Covariance matrix estimation under data–based loss

Anis. M. Haddouche[(1)], Dominique Fourdrinier[(2)] and Fatiha Mezoued[(3)]

Université de Sherbrook, département de Mathématiques

05 May 2021

(1) INSA de Rouen Normandie, France

(2) Université de Rouen Normandie, France

(3) ENSSEA, Algérie

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ROUEN

Outline

Introduction

Model

Let consider the multivariate linear regression model

$$Y = X\beta + \mathcal{E}\,, \tag{1.1}$$

where

- Y is an observed $n \times p$ matrix, X is an $n \times q$ matrix of known constants such that

$$\mathrm{rank}(X) = q \leq n. \tag{1.2}$$

- β is a $q \times p$ matrix of unknown parameters.
- $\mathcal{E}$ is an $n \times p$ elliptically symmetric noise.

We assume that $\mathcal{E}$ has a density, w.r.t the Lebesgue measure in $\mathbb{R}^{pn}$, of the form

$$\varepsilon \mapsto |\Sigma|^{-n/2} f\big(\mathrm{tr}(\varepsilon\,\Sigma^{-1}\varepsilon^{\top})\big)\,, \tag{1.3}$$

where Σ is a $p \times p$ unknown positive definite matrix and $f(\cdot)$ is a non–negative unknown function.

Model

Let consider the multivariate linear regression model

$$Y = X\beta + \mathcal{E}\,, \tag{1.1}$$

where

- Y is an observed $n \times p$ matrix, X is an $n \times q$ matrix of known constants such that

$$\operatorname{rank}(X) = q \leq n. \tag{1.2}$$

- β is a $q \times p$ matrix of unknown parameters.
- $\mathcal{E}$ is an $n \times p$ elliptically symmetric noise.

We assume that $\mathcal{E}$ has a density, w.r.t the Lebesgue measure in $\mathbb{R}^{pn}$, of the form

$$\varepsilon \mapsto |\Sigma|^{-n/2} f\left(\operatorname{tr}(\varepsilon\,\Sigma^{-1}\varepsilon^{\top})\right), \tag{1.3}$$

where Σ is a $p \times p$ unknown positive definite matrix and $f(\cdot)$ is a non–negative unknown function.

Model

Let consider the multivariate linear regression model

$$Y = X\beta + \mathcal{E}\,, \tag{1.1}$$

where

- Y is an observed $n \times p$ matrix, X is an $n \times q$ matrix of known constants such that

$$\operatorname{rank}(X) = q \leq n. \tag{1.2}$$

- β is a $q \times p$ matrix of unknown parameters.
- $\mathcal{E}$ is an $n \times p$ elliptically symmetric noise.

We assume that $\mathcal{E}$ has a density, w.r.t the Lebesgue measure in $\mathbb{R}^{pn}$, of the form

$$\varepsilon \mapsto |\Sigma|^{-n/2} f\left(\operatorname{tr}(\varepsilon\,\Sigma^{-1}\varepsilon^{\top})\right), \tag{1.3}$$

where Σ is a $p \times p$ unknown positive definite matrix and $f(\cdot)$ is a non–negative unknown function.

The canonical form

Although the matrix of regression coefficients β is also unknown, we are interested in estimating the invertible scale matrix Σ.

We address this problem under a decision–theoretic framework through a canonical form of the model (1.1).

The canonical form

Thanks to (1.2), the QR decomposition of X is of the form

$$X = Q_1 T^\top ,$$

where

- Q_1 is a $n \times q$ semi-orthogonal matrix.
- T a $q \times q$ lower triangular matrix with positive diagonal elements.

There exists an $n \times (n - q)$ semi-orthogonal matrix Q_2 such that

$$Q_2^\top X \beta = Q_2^\top Q_1 T^\top \beta = 0 .$$

Completes Q_1 with Q_2 such that $Q = (Q_1 Q_2)$ is an $n \times n$ orthogonal matrix. Then, we have

$$\begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} Y = \begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} X \beta + Q^\top \mathcal{E} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + Q^\top \mathcal{E} , \qquad (1.4)$$

The canonical form

Thanks to (1.2), the QR decomposition of X is of the form

$$X = Q_1 T^\top,$$

where

- Q_1 is a $n \times q$ semi-orthogonal matrix.
- T a $q \times q$ lower triangular matrix with positive diagonal elements.

There exists an $n \times (n-q)$ semi-orthogonal matrix Q_2 such that

$$Q_2^\top X \beta = Q_2^\top Q_1 T^\top \beta = 0 .$$

Completes Q_1 with Q_2 such that $Q = (Q_1 Q_2)$ is an $n \times n$ orthogonal matrix. Then, we have

$$\begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} Y = \begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} X \beta + Q^\top \mathcal{E} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + Q^\top \mathcal{E} , \tag{1.4}$$

Why the canonical form

Recall that

$$\begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} Y = \begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + Q^\top \mathcal{E}\,.$$

Inference on the $p \times p$ scale matrix Σ relies on the $(n-q) \times p$ matrix U which is of low dimension than the $n \times p$ observed matrix Y.

Note that

$$S = U^\top U$$

is a sufficient statistic for Σ and may serve as an estimate of Σ.

Note also that S is invertible when $p \leq n-q$ and is non-invertible when $p > n-q$.

In the following we set $m = n - q$

Why the canonical form

Recall that

$$\begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} Y = \begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + Q^\top \mathcal{E} \, .$$

Inference on the $p \times p$ scale matrix Σ relies on the $(n-q) \times p$ matrix U which is of low dimension than the $n \times p$ observed matrix Y.

Note that

$$S = U^\top U$$

is a sufficient statistic for Σ and may serve as an estimate of Σ.

Note also that S is invertible when $p \leq n - q$ and is non-invertible when $p > n - q$.

In the following we set $m = n - q$

Why the canonical form

Recall that

$$\begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} Y = \begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + Q^\top \mathcal{E} \,.$$

Inference on the $p \times p$ scale matrix Σ relies on the $(n-q) \times p$ matrix U which is of low dimension than the $n \times p$ observed matrix Y.

Note that

$$S = U^\top U$$

is a sufficient statistic for Σ and may serve as an estimate of Σ.

Note also that S is invertible when $p \leq n - q$ and is non-invertible when $p > n - q$.

In the following we set $m = n - q$

Why the canonical form

Recall that

$$\begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} Y = \begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + Q^\top \mathcal{E} \,.$$

Inference on the $p \times p$ scale matrix Σ relies on the $(n-q) \times p$ matrix U which is of low dimension than the $n \times p$ observed matrix Y.

Note that

$$S = U^\top U$$

is a sufficient statistic for Σ and may serve as an estimate of Σ.

Note also that S is invertible when $p \leq n-q$ and is non-invertible when $p > n-q$.

In the following we set $m = n - q$

Related expectations

Recall that

$$\varepsilon \mapsto |\Sigma|^{-n/2} f\big(\mathrm{tr}(\,\varepsilon\,\Sigma^{-1}\varepsilon^{\top})\big)\,.$$

The density of $Q^{\top}\mathcal{E}$ is the same as that of $\mathcal{E}$. It follows that the density of $(Z^{\top}\,U^{\top})^{\top} = Q^{\top}Y$ is

$$(z,u) \mapsto |\Sigma|^{-n/2} f\big(\,\mathrm{tr}\,(z-\theta)\,\Sigma^{-1}\,(z-\theta)^{\top} + \mathrm{tr}\,u\,\Sigma^{-1}\,u^{\top}\,\big). \tag{1.5}$$

Bellow $E_{\theta,\Sigma}$ will be the expectation w.r.t (1.5) and $E^{*}_{\theta,\Sigma}$ the expectation w.r.t

$$(z,u) \mapsto \frac{1}{K^{*}}|\Sigma|^{-n/2}\, F^{*}\big(\,\mathrm{tr}\,(z-\theta)\,\Sigma^{-1}\,(z-\theta)^{\top} + \mathrm{tr}\,u\,\Sigma^{-1}\,u^{\top}\,\big)\,,$$

where, for any $t \geq 0$,

$$F^{*}(t) = \frac{1}{2}\int_{t}^{\infty} f(\nu)\,d\nu\,.$$

Related expectations

Recall that

$$\varepsilon \mapsto |\Sigma|^{-n/2} f\big(\mathrm{tr}(\,\varepsilon\,\Sigma^{-1}\varepsilon^\top)\big)\,.$$

The density of $Q^\top \mathcal{E}$ is the same as that of $\mathcal{E}$. It follows that the density of $(Z^\top\, U^\top)^\top = Q^\top Y$ is

$$(z,u) \mapsto |\Sigma|^{-n/2} f\big(\,\mathrm{tr}\,(z-\theta)\,\Sigma^{-1}\,(z-\theta)^\top + \mathrm{tr}\,u\,\Sigma^{-1}\,u^\top\,\big). \tag{1.5}$$

Bellow $E_{\theta,\Sigma}$ will be the expectation w.r.t (1.5) and $E^*_{\theta,\Sigma}$ the expectation w.r.t

$$(z,u) \mapsto \frac{1}{K^*} |\Sigma|^{-n/2}\, F^*\big(\,\mathrm{tr}\,(z-\theta)\,\Sigma^{-1}\,(z-\theta)^\top + \mathrm{tr}\,u\,\Sigma^{-1}\,u^\top\,\big)\,,$$

where, for any $t \geq 0$,

$$F^*(t) = \frac{1}{2}\int_t^\infty f(\nu)\,d\nu\,.$$

Related expectations

Recall that

$$\varepsilon \mapsto |\Sigma|^{-n/2} f\big(\mathrm{tr}(\,\varepsilon\,\Sigma^{-1}\varepsilon^\top)\big)\,.$$

The density of $Q^\top \mathcal{E}$ is the same as that of $\mathcal{E}$. It follows that the density of $(Z^\top\, U^\top)^\top = Q^\top Y$ is

$$(z,u) \mapsto |\Sigma|^{-n/2} f\big(\,\mathrm{tr}\,(z-\theta)\,\Sigma^{-1}\,(z-\theta)^\top + \mathrm{tr}\,u\,\Sigma^{-1}\,u^\top\,\big). \tag{1.5}$$

Bellow $E_{\theta,\Sigma}$ will be the expectation w.r.t (1.5) and $E^*_{\theta,\Sigma}$ the expectation w.r.t

$$(z,u) \mapsto \frac{1}{K^*} |\Sigma|^{-n/2}\, F^*\big(\,\mathrm{tr}\,(z-\theta)\,\Sigma^{-1}\,(z-\theta)^\top + \mathrm{tr}\,u\,\Sigma^{-1}\,u^\top\,\big)\,,$$

where, for any $t \geq 0$,

$$F^*(t) = \frac{1}{2}\int_t^\infty f(\nu)\,d\nu\,.$$

What is wrong with the usual estimators ?

