_5)+2\Phi_5(0,\Delta_{j2},\Delta_{j1},\Delta_k,\Delta_k;\boldsymbol{R}_5)\Bigg]/\partial t\\
&>2\partial\big[2\Big\{1-\Phi_1(\Delta_{j1})\Big\}\Phi_2(\Delta_{j2},\Delta_k;R_{jk})+\Phi_1(\Delta_{j2})-\Phi_2(0,\Delta_{j2};R_{jk}/\sqrt{2})\\
&-2\Phi_2(\Delta_{j1},\Delta_k;R_{jk})\Phi_2(\Delta_{j2},\Delta_k;R_{jk})+2\Phi_3(0,\Delta_{j2},\Delta_k;\boldsymbol{R}_{3e})-2\Phi_2(\Delta_{j2},\Delta_k;R_{jk})\\
&+2\Phi_3(\Delta_{j2},\Delta_{j1},0;\boldsymbol{R}_{3d})+\Phi_1(\Delta_k)\Phi_2(\Delta_{j2},\Delta_k;R_{jk})+2\Phi_1(\Delta_{j2})\Phi_2(\Delta_{j1},\Delta_k;R_{jk})\\
&+2\Phi_4(0,\Delta_{j2},\Delta_k,-\Delta_k;\boldsymbol{R}_{4f})+2\Phi_4(\Delta_{j2},\Delta_{j1},\Delta_k,0;\boldsymbol{R}_{4g})+2\Phi_3(0,\Delta_{j2},\Delta_k;\boldsymbol{R}_{3i})\\
&-2\Phi_4(0,\Delta_{j2},\Delta_k,-\Delta_{j1};\boldsymbol{R}_{4i})-2\Phi_4(0,\Delta_{j1},\Delta_k,-\Delta_{j1};\boldsymbol{R}_{4i})\big]/\partial t\\
&>0
\end{aligned}
$$

(d) $\partial F(t;\Delta_{j1},\Delta_{j2},\Delta_{k1},\Delta_{k2})/\partial t$

$$
\begin{aligned}
&= 2\partial\Big\{\Phi_2(\Delta_{j2},\Delta_{k2};t)-\Phi_2(\Delta_{j2},\Delta_{k2};t)\Phi_1(\Delta_{k1})+\Phi_2(\Delta_{j2},\Delta_{k1};t)\Phi_1(\Delta_{k2})\\
&\qquad -\Phi_2(\Delta_{j2},\Delta_{k2};t)\Phi_1(\Delta_{j1})+\Phi_2(\Delta_{j1},\Delta_{k2};t)\Phi_1(\Delta_{j2})\\
&\qquad +\Phi_2(\Delta_{j2},\Delta_{k2};t)\Phi_2(\Delta_{j1},\Delta_{k1};t)-\Phi_2(\Delta_{j2},\Delta_{k1};t)\Phi_2(\Delta_{j1},\Delta_{k2};t)\Big\}/\partial t\\
&= 2\partial\Big[\Phi_2(\Delta_{j2},\Delta_{k2};t)\Big\{1-\Phi_1(\Delta_{k1})-\Phi_1(\Delta_{j1})+\Phi_2(\Delta_{j1},\Delta_{k1};t)\Big\}\\
&\qquad +\Phi_2(\Delta_{j1},\Delta_{k2};t)\Phi_1(\Delta_{j2})+\Phi_2(\Delta_{j2},\Delta_{k1};t)\big\{\Phi_1(\Delta_{k2})-\Phi_2(\Delta_{j1},\Delta_{k2};t)\big\}\Big]/\partial t\\
&= 2\frac{\partial\big\{\Phi_2(\Delta_{j2},\Delta_{k2};t)\Phi_2(-\Delta_{j1},-\Delta_{k1};t)\big\}}{\partial t}+2\Phi_1(\Delta_{j2})\frac{\partial\Phi_2(\Delta_{j1},\Delta_{k2};t)}{\partial t}\\
&\qquad +2\frac{\partial\Big[\Phi_2(\Delta_{j2},\Delta_{k1};t)\big\{\mathbf{P}(Z_k<\Delta_{k2})-\mathbf{P}(Z_j<\Delta_{j1},Z_k<\Delta_{k2})\big\}\Big]}{\partial t}\\
&= 2\frac{\partial\big\{\Phi_2(\Delta_{j2},\Delta_{k2};t)\Phi_2(-\Delta_{j1},-\Delta_{k1};t)\big\}}{\partial t}+2\Phi_1(\Delta_{j2})\frac{\partial\Phi_2(\Delta_{j1},\Delta_{k2};t)}{\partial t}\\
&\qquad +2\frac{\partial\Big\{\Phi_2(\Delta_{j2},\Delta_{k1};t)\mathbf{P}(Z_j\geqslant\Delta_{j1},Z_k<\Delta_{k2})\Big\}}{\partial t}\\
&> 0.
\end{aligned}
$$

