SUPPLEMENTARY MATERIAL
The minimum covariance determinant estimator for interval-valued data

Wan Tian[1] and Zhongfeng Qin[1,2*]

[1]School of Economics and Management, Beihang University, Beijing, 100191, China.
[2]Key Laboratory of Complex System Analysis, Management and Decision (Beihang University), Beijing, 100191, China.

*Corresponding author(s). E-mail(s): qin@buaa.edu.cn;
Contributing authors: wantian61@foxmail.com;

1 The partial results in Subsection 4.1

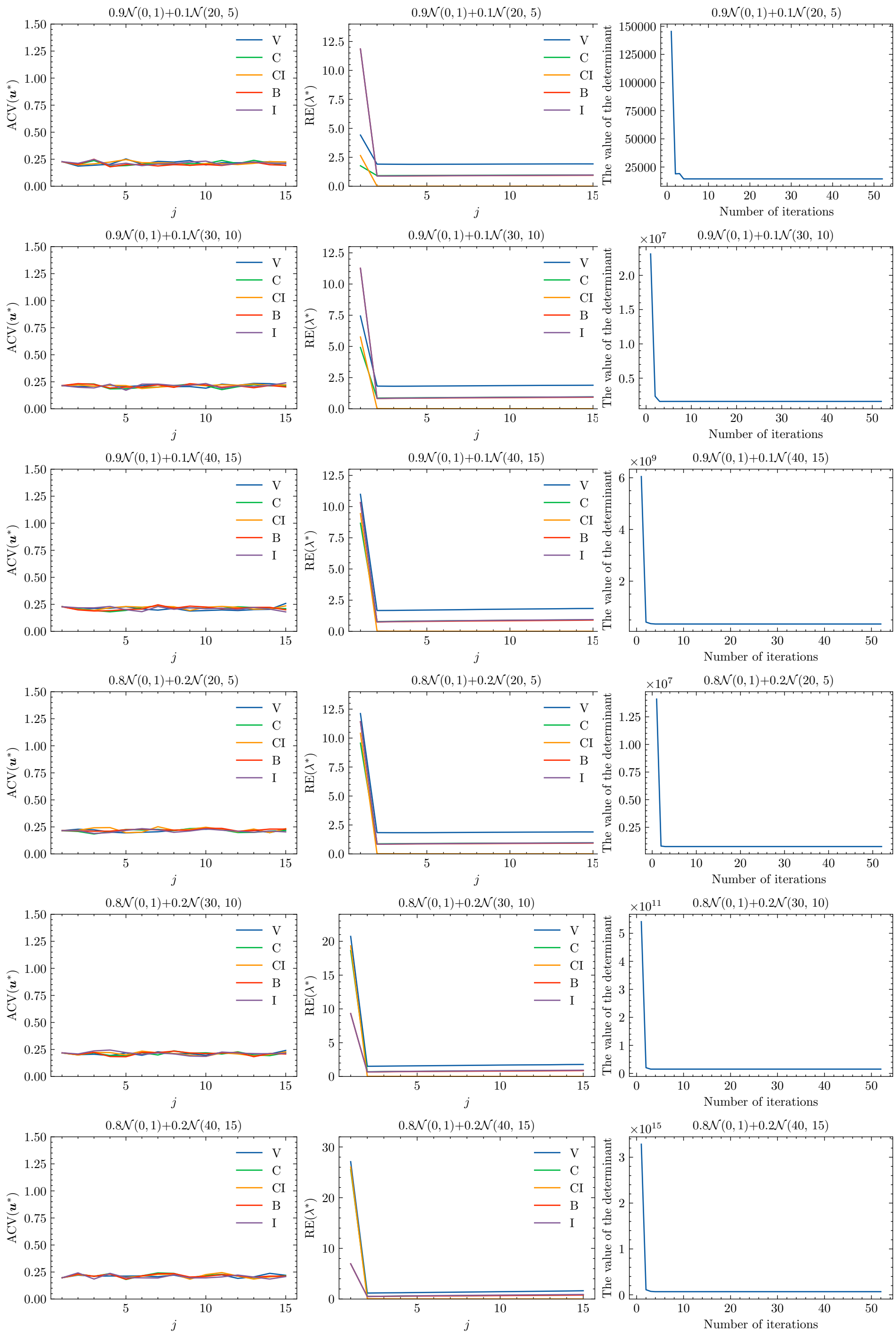

Fig. 1 Under configuration C_6, each row corresponds to different values of ω, δ_1, and ϕ_1. The first two columns represent the results of two metrics, with the horizontal coordinates indicating dimensions. The last column describes the variation of the determinant of the IMCD estimator with respect to the number of iterations. Here, V, C, CI, B, and I stand for VPCA, CPCA, CIPCA, BPCA, and IMCD, respectively.