The usual estimators are of the form

$$\hat{\Sigma}_a = a\,S, \quad \text{where} \quad a > 0.$$

In the Gaussian case $\hat{\Sigma}_{1/m}$ correspond respectively to the unbiased estimator.

In the standard asymptotic setting, when p is fixed and $m \to \infty$ the unbiased estimator $\hat{\Sigma}_{1/m}$ is a *good* estimator ; in particular, it is a *consistent* and invertible estimator.

In the general asymptotic setting, when $m, p \to \infty$ with $p/m \to c > 0$, $\hat{\Sigma}_{1/m}$ perform poorly and is non-invertible for $c > 1$.

[5] O. Ledoit and M. Wolf, *A well-conditioned estimator for large-dimensional covariance matrices, 2004.*

What is wrong with the usual estimators ?

The usual estimators are of the form

$$\hat{\Sigma}_a = a\,S, \quad \text{where} \quad a > 0.$$

In the Gaussian case $\hat{\Sigma}_{1/m}$ correspond respectively to the unbiased estimator.

In the standard asymptotic setting, when p is fixed and $m \to \infty$ the unbiased estimator $\hat{\Sigma}_{1/m}$ is a *good* estimator ; in particular, it is a *consistent* and invertible estimator.

In the general asymptotic setting, when $m, p \to \infty$ with $p/m \to c > 0$, $\hat{\Sigma}_{1/m}$ perform poorly and is non-invertible for $c > 1$.

[5] O. Ledoit and M. Wolf, *A well-conditioned estimator for large-dimensional covariance matrices, 2004.*

What is wrong with the usual estimators ?

The usual estimators are of the form

$$\hat{\Sigma}_a = a\,S, \quad \text{where} \quad a > 0.$$

In the Gaussian case $\hat{\Sigma}_{1/m}$ correspond respectively to the unbiased estimator.

In the standard asymptotic setting, when p is fixed and $m \to \infty$ the unbiased estimator $\hat{\Sigma}_{1/m}$ is a *good* estimator ; in particular, it is a *consistent* and invertible estimator.

In the general asymptotic setting, when $m, p \to \infty$ with $p/m \to c > 0$, $\hat{\Sigma}_{1/m}$ perform poorly and is non-invertible for $c > 1$.

[5] O. Ledoit and M. Wolf, *A well-conditioned estimator for large-dimensional covariance matrices, 2004.*

What is wrong with the usual estimators ?

The usual estimators are of the form

$$\hat{\Sigma}_a = a\,S, \quad \text{where} \quad a > 0.$$

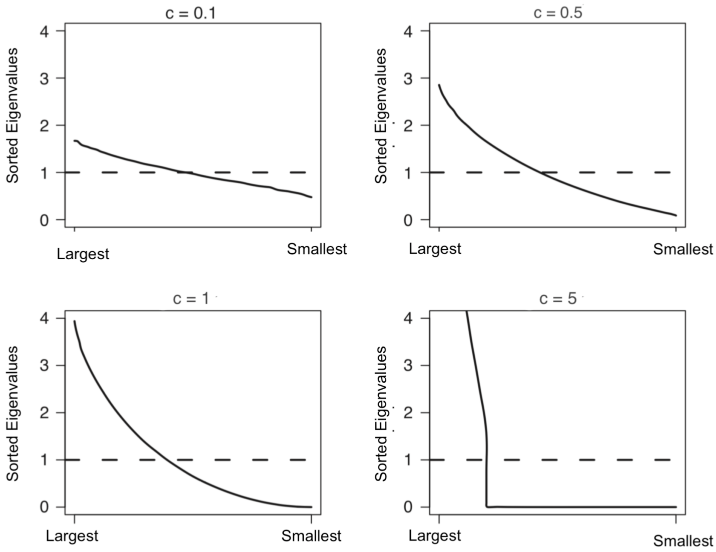
In the Gaussian case $\hat{\Sigma}_{1/m}$ correspond respectively to the unbiased estimator.

In the standard asymptotic setting, when p is fixed and $m \to \infty$ the unbiased estimator $\hat{\Sigma}_{1/m}$ is a *good* estimator ; in particular, it is a *consistent* and invertible estimator.

In the general asymptotic setting, when $m, p \to \infty$ with $p/m \to c > 0$, $\hat{\Sigma}_{1/m}$ perform poorly and is non-invertible for $c > 1$.

[5] O. Ledoit and M. Wolf, *A well-conditioned estimator for large-dimensional covariance matrices, 2004.*

What is wrong with the usual estimators



In the Gaussian setting, James and Stein

[3] W. James and C. Stein, Estimation with Quadratic Loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1961.

show that the usual estimators of the form

$$\hat{\Sigma}_a = a\,S\,, \qquad \text{where} \qquad a > 0,$$

are inadmissible in the general asymptotic setting, when $m, p \to \infty$ with $p/m \to c > 0$.

This phenomenon extends to the elliptical case.

In the following we set $r = \min(m, p)$.

Our objective

Based on the eigenvalue decomposition of $S = H L H^\top$, where

- H is a $p \times r$ semi–orthogonal matrix of eigenvectors.
- $L = \mathrm{diag}(l_1, \ldots, l_r)$, with $l_1 >, \ldots, > l_r$, is the diagonal matrix of the r positive corresponding eigenvalues of S.

We aim to *improve*

$$\hat{\Sigma}_a = a\,S\,, \qquad \text{where} \qquad a > 0,$$

by alternative estimators of the form

$$\hat{\Sigma}_\Psi = a\left(S + H\,L\,\Psi(L)\,H^\top\right) = a\,H\,L\left(I_r + \Psi(L)\right)H^\top\,,$$

with $\Psi(L) = \mathrm{diag}(\psi_1(L), \ldots, \psi_r(L))$, where $\psi_i = \psi_i(L)$ $(i = 1, \ldots, r)$ is a differentiable function of L, which are usually called orthogonally invariant estimators.

Our objective

Based on the eigenvalue decomposition of $S = H\,L\,H^\top$, where

- H is a $p \times r$ semi–orthogonal matrix of eigenvectors.
- $L = \mathrm{diag}(l_1, \ldots, l_r)$, with $l_1 >, \ldots, > l_r$, is the diagonal matrix of the r positive corresponding eigenvalues of S.

We aim to *improve*

$$\hat{\Sigma}_a = a\,S\,, \qquad \text{where} \qquad a > 0,$$

by alternative estimators of the form

$$\hat{\Sigma}_\Psi = a\left(S + H\,L\,\Psi(L)\,H^\top\right) = a\,H\,L\left(I_r + \Psi(L)\right)H^\top\,,$$

with $\Psi(L) = \mathrm{diag}(\psi_1(L), \ldots, \psi_r(L))$, where $\psi_i = \psi_i(L)$ $(i = 1, \ldots, r)$ is a differentiable function of L, which are usually called orthogonally invariant estimators.

Our objective

Based on the eigenvalue decomposition of $S = H\,L\,H^\top$, where

- H is a $p \times r$ semi–orthogonal matrix of eigenvectors.
- $L = \mathrm{diag}(l_1, \ldots, l_r)$, with $l_1 >, \ldots, > l_r$, is the diagonal matrix of the r positive corresponding eigenvalues of S.

We aim to *improve*

$$\hat{\Sigma}_a = a\,S\,, \qquad \text{where} \qquad a > 0,$$

by alternative estimators of the form

$$\hat{\Sigma}_\Psi = a\left(S + H\,L\,\Psi(L)\,H^\top\right) = a\,H\,L\left(I_r + \Psi(L)\right)H^\top\,,$$

with $\Psi(L) = \mathrm{diag}(\psi_1(L), \ldots, \psi_r(L))$, where $\psi_i = \psi_i(L)$ ($i = 1, \ldots, r$) is a differentiable function of L, which are usually called orthogonally invariant estimators.

Our objective

The performance of any estimators $\hat{\Sigma}$ is assessed through the data–based loss

$$L_S(\hat{\Sigma}, \Sigma) = \mathrm{tr}\left(S^+ \Sigma \left(\Sigma^{-1}\hat{\Sigma} - I_p\right)^2\right) \tag{1.6}$$

and its associated risk

$$R(\hat{\Sigma}, \Sigma) = E_{\theta,\Sigma}\left[L_S(\hat{\Sigma}, \Sigma)\right],$$

where

- $E_{\theta,\Sigma}$ denotes the expectation w.r.t. the density specified below in (1.5).
- S^+ is the Moore–Penrose inverse of S. Note that, when $c > 1$, S is non–invertible and, when $c < 1$, S is invertible so that S^+ coincides with the regular inverse S^{-1}.

Our objective

The performance of any estimators $\hat{\Sigma}$ is assessed through the data–based loss

$$L_S(\hat{\Sigma}, \Sigma) = \mathrm{tr}\left(S^{+}\Sigma\left(\Sigma^{-1}\hat{\Sigma} - I_p\right)^2\right) \tag{1.6}$$

and its associated risk

$$R(\hat{\Sigma}, \Sigma) = E_{\theta,\Sigma}\left[L_S(\hat{\Sigma}, \Sigma)\right],$$

where

- $E_{\theta,\Sigma}$ denotes the expectation w.r.t. the density specified below in (1.5).
- S^{+} is the Moore–Penrose inverse of S. Note that, when $c > 1$, S is non–invertible and, when $c < 1$, S is invertible so that S^{+} coincides with the regular inverse S^{-1}.

Why this data–based loss ?

