Comparison of Linear Shrinkage Estimators of a Large Covariance Matrix in Normal and Non-normal Distributions

Yuki Ikeda
Graduate School of Economics, The University of Tokyo

Tatsuya Kubokawa
The University of Tokyo

Muni S. Srivastava
University of Toronto

March 2015

Comparison of Linear Shrinkage Estimators of a Large Covariance Matrix in Normal and Non-normal Distributions

Yuki Ikeda[*], Tatsuya Kubokawa[†]and Muni S. Srivastava[‡]

University of Tokyo and University of Toronto

Abstract

The problem of estimating the large covariance matrix of both normal and non-normal distributions is addressed. In convex combinations of the sample covariance matrix and the identity matrix multiplied by a scalor statistic, we suggest a new estimator of the optimal weight based on exact or approximately unbiased estimators of the numerator and denominator of the optimal weight in non-normal cases. It is also demonstrated that the estimators given in the literature have second-order biases. It is numerically shown that the proposed estimator has a good risk performance.

Key words and phrases: Covariance matrix, high dimension, large sample, non-normal distribution, normal distribution, linear shrinkage estimator, risk function, shrinkage.

1 Introduction

Many applied problems in multivariate analysis require estimates of a covariance matrix and/or of its inverse. For example, the inverse of estimators of the covariance matrix is used in the Fisher linear discriminant analysis, confidence intervals based on the Mahalanobis distance and generalized least squares estimators in multivariate linear regression models. However, the unbiased estimator based on the sample covariance matrix is not invertible when the dimension p of the variables is larger than the sample size N. When p is large and close to N, the inverse of the unbiased estimator may be ill-conditioned

[*]Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: pt2y1003@gmail.com

[†]Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN. E-Mail: tatsuya@e.u-tokyo.ac.jp

[‡]Department of Statistics, University of Toronto, 100 St George Street, Toronto, Ontario, CANADA M5S 3G3, E-Mail: srivasta@utstat.toronto.edu

even if $N > p$. Thus, an estimator for the covariance matrix is required to be both invertible and well-conditioned. Many approaches to this goal have been considered in the literature. Among them, here we focus on linear shrinkage estimators such as Daniels and Kass (2001), Ledoit and Wolf (2003, 2004), Schafer and Strimmer (2005), Srivastava and Kubokawa (2007), Konno (2009), Chen, Wiesel, Eldar and Hero (2010), Fisher and Sun (2011) and Bai and Shi (2011). Their ideas are to shrink the sample covariance matrix in the direction of more stable target such as a diagonal matrix. Therefore it is crucial that to what degree the sample covariance should be shrunk, which is called a shrinkage intensity and it corresponds the weight appeared in linear shrinkage estimators. In this paper, we suggest a reasonable weight in the linear shrinkage estimator under general distributions and confirm the numerical performances.

To specify the problem, consider p-dimensional random vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ which are mutually independently and identically distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then, $\boldsymbol{\Sigma}$ is estimated unbiasedly by

$$\boldsymbol{S} = \frac{1}{n} \sum_{j=1}^{N} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_j - \overline{\boldsymbol{x}})^t, \tag{1.1}$$

for $n = N - 1$ and $\overline{\boldsymbol{x}} = N^{-1} \sum_{j=1}^{N} \boldsymbol{x}_j$. A linear shrinkage estimator we treat here is of the form

$$\widehat{\boldsymbol{\Sigma}}_w = w\boldsymbol{S} + (1 - w)\hat{a}_1 \boldsymbol{I}_p, \quad \hat{a}_1 = \operatorname{tr}[\boldsymbol{S}]/p,$$

for a constant w satisfying $0 \leq w \leq 1$. This shrinks $\boldsymbol{S}$ toward the target $\hat{a}_1 \boldsymbol{I}_p$.

The weight w is estimated based on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, and the performance of the linear shrinkage estimator depends on an estimator of w. We describe several estimators of w which have been suggested in the literature. Based on the optimal weight w in the sense of minimizing the risk function $E[\operatorname{tr}[(\boldsymbol{S}_w - \boldsymbol{\Sigma})^2]]$, Ledoit and Wolf (2004) suggested to estimate w by

$$\hat{w}_{LW} = 1 - \frac{\sum_{j=1}^{N} \operatorname{tr}[\{(\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_j - \overline{\boldsymbol{x}})^t - \boldsymbol{S}\}^2]}{n^2(\operatorname{tr}\boldsymbol{S}^2 - (\operatorname{tr}\boldsymbol{S})^2/p)}, \tag{1.2}$$

which yields the plug-in linear shrinkage estimator

$$\widehat{\boldsymbol{\Sigma}}_{LW} = \hat{w}_{LW}\boldsymbol{S} + (1 - \hat{w}_{LW})\hat{a}_1 \boldsymbol{I}_p.$$

Chen, Wiesel, Eldar and Hero (2010) considered to improve $\widehat{\boldsymbol{\Sigma}}_{LW}$ by the Rao-Blackwell theorem in the normal distribution. The Rao-Blackwell Ledoit-Wolf (RBLW) estimator is given by $\widehat{\boldsymbol{\Sigma}}_{RBLW} = E[\widehat{\boldsymbol{\Sigma}}_{LW}|\boldsymbol{S}] = \hat{w}_{RBLW}\boldsymbol{S} + (1 - \hat{w}_{RBLW})\hat{a}_1 \boldsymbol{I}_p$ for

$$\hat{w}_{RBLW} = \max\left\{0, 1 - \frac{(n-2)\operatorname{tr}(\boldsymbol{S}^2) + n(\operatorname{tr}\boldsymbol{S})^2}{n(n+2)(\operatorname{tr}\boldsymbol{S}^2 - (\operatorname{tr}\boldsymbol{S})^2/p)}\right\}. \tag{1.3}$$

2

Chen, Wiesel, Eldar and Hero (2010) also suggested another estimator, called the Oracle-Approximating Shrinkage (OAS) estimator, given by $\widehat{\boldsymbol{\Sigma}}_{OAS} = \hat{w}_{OAS}\boldsymbol{S} + (1 - \hat{w}_{OAS})\hat{a}_1\boldsymbol{I}_p$ for

$$\hat{w}_{OAS} = \max\left\{0, 1 - \frac{(p-2)\mathrm{tr}\,(\boldsymbol{S}^2) + p(\mathrm{tr}\,\boldsymbol{S})^2}{p(n+1-2/p)(\mathrm{tr}\,\boldsymbol{S}^2 - (\mathrm{tr}\,\boldsymbol{S})^2/p)}\right\}. \tag{1.4}$$

Fisher and Sun (2011) proposed the linear shrinkage estimator $\widehat{\boldsymbol{\Sigma}}_{FS} = \hat{w}_{FS}\boldsymbol{S} + (1 - \hat{w}_{FS})\hat{a}_1\boldsymbol{I}_p$ for

$$\hat{w}_{FS} = \frac{n(\hat{a}_2 - \hat{a}_1^2)}{n(\hat{a}_2 - \hat{a}_1^2) + p\hat{a}_1^2 + \hat{a}_2}, \tag{1.5}$$

where

$$\hat{a}_2 = \frac{n^2}{(n-1)(n+2)}\left[\frac{1}{p}\mathrm{tr}\,[\boldsymbol{S}^2] - \frac{p}{n}\hat{a}_1^2\right]. \tag{1.6}$$

As shown in Section 3, the OAS estimator $\widehat{\boldsymbol{\Sigma}}_{OAS}$ has almost the same form as the $\widehat{\boldsymbol{\Sigma}}_{FS}$ estimator of Fisher and Sun (2011) and this fact can be confirmed numerically in Section 4.

In the OAS estimator and the Fisher-Sun estimator, the estimator $\hat{a}_2$ is used for a_2. In the normal distribution, Srivastava (2005) showed that $\hat{a}_2$ is unbiased and that $\hat{a}_2 - a_2 = O_p((np)^{-1/2}) + O_p(n^{-1})$. That is, $\hat{a}_2$ is a good estimator of a_2 in the normal case. In the non-normal distributions, however, $\hat{a}_2$ is not unbiased as shown in Srivastava, Kollo and von Rosen (2011) and Srivastava, Yanagihara and Kubokawa (2014). This affects the performance of $\widehat{\boldsymbol{\Sigma}}_{OAS}$ and $\widehat{\boldsymbol{\Sigma}}_{FS}$. In fact, Chen, Wiesel, Eldar and Hero (2010) pointed out that the performance of their estimators is worse than $\widehat{\boldsymbol{\Sigma}}_{LW}$ in non-normal cases.

