

Article

Estimation of Large-Dimensional Covariance Matrices via Second-Order Stein-Type Regularization

Bin Zhang ¹, Hengzhen Huang ^{1,*} and Jianbin Chen ²

¹ College of Mathematics and Statistics, Guangxi Normal University, Guilin 541004, China

² School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China

* Correspondence: hzhuang@mailbox.gxnu.edu.cn

Abstract: This paper tackles the problem of estimating the covariance matrix in large-dimension and small-sample-size scenarios. Inspired by the well-known linear shrinkage estimation, we propose a novel second-order Stein-type regularization strategy to generate well-conditioned covariance matrix estimators. We model the second-order Stein-type regularization as a quadratic polynomial concerning the sample covariance matrix and a given target matrix, representing the prior information of the actual covariance structure. To obtain available covariance matrix estimators, we choose the spherical and diagonal target matrices and develop unbiased estimates of the theoretical mean squared errors, which measure the distances between the actual covariance matrix and its estimators. We formulate the second-order Stein-type regularization as a convex optimization problem, resulting in the optimal second-order Stein-type estimators. Numerical simulations reveal that the proposed estimators can significantly lower the Frobenius losses compared with the existing Stein-type estimators. Moreover, a real data analysis in portfolio selection verifies the performance of the proposed estimators.

Keywords: covariance matrix estimation; Stein-type regularization; unbiased estimate



Citation: Zhang, B.; Huang, H.; Chen, J. Estimation of Large-Dimensional Covariance Matrices via Second-Order Stein-Type Regularization. *Entropy* **2023**, *25*, 53. <https://doi.org/10.3390/e25010053>

Academic Editor: Yuehua Wu

Received: 11 November 2022

Revised: 20 December 2022

Accepted: 23 December 2022

Published: 27 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a fundamental problem in modern multivariate statistics and various practical applications, estimating the covariance matrix of a large-dimensional random vector has attracted significant interest in the last two decades [1,2]. The traditional sample covariance matrix (SCM) becomes unstable and ill-conditioned when the dimension increases proportionally with the sample size. The algorithm, which still employs SCM as the covariance matrix estimator, will result in a drop in performance or failure [3,4]. Although remaining unbiased, the SCM is not a satisfactory estimator of the actual covariance matrix anymore [5,6]. Therefore, it is of great concern to develop well-conditioned covariance matrix estimators in large-dimensional scenarios [7–9].

A comprehensive point of view to obtain well-conditioned estimators is by improving the SCM [10]. In the Stein-type regularization (linear shrinkage estimation), a target matrix is preset by rectifying the SCM according to prior information on the covariance structure. For example, the spherical target is a scalar matrix, with the coefficient being the average of the diagonal elements of the SCM [11]. The diagonal target is a diagonal matrix which retains the diagonal elements of the SCM [12,13]. Moreover, the Toeplitz-structured target is formed by averaging each diagonal of the SCM [14]. To some extent, the target matrices are covariance matrix estimators. Despite being biased, they usually enjoy low variances. The Stein-type estimator combines the SCM and the target matrix to balance the bias and variance. This method generates a well-conditioned estimator for the spherical target matrix by retaining the sample eigenvectors and shrinking the sample eigenvalues toward their grand mean. Moreover, the Stein-type regularization can be translated as a weighted average between the SCM and the target matrix. Many Stein-type estimators have been developed for various target matrices. Additionally, it is worth mentioning that the optimal

Stein-type estimator can be expressed in closed form and significantly outperforms the traditional SCM under some appropriate criteria [15–17].

In addition, the nonlinear shrinkage estimation is proposed based on the random matrix theory [18]. By taking spectral decomposition on the SCM, this method retains the sample eigenvectors and estimates the actual eigenvalues by taking a nonlinear transformation on the sample eigenvalues. Then, the nonlinear shrinkage estimator is obtained by assembling the estimated eigenvalues and the sample eigenvectors. In the mean squared error (MSE) sense, the resulting nonlinear shrinkage estimator enjoys a significant advantage over the SCM. It also outperforms the Stein-type estimator for the spherical target. However, both the sample eigenvalues and sample eigenvectors have serious deficiencies in a high-dimensional case [19]. Hence, the existing nonlinear shrinkage strategy, which modifies the sample eigenvalues while retaining the sample eigenvectors, has some limitations in improving the SCM performance. Moreover, the method can hardly handle the prior structure information employed in the Stein-type regularization. Hence, developing a new nonlinear shrinkage technique is essential to generate outperformed covariance matrix estimators.

This paper combines the SCM and the target matrix via a nonlinear shrinkage strategy to obtain well-conditioned estimators of a large-dimensional covariance matrix. The main contributions are the following:

1. The second-order Stein-type estimator is modeled as a quadratic polynomial concerning the SCM and an almost surely (a.s.) positive definite target matrix. For the spherical and diagonal target matrices, the MSEs between the second-order Stein-type estimator and the actual covariance matrix are unbiasedly estimated under Gaussian distribution.
2. We formulate the second-order Stein-type estimators for the two target matrices as convex quadratic programming problems. Then, the optimal second-order Stein-type estimators are immediately obtained.
3. Some numerical simulations and application examples are provided for comparing the proposed second-order Stein-type estimators with the existing linear and nonlinear shrinkage estimators.

The outline of this paper is as follows. Section 2 proposes the second-order Stein-type estimator based on the Stein-type regularization. In Section 3, the spherical and diagonal matrices are employed as the target matrices. We obtain unbiased estimates of the MSEs between the second-order Stein-type estimators and the actual covariance matrix. The optimal second-order Stein-type estimators are obtained by solving the corresponding optimization problems. Section 4 provides some numerical simulations and two examples to discover the performance of the proposed estimators in large-dimensional scenarios. Section 5 concludes the major work of this paper.

2. Notation, Motivation, and Formulation

The symbol \mathbb{R}^p denotes the set of entire p -dimensional real column vectors, $\mathbb{R}^{m \times n}$ denotes the set of entire $m \times n$ real matrices, and \mathbb{S}^p denotes the set of entire $p \times p$ real symmetric matrices. The bold symbol \mathbf{E} denotes the squared matrix having all entries 1 with appropriate dimensions. The symbol \mathbf{I}_p denotes the $p \times p$ identity matrix. For a matrix \mathbf{A} , \mathbf{A}^T , $\text{tr}(\mathbf{A})$, and $\|\mathbf{A}\|$, we denote its transpose, trace, and Frobenius matrix norm, respectively. For two matrices \mathbf{A} and \mathbf{B} , $\mathbf{A} \circ \mathbf{B}$ means their Hadamard (element-wise) product.

Assume that $x_1, x_2, \dots, x_n \in \mathbb{R}^p$ is an independent and identically distributed (i.i.d.) sample drawn from a certain distribution with mean 0 and covariance matrix Σ . The SCM \mathbf{S} is defined by

$$\mathbf{S} = (s_{ij})_{p \times p} = \frac{1}{n} \sum_{m=1}^n x_m x_m^T. \quad (1)$$

As is widely known, the SCM is ill conditioned in large-dimension scenarios and is even singular when $p > n$. The Stein-type regularization can produce a well-conditioned covariance matrix estimator based on the SCM [20–23].

For an a.s. positive definite target matrix \mathbf{T} representing the prior information of the covariance structure, the Stein-type estimator combines the SCM and the target matrix with a linear function

$$f(\mathbf{S}, \mathbf{T}) = (1 - w)\mathbf{S} + w\mathbf{T}. \tag{2}$$

Through an equivalent transformation, the expression given by (2) can be recombined as

$$f(\mathbf{S}, \mathbf{T}) = \mathbf{S} + w(\mathbf{T} - \mathbf{S}), \tag{3}$$

where $\mathbf{T} - \mathbf{S}$ is a regularization term and w is a tuning parameter. Moreover, the tuning parameter w lies in $(0, 1]$ so as to keep the Stein-type estimator a.s. positive definite even when $n < p$. An interesting fact is that the matrix $(\mathbf{T} - \mathbf{S})^2$ is still symmetric and positive definite, which motivates us to consider further a quadratic function of the SCM and the target matrix.

For an a.s. positive definite target matrix \mathbf{T} , we model the second-order Stein-type estimator of the covariance matrix Σ as

$$\hat{\Sigma} = \mathbf{S} + w_1(\mathbf{T} - \mathbf{S}) + w_2(\mathbf{T} - \mathbf{S})^2, \tag{4}$$

where w_1, w_2 are the tuning parameters. It is easy to find out that $\hat{\Sigma} \in \mathbb{S}^p$. In the same manner, we further assume $w_1 \in (0, 1]$ and $w_2 \geq 0$ to keep the covariance estimator given by (4) to be a.s. positive definite. We note that the constraint on the tuning parameters is an easy-to-implement condition but is not necessary. One can also consider alternative assumptions, such as the condition number constraint, to obtain positive definite estimators of the large-dimensional covariance matrix [24].