Web Appendix B. Derivation of Algorithm 1

Let $\mathbf{\Theta}=(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2,\mathbf{\Sigma}_U)$ and $\widehat{\boldsymbol{R}}=\widehat{\boldsymbol{R}}_1+\widehat{\boldsymbol{R}}_2$. We have $\ell(\mathbf{\Theta})=(1/2)\mathrm{tr}\{(\widehat{\boldsymbol{R}}-\mathbf{\Sigma}_1-\mathbf{\Sigma}_2-2\mathbf{\Sigma}_U)^T(\widehat{\boldsymbol{R}}-\mathbf{\Sigma}_1-\mathbf{\Sigma}_2-2\mathbf{\Sigma}_U)\}$. To start the iterations, we need to obtain an initial estimator $\widehat{\mathbf{\Theta}}^{(0)}=(\widehat{\mathbf{\Sigma}}_1^{(0)},\widehat{\mathbf{\Sigma}}_2^{(0)},\widehat{\mathbf{\Sigma}}_U^{(0)})$. First, we estimate r_g by

$$\widehat{r}_g=\mathrm{argmax}_{j\leqslant\min\{(n_g,p)\}}\lambda_{j-1}(\widehat{\boldsymbol{R}}_g)/\lambda_j(\widehat{\boldsymbol{R}}_g), \tag{S4}$$

where $\lambda_j(\widehat{\boldsymbol{R}}_g)$ is the j-th largest eigenvalue of $\widehat{\boldsymbol{R}}_g$. Let $\widehat{\boldsymbol{V}}_g=(\widehat{\boldsymbol{v}}_g^1,\ldots,\widehat{\boldsymbol{v}}_g^{\widehat{r}_g})$, where $\widehat{\boldsymbol{v}}_g^j$ is the eigenvector corresponding to $\lambda_j(\widehat{\boldsymbol{R}}_g)$ and $\widehat{\boldsymbol{D}}_g=\mathrm{diag}\{\lambda_1(\widehat{\boldsymbol{R}}_g),\ldots,\lambda_{\widehat{r}_g}(\widehat{\boldsymbol{R}}_g)\}$. Then, we let $\widehat{\mathbf{\Sigma}}_g^{(0)}=\widehat{\boldsymbol{V}}_g\widehat{\boldsymbol{D}}_g\widehat{\boldsymbol{V}}_g^T$ for $g\in\{1,2\}$ and $\widehat{\mathbf{\Sigma}}_U^{(0)}=(1/2)\big(\widehat{\boldsymbol{R}}-\widehat{\mathbf{\Sigma}}_1^{(0)}-\widehat{\mathbf{\Sigma}}_2^{(0)}\big)$. Denote the solution of $\mathbf{\Theta}$ at the h-th iteration as $\widehat{\mathbf{\Theta}}^{(h)}=(\widehat{\mathbf{\Sigma}}_1^{(h)},\widehat{\mathbf{\Sigma}}_2^{(h)},\widehat{\mathbf{\Sigma}}_U^{(h)})$. At the $(h+1)$-th iteration, we first fix $\widehat{\mathbf{\Sigma}}_2^{(h)},\widehat{\mathbf{\Sigma}}_U^{(h)}$ and solve for $\widehat{\mathbf{\Sigma}}_1^{(h+1)}$. This becomes a spectral regularization problem.

As shown in Mazumder et al. (2010), this problem can be solved by a hard-thresholding Singular Value Decomposition (SVD). In particular, let $\widehat{\boldsymbol{R}} - \widehat{\boldsymbol{\Sigma}}_2^{(h)} - 2\widehat{\boldsymbol{\Sigma}}_U^{(h)} = \boldsymbol{U}_{\widehat{r}_1}^{(h)} \boldsymbol{D}_{\widehat{r}_1}^{(h)} \boldsymbol{V}_{\widehat{r}_1}^{(h)T}$ be the rank-$\widehat{r}_1$ SVD. We have $\widehat{\boldsymbol{\Sigma}}_1^{(h+1)} = \boldsymbol{U}_{\widehat{r}_1}^{(h)} \boldsymbol{S}_{\nu_1}(\boldsymbol{D}_{\widehat{r}_1}^{(h)}) \boldsymbol{V}_{\widehat{r}_1}^{(h)T}$, where $\boldsymbol{S}_{\nu_1}(\boldsymbol{D}_{\widehat{r}_1}^{(h)}) = \mathrm{diag}\{(\lambda_1^{(h)} - \nu_1)_+, \ldots, (\lambda_{\widehat{r}_1}^{(h)} - \nu_1)_+\}$ and $\lambda_j^{(h)}$ is the j-th largest singular value of $\boldsymbol{D}_{\widehat{r}_1}^{(h)}$. Then, we check if $\widehat{\boldsymbol{\Sigma}}_1^{(h+1)}$ is positive semidefinite. If not, we project it to the nearest positive definite matrix by solving

$$\underset{\lambda_{\min}(\boldsymbol{A}) \geqslant 0}{\operatorname{argmin}} \|\widehat{\boldsymbol{\Sigma}}_1^{(h+1)} - \boldsymbol{A}\|_F, \tag{S5}$$