In this paper, we address the problem of improving the performance of the linear shrinkage estimators in non-normal distributions. To this end, we use the estimator

$$\hat{a}_{2C} = \frac{n}{(n+1)(n-1)(n-2)p}\{n(n-1)\mathrm{tr}\,\boldsymbol{S}^2 + (\mathrm{tr}\,\boldsymbol{S})^2 - (n+1)Q\} \tag{1.7}$$

instead of $\hat{a}_2$, where

$$Q = \frac{1}{n}\sum_{i=1}^{N}\{(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^t(\boldsymbol{x}_i - \overline{\boldsymbol{x}})\}^2. \tag{1.8}$$

As shown in Srivastava, Yanagihara and Kubokawa (2014) and Himeno and Yamada (2014), the estimator $\hat{a}_{2C}$ is an unbiased estimator of a_2 in non-normal distributions. Himeno and Yamada (2014) indicated the interesting fact that $\hat{a}_{2C}$ is of the same form as the estimator given in Chen, Zhang and Zhong (2010). Although the estimator of Chen, Zhang and Zhong (2010) is known to be computationally hard, the expression in $\hat{a}_{2C}$ is

simple and easily implementable. In Section 2, based on the optimal weight w in linear shrinkage estimators, we suggest the estimator of w, given by

$$\hat{w}_U = \frac{\hat{a}_{2C} - \hat{a}_1^2}{\operatorname{tr} \boldsymbol{S}^2/p - \hat{a}_1^2},$$

and the plug-in estimator $\widehat{\boldsymbol{\Sigma}}_U = \hat{w}_U \boldsymbol{S} + (1 - \hat{w}_U)\hat{a}_1 \boldsymbol{I}_p$. A motivation for this estimator and some asymptotic properties are provided in Section 2.

In Section 3, we compare the suggested weight function $\hat{w}_U$ with the others $\hat{w}_{LW}$, $\hat{w}_{RBLW}$, $\hat{w}_{OAS}$ and $\hat{w}_{FS}$ by deriving the second-order biases of their numerators. It is also seen that the OAS estimator in Chen, Wiesel, Eldar and Hero (2010) is close to the estimator in Fisher and Sun (2011).

In Section 4, we investigate the risk performance of the above estimators through simulation, and it is numerically shown that the estimator $\widehat{\boldsymbol{\Sigma}}_U$ is superior in the cases of non-normal distributions, while it has a bit small loss from some of the other estimators in the normal distributions, but the difference is quite small. The numerical results confirm the analytical results given in Section 3 that the OAS estimator is almost identical to the estimator in Fisher and Sun (2011). Concluding remarks are given in Section 5.

2   Estimation of the Optimal Weight under Non-normal Distributions

Consider p-dimensional random vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ which are mutually independently and identically distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Sigma}^{1/2})^t$, where $\boldsymbol{\Sigma}^{1/2}$ is the Cholesky decomposition with positive diagonal elements. Assume that the observation vectors $\boldsymbol{x}_j$ are generated as

$$\boldsymbol{x}_j = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{u}_j, \quad j = 1, \ldots, N, \tag{2.1}$$

with

$$E(\boldsymbol{u}_j) = \boldsymbol{0}, \ \ \mathbf{Cov}\,(\boldsymbol{u}_j) = \boldsymbol{I}_p, \tag{2.2}$$

and for integers $\gamma_1, \ldots, \gamma_k$ satisfying $0 \le \sum_{k=1}^{p} \gamma_k \le 4$,

$$E\left[\prod_{k=1}^{p} u_{jk}^{\gamma_k}\right] = \prod_{k=1}^{p} E(u_{jk}^{\gamma_k}), \quad j = 1, \ldots, N, \tag{2.3}$$

where u_{jk} is the kth component of the vector $\boldsymbol{u}_j = (u_{j1}, \ldots, u_{jk}, \ldots, u_{jp})^t$. This assumption is requested for existence of the risk function. We shall write the third and fourth moments of u_{jk} as $E[u_{jk}^3] = K_3$ and $E[u_{jk}^4] = K_4 + 3$. In the case of a normal distribution, we have $K_3 = K_4 = 0$. We use the notations

$$a_1 = \operatorname{tr}[\boldsymbol{\Sigma}]/p, \ a_2 = \operatorname{tr}[\boldsymbol{\Sigma}^2]/p \text{ and } a_{20} = \sum_{i=1}^{p} \sigma_{ii}^2/p,$$

4

for $\boldsymbol{\Sigma} = (\sigma_{ij})$.

We consider the problem of estimating $\boldsymbol{\Sigma}$ by an estimator $\widehat{\boldsymbol{\Sigma}}$ relative to the quadratic loss function $L_q(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \text{tr}\,[(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})^2]$. Estimator $\widehat{\boldsymbol{\Sigma}}$ is evaluated in terms of the risk function $R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) = E[\text{tr}\,[(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})^2]]$. An unbiased estimator of $\boldsymbol{\Sigma}$ is $\boldsymbol{S}$, but is not invertible in the case of $p > n$ nor well-conditioned in the case that p is close to n even if $n > p$. Thus, it is reasonable to consider convex combinations of $\boldsymbol{S}$ and a positive definite matrix based on $\boldsymbol{S}$. Under the assumption that $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_p$, an unbiased estimator of σ^2 is $\hat{a}_1 = \text{tr}\,[\boldsymbol{S}]/p$. Then, we consider a class of linear shrinkage estimators $\widehat{\boldsymbol{\Sigma}}_w = w\boldsymbol{S} + (1-w)\hat{a}_1 \boldsymbol{I}_p$ for $0 \le w \le 1$. A direct calculation shows that

$$R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_w) = E[\text{tr}\,[(\boldsymbol{S} - \hat{a}_1 \boldsymbol{I}_p)^2]w^2 - 2\text{tr}\,[(\boldsymbol{S} - \hat{a}_1 \boldsymbol{I}_p)(\boldsymbol{\Sigma} - \hat{a}_1 \boldsymbol{I}_p)]w + \text{tr}\,[(\boldsymbol{\Sigma} - \hat{a}_1 \boldsymbol{I}_p)^2]],$$

which implies that the optimal weight w minimizing the risk is

$$w^* = \frac{E\text{tr}\,[(\boldsymbol{S} - \hat{a}_1 \boldsymbol{I}_p)(\boldsymbol{\Sigma} - \hat{a}_1 \boldsymbol{I}_p)]}{E\text{tr}\,[(\boldsymbol{S} - \hat{a}_1 \boldsymbol{I}_p)^2]}.$$

Since $E(\boldsymbol{S}) = \boldsymbol{\Sigma}$ and $E(\hat{a}_1) = a_1$ for any underlying distributions, w^* is simplified as

$$w^* = \frac{E\text{tr}\,[(\boldsymbol{S} - \hat{a}_1 \boldsymbol{I}_p)\boldsymbol{\Sigma}]}{E\text{tr}\,[(\boldsymbol{S} - \hat{a}_1 \boldsymbol{I}_p)^2]} = \frac{a_2 - a_1^2}{E\text{tr}\,[(\boldsymbol{S} - \hat{a}_1 \boldsymbol{I}_p)^2]/p}. \tag{2.4}$$

Since w^* is a function of $\boldsymbol{\Sigma}$, we need to estimate w^* based on $\boldsymbol{S}$. An idea is to provide an unbiased estimator of $\boldsymbol{w}^*$. However, it is not easy to obain such an estimator. Thus, in this paper, we consider to estimate the numerator $a_2 - a_1^2$ and the denominator $E\text{tr}\,[(\boldsymbol{S} - \hat{a}_1 \boldsymbol{I}_p)^2]/p$ unbiasedly. For the denominator, $\text{tr}\,[(\boldsymbol{S} - \hat{a}_1 \boldsymbol{I}_p)^2]/p = \text{tr}\,\boldsymbol{S}^2/p - \hat{a}_1^2$, and clearly it is an unbiased estimator of $E\text{tr}\,[(\boldsymbol{S} - \hat{a}_1 \boldsymbol{I}_p)^2]/p$ for any distribution. We next want to find an unbiased or approximately unbiased estimator of $a_2 - a_1^2$.

