CIRJE-F-855

Estimation of Covariance and Precision Matrices in High Dimension

Tatsuya Kubokawa
University of Tokyo

Akira Inoue
Graduate School of Economics, University of Tokyo

July 2012

Estimation of Covariance and Precision Matrices in High Dimension

Tatsuya Kubokawa* and Akira Inoue†

University of Tokyo

July 21, 2012

Abstract

The problem of estimating covariance and precision matrices of multivariate normal distributions is addressed when both the sample size and the dimension of variables are large. The estimation of the precision matrix is important in various statistical inference including the Fisher linear discriminant analysis, confidence region based on the Mahalanobis distance and others. A standard estimator is the inverse of the sample covariance matrix, but it may be instable or can not be defined in the high dimension. Although (adaptive) ridge type estimators are alternative procedures which are useful and stable for large dimension. However, we are faced with questions about how to choose ridge parameters and their estimators and how to set up asymptotic order in ridge functions in high dimensional cases. In this paper, we consider general types of ridge estimators for covariance and precision matrices, and derive asymptotic expansions of their risk functions. Then we suggest the ridge functions so that the second order terms of risks of ridge estimators are smaller than those of risks of the standard estimators.

Key words and phrases: Asymptotic expansion, covariance matrix, high dimension, Moore-Penrose inverse, multivariate normal distribution, point estimation, precision matrix, ridge estimator, risk comparison, Stein-Haff identity, Stein loss, Wishart distribution.

1 Introduction

Statistical inference with high dimension has received much attention recent years and has been actively studied from both theoretical and practical aspects in the literature. Of these, estimate of the precision matrix is required in many multivariate inference procedures including the Fisher linear discriminant analysis, confidence intervals based

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

†Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN E-Mail: inoue-akira@hotmail.com

on the Mahalanobis distance and weighted least squares estimator in multivariate linear regression models. A standard estimator of the precision based on the sample covariance matrix is likely to be instable when the dimension p is large and close to the sample size N even if $N > p$. In the case of $p > N$, the inverse of the sample covariance matrix cannot be defined, and an estimator based on the Moore-Penrose generalized inverse of the sample covariance matrix has been used in Srivastava (2005). Another useful and stable estimator for the precision matrix is a ridge estimator, and its various variants have been used in literature. For example, see Ledoit and Wolf (2003, 2004), Fisher and Sun (2011) and Bai and Shi (2011). However, superiority of the ridge-type estimators over the standard estimators have not been studied except Kubokawa and Srivastava (2008), who obtained exact conditions for the ridge-type estimators to have uniformly smaller risks than the standard estimator. However, their results are limited to specific ridge functions and special loss functions.

To specify the problem considered here, let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ be independently and identically distributed (i.i.d.) as multivariate normal with mean vector $\boldsymbol{\mu}$ and $p \times p$ positive definite covariance matrix $\boldsymbol{\Sigma}$ denoted as $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} > 0$. Let $\overline{\boldsymbol{y}} = N^{-1}\sum_{i=1}^N \boldsymbol{y}_i$, $\boldsymbol{V} = \sum_{i=1}^N (\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^t$ and $n = N - p$. Then, in the case of $n \geq p$, $\boldsymbol{V}$ has a Wishart distribution with mean $n\boldsymbol{\Sigma}$ and degrees of freedom n, denoted as $\mathcal{W}_p(\boldsymbol{\Sigma}, n)$. When $n < p$, it is called a singular Wishart distribution, whose distribution has been recently studied by Srivastava (2003). In many inference procedures, an estimate of the precision matrix $\boldsymbol{\Sigma}^{-1}$ is required. In the case that $n > p$, the standard estimator of $\boldsymbol{\Sigma}^{-1}$ is $\widehat{\boldsymbol{\Sigma}}_0^{-1} = c\boldsymbol{V}^{-1}$ for a positive constant c, but it may not be stable when p is large and close to n. In the case of $p > n$, the estimator $c\boldsymbol{V}^{-1}$ cannot be defined. Srivastava (2005) used the estimator $c\boldsymbol{V}^+$ based on the Moore-Penrose inverse $\boldsymbol{V}^+$ of $\boldsymbol{V}$.

In this paper, we address the problems of estimating both covariance matrix $\boldsymbol{\Sigma}$ and precision matrix $\boldsymbol{\Sigma}^{-1}$, and consider general ridge-type estimators, respectively given by

$$\widehat{\boldsymbol{\Sigma}}_\Lambda = c(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}}), \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_\Lambda^{-1} = c(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})^{-1}, \tag{1.1}$$

where c and d are positive constant based on (n, p), and $\widehat{\boldsymbol{\Lambda}}$ is a $p \times p$ positive definite statistic based on $\boldsymbol{V}$. Examples of the ridge function $\boldsymbol{\Lambda}$ include $\widehat{\boldsymbol{\Lambda}} = \hat{\lambda}\boldsymbol{I}$, $\widehat{\boldsymbol{\Lambda}} = \mathrm{diag}\,(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$ and others, where $\hat{\lambda}$, $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ are functions of $\boldsymbol{V}$. We evaluate the difference of risk functions of the ridge-type and the standard estimators asymptotically for large n and p, where the risk functions are measured with respect to the quadratic loss and the Stein loss functions. Then we derive conditions on d and $\widehat{\boldsymbol{\Lambda}}$ such that the ridge-type estimators improve on the standard estimators asymptotically.

The paper is organized as follows. Section 2 treats estimation of the covariance matrix $\boldsymbol{\Sigma}$, and gives asymptotic evaluations for risks of ridge estimators when $(n, p) \to \infty$. The estimation of the precision matrix $\boldsymbol{\Sigma}^{-1}$ is dealt with in Section 3. For estimation of the covariance matrix relative to the quadratic loss, we can handle both cases of $n > p$ and $p > n$ in the unified framework. For the precision matrix, however, the ridge type estimator has different properties between the two cases, and the standard estimator is $c\boldsymbol{V}^+$ in the case of $p > n$, so that we need to treat the two cases separately. The asymptotic evaluations relative to the Stein loss functions are investigated in Section 4 when $n > p$.

Some examples of the ridge function $\widehat{\boldsymbol{\Lambda}}$ are given in Section 5. Risk performances of the ridge-type estimators are investigated by simulation in Section 6. Concluding remarks are given in Section 7. Some technical tools and proofs are given in the appendix.

2 A Unified Result in Estimation of Covariance under Quadratic Loss

Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a $p \times n$ random matrix such that $\boldsymbol{x}_i \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ for $i = 1, \ldots, n$, where $\boldsymbol{\Sigma}$ is an unknown positive definite matrix. Let $\boldsymbol{V} = \boldsymbol{X}\boldsymbol{X}'$. In the case of $n \geq p$, $\boldsymbol{V}$ is distributed as a Wishart distribution $\mathcal{W}_p(n, \boldsymbol{\Sigma})$ with n degrees of freedom. We first consider the estimation of the covariance matrix $\boldsymbol{\Sigma}$ in terms of the risk function $R_1(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) = E[L_1(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}]$, where $L_1(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}})$ is the quadratic loss

$$L_1(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) = \mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2].$$

The loss function is invariant under the scale transformation $\boldsymbol{\Sigma} \to \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$ and $\widehat{\boldsymbol{\Sigma}} \to \boldsymbol{A}\widehat{\boldsymbol{\Sigma}}\boldsymbol{A}'$ for any nonsingular matrix $\boldsymbol{A}$.

A standard estimator is of the form $c\boldsymbol{V}$ for $c \in \boldsymbol{R}_+$, where $\boldsymbol{R}_+$ is a set of real positive numbers, and the optimal c in terms of the risk is given by $c_1 = 1/(n + p + 1)$ and the risk of the estimator $\widehat{\boldsymbol{\Sigma}}_0 = c_1\boldsymbol{V}$ is $p(p + 1)/(n + p + 1)$. This can be easily seen for $n \geq p$ and it follows from Konno (2009) for $p > n$. To improve the estimator $c_1\boldsymbol{V}$, we consider a class of estimators given by

$$\widehat{\boldsymbol{\Sigma}}_\Lambda = c_1(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}}), \tag{2.1}$$

where $\widehat{\boldsymbol{\Lambda}}$ is a $p \times p$ positive definite matrix based on $\boldsymbol{V}$, and d is a positive constant. As several choices of d, we consider cases of $d = 1$, $d = p$, $d = n$ and $d = \max\{\sqrt{n}, \sqrt{p}\} \equiv d_{n,p}$ and investigate risk performances analytically and numerically. Let

$$\Delta_1 = R_1(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\Lambda) - R_1(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0).$$

To investigate Δ_1 asymptotically, we assume the following conditions:

(A1) Assume that $(n, p) \to \infty$. Throughout the paper, δ given in the following is a constant satisfying $0 < \delta \leq 1$. Assume either (A1-1) or (A1-2) for order of (n, p), where

(A1-1) $p = O(n^\delta)$ for $0 < \delta \leq 1$ in the case of $n \geq p$,

(A1-2) $n = O(p^\delta)$ for $0 < \delta \leq 1$ in the case of $p > n$.

(A2) There exist limiting values $\lim_{p\to\infty} \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})^i]/p$ and $\lim_{p\to\infty} \mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^i]/p$ for $i = 1, 2$, where $\boldsymbol{\Lambda}$ is a $p \times p$ symmetric matrix based on $\boldsymbol{\Sigma}$.

(A3) Assume that $\widehat{\boldsymbol{\Lambda}}$ is a $p \times p$ symmetric matrix based on $\boldsymbol{V}$ such that $\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2]/p = O_p(n^{-1})$, $E[\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})]]/p = O_p(n^{-1})$ and $E[\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})]]/p = O_p(n^{-1})$.

Some examples of statistics $\widehat{\boldsymbol{\Lambda}}$ satisfying condition (A3) will be given in Section 5.

In this paper, we use the notations

$$\boldsymbol{W} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}\boldsymbol{\Sigma}^{-1/2}, \boldsymbol{\Gamma} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1/2}, \widehat{\boldsymbol{\Gamma}} = \boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1/2},$$
$$m = n - p, \mathrm{Ch}_{max}(\boldsymbol{A}) = (\text{the largest eigenvalue of } \boldsymbol{A}).$$

Theorem 1 *Assume conditions* (A1)-(A3). *Then, the risk difference of the estimators* $\widehat{\boldsymbol{\Sigma}}_\Lambda = c_1(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})$ *and* $\widehat{\boldsymbol{\Sigma}}_0$ *is approximated as*

$$\begin{aligned}\Delta_1 =& \frac{pd}{(n+p)^2}\Big\{\frac{d}{p}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]\Big\} \\ &+ \frac{p}{n+p}\Big\{O\Big(d\frac{\sqrt{p/n}}{\sqrt{n+p}}\Big) + O\Big(\frac{d^2/n}{n+p}\Big)\Big\}. \end{aligned} \tag{2.2}$$

For $d = 1$, p, n and $\max\{\sqrt{n}, \sqrt{p}\}$, we get the following approximations from Theorem 1.

Corollary 1 *Assume conditions* (A1)-(A3).

(1) *In the case of* $d = 1$,

$$\Delta_1 = -\frac{2p}{(n+p)^2}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + R_{n,p}, \tag{2.3}$$

where $R_{n,p} = O(n^{-2+3\delta/2})$ *for* $p = O(n^\delta)$ *and* $R_{n,p} = O(p^{-\delta/2})$ *for* $n = O(p^\delta)$ *for* $0 < \delta \leq 1$.

(2) *In the case of* $d = p$,

$$\Delta_1 = \frac{p^2}{(n+p)^2}\Big\{\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]\Big\} + R_{n,p}, \tag{2.4}$$

where $R_{n,p} = O(n^{-2+5\delta/2})$ *for* $p = O(n^\delta)$ *and* $R_{n,p} = O(p^{1-\delta/2})$ *for* $n = O(p^\delta)$ *for* $0 < \delta \leq 1$.

(3) *In the case of* $d = n$,

$$\Delta_1 = \frac{pn}{(n+p)^2}\Big\{\frac{n}{p}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]\Big\} + R_{n,p}, \tag{2.5}$$

where $R_{n,p} = O(n^{-1+3\delta/2})$ *for* $p = O(n^\delta)$ *and* $R_{n,p} = O(p^{\delta/2})$ *for* $n = O(p^\delta)$ *for* $0 < \delta \leq 1$.

(4) *In the case of* $d = \max\{\sqrt{n}, \sqrt{p}\} \equiv d_{n,p}$,

$$\Delta_1 = -2\frac{pd_{n,p}}{(n+p)^2}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + R_{n,p}, \tag{2.6}$$

where $R_{n,p} = O(n^{-1+\delta})$ *for* $p = O(n^\delta)$ *and* $1/2 < \delta \leq 1$, *and* $R_{n,p} = O(1)$ *for* $n = O(p^\delta)$ *and* $0 < \delta \leq 1$.

As seen from (2.3) and (2.4), the leading term of the risk difference is always negative in the case of $d = 1$, but the sign of the leading term in the case of $d = p$ depends on the sign of $\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]$. It should be noted here that the order of the leading term in (2.3) is $O(p^2/(n+p)^2)$, while that in (2.4) is $O(p^3/(n+p)^2)$. This means that if the quantity $\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]$ is negative, then for large p the risk gain in the case of $d = p$ is much larger than that in the case of $d = 1$. This observation can be confirmed by simulation studies in Section 6.