where $\lambda_{\min}(\boldsymbol{A})$ is the smallest eigenvalue of $\boldsymbol{A}$. With a slight abuse of notation, we still denote the solution to (S5) as $\widehat{\boldsymbol{\Sigma}}_1^{(h+1)}$. Then, we fix $(\widehat{\boldsymbol{\Sigma}}_1^{(h+1)}, \widehat{\boldsymbol{\Sigma}}_U^{(h)})$ and solve for $\widehat{\boldsymbol{\Sigma}}_2^{(h+1)}$, which can be done using the same hard-thresholding SVD. Lastly, we fix $(\widehat{\boldsymbol{\Sigma}}_1^{(h+1)}, \widehat{\boldsymbol{\Sigma}}_2^{(h+1)})$ and solve for $\widehat{\boldsymbol{\Sigma}}_U^{(h+1)}$. In this step, we use the proximal gradient descent algorithm (Parikh and Boyd, 2014) to solve the corresponding L_1-penalization problem. The solution is given by $\widehat{\boldsymbol{\Sigma}}_U^{(h+1)} = s(\widehat{\boldsymbol{\Sigma}}_U^{(h)} - d\nabla_{\boldsymbol{\Sigma}_U} \ell(\widehat{\boldsymbol{\Sigma}}_1^{(h+1)}, \widehat{\boldsymbol{\Sigma}}_2^{(h+1)}, \widehat{\boldsymbol{\Sigma}}_U^{(h)}), \nu_3 d)$, where d is the step size of iterations, $\nabla_{\boldsymbol{\Sigma}_U} \ell(\widehat{\boldsymbol{\Sigma}}_1^{(h+1)}, \widehat{\boldsymbol{\Sigma}}_2^{(h+1)}, \widehat{\boldsymbol{\Sigma}}_U^{(h)}) = 4\widehat{\boldsymbol{\Sigma}}_U^{(h)} - 2(\widehat{\boldsymbol{R}} - \widehat{\boldsymbol{\Sigma}}_1^{(h+1)} - \widehat{\boldsymbol{\Sigma}}_2^{(h+1)})$ and $s(\boldsymbol{x}, \pi)$ is the element-wise soft-thresholding operator, whose (i,j)-th element is defined as $s(\boldsymbol{x}, \pi)_{i,j} = \mathrm{sign}(x_{i,j})(|x_{i,j}| - \pi)_+$. As for the choice of d, we follow Parikh & Boyd (2014, Section 4.2) to perform a backtracking line search. That is, we iteratively decrease d until $\ell(\widehat{\boldsymbol{\Theta}}^{(h+1)}) \leqslant Q_d\big(\widehat{\boldsymbol{\Theta}}^{(h+1)}; \widehat{\boldsymbol{\Theta}}^{(h)}\big)$, where $Q_d\big(\widehat{\boldsymbol{\Theta}}^{(h+1)}; \widehat{\boldsymbol{\Theta}}^{(h)}\big) = \ell(\widehat{\boldsymbol{\Sigma}}_1^{(h+1)}, \widehat{\boldsymbol{\Sigma}}_2^{(h+1)}, \widehat{\boldsymbol{\Sigma}}_U^{(h)}) + \langle \widehat{\boldsymbol{\Sigma}}_U^{(h+1)} - \widehat{\boldsymbol{\Sigma}}_U^{(h)}, \nabla_{\boldsymbol{\Sigma}_U} \ell(\widehat{\boldsymbol{\Sigma}}_1^{(h+1)}, \widehat{\boldsymbol{\Sigma}}_2^{(h+1)}, \widehat{\boldsymbol{\Sigma}}_U^{(h)}) \rangle + (1/2d)||\widehat{\boldsymbol{\Sigma}}_U^{(h+1)} - \widehat{\boldsymbol{\Sigma}}_U^{(h)}||_F^2$. We stop the iterations when the proportion of the maximal changes of $(\widehat{\boldsymbol{\Sigma}}_1, \widehat{\boldsymbol{\Sigma}}_2, \widehat{\boldsymbol{\Sigma}}_U)$ between two consecutive iterations is less than ζ, where $\zeta \in (0,1)$ is a user-defined stopping threshold, which is set to be 0.1 by us.

Web Appendix C. Numerical Convergence of Algorithm 1

We use Model 1 from Scenario 2 to investigate the numerical convergence of Algorithm 1. We run Algorithm 1 on this model with fixed ν_1, ν_2, ν_3, and plot values of the objective function (panel a), $\|\widehat{\mathbf{\Sigma}}_U - \mathbf{\Sigma}_U\|_F$ (panel b), $\|\widehat{\mathbf{\Sigma}}_g - \mathbf{\Sigma}_g\|_F$ (panels c and d), and $\|\widehat{\mathbf{\Sigma}}_1 + \widehat{\mathbf{\Sigma}}_2 - \mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_F$ (panel e) along the iterations of Algorithm 1; see Figure S1. In panel (a), we use red dots to represent the values after one full iteration, i.e., after updating $\widehat{\mathbf{\Sigma}}_1$, $\widehat{\mathbf{\Sigma}}_2$ and $\widehat{\mathbf{\Sigma}}_U$. These figures indicate that our algorithm can finally converge.

[Figure 1 about here.]