As shown by various authors such as Haddouche et al. (2021, [1]), Konno (2009, [4]) and (1980, Haff [2]), it is difficult to handle on the usual quadratic loss

two Σ^{-1}

$$L(\Sigma, \hat{\Sigma}) = \mathrm{tr}\left((\Sigma^{-1}\hat{\Sigma} - I_p)^2 \right) = \mathrm{tr}(\Sigma^{-1}\hat{\Sigma}\,\Sigma^{-1}\hat{\Sigma}) - 2\,\mathrm{tr}(\Sigma^{-1}\hat{\Sigma}) + p \,. \qquad (1.7)$$

We introduce the data, which give rise to the data–based loss

one Σ^{-1}

$$L_S(\Sigma, \hat{\Sigma}) = \mathrm{tr}(S^{+}\Sigma\,(\Sigma^{-1}\hat{\Sigma} - I_p)^2) = \mathrm{tr}(\Sigma^{-1}\hat{\Sigma}\,S^{+}\hat{\Sigma}) - 2\,\mathrm{tr}\,(S^{+}\hat{\Sigma}) + \mathrm{tr}\,(S^{+}\Sigma) \,.$$

Why this data–based loss ?

As shown by various authors such as Haddouche et al. (2021, [1]), Konno (2009, [4]) and (1980, Haff [2]), it is difficult to handle on the usual quadratic loss

two Σ^{-1}

$$L(\Sigma,\hat{\Sigma}) = \mathrm{tr}(\,(\Sigma^{-1}\,\hat{\Sigma} - I_p)^2) = \mathrm{tr}(\Sigma^{-1}\,\hat{\Sigma}\,\Sigma^{-1}\,\hat{\Sigma}) \; -2\,\mathrm{tr}(\Sigma^{-1}\,\hat{\Sigma}) + p\,. \tag{1.7}$$

We introduce the data, which give rise to the data–based loss

one Σ^{-1}

$$L_S(\Sigma,\hat{\Sigma}) = \mathrm{tr}(S^{+}\,\Sigma\,(\Sigma^{-1}\,\hat{\Sigma} - I_p)^2) = \mathrm{tr}(\Sigma^{-1}\,\hat{\Sigma}\,S^{+}\,\hat{\Sigma}) \; -2\,\mathrm{tr}\,(S^{+}\,\hat{\Sigma}) + \mathrm{tr}\,(S^{+}\,\Sigma)\,.$$

Improved estimators

Our approach

Consider the data–based risk function

$$R(\hat{\Sigma}, \Sigma) = E_{\theta,\Sigma}\left[\mathrm{tr}\left(S^{+}\Sigma\left(\Sigma^{-1}\hat{\Sigma} - I_p\right)^2\right)\right].$$

When $\hat{\Sigma} = \hat{\Sigma}_a = a\,S$, the best constant a is given by

$$a_o = \frac{1}{v\,K^*}, \qquad \text{where} \qquad v = \max(p, m). \tag{2.1}$$

Consider alternative estimators of the form

$$\hat{\Sigma}_\Psi = a_o\,H\,L\,(I_r + \Psi(L))\,H^\top. \tag{2.2}$$

The estimators $\hat{\Sigma}_\Psi$ improves over $\hat{\Sigma}_{a_o}$ as soon as

$$\Delta(G) = R(\hat{\Sigma}_\Psi, \Sigma) - R(\hat{\Sigma}_{a_o}, \Sigma) \leq 0$$

for all Σ, with strict inequality for some Σ.

Our approach

Consider the data–based risk function

$$R(\hat{\Sigma}, \Sigma) = E_{\theta,\Sigma}\left[\mathrm{tr}\left(S^{+}\Sigma\left(\Sigma^{-1}\hat{\Sigma} - I_p\right)^2\right)\right].$$

When $\hat{\Sigma} = \hat{\Sigma}_a = a\,S$, the best constant a is given by

$$a_o = \frac{1}{v\,K^*}\,, \qquad \text{where} \qquad v = \max(p, m)\,. \tag{2.1}$$

Consider alternative estimators of the form

$$\hat{\Sigma}_\Psi = a_o\,H\,L\left(I_r + \Psi(L)\right)H^\top\,. \tag{2.2}$$

The estimators $\hat{\Sigma}_\Psi$ improves over $\hat{\Sigma}_{a_o}$ as soon as

$$\Delta(G) = R(\hat{\Sigma}_\Psi, \Sigma) - R(\hat{\Sigma}_{a_o}, \Sigma) \leq 0$$

for all Σ, with strict inequality for some Σ.

Our approach

Consider the data–based risk function

$$R(\hat{\Sigma}, \Sigma) = E_{\theta,\Sigma}\left[\mathrm{tr}\left(S^{+}\Sigma\left(\Sigma^{-1}\hat{\Sigma} - I_p\right)^2\right)\right].$$

When $\hat{\Sigma} = \hat{\Sigma}_a = a\,S$, the best constant a is given by

$$a_o = \frac{1}{v\,K^*}, \qquad \text{where} \qquad v = \max(p, m). \tag{2.1}$$

Consider alternative estimators of the form

$$\hat{\Sigma}_\Psi = a_o\,H\,L\left(I_r + \Psi(L)\right)H^\top. \tag{2.2}$$

The estimators $\hat{\Sigma}_\Psi$ improves over $\hat{\Sigma}_{a_o}$ as soon as

$$\Delta(G) = R(\hat{\Sigma}_\Psi, \Sigma) - R(\hat{\Sigma}_{a_o}, \Sigma) \leq 0$$

for all Σ, with strict inequality for some Σ.

Our approach

The risk difference between $\hat{\Sigma}_\Psi$ and $\hat{\Sigma}_{a_o}$ is given by

$$\Delta(\Psi) = a_o^2 \, E_{\theta,\Sigma} \left[\mathrm{tr} \left(\Sigma^{-1} H L (2\,\Psi + \Psi^2) H^\top \right) \right] - 2\, a_o \, E_{\theta,\Sigma} \left[\mathrm{tr} \, (\Psi) \right] . \tag{2.3}$$

Replacing the integrand term of $\Delta(\Psi)$ by a random matrix $\delta(\Psi)$, which does not depends on Σ^{-1} such that

$$\Delta(\Psi) \leq E^*_{\theta,\Sigma}[\delta(\Psi)] .$$

A sufficient condition for $\Delta(\Psi)$ to be non–positive is that $\delta(\Psi)$ is non–positive.

To this end, we rely on the following Stein–Haff type identity.

Our approach

The risk difference between $\hat{\Sigma}_{\Psi}$ and $\hat{\Sigma}_{a_o}$ is given by

$$\Delta(\Psi) = a_o^2 \, E_{\theta,\Sigma} \left[\mathrm{tr} \left(\Sigma^{-1} H L (2\,\Psi + \Psi^2) H^{\top} \right) \right] - 2\, a_o \, E_{\theta,\Sigma} \left[\mathrm{tr} \, (\Psi) \right] . \tag{2.3}$$

Replacing the integrand term of $\Delta(\Psi)$ by a random matrix $\delta(\Psi)$, which does not depends on Σ^{-1} such that

$$\Delta(\Psi) \leq E^{*}_{\theta,\Sigma}[\delta(\Psi)] \, .$$

A sufficient condition for $\Delta(\Psi)$ to be non–positive is that $\delta(\Psi)$ is non–positive.

To this end, we rely on the following Stein–Haff type identity.

Our approach

The risk difference between $\hat{\Sigma}_{\Psi}$ and $\hat{\Sigma}_{a_o}$ is given by

$$\Delta(\Psi) = a_o^2\, E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1} HL(2\,\Psi + \Psi^2)H^\top\right)\right] - 2\,a_o\,E_{\theta,\Sigma}\left[\mathrm{tr}\,(\Psi)\right]\,. \tag{2.3}$$

Replacing the integrand term of $\Delta(\Psi)$ by a random matrix $\delta(\Psi)$, which does not depends on Σ^{-1} such that

$$\Delta(\Psi) \le E^*_{\theta,\Sigma}[\delta(\Psi)]\,.$$

A sufficient condition for $\Delta(\Psi)$ to be non–positive is that $\delta(\Psi)$ is non–positive.

To this end, we rely on the following Stein–Haff type identity.

Our approach

The risk difference between $\hat{\Sigma}_\Psi$ and $\hat{\Sigma}_{a_o}$ is given by

$$\Delta(\Psi) = a_o^2 \, E_{\theta,\Sigma} \left[\mathrm{tr} \left(\Sigma^{-1} H L (2\,\Psi + \Psi^2) H^\top \right) \right] - 2\, a_o \, E_{\theta,\Sigma} \left[\mathrm{tr}\,(\Psi) \right] . \tag{2.3}$$

Replacing the integrand term of $\Delta(\Psi)$ by a random matrix $\delta(\Psi)$, which does not depends on Σ^{-1} such that

$$\Delta(\Psi) \leq E^*_{\theta,\Sigma}[\delta(\Psi)] \,.$$

A sufficient condition for $\Delta(\Psi)$ to be non–positive is that $\delta(\Psi)$ is non–positive.

To this end, we rely on the following Stein–Haff type identity.

A Stein–Haff type identity

Lemma 1

Let $\Phi(L) = \mathrm{diag}(\phi_1, \dots, \phi_r)$ *where* $\phi_i = \phi_i(L)$ *(*$i = 1, \dots, r$*) is differentiable function of* L*. Assume that* $E_{\theta,\Sigma}\left[\left|\mathrm{tr}(\Sigma^{-1} H\, L\, \Phi(L)\, H^\top)\right|\right] < \infty$*. Then we have*

$$E_{\theta,\Sigma}\left[\mathrm{tr}(\Sigma^{-1}\, HL\, \Phi(L)\, H^\top)\right] = K^*\, E^*_{\theta,\Sigma}\left[\sum_{i=1}^{r}\left((v - r + 1)\,\phi_i + 2\, l_i\, \frac{\partial \phi_i}{\partial l_i} + \sum_{j \neq i}^{r} \frac{l_i\, \phi_i - l_j\, \phi_j}{l_i - l_j}\right)\right].$$

Dominance results

Theorem 1

Assume that $E_{\theta,\Sigma}\left[\mathrm{tr}(S)\right],\ E_{\theta,\Sigma}\left[\mathrm{tr}(S^{+})\right],\ E_{\theta,\Sigma}\left[\|HL\Psi(L)\,H^{\top}\|_F^2\right]$ *and* $E_{\theta,\Sigma}\left[\|H\Psi(L)H^{\top}\|_F^2\right]$ *are finite. Let* $\Psi(L) = \mathrm{diag}(\psi_1, \ldots, \psi_r)$ *with* $\mathrm{tr}\big(\Psi(L)\big) \geq \lambda$, *for a fixed positive constant* λ. *Then an upper bound of the risk difference in* (2.3) *is given by*

$$\Delta(\Psi(L)) \leq a_o^2\, K^{*}\, E^{*}_{\theta,\Sigma}\left[g(\Psi)\right],$$

where

$$g(\Psi) = \sum_{i=1}^{r}\Bigg\{2(v-r+1)\psi_i + (v-r+1)\psi_i^2 + 4l_i(1+\psi_i)\frac{\partial\psi_i}{\partial l_i} + \sum_{j\neq i}^{r}\frac{l_i\,(2\psi_i+\psi_i^2) - l_j(2\psi_j+\psi_i^2)}{l_i - l_j} - 2v\lambda\Bigg\}.$$

Also, $\hat{\Sigma}_{\Psi}$ *in* (2.2) *improves over* $\hat{\Sigma}_{a_o}$ *in* (2.1) *as soon as* $g(\Psi) \leq 0$.