In the case that $n > p$ and $p = O(n^\delta)$ for $0 < \delta \lneqq 1$, Corollary 1 (3) implies that

$$\Delta_1 = \frac{n^2}{n+p}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] + O(n^{-1+2\delta}),$$

which shows that $\widehat{\boldsymbol{\Sigma}}_\Lambda$ with $d = n$ and $\delta < 1$ has a larger risk than $\widehat{\boldsymbol{\Sigma}}_0$ asymptotically.

Since the leading term in (2.2) is a quadratic function of d, it can be minimized at

$$d = p\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]/\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2].$$

In the case of $\boldsymbol{\Lambda} = \lambda \boldsymbol{I}$ for a positive parameter λ, the minimizing $d\lambda$ is

$$d\lambda = p\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]/\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]. \tag{2.7}$$

A corresponding estimator of $d\lambda$ is given by $\widehat{\boldsymbol{\Lambda}}_3$ of Example 2 in Section 5 when $n > p$.

Proof of Theorem 1. The risk difference of the estimators $\widehat{\boldsymbol{\Sigma}}_\Lambda = c_1(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})$ and $\widehat{\boldsymbol{\Sigma}}_0$ is written as

$$\begin{aligned}\Delta_1 =& E\big[\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I} + c_1 d\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}\}^2\big] - R_1(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0)\\ =& E\big[2c_1 d\mathrm{tr}\,[(c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}] + c_1^2 d^2\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1})^2]\big].\end{aligned} \tag{2.8}$$

We shall evaluate each term in the r.h.s. of (2.8). The first term in the r.h.s. of the last equality in (2.8) is written as

$$\begin{aligned}&E\big[c_1 d\mathrm{tr}\,[(c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}]\\ &= c_1 d(nc_1 - 1)\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + c_1 dE\big[\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}]\big]\\ &= -\frac{p(p+1)d}{(n+p+1)^2}\frac{\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]}{p} + c_1 dE\big[\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}]\big].\end{aligned} \tag{2.9}$$

It is here noted from the Cauchy-Shwartz' inequality that the inequality

$$(\mathrm{tr}\,[\boldsymbol{A}\boldsymbol{B}])^2 \leq \mathrm{tr}\,[\boldsymbol{A}^2]\mathrm{tr}\,[\boldsymbol{B}^2] \tag{2.10}$$

holds for symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. It is also noted that $\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}^2] = O_p(p^2(n+p)^{-1})$ since $E[\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}^2]] = R_1(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0) = p(p+1)/(n+p+1)$. Then,

$$c_1 d\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}] \leq c_1 d\Big\{\mathrm{tr}\,[\{c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}\}^2]\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2]\Big\}^{1/2},$$

which is of order $O_p((n+p)^{-1}d[p^3n^{-1}(n+p)^{-1}]^{1/2})$. Thus, from (2.9),

$$\begin{aligned}&E\big[c_1 d\mathrm{tr}\,[(c_1\boldsymbol{V}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}]\\ &= -\frac{p(p+1)d}{(n+p+1)^2}\frac{\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]}{p} + O\Big(\frac{dp\sqrt{p/n}}{(n+p)^{3/2}}\Big).\end{aligned} \tag{2.11}$$

Finally, we estimate the term $c_1^2 d^2 E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1})^2]$. Note that

$$\begin{aligned}c_1^2 d^2 E\big[\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1})^2]\big] =& c_1^2 d^2\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] + 2c_1^2 d^2 E\big[\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}]\big]\\ &+ c_1^2 d^2 E\big[\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2]\big].\end{aligned} \tag{2.12}$$

Under condition (A3), it is observed that

$$
\begin{aligned}
c_1^2 d^2 E\big[\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}]\big] =& O\Big(\frac{d^2p/n}{(n+p)^2}\Big),\\
c_1^2 d^2 \mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2] =& O\Big(\frac{d^2p/n}{(n+p)^2}\Big),
\end{aligned}
$$

so that

$$
c_1^2 d^2 E\big[\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1})^2]\big] = c_1^2 d^2 \mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] + O\Big(\frac{d^2p/n}{(n+p)^2}\Big). \tag{2.13}
$$

Combining (2.11) and (2.13), we get

$$
\begin{aligned}
\Delta_1 =& -\frac{2p(p+1)d}{(n+p)^2}\frac{\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]}{p} + \frac{pd^2}{(n+p)^2}\frac{\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2]}{p}\\
&+\frac{p}{n+p}\Big\{O\Big(\frac{d\sqrt{p/n}}{\sqrt{n+p}}\Big)+O\Big(\frac{d^2/n}{n+p}\Big)\Big\},
\end{aligned} \tag{2.14}
$$

which yields the approximation in Theorem 1. ■

3 Estimation of Precision under Quadratic Loss

In this section we consider the estimation of the precision matrix $\boldsymbol{\Sigma}^{-1}$. For estimation of the covariance matrix, we have treated both cases of $n > p$ and $p > n$ in the unified framework. For the precision matrix, however, the ridge type estimator has different properties between the two cases, so that we need to treat the two cases separately.

3.1 Case of $n > p$

We begin by considering the case of $n > p$ in estimation of the precision matrix $\boldsymbol{\Sigma}^{-1}$ in terms of the risk function $R_1^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1}) = E[L_1^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1})]$, where

$$
L_1^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1}) = \mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})^2],
$$

which is invariant under the scale transformation. A standard estimator is of the form $c\boldsymbol{V}^{-1}$ for $c \in \boldsymbol{R}_+$, and the risk is $R_1^*(\boldsymbol{\Sigma}, c\boldsymbol{V}^{-1}) = E[c^2\mathrm{tr}\,[\boldsymbol{W}^{-2}] - 2c\mathrm{tr}\,[\boldsymbol{W}^{-1}] + p]$, which is

$$
R_1^*(\boldsymbol{\Sigma}, c\boldsymbol{V}^{-1}) = c^2\frac{p(m+p-1)}{m(m-1)(m-3)} - 2c\frac{p}{m-1} + p. \tag{3.1}
$$

Thus, the best constant c is $c_2 = m(m-3)/(n-1)$, and the risk is $R_1^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0^{-1}) = p(mp+2m-n+1)/\{(m-1)(n-1)\}$ for $\widehat{\boldsymbol{\Sigma}}_0^{-1} = c_2\boldsymbol{V}^{-1}$.

A drawback of $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ is that it may be close to be instable when p is large and $n-p$ is small. To modify the estimator $\widehat{\boldsymbol{\Sigma}}_0^{-1}$, we consider a class of estimators given by

$$
\widehat{\boldsymbol{\Sigma}}_\Lambda^{-1} = c_2(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}, \tag{3.2}
$$

where $\widehat{\mathbf{\Lambda}}$ is a $p \times p$ positive definite matrix based on $\boldsymbol{V}$ satisfying condition (A3). Then, we investigate whether $\widehat{\mathbf{\Sigma}}_\Lambda^{-1}$ improves $\widehat{\mathbf{\Sigma}}_0^{-1}$ Let $\Delta_1^* = R_1^*(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}}_\Lambda) - R_1^*(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}}_0)$. To give approximation of Δ_1^*, we assume the following condition:

(A4) There exist limiting values $\lim_{p\to\infty} \mathrm{tr}\,[\mathbf{\Sigma}^i]/p$, $i = 1, 2$, and $\lim_{p\to\infty} \mathrm{tr}\,[(\mathbf{\Lambda}\mathbf{\Sigma}^{-1})^j]/p$, $j = 1, 2, 3$.

Theorem 2 *Assume conditions* (A1-1), (A3) *and* (A4). *Also assume that* $m = n - p > 7$ *and* $\delta < 1$. *Then, the risk difference of* $\widehat{\mathbf{\Sigma}}_\Lambda^{-1}$ *and* $\widehat{\mathbf{\Sigma}}_0^{-1}$ *can be approximated as*

$$\begin{aligned}\Delta_1^* =& \frac{pd}{n^2}\Big\{\frac{d}{p}\mathrm{tr}\,[(\mathbf{\Lambda}\mathbf{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\mathbf{\Lambda}\mathbf{\Sigma}^{-1}]\Big\} + O(d^3 n^{-3+\delta}) + O(d^2 n^{-5/2+3\delta/2})\\ &+ O(dn^{-5/2+5\delta/2}) + O(dn^{-2+3\delta/2}).\end{aligned} \tag{3.3}$$

For the cases of $d = 1$, p and $\sqrt{n}$, from Theorem 2, we get the following corollary.

Corollary 2 *Under the same conditions as in Theorem 2, the following evaluations hold:*
(1) *For* $d = 1$ *and* $\delta < 1$, $\Delta_1^* = -2pn^{-2}\mathrm{tr}\,[\mathbf{\Lambda}\mathbf{\Sigma}^{-1}] + O(n^{-2+3\delta/2}) + O(n^{-5/2+5\delta/2})$.
(2) *For* $d = p$ *and* $\delta < 1$,

$$\Delta_1^* = \frac{p^2}{n^2}\Big\{\mathrm{tr}\,[(\mathbf{\Lambda}\mathbf{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\mathbf{\Lambda}\mathbf{\Sigma}^{-1}]\Big\} + O(n^{-5/2+7\delta/2}) + O(n^{-2+5\delta/2}).$$

(3) *For* $d = \sqrt{n}$ *and* $1/2 < \delta < 1$,

$$\Delta_1^* = -2\frac{p\sqrt{n}}{n^2}\mathrm{tr}\,[\mathbf{\Lambda}\mathbf{\Sigma}^{-1}] + O(n^{-1+\delta}) + O(n^{-2+5\delta/2}).$$

Proof of Theorem 2. For evaluating the risk $R_1^*(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}}_\Lambda^{-1})$, it is noted that

$$\begin{aligned}(\boldsymbol{V} + d\widehat{\mathbf{\Lambda}})^{-1} =& (\boldsymbol{V} + d\mathbf{\Lambda})^{-1} + \{(\boldsymbol{V} + d\widehat{\mathbf{\Lambda}})^{-1} - (\boldsymbol{V} + d\mathbf{\Lambda})^{-1}\}\\ =& (\boldsymbol{V} + d\mathbf{\Lambda})^{-1} + d(\boldsymbol{V} + d\widehat{\mathbf{\Lambda}})^{-1}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda})(\boldsymbol{V} + d\mathbf{\Lambda})^{-1},\end{aligned}$$

so that the risk of the estimator $\widehat{\mathbf{\Sigma}}_\Lambda^{-1} = c_2(\boldsymbol{V} + d\widehat{\mathbf{\Lambda}})^{-1}$ is written as

$$\begin{aligned}R_1^*(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}}_\Lambda^{-1}) =& E\big[\mathrm{tr}\,[\{c_2(\boldsymbol{V} + d\widehat{\mathbf{\Lambda}})^{-1}\mathbf{\Sigma} - \boldsymbol{I}\}^2]\big]\\ =& E\big[\mathrm{tr}\,[\{c_2(\boldsymbol{V} + d\mathbf{\Lambda})^{-1}\mathbf{\Sigma} - \boldsymbol{I} + c_2 d(\boldsymbol{V} + d\widehat{\mathbf{\Lambda}})^{-1}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda})(\boldsymbol{V} + d\mathbf{\Lambda})^{-1}\mathbf{\Sigma}\}^2]\big]\\ =& E\big[\mathrm{tr}\,[\{c_2(\boldsymbol{V} + d\mathbf{\Lambda})^{-1}\mathbf{\Sigma} - \boldsymbol{I}\}^2]\big]\\ &+ c_2^2 d^2 E\big[\mathrm{tr}\,[\{(\boldsymbol{V} + d\widehat{\mathbf{\Lambda}})^{-1}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda})(\boldsymbol{V} + d\mathbf{\Lambda})^{-1}\mathbf{\Sigma}\}^2]\big]\\ &+ 2c_2 dE\big[\mathrm{tr}\,[\{c_2(\boldsymbol{V} + d\mathbf{\Lambda})^{-1}\mathbf{\Sigma} - \boldsymbol{I}\}(\boldsymbol{V} + d\widehat{\mathbf{\Lambda}})^{-1}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda})(\boldsymbol{V} + d\mathbf{\Lambda})^{-1}\mathbf{\Sigma}]\big]\\ =& I_1 + I_2 + I_3.\end{aligned} \tag{3.4}$$

For I_2, it is estimated as

$$\begin{aligned}
&c_2^2 d^2 \mathrm{tr}\,[\{(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}\}^2\\
&\le c_2^2 d^2 \mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{V}^{-1}]\\
&= c_2^2 d^2 \mathrm{tr}\,[(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{W}^{-2}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{W}^{-2}]\\
&= \frac{d^2}{(m-1)^2}\mathrm{tr}\,[\{(\beta\boldsymbol{W}^{-2}-\boldsymbol{I}+\boldsymbol{I})(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\}^2].
\end{aligned}$$

for $\beta = m(m-1)(m-3)/(n-1)$. Since $E[\boldsymbol{W}^{-2}] = (1/\beta)\boldsymbol{I}$, it is noted that $E[\beta\boldsymbol{W}^{-2}] = \boldsymbol{I}$. Thus, I_2 is evaluated from above as

$$I_2 \le 2\frac{d^2}{(m-1)^2}E\big[\mathrm{tr}\,[\{(\beta\boldsymbol{W}^{-2}-\boldsymbol{I})(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\}^2]\big] + 2\frac{d^2}{(m-1)^2}E\big[\mathrm{tr}\,[(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})^2]\big]. \tag{3.5}$$

Using Lemmas 6-8 and condition (A3), we can demonstrate that $\mathrm{tr}\,[(\beta\boldsymbol{W}^{-2}-\boldsymbol{I})^2] = O_p(p^2/n)$, so that the first term in (3.5) is evaluated as $d^2n^{-2}E[\mathrm{tr}\,[\{(\beta\boldsymbol{W}^{-2}-\boldsymbol{I})(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\}^2]] \le d^2n^{-2}E[\mathrm{tr}\,[\{(\beta\boldsymbol{W}^{-2}-\boldsymbol{I})\}^2]\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\}^2]]$, which is of order $O(d^2n^{-2-1+2\delta-1+\delta})$, or $O(d^2n^{-4+3\delta})$. Since the third term is of order $O(d^2n^{-3+\delta})$, it is observed that $I_2 = O(d^2n^{-4+3\delta}) + O(d^2n^{-3+\delta})$.