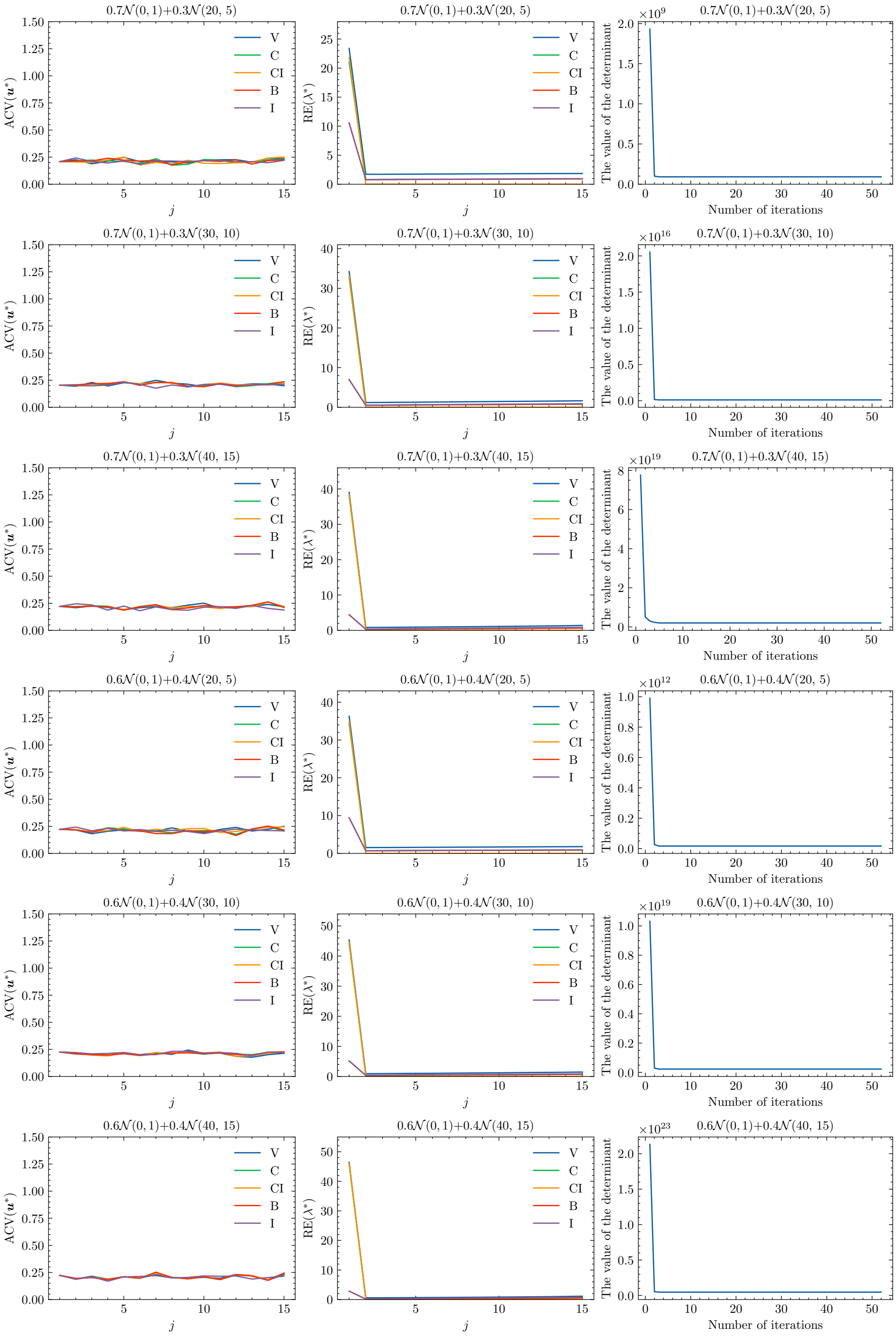

Fig. 2 Continuation of Figure 1.

2 Proofs for results

Leamma 2.1. *The covariance matrix estimated based on (5) and (6), that is,* $\widehat{\boldsymbol{\Sigma}} = (\widehat{\sigma}_{jk})_{1\le j,k\le p}$ *, is non-negative definite almost surely (a.s.), that is,*

$$\widehat{\boldsymbol{\Sigma}} \succeq 0, \ \ a.s..$$

Proof. First, for the estimated variance of each interval-valued variable, we have

$$\begin{aligned}
\widehat{\sigma}_{jj} &= \frac{1}{3n}\sum_{i=1}^{n}\left((y_{ij}^l)^2 + y_{ij}^l y_{ij}^u + (y_{ij}^u)^2\right) - \widehat{\mu}_j^2 \\
&= \frac{1}{6n}\sum_{i=1}^{n}\left(2(y_{ij}^l)^2 + 2y_{ij}^l y_{ij}^u + 2(y_{ij}^u)^2 - 6\widehat{\mu}_j^2\right) \\
&= \frac{1}{6n}\sum_{i=1}^{n}\left(2(y_{ij}^l - \widehat{\mu}_j)^2 + 2(y_{ij}^l - \widehat{\mu}_j)(y_{ij}^u - \widehat{\mu}_j) + 2(y_{ij}^u - \widehat{\mu}_j)^2\right) + A_{ij},
\end{aligned}$$

where $A_{ij} = \left(4y_{ij}^l\widehat{\mu}_j - 6\widehat{\mu}_j^2 + 4y_{ij}^u\widehat{\mu}_j + 2\widehat{\mu}_j(y_{ij}^l + y_{ij}^u)\right) - 6\widehat{\mu}_j^2$. Summing over the A_{ij} yields

$$\begin{aligned}
\sum_{i=1}^{n} A_{ij} &= \sum_{i=1}^{n}\left(4y_{ij}^l\widehat{\mu}_j - 6\widehat{\mu}_j^2 + 4y_{ij}^u\widehat{\mu}_j + 2\widehat{\mu}_j(y_{ij}^l + y_{ij}^u)\right) - 6\widehat{\mu}_j^2 \\
&= \sum_{i=1}^{n} 6\widehat{\mu}_j(y_{ij}^l + y_{ij}^u) - 12\widehat{\mu}_j^2 \\
&= \sum_{i=1}^{n} 12\widehat{\mu}_j\left(\frac{y_{ij}^l + y_{ij}^u}{2}\right) - 12\widehat{\mu}_j^2 \\
&= 0.
\end{aligned}$$

Denote the estimated covariance matrix $\widehat{\boldsymbol{\Sigma}}$ as a sum of four matrices as follows,

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}_1 + \widehat{\boldsymbol{\Sigma}}_2 + \widehat{\boldsymbol{\Sigma}}_3 + \widehat{\boldsymbol{\Sigma}}_4,$$

where

$$\widehat{\boldsymbol{\Sigma}}_1 = \left(\frac{1}{6n}\sum_{i=1}^{n} 2(y_{ij}^l - \widehat{\mu}_j)(y_{ik}^l - \widehat{\mu}_k)\right)_{1\le j,k\le p}, \ \widehat{\boldsymbol{\Sigma}}_2 = \left(\frac{1}{6n}\sum_{i=1}^{n}(y_{ij}^l - \widehat{\mu}_j)(y_{ik}^u - \widehat{\mu}_k)\right)_{1\le j,k\le p},$$

$$\widehat{\boldsymbol{\Sigma}}_3 = \left(\frac{1}{6n}\sum_{i=1}^{n}(y_{ij}^u - \widehat{\mu}_j)(y_{ik}^l - \widehat{\mu}_k)\right)_{1\le j,k\le p}, \ \widehat{\boldsymbol{\Sigma}}_4 = \left(\frac{1}{6n}\sum_{i=1}^{n} 2(y_{ij}^u - \widehat{\mu}_j)(y_{ik}^u - \widehat{\mu}_k)\right)_{1\le j,k\le p}.$$

Let $\boldsymbol{z}_i^l = (y_{i1}^l - \widehat{\mu}_1, y_{i2}^l - \widehat{\mu}_2, \cdots, y_{ip}^l - \widehat{\mu}_p)^\top, \boldsymbol{z}_i^u = (y_{i1}^u - \widehat{\mu}_1, y_{i2}^u - \widehat{\mu}_2, \cdots, y_{ip}^u - \widehat{\mu}_p)^\top, i = 1, 2, \cdots, n$. Then, we have

$$\widehat{\boldsymbol{\Sigma}}_1 = \frac{1}{6n}\sum_{i=1}^n 2\boldsymbol{z}_i^l(\boldsymbol{z}_i^l)^\top, \quad \widehat{\boldsymbol{\Sigma}}_2 = \frac{1}{6n}\sum_{i=1}^n \boldsymbol{z}_i^l(\boldsymbol{z}_i^u)^\top,$$
$$\widehat{\boldsymbol{\Sigma}}_3 = \frac{1}{6n}\sum_{i=1}^n \boldsymbol{z}_i^u(\boldsymbol{z}_i^l)^\top, \quad \widehat{\boldsymbol{\Sigma}}_4 = \frac{1}{6n}\sum_{i=1}^n 2\boldsymbol{z}_i^u(\boldsymbol{z}_i^u)^\top.$$