Dominance results

Theorem 1

Assume that $E_{\theta,\Sigma}\left[\operatorname{tr}(S)\right]$, $E_{\theta,\Sigma}\left[\operatorname{tr}(S^{+})\right]$, $E_{\theta,\Sigma}\left[\|HL\Psi(L)\,H^{\top}\|_F^2\right]$ *and* $E_{\theta,\Sigma}\left[\|H\Psi(L)H^{\top}\|_F^2\right]$ *are finite. Let* $\Psi(L)=\operatorname{diag}(\psi_1,\ldots,\psi_r)$ *with* $\operatorname{tr}\big(\Psi(L)\big)\geq\lambda$, *for a fixed positive constant* λ. *Then an upper bound of the risk difference in* (2.3) *is given by*

$$\Delta(\Psi(L)) \leq a_o^2\, K^*\, E^*_{\theta,\Sigma}\left[g(\Psi)\right],$$

where

$$g(\Psi)=\sum_{i=1}^{r}\left\{2(v-r+1)\psi_i+(v-r+1)\psi_i^2+4l_i(1+\psi_i)\frac{\partial\psi_i}{\partial l_i}+\sum_{j\neq i}^{r}\frac{l_i\,(2\psi_i+\psi_i^2)-l_j(2\psi_j+\psi_i^2)}{l_i-l_j}-2v\lambda\right\}.$$

Also, $\hat{\Sigma}_{\Psi}$ *in* (2.2) *improves over* $\hat{\Sigma}_{a_o}$ *in* (2.1) *as soon as* $g(\Psi)\leq 0$.

Dominance results

Theorem 1

Assume that $E_{\theta,\Sigma}\left[\mathrm{tr}(S)\right],\ E_{\theta,\Sigma}\left[\mathrm{tr}(S^{+})\right],\ E_{\theta,\Sigma}\left[\|HL\Psi(L)\,H^{\top}\|_F^2\right]$ *and* $E_{\theta,\Sigma}\left[\|H\Psi(L)H^{\top}\|_F^2\right]$ *are finite. Let* $\Psi(L) = \mathrm{diag}(\psi_1, \ldots, \psi_r)$ *with* $\mathrm{tr}\big(\Psi(L)\big) \geq \lambda$, *for a fixed positive constant* λ. *Then an upper bound of the risk difference in* (2.3) *is given by*

$$\Delta(\Psi(L)) \leq a_o^2\, K^*\, E^*_{\theta,\Sigma}\left[g(\Psi)\right],$$

where

$$g(\Psi) = \sum_{i=1}^{r} \Bigg\{ 2(v-r+1)\psi_i + (v-r+1)\psi_i^2 + 4l_i(1+\psi_i)\frac{\partial \psi_i}{\partial l_i} + \sum_{j\neq i}^{r} \frac{l_i\,(2\psi_i+\psi_i^2) - l_j(2\psi_j+\psi_i^2)}{l_i - l_j} - 2v\lambda \Bigg\}.$$

Also, $\hat{\Sigma}_{\Psi}$ *in* (2.2) *improves over* $\hat{\Sigma}_{a_o}$ *in* (2.1) *as soon as* $g(\Psi) \leq 0$.

Dominance results

Theorem 1

Assume that $E_{\theta,\Sigma}\left[\mathrm{tr}(S)\right]$, $E_{\theta,\Sigma}\left[\mathrm{tr}(S^{+})\right]$, $E_{\theta,\Sigma}\left[\|HL\Psi(L)\,H^{\top}\|_F^2\right]$ *and* $E_{\theta,\Sigma}\left[\|H\Psi(L)H^{\top}\|_F^2\right]$ *are finite. Let* $\Psi(L) = \mathrm{diag}(\psi_1,\ldots,\psi_r)$ *with* $\mathrm{tr}\big(\Psi(L)\big) \geq \lambda$*, for a fixed positive constant* λ*. Then an upper bound of the risk difference in* (2.3) *is given by*

$$\Delta(\Psi(L)) \leq a_o^2\, K^*\, E_{\theta,\Sigma}^*\left[g(\Psi)\right],$$

where

$$g(\Psi) = \sum_{i=1}^{r}\left\{2(v-r+1)\psi_i + (v-r+1)\psi_i^2 + 4l_i(1+\psi_i)\frac{\partial\psi_i}{\partial l_i} + \sum_{j\neq i}^{r}\frac{l_i\,(2\psi_i+\psi_i^2) - l_j(2\psi_j+\psi_i^2)}{l_i - l_j} - 2v\lambda\right\}.$$

Also, $\hat{\Sigma}_{\Psi}$ *in* (2.2) *improves over* $\hat{\Sigma}_{a_o}$ *in* (2.1) *as soon as* $g(\Psi) \leq 0$.

Sketch of proof

Recall that

$$\Delta(\Psi) = a_o^2\, E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1} H L (2\,\Psi + \Psi^2) H^\top\right)\right] - 2\, a_o\, E_{\theta,\Sigma}\left[\mathrm{tr}\,(\Psi)\right]\,.$$

In order to the rid of Σ^{-1} in the integrand term, we apply the Stein–Haff type identity in Lemma 1. Then we have

$$\Delta(\Psi) = a_o^2\, K^*\, E^*_{\theta,\Sigma}\left[\sum_{i=1}^{r}\left\{(v-r+1)\,(2\,\psi_i + \psi_i^2) + 2\, l_i\, \frac{\partial(2\,\psi_i + \psi_i^2)}{\partial l_i} + \sum_{j\neq i}^{r} \frac{l_i\,(2\,\psi_i + \psi_i^2) - l_j\,(2\,\psi_j + \psi_i^2)}{l_i - l_j}\right\}\right] - 2\, a_o E_{\theta,\Sigma}\left[\mathrm{tr}(\Psi)\right].$$

Therefore, using the fact that

$$\mathrm{tr}(\Psi(L)) \geq \lambda > 0,$$

an upper bound of the risk difference $\Delta(\Psi)$ is given by

$$\Delta(\Psi) \leq a_o^2\, K^*\, E^*_{\theta,\Sigma}\left[g(\Psi)\right]\,,$$

Sketch of proof

Recall that

$$\Delta(\Psi) = a_o^2 \, E_{\theta,\Sigma} \left[\mathrm{tr} \left(\Sigma^{-1} HL(2\,\Psi + \Psi^2) H^\top \right) \right] - 2\, a_o \, E_{\theta,\Sigma} \left[\mathrm{tr} \, (\Psi) \right] .$$

In order to the rid of Σ^{-1} in the integrand term, we apply the Stein–Haff type identity in Lemma 1. Then we have

$$\Delta(\Psi) = a_o^2 \, K^* \, E^*_{\theta,\Sigma} \left[\sum_{i=1}^{r} \left\{ (v - r + 1)\,(2\,\psi_i + \psi_i^2) + 2\, l_i \, \frac{\partial (2\,\psi_i + \psi_i^2)}{\partial l_i} + \sum_{j \neq i}^{r} \frac{l_i\,(2\,\psi_i + \psi_i^2) - l_j\,(2\,\psi_j + \psi_i^2)}{l_i - l_j} \right\} \right] - 2\, a_o E_{\theta,\Sigma} \left[\mathrm{tr} (\Psi) \right] .$$

Therefore, using the fact that

$$\mathrm{tr}(\Psi(L)) \geq \lambda > 0,$$

an upper bound of the risk difference $\Delta(\Psi)$ is given by

$$\Delta(\Psi) \leq a_o^2 \, K^* \, E^*_{\theta,\Sigma} \left[g(\Psi) \right] ,$$

Sketch of proof

Recall that

$$\Delta(\Psi) = a_o^2\, E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1} HL(2\,\Psi + \Psi^2)H^\top\right)\right] - 2\,a_o\,E_{\theta,\Sigma}\left[\mathrm{tr}\,(\Psi)\right] .$$

In order to the rid of Σ^{-1} in the integrand term, we apply the Stein–Haff type identity in Lemma 1. Then we have

$$\Delta(\Psi) = a_o^2\, K^*\, E^*_{\theta,\Sigma}\left[\sum_{i=1}^{r}\left\{(v-r+1)\,(2\,\psi_i+\psi_i^2) + 2\,l_i\,\frac{\partial(2\,\psi_i+\psi_i^2)}{\partial l_i} + \sum_{j\neq i}^{r}\frac{l_i\,(2\,\psi_i+\psi_i^2) - l_j\,(2\,\psi_j+\psi_i^2)}{l_i - l_j}\right\}\right] - 2\,a_o E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Psi\right)\right] .$$

Therefore, using the fact that

$$\mathrm{tr}(\Psi(L)) \geq \lambda > 0,$$

an upper bound of the risk difference $\Delta(\Psi)$ is given by

$$\Delta(\Psi) \leq a_o^2\, K^*\, E^*_{\theta,\Sigma}\left[g(\Psi)\right] ,$$

Examples

Note that Theorem 1 is well adapted to deal with:

- The James Stein (1961, [3]) estimator where

$$\psi_i(L) = \frac{1}{(v + r - 2i + 1)},$$

for $i = 1, \ldots, r$, since

$$\mathrm{tr}\big(\Psi(L)\big) > \lambda = 1/(v + r - 1).$$

- The Efron-Morris-Dey estimator, considered by Tsukuma and Kubokawa (2020, [6]), where

$$\psi_i(L) = \frac{1}{\left(1 + b\frac{l_i^\alpha}{\mathrm{tr}(L^\alpha)}\right)v},$$

for $i = 1, \ldots, r$ and for positive constants b and α, since

$$\mathrm{tr}\big(\Psi(L)\big) > \lambda = r\,/(b + 1)\,v\,.$$

Examples

Note that Theorem 1 is well adapted to deal with:

- The James Stein (1961, [3]) estimator where

$$\psi_i(L) = \frac{1}{(v + r - 2i + 1)},$$

  for $i = 1, \ldots, r$, since

$$\mathrm{tr}\big(\Psi(L)\big) > \lambda = 1/(v + r - 1).$$

- The Efron-Morris-Dey estimator, considered by Tsukuma and Kubokawa (2020, [6]), where

$$\psi_i(L) = \frac{1}{\left(1 + b\frac{l_i^\alpha}{\mathrm{tr}(L^\alpha)}\right)v},$$

  for $i = 1, \ldots, r$ and for positive constants b and α, since

$$\mathrm{tr}\big(\Psi(L)\big) > \lambda = r\,/(b + 1)\,v\,.$$

Haff estimators : the Gaussian case

When $c \leq 1$, Haff (1980, [2]) considered an empirical Bayes estimation of Σ.