Next, we choose the optimal tuning parameters in (4). In the Stein-type regularization, the MSE between the actual covariance matrix and its estimator is the most commonly used loss function. It includes unknown scalars concerning the expectation operator and the actual covariance matrix. One practical way is to make the MSE available by estimating the unknown scalars and obtaining the optimal shrinkage intensity by minimizing the available MSE [15]. Therefore, we still follow the above steps to find optimal tuning parameters in the second-order Stein-type estimator. To be specific, the loss function of the second-order Stein-type estimator $\hat{\Sigma}$ is defined by

$$\mathcal{M}_{\mathbf{T}}(\mathbf{w}|\Sigma) = \mathbb{E}[\|\hat{\Sigma} - \Sigma\|^2], \tag{5}$$

where $\mathbf{w} = (w_1, w_2)^T$. Through substituting the expression of $\hat{\Sigma}$ into (5), we can obtain

$$\mathcal{M}_{\mathbf{T}}(\mathbf{w}|\Sigma) = \frac{1}{2} \mathbf{w}^T \mathbf{H}(\mathbf{T}) \mathbf{w} - 2\mathbf{w}^T \mathbf{b}(\mathbf{T}) + c, \tag{6}$$

where

$$\mathbf{H}(\mathbf{T}) = 2 \begin{bmatrix} \mathbb{E}[\text{tr}(\mathbf{S} - \mathbf{T})^2] & -\mathbb{E}[\text{tr}(\mathbf{S} - \mathbf{T})^3] \\ -\mathbb{E}[\text{tr}(\mathbf{S} - \mathbf{T})^3] & \mathbb{E}[\text{tr}(\mathbf{S} - \mathbf{T})^4] \end{bmatrix}, \tag{7}$$

$$\mathbf{b}(\mathbf{T}) = \begin{bmatrix} \mathbb{E}[\text{tr}(\mathbf{S} - \Sigma)(\mathbf{S} - \mathbf{T})] \\ -\mathbb{E}[\text{tr}(\mathbf{S} - \Sigma)(\mathbf{S} - \mathbf{T})^2] \end{bmatrix}, \tag{8}$$

$$c = \mathbb{E}[\text{tr}(\mathbf{S} - \Sigma)^2]. \tag{9}$$

Therefore, the second-order Stein-type regularization can be modeled as the following optimization problem:

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{w}^T \mathbf{H}(\mathbf{T}) \mathbf{w} - 2 \mathbf{w}^T \mathbf{b}(\mathbf{T}) + c \\ \text{s.t.} \quad & 0 \leq \mathbf{e}_1 \mathbf{w} \leq 1, \\ & \mathbf{e}_2 \mathbf{w} \geq 0, \end{aligned} \tag{10}$$

where $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$. It is easy to see that the loss function in problem (10) is a binary quadratic polynomial function about \mathbf{w} , and $\mathbf{H}(\mathbf{T})$ is the Hessian matrix. By the Cauchy–Schwarz inequality, we have

$$\left(\mathbb{E}[\text{tr}(\mathbf{S} - \mathbf{T})^3] \right)^2 \leq \mathbb{E}[\text{tr}(\mathbf{S} - \mathbf{T})^2] \mathbb{E}[\text{tr}(\mathbf{S} - \mathbf{T})^4]. \tag{11}$$

Therefore, the Hessian matrix $\mathbf{H}(\mathbf{T})$ is positive definite. Then, the optimization problem (10) is a convex quadratic program. However, it cannot be solved because the quantities $\mathbf{H}(\mathbf{T})$, $\mathbf{b}(\mathbf{T})$, and c in the objective function are unknown. When the underlying distribution is given and the target matrix is prespecified, we can estimate the unknown quantities $\mathbf{H}(\mathbf{T})$, $\mathbf{b}(\mathbf{T})$, and c . Then, the optimization problem (10) turns out to be available based on plug-in strategy and can be effectively solved.

Remark 1. *The unknown quantity c does not affect the choice of optimal tuning parameter \mathbf{w} in the optimization problem (10). Moreover, it is the theoretical MSE between the actual covariance matrix and the classic SCM and plays an important role in evaluating the performance of improved covariance estimators based on the SCM.*

3. Optimal Second-Order Stein-Type Estimators

In this section, as the target matrix is specified, we estimate the corresponding unknown quantities under Gaussian distribution, then establish the available version of the optimization problem to obtain the optimal second-order Stein-type estimator.

3.1. Target Matrices

As mentioned before, the target matrix represents the prior information of the actual covariance structure. In the Stein-type regularization, the commonly used target matrices include the spherical target, the diagonal target, the Toeplitz-structured target, and the tapered SCM. Among these, the spherical target and the diagonal target are a.s. positive definite, whereas both the Toeplitz-structured target and the tapered SCM are unnecessary. Thereby, in the second-order Stein-type regularization, we employ the spherical target and the diagonal target, given by

$$\mathbf{T}_1 = \frac{\text{tr}(\mathbf{S})}{p} \mathbf{I}_p, \quad \mathbf{T}_2 = \text{diag}(s_{11}, \dots, s_{pp}). \tag{12}$$

The diagonal target \mathbf{T}_2 is also denoted as \mathbf{D}_S because it consists of the diagonal elements of the SCM.

3.2. Available Loss Functions

For the target matrices \mathbf{T}_1 and \mathbf{T}_2 , we unbiasedly estimate the loss functions given by (6) through plugging in the estimates of unknown quantities $\mathbf{H}(\mathbf{T}_i)$, $\mathbf{b}(\mathbf{T}_i)$, $i = 1, 2$ and c under Gaussian distribution.

First of all, by directly removing the expectation operator, the Hessian matrices $\mathbf{H}(\mathbf{T}_i)$ can be estimated by

$$\hat{\mathbf{H}}(\mathbf{T}_i) = 2 \begin{bmatrix} \text{tr}(\mathbf{S} - \mathbf{T}_i)^2 & -\text{tr}(\mathbf{S} - \mathbf{T}_i)^3 \\ -\text{tr}(\mathbf{S} - \mathbf{T}_i)^3 & \text{tr}(\mathbf{S} - \mathbf{T}_i)^4 \end{bmatrix}. \tag{13}$$

Furthermore, $\hat{\mathbf{H}}(\mathbf{T}_i)$, $i = 1, 2$ are, respectively, unbiased estimates of $\mathbf{H}(\mathbf{T}_i)$, $i = 1, 2$.

Next, for $i = 1, 2$, we decompose the unknown vectors $\mathbf{b}(\mathbf{T}_i)$ into two terms,

$$\mathbf{b}(\mathbf{T}_i) = \begin{bmatrix} \mathbb{E}[\text{tr}(\mathbf{S} - \boldsymbol{\Sigma})(\mathbf{S} - \mathbf{T}_i)] \\ -\mathbb{E}[\text{tr}(\mathbf{S} - \boldsymbol{\Sigma})(\mathbf{S} - \mathbf{T}_i)^2] \end{bmatrix} \tag{14}$$

$$= \begin{bmatrix} \mathbb{E}[\text{tr}(\mathbf{S}(\mathbf{S} - \mathbf{T}_i))] \\ -\mathbb{E}[\text{tr}(\mathbf{S}(\mathbf{S} - \mathbf{T}_i)^2)] \end{bmatrix} + \begin{bmatrix} -\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}(\mathbf{S} - \mathbf{T}_i))] \\ \mathbb{E}[\text{tr}(\boldsymbol{\Sigma}(\mathbf{S} - \mathbf{T}_i)^2)] \end{bmatrix} \\ \triangleq \mathbf{u}(\mathbf{T}_i) + \mathbf{v}(\mathbf{T}_i). \tag{15}$$

Similar to the Hessian matrices $\mathbf{H}(\mathbf{T}_1)$ and $\mathbf{H}(\mathbf{T}_2)$, the first term $\mathbf{u}(\mathbf{T}_i)$ can be unbiasedly estimated by

$$\hat{\mathbf{u}}(\mathbf{T}_i) = \begin{bmatrix} \text{tr}(\mathbf{S}(\mathbf{S} - \mathbf{T}_i)) \\ -\text{tr}(\mathbf{S}(\mathbf{S} - \mathbf{T}_i)^2) \end{bmatrix}. \tag{16}$$

Therefore, we only need to estimate the second term $\mathbf{v}(\mathbf{T}_i)$. It is challenging to estimate $\mathbf{v}(\mathbf{T}_i)$ unbiased because it includes both the expectation operator and the actual covariance matrix $\boldsymbol{\Sigma}$. We need the following moment properties about the Wishart distribution [25].

Lemma 1. Denote \mathbf{A} and \mathbf{B} as arbitrary symmetric nonrandom matrices, \mathbf{S} is the sample covariance matrix given by (1), and then the following equalities hold under Gaussian distribution:

$$\mathbb{E}[\mathbf{SAS}] = \frac{n+1}{n}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma} + \frac{1}{n}\text{tr}(\boldsymbol{\Sigma}\mathbf{A})\boldsymbol{\Sigma}, \tag{17}$$

$$\mathbb{E}[\text{tr}(\mathbf{AS})\mathbf{S}] = \text{tr}(\mathbf{A}\boldsymbol{\Sigma})\boldsymbol{\Sigma} + \frac{2}{n}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}, \tag{18}$$

$$\mathbb{E}[\text{tr}(\mathbf{AS})\text{tr}(\mathbf{BS})] = \text{tr}(\mathbf{A}\boldsymbol{\Sigma})\text{tr}(\mathbf{B}\boldsymbol{\Sigma}) + \frac{2}{n}\text{tr}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}). \tag{19}$$

By Lemma 1, letting $\mathbf{A} = \mathbf{B} = \mathbf{I}_p$, we can obtain

$$\mathbb{E}[\text{tr}(\mathbf{S}^2)] = \frac{n+1}{n}\text{tr}(\boldsymbol{\Sigma}^2) + \frac{1}{n}\text{tr}^2(\boldsymbol{\Sigma}), \quad \mathbb{E}[\text{tr}^2(\mathbf{S})] = \text{tr}^2(\boldsymbol{\Sigma}) + \frac{2}{n}\text{tr}(\boldsymbol{\Sigma}^2). \tag{20}$$

Moreover, letting $\mathbf{A} = \boldsymbol{\Sigma}$ and $\mathbf{B} = \mathbf{I}_p$, we have

$$\mathbb{E}[\text{tr}(\mathbf{S}^2\boldsymbol{\Sigma})] = \frac{n+1}{n}\text{tr}(\boldsymbol{\Sigma}^3) + \frac{1}{n}\text{tr}(\boldsymbol{\Sigma})\text{tr}(\boldsymbol{\Sigma}^2), \tag{21}$$

$$\mathbb{E}[\text{tr}(\mathbf{S}\boldsymbol{\Sigma})\text{tr}(\mathbf{S})] = \text{tr}(\boldsymbol{\Sigma})\text{tr}(\boldsymbol{\Sigma}^2) + \frac{2}{n}\text{tr}(\boldsymbol{\Sigma}^3). \tag{22}$$