Let $\hat{a}_2$ be defined in (1.6). In the case of normal distributions, Srivastava (2005) showed that $\hat{a}_2$ is an unbiased estimator of a_2. Under non-normality, however, Srivastava, *et al.* (2014) demonstrated that

$$E[\hat{a}_2] = a_2 + \frac{n}{(n+1)(n+2)}K_4 a_{20}.$$

This implies that the estimator $\hat{a}_2$ has a second-order bias in non-normal distributions when $a_{20} = O(1)$. Srivastava, *et al.* (2014) and Himeno and Yamada (2014) suggested the unbiased estimator $\hat{a}_{2C}$ given in (1.7) for a_2, which is rewritten as

$$\hat{a}_{2C} = \frac{n}{(n+1)(n-1)(n-2)}\{n(n-1)\text{tr}\,\boldsymbol{S}^2/p + p\hat{a}_1^2 - (n+1)Q/p\}.$$

Although they derived the estimator independently, we can verify that their proposals are identical. Himeno and Yamada (2014) also demonstrated that the estimator $\hat{a}_{2C}$ is identical to the one given in Chen, Zhang and Zhong (2010). Compared with the

form given in Chen, Zhang and Zhong (2010), the expression in $\hat{a}_{2C}$ is simple and easily implementable.

Concerning the estimation of a_1^2, Srivastava, *et al.* (2011) showed that

$$Var(\hat{a}_1) = \frac{2}{np}a_2 + \frac{1}{(n+1)p}K_4 a_{20}, \tag{2.5}$$

in non-normal cases. Thus,

$$E[\hat{a}_1^2] = a_1^2 + \frac{2}{np}a_2 + \frac{1}{(n+1)p}K_4 a_{20}, \tag{2.6}$$

which implies that $E[\hat{a}_1^2] = a_1^2 + O((np)^{-1})$ when $a_2 = O(1)$ and $a_{20} = O(1)$. In the non-sparse case of $a_2 = O(p)$, $\hat{a}_1^2$ has a second-order bias, and we should use the exact unbiased estimator

$$\widehat{a_{1C}^2} = \frac{n}{(n+1)(n-1)(n-2)}\left\{\frac{2}{p^2}\operatorname{tr} \boldsymbol{S}^2 + (n^2 - n - 1)\hat{a}_1^2 - \frac{n+1}{p^2}Q\right\},$$

which was derived in Himeno and Yamada (2014). However, the two estimators $\hat{a}_1^2$ and $\widehat{a_{1C}^2}$ give little difference in numerical risk performances in most settings. Hence, we use the simpler estimator $\hat{a}_1^2$. Thus, the numerator $a_2 - a_1^2$ in w^* is estimated by $\hat{a}_{2C} - \hat{a}_1^2$, and we can suggest the estimator

$$\widehat{\boldsymbol{\Sigma}}_U = \hat{w}_U \boldsymbol{S} + (1 - \hat{w}_U)\hat{a}_1 \boldsymbol{I}_p, \tag{2.7}$$

for

$$\hat{w}_U = \frac{\hat{a}_{2C} - \hat{a}_1^2}{\operatorname{tr} \boldsymbol{S}^2 / p - \hat{a}_1^2} = \frac{\hat{a}_{2C} - \hat{a}_1^2}{\hat{a}_2 - \hat{a}_1^2 + p\hat{a}_1^2/n + (n-2)\hat{a}_2/n^2}, \tag{2.8}$$

where the second equality follows from

$$\operatorname{tr} \boldsymbol{S}^2 / p = \frac{(n-1)(n+2)}{n^2}\hat{a}_2 + \frac{p}{n}\hat{a}_1^2. \tag{2.9}$$

We here give an exact expression of w^*. It follows from Himeno and Yamada (2014) and Lemma 7.1 in Srivastava, *et al.* (2014) that

$$\begin{aligned}
E[\operatorname{tr} \boldsymbol{S}^2 / p] &= a_2 + \frac{1}{n}a_2 + \frac{p}{n}a_1^2 + \frac{1}{n+1}K_4 a_{20}, \\
E[Q/p] &= \frac{n^2 - n + 1}{(n+1)^2}K_4 a_{20} + \frac{2n}{n+1}a_2 + \frac{np}{n+1}a_1^2.
\end{aligned} \tag{2.10}$$

Using (2.6) and (2.10), we can see that

$$E[\operatorname{tr} \boldsymbol{S}^2 / p - \hat{a}_1^2] = a_2 - a_1^2 + \frac{p}{n}a_1^2 + \frac{p-2}{np}a_2 + \frac{p-1}{(n+1)p}K_4 a_{20},$$

6

which means that the optimal weight is expressed as

$$w^* = \frac{a_2 - a_1^2}{a_2 - a_1^2 + pa_1^2/n + (p-2)a_2/(np) + (p-1)K_4 a_{20}/(Np)}. \quad (2.11)$$

In the case of normal distributions, $w^* = (a_2 - a_1^2)/\{a_2 - a_1^2 + pa_1^2/n + (p-2)a_2/(np)\}$ and it is estimated by $(\hat{a}_2 - \hat{a}_1^2)/\{\hat{a}_2 - \hat{a}_1^2 + p\hat{a}_1^2/n + (p-2)\hat{a}_2/(np)\}$. Comparing this estimator with (2.8) in non-normal distributions, we can see that there is a second-order bias in the numerator, while there is no bias at all in the denominator.

When w^* is close to one, the linear shrinkage estimator approches to the sample covariance matrix $\boldsymbol{S}$, namely it may be ill-conditioned. Thus, it is important to investigate the limit value of w^*.

Proposition 2.1 *For large p, assume that $a_1 = O(1)$, $a_{20} = O(1)$ and $a_2 = O(p^\delta)$ for $\delta \geq 0$.*
   (Case 1) If $na_2/p \to \infty$, then $w^* \to 1$.
   (Case 2) If $na_2/p \to 0$, then $w^* \to 0$.
   (Case 3) If $na_2/p \to C$ for positive constant C, then w^* approaches to

$$\left(C - \frac{n}{p}a_1^2\right) \Big/ \left(C - \frac{n}{p}a_1^2 + a_1^2 + \frac{1}{n}C\right).$$

We can verify Proposition 2.1 from the expression

$$w^* = \left(\frac{na_2}{p} - \frac{n}{p}a_1^2\right) \Big/ \left(\frac{na_2}{p} - \frac{n}{p}a_1^2 + a_1^2 + \frac{p-2}{pn}\frac{na_2}{p} + \frac{n(p-1)}{Np^2}K_4 a_{20}\right).$$

We get some implications from Proposition 2.1. First, consider the case of $a_2 = O(1)$, which means that $\boldsymbol{\Sigma}$ is sparce or not dense. Then, Cases 1, 2 and 3 correspond to
   (Case 1) $n \to \infty$ and $p = O(n^\delta)$ for constant δ satisfying $0 \leq \delta < 1$,
   (Case 2) $p \to \infty$ and $n = O(p^\delta)$ for constant δ satisfying $0 \leq \delta < 1$,
   (Case 3) $(p, n) \to \infty$ and $n/p \to \gamma$ for constant γ.
   Second, consider the case of $a_2 = O(p)$, which means that $\boldsymbol{\Sigma}$ is non-sparce or dense. Then, Cases 1, 2 and 3 correspond to
   (Case 1) $n \to \infty$,
   (Case 2) NA,
   (Case 3) n is bounded.
   When $p \to \infty$, but n is bounded, we have $w^* \to 0$ for $a_2 = O(1)$, and $w^* \to C/(C + a_1^2 + n^{-1}C)$ for $a_2 = O(p)$. When $n \to \infty$, but p is bounded, we have always $w^* \to 1$. When $(p, n) \to \infty$ and $n/p \to \gamma$ for constant γ, we have

$$w^* \to \begin{cases} \gamma(a_2 - a_1^2)/\{\gamma(a_2 - a_1^2) + a_1^2\} & \text{for } a_2 = O(1), \\ 1 & \text{for } a_2 = O(p). \end{cases}$$

As in Ledoit and Wolf (2004), we define the percentage relative improvement in average loss (PRIAL) over the sample covariance as

$$\text{PRIAL} = \frac{E[\text{tr}\,(\boldsymbol{S} - \boldsymbol{\Sigma})^2] - E[\text{tr}\,(\widehat{\boldsymbol{\Sigma}}_{w^*} - \boldsymbol{\Sigma})^2]}{E[\text{tr}\,(\boldsymbol{S} - \boldsymbol{\Sigma})^2]}.$$

Then the following proposition can be established similarly to Proposition 2.1.