Let a prior of Σ^{-1} be

$$\Sigma^{-1} \sim \mathcal{W}_p(p+1, \gamma^{-1} I_p),$$

where γ is an unknown hyperparameter.

The resulting posterior distribution of Σ^{-1} given S is

$$\Sigma^{-1} | S \sim \mathcal{W}_p(m+p+1, (S+\gamma I_p)^{-1}).$$

So that, the posterior mean of Σ is

$$E[\Sigma | S] = \frac{1}{m}(S + \gamma I_p) \, .$$

Haff estimators : the Gaussian case

When $c \leq 1$, Haff (1980, [2]) considered an empirical Bayes estimation of Σ.

Let a prior of Σ^{-1} be

$$\Sigma^{-1} \sim \mathcal{W}_p(p+1, \gamma^{-1} I_p),$$

where γ is an unknown hyperparameter.

The resulting posterior distribution of Σ^{-1} given S is

$$\Sigma^{-1}|S \sim \mathcal{W}_p(m+p+1, (S+\gamma I_p)^{-1}).$$

So that, the posterior mean of Σ is

$$E[\Sigma|S] = \frac{1}{m}(S+\gamma I_p) \, .$$

Haff estimators : the Gaussian case

When $c \leq 1$, Haff (1980, [2]) considered an empirical Bayes estimation of Σ.

Let a prior of Σ^{-1} be

$$\Sigma^{-1} \sim \mathcal{W}_p(p+1, \gamma^{-1} I_p),$$

where γ is an unknown hyperparameter.

The resulting posterior distribution of Σ^{-1} given S is

$$\Sigma^{-1} | S \sim \mathcal{W}_p(m+p+1, (S+\gamma I_p)^{-1}).$$

So that, the posterior mean of Σ is

$$E[\Sigma|S] = \frac{1}{m}(S+\gamma I_p)\,.$$

Haff estimators : the Gaussian case

The hyperparameter γ is estimated from the marginal density of S proportional to

$$\gamma^{p(p+1)/2}|S|^{(m-p-1)/2}|S+\gamma I_p|^{-(m+p+1)/2} .$$

The resulting maximum likelihood estimator is

$$\hat{\gamma} = \frac{p(p+1)}{m+p+1}\frac{1}{\mathrm{tr}(S^{-1})} .$$

Thus we obtain an empirical Bayes estimator of the form

$$\hat{\Sigma}^{EB} = \frac{1}{m}(S+\hat{\gamma}\, I_p) = \frac{1}{m} H\, L\left(I_p + \frac{p(p+1)}{m+p+1}\frac{L^{-1}}{\mathrm{tr}(L^{-1})}\right) H^{\top}$$

since $S = H\, L\, H^{\top}$ and $\mathrm{tr}(S^{-1}) = \mathrm{tr}(L^{-1})$.

Haff estimators : the Gaussian case

The hyperparameter γ is estimated from the marginal density of S proportional to

$$\gamma^{p(p+1)/2} |S|^{(m-p-1)/2} |S + \gamma I_p|^{-(m+p+1)/2} .$$

The resulting maximum likelihood estimator is

$$\hat{\gamma} = \frac{p(p+1)}{m+p+1} \frac{1}{\mathrm{tr}(S^{-1})} .$$

Thus we obtain an empirical Bayes estimator of the form

$$\hat{\Sigma}^{EB} = \frac{1}{m} (S + \hat{\gamma} I_p) = \frac{1}{m} H L \left(I_p + \frac{p(p+1)}{m+p+1} \frac{L^{-1}}{\mathrm{tr}(L^{-1})} \right) H^\top$$

since $S = H L H^\top$ and $\mathrm{tr}(S^{-1}) = \mathrm{tr}(L^{-1})$.

Haff estimators : the Gaussian case

Taking into account both cases where $c \leq 1$ and $c > 1$, we define a Haff type estimators as

$$\hat{\Sigma}_{\alpha,b} = a_o \, H \left(I_r + b \frac{L^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} \right) H^\top \,, \quad \text{where} \quad \alpha \geq 1 \text{ and } b > 0 \,. \qquad (2.4)$$

For $\alpha = 1$, this is the estimator considered by Konno (2009, [4]), who deals with the Gaussian case and the quadratic loss

Tsukuma and Kubokawa (2020, [6]) used an extended Stein loss.

An elliptical setting was also considered by Haddouche et al. (2021, [1]) under the usual quadratic loss function.

Haff estimators : the Gaussian case

Taking into account both cases where $c \leq 1$ and $c > 1$, we define a Haff type estimators as

$$\hat{\Sigma}_{\alpha,b} = a_o\, H \left(I_r + b \frac{L^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} \right) H^\top\,, \quad \text{where} \quad \alpha \geq 1 \text{ and } b > 0\,. \qquad (2.4)$$

For $\alpha = 1$, this is the estimator considered by Konno (2009, [4]), who deals with the Gaussian case and the quadratic loss

Tsukuma and Kubokawa (2020, [6]) used an extended Stein loss.

An elliptical setting was also considered by Haddouche et al. (2021, [1]) under the usual quadratic loss function.

Haff estimators : the Gaussian case

Taking into account both cases where $c \leq 1$ and $c > 1$, we define a Haff type estimators as

$$\hat{\Sigma}_{\alpha,b} = a_o \, H \left(I_r + b \frac{L^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} \right) H^\top \,, \quad \text{where} \quad \alpha \geq 1 \text{ and } b > 0 \,. \tag{2.4}$$

For $\alpha = 1$, this is the estimator considered by Konno (2009, [4]), who deals with the Gaussian case and the quadratic loss

Tsukuma and Kubokawa (2020, [6]) used an extended Stein loss.

An elliptical setting was also considered by Haddouche et al. (2021, [1]) under the usual quadratic loss function.

Haff estimators : the Gaussian case

Taking into account both cases where $c \leq 1$ and $c > 1$, we define a Haff type estimators as

$$\hat{\Sigma}_{\alpha,b} = a_o \, H \left(I_r + b \frac{L^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} \right) H^\top , \quad \text{where} \quad \alpha \geq 1 \text{ and } b > 0 \, . \tag{2.4}$$

For $\alpha = 1$, this is the estimator considered by Konno (2009, [4]), who deals with the Gaussian case and the quadratic loss

Tsukuma and Kubokawa (2020, [6]) used an extended Stein loss.

An elliptical setting was also considered by Haddouche et al. (2021, [1]) under the usual quadratic loss function.

Haff estimators : the elliptical case

Recall that the Haff estimators are of the form

$$\hat{\Sigma}_{\alpha,b} = a_o\, H \left(I_r + b \frac{L^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} \right) H^\top \quad \text{where} \quad \alpha \geq 1 \text{ and } b > 0\,.$$

Proposition 1

Assume that the expectations $E_{\theta,\Sigma}\left[\mathrm{tr}(S^+)\right]$ and $E_{\theta,\Sigma}\left[\mathrm{tr}^2(S)\right]$ are finite. Then the Haff type estimators $\hat{\Sigma}_{\alpha,b}$ in (2.4) *improves on the usual estimator $\hat{\Sigma}_{a_o}$ in* (2.1) *under the data–based loss* (1.6) *as soon as*

$$0 < b \leq \frac{2\,(r-1)}{(v-r+1)}\,.$$

Haff estimators : the elliptical case

Recall that the Haff estimators are of the form

$$\hat{\Sigma}_{\alpha,b} = a_o\, H \left(I_r + b \frac{L^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} \right) H^\top \quad \text{where} \quad \alpha \geq 1 \text{ and } b > 0\,.$$

Proposition 1

Assume that the expectations $E_{\theta,\Sigma}\,[\mathrm{tr}(S^+)]$ *and* $E_{\theta,\Sigma}\,\left[\mathrm{tr}^2(S)\right]$ *are finite. Then the Haff type estimators* $\hat{\Sigma}_{\alpha,b}$ *in* (2.4) *improves on the usual estimator* $\hat{\Sigma}_{a_o}$ *in* (2.1) *under the data–based loss* (1.6) *as soon as*

$$0 < b \leq \frac{2\,(r-1)}{(v-r+1)}\,.$$

Sketch of Proof 1/3

Applying Theorem 1, an upper bound of the risk difference is given by

$$\Delta(\Psi) \leq a_o^2 \, K^* \, E^*_{\theta,\Sigma}\big(g(\Psi)\big)\,, \tag{2.5}$$

where $g(\Psi) = g_1(\Psi) + g_2(\Psi)$ with

$$g_1(\Psi) = -2\,(r-1)\,b \sum_{i=1}^{r} \frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} + (v-r+1)\,b^2 \sum_{i=1}^{r} \frac{l_i^{-2\alpha}}{\mathrm{tr}^2(L^{-\alpha})}$$

and

$$g_2(\Psi) = 4 l_i b \left(1 + b \frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) \frac{\partial}{\partial l_i} \left(\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) + \frac{2b}{\mathrm{tr}(L^{-\alpha})} \sum_{i=1}^{r} \sum_{j \neq i}^{r} \frac{l_i^{1-\alpha} - l_j^{1-\alpha}}{l_i - l_j}$$
$$+ \frac{b^2}{\mathrm{tr}^2(L^{-\alpha})} \sum_{i=1}^{r} \sum_{j \neq i}^{r} \frac{l_i^{1-2\alpha} - l_j^{1-2\alpha}}{l_i - l_j}\,.$$

The proof consists in proving that $g_2(\Psi)$ is non–positive.