Web Appendix D. Rank Estimation

We inspect how well our method estimates the true ranks and how sensitive our decomposition method is to the rank estimation. Our method obtains an initial estimator of r_g by letting $\widehat{r}_g = \operatorname{argmax}_{j \leqslant \min\{(n_g,p)\}} \lambda_{j-1}(\widehat{\boldsymbol{R}}_g)/\lambda_j(\widehat{\boldsymbol{R}}_g)$, where $\lambda_j(\widehat{\boldsymbol{R}}_g)$ is the j-th largest eigenvalue of $\widehat{\boldsymbol{R}}_g$. Such a rank estimator is commonly used in factor analysis literature (Lam and Yao, 2012; Ahn and Horenstein, 2013). We remark that $\widehat{r}_g$ is allowed to over-estimate r_g as the nuclear norm penalty in (5) can further shrink the rank estimator; see Algorithm 1. To see how our method estimates the true ranks, we report in Table S1 the percentage of simulation runs that correctly estimated, over-estimated or under-estimated the ranks in the two groups in all numerical studies. It can be seen that our method can mostly correctly estimate the true ranks in Scenarios 1, 3, 4 and 6. In Scenario 2 and 5, it tends to under- or over- estimate the ranks. The reason is because in these two scenarios $\mathbf{\Sigma}_g$ takes a smaller proportion of $\boldsymbol{R}_g$ so that the rank estimation is harder compared with the other settings. However, if we compare the simulation results in these two scenarios with the counterparts in other scenarios; see Figures 1(b) and 2 in the main manuscript, we find that our method's performance is not sensitive to rank estimation. Even in Scenarios 2 and 5, it still outperforms the competitors.

[Table 1 about here.]

Web Appendix E. Alternative loss function

In our paper, we choose $\|\widehat{\boldsymbol{R}}_1 + \widehat{\boldsymbol{R}}_2 - \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2 - 2\boldsymbol{\Sigma}_U\|_F^2$ as our loss function. As one reviewer pointed out, an alternative loss function could be the sum-of-norms loss: $\sum_{g=1}^{2} \|\widehat{\boldsymbol{R}}_g - \boldsymbol{\Sigma}_g - \boldsymbol{\Sigma}_U\|_F^2$. We apply both loss functions to Model 1 of Scenario 1 with 100 simulations and compare their numerical performance on estimating $\boldsymbol{\Sigma}_g$ and $\boldsymbol{\Sigma}_U$; see Figure S2. It is seen that our loss function has better estimation error for estimating $\boldsymbol{\Sigma}_g$ and higher sensitivity for estimating $\boldsymbol{\Sigma}_U$, even though its specificity is slightly worse. In practice, both loss functions can be used for variation decomposition.

[Figure 2 about here.]

Web Appendix F. The set-up of each scenario in simulation studies

In all cases, we set $r_1 = \text{rank}(\boldsymbol{\Sigma}_1) = 3$ and $r_2 = \text{rank}(\boldsymbol{\Sigma}_2) = 2$. In the low-dimensional settings, we set $n_g = 100$ for $g \in \{1, 2\}$, $p = 60$ and consider the following choices of $\boldsymbol{\Sigma}_g$ and $\boldsymbol{\Sigma}_U$.

Scenario 1: We set $\boldsymbol{\Sigma}_1 = \boldsymbol{Q}_1\boldsymbol{D}_1\boldsymbol{Q}_1^T$, where $\boldsymbol{D}_1 = \text{diag}(7, 6.5, 6)$ and $\boldsymbol{Q}_1 \in \mathbb{R}^{p\times r_1}$ is an orthonormal matrix. To generate $\boldsymbol{Q}_1$, we start with $\boldsymbol{A}_1 = (\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3) \in \mathbb{R}^{p\times r_1}$, where the first 30 elements of $\boldsymbol{a}_1$ range from 0.02 to 0.6 with increments of 0.02 and the next 30 elements of $\boldsymbol{a}_1$ range from 0.71 to 1 with increments of 0.01, elements of $\boldsymbol{a}_2$ range from -1.16 to 1.2 with increments of 0.04, and elements of $\boldsymbol{a}_3$ range from 0.12 to 1.3 with increments of 0.02. We then apply the Gram-Schmidt normalization to $\boldsymbol{A}_1$ to obtain $\boldsymbol{Q}_1$. We set diagonal elements of $\boldsymbol{\Sigma}_U$ to be 1 minus the diagonal elements of $\boldsymbol{\Sigma}_1$, and its (i, j)-th off-diagonal element as $\sigma_{u,ij} = \sigma_{u,i}\sigma_{u,j}\rho^{|i-j|}$ if $|i - j| = 1$ and $\sigma_{u,ij} = 0$ otherwise, where $\rho = 0.5$ and $\sigma_{u,j}$ denotes the j-th diagonal element of $\boldsymbol{\Sigma}_U$. We set $\boldsymbol{\Sigma}_2 = \mathbf{w}_2\mathbf{w}_2^T$, where $\mathbf{w}_2 = (\mathbf{w}_{21}, \mathbf{w}_{22})$ with the (1:4, 31:35, 54:60)-th the elements of $\mathbf{w}_{21}$ being 0.65 and the rest being zeros. Here, we use $i : j$ to denote a sequence of consecutive integers from i to j. The j-th element of $\mathbf{w}_{22}$ is set as

$w_{22;j} = \sqrt{1 - \sigma_{u,j} - w_{21;j}^2}$, for $1 \leqslant j \leqslant 60$, where $w_{22;j}$ denotes the j-th element of w_{22}. Under this scenario, $\|\mathbf{\Sigma}_U\|_F \approx 44.89\% \|\boldsymbol{R}_1\|_F$ and $\|\mathbf{\Sigma}_U\|_F \approx 36.82\% \|\boldsymbol{R}_2\|_F$, meaning the common variation captures 44.89% and 36.82% of total variation in these two groups.