Since $I_1 = O(n^{-1+2\delta})$ as seen below, for I_3, we have

$$\begin{aligned}
&c_2 d\,\mathrm{tr}\,[\{c_2(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}-\boldsymbol{I}\}(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}]\\
&\le \Big[O_p(n^{-1+2\delta})\times\{O_p(d^2n^{-4+3\delta})+O_p(d^2n^{-3+\delta})\}\Big]^{1/2}\\
&= O_p(dn^{-5/2+5\delta/2}) + O_p(dn^{-2+3\delta/2}),
\end{aligned}$$

so that

$$I_2 + I_3 = O(d^2n^{-4+3\delta}) + O(d^2n^{-3+\delta}) + O_p(dn^{-5/2+5\delta/2}) + O_p(dn^{-2+3\delta/2}). \tag{3.6}$$

Finally, we estimate I_1. Since

$$(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1} = \boldsymbol{V}^{-1} - (\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}\boldsymbol{V}^{-1}, \tag{3.7}$$

the term $c_2(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}-\boldsymbol{I}$ is rewritten as

$$\begin{aligned}
c_2(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}-\boldsymbol{I} =& c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma} - c_2(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I}\\
=&(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I}) - (\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I}) - (\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda},
\end{aligned}$$

so that the first term I_1 is expressed as

$$\begin{aligned}
I_1 =& E\Big[\mathrm{tr}\,[(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})^2] + \mathrm{tr}\,[\{(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\}^2]\\
&+\mathrm{tr}\,[\{(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}\}^2] - 2\mathrm{tr}\,[(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})^2(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}]\\
&-2\mathrm{tr}\,[(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}] + 2\mathrm{tr}\,[(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\{(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}\}^2]\Big]\\
=& R_1^*(\boldsymbol{\Sigma},\widehat{\boldsymbol{\Sigma}}_0^{-1}) + I_{11} + I_{12} - 2I_{13} - 2I_{14} + 2I_{15}.
\end{aligned} \tag{3.8}$$

We shall evaluate each term in (3.8). Using Lemmas 6-8, we can evaluate I_{11} as

$$
\begin{aligned}
I_{11} \le& d^2 E\big[\mathrm{tr}\,[\{(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\boldsymbol{V}^{-1}\boldsymbol{\Lambda}\}^2] = d^2 E\big[\mathrm{tr}\,[\{(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})\boldsymbol{W}^{-1}\boldsymbol{\Gamma}\}^2]\\
=& d^2 E\big[\mathrm{tr}\,[c_2^2\boldsymbol{W}^{-2}\boldsymbol{\Gamma}\boldsymbol{W}^{-2}\boldsymbol{\Gamma} - 2c_2\boldsymbol{W}^{-2}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}\boldsymbol{\Gamma} + \boldsymbol{W}^{-1}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]\big] = O(d^2 n^{-3+2\delta}).
\end{aligned}
$$

Similarly, from (3.7), Lemma 10 and condition (A4),

$$
\begin{aligned}
I_{12} =& d^2 E\big[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^2] - 2d\mathrm{tr}\,[(\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(d\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^2]\\
&+ d^2\mathrm{tr}\,[\{(\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}d\boldsymbol{W}^{-1}\boldsymbol{\Gamma}\}^2]\big]\\
=& \frac{d^2}{n^2}\mathrm{tr}\,[\boldsymbol{\Gamma}^2] + O(d^3 n^{-3+\delta}),
\end{aligned}
$$

since $d^2\mathrm{tr}\,[\{(\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}d\boldsymbol{W}^{-1}\boldsymbol{\Gamma}\}^2 \le d^3\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^3]$ and $\mathrm{tr}\,[\boldsymbol{\Gamma}^3]/p = O(1)$. For I_{13}, from (3.7),

$$
\begin{aligned}
I_{13} =& dE\big[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})^2(\boldsymbol{W}^{-1} - (\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}d\boldsymbol{\Gamma}\boldsymbol{W}^{-1})\boldsymbol{\Gamma}]\big]\\
=& dE\big[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})^2\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]\big]\\
&- d^2E\big[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})^2(\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]\big]\\
=& I_{131} - I_{132}.
\end{aligned}
$$

It can be seen that $I_{131} = pdn^{-2}\mathrm{tr}\,[\boldsymbol{\Gamma}] + O(dn^{-3+2\delta})$. Also, it is observed that

$$
\begin{aligned}
I_{132} \le& d^2E[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})^2(\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^2]]\\
=& d^2c_2E[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})\boldsymbol{W}^{-2}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]] - E[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})(\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^2]]\\
\le& d^2c_2E[\mathrm{tr}\,[\{(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})\boldsymbol{W}^{-1}\}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^4]] + O(d^2n^{-3+2\delta}),
\end{aligned}
$$

where the first term in the last equality can be estimated as $O(d^2n^{-5/2+3\delta/2})$. Thus,

$$
I_{13} = \frac{pd}{n^2}\mathrm{tr}\,[\boldsymbol{\Gamma}] + O(dn^{-3+2\delta}) + O(d^2n^{-5/2+3\delta/2}).
$$

The term I_{14} is evaluated as

$$
\begin{aligned}
I_{14} =& dE[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})(\boldsymbol{W}^{-1} - (\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}d\boldsymbol{\Gamma}\boldsymbol{W}^{-1})\boldsymbol{\Gamma}]]\\
=& dE[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]] - d^2E[\mathrm{tr}\,[(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})(\boldsymbol{W}+d\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\boldsymbol{W}^{-1})\boldsymbol{\Gamma}]]\\
=& I_{141} + I_{142}.
\end{aligned}
$$

It can be seen that $I_{141} = 0$ and $I_{142} = O(d^2n^{-5/2+3\delta/2})$. Thus, $I_{14} = O(d^2n^{-5/2+3\delta/2})$.

For I_{15}, it is noted that $\mathrm{tr}\,[(c_2\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\{(\boldsymbol{V}+d\boldsymbol{\Lambda})^{-1}d\boldsymbol{\Lambda}\}^2] \le d^2\big[\mathrm{tr}\,[\{\boldsymbol{\Gamma}(c_2\boldsymbol{W}^{-1} - \boldsymbol{I})\}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-2}\boldsymbol{\Gamma})^2]\big]^{1/2} = O_p(d^2n^{-5/2+3\delta/2})$ since $\mathrm{tr}\,[\{\boldsymbol{\Gamma}(c_2\boldsymbol{W}^{-1}-\boldsymbol{I})\}^2] = O_p(n^{-1+2\delta})$.

Combining these evaluations gives that

$$
\begin{aligned}
I_1 =& R_1^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0^{-1}) + \frac{pd}{n^2}\Big\{\frac{d}{p}\mathrm{tr}\,[\boldsymbol{\Gamma}^2] - 2\mathrm{tr}\,[\boldsymbol{\Gamma}]\Big\}\\
&+ O(d^3n^{-3+\delta}) + O(d^2n^{-5/2+3\delta/2}) + O(dn^{-3+2\delta}). \qquad (3.9)
\end{aligned}
$$

Together with (3.6), we get Theorem 2. ■

3.2 Case of $p > n$

We next consider the case of $p > n$ in the estimation of the precision matrix $\boldsymbol{\Sigma}^{-1}$. In this case, $\boldsymbol{V}$ is singular, and there does not exist the inverse of $\boldsymbol{V}$. A possible estimator of $\boldsymbol{\Sigma}^{-1}$ is $c\boldsymbol{V}^+$ for $c \in \boldsymbol{R}_+$, where $\boldsymbol{V}^+$ is the Moore-Penrose generalized inverse of $\boldsymbol{V}$. To improve on the estimator $c\boldsymbol{V}^+$, we consider estimators of the form

$$\widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1} = c(\boldsymbol{V} + p\widehat{\boldsymbol{\Lambda}})^{-1}. \tag{3.10}$$

A loss function treated here is the quadratic loss $L_1^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1}) = \mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})^2]$. Relative to this loss function, an approximation of the risk is provided under the following condition:

(A5) There exist the limiting values $\lim_{p\to\infty} \mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^i]/p$ for $i = 1, 2$.

(A6) Assume that $\widehat{\boldsymbol{\Lambda}}$ satisfies that $\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}}^{-1} - \boldsymbol{\Lambda}^{-1})\boldsymbol{\Sigma}\}^2]/p = O_p(n^{-1})$, $E[\mathrm{tr}\,[\boldsymbol{\Sigma}(\widehat{\boldsymbol{\Lambda}}^{-1} - \boldsymbol{\Lambda}^{-1})]]/p = O_p(n^{-1})$ and $E[\mathrm{tr}\,[\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\Lambda}}^{-1} - \boldsymbol{\Lambda}^{-1})]]/p = O_p(n^{-1})$.

Theorem 3 *Assume conditions* (A1-2), (A5) *and* (A6) *with* $c = c_{n,p} = O(p)$. *Also assume that* $\widehat{\boldsymbol{\Lambda}}$ *satisfies the following condition:*

$$\mathrm{tr}\,[\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}] = O_p(np) \quad \text{and} \quad \mathrm{tr}\,[\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}] = O_p(np). \tag{3.11}$$

Then, the risk of the estimator $\widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1}$ *given in* (3.10) *is approximated as*

$$R_1^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1}) = p\Big\{1 - 2\frac{c}{p}\frac{\mathrm{tr}\,[\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}]}{p} + \frac{c^2}{p^2}\frac{\mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^2]}{p} + O(n^{-1})\Big\} + O(n). \tag{3.12}$$

Proof. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a $p \times n$ random matrix such that $\boldsymbol{V} = \boldsymbol{X}\boldsymbol{X}'$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are i.i.d. as $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. Note that

$$(\boldsymbol{V} + p\widehat{\boldsymbol{\Lambda}})^{-1} = p^{-1}\widehat{\boldsymbol{\Lambda}}^{-1} - p^{-2}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}(\boldsymbol{I}_n + p^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}.$$

The quadratic loss of $\widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1}$ is written as

$$\begin{aligned}
&\mathrm{tr}\,[\{c(\boldsymbol{V} + p\widehat{\boldsymbol{\Lambda}})^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}\}^2] \\
&= \mathrm{tr}\,\Big[\big\{\frac{c}{p}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I} - \frac{c}{p^2}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}(\boldsymbol{I}_n + p^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\big\}^2\Big] \\
&= \mathrm{tr}\,\Big[\big\{\frac{c}{p}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}\big\}^2\Big] \\
&\quad - 2\frac{c}{p^2}\mathrm{tr}\,\Big[(\boldsymbol{I}_n + p^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\{\frac{c}{p}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}\}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\Big] \\
&\quad + \frac{c^2}{p^4}\mathrm{tr}\,\Big[\big\{(\boldsymbol{I}_n + p^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\big\}^2\Big],
\end{aligned}$$

where the second term in the r.h.s. of the last equality is of order $O_p(n)$ from condition (3.11). For the third term, it is observed that

$$
\begin{aligned}
&p^{-2}\mathrm{tr}\left[\left\{(\boldsymbol{I}_n+p^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\right\}^2\right]\\
&\leq p^{-1}\mathrm{tr}\left[(\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\left\{\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\right\}^2\right]\\
&=p^{-1}\mathrm{tr}\left[\boldsymbol{P}\widehat{\boldsymbol{\Lambda}}^{-1/2}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1/2}\right]\\
&\leq p^{-1}\mathrm{tr}\left[\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X}\right],
\end{aligned}
$$

which is of order $O_p(n)$ from condition (3.11), where $\boldsymbol{P}=\widehat{\boldsymbol{\Lambda}}^{-1/2}\boldsymbol{X}(\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Lambda}}^{-1/2}$, being idempotent. Thus,

$$
\mathrm{tr}\,[\{c(\boldsymbol{V}+p\widehat{\boldsymbol{\Lambda}})^{-1}\boldsymbol{\Sigma}-\boldsymbol{I}\}^2]=\mathrm{tr}\,[\{(c/p)\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I}\}^2]+O_p(n). \tag{3.13}
$$

We next evaluate the first term in the r.h.s. of (3.13) as

$$
\begin{aligned}
&\frac{c^2}{p^2}\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma})^2]-2\frac{c}{p}\mathrm{tr}\,[\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Sigma}]\\
&=\frac{c^2}{p^2}\mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^2]-2\frac{c}{p}\mathrm{tr}\,[\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}]+2\frac{c^2}{p^2}\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}^{-1}-\boldsymbol{\Lambda}^{-1})\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}]\\
&\quad+\frac{c^2}{p^2}\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}}^{-1}-\boldsymbol{\Lambda}^{-1})\boldsymbol{\Sigma}\}^2]-2\frac{c}{p}\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}^{-1}-\boldsymbol{\Lambda}^{-1})\boldsymbol{\Sigma}]\\
&=\frac{c^2}{p^2}\mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^2]-2\frac{c}{p}\mathrm{tr}\,[\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}]+O(pn^{-1}),
\end{aligned}
$$

from condition (A6). This shows (3.12). ■

Lemma 1 *If* $\widehat{\boldsymbol{\Lambda}}$ *satisfies* $\mathrm{Ch}_{max}(\widehat{\boldsymbol{\Lambda}}^{-1})=O_p(1)$ *for large* (n,p) *satisfying* (A1-2)*, then condition* (3.11) *is satisfied under condition* $\mathrm{tr}\,[\boldsymbol{\Sigma}^i]/p=O(1)$ *for* $i=1,2,3$.