Proposition 2.2 *For large p, assume that $a_1 = O(1)$, $a_{20} = O(1)$ and $a_2 = O(p^\delta)$ for $\delta \geq 0$.*
   (Case 1) If $na_2/p \to \infty$, then PRIAL$\to 0$.
   (Case 2) If $na_2/p \to 0$, then PRIAL$\to 1$.
   (Case 3) If $na_2/p \to C$ for positive constant C, then PRIAL approaches to

$$\left(a_1^2 + \frac{1}{n}C\right) \Big/ \left(C - \frac{n}{p}a_1^2 + a_1^2 + \frac{1}{n}C\right).$$

Proof Since $E[\text{tr}\,(\widehat{\boldsymbol{\Sigma}}_{w^*} - \boldsymbol{\Sigma})^2] = E[\text{tr}\,(\boldsymbol{S} - \boldsymbol{\Sigma})^2 - 2(1-w^*)\text{tr}\,(\boldsymbol{S} - \boldsymbol{\Sigma})(\boldsymbol{S} - \hat{a}_1\boldsymbol{I}_p) + (1-w^*)^2\text{tr}\,(\boldsymbol{S} - \hat{a}_1\boldsymbol{I}_p)^2]$, PRIAL is written as

$$\text{PRIAL} = (1 - w^*)\frac{2E[\text{tr}\,(\boldsymbol{S} - \boldsymbol{\Sigma})(\boldsymbol{S} - \hat{a}_1\boldsymbol{I}_p)] - (1 - w^*)E\text{tr}\,(\boldsymbol{S} - \hat{a}_1\boldsymbol{I}_p)^2]}{E[\text{tr}\,(\boldsymbol{S} - \boldsymbol{\Sigma})^2]}.$$

Since $w^* = E\text{tr}\,[(\boldsymbol{S} - \hat{a}_1\boldsymbol{I}_p)\boldsymbol{\Sigma}]/E\text{tr}\,[(\boldsymbol{S} - \hat{a}_1\boldsymbol{I}_p)^2]$ from (2.4), PRIAL can be further rewritten as

$$\text{PRIAL} = \frac{\{E[\text{tr}\,(\boldsymbol{S} - \boldsymbol{\Sigma})(\boldsymbol{S} - \hat{a}_1\boldsymbol{I}_p)]\}^2}{E[\text{tr}\,(\boldsymbol{S} - \boldsymbol{\Sigma})^2]E[\text{tr}\,(\boldsymbol{S} - \hat{a}_1\boldsymbol{I}_p)^2]}.$$

Note that

$$E[\text{tr}\,(\boldsymbol{S} - \boldsymbol{\Sigma})^2]/p = \frac{1}{n}a_2 + \frac{p}{n}a_1^2 + \frac{1}{n+1}K_4 a_{20},$$

$$E[\text{tr}\,(\boldsymbol{S} - \hat{a}_1\boldsymbol{I}_p)^2]/p = a_2 - a_1^2 + \frac{p}{n}a_1^2 + \frac{p-2}{np}a_2 + \frac{p-1}{(n+1)p}K_4 a_{20},$$

$$E[\text{tr}\,(\boldsymbol{S} - \boldsymbol{\Sigma})(\boldsymbol{S} - \hat{a}_1\boldsymbol{I}_p)]/p = \frac{p}{n}a_1^2 + \frac{p-2}{np}a_2 + \frac{p-1}{(n+1)p}K_4 a_{20}.$$

Then, PRIAL is expressed as

$$\frac{\{a_1^2 + (1 - \frac{2}{p})\frac{1}{n}(\frac{n}{p}a_2) + \frac{n(p-1)}{(n+1)p}\frac{1}{p}K_4 a_{20}\}^2}{\{\frac{1}{n}(\frac{n}{p}a_2) + a_1^2 + \frac{n}{n+1}\frac{1}{p}K_4 a_{20}\}\{(\frac{n}{p}a_2) - \frac{n}{p}a_1^2 + a_1^2 + (1 - \frac{2}{p})\frac{1}{n}(\frac{n}{p}a_2) + \frac{n(p-1)}{(n+1)p}\frac{1}{p}K_4 a_{20}\}}.$$

In this expression, we can obtain the limiting values of PRIAL for the three cases. For example, (Case 3) implies that $p \to \infty$, so that it follows from the above expression that PRIAL approaches to

$$\frac{\{a_1^2 + \frac{1}{n}C\}^2}{\{\frac{1}{n}C + a_1^2\}\{C - \frac{n}{p}a_1^2 + a_1^2 + \frac{1}{n}C\}},$$

which yields the result in (Case 3). The other cases can be easily verified. $\square$

8

3 Comparison of the Estimators of the Optimal Weight

In this section, we compare the suggested estimator of the weight w with the estimators given in the literature. The major difference between $\hat{w}_U$ and the other estimators is that the numerator $\hat{a}_{2C} - \hat{a}_1^2$ in $\hat{w}_U = (\hat{a}_{2C} - \hat{a}_1^2)/(\text{tr}\,\boldsymbol{S}^2/p - \hat{a}_1^2)$ is a second-order unbiased estimator of $a_2 - a_1^2$ when $a_2 = O(1)$, namely,

$$E[\hat{a}_{2C} - \hat{a}_1^2] = a_2 - a_1^2 + O((np)^{-1}).$$

In what follows, we compare the estimators given in Section 1 in light of the expectations of their numerators under the assumption that $a_i = O(1)$, $i = 1, 2$, and $a_{20} = O(1)$.

For the weight function $\hat{w}_{LW}$ of Ledoit and Wolf (2004) given in (1.2), it can be rewritten as

$$\hat{w}_{LW} = \left\{ \left(1 + \frac{n-1}{n^2} \right) \text{tr}\,\boldsymbol{S}^2/p - \hat{a}_1^2 - \frac{1}{np}Q \right\} / \left(\text{tr}\,\boldsymbol{S}^2/p - \hat{a}_1^2 \right).$$

Using the moments given in (2.6) and (2.10), we can evaluate the expectation of the numerator as

$$E\left[\left(1 + \frac{n-1}{n^2} \right) \text{tr}\,\boldsymbol{S}^2/p - \hat{a}_1^2 - \frac{1}{np}Q \right]$$
$$= a_2 - a_1^2 + \frac{p}{n^2}a_1^2 - \frac{2}{np}a_2 - \frac{1}{np}K_4 a_{20}$$
$$= a_2 - a_1^2 + \frac{p}{n^2}a_1^2 + O((np)^{-1}).$$

The leading term of the bias $(p/n^2)a_1^2$ is enhanced in the case that n is small, but p is large.

For the weight $\hat{w}_{RBLW}$ of Chen, Wiesel, Eldar and Hero (2010) given in (1.3), it can be rewritten as $\hat{w}_{RBLW} = \max(0, w_{RBLW}^*)$, where

$$w_{RBLW}^* = \left\{ \left(1 - \frac{n-2}{n(n+2)} \right) \text{tr}\,\boldsymbol{S}^2/p - \left(1 + \frac{p}{n+2} \right) \hat{a}_1^2 \right\} / \left(\text{tr}\,\boldsymbol{S}^2/p - \hat{a}_1^2 \right).$$

The expectation of the numerator is evaluated as

$$E\left[\left(1 - \frac{n-2}{n(n+2)} \right) \text{tr}\,\boldsymbol{S}^2/p - \left(1 + \frac{p}{n+2} \right) \hat{a}_1^2 \right]$$
$$= a_2 - a_1^2 + \frac{p}{n^2}a_1^2 + \frac{1}{n}K_4 a_{20} + O((np)^{-1}) + O(n^{-2}),$$

which implies that the leading term of the bias $(p/n^2)a_1^2 + n^{-1}K_4 a_{20}$ may affect the performance for small n.

For the weight $\hat{w}_{OAS}$ of Chen, Wiesel, Eldar and Hero (2010) given in (1.4), it can be rewritten as $\hat{w}_{OAS} = \max(0, w_{OAS}^*)$, where

$$w_{OAS}^* = \left\{ \left(1 - \frac{p-2}{p(n+1-2/p)} \right) \text{tr}\,\boldsymbol{S}^2/p - \left(1 + \frac{p}{n+1-2/p} \right) \hat{a}_1^2 \right\} / \left(\text{tr}\,\boldsymbol{S}^2/p - \hat{a}_1^2 \right).$$

The expectation of the numerator is evaluated as

$$E\left[\left(1 - \frac{p-2}{p(n+1-2/p)}\right)\operatorname{tr}\boldsymbol{S}^2/p - \left(1 + \frac{p}{n+1-2/p}\right)\hat{a}_1^2\right]$$
$$= a_2 - a_1^2 + \frac{1}{n}K_4 a_{20} + O((np)^{-1}) + O(n^{-2}),$$

which implies that the leading term of the bias $n^{-1}K_4 a_{20}$ may affect the performance for small n in non-normal distributions.