For a nonzero vector $\boldsymbol{\beta} \in \mathbb{R}^p$, we have

$$\begin{aligned}
\boldsymbol{\beta}^\top \widehat{\boldsymbol{\Sigma}} \boldsymbol{\beta} &= \boldsymbol{\beta}^\top (\widehat{\boldsymbol{\Sigma}}_1 + \widehat{\boldsymbol{\Sigma}}_2 + \widehat{\boldsymbol{\Sigma}}_3 + \widehat{\boldsymbol{\Sigma}}_4)\boldsymbol{\beta} \\
&= \frac{1}{6n}\sum_{i=1}^n \boldsymbol{\beta}^\top \left(2\boldsymbol{z}_i^l(\boldsymbol{z}_i^l)^\top + \boldsymbol{z}_i^l(\boldsymbol{z}_i^u)^\top + \boldsymbol{z}_i^u(\boldsymbol{z}_i^l)^\top + \boldsymbol{z}_i^u(\boldsymbol{z}_i^u)^\top\right)\boldsymbol{\beta} \\
&= \frac{1}{6n}\sum_{i=1}^n 2\left((\boldsymbol{\beta}^\top \boldsymbol{z}_i^l)^2 + (\boldsymbol{\beta}^\top \boldsymbol{z}_i^l)(\boldsymbol{\beta}^\top \boldsymbol{z}_i^u) + (\boldsymbol{\beta}^\top \boldsymbol{z}_i^u)^2\right) \\
&\geq \frac{1}{6n}\sum_{i=1}^n 2\left((\boldsymbol{\beta}^\top \boldsymbol{z}_i^l)^2 - 2|\boldsymbol{\beta}^\top \boldsymbol{z}_i^l||\boldsymbol{\beta}^\top \boldsymbol{z}_i^u| + (\boldsymbol{\beta}^\top \boldsymbol{z}_i^u)^2\right) \\
&= \frac{1}{6n}\sum_{i=1}^n 2\left(|\boldsymbol{\beta}^\top \boldsymbol{z}_i^l| - |\boldsymbol{\beta}^\top \boldsymbol{z}_i^u|\right)^2 \\
&\geq 0, \ \ a.s.,
\end{aligned}$$

which completes the proof. □

Theorem 2.1. *For any $H_1 \subset \{1, 2, \cdots, n\}$ with $|H_1| = h$, if $\det(\widehat{\boldsymbol{\Sigma}}(H_1)) \neq 0$, we select H_2 such that $\{d_1^2(i), i \in H_2\} = \{(d_1^2)_{1:n}, (d_1^2)_{2:n}, \cdots, (d_1^2)_{h:n}\}$, where $(d_1^2)_{1:n} \leq (d_1^2)_{2:n} \leq \cdots \leq (d_1^2)_{n:n}$ represent the ordered distances. In this case, $\det(\widehat{\boldsymbol{\Sigma}}(H_2)) \leq \det(\widehat{\boldsymbol{\Sigma}}(H_1))$, and equality holds only if $\widehat{\boldsymbol{\mu}}(H_1) = \widehat{\boldsymbol{\mu}}(H_2)$ and $\widehat{\boldsymbol{\Sigma}}(H_1) = \widehat{\boldsymbol{\Sigma}}(H_2)$.*

Proof. Firstly, it holds trivially if $\det(\widehat{\boldsymbol{\Sigma}}(H_2)) = 0$. We consider the case of $\det(\widehat{\boldsymbol{\Sigma}}(H_2)) > 0$ and compute the distances $d_2^2(i), i = 1, 2, \cdots, n$ based on $\widehat{\boldsymbol{\mu}}(H_2)$ and

$\widehat{\mathbf{\Sigma}}(H_2)$. Then we have

$$
\begin{aligned}
\frac{2}{3hp}\sum_{i\in H_2} d_2^2(i) &= \frac{2}{3hp}\sum_{i\in H_2}(\boldsymbol{y}_i^c-\widehat{\boldsymbol{\mu}}(H_2))^\top\widehat{\mathbf{\Sigma}}(H_2)^{-1}(\boldsymbol{y}_i^c-\widehat{\boldsymbol{\mu}}(H_2))\\
&= \frac{2}{3hp}\operatorname{Tr}\left(\sum_{i\in H_2}(\boldsymbol{y}_i^c-\widehat{\boldsymbol{\mu}}(H_2))^\top\widehat{\mathbf{\Sigma}}(H_2)^{-1}(\boldsymbol{y}_i^c-\widehat{\boldsymbol{\mu}}(H_2))\right)\\
&= \frac{2}{3hp}\operatorname{Tr}\left(\sum_{i\in H_2}\widehat{\mathbf{\Sigma}}(H_2)^{-1}(\boldsymbol{y}_i^c-\widehat{\boldsymbol{\mu}}(H_2))(\boldsymbol{y}_i^c-\widehat{\boldsymbol{\mu}}(H_2))^\top\right)\\
&= \frac{2}{3hp}\operatorname{Tr}\left(\widehat{\mathbf{\Sigma}}(H_2)^{-1}\sum_{i\in H_2}(\boldsymbol{y}_i^c-\widehat{\boldsymbol{\mu}}(H_2))(\boldsymbol{y}_i^c-\widehat{\boldsymbol{\mu}}(H_2))^\top\right)\\
&= \frac{2}{3hp}\operatorname{Tr}\left(\widehat{\mathbf{\Sigma}}(H_2)^{-1}\sum_{i\in H_2}((\boldsymbol{y}_i^l+\boldsymbol{y}_i^u)/2-\widehat{\boldsymbol{\mu}}(H_2))((\boldsymbol{y}_i^l+\boldsymbol{y}_i^u)/2-\widehat{\boldsymbol{\mu}}(H_2))^\top\right)\\
&= \frac{1}{6hp}\operatorname{Tr}\left(\widehat{\mathbf{\Sigma}}(H_2)^{-1}\sum_{i\in H_2}(\boldsymbol{y}_i^l-\widehat{\boldsymbol{\mu}}(H_2)+\boldsymbol{y}_i^u-\widehat{\boldsymbol{\mu}}(H_2))(\boldsymbol{y}_i^l-\widehat{\boldsymbol{\mu}}(H_2)+\boldsymbol{y}_i^u-\widehat{\boldsymbol{\mu}}(H_2))^\top\right),
\end{aligned}
\tag{1}
$$

with the notations in the proof of Lemma 2.1 , we can obtain

$$
\begin{aligned}
\frac{2}{3hp}\sum_{i\in H_2} d_2^2(i) &= \frac{1}{6hp}\operatorname{Tr}\left(\widehat{\mathbf{\Sigma}}(H_2)^{-1}\sum_{i\in H_2}\boldsymbol{z}_i^l(\boldsymbol{z}_i^l)^\top+\boldsymbol{z}_i^l(\boldsymbol{z}_i^u)^\top+\boldsymbol{z}_i^u(\boldsymbol{z}_i^l)^\top+\boldsymbol{z}_i^u(\boldsymbol{z}_i^u)^\top\right)\\
&= \frac{1}{p}\operatorname{Tr}\left(\widehat{\mathbf{\Sigma}}(H_2)^{-1}\frac{1}{6h}\sum_{i\in H_2}\boldsymbol{z}_i^l(\boldsymbol{z}_i^l)^\top+\boldsymbol{z}_i^l(\boldsymbol{z}_i^u)^\top+\boldsymbol{z}_i^u(\boldsymbol{z}_i^l)^\top+\boldsymbol{z}_i^u(\boldsymbol{z}_i^u)^\top\right)\\
&= \frac{1}{p}\operatorname{Tr}\left(\widehat{\mathbf{\Sigma}}(H_2)^{-1}\left(\widehat{\mathbf{\Sigma}}_1(H_2)+\widehat{\mathbf{\Sigma}}_2(H_2)+\widehat{\mathbf{\Sigma}}_3(H_2)+\widehat{\mathbf{\Sigma}}_4(H_2)\right)\right)\\
&= \frac{1}{p}\operatorname{Tr}\left(\left(\widehat{\mathbf{\Sigma}}_1(H_2)+\widehat{\mathbf{\Sigma}}_2(H_2)+\widehat{\mathbf{\Sigma}}_3(H_2)+\widehat{\mathbf{\Sigma}}_4(H_2)\right)^{-1}\right.\\
&\qquad\qquad \left.\times\left(\widehat{\mathbf{\Sigma}}_1(H_2)+\widehat{\mathbf{\Sigma}}_2(H_2)+\widehat{\mathbf{\Sigma}}_3(H_2)+\widehat{\mathbf{\Sigma}}_4(H_2)\right)\right)\\
&= \frac{1}{p}\operatorname{Tr}(\boldsymbol{I}_{p\times p})\\
&= 1,
\end{aligned}
\tag{2}
$$