Scenario 2: We set $\boldsymbol{D}_1 = \text{diag}(3.5, 3.5, 3.5)$. Using the same $\boldsymbol{Q}_1$ as in Scenario 1, we let $\mathbf{\Sigma}_1 = \boldsymbol{Q}_1 \boldsymbol{D}_1 \boldsymbol{Q}_1^T$. We set the diagonal elements of $\mathbf{\Sigma}_U$ to be 1 minus the diagonal elements of $\mathbf{\Sigma}_1$, and its (i,j)-th off-diagonal element $\sigma_{u,ij} = \sigma_{u,i}\sigma_{u,j}\rho^{|i-j|}$ if $|i-j| \leqslant 2$ and $\sigma_{u,ij} = 0$ otherwise, where $\rho = 0.5$. We set $\mathbf{\Sigma}_2$ the same as in Scenario 1 with the (1:3, 31:35, 55:60)-th elements of $\mathbf{w}_{21}$ chosen as 0.503 and the rest as 0. Under this scenario, $\|\mathbf{\Sigma}_U\|_F \approx 68.41\% \|\boldsymbol{R}_1\|_F$ and $\|\mathbf{\Sigma}_U\|_F \approx 62.72\% \|\boldsymbol{R}_2\|_F$.

Scenario 3: We set $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ the same as in Scenario 1 with the (1:4, 31:35, 54:60)-th elements of $\mathbf{w}_{21}$ to be 0.65 and the rest to be 0. We set the diagonal elements of $\mathbf{\Sigma}_U$ to be 1 minus the diagonal elements of $\mathbf{\Sigma}_1$. The off-diagonal elements of $\mathbf{\Sigma}_U$ are block-wise sparse structure and given in Figure S3(a). In this scenario, $\|\mathbf{\Sigma}_U\|_F \approx 42.09\% \|\boldsymbol{R}_1\|_F$ and $\|\mathbf{\Sigma}_U\|_F \approx 33.31\% \|\boldsymbol{R}_2\|_F$.

In the high-dimensional settings, we set $n_g = 50$ for $g \in \{1, 2\}$, $p = 90$ and consider the following choices of $\mathbf{\Sigma}_g$ and $\mathbf{\Sigma}_U$.

Scenario 4: We set $\mathbf{\Sigma}_1 = \boldsymbol{Q}_1 \boldsymbol{D}_1 \boldsymbol{Q}_1^T$, where $\boldsymbol{D}_1 = \text{diag}(11, 10.5, 10)$ and $\boldsymbol{Q}_1 \in \mathbb{R}^{p \times r_1}$ is an orthonormal matrix. To generate $\boldsymbol{Q}_1$, we start with $\boldsymbol{A}_1 = (\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3)$, where the first 45 elements of $\boldsymbol{a}_1$ range from 0.02 to 0.9 with increments of 0.02 and the rest elements range from 1.01 to 1.45 with increments of 0.01, elements of $\boldsymbol{a}_2$ range from -1.76 to 1.8 with increments of 0.04, and elements of $\boldsymbol{a}_3$ range from 0.12 to 1.9 with increments of 0.02. We then apply Gram-Schmidt normalization to $\boldsymbol{A}_1$ to obtain $\boldsymbol{Q}_1$. We set diagonal elements of $\mathbf{\Sigma}_U$ to be 1 minus the diagonal elements of $\mathbf{\Sigma}_1$, and its (i,j)-th off-diagonal element as $\sigma_{u,ij} = \sigma_{u,i}\sigma_{u,j}\rho^{|i-j|}$ if $|i-j| = 1$ and $\sigma_{u,ij} = 0$ otherwise, where $\rho = 0.5$. We set $\mathbf{\Sigma}_2 = \mathbf{w}_2 \mathbf{w}_2^T$, where $\mathbf{w}_2 = (\mathbf{w}_{21}, \mathbf{w}_{22})$ with the (1:8, 46:52, 81:90)-th elements of $\mathbf{w}_{21}$ being set as 0.65 and the rest as 0. The j-th element of $\mathbf{w}_{22}$ is set as $w_{22;j} = \sqrt{1 - \sigma_{u,j} - w_{21;j}^2}$. Under this scenario, $\|\mathbf{\Sigma}_U\|_F \approx 36.37\% \|\boldsymbol{R}_1\|_F$ and $\|\mathbf{\Sigma}_U\|_F \approx 28.83\% \|\boldsymbol{R}_2\|_F$.

Scenario 5: This scenario is the same as Scenario 2, except that we set $\boldsymbol{D}_1 = \text{diag}(4.5, 4.5, 4.5)$ and use the same $\boldsymbol{Q}_1$ as in Scenario 4. We set $\mathbf{\Sigma}_2 = \mathbf{w}_2\mathbf{w}_2^T$, where $\mathbf{w}_2 = (\mathbf{w}_{21}, \mathbf{w}_{22})$ with the (1:6, 46:51, 82:90)-th elements of $\mathbf{w}_{21}$ set to be 0.459 and the rest to be 0. Under this scenario, $\|\mathbf{\Sigma}_U\|_F \approx 69.39\% \|\boldsymbol{R}_1\|_F$ and $\|\mathbf{\Sigma}_U\|_F \approx 62.82\% \|\boldsymbol{R}_2\|_F$.