For the weight $\hat{w}_{FS}$ of Fisher and Sun (2011) given in (1.5), from (2.9), it can be rewritten as

$$\hat{w}_{FS} = \frac{n(\hat{a}_2 - \hat{a}_1^2)}{n(\hat{a}_2 - \hat{a}_1^2) + p\hat{a}_1^2 + \hat{a}_2} = \frac{\hat{a}_2 - \hat{a}_1^2}{\hat{a}_2 - \hat{a}_1^2 + (p/n)\hat{a}_1^2 + \hat{a}_2/n},$$

where the denomenator is rewritten as $\hat{a}_2 - \hat{a}_1^2 + (p/n)\hat{a}_1^2 + \hat{a}_2/n = \operatorname{tr}\boldsymbol{S}^2/p - \hat{a}_1^2 + 2\hat{a}_2/n^2$, which is close to $\operatorname{tr}\boldsymbol{S}^2/p - \hat{a}_1^2$, the denomenator of the other estimators of the weight. The expectation of the numerator is evaluated as

$$E[\hat{a}_2 - \hat{a}_1^2] = a_2 - a_1^2 + \frac{1}{n}K_4 a_{20} + O((np)^{-1}) + O(n^{-2}),$$

which implies that the leading term of the bias $n^{-1}K_4 a_{20}$ may affect the performance for small n in non-normal distributions.

In the above arguments, it is noted that the expectations of the numerators in $\hat{w}_{OAS}$ and $\hat{w}_{FS}$ are identical up to $O(n^{-1})$. Then, it is expected that the OAS estimator of Chen, Wiesel, Eldar and Hero (2010) is asymptotically close to the estimator of Fisher and Sun (2011). In fact, it follows from (1.6) that $\hat{w}_{OAS}$ can be rewritten as

$$\hat{w}_{OAS} = \min\left(1, \frac{c_1\hat{a}_2 - n\hat{a}_1^2}{c_2\hat{a}_2 + (p-n)\hat{a}_1^2}\right),$$

where $c_1 = (n-1)(n+2)/(n+1-2/p)$ and $c_2 = (n-1)(n+2)/n$. Since $c_1 = n + O(n^{-1}) + O(p^{-1})$ and $c_2 = n + 1 + O(n^{-1})$, it is seen that

$$\hat{w}_{OAS} = \min\left(1, \frac{(n + O(n^{-1}) + O(p^{-1}))\hat{a}_2 - n\hat{a}_1^2}{(n + 1 + O(n^{-1}))\hat{a}_2 + (p-n)\hat{a}_1^2}\right)$$
$$= \min\left(1, \hat{w}_{FS} + O_p(n^{-1}p^{-1}) + O_p(n^{-2})\right),$$

when $a_2 = O(1)$. This demonstrates that $\hat{w}_{OAS}$ and $\hat{w}_{FS}$ are almost equivalent.

The numerators of the weights in $\widehat{\boldsymbol{\Sigma}}_{FS}$ and $\boldsymbol{\Sigma}_{OAS}$ are unbiased in the normal case, but not unbiased in non-normal cases. As illustrated in the next section, this property affects the risk performance of $\widehat{\boldsymbol{\Sigma}}_{FS}$ and $\boldsymbol{\Sigma}_{OAS}$.

4　Simulation Studies

We now investigate the numerical performances of the risk functions of the ridge-type estimators through simulation.

As a structure of the covariance matrix, we follow Chen, Wiesel, Eldar and Hero (2010). Namely we set

(Model 1) $\boldsymbol{\Sigma}_{i,j} = \rho^{|i-j|}$,

(Model 2) $\boldsymbol{\Sigma}_{i,j} = (|i-j+1|^{2h} - 2|i-j|^{2h} + |i-j-1|^{2h})/2$,

which, respectively, correspond to the covariance structure of an autoregressive process and a fractional Brownian Motion (fBM), where $\boldsymbol{\Sigma}_{i,j}$ is the (i,j)-th element of $\boldsymbol{\Sigma}$. Note that the above settings satisfy that $a_i = O(1)$, $i = 1, 2$, and $a_{20} = O(1)$.

We vary the value of ρ and h as $\rho = 0$, 0.3, 0.6 and 0.9 in Model 1 and $h = 0.2$, 0.5 and 0.8 in Model 2. Note that the case $\rho = 0$ in (Model 1) is identical to the case $h = 0.5$ in (Model 2), since in both cases $\boldsymbol{\Sigma}$ reduces to the identity matrix. As models for underlying distributions, we treat the following three cases: Random observations $\boldsymbol{x}_i$'s, $i = 1, \ldots, n$, are generated as $\boldsymbol{x}_i = \boldsymbol{\Sigma}^{1/2} \boldsymbol{z}_i$ for $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{pi})$ with $z_{i1}, \ldots, z_{pi}$ being mutually independent:

(Case 1) $z_{ij} \sim \mathcal{N}(0,1)$,

(Case 2) $z_{ij} = (u_{ij} - m)/\sqrt{2m}$, $u_{ij} \sim \chi_m^2$ for $m = 2$.

Case 2 is a non-normal case. Note that the skewness and kurtosis $(K_4 + 3)$ of χ_m^2 is, respectively, $(8/m)^{1/2}$ and $3 + 12/m$.

We investigate the risk performance of the estimators given in Section 2, namely $\widehat{\boldsymbol{\Sigma}}_{LW}$ given in (1.2), RBLW estimator $\widehat{\boldsymbol{\Sigma}}_{RBLW}$ in (1.3), $\widehat{\boldsymbol{\Sigma}}_{FS}$ given in (1.5), $\widehat{\boldsymbol{\Sigma}}_{OAS}$ given in (1.4) and the proposed estimator $\widehat{\boldsymbol{\Sigma}}_U$ in (2.8). Additionally, we add the oracle estimator $\widehat{\boldsymbol{\Sigma}}_0$, i.e., the estimator whose weight is the optimal one given in (2.11).

The simulation experiments are carried out under the above model for fixed $p = 100$ and $N = 5, 10, 15, \ldots, 50$. Based on 5,000 replications, we calculate empirical risk of these estimators and show them on graphs in the two cases of the underlying distributions. The results are shown in several figures at the end of the paper, where $\widehat{\boldsymbol{\Sigma}}_{LW}$, $\widehat{\boldsymbol{\Sigma}}_{RBLW}$, $\widehat{\boldsymbol{\Sigma}}_{FS}$, $\widehat{\boldsymbol{\Sigma}}_{OAS}$, $\widehat{\boldsymbol{\Sigma}}_U$ and the oracle estimator are denoted by LW, RBLW, FS, OAS, U and Oracle.

In the cases of normal distributions, as shown in Figure 1 in which $\boldsymbol{\Sigma}$ is identity matrix, $\widehat{\boldsymbol{\Sigma}}_{LW}$ looks to be unstable for small N. Our proposed estimator $\widehat{\boldsymbol{\Sigma}}_U$ performs some what worse than $\widehat{\boldsymbol{\Sigma}}_{FS}$ and $\widehat{\boldsymbol{\Sigma}}_{OAS}$, but the difference converges to zero as N increases. In other cases with normal distributions, we get similar results and omit the details here.

In the non-normal cases , on the other hand, one sees that the proposed estimator $\widehat{\boldsymbol{\Sigma}}_U$ performs best of these in most cases (other than the oracle estimator) as shown in Figure 2-8. Especially in the case close to the sphericity (i.e., small ρ and h), only $\widehat{\boldsymbol{\Sigma}}_U$ performs closely to the oracle estimator even when the sample size is small. Note that, however, when $\rho = 0.9$ in (Model 1), $\widehat{\boldsymbol{\Sigma}}_{FS}$ and $\widehat{\boldsymbol{\Sigma}}_{OAS}$ perform some what better than $\widehat{\boldsymbol{\Sigma}}_U$ with $N = 5$. However, in such cases, some other shrinkage target than spherical one is more appropriate. Compared to $\widehat{\boldsymbol{\Sigma}}_{FS}$ and $\widehat{\boldsymbol{\Sigma}}_{OAS}$, the estimator $\widehat{\boldsymbol{\Sigma}}_{LW}$ performs better in

that they converge to the oracle estimator faster. It is probably because $\widehat{\mathbf{\Sigma}}_{OAS}$ and $\widehat{\mathbf{\Sigma}}_{FS}$ is more affected by the outliers from non-normal distributions.