Sketch of Proof 1/3

Applying Theorem 1, an upper bound of the risk difference is given by

$$\Delta(\Psi) \leq a_o^2 \, K^* \, E^*_{\theta,\Sigma}\big(g(\Psi)\big)\,, \tag{2.5}$$

where $g(\Psi) = g_1(\Psi) + g_2(\Psi)$ with

$$g_1(\Psi) = -2\,(r-1)\,b\,\sum_{i=1}^{r} \frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} + (v-r+1)\,b^2\,\sum_{i=1}^{r} \frac{l_i^{-2\alpha}}{\mathrm{tr}^2(L^{-\alpha})}$$

and

$$g_2(\Psi) = 4 l_i b \left(1 + b\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) \frac{\partial}{\partial l_i}\left(\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) + \frac{2b}{\mathrm{tr}(L^{-\alpha})} \sum_{i=1}^{r}\sum_{j\neq i}^{r} \frac{l_i^{1-\alpha} - l_j^{1-\alpha}}{l_i - l_j}$$
$$+ \frac{b^2}{\mathrm{tr}^2(L^{-\alpha})} \sum_{i=1}^{r}\sum_{j\neq i}^{r} \frac{l_i^{1-2\alpha} - l_j^{1-2\alpha}}{l_i - l_j}\,.$$

The proof consists in proving that $g_2(\Psi)$ is non–positive.

Sketch of Proof 2/3

Recall that

$$g_2(\Psi) = 4 l_i b \left(1 + b \frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) \frac{\partial}{\partial l_i} \left(\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) + \frac{2b}{\mathrm{tr}(L^{-\alpha})} \sum_{i=1}^{r} \sum_{j \neq i}^{r} \frac{l_i^{1-\alpha} - l_j^{1-\alpha}}{l_i - l_j}$$

$$+ \frac{b^2}{\mathrm{tr}^2(L^{-\alpha})} \sum_{i=1}^{r} \sum_{j \neq i}^{r} \frac{l_i^{1-2\alpha} - l_j^{1-2\alpha}}{l_i - l_j} .$$

It can be shown that, for $\alpha \geq 1$,

$$\sum_{i=1}^{r} \sum_{j \neq i}^{r} \frac{l_i^{1-\alpha} - l_j^{1-\alpha}}{l_i - l_j} \leq 0 , \quad \sum_{i=1}^{r} \sum_{j \neq i}^{r} \frac{l_i^{1-2\alpha} - l_j^{1-2\alpha}}{l_i - l_j} \leq 0$$

and

$$\frac{\partial}{\partial l_i} \left(\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) = \alpha \frac{l_i^{-\alpha-1}}{\mathrm{tr}(L^{-\alpha})} \left(\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} - 1\right) \leq 0 .$$

Sketch of Proof 2/3

Recall that

$$g_2(\Psi) = 4 l_i b \left(1 + b \frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) \frac{\partial}{\partial l_i}\left(\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) + \frac{2b}{\mathrm{tr}(L^{-\alpha})} \sum_{i=1}^{r} \sum_{j \neq i}^{r} \frac{l_i^{1-\alpha} - l_j^{1-\alpha}}{l_i - l_j} + \frac{b^2}{\mathrm{tr}^2(L^{-\alpha})} \sum_{i=1}^{r} \sum_{j \neq i}^{r} \frac{l_i^{1-2\alpha} - l_j^{1-2\alpha}}{l_i - l_j} .$$

It can be shown that, for $\alpha \geq 1$,

$$\sum_{i=1}^{r} \sum_{j \neq i}^{r} \frac{l_i^{1-\alpha} - l_j^{1-\alpha}}{l_i - l_j} \leq 0 , \quad \sum_{i=1}^{r} \sum_{j \neq i}^{r} \frac{l_i^{1-2\alpha} - l_j^{1-2\alpha}}{l_i - l_j} \leq 0$$

and

$$\frac{\partial}{\partial l_i}\left(\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) = \alpha \frac{l_i^{-\alpha-1}}{\mathrm{tr}(L^{-\alpha})} \left(\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} - 1\right) \leq 0 .$$

Sketch of Proof 2/3

Recall that

$$g_2(\Psi) = 4l_i b \left(1 + b\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) \frac{\partial}{\partial l_i}\left(\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) + \frac{2b}{\mathrm{tr}(L^{-\alpha})}\sum_{i=1}^{r}\sum_{j\neq i}^{r} \frac{l_i^{1-\alpha} - l_j^{1-\alpha}}{l_i - l_j}$$

$$+ \frac{b^2}{\mathrm{tr}^2(L^{-\alpha})}\sum_{i=1}^{r}\sum_{j\neq i}^{r} \frac{l_i^{1-2\alpha} - l_j^{1-2\alpha}}{l_i - l_j}\,.$$

It can be shown that, for $\alpha \geq 1$,

$$\sum_{i=1}^{r}\sum_{j\neq i}^{r} \frac{l_i^{1-\alpha} - l_j^{1-\alpha}}{l_i - l_j} \leq 0\,, \quad \sum_{i=1}^{r}\sum_{j\neq i}^{r} \frac{l_i^{1-2\alpha} - l_j^{1-2\alpha}}{l_i - l_j} \leq 0$$

and

$$\frac{\partial}{\partial l_i}\left(\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})}\right) = \alpha\frac{l_i^{-\alpha-1}}{\mathrm{tr}(L^{-\alpha})}\left(\frac{l_i^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} - 1\right) \leq 0\,.$$

Sketch of Proof 3/3

Therefore, since $l_i^{-\alpha} \leq \mathrm{tr}(L^{-\alpha})$, the integrand term $g_2(\Psi) \leq 0$. Then

$$g(\Psi) \leq g_1(\Psi) = -2\,(r-1)\,b \,+ (v-r+1)\,b^2\,\frac{\mathrm{tr}(L^{-2\alpha})}{\mathrm{tr}^2(L^{-\alpha})}\,.$$

Now, using the fact that $\mathrm{tr}(L^{-2\alpha}) \leq \mathrm{tr}^2(L^{-\alpha})$, we have

$$g(\Psi) \leq -2\,(r-1)\,b + (v-r+1)\,b^2\,.$$

Hence, an upper bound for the risk difference in (2.5) is given by

$$\Delta(\Psi) \leq a_o^2\, b\, K^*\, E^*_{\theta,\Sigma}\big[-2\,(r-1)\, + (v-r+1)\,b\big]\,.$$

Therefore, $\hat{\Sigma}_{\alpha,b}$ improves over $\hat{\Sigma}_{a_o}$ under the data-based loss (1.6) as soon as $0 < b \leq b_0 = 2\,(r-1)/(v-r+1)\,.$

Sketch of Proof 3/3

Therefore, since $l_i^{-\alpha} \leq \mathrm{tr}(L^{-\alpha})$, the integrand term $g_2(\Psi) \leq 0$. Then

$$g(\Psi) \leq g_1(\Psi) = -2\,(r-1)\,b \;+ (v-r+1)\,b^2\,\frac{\mathrm{tr}(L^{-2\alpha})}{\mathrm{tr}^2(L^{-\alpha})}\,.$$

Now, using the fact that $\mathrm{tr}(L^{-2\alpha}) \leq \mathrm{tr}^2(L^{-\alpha})$, we have

$$g(\Psi) \leq -2\,(r-1)\,b + (v-r+1)\,b^2\,.$$

Hence, an upper bound for the risk difference in (2.5) is given by

$$\Delta(\Psi) \leq a_o^2\,b\,K^*\,E^*_{\theta,\Sigma}\big[-2\,(r-1)\;+(v-r+1)\,b\big]\,.$$

Therefore, $\hat{\Sigma}_{\alpha,b}$ improves over $\hat{\Sigma}_{a_o}$ under the data-based loss (1.6) as soon as $0 < b \leq b_0 = 2\,(r-1)/(v-r+1)\,.$

Sketch of Proof 3/3

Therefore, since $l_i^{-\alpha} \leq \mathrm{tr}(L^{-\alpha})$, the integrand term $g_2(\Psi) \leq 0$. Then

$$g(\Psi) \leq g_1(\Psi) = -2\,(r-1)\,b \;+ (v-r+1)\,b^2\,\frac{\mathrm{tr}(L^{-2\alpha})}{\mathrm{tr}^2(L^{-\alpha})}\,.$$

Now, using the fact that $\mathrm{tr}(L^{-2\alpha}) \leq \mathrm{tr}^2(L^{-\alpha})$, we have

$$g(\Psi) \leq -2\,(r-1)\,b + (v-r+1)\,b^2\,.$$

Hence, an upper bound for the risk difference in (2.5) is given by

$$\Delta(\Psi) \leq a_o^2\,b\,K^*\,E^*_{\theta,\Sigma}\big[-2\,(r-1) \;+ (v-r+1)\,b\big]\,.$$

Therefore, $\hat{\Sigma}_{\alpha,b}$ improves over $\hat{\Sigma}_{a_o}$ under the data-based loss (1.6) as soon as $0 < b \leq b_0 = 2\,(r-1)/(v-r+1)$.

Sketch of Proof 3/3

Therefore, since $l_i^{-\alpha} \leq \mathrm{tr}(L^{-\alpha})$, the integrand term $g_2(\Psi) \leq 0$. Then

$$g(\Psi) \leq g_1(\Psi) = -2\,(r-1)\,b \,+ (v-r+1)\,b^2\,\frac{\mathrm{tr}(L^{-2\alpha})}{\mathrm{tr}^2(L^{-\alpha})}\,.$$

Now, using the fact that $\mathrm{tr}(L^{-2\alpha}) \leq \mathrm{tr}^2(L^{-\alpha})$, we have

$$g(\Psi) \leq -2\,(r-1)\,b + (v-r+1)\,b^2\,.$$

Hence, an upper bound for the risk difference in (2.5) is given by

$$\Delta(\Psi) \leq a_o^2\,b\,K^*\,E^*_{\theta,\Sigma}\big[-2\,(r-1)\,+(v-r+1)\,b\big]\,.$$

Therefore, $\hat{\Sigma}_{\alpha,b}$ improves over $\hat{\Sigma}_{a_o}$ under the data-based loss (1.6) as soon as $0 < b \leq b_0 = 2\,(r-1)/(v-r+1)$.