where $\operatorname{Tr}(\cdot)$ is the trace operation of the matrix. In the observation selection procedure of the fast-MCD algorithm, $\frac{2}{3hp}\sum_{i\in H_k} d_k^2(i)=1$ is fixed for any k. Moreover, we have

$$
\Delta \coloneqq \frac{2}{3hp}\sum_{i\in H_2} d_1^2(i) = \frac{2}{3hp}\sum_{i=1}^{h}(d_1^2)_{i:n} \le \frac{2}{3hp}\sum_{i\in H_1} d_1^2(i) = 1, \tag{3}
$$

where both the equality and inequality of the above equation hold due to the subset H_2 being chosen based on $\widehat{\boldsymbol{\mu}}(H_1)$ and $\widehat{\boldsymbol{\Sigma}}(H_1)$. On the other hand

$$\begin{aligned}
\Delta &:= \frac{2}{3hp}\sum_{i\in H_2} d_1^2(i) = \frac{2}{3hp}\sum_{i\in H_2}(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_1))^\top \widehat{\boldsymbol{\Sigma}}(H_1)^{-1}(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_1)) \\
&= \frac{2}{3hp}\sum_{i\in H_2}(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2) + \widehat{\boldsymbol{\mu}}(H_2) - \widehat{\boldsymbol{\mu}}(H_1))^\top \widehat{\boldsymbol{\Sigma}}(H_1)^{-1}(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2) + \widehat{\boldsymbol{\mu}}(H_2) - \widehat{\boldsymbol{\mu}}(H_1)) \\
&= \frac{4}{p}\operatorname{Tr}\left(\widehat{\boldsymbol{\Sigma}}(H_1)^{-1}\widehat{\boldsymbol{\Sigma}}(H_2)\right) + \frac{2}{3p}(\widehat{\boldsymbol{\mu}}(H_2) - \widehat{\boldsymbol{\mu}}(H_1))^\top \widehat{\boldsymbol{\Sigma}}(H_1)^{-1}(\widehat{\boldsymbol{\mu}}(H_2) - \widehat{\boldsymbol{\mu}}(H_1)) \\
&> 0,
\end{aligned}$$

where the last inequality holds because both $\widehat{\boldsymbol{\Sigma}}(H_1)$ and $\widehat{\boldsymbol{\Sigma}}(H_2)$ are positive definite and symmetric. Considering the distances $d^2_{(\widehat{\mu}(H_1),\Delta\widehat{\boldsymbol{\Sigma}}(H_1))}(i)$ based on $\widehat{\mu}(H_1)$ and $\widehat{\boldsymbol{\Sigma}}(H_1)$, it follows from (2) and (3) that

$$\begin{aligned}
\frac{2}{3hp}\sum_{i\in H_2} d^2_{(\widehat{\mu}(H_1),\Delta\widehat{\boldsymbol{\Sigma}}(H_1))}(i) &= \frac{2}{3hp}\sum_{i\in H_2}(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_1))^\top \frac{1}{\Delta}\widehat{\boldsymbol{\Sigma}}(H_1)^{-1}(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_1)) \\
&= \frac{2}{3\Delta hp}\sum_{i\in H_2} d_1^2(i) \\
&= 1.
\end{aligned}$$

The Lemma 2.1 suggests that $(\widehat{\boldsymbol{\mu}}(H_2), \widehat{\boldsymbol{\Sigma}}(H_2))$ is the unique minimizer of $\det(\boldsymbol{\Lambda})$ among all $(\boldsymbol{\gamma}, \boldsymbol{\Lambda})$ that satisfy $\frac{2}{3hp}\sum_{i\in H_2} d^2_{(\boldsymbol{\gamma},\boldsymbol{\Lambda})}(i) = 1$. Thus, we have $\det(\widehat{\boldsymbol{\Sigma}}(H_2)) \leq \det(\Delta\widehat{\boldsymbol{\Sigma}}(H_1))$. Furthermore, from the inequality in (3), we have

$$\det(\widehat{\boldsymbol{\Sigma}}(H_2)) \leq \det(\Delta\widehat{\boldsymbol{\Sigma}}(H_1)) \leq \det(\widehat{\boldsymbol{\Sigma}}(H_1)). \tag{4}$$

As it follows from (11), $\det(\widehat{\boldsymbol{\Sigma}}(H_1)) = \det(\widehat{\boldsymbol{\Sigma}}(H_2))$ if and only if both inequalities hold. For the first, Lemma 2.1 suggests that $\det(\widehat{\boldsymbol{\Sigma}}(H_2)) = \det(\Delta\widehat{\boldsymbol{\Sigma}}(H_1))$ if and only if $(\widehat{\mu}(H_2), \widehat{\boldsymbol{\Sigma}}(H_2)) = (\widehat{\mu}(H_1), \Delta\widehat{\boldsymbol{\Sigma}}(H_1))$. For the second, $\det(\Delta\widehat{\boldsymbol{\Sigma}}(H_1)) = \det(\widehat{\boldsymbol{\Sigma}}(H_1))$ implies that $\Delta = 1$. Combining both yields $(\widehat{\mu}(H_2), \widehat{\boldsymbol{\Sigma}}(H_2)) = (\widehat{\mu}(H_1), \widehat{\boldsymbol{\Sigma}}(H_1))$.

□

Lemma 2.1. *The optimization problem*

$$\begin{aligned}
&\min_{\boldsymbol{\gamma}\in\mathbb{R}^p,\boldsymbol{\Lambda}\in\mathcal{S}^p_{++}} \det(\boldsymbol{\Lambda}), \\
&\text{subject to } \frac{2}{3hp}\sum_{i\in H_2} d^2_{(\boldsymbol{\gamma},\boldsymbol{\Lambda})}(i) = 1
\end{aligned}$$

has a unique minimizer $(\widehat{\boldsymbol{\mu}}(H_2), \widehat{\boldsymbol{\Sigma}}(H_2))$, where $\mathcal{S}^p_{++}$ denotes the set of symmetric positive definite $p \times p$ matrices.

Proof. From the proof of Theorem 2.1, it is clear that $(\widehat{\boldsymbol{\mu}}(H_2), \widehat{\boldsymbol{\Sigma}}(H_2))$ satisfies the constraint of the optimization problem. In the following, we prove that $(\widehat{\boldsymbol{\mu}}(H_2), \widehat{\boldsymbol{\Sigma}}(H_2))$ is a unique minimizer. Without loss of generality, we assume that $\widehat{\boldsymbol{\mu}}(H_2) = \mathbf{0}$ and $\widehat{\boldsymbol{\Sigma}}(H_2) = \boldsymbol{I}_{p\times p}$. Given $\boldsymbol{\Lambda} \in \mathcal{S}^p_{++}$, then there exist orthogonal matrix $\boldsymbol{U}$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ such that

$$\boldsymbol{\Lambda} = \boldsymbol{U}\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_p)\boldsymbol{U}^\top$$