Although $\widehat{\mathbf{\Sigma}}_{OAS}$ and $\widehat{\mathbf{\Sigma}}_{FS}$ are slightly better than $\widehat{\mathbf{\Sigma}}$ in the case of normal distributions, they are more unstable than $\widehat{\mathbf{\Sigma}}_U$ in the cases of non-normal distributions. In practice we often encounter the situation in which we do not know whether underlying distribution is normal or not. Thus, we recommend the suggested estimator $\widehat{\mathbf{\Sigma}}_U$, since $\widehat{\mathbf{\Sigma}}_{LW}$ performs worse than $\widehat{\mathbf{\Sigma}}_U$ with small number of samples.

Lastly, Figure 9 shows sample average estimated weights of $\mathbf{S}$; $\hat{w}_{LW}$, $\hat{w}_{RBLW}$, $\hat{w}_{OAS}$, $\hat{w}_{FS}$ and $\hat{w}_U$ with the optimal weight w^* given in (2.11). It is seen that the optimal weight w^* are estimated more unbiasedly by $\hat{w}_U$ in every sample size N than the other estimators of the weight. Results in other non-normal cases are similar and omitted.

5   Concluding Remarks

In this paper, we have addressed the problem of constructing an invertible and well-conditioned estimator of a large covariance matrix in both normal and non-normal cases, which is the plug-in estimator based on the optimal convex combination of $\mathbf{S}$ and $\hat{a}_1 \mathbf{I}_p$. The performance of such a linear shrinkage estimator depends on an estimator of the optimal weight. Our proposal for estimation of the weight is the ratio estimator based on approximated or exact unbiased estimators of the numerator and denominator of the optimal weight in non-normal distributions. We have shown that the estimator of Ledoit and Wolf (2004) and the RBLW estimator have second-order biases in normal and non-normal distributions. It has been also shown that the estimator of Fisher and Sun (2011) and the OAS estimator are almost indentical and that they are second-order unbiased in the normal case, but have second-order biases in non-normal cases. This is why, in simulation studies, their estimators perform slightly better in the normal case than our proposal, but worse in non-normal cases. Some simulation results tell us about how important the unbiased estimation of the optimal weight is, and our proposal is recommended as an estimator of the optimal weight.

Although we treat the sphericity assumption $a_1 \mathbf{I}_p$ as the shrinkage target. The shrinkage target may be extended to the general assumption $\mathbf{\Lambda}(\boldsymbol{\theta})$ for unknown parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$. For example, we consider a diagonal matrix, an autoregressive structure and an intra-class correlation structre as $\mathbf{\Lambda}(\boldsymbol{\theta})$. A consistent estimator of $\boldsymbol{\theta}$ under the constraint $\mathbf{\Sigma} = \mathbf{\Lambda}(\boldsymbol{\theta})$ is denoted by $\widehat{\boldsymbol{\theta}}$. Then, linear shrinkage estimators which shrinkage $\mathbf{S}$ toward $\mathbf{\Lambda}(\widehat{\boldsymbol{\theta}})$ is written as $\widehat{\mathbf{\Sigma}}(\mathbf{\Lambda}) = w\mathbf{S} + (1-w)\mathbf{\Lambda}(\widehat{\boldsymbol{\theta}})$, and the optimal weight is given by

$$w^*(\mathbf{\Lambda}) = E[\operatorname{tr}[(\mathbf{S} - \mathbf{\Lambda}(\widehat{\boldsymbol{\theta}}))(\mathbf{\Sigma} - \mathbf{\Lambda}(\widehat{\boldsymbol{\theta}}))]]/E[\operatorname{tr}(\mathbf{S} - \mathbf{\Lambda}(\widehat{\boldsymbol{\theta}}))^2]. \tag{5.1}$$

Assume that the numerator $E[\operatorname{tr}[(\mathbf{S} - \mathbf{\Lambda}(\widehat{\boldsymbol{\theta}}))(\mathbf{\Sigma} - \mathbf{\Lambda}(\widehat{\boldsymbol{\theta}}))]]$ can be unbiasedly estimated by $G(\mathbf{S})$ in non-normal distributions. It follows from the arguments in Section 2 that the optimal weight is estimated by

$$\hat{w}_U(\mathbf{\Lambda}) = G(\mathbf{S})/\operatorname{tr}(\mathbf{S} - \mathbf{\Lambda}(\widehat{\boldsymbol{\theta}}))^2, \tag{5.2}$$

12

which is an extension of our proposal to the general target.

For example, consider the diagonal matrix $\boldsymbol{\Lambda} = \boldsymbol{D} = \operatorname{diag}(\theta_1, \ldots, \theta_p)$ as a target. Under the constraint $\boldsymbol{\Sigma} = \boldsymbol{D}$, the restricted estimator is $\hat{\theta}_i = s_{ii}$ for $\boldsymbol{S} = (s_{ij})$. Let $\widehat{\boldsymbol{D}} = \operatorname{diag}(\hat{\theta}_1, \ldots, \hat{\theta}_p)$. Then, the numerator in (5.1) is written as $E[\operatorname{tr}[(\boldsymbol{S} - \widehat{\boldsymbol{D}})(\boldsymbol{\Sigma} - \widehat{\boldsymbol{D}})]] = E[\operatorname{tr}[(\boldsymbol{S} - \widehat{\boldsymbol{D}})\boldsymbol{\Sigma}]] = pa_2 - pa_{20}$, which is unbiasedly estimated as $G(\boldsymbol{S}) = p(\hat{a}_{2C} - \hat{a}_{20})$. Since the denomenator of (5.1) is $\operatorname{tr}\boldsymbol{S}^2 - \operatorname{tr}\widehat{\boldsymbol{D}}^2 = \operatorname{tr}\boldsymbol{S}^2 - p\hat{a}_{20}$, we have

$$\hat{w}_U(\boldsymbol{D}) = \frac{\hat{a}_{2C} - \hat{a}_{20}}{\operatorname{tr}\boldsymbol{S}^2/p - \hat{a}_{20}}$$

and the estimator of $\boldsymbol{\Sigma}$ is $\hat{w}_U(\boldsymbol{D})\boldsymbol{S} + \{1 - \hat{w}_U(\boldsymbol{D})\}\widehat{\boldsymbol{D}}$. Asymptotic properties of linear shrinkage estimators toward such a general target would be worth investigating in a future.

Acknowledgments.

References

[1] Bai, J., and Shi, S. (2011). Estimating high dimensional covariance matrices and its applications. *Ann. Economics and finance*, **12**, 199-215.

[2] Chen, Y., Wiesel, A., Eldar, C.Y., and Hero, A.O. (2010). Shrinkage Algorithms for MMSE Covariance Estimation. *IEEE Trans. on Sig. Process.*, **58**, 5016-5029.

[3] Chen, S. X., Zhang, L.-X. and Zhong, P.-S. (2010) Tests for high-dimensional covariance matrices. *J. Am. Statist. Assoc.*, **105**, 810-819.

[4] Daniels, M., and Kass, R.E. (2001). Shrinkage estimators for covariance matrices. *Biometrics*, **57**, 1173-1184.

[5] Fisher, T.J., and Sun, X. (2011). Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix. *Comp. Statist. Data Analysis*, **55**, 1909-1918.

[6] Himeno, T., and Yamada, T. (2014). Estimation for some functions of covariance matrix in high dimension under non-normality and its applications. *J. Multivariate Analysis*, **130**, 27-44.

[7] Konno, Y. (2009). Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss. *J. Multivariate Analysis*, **100**, 2237-2253.

[8] Ledoit, O., and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empirical Finance*, **10**, 603-621.

[9] Ledoit, O., and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Analysis*, **88**, 365-411.

[10] Srivastava, M.S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.*, **35**, 251-272.

[11] Srivastava, M.S., Kollo, T., and von Rosen, D. (2011). Some tests for the covariance matrix with fewer observations than the dimension under non-normality. *J. Multivariate Analysis*, **102**, 1090-1103.

[12] Srivastava, M.S., and Kubokawa, T. (2007). Comparison of discrimination methods for high dimensional data. *J. Japan Statist. Soc.*, **37**, 123-134.

[13] Srivastava, M.S., Yanagihara, H., and Kubokawa, T. (2014). Tests for covariance matrices in high dimension with less sample size. *J. Multivariate Analysis*, **130**, 289-309.