Lemma 1 is very helpful to compute the second term $\mathbf{v}(\mathbf{T}_i)$ in (14). For the spherical target matrix \mathbf{T}_1 , we have

$$\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}(\mathbf{S} - \mathbf{T}_1))] = \text{tr}(\boldsymbol{\Sigma}^2) - \frac{1}{p}\text{tr}^2(\boldsymbol{\Sigma}), \tag{23}$$

and

$$\begin{aligned} \mathbb{E}[\text{tr}(\boldsymbol{\Sigma}(\mathbf{S} - \mathbf{T}_1)^2)] &= \mathbb{E}[\text{tr}(\mathbf{S}^2\boldsymbol{\Sigma})] - \frac{2}{p}\mathbb{E}[\text{tr}(\mathbf{S})\text{tr}(\mathbf{S}\boldsymbol{\Sigma})] + \frac{1}{p^2}\mathbb{E}[\text{tr}^2(\mathbf{S})]\text{tr}(\boldsymbol{\Sigma}) \\ &= \frac{np+p-4}{np}\text{tr}(\boldsymbol{\Sigma}^3) + \frac{p^2-2np+2}{np^2}\text{tr}(\boldsymbol{\Sigma})\text{tr}(\boldsymbol{\Sigma}^2) + \frac{1}{p^2}\text{tr}^3(\boldsymbol{\Sigma}). \end{aligned} \tag{24}$$

For the diagonal target matrix \mathbf{T}_2 , we have

$$\mathbb{E}[\text{tr}(\boldsymbol{\Sigma}(\mathbf{S} - \mathbf{T}_2))] = \text{tr}(\boldsymbol{\Sigma}^2) - \text{tr}(\mathbf{D}_{\boldsymbol{\Sigma}}^2), \tag{25}$$

where $\mathbf{D}_\Sigma = \text{diag}(\Sigma_{11}, \dots, \Sigma_{pp})$. Moreover, we can obtain

$$\begin{aligned} & \mathbb{E}[\text{tr}(\Sigma(\mathbf{S} - \mathbf{T}_2)^2)] \\ &= \mathbb{E}[\text{tr}(\Sigma^2 \Sigma)] - 2\mathbb{E}[\text{tr}(\Sigma \mathbf{T}_2 \Sigma)] + \mathbb{E}[\text{tr}(\mathbf{T}_2^2 \Sigma)] \\ &= \frac{n+1}{n} \text{tr}(\Sigma^3) + \frac{1}{n} \text{tr}(\Sigma) \text{tr}(\Sigma^2) - \frac{2n+4}{n} \text{tr}(\mathbf{D}_\Sigma \Sigma^2) + \frac{n+2}{n} \text{tr}(\mathbf{D}_\Sigma^3). \end{aligned} \tag{26}$$

Denote

$$\begin{aligned} a_1 &= \text{tr}(\Sigma^2), & a_2 &= \text{tr}^2(\Sigma), & a_3 &= \text{tr}(\mathbf{D}_\Sigma^2), \\ b_1 &= \text{tr}(\Sigma^3), & b_2 &= \text{tr}^3(\Sigma), & b_3 &= \text{tr}(\Sigma) \text{tr}(\Sigma^2), \\ c_1 &= \text{tr}(\mathbf{D}_\Sigma^3), & c_2 &= \text{tr}(\mathbf{D}_\Sigma \Sigma^2), \end{aligned} \tag{27}$$

then the vectors $\mathbf{v}(\mathbf{T}_i), i = 1, 2$ can be rewritten as

$$\mathbf{v}(\mathbf{T}_1) = \begin{bmatrix} \frac{1}{p} a_2 - a_1 \\ \frac{np+p-4}{np} b_1 + \frac{1}{p^2} b_2 + \frac{p^2-2np+2}{np^2} b_3 \end{bmatrix}, \tag{28}$$

$$\mathbf{v}(\mathbf{T}_2) = \begin{bmatrix} a_3 - a_1 \\ \frac{n+1}{n} b_1 + \frac{1}{n} b_3 + \frac{n+2}{n} c_1 - \frac{2n+4}{n} c_2 \end{bmatrix}. \tag{29}$$

It is worth noting that each element in $\mathbf{v}(\mathbf{T}_i), i = 1, 2$ is a linear combination of the quantities in (27). Therefore, we only need to find out the estimates of the quantities in (27).

Firstly, the unbiased estimates of the quantities $a_i, i = 1, 2, 3$ were proposed in [12,26,27],

$$\alpha_1 = \tau_a (n \text{tr}(\mathbf{S}^2) - \text{tr}^2(\mathbf{S})), \quad \alpha_2 = \tau_a ((n+1) \text{tr}^2(\mathbf{S}) - 2 \text{tr}(\mathbf{S}^2)), \tag{30}$$

$$\alpha_3 = \tau_a (n-1) \text{tr}(\mathbf{D}_\mathbf{S}^2), \tag{31}$$

where $\tau_a = \frac{n}{(n-1)(n+2)}$.

Secondly, denote the matrix \mathbf{W} as

$$\mathbf{W} = \begin{bmatrix} n^2 & 2 & -3n \\ 16 & n^2 + 3n - 2 & -6(n+2) \\ -4n & -(n+2) & n^2 + 2n + 4 \end{bmatrix}, \tag{32}$$

Then, the unbiased estimates of $b_i, i = 1, 2, 3$ can be obtained by the following theorem.

Theorem 1. Under Gaussian distribution, the following equation holds when $n \geq 3$:

$$\mathbb{E} \begin{bmatrix} \text{tr}(\mathbf{S}^3) \\ \text{tr}^3(\mathbf{S}) \\ \text{tr}(\mathbf{S}) \text{tr}(\mathbf{S}^2) \end{bmatrix} = (\tau_b \mathbf{W})^{-1} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \tag{33}$$

where $\tau_b = \frac{n^2}{(n-1)(n-2)(n+2)(n+4)}$.

Proof. The actual covariance matrix Σ has the spectral decomposition which is described as $\Sigma = \Gamma^T \Lambda \Gamma$, where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ is a diagonal matrix consisting of the eigenvalues and Γ is the corresponding unitary matrix. Define $\mathbf{F} = \Sigma^{\frac{1}{2}}$, and \mathbf{F} is a symmetric matrix; then we have $\mathbf{F}^2 = \Sigma$. For $m = 1, \dots, n$, denote $\mathbf{z}_m = \mathbf{F}^{-1} x_m$, then \mathbf{z}_m is an i.i.d. sample and $\mathbf{z}_m \sim N(\mathbf{0}, \mathbf{I}_p)$. Let $\mathbf{X} = (x_1, \dots, x_n)$ and $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)$; then, we have $\mathbf{X} = \mathbf{FZ}$. Notice that the SCM is $\mathbf{S} = \frac{1}{n} \sum_{m=1}^n x_m x_m^T$; therefore, we have

$$n\mathbf{S} = \mathbf{X}\mathbf{X}^T = \mathbf{F}\mathbf{Z}\mathbf{Z}^T\mathbf{F}. \tag{34}$$

Moreover, we can obtain

$$\text{tr}(n\mathbf{S}) = \text{tr}(\mathbf{FZZ}^T\mathbf{F}) = \text{tr}(\mathbf{Z}^T\boldsymbol{\Sigma}\mathbf{Z}) = \text{tr}(\mathbf{Z}^T\Gamma^T\Lambda\Gamma\mathbf{Z}). \tag{35}$$

Define a matrix $\mathbf{Q}^T = \mathbf{Z}^T\Gamma^T = (\mathbf{q}_1, \dots, \mathbf{q}_p)$ and denote $v_{ii} = \mathbf{q}_i^T\mathbf{q}_i$ and $v_{ij} = \mathbf{q}_i^T\mathbf{q}_j = \mathbf{q}_j^T\mathbf{q}_i$ for $i, j \in \{1, \dots, p\}$, then the above equation can be rewritten as

$$\text{tr}(n\mathbf{S}) = \text{tr}(\mathbf{Q}^T\Lambda\mathbf{Q}) = \sum_{i=1}^p \lambda_i v_{ii}. \tag{36}$$

In a same manner, we can obtain

$$\text{tr}(n\mathbf{S})^2 = \sum_{i,j=1}^p \lambda_i \lambda_j v_{ij}^2, \quad \text{tr}^2(n\mathbf{S}) = \sum_{i,j=1}^p \lambda_i \lambda_j v_{ii} v_{jj}, \tag{37}$$

and

$$\text{tr}(n\mathbf{S})^3 = \sum_{i,j,k=1}^p \lambda_i \lambda_j \lambda_k v_{ij} v_{ik} v_{jk}, \tag{38}$$

$$\text{tr}^3(n\mathbf{S}) = \sum_{i,j,k=1}^p \lambda_i \lambda_j \lambda_k v_{ii} v_{jj} v_{kk}, \tag{39}$$

$$\text{tr}(n\mathbf{S}) \text{tr}(n\mathbf{S})^2 = \sum_{i,j,k=1}^p \lambda_i \lambda_j \lambda_k v_{ij}^2 v_{kk}. \tag{40}$$

By the moment properties of random variables v_{ii} and v_{ij} in [26,28], we have

$$\begin{aligned} \mathbb{E}[v_{ii}] &= n, & \mathbb{E}[v_{ii}^2] &= n(n+2), & \mathbb{E}[v_{ij}^2] &= n, \\ \mathbb{E}[v_{ii}^3] &= n(n+2), & \mathbb{E}[v_{ij}v_{ik}v_{jk}] &= n & \mathbb{E}[v_{ii}^3] &= n(n+2)(n+4). \end{aligned} \tag{41}$$

where i, j, k are arbitrary mutually unequal numbers in $\{1, \dots, p\}$. Next, we compute the mathematical expectation of Equations (38)–(40) based on the moment properties in (41). Denote $\mu_1 = \sum_{i=1}^p \lambda_i^3$, $\mu_2 = \sum_{i \neq j}^p \lambda_i^2 \lambda_j$ and $\mu_3 = \sum_{i \neq j \neq k}^p \lambda_i \lambda_j \lambda_k$, where $i \neq j \neq k$ means that i, j, k are mutually unequal, and we have

$$\mathbb{E}[\text{tr}(n\mathbf{S})^3] = n(n+2)(n+4)\mu_1 + 3n(n+2)\mu_2 + n\mu_3, \tag{42}$$

$$\mathbb{E}[\text{tr}^3(n\mathbf{S})] = n(n+2)(n+4)\mu_1 + 3n^2(n+2)\mu_2 + n^3\mu_3, \tag{43}$$

$$\mathbb{E}[\text{tr}(n\mathbf{S}) \text{tr}(n\mathbf{S})^2] = n(n+2)(n+4)\mu_1 + n(n+2)^2\mu_2 + n^2\mu_3. \tag{44}$$