Scenario 6: We set $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ the same as in Scenario 4 with the (1:5, 46:50, 83:90)-th elements of $\mathbf{w}_{21}$ to be 0.72 and the rest to be 0. $\mathbf{\Sigma}_U$ is block-wise sparse and given in Figure S3(b). In this scenario, $\|\mathbf{\Sigma}_U\|_F \approx 33.2\% \|\boldsymbol{R}_1\|_F$ and $\|\mathbf{\Sigma}_U\|_F \approx 25.47\% \|\boldsymbol{R}_2\|_F$.

[Figure 3 about here.]

Web Appendix G. Additional results for real data analysis

Estimates of non-zero gene-gene and gene-SNP correlations in $\widehat{\mathbf{\Sigma}}_U$ are given in Tables S2 and S3.

[Table 2 about here.]

[Table 3 about here.]

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81,** 1203–1227.

Fan, J., Liu, H., Ning, Y., and Zou, H. (2017). High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society: Series B* **79,** 405–421.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics* **40,** 694–726.

Mazumder, R., Hastie, T., and Tibshirani, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research* **11,** 2287–2322.

Parikh, N. and Boyd, S. (2014). Proximal algorithms. *Foundations and Trends in Optimization* **1,** 127–239.

Plackett, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika* **41,** 351–360.

Yoon, G., Carroll, R. J., and Gaynanova, I. (2020). Sparse semiparametric canonical correlation analysis for data of mixed types. *Biometrika* **107,** 609–625.