In fact, since $\mathrm{Ch}_{max}(\widehat{\boldsymbol{\Lambda}}^{-1})=O_p(1)$, it is sufficient to show that $E[\mathrm{tr}\,[\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\Sigma}^i]]=O(np)$, $i=1,2$, which can be easily verified if $\mathrm{tr}\,[\boldsymbol{\Sigma}^i]/p=O(1)$, $i=1,2,3$.

Finally, we compare the risk functions of the two estimators $\widehat{\boldsymbol{\Sigma}}_{\Lambda}^{-1}$ and $\widehat{\boldsymbol{\Sigma}}_0^{-1}=c\boldsymbol{V}^+$ for the Moore-Penrose generalized inverse $\boldsymbol{V}^+$. The risk function of $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ is

$$
\begin{aligned}
R_1^*(\boldsymbol{\Sigma},\widehat{\boldsymbol{\Sigma}}_0^{-1})=&p-2\frac{c}{p}E[\mathrm{tr}\,[p\boldsymbol{V}^+\boldsymbol{\Sigma}]]+\frac{c^2}{p^2}E[\mathrm{tr}\,[(p\boldsymbol{V}^+\boldsymbol{\Sigma})^2]]\\
=&p\left\{-2\frac{c}{p}E[\mathrm{tr}\,[\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1]]+\frac{c^2}{p^2}E[p\mathrm{tr}\,[(\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1)^2]]\right\},
\end{aligned}
$$

for $\boldsymbol{H}_1$ and $\boldsymbol{L}$ defined in (A.3). It follows from Lemma 5 that

$$
\begin{aligned}
E[\mathrm{tr}\,[\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1]]\leq&\mathrm{Ch}_{max}(\boldsymbol{\Sigma})\mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})\frac{n}{p-n-1},\\
E[p\mathrm{tr}\,[(\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}\boldsymbol{H}_1)^2]]\leq&\frac{\mathrm{Ch}_{max}(\boldsymbol{\Sigma})\{\mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})\}^2np(p-1)}{\{(p-n-1)(p-n-3)-2\}(p-n-1)},
\end{aligned}
$$

both of which are of order $O(p^{\delta-1})$ when $n = O(p^\delta)$ for $0 < \delta < 1$. Hence, we get the following proposition.

Proposition 1 *Assume that* $\mathrm{Ch}_{max}(\boldsymbol{\Sigma})$ *and* $\mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})$ *are bounded for large* p*, and that* $p - n \geq 4$*,* $(n, p) \to \infty$ *and* $n = O(p^\delta)$ *for* $0 < \delta < 1$*. Then,* $R_1^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0^{-1}) = p + O(p^\delta)$ *when* $c/p = O(1)$*.*

Combining Theorem 3 and Proposition 1 gives the following asymptotic approximation for $\Delta_1^* = R_1^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\Lambda^{-1}) - R_1^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0^{-1})$.

Corollary 3 *Assume the conditions given in both Theorem 3 and Proposition 1. Then,*

$$\frac{\Delta_1^*}{p} = -2\frac{c}{p}\frac{\mathrm{tr}\,[\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}]}{p} + \frac{c^2}{p^2}\frac{\mathrm{tr}\,[(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma})^2]}{p} + O(p^{-\delta}) + O(p^{\delta-1}). \tag{3.14}$$

4 Extensions to the Stein Loss

We have investigated the risk improvement of the ridge-type estimators relative to the quadratic loss function so far. We next address the query whether the observed properties given in the previous sections hold for another loss function. As a loss function in estimation of the covariance matrix, we treat the Stein loss

$$L_2(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) = \mathrm{tr}\,[\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}] - \log|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p,$$

which has been used frequently in the litrerature. It is noted that $L_2(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}})$ is available in the case that $\widehat{\boldsymbol{\Sigma}}$ is positive definite. Thus, we can handle only the case of $n > p$. For the Stein loss, the best constant c among estimators $c\boldsymbol{V}$ is $c_3 = 1/n$, and let $\widehat{\boldsymbol{\Sigma}}_0 = c_3\boldsymbol{V}$. Let $\Delta_2 = R_2(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\Lambda) - R_2(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0)$. To derive asymptotic approximation of Δ_2, we assume the following condition:

(A7) Either $\mathrm{Ch}_{max}(\widehat{\boldsymbol{\Lambda}}) = O_p(1)$ or $\mathrm{Ch}_{max}(n\boldsymbol{W}^{-1}) = O_p(1)$ holds for large n and p, $n > p$. Also assume that $\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^3]/p = O(1)$.

It is noted that condition (A7) is satisfied when $p/n \to \gamma$ for $0 < \gamma < 1$, for Bai and Yin (1993) showed that the smallest and largest eigenvalues of $\boldsymbol{W}/n$ are almost surely bounded by a constant.

Theorem 4 *Assume conditions* (A1-1)*,* (A2)*,* (A3) *and* (A7) *with* $m = n - p > 7$*. Then, the risk difference of the estimator* $\widehat{\boldsymbol{\Sigma}}_\Lambda = c_3(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})$ *relative to the Stein loss* $L_2(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}})$ *is evaluated as*

$$\begin{aligned}\Delta_2 =& \frac{pd}{2n^2}\Big\{\frac{d}{p}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]\Big\}\\ &+ O(d^3n^{-3+\delta}) + O(d^2n^{-5/2+3\delta/2}) + O_p(dn^{-2+3\delta/2}).\end{aligned} \tag{4.1}$$

The proof is given in the appendix. Theorem 4 gives the following corollary.

Corollary 4 *Under the same conditions as in Theorem 4, the following evaluations hold for $d = 1$ and p:*

(1) *For $d = 1$, $\Delta_2 = -pn^{-2}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + O_p(n^{-2+3\delta/2})$.*

(2) *For $d = p$ and $\delta < 1$,*

$$\Delta_2 = \frac{p^2}{2n^2}\Big\{\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]\Big\} + O(n^{-5/2+7\delta/2}) + O(n^{-2+5\delta/2}) + O(n^{-3+4\delta}).$$

We next treat the estimation of the precision matrix in terms of the risk $R_2^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1}) = E[L_2^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1})]$, where the Stein loss for estimating $\boldsymbol{\Sigma}^{-1}$ is $L_2^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{-1}) = \mathrm{tr}\,[\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}] - \log|\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}| - p$. Relative to the Stein loss, the best estimator among $c\boldsymbol{V}^{-1}$ is the unbiased estimator $\widehat{\boldsymbol{\Sigma}}_0^{-1} = c_4\boldsymbol{V}^{-1}$, where $c_4 = m-1$. Let $\Delta_2^* = R_2^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\Lambda) - R_2^*(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0)$. For the loss L_2^*, we get a similar dominance result in estimation of the precision matrix, which will be shown in the appendix.

Theorem 5 *Under conditions (A1-1), (A2), (A3) and (A7) with $m = n - p > 7$, the risk function of $\widehat{\boldsymbol{\Sigma}}_\Lambda^{-1}$ with $c_4 = m-1$ is approximated as*

$$\begin{aligned}\Delta_2^* =& \frac{pd}{2n^2}\Big\{\frac{d}{p}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]\Big\}\\ &+ O(d^3n^{-3+\delta}) + O(d^2n^{-5/2+3\delta/2}) + O(dn^{-2+3\delta/2}),\end{aligned} \tag{4.2}$$

Corollary 5 *Under the same conditions as in Theorem 5, the following evaluations hold for $d = 1$ and p:*

(1) *For $d = 1$, $\Delta_2^* = -pn^{-2}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + O_p(n^{-2+3\delta/2})$.*

(2) *For $d = p$ and $\delta < 1$,*

$$\Delta_2^* = \frac{p^2}{2n^2}\Big\{\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] - 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]\Big\} + O(n^{-5/2+7\delta/2}) + O(n^{-2+5\delta/2}) + O(n^{-3+4\delta}).$$

5 Examples of Statistic $\widehat{\boldsymbol{\Lambda}}$ for Estimating $\boldsymbol{\Lambda}$

As shown so far, the asymptotic approximations of the risk differences are based on values of d and $\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}$, and in the case of $d = p$, the improvement of the ridge type estimator over the standard one depends on a choice of $\boldsymbol{\Lambda}$. In this section, we provide some examples for statistics $\widehat{\boldsymbol{\Lambda}}$ satisfying condition (A3) or (A6).

Example 1 Consider the statistic given by

$$\widehat{\boldsymbol{\Lambda}}_1 = \hat{a}_1\boldsymbol{I} \quad \text{for} \quad \hat{a}_1 = \mathrm{tr}\,[\boldsymbol{V}]/(np). \tag{5.1}$$

This is an unbiased estimator of $\boldsymbol{\Lambda}_1 = a_1\boldsymbol{I}$ for $a_1 = \mathrm{tr}\,[\boldsymbol{\Sigma}]/p$. As given in Lemma 12, Srivastava (2005) showed that $\hat{a}_1 - a_1 = O_p((np)^{-1/2})$ under condition (A2) for large n or p. This shows that $\widehat{\boldsymbol{\Lambda}}_1 = \hat{a}_1\boldsymbol{I}$ satisfies conditions (A3) and (A7). Thus, in the case that both n and p are large under condition (A1), the results given in Theorems 1-5 follow. For example, from Corollary 1, it is seen that for (2.3) with $d = 1$,

$$\Delta_1 = -\frac{2}{(n+p)^2}\mathrm{tr}\,[\boldsymbol{\Sigma}]\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}] + R_{n,p},$$

and for (2.4) with $d = p$,

$$\Delta_1 = \frac{1}{(n+p)^2}\mathrm{tr}\,[\boldsymbol{\Sigma}]\big\{\mathrm{tr}\,[\boldsymbol{\Sigma}]\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}] - 2p\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]\big\} + R_{n,p}.$$

The second-order term in the former case is negative, while the second-order term in the latter case is not necessarily negative.

We next investigate the conditions in Theorem 3 when $p > n$. Note that

$$\hat{a}_1^{-1} - a_1^{-1} = -a_1^{-1}(\hat{a}_1 - a_1) + a_1^{-2}(\hat{a}_1 - a_1)^2 + o_p((np)^{-1}).$$

Since $E[\hat{a}_1 - a_1] = 0$, $E[(\hat{a}_1 - a_1)^2] = O((np)^{-1})$ and $\hat{a}_1 = O_p(1)$, it is easily verified that condition (A6) is satisfied, and the results given in Theorem 3 hold. Especially, from (3.14),

$$\lim_{p\to\infty} \Delta_1^*/p = \lim_{p\to\infty} \frac{1}{(\mathrm{tr}\,[\boldsymbol{\Sigma}])^2}\big\{p\mathrm{tr}\,[\boldsymbol{\Sigma}^2] - 2(\mathrm{tr}\,[\boldsymbol{\Sigma}])^2\big\},$$

for $c = p$. The inequality $\lim_{p\to\infty} \Delta_1^*/p \le 0$ is satisfied when $\boldsymbol{\Sigma}$ is close to $(const.)\boldsymbol{I}$, but it does not hold for $\boldsymbol{\Sigma}$ away from $(const.)\boldsymbol{I}$. ■

Example 2 Consider the statistic given by

$$\widehat{\boldsymbol{\Lambda}}_2 = (\hat{a}_2/\hat{a}_1)\boldsymbol{I} \quad \text{for} \quad \hat{a}_2 = \frac{\mathrm{tr}\,[\boldsymbol{V}^2]}{pn^2} - \frac{(\mathrm{tr}\,[\boldsymbol{V}])^2}{pn^3}. \tag{5.2}$$

As given in Lemma 12, Srivastava (2005) showed that $\hat{a}_2$ is an unbiased estimator of $a_2 = \mathrm{tr}\,[\boldsymbol{\Sigma}^2]/p$ and $\hat{a}_2 - a_2 = O_p((np)^{-1/2})$ under condition (A2) for large n or p. Note that

$$\begin{aligned}\frac{\hat{a}_2}{\hat{a}_1} - \frac{a_2}{a_1} =& \frac{a_2}{a_1}\Big\{\frac{\hat{a}_2 - a_2}{a_2} - \frac{\hat{a}_1 - a_1}{a_1}\Big\} \\ &- \frac{a_2}{a_1}\Big\{\frac{(\hat{a}_1 - a_1)^2}{a_1^2} + \frac{(\hat{a}_1 - a_1)(\hat{a}_2 - a_2)}{a_1 a_2}\Big\} + O_p(n^{-3/2}),\end{aligned} \tag{5.3}$$

which implies that $E[\hat{a}_2/\hat{a}_1 - a_2/a_1] = O(n^{-1})$ and $E[(\hat{a}_2/\hat{a}_1 - a_2/a_1)^2] = O(n^{-1})$. This shows that $\widehat{\boldsymbol{\Lambda}}_2$ satisfies conditions (A3) and (A7). Thus, in the case that both n and p are large under condition (A1), we have the results given in Theorems 1-5. For example, from Corollary 1, it is seen that for (2.3) with $d = 1$,

$$\Delta_1 = -\frac{2p}{(n+p)^2}\frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^2]\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]}{\mathrm{tr}\,[\boldsymbol{\Sigma}]} + R_{n,p},$$

and for (2.4) with $d = p$,

$$\Delta_1 = \frac{p^2}{(n+p)^2}\frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^2]}{(\mathrm{tr}\,[\boldsymbol{\Sigma}])^2}\big\{\mathrm{tr}\,[\boldsymbol{\Sigma}^2]\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}] - 2\mathrm{tr}\,[\boldsymbol{\Sigma}]\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]\big\} + R_{n,p}.$$

The second-order term in the former case is negative, while the second-order term in the latter case is not necessarily negative.