Denote a matrix \mathbf{D} as

$$\mathbf{D} = \begin{bmatrix} n(n+2)(n+4) & 3n(n+2) & n \\ n(n+2)(n+4) & 3n^2(n+2) & n^3 \\ n(n+2)(n+4) & n(n+2)^2 & n^2 \end{bmatrix}, \tag{45}$$

then the above Equations (42)–(44) can be rewritten in the form of a matrix equation:

$$\begin{bmatrix} \mathbb{E}[\text{tr}(n\mathbf{S})^3] \\ \mathbb{E}[\text{tr}^3(n\mathbf{S})] \\ \mathbb{E}[\text{tr}(n\mathbf{S}) \text{tr}(n\mathbf{S})^2] \end{bmatrix} = \mathbf{D} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}. \tag{46}$$

Furthermore, the unknown quantities $b_i, i = 1, 2, 3$ can be decomposed as

$$b_1 = \mu_1, \quad b_2 = \mu_1 + 3\mu_2 + \mu_3, \quad b_3 = \mu_1 + \mu_2. \tag{47}$$

Therefore, we have

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}. \tag{48}$$

By the Equations (46) and (48), we have

$$\begin{bmatrix} \mathbb{E}[\text{tr}(n\mathbf{S})^3] \\ \mathbb{E}[\text{tr}^3(n\mathbf{S})] \\ \mathbb{E}[\text{tr}(n\mathbf{S}) \text{tr}(n\mathbf{S})^2] \end{bmatrix} = \mathbf{D} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}. \tag{49}$$

For $n \geq 3$, the following equality holds:

$$\mathbf{D} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 1 & 0 \end{bmatrix}^{-1} = (\tau_b \mathbf{W})^{-1}. \tag{50}$$

Therefore, we can obtain

$$\mathbb{E} \begin{bmatrix} \text{tr}(n\mathbf{S})^3 \\ \text{tr}^3(n\mathbf{S}) \\ \text{tr}(n\mathbf{S}) \text{tr}(n\mathbf{S})^2 \end{bmatrix} = (\tau_b \mathbf{W})^{-1} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}. \tag{51}$$

This completes the proof. \square

Then, the unknown scalars $b_i, i = 1, 2, 3$ can be unbiasedly estimated by

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \tau_b \mathbf{W} \begin{bmatrix} \text{tr}(\mathbf{S}^3) \\ \text{tr}^3(\mathbf{S}) \\ \text{tr}(\mathbf{S}) \text{tr}(\mathbf{S}^2) \end{bmatrix}. \tag{52}$$

Remark 2. In a large sample scenario, the unknown scalars $b_i, i = 1, 2, 3$ can be consistently estimated by $\text{tr}(\mathbf{S}^3)$, $\text{tr}^3(\mathbf{S})$ and $\text{tr}(\mathbf{S}) \text{tr}(\mathbf{S}^2)$ [29]. Theorem 1 shows that these estimates are biased. Moreover, the biases become non-ignorable in high-dimensional situations [30]. Furthermore, by Theorem 1, the biases can be eliminated by the linear combinations of $\text{tr}(\mathbf{S}^3)$, $\text{tr}^3(\mathbf{S})$, and $\text{tr}(\mathbf{S}) \text{tr}(\mathbf{S}^2)$.

Thirdly, denote the matrices \mathbf{G} , \mathbf{R} , and \mathbf{K} as follows:

$$\mathbf{G} = (g_{ij}) \text{ with } g_{ij} = \|x^i \circ x^j\|^2, \tag{53}$$

$$\mathbf{R} = (r_{ij}) \text{ with } r_{ij} = \|x^i \circ x^i \circ x^j\|^2, \tag{54}$$

$$\mathbf{K} = (k_{ij}) \text{ with } k_{ij} = (x^i \circ x^i)(x^i \circ x^j)^T, \tag{55}$$

where x^i is the observations of i -th variable for $i = 1, \dots, p$, then the following theorem holds.

Theorem 2. Under Gaussian distribution, the following equations hold when $n \geq 3$:

$$\mathbb{E}[n^3 \text{tr}(\mathbf{D}_S^3) - 3n \text{tr}(\mathbf{G}\mathbf{D}_S) + 2 \text{tr}(\mathbf{R})] = \tau_c^{-1} c_1, \tag{56}$$

$$\mathbb{E}[n^3 \text{tr}(\mathbf{D}_S \mathbf{S}^2) - 2n \text{tr}(\mathbf{K}\mathbf{S}) - n \text{tr}(\mathbf{G}\mathbf{D}_S \mathbf{E}) + 2 \text{tr}(\mathbf{R}\mathbf{E})] = \tau_c^{-1} c_2, \tag{57}$$

where $\tau_c = \frac{1}{n(n-1)(n-2)}$.

Proof. Let $\mathbf{F} = (f_{ij})$, and \mathbf{F} is a symmetric matrix; then, we have $\mathbf{F}^2 = \mathbf{\Sigma} = (\sigma_{ij})$. Therefore, for arbitrary $i, j \in \{1, \dots, p\}$, the equalities $\sigma_{ij} = \sum_{k=1}^p f_{ik}f_{jk}$ and $\sigma_{ii} = \sum_{k=1}^p f_{ik}^2$ hold. For $m = 1, \dots, n$, denote $x_m = (x_{m1}, x_{m2}, \dots, x_{mp})^T$ and $\mathbf{z}_m = \mathbf{F}^{-1}x_m$, then \mathbf{z}_m is an i.i.d. sample and $\mathbf{z}_m \sim N(\mathbf{0}, \mathbf{I}_p)$. Let $\mathbf{z}_m = (z_{m1}, z_{m2}, \dots, z_{mp})^T$, then $z_{mk}, m = 1, \dots, n, k = 1, \dots, p$ are mutually independent standard Gaussian random variables. For arbitrary $m \in \{1, \dots, n\}$ and $i, j \in \{1, \dots, p\}$, we have $x_{mi} = \sum_{k=1}^p f_{ik}z_{mk}$ and $x_{mj} = \sum_{k=1}^p f_{jk}z_{mk}$. Denote that SCM $\mathbf{S} = (s_{ij})$, then s_{ij} can be decomposed as follows:

$$s_{ij} = \frac{1}{n} \sum_{m=1}^n x_{mi}x_{mj} = \frac{1}{n} \sum_{m=1}^n \sum_{k_1, k_2=1}^p f_{ik_1}f_{jk_2}z_{mk_1}z_{mk_2}. \tag{58}$$

Then, for arbitrary $m \in \{1, \dots, n\}$ and $i, j \in \{1, \dots, p\}$, we have

$$\mathbb{E}[x_{mi}x_{mj}] = \sum_{k_1, k_2=1}^p f_{ik_1}f_{jk_2}\mathbb{E}[z_{mk_1}z_{mk_2}] = \sum_{k=1}^p f_{ik}f_{jk} = \sigma_{ij}. \tag{59}$$

Especially when $i = j$, we have $\mathbb{E}[x_{mi}^2] = \sigma_{ii}$. Then, we can obtain

$$\begin{aligned} \mathbb{E}[n^3 \text{tr}(\mathbf{D}_S^3) - 3n \text{tr}(\mathbf{GD}_S) + 2 \text{tr}(\mathbf{R})] &= \sum_{m_1 \neq m_2 \neq m_3}^n \sum_{i=1}^p \mathbb{E}[x_{m_1 i}^2 x_{m_2 i}^2 x_{m_3 i}^2] \\ &= \sum_{m_1 \neq m_2 \neq m_3}^n \sum_{i=1}^p \mathbb{E}[x_{m_1 i}^2] \mathbb{E}[x_{m_2 i}^2] \mathbb{E}[x_{m_3 i}^2] = \sum_{m_1 \neq m_2 \neq m_3}^n \sum_{i=1}^p \sigma_{ii}^3 \\ &= n(n-1)(n-2) \sum_{i=1}^p \sigma_{ii}^3 = \tau_c^{-1} \text{tr}(\mathbf{D}_S^3). \end{aligned} \tag{60}$$

Furthermore, we have

$$\begin{aligned} &\mathbb{E}[n^3 \text{tr}(\mathbf{D}_S \mathbf{S}^2) - 2n \text{tr}(\mathbf{KS}) - n \text{tr}(\mathbf{GD}_S \mathbf{E}) + 2 \text{tr}(\mathbf{RE})] \\ &= \sum_{m_1 \neq m_2 \neq m_3}^n \sum_{i, j=1}^p \mathbb{E}[x_{m_1 i}^2 x_{m_2 i} x_{m_2 j} x_{m_3 i} x_{m_3 j}] = \sum_{m_1 \neq m_2 \neq m_3}^n \sum_{i, j=1}^p \mathbb{E}[x_{m_1 i}^2] \mathbb{E}[x_{m_2 i} x_{m_2 j}] \mathbb{E}[x_{m_3 i} x_{m_3 j}] \\ &= \sum_{m_1 \neq m_2 \neq m_3}^n \sum_{i, j=1}^p \sigma_{ii} \sigma_{ij}^2 = n(n-1)(n-2) \sum_{i, j=1}^p \sigma_{ii} \sigma_{ij}^2 = \tau_c^{-1} \text{tr}(\mathbf{D}_S \mathbf{\Sigma}^2). \end{aligned} \tag{61}$$