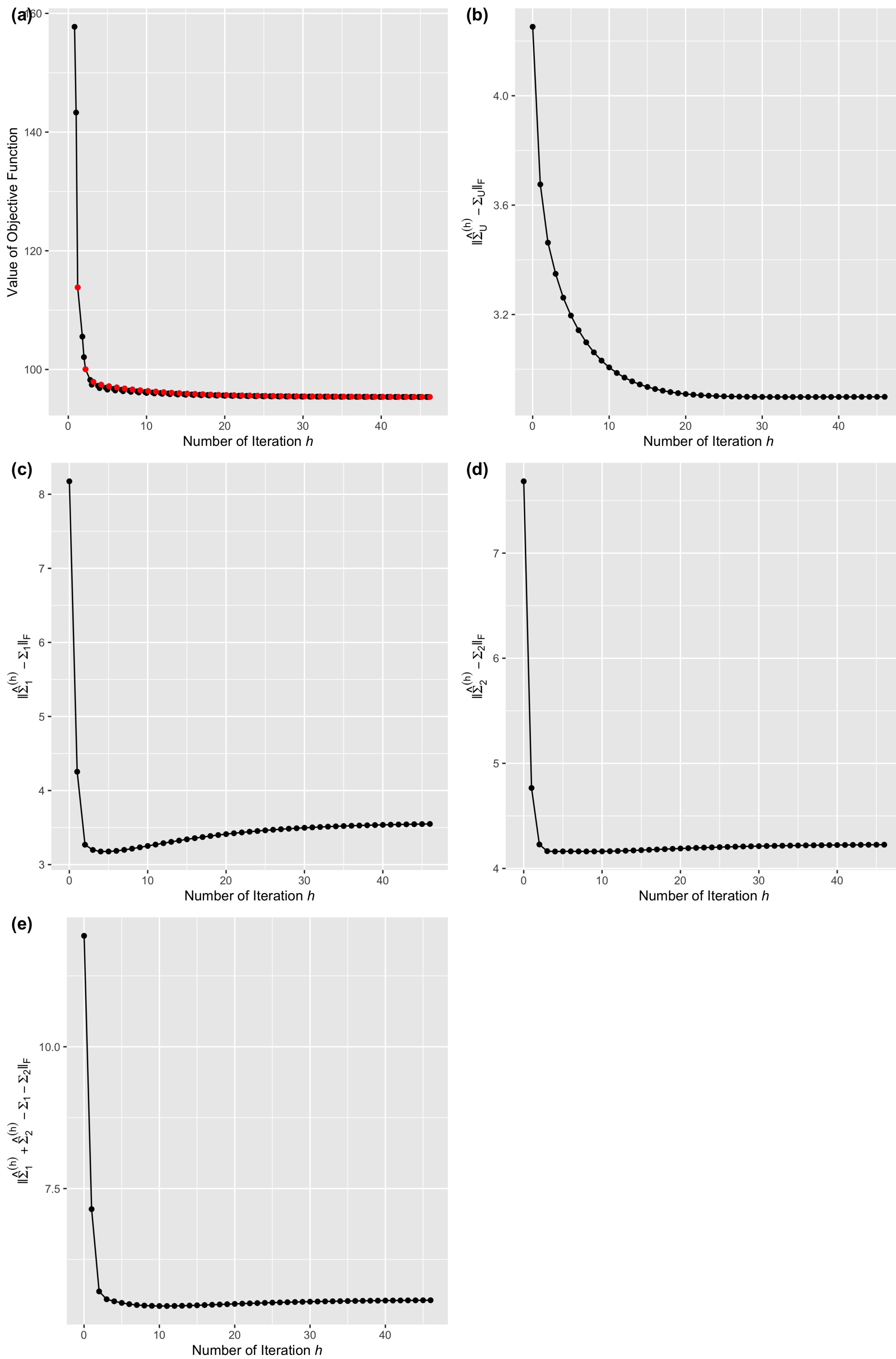

Figure S1: Numerical convergence of our algorithm

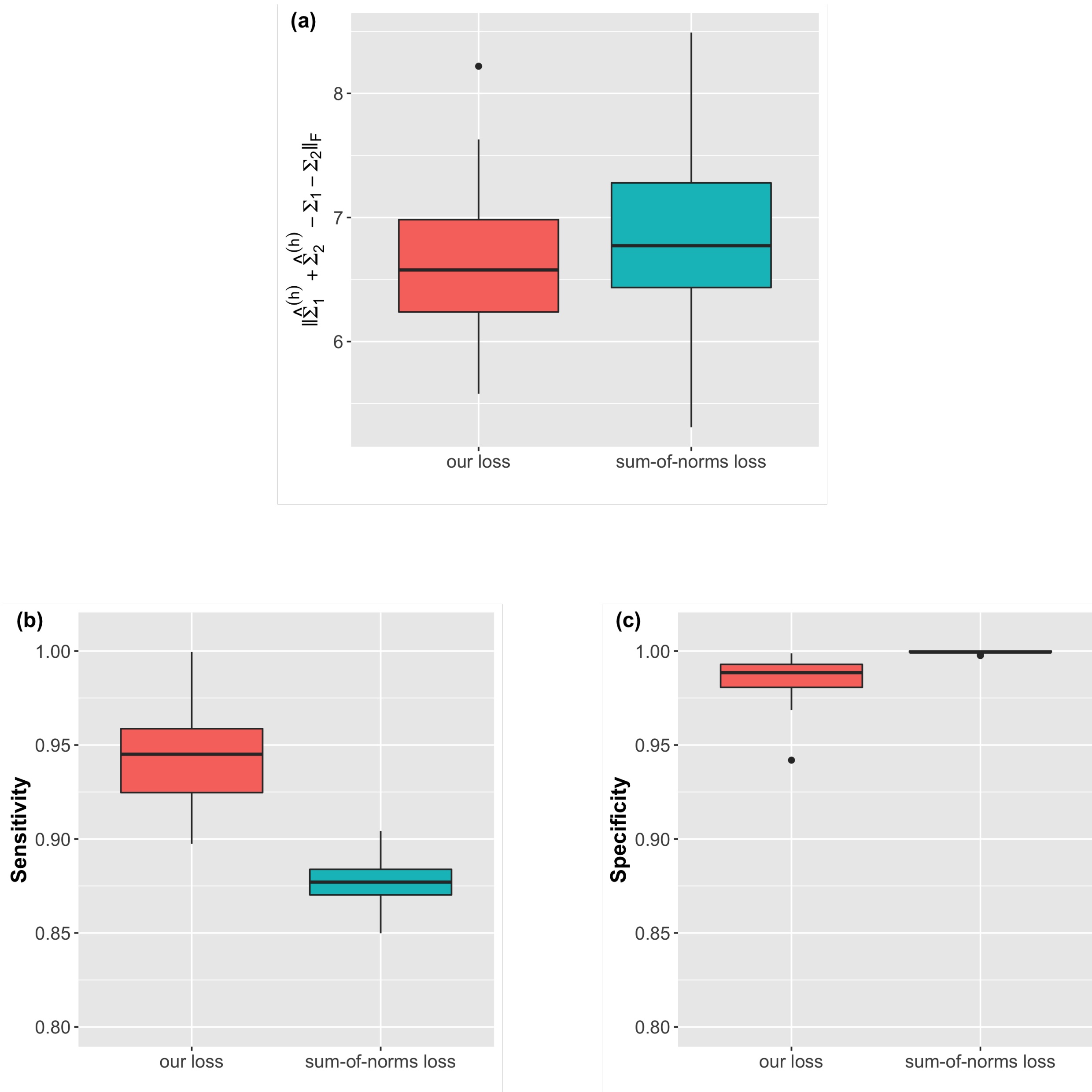

Figure S2: Comparison of two loss functions

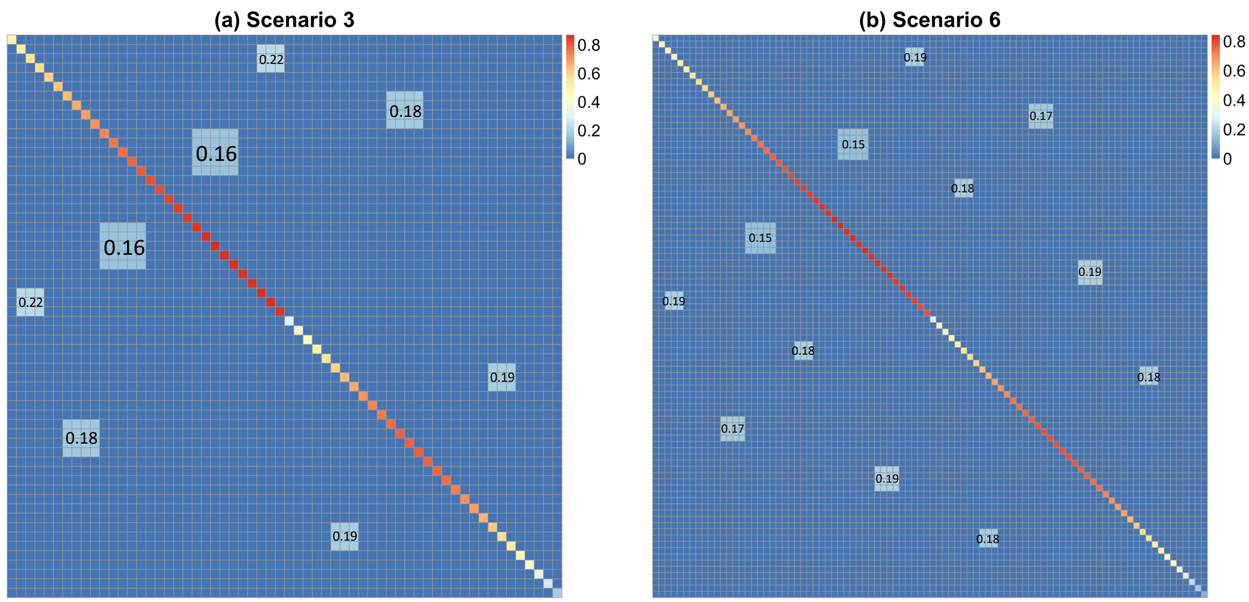

Figure S3: Heatmaps of $\mathbf{\Sigma}_U$ for Scenarios 3 and 6. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

Table S1: Percentage of simulation runs that have correctly estimated, over-estimated or under-estimated the numbers of ranks in the two groups.

		Group 1			Group 2		
Scenario	Model	Under-estimated	Correctly estimated	Over-estimated	Under-estimated	Correctly estimated	Over-estimated
Scenario 1	Model 1	0.00%	100.0%	0.0%	17.00%	81.0%	2%
	Model 2	0.00%	100.0%	0.0%	17.00%	83.0%	0%
	Model 3	0.00%	100.0%	0.0%	24.00%	76.0%	0%
Scenario 2	Model 1	34.00%	62.0%	4.0%	25.00%	50.0%	25%
	Model 2	27.00%	71.0%	2.0%	26.00%	51.0%	23%
	Model 3	24.00%	74.0%	2.0%	26.00%	58.0%	16%
Scenario 3	Model 1	0.00%	100.0%	0.0%	7.00%	93.0%	0%
	Model 2	1.00%	99.0%	0.0%	12.00%	88.0%	0%
	Model 3	0.00%	100.0%	0.0%	18.00%	82.0%	0%
Scenario 4	Model 1	5.00%	95.0%	0.0%	6.00%	75.0%	19%
	Model 2	7.00%	93.0%	0.0%	5.00%	75.0%	20%
	Model 3	7.00%	93.0%	0.0%	3.00%	75.0%	22%
Scenario 5	Model 1	48.00%	25.0%	27.0%	57.00%	32.0%	11%
	Model 2	39.00%	27.0%	34.0%	46.00%	35.0%	19%