We next investigate the conditions in Theorem 3 when $p > n$. Similar to (5.3), it can be verified that condition (A6) is satisfied, since $\hat{a}_2/\hat{a}_1 = O_p(1)$. The results given in Theorem 3 hold. Especially, from (3.14),

$$\lim_{p\to\infty} \Delta_1^*/p = -\lim_{p\to\infty} \frac{a_1^2}{a_2},$$

for $c = p$. The inequality $\lim_{p\to\infty} \Delta_1^*/p \le 0$ is always satisfied. ■

Example 3 Consider the case of $n > p$ and $p = O(n^\delta)$ for $0 < \delta \le 1$. Then, Lemma 13 proves that the estimators given by

$$\begin{aligned}\hat{b}_1 =& \frac{m}{p}\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}],\\ \hat{b}_2 =& \frac{m^2}{p}\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-2}] - \frac{p}{m}(\hat{b}_1)^2,\end{aligned} \tag{5.4}$$

are consistent estimators of b_1 and b_2, respectively, where $\hat{a}_1$ is given in (5.1) and $b_i = \mathrm{tr}\,[\boldsymbol{\Sigma}^{-i}]/p$ for $i = 1, 2$. That is , $\hat{b}_1 - b_1 = O_p((np)^{-1/2})$ and $\hat{b}_2 - b_2 = O_p(n^{-1/2})$. Based on these statistics, we consider the statistic

$$\widehat{\boldsymbol{\Lambda}}_3 = (\hat{b}_1/\hat{b}_2)\boldsymbol{I}. \tag{5.5}$$

Similarly to (5.3), we can see that $E[\hat{b}_1/\hat{b}_2 - b_1/b_2] = O(n^{-1})$ and $E[(\hat{b}_1/\hat{b}_2 - b_1/b_2)^2] = O(n^{-1})$. This shows that $\widehat{\boldsymbol{\Lambda}}_3$ satisfies conditions (A3) and (A7). Thus, in the case that both n and p are large under condition (A1), we have the results given in Theorems 1-5. For example, from Corollary 1, it is seen that for (2.3) with $d = 1$,

$$\Delta_1 = -\frac{2p}{(n+p)^2}\frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2}{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]} + R_{n,p},$$

and for (2.4) with $d = p$,

$$\Delta_1 = -\frac{p^2}{(n+p)^2}\frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2}{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}])^2} + R_{n,p}.$$

The second-order terms in the two cases are negative.

We next investigate the conditions in Theorem 3 when $p > n$. Similar to (5.3), it can be verified that condition (A6) is satisfied, since $\hat{b}_1/\hat{b}_2 = O_p(1)$. The condition (3.11) is satisfied, and the results given in Theorem 3 hold. Especially, from (3.14),

$$\lim_{p\to\infty} \Delta_1^*/p = \lim_{p\to\infty} \frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]}{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2}\big\{\mathrm{tr}\,[\boldsymbol{\Sigma}^2]\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}] - 2\mathrm{tr}\,[\boldsymbol{\Sigma}]\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]\big\},$$

for $c = p$. The inequality $\lim_{p\to\infty} \Delta_1^*/p \le 0$ is satisfied when $\boldsymbol{\Sigma}$ is close to $(const.)\boldsymbol{I}$, but it does not hold for $\boldsymbol{\Sigma}$ away from $(const.)\boldsymbol{I}$. ■

Example 4 Consider the statistic given by

$$\widehat{\mathbf{\Lambda}}_4 = n^{-1}\text{diag}\,(v_{11},\dots,v_{pp}), \tag{5.6}$$

where v_{ii} is the i-th diagonal element of $\boldsymbol{V}$. Then, $\widehat{\mathbf{\Lambda}}_4$ is an unbiased estimator of $\mathbf{\Lambda} = \text{diag}\,(\sigma_{11},\dots,\sigma_{pp})$.

We shall verify conditions (A3) and (A6). Note that $v_{ii}/\sigma_{ii} \sim \chi^2_n$ and that $E[(v_{ii}/n - \sigma_{ii})^2] = 2\sigma^2_{ii}/n$. For (A3), it is seen that for $\mathbf{\Sigma}^{-1} = (\sigma^{ij})$,

$$\begin{aligned} E[\text{tr}\,[\{(\widehat{\mathbf{\Lambda}}_4 - \mathbf{\Lambda})\mathbf{\Sigma}^{-1}\}^2]] =& \sum_{i,j}\sigma^{ij}\sigma^{ji}E[(v_{ii}/n-\sigma_{ii})(v_{jj}/n-\sigma_{jj})] \\ \le& \sum_{i,j}\sigma^{ij}\sigma^{ji}\{E[(v_{ii}/n-\sigma_{ii})^2]E[(v_{jj}/n-\sigma_{jj})^2]\}^{1/2} \\ =& \frac{2}{n}\sum_{i,j}\sigma^{ij}\sigma^{ji}\sigma_{ii}\sigma_{jj} = \frac{2}{n}\text{tr}\,[(\mathbf{\Lambda}\mathbf{\Sigma}^{-1})^2]. \end{aligned} \tag{5.7}$$

Thus, condition (A3) holds if $\text{tr}\,[(\mathbf{\Lambda}\mathbf{\Sigma}^{-1})^2]/p = O(1)$ for lage p. For (A6), it is seen that

$$\begin{aligned} E[\text{tr}\,[\{(\widehat{\mathbf{\Lambda}}_4^{-1} - \mathbf{\Lambda}^{-1})\mathbf{\Sigma}\}^2]] =& \sum_{i,j}\frac{\sigma^{ij}\sigma^{ji}}{\sigma_{ii}\sigma_{jj}}E[(1-n\sigma_{ii}/v_{ii})(1-n\sigma_{jj}/v_{jj})] \\ \le& \sum_{i,j}\frac{\sigma^{ij}\sigma^{ji}}{\sigma_{ii}\sigma_{jj}}\{E[(1-n\sigma_{ii}/v_{ii})^2]E[(1-n\sigma_{jj}/v_{jj})^2]\}^{1/2} \\ =& \frac{2(n+4)}{(n-2)(n-4)}\sum_{i,j}\frac{\sigma_{ij}\sigma_{ji}}{\sigma_{ii}\sigma_{jj}} = \frac{2(n+4)}{(n-2)(n-4)}\text{tr}\,[(\mathbf{\Lambda}^{-1}\mathbf{\Sigma})^2], \end{aligned}$$

which is of order $O(p/n)$ if $\text{tr}\,[(\mathbf{\Lambda}^{-1}\mathbf{\Sigma})^2]/p = O(1)$ for lage p. Similarly, $E[\text{tr}\,[(\widehat{\mathbf{\Lambda}}_4^{-1} - \mathbf{\Lambda}^{-1})\mathbf{\Sigma}] = 2(n-2)^{-1}p$ and $E[\text{tr}\,[(\widehat{\mathbf{\Lambda}}_4^{-1} - \mathbf{\Lambda}^{-1})\mathbf{\Sigma}\mathbf{\Lambda}^{-1}\mathbf{\Sigma}] = 2(n-2)^{-1}\text{tr}\,[(\mathbf{\Lambda}^{-1}\mathbf{\Sigma})^2]$. Hence, conditions (A6) holds for $\widehat{\mathbf{\Lambda}}_4$. For condition (3.11), it is noted that if $\text{tr}\,[\mathbf{\Sigma}^{-1}]/p = O(1)$, then

$$\text{tr}\,[\boldsymbol{X}'\widehat{\mathbf{\Lambda}}^{-1}\mathbf{\Sigma}\widehat{\mathbf{\Lambda}}^{-1}\boldsymbol{X}] = \text{Ch}_{max}(\mathbf{\Sigma})\text{tr}\,[\widehat{\mathbf{\Lambda}}^{-2}\boldsymbol{X}\boldsymbol{X}'] = \text{Ch}_{max}(\mathbf{\Sigma})\sum_{i=1}^{p} n^2/v_{ii},$$

where v_{ij} denotes the (i,j) element of $\boldsymbol{X}\boldsymbol{X}'$. Here, $\sum_{i=1}^p E[n^2/v_{ii}] = n^2\sum_{i=1}^p E[(\sigma_{ii}\chi^2_n)^{-1}] = n^2(n-2)^{-1}\text{tr}\,[\mathbf{\Sigma}^{-1}] = O(np)$, so that $\text{tr}\,[\boldsymbol{X}'\widehat{\mathbf{\Lambda}}^{-1}\mathbf{\Sigma}\widehat{\mathbf{\Lambda}}^{-1}\boldsymbol{X}] = O_p(np)$ if $\text{tr}\,[\mathbf{\Sigma}^{-1}]/p = O(1)$ and $\text{Ch}_{max}(\mathbf{\Sigma}) = O(1)$. Similarly, it can be seen that $E[\text{tr}\,[\boldsymbol{X}'\widehat{\mathbf{\Lambda}}^{-1}\mathbf{\Sigma}\widehat{\mathbf{\Lambda}}^{-1}\mathbf{\Sigma}\widehat{\mathbf{\Lambda}}^{-1}\boldsymbol{X}]] = O(np)$ if $\text{tr}\,[\mathbf{\Sigma}^{-2}]/p = O(1)$. Thus, condition (3.11) is satisfied for $\widehat{\mathbf{\Lambda}}_4$ if $\text{Ch}_{max}(\mathbf{\Sigma}) = O(1)$.

Concerning the risk difference in estimation of $\mathbf{\Sigma}^{-1}$ for $p > n$, it follows from Corollary 3 that $\lim_{p\to\infty}\Delta_1^*/p = \lim_{p\to\infty}\{-2\text{tr}\,[\mathbf{\Lambda}^{-1}\mathbf{\Sigma}]/p + \text{tr}\,[(\mathbf{\Lambda}^{-1}\mathbf{\Sigma})^2]/p\}$ for $c = p$. This means that in the case of $\mathbf{\Lambda} = \lambda\boldsymbol{I}$ for $\lambda > 0$, the improvement of the ridge type estimator is attained around the spherical point $\mathbf{\Sigma} = \lambda\boldsymbol{I}$, while in the case of $\mathbf{\Lambda} = \text{diag}\,(\sigma_{11},\dots,\sigma_{pp})$, the improvement can be realized near the diagonal point $\mathbf{\Sigma} = \text{diag}\,(\sigma_{11},\dots,\sigma_{pp})$. ■

6 Simulation Studies

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

As a model for simulation experiments, we consider the random variables $\boldsymbol{x}_i$ from $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ for $i = 1, \ldots, n$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_p \end{pmatrix} \begin{pmatrix} \rho^{|1-1|} & \rho^{|1-2|} & \cdots & \rho^{|1-p|} \\ \rho^{|2-1|} & \rho^{|2-2|} & \cdots & \rho^{|2-p|} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{|p-1|} & \rho^{|p-2|} & \cdots & \rho^{|p-p|} \end{pmatrix} \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_p \end{pmatrix},$$

for a constant ρ on the interval $(-1, 1)$ and $\sigma_i = 5 + (-1)^{i-1}(p - i + 1)/p$. Let $\boldsymbol{V} = \boldsymbol{X}\boldsymbol{X}'$ for $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Then for estimation of $\boldsymbol{\Sigma}$, we can calculate the four kinds of ridge estimators $\widehat{\boldsymbol{\Sigma}}_{\Lambda,i} = c(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}}_i)$ for $\widehat{\boldsymbol{\Lambda}}_i$'s given in (5.1), (5.2), (5.5) and (5.6), which are denoted by Rid_1, Rid_2, Rid_3 and Rid_4. As values of d, we treat the three cases: $d = 1$, p and $d_{n,p}$ for $d_{n,p} = \max\{\sqrt{n}, \sqrt{p}\}$. We use these notations for estimation of $\boldsymbol{\Sigma}^{-1}$. It is noted that $\widehat{\boldsymbol{\Lambda}}_3$ or Rid_3 is available only for $n > p$.