and $\det(\boldsymbol{\Lambda}) = \lambda_1\lambda_2\cdots\lambda_p, \boldsymbol{\Lambda}^{-1} = \boldsymbol{U}\mathrm{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \cdots, \lambda_p^{-1})\boldsymbol{U}^\top$. Assume that $(\boldsymbol{\gamma}, \boldsymbol{\Lambda})$ satisfies the constraint in the optimization problem and let $\boldsymbol{w}_i^c = (w_{i1}^c, w_{i2}^c, \cdots, w_{ip}^c)^\top = \boldsymbol{U}^\top(\boldsymbol{y}_i^c - \boldsymbol{\gamma}), i \in H_2$. First, we have

$$\frac{1}{h}\sum_{i\in H_2}\boldsymbol{w}_i^c = \frac{1}{h}\sum_{i\in H_2}\boldsymbol{U}^\top(\boldsymbol{y}_i^c - \boldsymbol{\gamma}) = \boldsymbol{U}^\top\widehat{\boldsymbol{\mu}}(H_2) - \boldsymbol{U}^\top\boldsymbol{\gamma} = -\boldsymbol{U}^\top\boldsymbol{\gamma},$$

and

$$\begin{aligned}
&\frac{1}{h}\sum_{i\in H_2}\left(\boldsymbol{w}_i^c - \frac{1}{h}\sum_{i\in H_2}\boldsymbol{w}_i^c\right)\left(\boldsymbol{w}_i^c - \frac{1}{h}\sum_{i\in H_2}\boldsymbol{w}_i^c\right)^\top \\
&= \frac{1}{h}\sum_{i\in H_2}\left(\boldsymbol{w}_i^c + \boldsymbol{U}^\top\boldsymbol{\gamma}\right)\left(\boldsymbol{w}_i^c + \boldsymbol{U}^\top\boldsymbol{\gamma}\right)^\top \\
&= \frac{1}{h}\sum_{i\in H_2}\boldsymbol{U}^\top\boldsymbol{y}_i^c\left(\boldsymbol{U}^\top\boldsymbol{y}_i^c\right)^\top = \boldsymbol{U}^\top\left(\frac{1}{h}\sum_{i\in H_2}\boldsymbol{y}_i^c(\boldsymbol{y}_i^c)^\top\right)\boldsymbol{U} \\
&= \frac{3}{2}\boldsymbol{U}^\top\left(\frac{1}{6h}\sum_{i\in H_2}(\boldsymbol{y}_l^c + \boldsymbol{y}_u^c)(\boldsymbol{y}_l^c + \boldsymbol{y}_u^c)^\top\right)\boldsymbol{U} \\
&= \frac{3}{2}\boldsymbol{U}^\top\left(\frac{1}{6h}\sum_{i\in H_2}\boldsymbol{y}_l^c(\boldsymbol{y}_l^c)^\top + \boldsymbol{y}_l^c(\boldsymbol{y}_u^c)^\top + \boldsymbol{y}_u^c(\boldsymbol{y}_l^c)^\top + \boldsymbol{y}_u^c(\boldsymbol{y}_u^c)^\top\right)\boldsymbol{U} \\
&= \frac{3}{2}\boldsymbol{U}^\top\widehat{\boldsymbol{\Sigma}}(H_2)\boldsymbol{U} = \frac{3}{2}\boldsymbol{U}^\top\boldsymbol{I}\boldsymbol{U} = \frac{3}{2}\boldsymbol{I}.
\end{aligned} \tag{5}$$

Then we have

$$
\begin{aligned}
\frac{2}{3hp}\sum_{i\in H_2} d^2_{(\boldsymbol{\gamma},\boldsymbol{\Lambda})}(i) &= \frac{2}{3hp}\sum_{i\in H_2}(\boldsymbol{y}_i^c-\boldsymbol{\gamma})^\top\boldsymbol{\Lambda}^{-1}(\boldsymbol{y}_i^c-\boldsymbol{\gamma}) \\
&= \frac{2}{3hp}\sum_{i\in H_2}(\boldsymbol{y}_i^c-\boldsymbol{\gamma})^\top\boldsymbol{U}\mathrm{diag}(\lambda_1^{-1},\lambda_2^{-1},\cdots,\lambda_p^{-1})\boldsymbol{U}^\top(\boldsymbol{y}_i^c-\boldsymbol{\gamma}) \\
&= \frac{2}{3hp}\sum_{i\in H_2}\left(\boldsymbol{U}^\top(\boldsymbol{y}_i^c-\boldsymbol{\gamma})\right)^\top\mathrm{diag}(\lambda_1^{-1},\lambda_2^{-1},\cdots,\lambda_p^{-1})\left(\boldsymbol{U}^\top(\boldsymbol{y}_i^c-\boldsymbol{\gamma})\right) \\
&= \frac{2}{3hp}\sum_{i\in H_2}\left(\boldsymbol{w}_i^c+\boldsymbol{U}^\top\boldsymbol{\gamma}-\boldsymbol{U}^\top\boldsymbol{\gamma}\right)^\top\mathrm{diag}(\lambda_1^{-1},\lambda_2^{-1},\cdots,\lambda_p^{-1})\left(\boldsymbol{w}_i^c+\boldsymbol{U}^\top\boldsymbol{\gamma}-\boldsymbol{U}^\top\boldsymbol{\gamma}\right)^\top \\
&= \frac{2}{3hp}\sum_{i\in H_2}(\boldsymbol{w}_i^c+\boldsymbol{U}^\top\boldsymbol{\gamma})^\top\mathrm{diag}(\lambda_1^{-1},\lambda_2^{-1},\cdots,\lambda_p^{-1})(\boldsymbol{w}_i^c+\boldsymbol{U}^\top\boldsymbol{\gamma}) \\
&- 2(\boldsymbol{U}^\top\boldsymbol{\gamma})^\top\mathrm{diag}(\lambda_1^{-1},\lambda_2^{-1},\cdots,\lambda_p^{-1})(\boldsymbol{w}_i^c+\boldsymbol{U}^\top\boldsymbol{\gamma}) \\
&+ (\boldsymbol{U}^\top\boldsymbol{\gamma})^\top\mathrm{diag}(\lambda_1^{-1},\lambda_2^{-1},\cdots,\lambda_p^{-1})(\boldsymbol{U}^\top\boldsymbol{\gamma}) \\
&= \frac{1}{p}\sum_{j=1}^{p}\frac{1}{\lambda_j}+\frac{2}{3hp}\sum_{j=1}^{p}\frac{(\boldsymbol{U}^\top\boldsymbol{\gamma})^\top(\boldsymbol{U}^\top\boldsymbol{\gamma})}{\lambda_j},
\end{aligned}
\tag{6}
$$

where the last equality is due to (5), and i_j denotes the index of the observation in H_2 corresponding to the jth eigenvalue. Eq.(6) implies $\sum_{j=1}^{p}\lambda_j^{-1} < p$ if $\boldsymbol{\gamma}\neq\mathbf{0}$. Using geometric and arithmetic inequalities, we have

$$
\det(\boldsymbol{\Lambda}) = \lambda_1\lambda_2\cdots\lambda_p \geq \left(\frac{p}{\lambda_1^{-1}+\lambda_2^{-1}+\cdots+\lambda_p^{-1}}\right)^p > 1.
$$

Thus any solution $(\boldsymbol{\gamma},\boldsymbol{\Lambda})$ with $\boldsymbol{\gamma}\neq 0$ does not minimize $\det(\boldsymbol{\Lambda})$. If $\boldsymbol{\gamma}=\mathbf{0}$, then $\sum_{j=1}^{p}\lambda_j^{-1}=p$, which implies that $\lambda_1=\lambda_2=\cdots=\lambda_p=0$ and therefore $(\mathbf{0},\boldsymbol{I}_{p\times p})$ is the unique minimizer. □