Figure 1: Comparison of estimators in normal case; $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ ($\rho = 0$ in (Model 1) or $h = 0.5$ in (Model 2))

14
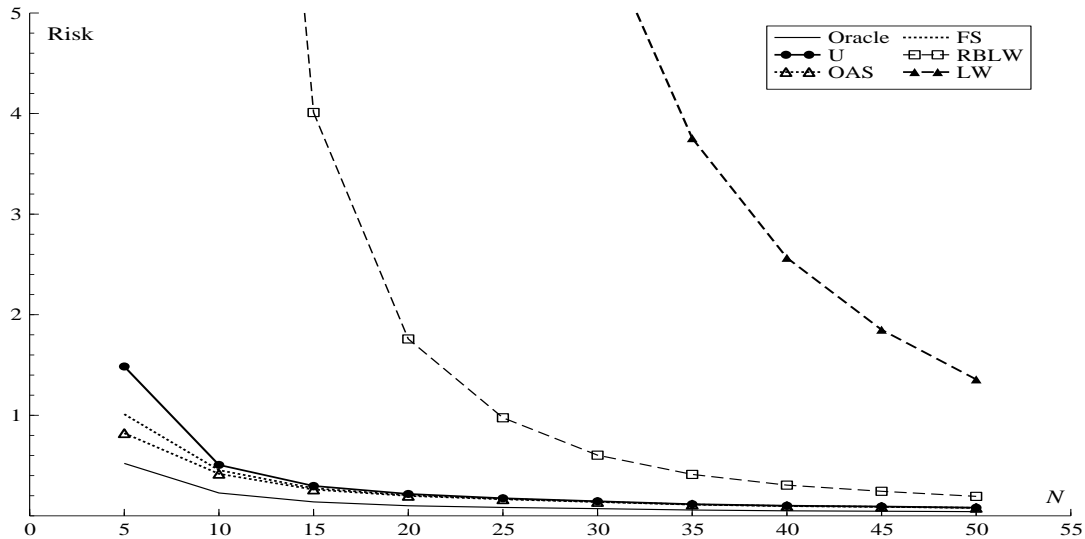
Figure 2: Comparison of estimators in non-normal case; $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ ($\rho = 0$ in (Model 1) or $h = 0.5$ in (Model 2))
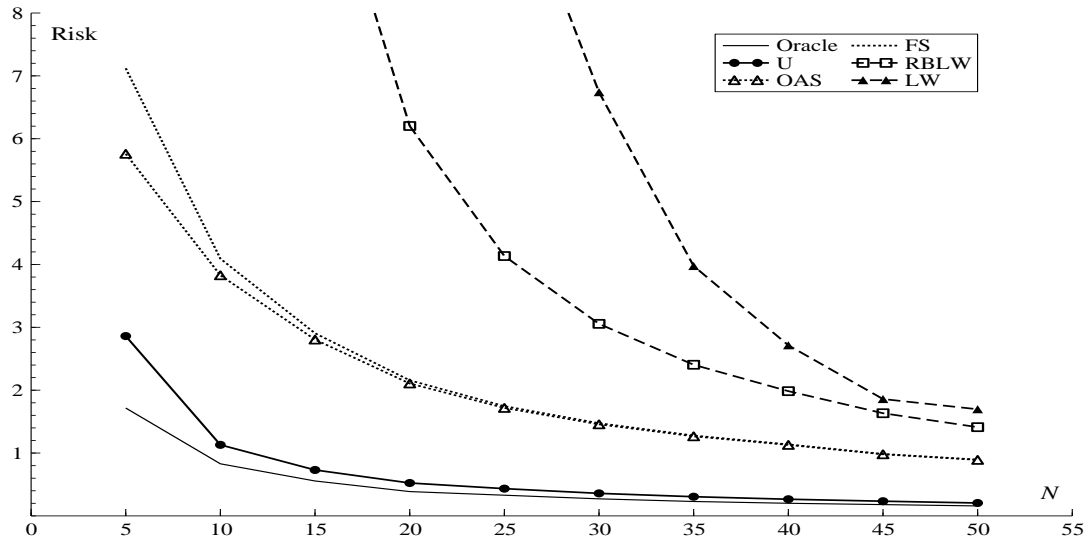

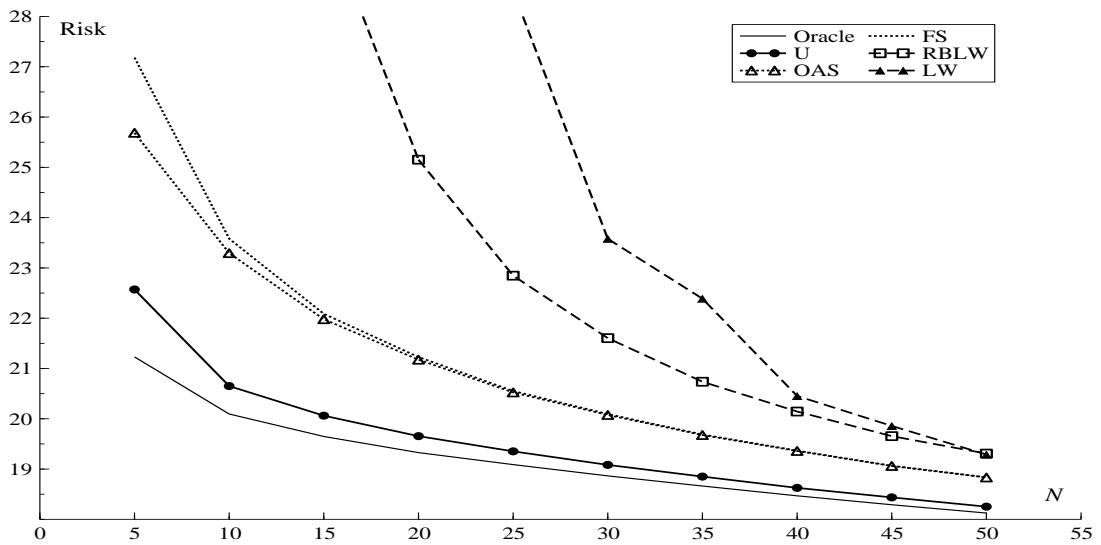Figure 3: Comparison of estimators in non-normal case; $\rho = 0.3$ in (Model 1)

Figure 4: Comparison of estimators in non-normal case; $\rho = 0.6$ in (Model 1)
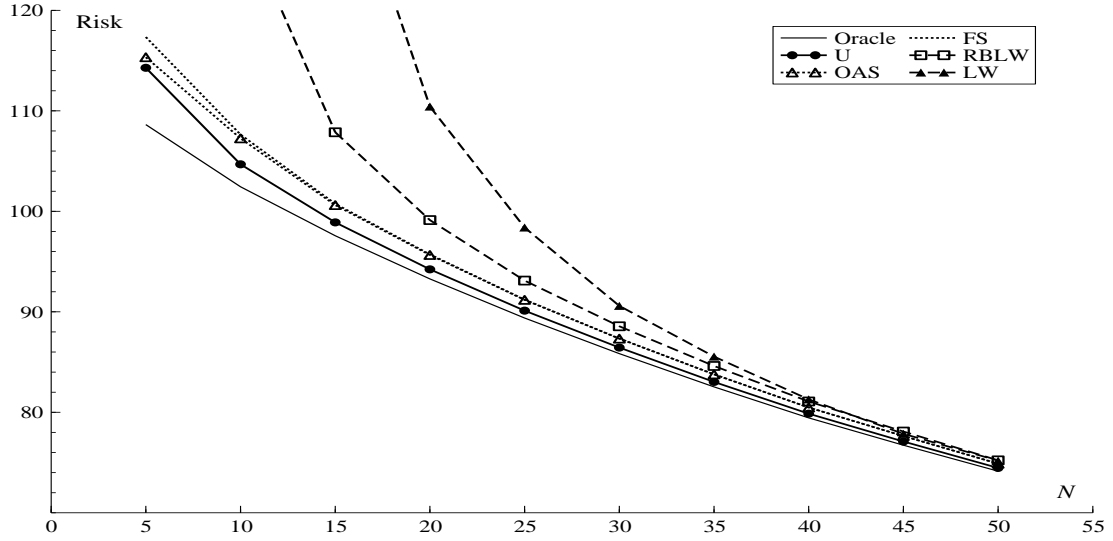