The simulation experiments are carried out under the above model for $(n, p) = (200, 80)$, $(120, 80)$, $(60, 80)$ and $(40, 80)$ and $\rho = 0.1$ and 0.5. Based on 10,000 replications, we calculate averages of the following *Relative Risk Gain* of the ridge estimators:

$$\begin{aligned} RRG_i =& 100 \times \{E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2]] - E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_{\Lambda,i}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2]]\}/E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_0\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2]], \\ RRG_i^* =& 100 \times \{E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})^2]] - E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_{\Lambda,i}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})^2]]\}/E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{\Sigma} - \boldsymbol{I})^2]], \end{aligned}$$

where $\widehat{\boldsymbol{\Sigma}}_0^{-1} = p\boldsymbol{V}^+$ in the case of $p > n$.

The simulation results for estimation of $\boldsymbol{\Sigma}$ are reported in Table 1. In the case of $d = 1$, the ridge-type estimators are always better than the standard estimators, but the amounts of improvement are small. In the case of $d = p$, the values of Rid_1, Rid_2 and Rid_4 are very large for $\rho = 0.1$, but negative and very small for $\rho = 0.5$. This phenomenon is understandable from (2.4), since sign of Rid_i, $i = 1, 2, 4$, may be due to sign of $-\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] + 2\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]$, which is not always positive for $\widehat{\boldsymbol{\Lambda}}_1$, $\widehat{\boldsymbol{\Lambda}}_2$ and $\widehat{\boldsymbol{\Lambda}}_4$. However, $\widehat{\boldsymbol{\Lambda}}_3$ is taken so that this quantity is always positive. This may be reason that the values of Rid_3 for $\rho = 0.5$ are positive and large. It is, however, noted that $\widehat{\boldsymbol{\Lambda}}_3$ is not available in the case of $p > n$. All the values of Rid_i with $d = d_{n,p}$ are positive and larger than those for $d = 1$. These observations show that $\widehat{\boldsymbol{\Lambda}}_3$ with $d = p$ is recommended for $n > p$, while $\widehat{\boldsymbol{\Lambda}}_1$, $\widehat{\boldsymbol{\Lambda}}_2$ and $\widehat{\boldsymbol{\Lambda}}_4$ with $d = d_{n,p}$ are recommendable for any cases of n and p.

The simulation results for estimation of $\boldsymbol{\Sigma}^{-1}$ are reported in Table 2, which clarifies that the performances of the ridge-type estimators of $\boldsymbol{\Sigma}^{-1}$ are quite different from those for $\boldsymbol{\Sigma}$ given in Table 1. The values of Rid_i, $i = 1, 2, 3$, for $d = 1$ are always positive for $n > p$, but very small for $p > n$. In the case of $n > p$, $\widehat{\boldsymbol{\Lambda}}_3$ with $d = d_{n,p}$ has a relatively good performance. In the case of $p > n$, $\widehat{\boldsymbol{\Lambda}}_1$, $\widehat{\boldsymbol{\Lambda}}_2$ and $\widehat{\boldsymbol{\Lambda}}_4$ with $d = p$ are recommended.

Table 1: Values of RRG_i in Estimation of $\boldsymbol{\Sigma}$ for $\rho = 0.1, 0.5$, where $d_{n,p} = \max(\sqrt{n}, \sqrt{p})$ and $\widehat{\boldsymbol{\Lambda}}_3$ is not available for $p > n$ (values of the proposed estimators are given with boldface)

			$d = 1$				$d = p$				$d = d_{n,p}$			
ρ	n	p	Rid_1	Rid_2	Rid_3	Rid_4	Rid_1	Rid_2	Rid_3	Rid_4	Rid_1	Rid_2	Rid_3	Rid_4
	200	80	0.8	0.8	0.2	0.7	26.3	25.4	**11.3**	26.6	**9.8**	**10.4**	2.2	**9.1**
0.1	120	80	1.1	1.1	0.2	1.0	36.7	35.4	**12.1**	36.9	**10.8**	**11.6**	1.8	**10.1**
	60	80	1.5	1.6	NA	1.4	52.3	50.1	NA	51.6	**12.8**	**13.8**	NA	**11.9**
	40	80	1.8	1.9	NA	1.6	60.9	57.8	NA	58.9	**14.9**	**16.2**	NA	**13.8**
	200	80	1.2	2.1	0.4	1.1	−17.9	−168	**20.3**	−9.8	**14.0**	**19.4**	5.5	**13.2**
0.5	120	80	1.7	2.9	0.4	1.6	−25.1	−238	**23.9**	−14.6	**16.0**	**23.6**	4.4	**15.1**
	60	80	2.4	4.1	NA	2.3	−35.8	−347	NA	−23.6	**19.3**	**29.4**	NA	**18.1**
	40	80	2.9	4.9	NA	2.7	−42.1	−418	NA	−30.9	**22.5**	**34.5**	NA	**21.1**

Table 2: Values of RRG_i^* in Estimation of $\boldsymbol{\Sigma}^{-1}$ for $\rho = 0.1, 0.5$, where $d_{n,p} = \max(\sqrt{n}, \sqrt{p})$, $\widehat{\boldsymbol{\Lambda}}_3$ is not available for $p > n$ and the notation $*$ denotes a very bad value beyond -100.0 (values of the proposed estimators are given with boldface)

			$d = 1$				$d = p$				$d = d_{n,p}$			
ρ	n	p	Rid_1	Rid_2	Rid_3	Rid_4	Rid_1	Rid_2	Rid_3	Rid_4	Rid_1	Rid_2	Rid_3	Rid_4
	200	80	**1.7**	**1.8**	0.4	**1.6**	−23.8	−26.6	7.8	−21.9	8.3	8.1	**4.3**	8.3
0.1	120	80	**3.8**	**4.0**	0.9	**3.6**	−25.5	−26.4	−1.9	−24.9	−0.7	−1.4	**4.9**	−0.1
	60	80	$*$	$*$	NA	$*$	**99.7**	**99.6**	NA	**99.7**	67.8	72.9	NA	63.0
	40	80	$*$	$*$	NA	$*$	**95.9**	**95.6**	NA	**95.8**	$*$	$*$	NA	$*$
	200	80	**2.6**	**4.0**	1.0	**2.4**	−38.3	−58.5	−4.6	−36.5	5.3	−2.1	**7.4**	5.8
0.1	120	80	**4.7**	**5.3**	1.9	**4.6**	−28.6	−33.4	−11.9	−28.1	−5.3	−11.6	**4.9**	−4.6
	60	80	$*$	$*$	NA	$*$	**99.5**	**99.3**	NA	**99.5**	83.3	94.3	NA	81.1
	40	80	$*$	$*$	NA	$*$	**93.7**	**92.8**	NA	**93.2**	$*$	−41.6	NA	$*$

7 Concluding Remarks

In this paper, we have considered estimation of the covariance and precision matrices by the ridge-type estimators, and have derived asymptotic expansions of their risk functions relative to the quadratic and Stein loss functions when the sample size and the dimension are very large. These expansions clarify the conditions for the ridge-type estimators to have smaller risks than the standard estimators in terms of the second-order terms.

The conditions for the improvement depend on the choice of the ridge function $\widehat{\mathbf{\Lambda}}$ and the order of d, namely, in estimation of the covariance matrix, if the following inequality holds

$$(d/p)\text{tr}\,[(\mathbf{\Lambda\Sigma}^{-1})^2] \leq 2\text{tr}\,[\mathbf{\Lambda\Sigma}^{-1}], \tag{7.1}$$

then the ridge-type estimators improve on the standard estimators asymptotically relative to the quadratic loss functions in both cases of $n > p$ and $p > n$, and relative to the Stein loss in the case of $n > p$. It is interesting to note that in estimation of the precision matrix, under the same condition as in (7.1), the ridge-type estimator improves on the standard estimator asymptotically relative to both quadratic and Stein loss functions in the case of $n > p$. However, the condition for the improvement in estimation of the precision matrix in the case of $p > n$ is slightly different from (7.1). Although condition (7.1) always holds asymptotically when $d = 1$, it depends on $\mathbf{\Lambda\Sigma}^{-1}$ in the case of $d = p$. In the case of $n > p$, we have provided the statistic $\widehat{\mathbf{\Lambda}}_3$ which alwayes satisfies condition (7.1) for $d = p$. Thus, an interesting issue is whether we can construct a statistic $\widehat{\mathbf{\Lambda}}$ which satisfies (7.1) for $d = p$ in the case of $p > n$. We shall address this issue in a future. Various variants of the ridge-type estimators have been investigated through the performances of the risk functions by simulation.

Finally, it is noted that the validity of the asymptotic expansions will not be discussed here. All the results in this paper are based on major terms obtained by Taylor series expansions. Although this paper provides the second order approximations without the validity, we need more conditions and many more steps for establishing the validity of the second-order approximations.

Acknowledgments. The research of the first author was supported in part by Grant-in-Aid for Scientific Research Nos. 21540114 and 23243039 from Japan Society for the Promotion of Science.

References

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

[2] Bai, Z.D., and Yin, Y.Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Statist.*, **21**, 1275-1294.

[3] 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.

[4] Haff, L.R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, **9**, 531-542.

[5] 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.

[6] Kubokawa, T., and Srivastava, M.S. (2008). Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *J. Multivariate Analysis*, **99**, 1906-1928.

[7] Kubokawa, T, Hyodo, M., and Srivastava, M.S. (2011). Asymptotic expansion and estimation of EPMC for linear classification rules in high dimension. Discussion Paper Series, CIRJE-F-818.

[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. (2003). Singular Wishart and multivariate beta distributions. *Ann. Statist.*, **31**, 1537-1560.

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

[12] Stein, C. (1977). Lectures on multivariate estimation theory. (In Russian.) In *Investigation on Statistical Estimation Theory* I, 4-65. Zapiski Nauchych Seminarov LOMI im. V.A. Steklova AN SSSR vol. 74, Leningrad.(In Russian)

A Appendix

A.1 Identities useful for evaluation of moments

The following identity derived by Konno (2009) is useful. It is related to the Stein-Haff identity given by Stein (1977) and Haff (1979) for $n > p$, but it can be used in both cases of $n > p$ and $n \leq p$. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a $p \times n$ random matrix such that $\boldsymbol{V} = \boldsymbol{X}\boldsymbol{X}'$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are i.i.d. as $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$.

Lemma 2 (Konno (2009)) *Let $\boldsymbol{G}(\boldsymbol{V})$ be a $p \times p$ matrix of functions of $\boldsymbol{V}$. Then, the following identity holds:*

$$E\big[\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{G}(\boldsymbol{V})]\big] = E\big[n\mathrm{tr}\,[\boldsymbol{G}(\boldsymbol{V})] + \mathrm{tr}\,[\boldsymbol{X}\boldsymbol{\nabla}'_X\boldsymbol{G}(\boldsymbol{V})']\big], \tag{A.1}$$

where $\boldsymbol{\nabla}_X = (\partial/\partial X_{ij})$ *for* $\boldsymbol{X} = (X_{ij})$.

In the case of $n > p$, we can use the Stein-Haff identity to evaluate higher moments of $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}\boldsymbol{\Sigma}^{-1/2}$. Let $\boldsymbol{G}(\boldsymbol{W})$ be a $p \times p$ matrix such that the (i, j) element $g_{ij}(\boldsymbol{W})$ is a differentiable function of $\boldsymbol{W} = (w_{ij})$ and denote

$$\{\mathcal{D}_W \boldsymbol{G}(\boldsymbol{W})\}_{ac} = \sum_b d_{ab} g_{bc}(\boldsymbol{W}),$$

where $d_{ab} = 2^{-1}(1 + \delta_{ab})\partial/\partial w_{ab}$ with $\delta_{ab} = 1$ for $a = b$ and $\delta_{ab} = 0$ for $a \neq b$.

Lemma 3 (Stein (1977), Haff (1979)) *In the case of* $n > p$, *the Stein-Haff identity is given by*

$$E\left[\mathrm{tr}\,\{\boldsymbol{G}(\boldsymbol{W})\}\right] = E\left[(m-1)\mathrm{tr}\,\left\{\boldsymbol{G}(\boldsymbol{W})\boldsymbol{W}^{-1}\right\} + 2\mathrm{tr}\,\{\mathcal{D}_W \boldsymbol{G}(\boldsymbol{W})\}\right], \tag{A.2}$$

for $m = n - p$.

In the case of $p > n$, the corresponding identity was provided by Kubokawa and Srivastava (2008). This identity was also derived from Lemma 2 by Konno (2009). Let $\boldsymbol{H} = (\boldsymbol{H}_1, \boldsymbol{H}_2)$ be a $p \times p$ orthogonal matrix such that

$$\boldsymbol{V} = \boldsymbol{H}\begin{pmatrix} \boldsymbol{L} & \boldsymbol{0} \\ \boldsymbol{0}' & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}' = \boldsymbol{H}_1\boldsymbol{L}\boldsymbol{H}_1', \quad \boldsymbol{L} = \mathrm{diag}\,(\ell_1, \ldots, \ell_n), \quad \ell_1 \geq \cdots \geq \ell_n, \tag{A.3}$$

where $\boldsymbol{H}_1$ is a $p \times n$ matrix satisfying $\boldsymbol{H}_1'\boldsymbol{H}_1 = \boldsymbol{I}_n$. Let $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n)'$, and $\boldsymbol{\Phi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\phi_1(\boldsymbol{\ell}), \ldots, \phi_n(\boldsymbol{\ell}))$.