	Model 3	52.00%	24.0%	24.0%	62.00%	31.0%	7%
Scenario 6	Model 1	2.00%	98.0%	0.0%	14.00%	86.0%	0%
	Model 2	5.00%	95.0%	0.0%	8.00%	92.0%	0%
	Model 3	8.00%	92.0%	0.0%	19.00%	81.0%	0%

Table S2: Estimate of all non-zero gene-gene correlations in $\widehat{\boldsymbol{\Sigma}}_U$ in real data analysis

Gene, Gene	Estimate
TNFSF13, CXCL5	-0.037
TNFSF13, CCL23	0.203
CXCL14, PDGFB	-0.004
CXCL9, CCL2	-0.026
EGF, CCL7	0.001
EGF, TNF	-0.025
CXCL5, IFNA2	-0.063
CCL11, CCL7	0.057
FGF2, IL15	-0.009
FGF2, PDGFA	-0.018
FGF2, PDGFB	-0.015
CX3CL1, CXCL1	-0.021
IFNA2, IL13	0.021
IL13, IL6	-0.080
IL13, PDGFB	0.076
IL16, PDGFA	-0.106
IL6, CCL23	0.026
IL6, PDGFB	-0.024
CCL3, CCL4	0.097

Table S3: Estimate of all non-zero gene-SNP correlations in $\widehat{\mathbf{\Sigma}}_U$ in real data analysis

Gene, SNP	Estimate
TNFSF13, rs4227	-0.006
CXCL14, rs2112186	-0.041
EGF, rs2081466	-0.060
CXCL5, rs10002688	0.035
CXCL5, rs13139174	0.075
CCL11, rs280045	0.088
CCL11, rs8070999	0.003
CX3CL1, rs9935360	0.003
CX3CL1, rs1466133	-0.005
IFNG, rs11176892	-0.030
IL12B, rs11952950	0.031
IL12B, rs7734683	0.043
IL15, rs12331218	-0.012
IL15, rs1425520	0.041
IL16, rs7178382	-0.001
IL16, rs6495518	0.005
IL16, rs4778906	-0.071
CCL22, rs11076198	-0.003
TGFA, rs1871241	-0.002