□

Then, the unknown scalars c_1 and c_2 can be unbiasedly estimated by

$$\gamma_1 = \tau_c [n^3 \text{tr}(\mathbf{D}_S^3) - 3n \text{tr}(\mathbf{GD}_S) + 2 \text{tr}(\mathbf{R})], \tag{62}$$

$$\gamma_2 = \tau_c [n^3 \text{tr}(\mathbf{D}_S \mathbf{S}^2) - 2n \text{tr}(\mathbf{KS}) - n \text{tr}(\mathbf{GD}_S \mathbf{E}) + 2 \text{tr}(\mathbf{RE})]. \tag{63}$$

By plugging the estimates of quantities in (27) into (28) and (29), the unbiased estimates of $\mathbf{v}(\mathbf{T}_1)$ and $\mathbf{v}(\mathbf{T}_2)$ are given by

$$\hat{\mathbf{v}}(\mathbf{T}_1) = \begin{bmatrix} \frac{1}{p} \alpha_2 - \alpha_1 \\ \frac{np+p-4}{np} \beta_1 + \frac{1}{p^2} \beta_2 + \frac{p^2-2np+2}{np^2} \beta_3 \end{bmatrix}, \tag{64}$$

$$\hat{\mathbf{v}}(\mathbf{T}_2) = \begin{bmatrix} \alpha_3 - \alpha_1 \\ \frac{n+1}{n} \beta_1 + \frac{1}{n} \beta_3 + \frac{n+2}{n} \gamma_1 - \frac{2n+4}{n} \gamma_2 \end{bmatrix}. \tag{65}$$

By the Equations (16), (64), and (65), the unbiased estimates of the vectors $\mathbf{b}(\mathbf{T}_i), i = 1, 2$ are given by

$$\hat{\mathbf{b}}(\mathbf{T}_i) = \hat{\mathbf{u}}(\mathbf{T}_i) + \hat{\mathbf{v}}(\mathbf{T}_i), i = 1, 2. \tag{66}$$

In addition, the constant c in (10) can be further calculated under Gaussian distribution, which is

$$c = \frac{1}{n}a_1 + \frac{1}{n}a_2. \tag{67}$$

Therefore, we can obtain that the unbiased estimate of c is given by

$$\hat{c} = \frac{1}{n}\alpha_1 + \frac{1}{n}\alpha_2 = \tau_a \left(\frac{n-2}{n} \text{tr}(\mathbf{S}^2) + \text{tr}^2(\mathbf{S}) \right). \tag{68}$$

To sum up, by Equations (13), (66) and (68), we can obtain that the unbiased estimates of the loss functions $\mathcal{M}_{\mathbf{T}_i}(\mathbf{w}|\Sigma)$ are given by

$$\hat{\mathcal{M}}_{\mathbf{T}_i}(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T \hat{\mathbf{H}}(\mathbf{T}_i)\mathbf{w} - 2\mathbf{w}^T \hat{\mathbf{b}}(\mathbf{T}_i) + \hat{c}. \tag{69}$$

3.3. Optimal Second-Order Stein-Type Estimators

For the target matrices $\mathbf{T}_i, i = 1, 2$, through replacing the objective function in (10) with its unbiased estimate given by (69), we further formulate the second-order Stein-type estimators as the following optimization problems:

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{w}^T \hat{\mathbf{H}}(\mathbf{T}_i)\mathbf{w} - 2\mathbf{w}^T \hat{\mathbf{b}}(\mathbf{T}_i) + \hat{c} \\ \text{s.t.} \quad & 0 \leq \mathbf{e}_1 \mathbf{w} \leq 1, \\ & \mathbf{e}_2 \mathbf{w} \geq 0. \end{aligned} \tag{70}$$

For $i = 1, 2$, the Hessian matrix of the objective function in (70) is $\hat{\mathbf{H}}(\mathbf{T}_i)$. By the following inequality

$$\left(\text{tr}(\mathbf{S} - \mathbf{T}_i)^3 \right)^2 \leq \text{tr}(\mathbf{S} - \mathbf{T}_i)^2 \text{tr}(\mathbf{S} - \mathbf{T}_i)^4, \tag{71}$$

we can obtain that $\hat{\mathbf{H}}(\mathbf{T}_i)$ is positive definite. Therefore, the optimization problem (70) is a convex quadratic program. Furthermore, we can obtain the globally optimal solution by an efficient algorithm.

For the target matrices $\mathbf{T}_i, i = 1, 2$, by denoting the corresponding optimal tuning parameters as $\mathbf{w}^i = (w_1^i, w_2^i)^T$, the optimal second-order Stein-type estimators can be expressed as

$$\hat{\Sigma}_i = \mathbf{S} + w_1^i(\mathbf{T} - \mathbf{S}) + w_2^i(\mathbf{T} - \mathbf{S})^2. \tag{72}$$

Remark 3. The proposed second-order Stein-type estimators are both well conditioned. Moreover, by taking the spectral decomposition, the SCM can be expressed as

$$\mathbf{S} = \mathbf{U}^T \Delta \mathbf{U}, \tag{73}$$

where $\Delta = \text{diag}(\delta_1, \dots, \delta_p)$ is a diagonal matrix consisting of the eigenvalues and \mathbf{U} is the corresponding unitary matrix. Then, the second-order Stein-type estimator $\hat{\Sigma}_1$ has the following spectral decomposition:

$$\hat{\Sigma}_1 = \mathbf{U}^T \tilde{\Delta} \mathbf{U}, \tag{74}$$

where $\tilde{\Delta} = \text{diag}(\tilde{\delta}_1, \dots, \tilde{\delta}_p)$ with

$$\tilde{\delta}_i = \delta_i + w_1^1(\theta - \delta_i) + w_2^1(\theta - \delta_i)^2. \tag{75}$$

where $\theta = \frac{\text{tr}(\mathbf{S})}{p}$ is the mean of sample eigenvalues. Therefore, the proposed estimator $\hat{\Sigma}_1$ shrinks the sample eigenvalues by a nonlinear transformation whileretaining the sample eigenvectors.

4. Numerical Simulations and Real Data Analysis

This section presents numerical simulations and two application examples to discover the performance of the proposed second-order Stein-type estimators. The proposed covariance matrix estimators for the target matrices T_1 and T_2 are denoted as QS-T1 and QS-T2, respectively. The control estimators include the Stein-type estimator LS-T1 for T_1 in [11], and the Stein-type estimator LS-T2 for T_2 in [12] and the nonlinear shrinkage estimator NS developed in [18].

4.1. MSE Performance

We assume that the actual distribution is $\mathcal{N}(\mathbf{0}, \Sigma)$, where the following models generate the covariance matrix:

- (1) Model 1: $\Sigma = (\sigma_{ij})_{p \times p}$ with $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.1^{|i-j|}$ for $i \neq j$,
- (2) Model 2: $\Sigma = (\sigma_{ij})_{p \times p}$ with $\sigma_{ii} = \max\{50 - i, 0\} + 2$ and $\sigma_{ij} = 0.6$ for $i \neq j$.

Under Model 1, the diagonal elements are equal to 1, and the off-diagonal elements are tiny. Therefore, the covariance matrix is close to a spherical matrix. Under Model 2, the diagonal elements are dispersive, and the off-diagonal elements correspond to weak correlations. Therefore, the covariance matrix is close to a diagonal matrix. We carry out random sampling in each Monte Carlo run and compute the Frobenius loss of each covariance matrix estimator. The MSE performance of each covariance matrix estimator is evaluated by averaging the Frobenius losses of 5×10^3 runs.

Figures 1 and 2 report the logarithmic Frobenius loss of each estimator in large-dimensional scenarios where the dimension is 180 and the sample size varies from 10 to 100. Under Model 1, the Stein-type estimator LS-T1 and the second-order Stein-type estimator QS-T1 outperform the nonlinear shrinkage estimator NS and the shrinkage estimators LS-T2 and QS-T2, which employ the diagonal target matrix. Furthermore, the proposed second-order Stein-type estimator QS-T1 shows a significant advantage over the Stein-type estimator LS-T1, especially when the sample size is tiny. Similarly, under Model 2, the Stein-type estimator LS-T2 and the second-order Stein-type estimator QS-T2 perform better than the other three estimators. Moreover, the proposed estimator QS-T2 outperforms the corresponding Stein-type estimator LS-T2. Therefore, when the correct target matrix is employed, the second-order Stein-type estimators enjoy lower Frobenius losses than the linear and nonlinear shrinkage estimators in large-dimensional scenarios.

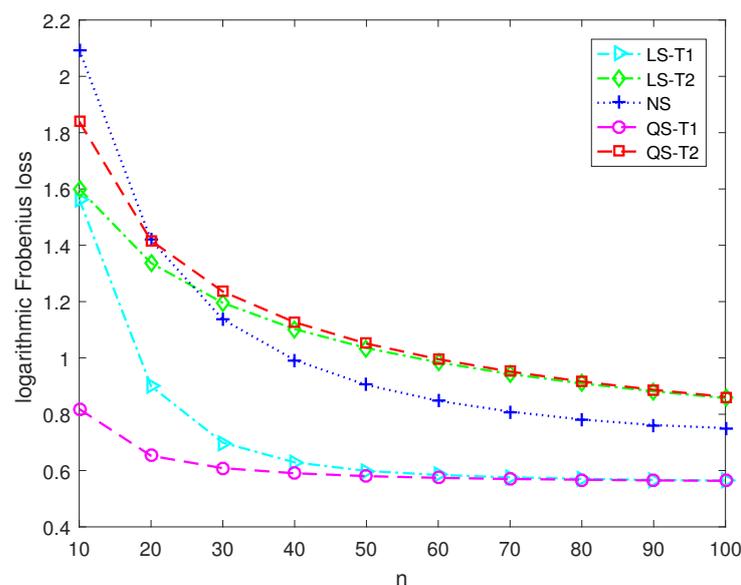


Figure 1. The logarithmic Frobenius losses of shrinkage estimators under Model 1 with $p = 180$.