Lemma 4 (Kubokawa and Srivastava (2008)) *In the case of* $p > n$, *the Stein-Haff identity is given by*

$$E\left[\mathrm{tr}\,\{\boldsymbol{H}_1\boldsymbol{\Phi}(\boldsymbol{\ell})\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\}\right] = \sum_{i=1}^n E\Big[(p-n-1)\frac{\phi_i}{\ell_i} + 2\frac{\partial}{\partial \ell_i}\phi_i + 2\sum_{j>i}\frac{\phi_i - \phi_j}{\ell_i - \ell_j}\Big]. \tag{A.4}$$

Using Lemma 4, we can evaluate the moments of $\mathrm{tr}\,[\boldsymbol{L}^{-1}]$ and $\mathrm{tr}\,[\boldsymbol{L}^{-2}]$ from above.

Lemma 5 *In the case of* $p > n$, *the following inequalities hold:*

$$\begin{aligned} E[\mathrm{tr}\,[\boldsymbol{L}^{-1}]] \leq& \mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})\frac{n}{p-n-1}, \\ E[\mathrm{tr}\,[\boldsymbol{L}^{-2}]] \leq& \{\mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})\}^2\frac{n(p-1)}{\{(p-n-1)(p-n-3)-2\}(p-n-1)}, \\ E[(\mathrm{tr}\,[\boldsymbol{L}^{-1}])^2] \leq& \{\mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})\}^2\frac{n}{(p-n-1)(p-n-3)-2}. \end{aligned}$$

Proof. Putting $\boldsymbol{\Phi}(\boldsymbol{\ell}) = \boldsymbol{I}$, $\boldsymbol{\Phi}(\boldsymbol{\ell}) = \boldsymbol{L}^{-1}$ and $\boldsymbol{\Phi}(\boldsymbol{\ell}) = (\mathrm{tr}\,[\boldsymbol{L}^{-1}])\boldsymbol{I}$ in the identity (A.4), we get

$$E[\mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]] = (p-n-1)E[\mathrm{tr}\,[\boldsymbol{L}^{-1}]], \tag{A.5}$$

$$E[\mathrm{tr}\,[\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]] = (p-n-3)E[\mathrm{tr}\,[\boldsymbol{L}^{-2}]] - E[(\mathrm{tr}\,[\boldsymbol{L}^{-1}])^2], \tag{A.6}$$

$$E[\mathrm{tr}\,[\boldsymbol{L}^{-1}]\mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]] = (p-n-1)E[(\mathrm{tr}\,[\boldsymbol{L}^{-1}])^2] - 2E[\mathrm{tr}\,[\boldsymbol{L}^{-2}]], \tag{A.7}$$

respectively, where the second equality follows from the fact that $2\sum_{i=1}^{n}\sum_{j=i+1}^{p}(\ell_i\ell_j)^{-1} = (\mathrm{tr}\,[\boldsymbol{L}^{-1}])^2$. The equality (A.5) yields the first inequality in Lemma 5. Combining (A.6) and (A.7) gives the equalities

$$
\begin{aligned}
E[\mathrm{tr}\,[\boldsymbol{L}^{-2}]] =& \frac{(p-n-1)E[\mathrm{tr}\,[\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]] + E[\mathrm{tr}\,[\boldsymbol{L}^{-1}]\mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]]}{(p-n-1)(p-n-3)-2},\\
E[(\mathrm{tr}\,[\boldsymbol{L}^{-1}])^2] =& \frac{2E[\mathrm{tr}\,[\boldsymbol{L}^{-1}\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]] + (p-n-3)E[\mathrm{tr}\,[\boldsymbol{L}^{-1}]\mathrm{tr}\,[\boldsymbol{H}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1]]}{(p-n-1)(p-n-3)-2},
\end{aligned}
$$

which, together with (A.5), provides the second and third inequalities in Lemma 5. ■

A.2 Evaluations of moments

Let $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}\boldsymbol{\Sigma}^{-1/2}$, and $\boldsymbol{W}$ has $\mathcal{W}_p(n, \boldsymbol{I})$ for $n > p$. The following lemmas provide exact moments of the inverted Wishart matrix $\boldsymbol{W}^{-1}$, where$\boldsymbol{A}$ and $\boldsymbol{B}$ in the lemmas are any symmetric matrices. For the proofs, see Kubokawa, Hyodo and Srivastava (2011).

Lemma 6 *Assume that $m = n - p > 3$. Let $\alpha_2 = [m(m-1)(m-3)]^{-1}$. Then,*

$$
\begin{aligned}
E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}] =& \alpha_2[(m-1)\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})], &\text{(A.8)}\\
E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{B})] =& \alpha_2[2\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (m-2)(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})]. &\text{(A.9)}
\end{aligned}
$$

Lemma 7 *Assume that $m = n - p > 5$. Let $\alpha_3 = \alpha_2[m(m-1)(m-3)]^{-1} = [(m+1)m(m-1)(m-3)(m-5)]^{-1}$. Then,*

$$
\begin{aligned}
E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}] =& \alpha_3(n-1)[(m-1)\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + 2(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})], &\text{(A.10)}\\
E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})] =& \alpha_3(n-1)[4\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (m-3)(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})]. &\text{(A.11)}
\end{aligned}
$$

Lemma 8 *Assume that $m = n - p > 7$. Let $\alpha_4 = \alpha_3[(m+2)(m-2)(m-7)]^{-1} = [(m+2)(m+1)m(m-1)(m-2)(m-3)(m-5)(m-7)]^{-1}$. Then,*

$$
\begin{aligned}
&E[\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}]\\
&=\alpha_4(n-1)\Big\{\{(m-1)(n-2)-6\}[(m-1)\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + 2(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})]\\
&\qquad + (2m+3p-2)[4\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (m-3)(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})]\Big\}, &\text{(A.12)}\\
&E[(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})]\\
&=\alpha_4(n-1)\Big\{2(2m+3p-2)[(m-1)\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + 2(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})]\\
&\qquad + \{(m-4)(n-1)-6\}[4\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (m-3)(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})]\Big\}. &\text{(A.13)}
\end{aligned}
$$

Lemma 9 *Let k_1, k_2, ℓ_1 and ℓ_2 be nonnegative integers satisfying $k_1\ell_1 + k_2\ell_2 = m$ for $m \le 4$. Assume that there exist limiting values of $\mathrm{tr}\,[\boldsymbol{A}^{k_1}]/p$ and $\mathrm{tr}\,[\boldsymbol{B}^{k_2}]/p$ for nonnegative definite matrices A and B. Also, assume that $\mathrm{tr}\,[\boldsymbol{W}^{-m}] < \infty$. Then, the moment*

$$M_{n,p} = E[\{\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{A})^{k_1}]/p\}^{\ell_1}\{\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{B})^{k_2}]/p\}^{\ell_2}]$$

is evaluated as $M_{n,p} = O(pn^{-m})$ for large n and p. In the special case of $p/n \to \gamma$ for $0 < \gamma < 1$, $M_{n,p}$ is of order $M_{n,p} = O(n^{-m})$.

Proof. It is noted that

$$\begin{aligned}M_{n,p} \leq & n^{-m}E[\{\mathrm{Ch}_{max}(n\boldsymbol{W}^{-1})\}^m]\{\mathrm{tr}\,[\boldsymbol{A}^{k_1}]/p\}^{\ell_1}\{\mathrm{tr}\,[\boldsymbol{B}^{k_2}]/p\}^{\ell_2}\\ \leq & pn^{-m}E[\mathrm{tr}\,[(n\boldsymbol{W}^{-1})^m]]\{\mathrm{tr}\,[\boldsymbol{A}^{k_1}]/p\}^{\ell_1}\{\mathrm{tr}\,[\boldsymbol{B}^{k_2}]/p\}^{\ell_2}\\ = & O(pn^{-m}),\end{aligned} \tag{A.14}$$

from Lemmas 6-8. In the case that $p/n \to \gamma$ for $0 < \gamma < 1$, we can use the result of Bai and Yin (1993), namely, $\mathrm{Ch}_{max}(n\boldsymbol{W}^{-1}) = O_p(1)$. Thus, from (A.14), it can be seen that $M_{n,p} = O(n^{-m})$. ■

Lemma 10 *Assume that $m = n - p > 5$. Then,*

$$\begin{aligned}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1}])^3] =& \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^3}{m^3} + O(p^3n^{-4}),\\ E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^2]] =& \frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]}{m^3} + \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^3}{m^4} + O(p^2n^{-4}),\\ E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-2}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^2]] =& \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}])^2}{m^3} + \frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}](\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2}{m^4} + O(p^2n^{-4}),\\ E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^3]] =& \frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-3}]}{m^3} + O(p^3n^{-4}).\end{aligned}$$

Proof. Let $\boldsymbol{D}$ be a $p\times p$ diagonal matrix of eigenvalues of $\boldsymbol{\Sigma}^{-1}$. Letting $\boldsymbol{G} = \boldsymbol{D}(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2$, $\boldsymbol{G} = \boldsymbol{D}\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]$, $\boldsymbol{G} = \boldsymbol{D}^2\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]$ and $\boldsymbol{G} = \boldsymbol{D}(\boldsymbol{W}^{-1}\boldsymbol{D})^2$ in Lemma 3, we have

$$\begin{aligned}\mathrm{tr}\,[\boldsymbol{D}]E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2] =& (m-1)E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3] - 4E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]],\\ \mathrm{tr}\,[\boldsymbol{D}]E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] =& (m-1)E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] - 4E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]],\\ \mathrm{tr}\,[\boldsymbol{D}^2]E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] =& (m-1)E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] - 4E[\mathrm{tr}\,[\boldsymbol{D}(\boldsymbol{W}^{-1}\boldsymbol{D})^3]],\\ E[\mathrm{tr}\,[\boldsymbol{D}(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] =& (m-3)E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]] - 2E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]],\end{aligned}$$

respectively. These can be rewritten as

$$\begin{aligned}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3] =& \frac{1}{m-1}\{\mathrm{tr}\,[\boldsymbol{D}]E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2] + 4E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]]\},\\ E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] =& \frac{1}{m-1}\{\mathrm{tr}\,[\boldsymbol{D}]E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] + 4E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]]\},\\ E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] =& \frac{1}{m-1}\{\mathrm{tr}\,[\boldsymbol{D}^2]E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] + 4E[\mathrm{tr}\,[\boldsymbol{D}(\boldsymbol{W}^{-1}\boldsymbol{D})^3]]\},\\ E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]] =& \frac{1}{m-3}\{E[\mathrm{tr}\,[\boldsymbol{D}(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] + 2E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]]\}.\end{aligned}$$

Further, from Lemmas 6 and 9, these third-order terms can be evaluated as

$$\begin{aligned}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3] =& \frac{(\mathrm{tr}\,[\boldsymbol{D}])^3}{m^3} + O(p^3n^{-4}),\\ E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] =& \frac{\mathrm{tr}\,[\boldsymbol{D}]\mathrm{tr}\,[\boldsymbol{D}^2]}{m^3} + \frac{(\mathrm{tr}\,[\boldsymbol{D}])^3}{m^4} + O(p^2n^{-4}),\\ E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] =& \frac{(\mathrm{tr}\,[\boldsymbol{D}^2])^2}{m^3} + \frac{\mathrm{tr}\,[\boldsymbol{D}^2](\mathrm{tr}\,[\boldsymbol{D}])^2}{m^4} + O(p^2n^{-4}),\\ E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]] =& \frac{\mathrm{tr}\,[\boldsymbol{D}^3]}{m^3} + O(p^3n^{-4}),\end{aligned}$$

which yields the evaluations in Lemma 10. ■

Lemma 11 *Assume that $m = n - p > 7$. Then,*

$$\frac{m^2}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1}])^4] = \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^4}{p^2m^2} + O(p^2n^{-3}) + O(pn^{-2}),$$
$$\frac{m^3}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1}])^2\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^2]] = \frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}](\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2}{p^2m} + \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^4}{p^2m^2} + O(pn^{-2}),$$
$$\frac{m^4}{p^2}E[(\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^2])^2] = \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}])^2}{p^2} + 2\frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}](\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2}{p^2m} + \frac{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^4}{p^2m^2} + O(n^{-1}).$$

Proof. It is hard to obtain exact expressions of the requested expectations in Lemma 11. Instead of that, we derive the leading terms and orders of the remainder terms using the same arguments as in the proof of Lemma 10. Letting $\boldsymbol{G} = \boldsymbol{D}(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3$, $\boldsymbol{G} = \boldsymbol{D}\boldsymbol{W}^{-1}\boldsymbol{D}\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]$ and $\boldsymbol{G} = \boldsymbol{D}\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]$ in (A.2) gives, respectively,

$$\begin{aligned}
&E[\mathrm{tr}\,[\boldsymbol{D}](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3] \\
&\quad = (m-1)E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^4] - 6E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2], \\
&E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}^2](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2] \\
&\quad = (m-2)E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2] - E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2]] \\
&\qquad - 4E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^4]], \\
&E[\mathrm{tr}\,[\boldsymbol{D}]\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] \\
&\quad = (m-1)E[(\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2])^2] - 2E[\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2](\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2]] \\
&\qquad - 4E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^3]].
\end{aligned}$$