Numerical study

Case of a t-distribution

Let the elliptical density in (1.3) be a variance mixture of normal distributions where the mixing variable, with density h, has the inverse–gamma distribution $\mathcal{IG}(k/2, k/2)$

Thus, for any $t \geq 0$, the generating function f in (1.3) has the form

$$f(t) = \int_0^\infty \frac{1}{(2\mathrm{v}\pi)^{np/2}} \exp\left(\frac{-t}{2\mathrm{v}}\right) h(\mathrm{v})\, d\mathrm{v}\,,$$

which corresponds to the t–distribution with k degrees of freedom.

Then the primitive F^* of f in is, for any $t \geq 0$,

$$F^*(t) = \int_0^\infty \frac{\mathrm{v}}{(2\mathrm{v}\pi)^{np/2}} \exp\left(\frac{-t}{2\mathrm{v}}\right) h(\mathrm{v})\, d\mathrm{v}\,.$$

The normalizing constant K^* is given by

$$K^* = \frac{k}{k-2}.$$

Case of a t-distribution

Let the elliptical density in (1.3) be a variance mixture of normal distributions where the mixing variable, with density h, has the inverse–gamma distribution $\mathcal{IG}(k/2, k/2)$

Thus, for any $t \geq 0$, the generating function f in (1.3) has the form

$$f(t) = \int_0^\infty \frac{1}{(2\mathrm{v}\pi)^{np/2}} \exp\left(\frac{-t}{2\mathrm{v}}\right) h(\mathrm{v})\, d\mathrm{v}\,,$$

which corresponds to the t–distribution with k degrees of freedom.

Then the primitive F^* of f in is, for any $t \geq 0$,

$$F^*(t) = \int_0^\infty \frac{\mathrm{v}}{(2\mathrm{v}\pi)^{np/2}} \exp\left(\frac{-t}{2\mathrm{v}}\right) h(\mathrm{v})\, d\mathrm{v}\,.$$

The normalizing constant K^* is given by

$$K^* = \frac{k}{k-2}.$$

Case of a t-distribution

Let the elliptical density in (1.3) be a variance mixture of normal distributions where the mixing variable, with density h, has the inverse–gamma distribution $\mathcal{IG}(k/2, k/2)$

Thus, for any $t \geq 0$, the generating function f in (1.3) has the form

$$f(t) = \int_0^\infty \frac{1}{(2\mathrm{v}\pi)^{np/2}} \exp\left(\frac{-t}{2\mathrm{v}}\right) h(\mathrm{v})\, d\mathrm{v}\,,$$

which corresponds to the t–distribution with k degrees of freedom.

Then the primitive F^* of f in is, for any $t \geq 0$,

$$F^*(t) = \int_0^\infty \frac{\mathrm{v}}{(2\mathrm{v}\pi)^{np/2}} \exp\left(\frac{-t}{2\mathrm{v}}\right) h(\mathrm{v})\, d\mathrm{v}\,.$$

The normalizing constant K^* is given by

$$K^* = \frac{k}{k-2}.$$

Case of a t-distribution

Let the elliptical density in (1.3) be a variance mixture of normal distributions where the mixing variable, with density h, has the inverse–gamma distribution $\mathcal{IG}(k/2, k/2)$

Thus, for any $t \geq 0$, the generating function f in (1.3) has the form

$$f(t) = \int_0^\infty \frac{1}{(2\mathrm{v}\pi)^{np/2}} \exp\left(\frac{-t}{2\mathrm{v}}\right) h(\mathrm{v})\, d\mathrm{v}\,,$$

which corresponds to the t–distribution with k degrees of freedom.

Then the primitive F^* of f in is, for any $t \geq 0$,

$$F^*(t) = \int_0^\infty \frac{\mathrm{v}}{(2\mathrm{v}\pi)^{np/2}} \exp\left(\frac{-t}{2\mathrm{v}}\right) h(\mathrm{v})\, d\mathrm{v}\,.$$

The normalizing constant K^* is given by

$$K^* = \frac{k}{k-2}.$$

Improved estimators

We study numerically the performance of

$$\hat{\Sigma}_{\alpha,b} = a_o\, H \left(I_r + b \frac{L^{-\alpha}}{\operatorname{tr}(L^{-\alpha})} \right) H^\top \,, \tag{3.1}$$

where

$$0 \leq b \leq b_0 = \frac{2\,(r-1)}{v-r+1} \quad \text{and} \quad \alpha \geq 1.$$

Konno (2009, [4]) consider the case $\alpha = 1$, in the Gaussian setting and under the usual quadratic loss, for which its improvement condition is $0 \leq b \leq b_1 = 2\,(r-1)\,(v+r+1)/(v-r+1)\,(v-r+3)$. Although $b_0 < b_1$, the improvement condition in (3.1) is valid for any $\alpha \geq 1$ and all the class of elliptically symmetric distributions.

It was shown numerically by Haddouche et al. (2021, [1]) that b_1 is optimal in the Gaussian context.

Improved estimators

We study numerically the performance of

$$\hat{\Sigma}_{\alpha,b} = a_o\, H \left(I_r + b \frac{L^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} \right) H^\top \,, \tag{3.1}$$

where

$$0 \le b \le b_0 = \frac{2\,(r-1)}{v-r+1} \quad \text{and} \quad \alpha \ge 1.$$

Konno (2009, [4]) consider the case $\alpha = 1$, in the Gaussian setting and under the usual quadratic loss, for which its improvement condition is $0 \le b \le b_1 = 2\,(r-1)\,(v+r+1)/(v-r+1)\,(v-r+3)$. Although $b_0 < b_1$, the improvement condition in (3.1) is valid for any $\alpha \ge 1$ and all the class of elliptically symmetric distributions.

It was shown numerically by Haddouche et al. (2021, [1]) that b_1 is optimal in the Gaussian context.

Improved estimators

We study numerically the performance of

$$\hat{\Sigma}_{\alpha,b} = a_o\, H \left(I_r + b \frac{L^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} \right) H^\top \,, \tag{3.1}$$

where

$$0 \le b \le b_0 = \frac{2\,(r-1)}{v-r+1} \quad \text{and} \quad \alpha \ge 1.$$

Konno (2009, [4]) consider the case $\alpha = 1$, in the Gaussian setting and under the usual quadratic loss, for which its improvement condition is $0 \le b \le b_1 = 2\,(r-1)\,(v+r+1)/(v-r+1)\,(v-r+3)$. Although $b_0 < b_1$, the improvement condition in (3.1) is valid for any $\alpha \ge 1$ and all the class of elliptically symmetric distributions.

It was shown numerically by Haddouche et al. (2021, [1]) that b_1 is optimal in the Gaussian context.

Improved estimators

We study numerically the performance of

$$\hat{\Sigma}_{\alpha,b} = a_o \, H \left(I_r + b \frac{L^{-\alpha}}{\mathrm{tr}(L^{-\alpha})} \right) H^\top \, , \tag{3.1}$$

where

$$0 \le b \le b_0 = \frac{2\,(r-1)}{v-r+1} \quad \text{and} \quad \alpha \ge 1.$$

Konno (2009, [4]) consider the case $\alpha = 1$, in the Gaussian setting and under the usual quadratic loss, for which its improvement condition is $0 \le b \le b_1 = 2\,(r-1)\,(v+r+1)/(v-r+1)\,(v-r+3)$. Although $b_0 < b_1$, the improvement condition in (3.1) is valid for any $\alpha \ge 1$ and all the class of elliptically symmetric distributions.

It was shown numerically by Haddouche et al. (2021, [1]) that b_1 is optimal in the Gaussian context.

The considered structures of Σ

We consider the following structures of Σ :

- (i) the identity matrix I_p
- (ii) an autoregressive structure with coefficient 0.9 (i.e. a $p \times p$ matrix where the (i,j)th element is $0.9^{|i-j|}$).

To assess how an alternative estimator $\hat{\Sigma}_{\alpha,b}$ improves over $\hat{\Sigma}_{a_o}$, we compute the Percentage Relative Improvement in Average Loss (PRIAL) defined as

$$PRIAL(\hat{\Sigma}_{\alpha,b}) = \frac{R(\hat{\Sigma}_{a_o},\Sigma) - R(\hat{\Sigma}_{\alpha,b},\Sigma)}{R(\hat{\Sigma}_{a_o},\Sigma)} \times 100$$

The considered structures of Σ

We consider the following structures of Σ :

- (i) the identity matrix I_p
- (ii) an autoregressive structure with coefficient 0.9 (i.e. a $p \times p$ matrix where the (i,j)th element is $0.9^{|i-j|}$).

To assess how an alternative estimator $\hat{\Sigma}_{\alpha,b}$ improves over $\hat{\Sigma}_{a_o}$, we compute the Percentage Relative Improvement in Average Loss (PRIAL) defined as

$$PRIAL(\hat{\Sigma}_{\alpha,b}) = \frac{R(\hat{\Sigma}_{a_o}, \Sigma) - R(\hat{\Sigma}_{\alpha,b}, \Sigma)}{R(\hat{\Sigma}_{a_o}, \Sigma)} \times 100$$

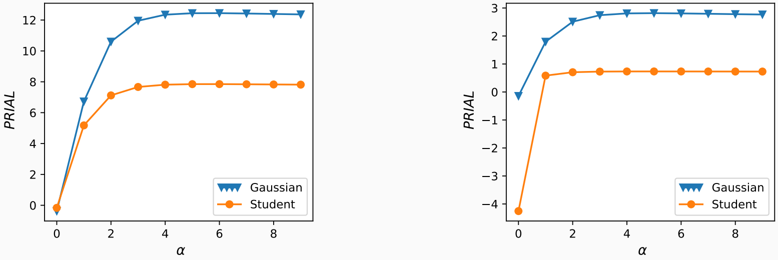
Effect of α

We study the effect of α on the prial's of the estimator $\hat{\Sigma}_{\alpha,b_0}$ over

- $\hat{\Sigma}_{a_o} = S/v$ when the sampling distribution is Gaussian,
- $\hat{\Sigma}_{a_o} = S(k-2)/vk$ when it is the t–distribution (K^* in (2.1) equals $(k-2)/k$).