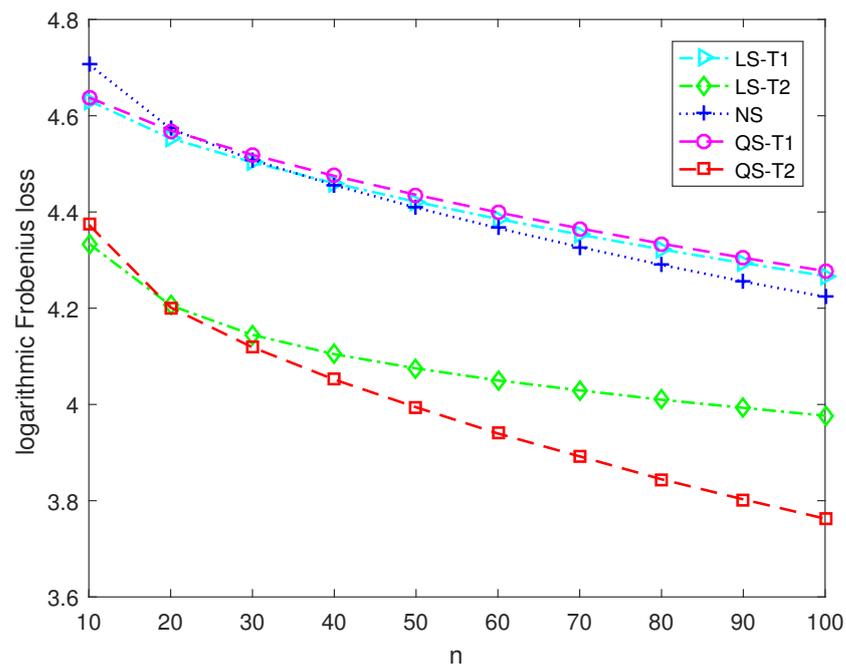


Figure 2. The logarithmic Frobenius losses of shrinkage estimators under Model 2 with $p = 180$.

Figures 3 and 4 report the logarithmic Frobenius loss of each estimator versus the dimension. The sample size is 30. We can find that the logarithmic Frobenius losses become more prominent as the dimension increases. In Figure 3, the actual covariance matrix is close to being spherical. The proposed second-order Stein-type estimator QS-T1 significantly outperforms the other estimators. In Figure 4, the actual covariance matrix is close to being diagonal. The proposed second-order Stein-type estimator QS-T2 enjoys lower Frobenius loss when the dimension exceeds 120.

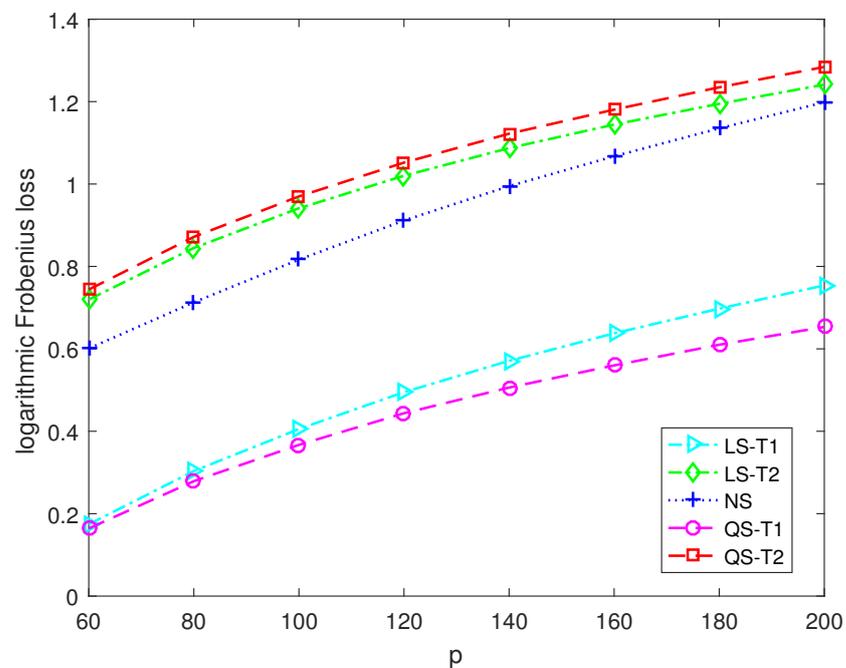


Figure 3. The logarithmic Frobenius losses of shrinkage estimators under Model 1 with $n = 30$.

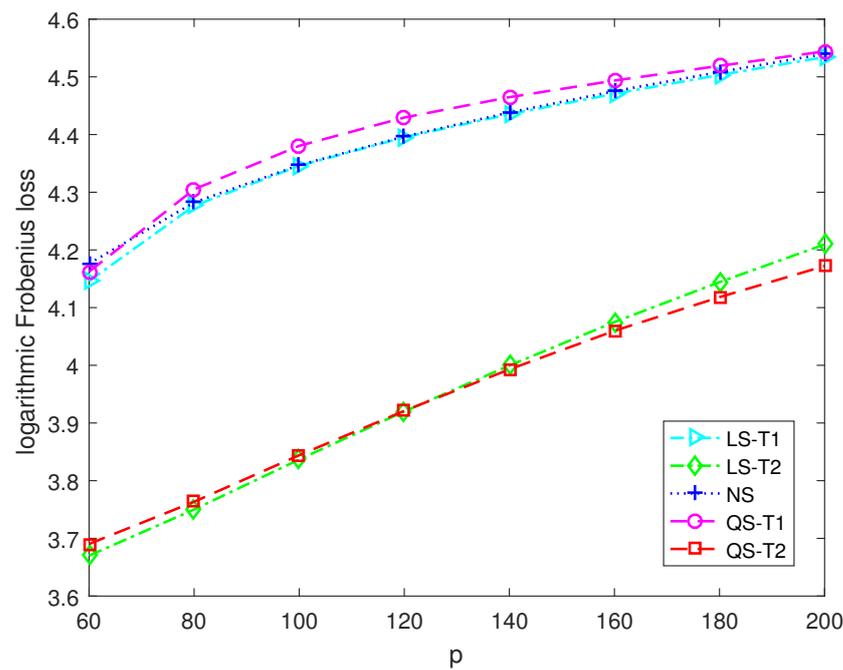


Figure 4. The logarithmic Frobenius losses of shrinkage estimators under Model 2 with $n = 30$.

The proposed second-order Stein-type estimators take significant advantage of the MSE performance over the linear and nonlinear shrinkage estimators, especially when the dimension is large compared to the sample size.

4.2. Portfolio Selection

In finance, assets with higher expected returns generally involve higher risks. Therefore, investors must constantly balance the expected return and the risk tolerance. The portfolio strategy is a popular way to reduce risk and enhance return. Therefore, portfolio selection plays a vital role in asset investment.

In 1952, Markowitz introduced the famous mean-variance optimization to determine the optimal portfolio weights [31]. Let \mathbf{m} and Σ be the expectation and covariance matrix of the daily returns. For portfolio weight \mathbf{k} , the variance of the portfolio is defined as $\sigma^2 = \mathbf{k}^T \Sigma \mathbf{k}$ in the Markowitz framework. As short selling is forbidden, the Markowitz portfolio optimization is formulated as the following mean-variance problem:

$$\begin{aligned} \min \quad & \mathbf{k}^T \Sigma \mathbf{k} \\ \text{s.t.} \quad & \mathbf{k}^T \mathbf{m} = r, \mathbf{k}^T \mathbf{1} = 1, \mathbf{k} \geq 0, \end{aligned} \tag{76}$$

where r is a given expected return. By, respectively, replacing \mathbf{m} and Σ with their estimates $\hat{\mathbf{m}}$ and $\hat{\Sigma}$, the optimal weight \mathbf{k}_r can be solved by efficient quadratic programming algorithm. It is obvious that the Markowitz optimization only depends on the estimates of the first and second moments of the daily returns. The sample mean and the SCM in the classic portfolio perform well in the portfolio risk measurement; however, the SCM becomes unstable as the number of stocks is large, resulting in significant property loss [32]. Therefore, a well-performed covariance matrix estimator is important in current portfolio selection [2,33].

In practice, we consider a portfolio consisting of $p = 95$ highly capitalized stocks from the New York stock exchange with ticker symbols AA, ABT, AIG, AIR, ALL, AMD, AP, APA, AXP, BA, BAC, BAX, BEN, BK, BMY, C, CAT, CCL, CHK, CL, COP, CPE, CVS, CVX, D, DB, DD, DE, DNR, DVN, EAT, EME, EMR, EXC, FCX, FDX, FNMA, GD, GE, GILD, GLW, HAL, HD, HIG, HON, HPQ, IBM, INTC, ITW, JNJ, JPM, KMB, KO, L, LLY, LMT, LOW, M, MCD, MDT, MMM, MO, MRK, MRO, MS, NBR, NC, NE, NEE, NL, NNN, ODC, OXM, OXY, PCG, PEP, PFE, PG, RIG, SLB, SO, T, TGT, TRV, TXN, UNH, UNP,

USB, VLO, VZ, WFC, WMT, WWW, X, XOM. The dataset \mathbf{X} contains $n = 536$ daily close prices from 11 November 2016 to 31 December 2018 collected via Yahoo! Finance at <https://au.finance.yahoo.com/lookup?s=DATA>, accessed on 10 November 2022. For each stock i , the daily close price is preprocessed as the daily return by

$$\tilde{\mathbf{X}}(i, j) = \frac{\mathbf{X}(i, j + 1)}{\mathbf{X}(i, j)} - 1, j = 1, \dots, n - 1. \tag{77}$$

The covariance matrix estimators LS-T1, LS-T2, NS, QS-T1, and QS-T2 are generated from the daily return $\tilde{\mathbf{X}}$. For an expected return r , the realized risk is defined as $\sigma_r = \sqrt{\mathbf{k}_r^T \hat{\Sigma} \mathbf{k}_r}$. Next, we employ the realized risks, one key index to evaluate the portfolio, to verify the performance of the covariance matrix estimators.

Figures 5 and 6 plot the realized risks of three kinds of shrinkage estimators for different investment horizons. For a short-term investment of 44 trading days (2 months), the proposed QS-T2 has the lowest realized risk when the expected return is less than 0.5% and has the highest realized risk when the expected return exceeds 0.6%. The proposed QS-T1 and the Stein-type estimator LS-T1 have the lowest realized risk when the expected return exceeds 0.5%. For a long-term investment of 280 trading days (13 months), the expected return of the five estimators becomes similar for the short-term investment.