Theorem 2.1. *For any interval-valued dataset $\boldsymbol{Y}\subset\mathbb{R}^p$ satisfying $k(\boldsymbol{Y})<h$ with $p>1$,*

$$
\epsilon_n(\widehat{\boldsymbol{\mu}}_{\mathrm{IMCD}},\boldsymbol{Y}) = \epsilon_n(\widehat{\boldsymbol{\Sigma}}_{\mathrm{IMCD}},\boldsymbol{Y}) = \frac{\min(n-h+1,h-k(\boldsymbol{Y}))}{n}.
$$

Proof. Following the argument of Agulló et al. [1], which concerns the least-trimmed squares estimation of the multivariate regression, we first prove that $\epsilon_n(\widehat{\boldsymbol{\Sigma}}_{\mathrm{IMCD}},\boldsymbol{Y})\geq\min(n-h+1,h-k(\boldsymbol{Y}))/n$. We show that there exists a value C_1, which only depends on $\boldsymbol{Y}$, such that for every $\boldsymbol{Y}'=(\boldsymbol{y}_1',\boldsymbol{y}_2',\cdots,\boldsymbol{y}_n')^\top$ obtained by replacing at most $m=\min(n-h+1,h-k(\boldsymbol{Y}))-1$ observations from $\boldsymbol{Y}$ we have that $\widehat{\boldsymbol{\Sigma}}'_{\mathrm{IMCD}}\leq C_1$, where $\widehat{\boldsymbol{\Sigma}}'_{\mathrm{IMCD}}$ is the IMCD estimator based on $\boldsymbol{Y}'$. The matrix norm we use here is $\|\boldsymbol{A}\|=\sup_{\|\boldsymbol{u}\|=1}\|\boldsymbol{A}\boldsymbol{u}\|$ where $\boldsymbol{u}\in\mathbb{R}^q$ and $\boldsymbol{A}\in\mathbb{R}^{p\times q}$. The ℓ_2-norm of the matrix is $\|\boldsymbol{A}\|_2=\left(\sum_{i,j}|\boldsymbol{A}(i,j)|^2\right)^{1/2}$. One should note that all matrix norms are topologically equivalent on $\mathbb{R}^{p\times q}$, that is, there exist constants $C_2, C_3 > 0$ such that $C_2\|\boldsymbol{A}\| \leq \|\boldsymbol{A}\|_2 \leq C_3\|\boldsymbol{A}\|$ for all $\boldsymbol{A} \in \mathbb{R}^{p\times q}$.

Let $\mathcal{J} = \{J \subset \{1, 2, \cdots, n\} : |J| = k(\boldsymbol{Y}) + 1\}$ be the collection of all subsets of size $k(\boldsymbol{Y}) + 1$, we have that

$$C_4 = \inf_{J\in\mathcal{J}} \lambda_p(\widehat{\boldsymbol{\Sigma}}(J)) > 0,$$

where $\widehat{\boldsymbol{\Sigma}}(J)$ is the covariance matrix estimated based on the observations with indices set J and definite

$$C_5 = \sup_{H\in\mathcal{H}} \|\widehat{\boldsymbol{\Sigma}}(H)\| < \infty.$$

Let H_2 be the optimal subset of size h corresponding to $\widehat{\boldsymbol{\Sigma}}'_{\mathrm{IMCD}}$ such that $\widehat{\boldsymbol{\Sigma}}'_{\mathrm{IMCD}} = \widehat{\boldsymbol{\Sigma}}'(H_2)$ where $\widehat{\boldsymbol{\Sigma}}'(H_2)$ denotes the empirical covariance matrix estimated based on the h observations referenced to H_2 from $\boldsymbol{Y}'$. Define C_1 as

$$C_1 = C_5^p \left(\frac{h}{C_4(k(\boldsymbol{Y}) + 1)}\right)^{p-1},$$

and suppose $\|\widehat{\boldsymbol{\Sigma}}'(H_2)\| > C_1$. First of all, there exists a subset $H_1 \in \mathcal{H}$ containing indices only corresponding to data points of data $\boldsymbol{Y}$. Using Lemma 5.1 in Lopuhaa and Rousseeuw [2] and the properties of matrix norms it follows that

$$\det(\widehat{\boldsymbol{\Sigma}}(H_1)) = \prod_{i=1}^{p} \lambda_i(\widehat{\boldsymbol{\Sigma}}(H_1)) \leq \lambda_1(\widehat{\boldsymbol{\Sigma}}(H_1))^p \leq C_5^p. \tag{7}$$

Since $m = \min(n - h + 1, h - k(\boldsymbol{Y})) - 1 \leq h - k(\boldsymbol{Y}) - 1$, the set H_2 contains a subset J_0 of size $k(\boldsymbol{Y}) + 1$ corresponding to the original observations of $\boldsymbol{Y}$. Using Lemma 5.1 in Lopuhaa and Rousseeuw [2] we have

$$\begin{aligned}
\lambda_p(\widehat{\boldsymbol{\Sigma}}'(H_2)) &= \lambda_p\left(\frac{1}{6h}\sum_{i\in H_2}\left(2\boldsymbol{z}_i^{l\prime}(\boldsymbol{z}_i^{l\prime})^\top + \boldsymbol{z}_i^{l\prime}(\boldsymbol{z}_i^{u\prime})^\top + \boldsymbol{z}_i^{u\prime}(\boldsymbol{z}_i^{l\prime})^\top + \boldsymbol{z}_i^{u\prime}(\boldsymbol{z}_i^{u\prime})^\top\right)\right) \\
&= \lambda_p\left(\frac{k(\boldsymbol{Y})+1}{h}\frac{1}{6(k(\boldsymbol{Y})+1)}\sum_{i\in H_2}\left(2\boldsymbol{z}_i^{l\prime}(\boldsymbol{z}_i^{l\prime})^\top + \boldsymbol{z}_i^{l\prime}(\boldsymbol{z}_i^{u\prime})^\top + \boldsymbol{z}_i^{u\prime}(\boldsymbol{z}_i^{l\prime})^\top + \boldsymbol{z}_i^{u\prime}(\boldsymbol{z}_i^{u\prime})^\top\right)\right) \\
&\geq \frac{k(\boldsymbol{Y})+1}{h}\lambda_p\left(\frac{1}{6(k(\boldsymbol{Y})+1)}\sum_{i\in J_0}\left(2\boldsymbol{z}_i^{l}(\boldsymbol{z}_i^{l})^\top + \boldsymbol{z}_i^{l}(\boldsymbol{z}_i^{u})^\top + \boldsymbol{z}_i^{u}(\boldsymbol{z}_i^{l})^\top + \boldsymbol{z}_i^{u}(\boldsymbol{z}_i^{u})^\top\right)\right) \\
&= \frac{k(\boldsymbol{Y})+1}{h}\lambda_p(\widehat{\boldsymbol{\Sigma}}(J_0)) \\
&\geq \frac{C_4(k(\boldsymbol{Y})+1)}{h},
\end{aligned}$$