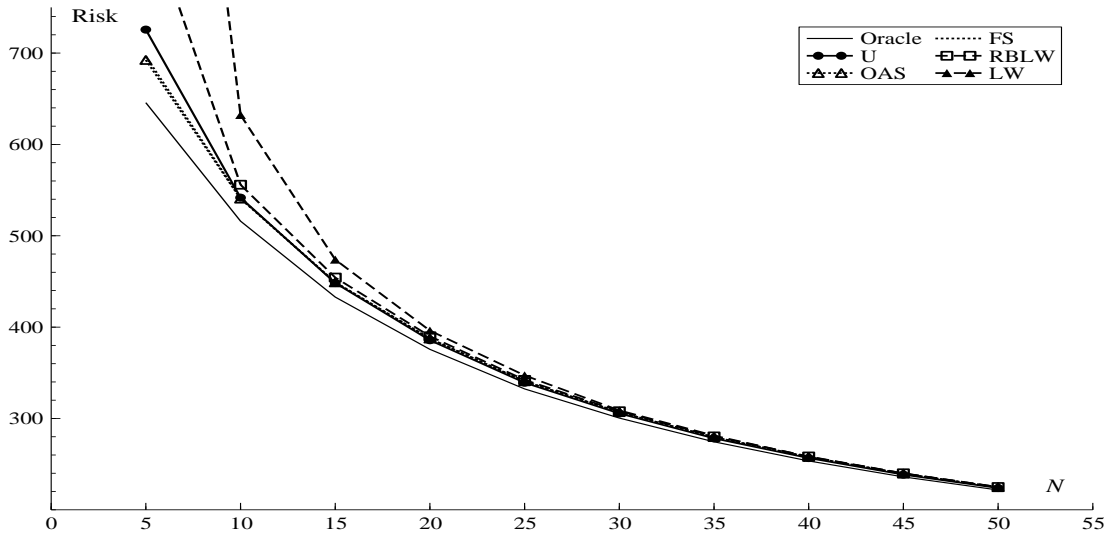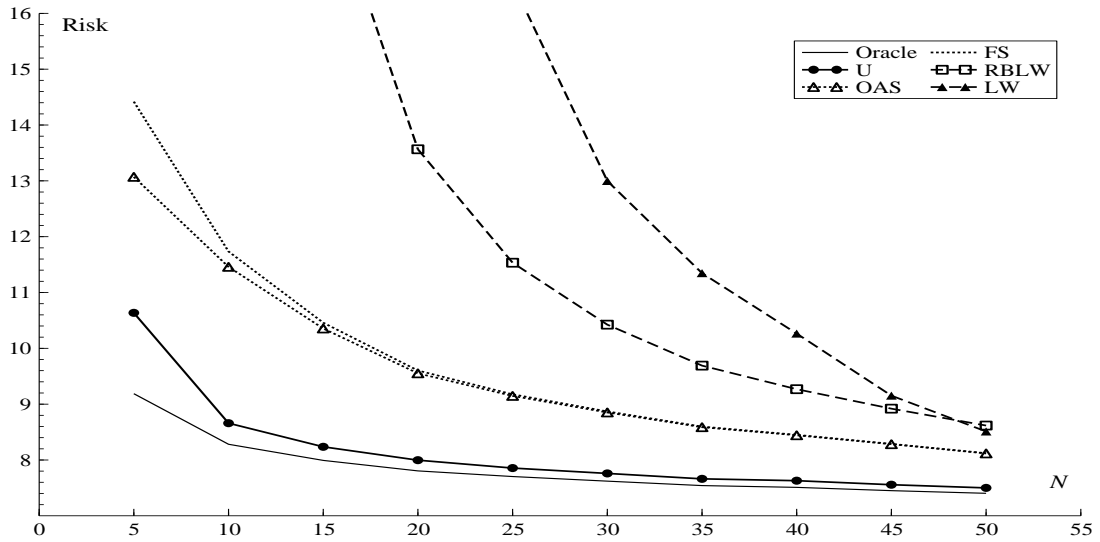Figure 5: Comparison of estimators in non-normal case; $\rho = 0.9$ in (Model 1)

16

Figure 6: Comparison of estimators in non-normal case; $h = 0.6$ in (Model 2)

Figure 7: Comparison of estimators in non-normal case; $h = 0.7$ in (Model 2)
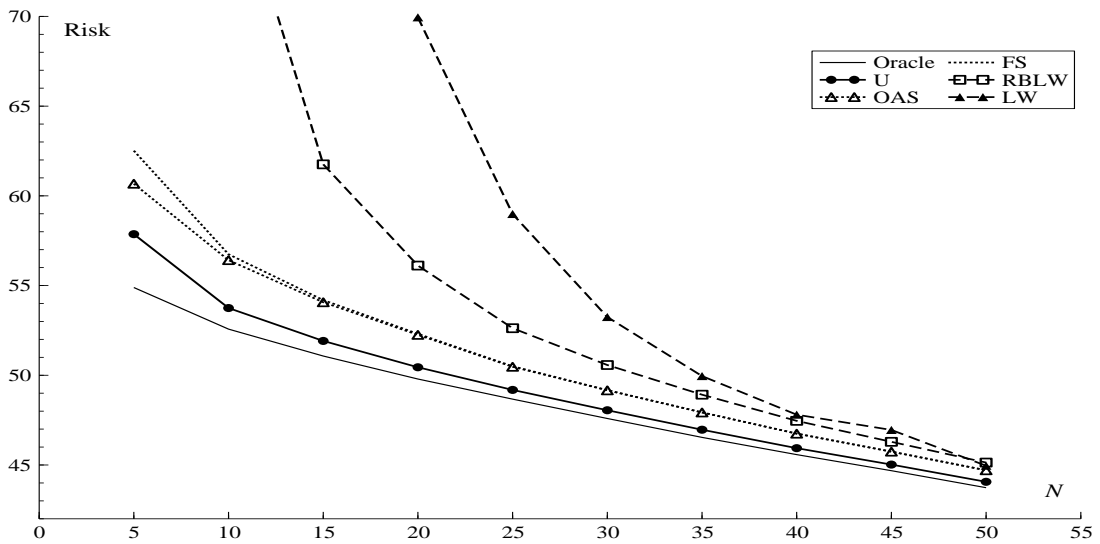
17

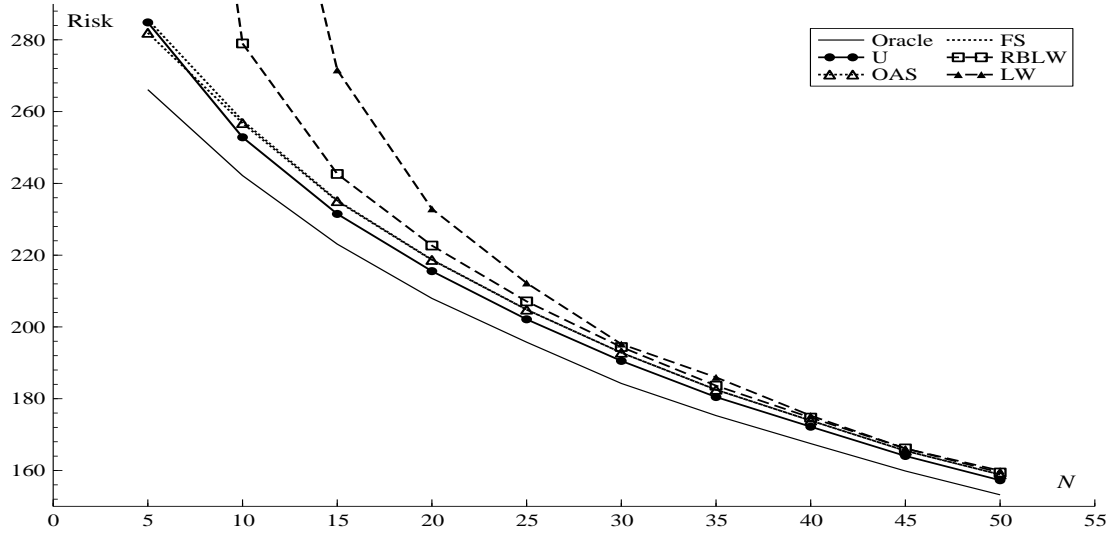Figure 8: Comparison of estimators in non-normal case; $h = 0.8$ in (Model 2)

Figure 9: Comparison of sample average of estimated weights of $\boldsymbol{S}$ in non-normal case with the optimal weight; $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ ($\rho = 0$ in (Model 1) or $h = 0.5$ in (Model 2))