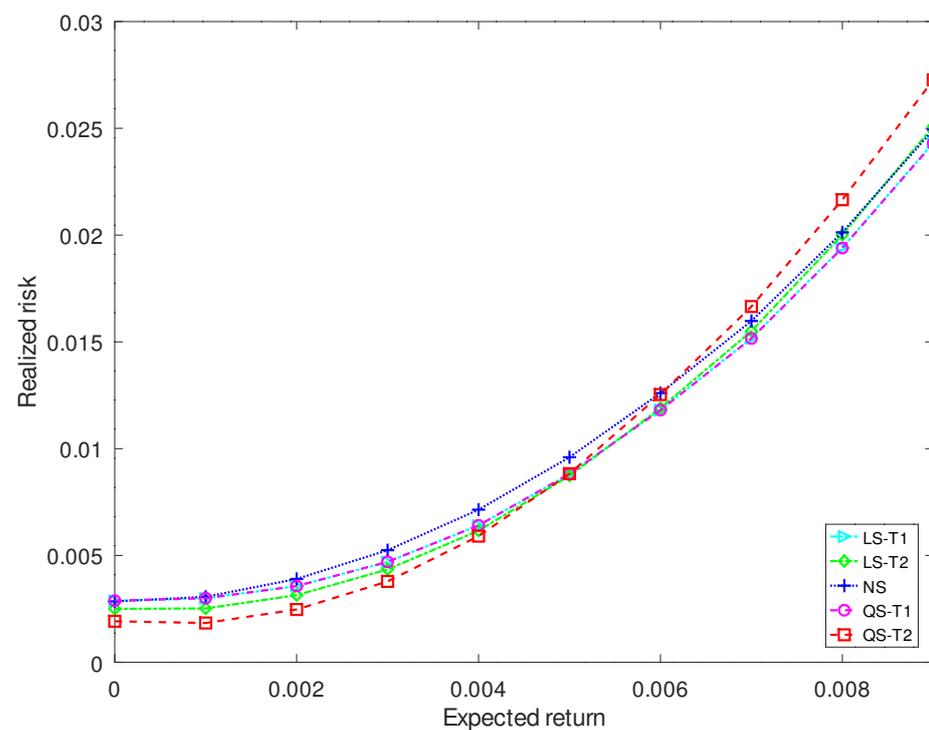


Figure 5. Realized risk of each covariance matrix estimator when the investment horizon contains 44 trading days.

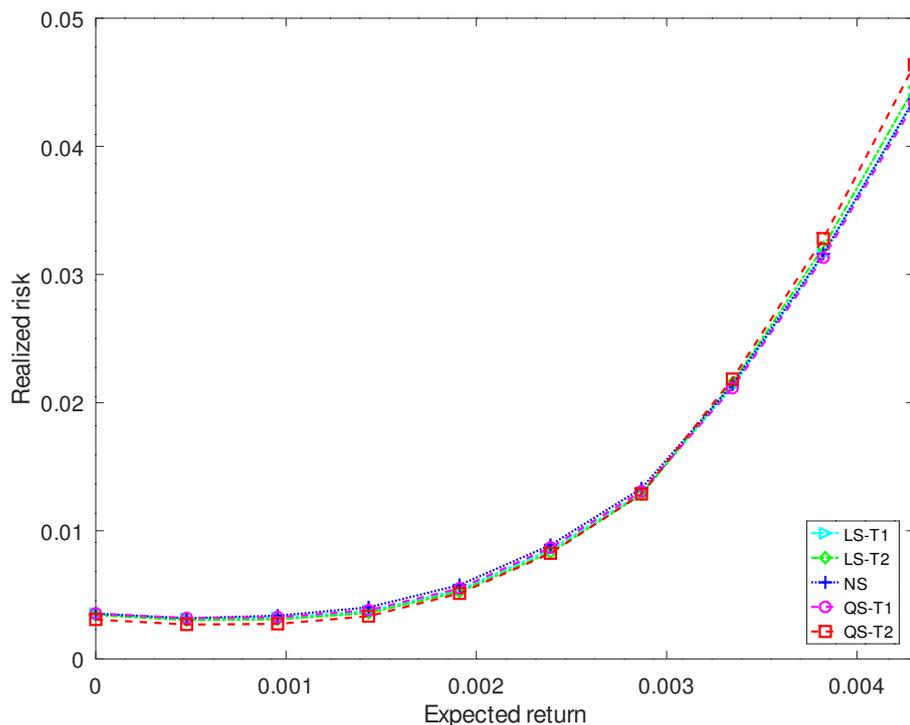


Figure 6. Realized risk of each covariance matrix estimator when the investment horizon contains 280 trading days.

Figures 7 and 8 plot the realized risks of shrinkage estimators for different expected return levels. The proposed estimator QS-T2 enjoys the lowest realized risk for a low expected return. The nonlinear shrinkage estimator NS has the highest realized risk. However, the proposed estimator QS-T2 performs worst when the expected return becomes high. The proposed estimator QS-T1, together with NS and LS-T1, performs best in this scenario.

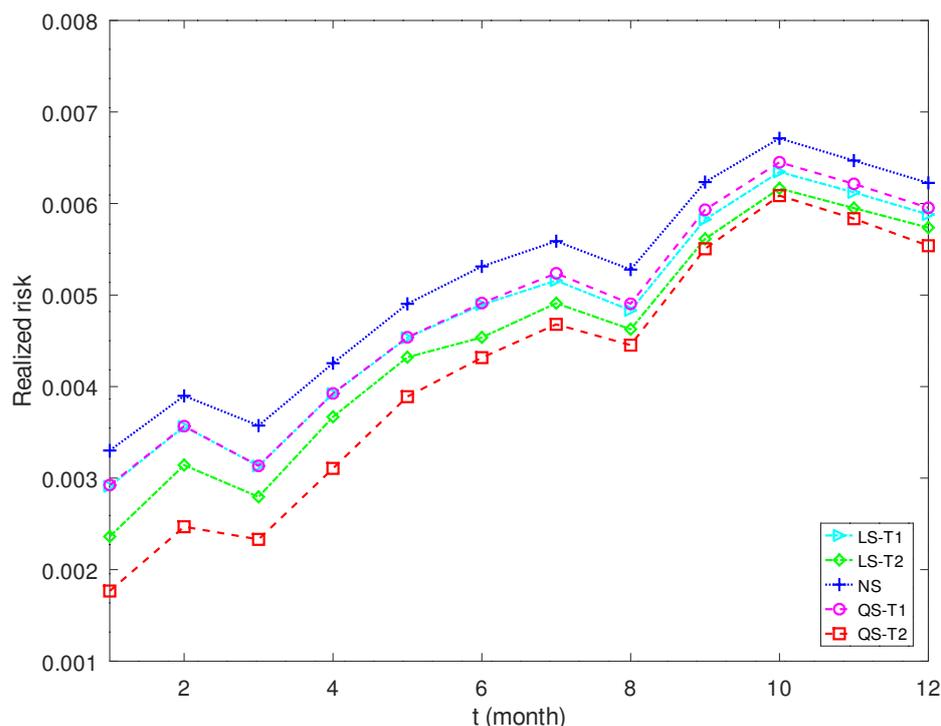


Figure 7. Realized risk of each covariance matrix estimator when the expected return is 0.2%.

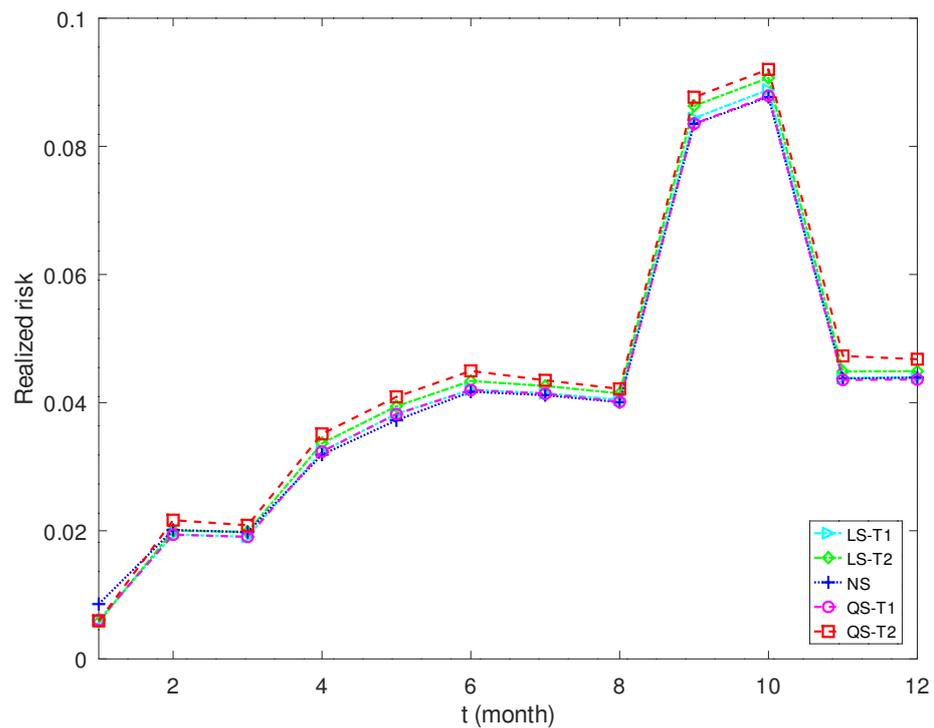


Figure 8. Realized risk of each covariance matrix estimator when the expected return is 0.8%.

The proposed second-order Stein-type estimator QS-T2 enjoys good portfolio selection for short-term investment and prudent return cases. The proposed second-order Stein-type estimator QS-T1 is recommended in long-term investment and high-return scenarios.