on the other hand,

$$\lambda_1(\widehat{\boldsymbol{\Sigma}}'(H_2)) = \sup_{\|\boldsymbol{u}\|=1} \frac{1}{6h} \sum_{i\in H_2} \boldsymbol{u}^\top \left(2\boldsymbol{z}_i^{l\prime}(\boldsymbol{z}_i^{l\prime})^\top + \boldsymbol{z}_i^{l\prime}(\boldsymbol{z}_i^{u\prime})^\top + \boldsymbol{z}_i^{u\prime}(\boldsymbol{z}_i^{l\prime})^\top + 2\boldsymbol{z}_i^{u\prime}(\boldsymbol{z}_i^{u\prime})^\top\right) \boldsymbol{u} > C_1,$$

thus

$$\det(\widehat{\boldsymbol{\Sigma}}'(H_2)) \geq \lambda_1(\widehat{\boldsymbol{\Sigma}}'(H_2))\lambda_p(\widehat{\boldsymbol{\Sigma}}'(H_2))^{p-1} > C_1 \left(\frac{C_4(k(\boldsymbol{Y})+1)}{h}\right)^{p-1} = C_5^p$$

by the definition of C_1. Together with (7) this implies $\det(\widehat{\boldsymbol{\Sigma}}'(H_2)) > \det(\widehat{\boldsymbol{\Sigma}}(H_1))$ which contradicts the definition of $\widehat{\boldsymbol{\Sigma}}'(H_2)$, so we conclude that $\|\widehat{\boldsymbol{\Sigma}}'(H_2)\| \leq C_1$.

In the following we prove that $\epsilon_n(\widehat{\boldsymbol{\Sigma}}_{\text{IMCD}}, \boldsymbol{Y}) \leq \min(n-h+1, h-k(\boldsymbol{Y}))/n$. We first show that $\epsilon_n(\widehat{\boldsymbol{\Sigma}}_{\text{IMCD}}, \boldsymbol{Y}) \leq (n-h+1)/n$. Actually, if we replace $n-h+1$ observations of the original data $\boldsymbol{Y}$ with arbitrary interval-values, then the optimal subset H_2 of $\boldsymbol{Y}'$ contains at least one outlier, and we know that the least squares will explode even with a single outlier. It then follows that also $\widehat{\boldsymbol{\Sigma}}'_{\text{IMCD}}$ explodes. □

Theorem 3.1. *For any $H_1 \subset \{1, 2, \cdots, n\}$ with $|H_1| = h$, projecting $\widehat{\boldsymbol{\Sigma}}(H_1)$ onto the convex cone $\{\boldsymbol{\Theta} \succeq \nu_1\boldsymbol{I}\}$ as $(\widehat{\boldsymbol{\Sigma}}(H_1))_+$. Then, take H_2 such that $\{d_1^2(i), i \in H_2\} = \{(d_1^2)_{1:n}, (d_1^2)_{2:n}, \cdots, (d_1^2)_{h:n}\}$, where $(d_1^2)_{1:n} \leq (d_1^2)_{2:n} \leq \cdots \leq (d_1^2)_{n:n}$ are the ordered distances. Calculate $\widehat{\boldsymbol{\mu}}(H_2)$ and $\widehat{\boldsymbol{\Sigma}}(H_2)$ and project $\widehat{\boldsymbol{\Sigma}}(H_2)$ onto the convex cone $\{\boldsymbol{\Theta} \succeq \nu_2\boldsymbol{I}\}$ as $(\widehat{\boldsymbol{\Sigma}}(H_2))_+$, then $\det((\widehat{\boldsymbol{\Sigma}}(H_2))_+) \leq \det((\widehat{\boldsymbol{\Sigma}}(H_1))_+)$.*

Proof. It follows from the definition of the projection estimator that $(\widehat{\boldsymbol{\Sigma}}(H_2))_+$ must be positive definite, that is, $\det((\widehat{\boldsymbol{\Sigma}}(H_2))_+) > 0$. Before the proof, we first address the value of $\operatorname{Tr}\left((\widehat{\boldsymbol{\Sigma}}(H_2))_+^{-1}\widehat{\boldsymbol{\Sigma}}(H_2)\right)$. It is assume, without loss of generality, that $\widehat{\boldsymbol{\Sigma}}(H_2)$ has the following spectral decomposition

$$\widehat{\boldsymbol{\Sigma}}(H_2) = \boldsymbol{P}_2 \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_k, \cdots, \lambda_p) \boldsymbol{P}_2^\top,$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the eigenvalues corresponding to the eigenvectors in the orthogonal matrix $\boldsymbol{P}$. By the discussion in Section 3.1, we know that the threshold value $\nu_2 = \max(\lambda_p, \lambda_1/\kappa)$. Without loss of generality, assuming that the last k eigenvalues in the regularized estimator $(\widehat{\boldsymbol{\Sigma}}(H_2))_+$ are replaced by the threshold ν_2, then we have

$$(\widehat{\boldsymbol{\Sigma}}(H_2))_+^{-1} = \boldsymbol{P}_2 \operatorname{diag}(1/\lambda_1, 1/\lambda_2, \cdots, 1/\lambda_{k-1}, 1/\nu_2, \cdots, 1/\nu_2)\boldsymbol{P}_2^\top.$$

In this case,

$$\operatorname{Tr}\left((\widehat{\boldsymbol{\Sigma}}(H_2))_+^{-1}\widehat{\boldsymbol{\Sigma}}(H_2)\right) = \operatorname{Tr}\left(\boldsymbol{P}_2 \operatorname{diag}(1, 1, \cdots, 1, \lambda_k/\nu_2, \cdots, \lambda_p/\nu_2)\boldsymbol{P}_2^\top\right) = k-1+\sum_{i=k}^{p} \lambda_i/\nu_2 := T_2.$$

We now define a modified Mahalanobis Distance based on $\widehat{\boldsymbol{\mu}}(H_2)$ and $(\widehat{\boldsymbol{\Sigma}}(H_2))_+$,

$$D_2^2(i) = (\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2))^\top \left(\frac{T_2}{p}(\widehat{\boldsymbol{\Sigma}}(H_2))_+\right)^{-1} (\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2)), i = 1, 2, \cdots, n.$$

We compute those sorted distances $D_2^2(i), i = 1, 2, 3, \cdots, n$ based on $\widehat{\boldsymbol{\mu}}(H_2)$ and $(\widehat{\boldsymbol{\Sigma}}(H_2))_+$, then we have

$$\begin{aligned}
\frac{2}{3hp}\sum_{i\in H_2} D_2^2(i) &= \frac{2}{3hp}\sum_{i\in H_2} (\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2))^\top \left(\frac{T_2}{p}(\widehat{\boldsymbol{\Sigma}}(H_2))_+\right)^{-1} (\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2)) \\
&= \frac{2}{3hT_2}\operatorname{Tr}\left(\sum_{i\in H_2} (\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2))^\top (\widehat{\boldsymbol{\Sigma}}(H_2))_+^{-1} (\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2))\right) \\
&= \frac{2}{3hT_2}\operatorname{Tr}\left(\sum_{i\in H_2} (\widehat{\boldsymbol{\Sigma}}(H_2))_+^{-1} (\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2))(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2))^\top\right) \\
&= \frac{2}{3hT_2}\operatorname{Tr}\left((\widehat{\boldsymbol{\Sigma}}(H_2))_+^{-1}\sum_{i\in H_2} (\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2))(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_2))^\top\right) \\
&= \frac{2}{3hT_2}\operatorname{Tr}\left((\widehat{\boldsymbol{\Sigma}}(H_2))_+^{-1}\sum_{i\in H_2} ((\boldsymbol{y}_i^l + \boldsymbol{y}_i^u)/2 - \widehat{\boldsymbol{\mu}}(H_2))((\boldsymbol{y}_i^l + \boldsymbol{y}_i^u)/2 - \widehat{\boldsymbol{\mu}}(H_2))^\top\right) \\
&= \frac{1}{6hT_2}\operatorname{Tr}\left((\widehat{\boldsymbol{\Sigma}}(H_2))_+^{-1}\sum_{i\in H_2} (\boldsymbol{y}_i^l - \widehat{\boldsymbol{\mu}}(H_2) + \boldsymbol{y}_i^u - \widehat{\boldsymbol{\mu}}(H_2))(\boldsymbol{y}_i^l - \widehat{\boldsymbol{\mu}}(H_2) + \boldsymbol{y}_i^u - \widehat{\boldsymbol{\mu}}(H_2))^\top\right),
\end{aligned} \tag{8}$$