We consider the non–invertible case where $p/m = c > 1$, with $(p, m) = (50, 20)$, for the structures (i) and (ii) of Σ for the t–distribution, with $k = 5$, and the Gaussian distribution

The data–based loss

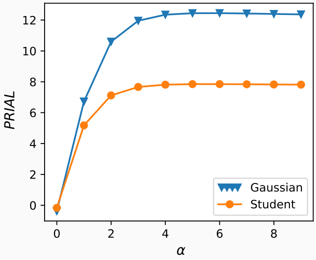


(i) (ii)

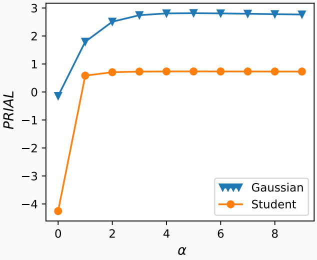
Fig. 2 – PRIAL's of $\hat{\Sigma}_{\alpha, b_0}$ in (3.1) under the data–based loss (1.6).

- For the structure (i) of Σ, note that, for $\alpha \geq 6$, the prial's stabilize at 12.5%, in the Gaussian case, and at 8.5%, in the Student case.
- Similarly, the prial's are better in the Gaussian setting for the structure (ii)
- When α is close to zero, the prial's are small for the structure (i) and may be negative for the structure (ii).

The data–based loss



(i)

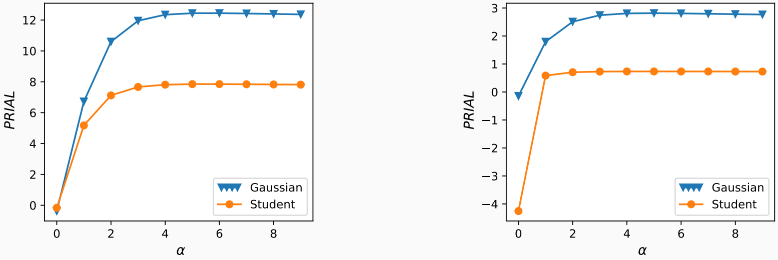


(ii)

Fig. 2 – PRIAL's of $\hat{\Sigma}_{\alpha,b_0}$ in (3.1) under the data–based loss (1.6).

- For the structure (i) of Σ, note that, for $\alpha \geq 6$, the prial's stabilize at 12.5%, in the Gaussian case, and at 8.5%, in the Student case.
- Similarly, the prial's are better in the Gaussian setting for the structure (ii)
- When α is close to zero, the prial's are small for the structure (i) and may be negative for the structure (ii).

The data–based loss

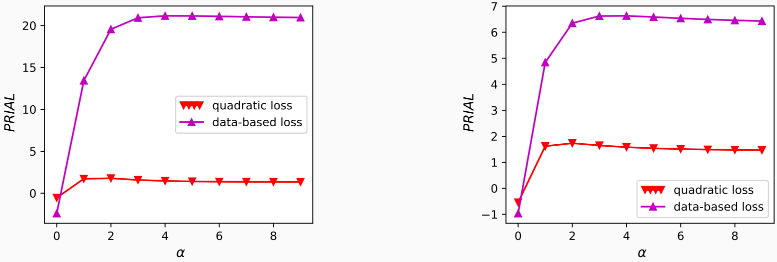


(i) (ii)

Fig. 2 – PRIAL's of $\hat{\Sigma}_{\alpha,b_0}$ in (3.1) under the data–based loss (1.6).

- For the structure (i) of Σ, note that, for $\alpha \geq 6$, the prial's stabilize at 12.5%, in the Gaussian case, and at 8.5%, in the Student case.
- Similarly, the prial's are better in the Gaussian setting for the structure (ii)
- When α is close to zero, the prial's are small for the structure (i) and may be negative for the structure (ii).

Loss functions comparison under the Gaussian assumption

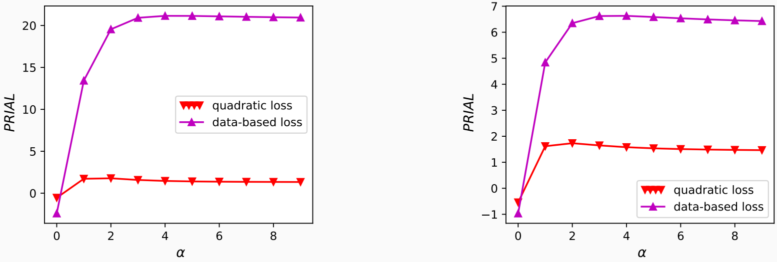


(i) **(ii)**

Fig. 3 – PRIAL's of $\hat{\Sigma}_{\alpha,b_0}$ under data–based loss and PRIAL's of $\hat{\Sigma}_{\alpha,b_1}$ under quadratic loss.

- Prial's of $\hat{\Sigma}_{\alpha,b_0}$ w.r.t $\hat{\Sigma}_{a_o} = S/v$ under the data–based loss (1.6) and the prial's of $\hat{\Sigma}_{\alpha,b_1}$ w.r.t $\hat{\Sigma}_{a_o} = S/(v+r+1)$ under the quadratic loss (1.7).
- For (i) and (ii), the prial's are better under the data–based loss.
- For the structure (i) with $\alpha = 1$ (the Konno's estimator), we observe a prial equal to 1.73% which is similar to that of [4].

Loss functions comparison under the Gaussian assumption

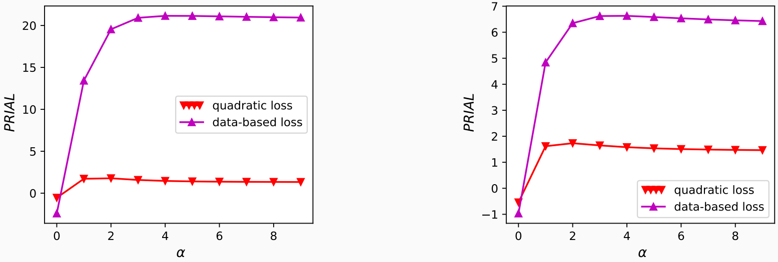


(i) **(ii)**

Fig. 3 – PRIAL's of $\hat{\Sigma}_{\alpha,b_0}$ under data–based loss and PRIAL's of $\hat{\Sigma}_{\alpha,b_1}$ under quadratic loss.

- Prial's of $\hat{\Sigma}_{\alpha,b_0}$ w.r.t $\hat{\Sigma}_{a_o} = S/v$ under the data–based loss (1.6) and the prial's of $\hat{\Sigma}_{\alpha,b_1}$ w.r.t $\hat{\Sigma}_{a_o} = S/(v+r+1)$ under the quadratic loss (1.7).
- For (i) and (ii), the prial's are better under the data–based loss.
- For the structure (i) with $\alpha = 1$ (the Konno's estimator), we observe a prial equal to 1.73% which is similar to that of [4].

Loss functions comparison under the Gaussian assumption



(i) **(ii)**

Fig. 3 – PRIAL's of $\hat{\Sigma}_{\alpha,b_0}$ under data–based loss and PRIAL's of $\hat{\Sigma}_{\alpha,b_1}$ under quadratic loss.

- Prial's of $\hat{\Sigma}_{\alpha,b_0}$ w.r.t $\hat{\Sigma}_{a_o} = S/v$ under the data–based loss (1.6) and the prial's of $\hat{\Sigma}_{\alpha,b_1}$ w.r.t $\hat{\Sigma}_{a_o} = S/(v+r+1)$ under the quadratic loss (1.7).
- For (i) and (ii), the prial's are better under the data–based loss.
- For the structure (i) with $\alpha = 1$ (the Konno's estimator), we observe a prial equal to 1.73% which is similar to that of [4].

Conclusion

Conclusion

- For a wide class of e.s.d, we provide a large class of estimators of the scale matrix Σ for the elliptical multivariate linear model (1.1) which improve over the usual estimators $a\,S$.
- The use of the data–based loss is more attractive than the use of the classical quadratic loss.
- The data–based loss brings more improved estimators and their improvement is valid within a larger class of distributions.

Conclusion

- For a wide class of e.s.d, we provide a large class of estimators of the scale matrix Σ for the elliptical multivariate linear model (1.1) which improve over the usual estimators $a\,S$.
- The use of the data–based loss is more attractive than the use of the classical quadratic loss.
- The data–based loss brings more improved estimators and their improvement is valid within a larger class of distributions.

Conclusion

- For a wide class of e.s.d, we provide a large class of estimators of the scale matrix Σ for the elliptical multivariate linear model (1.1) which improve over the usual estimators $a\,S$.
- The use of the data–based loss is more attractive than the use of the classical quadratic loss.
- The data–based loss brings more improved estimators and their improvement is valid within a larger class of distributions.

Thank you for your attention !

For references and other details, I can be reached at

mohamed.haddouche@insa-rouen.fr

References i

[1] Anis M. Haddouche, Dominique Fourdrinier, and Fatiha Mezoued.
Scale matrix estimation of an elliptically symmetric distribution in high and low dimensions.
J. Multivariate Anal., 181 :104680, 2021.

[2] L.R. Haff.
Empirical bayes estimation of the multivariate normal covariance matrix.
The Annals of Statistics, 8(3) :586–597, 1980.

[3] W. James and C. Stein.
Estimation with quadratic loss.
In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1 : Contributions to the Theory of Statistics*, pages 361–379, Berkeley, Calif., 1961. University of California Press.

References ii

[4] Y. Konno.
Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss.
Journal of Multivariate Analysis, 100(10) :2237–2253, 2009.

[5] O. Ledoit and M. Wolf.
A well-conditioned estimator for large-dimensional covariance matrices.
Journal of Multivariate Analysis, 88(2) :365 – 411, 2004.

[6] H. Tsukuma and T. Kubokawa.
***Shrinkage estimation for mean and covariance matrices*.**
JSS research series in statistics. Springer, Singapore, 2020.