4.3. Discriminant Analysis

We further discover the performance of the second-order regularized estimators in small-sample-size situations. The Parkinson’s data are collected on the website <https://archive-beta.ics.uci.edu/>, accessed on 10 November 2022. $p = 160$ biomedical voice attitudes are measured from n_1 patients and n_2 healthy individuals. Let $\hat{\Sigma}$ be the pooling estimator based on a certain estimation strategy. We use the following quadratic discriminant rule $M = (x_i - \bar{x})^T \hat{\Sigma}^{-1} (x_i - \bar{x})$ to make a diagnosis for each x_i . M denotes the Mahalanobis distance between individual x_i and the sample center \bar{x} . The individual x_i is classified as Parkinson’s patient if x_i is closer to the sample center of patients in the sense of Mahalanobis distance.

Table 1 reports the classification accuracy rates of different estimators. We can see that the Stein-type estimators LS-T1, LS-T2, QS-T1, and LS-T2 perform better than NS. Moreover, the accuracy rate of the Stein-type estimators becomes larger as the sample size increases. Moreover, the proposed second-order Stein-type estimator QS-T2 enjoys the largest accuracy rate when $n \leq 40$, and the Stein-type estimator LS-T2 has the best performance when $n \geq 45$. Therefore, we can see that the proposed second-order Stein-type estimation performs better than the classic Stein-type estimation in this application.

Table 1. The classification accuracy rate of Parkinson’s data.

n_2	15	20	25	30	35	40	45	50
LS-T1	0.5165	0.5530	0.5646	0.5826	0.5870	0.5956	0.5981	0.6054
LS-T2	0.6262	0.6459	0.6607	0.6714	0.6796	0.6892	0.6918	0.6971
NS	0.4931	0.4842	0.4701	0.4688	0.4544	0.4458	0.4285	0.4304
QS-T1	0.5124	0.5465	0.5600	0.5761	0.5833	0.5938	0.5953	0.6025
QS-T2	0.6336	0.6539	0.6678	0.6747	0.6840	0.6902	0.6916	0.6957

5. Conclusions and Discussion

This paper investigated the problem of estimating a large-dimensional covariance matrix. Motivated by Stein's strategy, we developed a novel strategy named the second-order Stein-type estimation. The proposed estimator is expected to be positive definite in the form of a quadratic binomial of the SCM and the target matrix. Firstly, we specified the spherical and diagonal targets in the second-order Stein-type regularization. The mean squared errors were, respectively, obtained for the two targets. Secondly, we unbiasedly estimated the two mean squared errors under the Gaussian distribution. Thirdly, the optimal parameters were obtained by solving the convex quadratic programming. The optimal second-order Stein-type estimators were obtained for the two target matrices. Finally, we verified the performance of the proposed estimators in numerical simulations and real data applications.

It is worth mentioning that the second-order Stein-type estimators were proposed under Gaussian distribution. In practical applications, the data may often deviate from the Gaussian distribution. Therefore, the problem of investigating the second-order Stein-type regularization under non-Gaussian distributions remains open and important.

Author Contributions: Conceptualization, B.Z.; methodology, B.Z.; software, B.Z.; validation, B.Z.; formal analysis, B.Z.; investigation, H.H.; resources B.Z.; data curation, B.Z.; writing—original draft preparation, B.Z.; writing—review and editing, H.H. and J.C.; visualization, B.Z.; supervision, B.Z.; project administration, B.Z. and H.H.; funding acquisition, B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Guangxi Science and Technology Planning Project (2022AC21276), Science and Technology Project of Guangxi Guike(AD21220114), National Natural Science Foundation of China (Grant Nos. 11861071 and 12261011), and Beijing Institute of Technology Research Fund Program for Young Scholars.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: All data included in this study are available upon request by contact with the first author.

Conflicts of Interest: The authors declare that there are no conflict of interest regarding the publication of this paper.

References

1. Fan, J.; Liao, Y.; Mincheva, M. Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2013**, *75*, 603–680. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Bodnar, T.; amd Yarema Okhrin, S.D.; Parolya, N.; Schmid, W. Statistical inference for the expected utility portfolio in high dimensions. *IEEE Trans. Signal Process.* **2021**, *69*. [\[CrossRef\]](#)
3. Vershynin, R. How close is the sample covariance matrix to the actual covariance matrix? *J. Theor. Probab.* **2012**, *25*, 655–686. [\[CrossRef\]](#)
4. Cai, T.T.; Han, X.; Pan, G. Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices. *Ann. Stat.* **2020**, *48*, 1255–1280. [\[CrossRef\]](#)
5. Wu, W.; Pourahmadi, M. Banding sample autocovariance matrices of stationary progress. *Stat. Sin.* **2009**, *19*, 1755–1768.
6. Fan, J.; Liu, H.; Wang, W. Large covariance estimation through elliptical factor models. *Ann. Stat.* **2018**, *46*, 1383–1414. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Cai, T.T.; Yuan, M. Adaptive covariance matrix estimation through block thresholding. *Ann. Stat.* **2012**, *40*, 2014–2042. [\[CrossRef\]](#)
8. Cao, Y.; Lin, W.; Li, H. Large covariance estimation for compositional data via composition-adjusted thresholding. *J. Am. Stat. Assoc.* **2019**, *114*, 759–772. [\[CrossRef\]](#)
9. Bodnar, O.; Bodnar, T.; Parolya, N. Recent advances in shrinkage-based high-dimensional inference. *J. Multivar. Anal.* **2021**, *188*, 104826. [\[CrossRef\]](#)
10. Raninen, E.; Ollila, E. Coupled regularized sample covariance matrix estimator for multiple classes. *IEEE Trans. Signal Process.* **2021**, *69*, 5681–5692. [\[CrossRef\]](#)
11. Chen, Y.; Wiesel, A.; Eldar, Y.C.; Hero, A.O. Shrinkage algorithms for MMSE covariance estimation. *IEEE Trans. Signal Process.* **2010**, *58*, 5016–5029. [\[CrossRef\]](#)
12. Fisher, T.J.; Sun, X. Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix. *Comput. Stat. Data Anal.* **2011**, *55*, 1909–1918.

13. Hannart, A.; Naveau, P. Estimating high dimensional covariance matrices: A new look at the Gaussian conjugate framework. *J. Multivar. Anal.* **2014**, *131*, 149–162. [[CrossRef](#)]
14. Liu, Y.; Sun, X.; Zhao, S. A covariance matrix shrinkage method with Toeplitz rectified target for DOA estimation under the uniform linear array. *Int. J. Electron. Commun. (AEÜ)* **2017**, *81*, 50–55. [[CrossRef](#)]
15. Lancewicki, T.; Aladjem, M. Multi-target shrinkage estimation for covariance matrices. *Trans. Signal Process.* **2014**, *62*, 6380–6390. [[CrossRef](#)]
16. Tong, J.; Hu, R.; Xi, J.; Xiao, Z.; Guo, Q.; Yu, Y. Linear shrinkage estimation of covariance matrices using low-complexity cross-validation. *Signal Process.* **2018**, *148*, 223–233. [[CrossRef](#)]
17. Yuasa, R.; Kubokawa, T. Ridge-type linear shrinkage estimation of the mean matrix of a high-dimensional normal distribution. *J. Multivar. Anal.* **2020**, *178*, 104608. [[CrossRef](#)]
18. Ledoit, O.; Wolf, M. Analytical nonlinear shrinkage of large-dimensional covariance matrices. *Ann. Stat.* **2020**, *48*, 3043–3065. [[CrossRef](#)]
19. Mestre, X. On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices. *IEEE Trans. Signal Process.* **2008**, *56*, 5353–5368. [[CrossRef](#)]
20. Ledoit, O.; Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* **2004**, *88*, 365–411. [[CrossRef](#)]
21. Ikeda, Y.; Kubokawa, T.; Srivastava, M.S. Comparison of linear shrinkage estimators of a large covariance matrix in normal and non-normal distributions. *Comput. Stat. Data Anal.* **2016**, *95*, 95–108.
22. Cabana, E.; Lillo, R.E.; Laniado, H. Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators. *Stat. Pap.* **2021**, *62*, 1583–1609. [[CrossRef](#)]
23. Ledoit, O.; Wolf, M. Shrinkage estimation of large covariance matrices: Keep it simple, statistician? *J. Multivar. Anal.* **2021**, *186*, 104796. [[CrossRef](#)]
24. Tanaka, M.; Nakata, K. Positive definite matrix approximation with condition number constraint. *Optim. Lett.* **2014**, *8*, 939–947. [[CrossRef](#)]
25. Gupta, A.K.; Nagar, D.K. *Matrix Variate Distributions*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2000.
26. Srivastava, M.S. Some tests concerning the covariance matrix in high dimensional data. *J. Jpn. Stat. Soc.* **2005**, *35*, 251–272. [[CrossRef](#)]
27. Li, J.; Zhou, J.; Zhang, B. Estimation of large covariance matrices by shrinking to structured target in normal and non-normal distributions. *IEEE Access* **2018**, *6*, 2158–2169. [[CrossRef](#)]
28. Fisher, T.J.; Sun, X.; Gallagher, C.M. A new test for sphericity of the covariance matrix for high dimensional data. *J. Multivar. Anal.* **2010**, *101*, 2554–2570. [[CrossRef](#)]
29. Lehmann, E.L. *Elements of Large-Sample Theory*; Springer: New York, NY, USA, 1999.
30. Bai, Z.; Silverstein, J.W. *Spectral Analysis of Large Dimensional Random Matrices*; Springer: New York, NY, USA, 2010.
31. Markowitz, H. Portfolio selection. *J. Financ.* **1952**, *7*, 77–91.
32. Pantaleo, E.; Tumminello, M.; Lillo, F.; Mantegna, R. When do improved covariance matrix estimators enhance portfolio optimization? An empirical comparative study of nine estimators. *Quant. Financ. Pap.* **2010**, *11*, 1067–1080. [[CrossRef](#)]
33. Joo, Y.C.; Park, S.Y. Optimal portfolio selection using a simple double-shrinkage selection rule. *Financ. Res. Lett.* **2021**, *43*, 102019. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.