using the notations in the proof of Lemma 2.1, we have

$$\begin{aligned}
\frac{2}{3hT_2}\sum_{i\in H_2} d_2^2(i) &= \frac{1}{6hT_2}\operatorname{Tr}\left((\widehat{\boldsymbol{\Sigma}}(H_2))_+^{-1}\sum_{i\in H_2} \boldsymbol{z}_i^l(\boldsymbol{z}_i^l)^\top + \boldsymbol{z}_i^l(\boldsymbol{z}_i^u)^\top + \boldsymbol{z}_i^u(\boldsymbol{z}_i^l)^\top + \boldsymbol{z}_i^u(\boldsymbol{z}_i^u)^\top\right) \\
&= \frac{1}{T_2}\operatorname{Tr}\left((\widehat{\boldsymbol{\Sigma}}(H_2))_+^{-1}\frac{1}{6h}\sum_{i\in H_2} \boldsymbol{z}_i^l(\boldsymbol{z}_i^l)^\top + \boldsymbol{z}_i^l(\boldsymbol{z}_i^u)^\top + \boldsymbol{z}_i^u(\boldsymbol{z}_i^l)^\top + \boldsymbol{z}_i^u(\boldsymbol{z}_i^u)^\top\right) \\
&= \frac{1}{T_2}\operatorname{Tr}\left((\widehat{\boldsymbol{\Sigma}}(H_2))_+^{-1}\widehat{\boldsymbol{\Sigma}}(H_2)\right) \\
&= 1.
\end{aligned} \tag{9}$$

Moreover, we have

$$\Delta \coloneqq \frac{2}{3hp}\sum_{i\in H_2} D_1^2(i) = \frac{2}{3hp}\sum_{i=1}^{h}(D_1^2)_{i:n} \le \frac{2}{3hp}\sum_{i\in H_1} D_1^2(i) = 1, \tag{10}$$

where $T_1 = \mathrm{Tr}\left((\widehat{\mathbf{\Sigma}}(H_1))_+^{-1}\widehat{\mathbf{\Sigma}}(H_1)\right)$. Additionally,

$$\Delta \coloneqq \frac{2}{3hp}\sum_{i\in H_2} D_1^2(i) = \frac{2}{3hp}\sum_{i\in H_2}(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_1))^\top \left(\frac{T_1}{p}(\widehat{\mathbf{\Sigma}}(H_1))_+\right)^{-1}(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_1)) > 0,$$

where the last inequality holds because $(\widehat{\mathbf{\Sigma}}(H_1))_+$ is positive definite. Considering the distances $D^2_{(\widehat{\mu}(H_1),\Delta\frac{T_1}{p}(\widehat{\mathbf{\Sigma}}(H_1))_+)}(i)$ based on $\widehat{\mu}(H_1)$ and $\left(\frac{T_1}{p}(\widehat{\mathbf{\Sigma}}(H_1))_+\right)^{-1}$, it follows from (9) and (10) that

$$\begin{aligned}\frac{2}{3hp}\sum_{i\in H_2} D^2_{(\widehat{\mu}(H_1),\Delta\frac{T_1}{p}(\widehat{\mathbf{\Sigma}}(H_1))_+)}(i) &= \frac{2}{3hp}\sum_{i\in H_2}(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_1))^\top \frac{1}{\Delta}\left(\frac{T_1}{p}(\widehat{\mathbf{\Sigma}}(H_1))_+\right)^{-1}(\boldsymbol{y}_i^c - \widehat{\boldsymbol{\mu}}(H_1))\\ &= \frac{2}{3\Delta h T_1}\sum_{i\in H_2} D_1^2(i)\\ &= 1.\end{aligned}$$

The Lemma 2.1 suggests that $(\widehat{\boldsymbol{\mu}}(H_2), \frac{T_2}{p}(\widehat{\mathbf{\Sigma}}(H_2))_+)$ is the unique minimizer of $\det(\mathbf{\Lambda})$ among all $(\boldsymbol{\gamma}, \mathbf{\Lambda})$ that satisfy $\frac{2}{3hp}\sum_{i\in H_2} D^2_{(\boldsymbol{\gamma},\mathbf{\Lambda})}(i) = 1$. Thus, we have $\det(\frac{T_2}{p}(\widehat{\mathbf{\Sigma}}(H_2))_+) \le \det(\Delta\frac{T_1}{p}(\widehat{\mathbf{\Sigma}}(H_1))_+)$. Furthermore, from the inequality in (10), we have

$$\det(\frac{T_2}{p}(\widehat{\mathbf{\Sigma}}(H_2))_+) \le \det(\Delta\frac{T_1}{p}(\widehat{\mathbf{\Sigma}}(H_1))_+) \le \det(\frac{T_1}{p}(\widehat{\mathbf{\Sigma}}(H_1))_+), \tag{11}$$

and we can have $\det(\frac{T_2}{p}(\widehat{\mathbf{\Sigma}}(H_2))_+) = \det(\frac{T_1}{p}(\widehat{\mathbf{\Sigma}}(H_1))_+)$ if both of these inequalities are equalities. Since the condition number is relatively small, resulting in both $\frac{T_1}{p}$ and $\frac{T_2}{p}$ being approximately 1, thus $\det((\widehat{\mathbf{\Sigma}}(H_2))_+) \le \det((\widehat{\mathbf{\Sigma}}(H_1))_+)$.

□

Theorem 3.2 *Let $H_1 \subset \{1, 2, \cdots, n\}$ with $|H_1| = h$, obtain the regularization estimator $\widehat{\boldsymbol{S}}(H_1)$ corresponding to $\widehat{\mathbf{\Sigma}}(H_1)$. Then, take H_2 such that $\{d_1^2(i), i \in H_2\} = \{(d_1^2)_{1:n}, (d_1^2)_{2:n}, \cdots, (d_1^2)_{h:n}\}$, where $(d_1^2)_{1:n} \le (d_1^2)_{2:n} \le \cdots \le (d_1^2)_{n:n}$ are the ordered distances. Calculate $\widehat{\boldsymbol{\mu}}(H_2)$ and $\widehat{\mathbf{\Sigma}}(H_2)$ based on H_2,and obtain its corresponding regularized estimator $\widehat{\mathbf{\Sigma}}(H_2)$, then $\det(\widehat{\boldsymbol{S}}(H_2)) \le \det(\widehat{\boldsymbol{S}}(H_1))$.*

The proof procedure is similar to that of Theorem 3.1 and thus omitted.

References

[1] Agulló J, Croux C, Van Aelst S. The multivariate least-trimmed squares estimator. Journal of Multivariate Analysis. 2008;99(3):311–338.

[2] Lopuhaa HP, Rousseeuw PJ. Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. The Annals of Statistics. 1991;p. 229–248.