Then, from Lemma 9, the fourth-order moments can be evaluated as

$$\begin{aligned}
\frac{m^2}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^4] &= \frac{m}{p^2}\mathrm{tr}\,[\boldsymbol{D}]E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^3] + O(p^2n^{-3}), \\
\frac{m^3}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] &= \frac{m^2}{p^2}\mathrm{tr}\,[\boldsymbol{D}]E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] + Op(n^{-2}), \\
\frac{m^4}{p^2}E[(\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2])^2] &= \frac{m^3}{p^2}E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}^2]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] \\
&\quad + \frac{m^2}{p^2}\mathrm{tr}\,[\boldsymbol{D}]E[\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}]\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] + O(n^{-1}).
\end{aligned}$$

Hence, from Lemma 10, we have

$$\frac{m^2}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^4] = \frac{(\mathrm{tr}\,[\boldsymbol{D}])^4}{p^2m^2} + O(p^2n^{-3}) + O(pn^{-2}),$$
$$\frac{m^3}{p^2}E[(\mathrm{tr}\,[\boldsymbol{W}^{-1}\boldsymbol{D}])^2\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2]] = \frac{\mathrm{tr}\,[\boldsymbol{D}^2](\mathrm{tr}\,[\boldsymbol{D}])^2}{p^2m} + \frac{(\mathrm{tr}\,[\boldsymbol{D}])^4}{p^2m^2} + O(pn^{-2}),$$
$$\frac{m^4}{p^2}E[(\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{D})^2])^2] = \frac{(\mathrm{tr}\,[\boldsymbol{D}^2])^2}{p^2} + 2\frac{\mathrm{tr}\,[\boldsymbol{D}^2](\mathrm{tr}\,[\boldsymbol{D}])^2}{p^2m} + \frac{(\mathrm{tr}\,[\boldsymbol{D}])^4}{p^2m^2} + O(n^{-1}),$$

which yields the results in Lemma 11. ■

A.3 Asymptotic properties of $\hat{a}_i$ and $\hat{b}_i$

Lemma 12 (Srivastava (2005)) *$E[\hat{a}_1] = 0$, $E[\hat{a}_2] = a_2$, $Var[\hat{a}_1] = 2a_2/(np)$ and*

$$Var[\hat{a}_2] = \frac{8(n+2)(n+3)(n-1)^2}{pn^5}a_4 + \frac{4(n+2)(n-1)}{n^4}\{a_2^2 - p^{-1}a_4\},$$

where $a_4 = \mathrm{tr}\,[\boldsymbol{\Sigma}^4]/p$. That is, $\hat{a}_1 - a_1 = O_p((np)^{-1/2})$ and $\hat{a}_2 - a_2 = O_p((np)^{-1/2}) + O_p(n^{-1})$ for large n and p.

Lemma 13 *$E[\hat{b}_1 - b_1] = O(n^{-1})$, $E[\hat{b}_2 - b_2] = O(n^{-1})$, $Var[\hat{b}_1] = O((np)^{-1})$ and $Var[\hat{b}_2] = O(n^{-1})$ for large n and p.*

Proof. It follows from (3.7) and Lemma 6 that

$$E[\hat{b}_1] = p^{-1}E[m\mathrm{tr}\,[\boldsymbol{V}^{-1}] - m\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]] = b_1 + O(n^{-1}).$$

For $Var[\hat{b}_1]$, it is written as

$$\begin{aligned}
Var[\hat{b}_1] =& p^{-2}E[\{m\mathrm{tr}\,[\boldsymbol{V}^{-1}] - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}] - m\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]\}^2] + O(n^{-2})\\
=& p^{-2}E[\{m\mathrm{tr}\,[\boldsymbol{V}^{-1}] - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]\}^2]\\
& - 2p^{-2}mE[\{m\mathrm{tr}\,[\boldsymbol{V}^{-1}] - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]\}\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]]\\
& + p^{-2}m^2E[\hat{a}_1^2\{\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]\}^2] + O(n^{-2})\\
=& J_1 - 2J_2 + J_3 + O(n^{-2}).
\end{aligned}$$

It can be easily seen that

$$J_1 = 2\frac{m}{p^2(m-1)(m-3)}\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}] + \frac{3}{p^2(m-1)(m-3)}(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]),$$

which is of order $O((np)^{-1})$. For J_3, it is noted that

$$J_3^* = \frac{m^2}{p^2}\hat{a}_1^2\{\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]\}^2 \le \frac{m^2}{p^2}\hat{a}_1^2\{\mathrm{tr}\,[\boldsymbol{V}^{-2}]\}^2 = \frac{m^2}{p^2}\hat{a}_1^2\{\mathrm{tr}\,[(\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1})^2]\}^2,$$

which is of order $O_p(n^{-2})$ as seen from Lemma 10. Since $J_2 = O((pn^3)^{-1/2})$, it is seen that $Var(\hat{b}_1) = O((np)^{-1})$, which implies that $\hat{b}_1 - b_1 = O_p((np)^{-1/2})$.

For $\hat{b}_2$, it is noted that

$$\begin{aligned}
\hat{b}_2 =& p^{-1}\Big\{m^2\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-2}] - m\{\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}]\}^2\Big\}\\
=& p^{-1}\Big\{m^2\mathrm{tr}\,[\boldsymbol{V}^{-2}] - m^2\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-2}] - m^2\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-2}\boldsymbol{V}^{-1}]\\
& - m(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 + 2m\hat{a}_1\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]\mathrm{tr}\,[\boldsymbol{V}^{-1}]\\
& - m\hat{a}_1^2\{\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-1}]\}^2\Big\}.
\end{aligned}$$

It here follows from Lemma 10 that $\mathrm{tr}\,[(\boldsymbol{V} + \hat{a}_1\boldsymbol{I})^{-1}\boldsymbol{V}^{-2}] \le \mathrm{tr}\,[\boldsymbol{V}^{-3}] = O_p(n^{-1})$. The same arguments can be used to approximate $\hat{b}_2$ as

$$\hat{b}_2 = \frac{m^2}{p}\mathrm{tr}\,[\boldsymbol{V}^{-2}] - \frac{m}{p}(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 + O_p(n^{-1}).$$

Using Lemma 6, we can verify that $E[\hat{b}_2 - b_2] = O(n^{-1})$. Also,

$$\begin{aligned}Var(\hat{b}_2) =& \frac{1}{p^2}E[\{m^2\mathrm{tr}\,[\boldsymbol{V}^{-2}] - m(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]\}^2] \\ &+ 2E[\{m^2\mathrm{tr}\,[\boldsymbol{V}^{-2}] - m(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]\} \times O_p(n^{-1})] + O(n^{-2}).\end{aligned}$$

Here, using Lemma 11, we can show that

$$\begin{aligned}&\frac{1}{p^2}E[\{m^2\mathrm{tr}\,[\boldsymbol{V}^{-2}] - m(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]\}^2] \\ &= \frac{1}{p^2}E\Big[m^4(\mathrm{tr}\,[\boldsymbol{V}^{-2}])^2 + m^2(\mathrm{tr}\,[\boldsymbol{V}^{-1}])^4 + (\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2})^2 \\ &- 2m^3\mathrm{tr}\,[\boldsymbol{V}^{-2}](\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2 - 2m^2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]\mathrm{tr}\,[\boldsymbol{V}^{-2}] + 2m\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}](\mathrm{tr}\,[\boldsymbol{V}^{-1}])^2\Big],\end{aligned}$$

which is of order $O(n^{-1})$. Therefore, the proof of Lemma 13 is complete. ■

A.4 Proofs of Theorems 4 and 5

We here give proofs of Theorems 4 and 5.

Proof of Theorem 4. The risk of the estimator $\widehat{\boldsymbol{\Sigma}}_\Lambda = c_3(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})$ is expressed as

$$\begin{aligned}R_2(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_\Lambda) =& E[c_3\mathrm{tr}\,[(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})\boldsymbol{\Sigma}^{-1}] - \log|c_3(\boldsymbol{V} + d\widehat{\boldsymbol{\Lambda}})\boldsymbol{\Sigma}^{-1}| - p] \\ =& R_2(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}_0) + E[n^{-1}d\mathrm{tr}\,[\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}] - \log|\boldsymbol{I} + d\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}|]. \end{aligned} \tag{A.15}$$

It is here noted that for a symmetric matrix $\boldsymbol{A}$,

$$\log|\boldsymbol{I} + t\boldsymbol{A}| = t\mathrm{tr}\,[\boldsymbol{A}] - \frac{t^2}{2}\mathrm{tr}\,[\boldsymbol{A}(\boldsymbol{I} + t^*\boldsymbol{A})^{-2}\boldsymbol{A}],$$

where t^* is a constant between 0 and t. Thus, for some t^* between 0 and 1,

$$\begin{aligned}\log|\boldsymbol{I} + d\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}| =& d\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}] - \frac{d^2}{2}\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2}(\boldsymbol{I} + dt^*\widehat{\boldsymbol{\Lambda}}^{1/2}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2})^{-2}\widehat{\boldsymbol{\Lambda}}^{1/2}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}] \\ =& d\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}] - \frac{d^2}{2}\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}] \\ &+ \frac{d^2}{2}\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2}\{\boldsymbol{I} - (\boldsymbol{I} + dt^*\widehat{\boldsymbol{\Lambda}}^{1/2}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2})^{-2}\}\widehat{\boldsymbol{\Lambda}}^{1/2}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}] \\ =& I_1 + I_2 + 2^{-1}I_3.\end{aligned}$$

To evaluate I_2, note that

$$\begin{aligned}d^2\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}] =& d^2\mathrm{tr}\,[\boldsymbol{V}^{-1}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\boldsymbol{\Lambda}] + 2d^2\mathrm{tr}\,[\boldsymbol{V}^{-1}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})] \\ &+ d^2\mathrm{tr}\,[\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})].\end{aligned}$$

First, from Lemma 6, it follows that

$$d^2E[\mathrm{tr}\,[\boldsymbol{V}^{-1}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\boldsymbol{\Lambda}]] = d^2\frac{(m-1)\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^2] + (\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}])^2}{m(m-1)(m-3)} = O(d^2n^{-2+\delta}). \tag{A.16}$$

Next, for the thrid term, it is noted that

$$d^2\mathrm{tr}\,[\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})] \le d^2\{\mathrm{Ch}_{max}(\boldsymbol{W}^{-1})\}^2\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2]$$
$$\le d^2\mathrm{tr}\,[\boldsymbol{W}^{-2}]\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2]. \tag{A.17}$$

Since $E[\mathrm{tr}\,[\boldsymbol{W}^{-2}] = O(n^{-2+\delta})$ from Lemma 6, under condition (A3), it is seen that

$$d^2\mathrm{tr}\,[\boldsymbol{W}^{-2}]\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2] = d^2O_p(n^{-2+\delta})O_p(n^{-1+\delta}) = O_p(d^2n^{-3+2\delta}). \tag{A.18}$$

Also, combining (A.16) and (A.18) yields that

$$d^2\mathrm{tr}\,[\boldsymbol{V}^{-1}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})] \le d^2\Big[\mathrm{tr}\,[\{\boldsymbol{V}^{-1}\boldsymbol{\Lambda}\}^2]\mathrm{tr}\,[\{\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\}^2]\Big]^{1/2}$$
$$= d^2\Big[O_p(n^{-2+\delta})O_p(n^{-3+2\delta})\Big]^{1/2} = O_p(d^2n^{-5/2+(3/2)\delta}).$$

Thus,

$$I_2 = -\frac{d^2}{2n^2}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] + O_p(d^2n^{-5/2+(3/2)\delta}). \tag{A.19}$$

To evaluate I_3, it is noted that

$$I_3 = d^2\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2}\{2dt^*\widehat{\boldsymbol{\Lambda}}^{1/2}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2} + (dt^*)^2\{\widehat{\boldsymbol{\Lambda}}^{1/2}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2}\}$$
$$\times(\boldsymbol{I}+dt^*\widehat{\boldsymbol{\Lambda}}^{1/2}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2})^{-2}\widehat{\boldsymbol{\Lambda}}^{1/2}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}]$$
$$\le 2d^3\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2}(\boldsymbol{I}+dt^*\widehat{\boldsymbol{\Lambda}}^{1/2}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2})^{-1}\widehat{\boldsymbol{\Lambda}}^{1/2}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}]$$
$$\le 2d^3\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}].$$

Under condition (A7), it is demonstrated that

$$\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}] = O_p(n^{-3+\delta}). \tag{A.20}$$

In fact, it can be verified that $\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1})^3]/p = O_p(1)$ if $\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^3]/p = O(1)$. If $\mathrm{Ch}_{max}(n\boldsymbol{W}^{-1}) = O_p(1)$ and $\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^3]/p = O(1)$, then

$$\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}] \le n^{-3}\{\mathrm{Ch}_{max}(n\boldsymbol{W}^{-1})\}^3\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1})^3] = O_p(n^{-3+\delta}).$$

If $\mathrm{Ch}_{max}(\widehat{\boldsymbol{\Lambda}}) = O_p(1)$, then $\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}] \le \{\mathrm{Ch}_{max}(\widehat{\boldsymbol{\Lambda}})\}^3\mathrm{tr}\,[\boldsymbol{V}^{-3}] = O_p(n^{-3+\delta})$. Thus, $I_3 = O(d^3n^{-3+\delta})$. Hence,

$$\log|\boldsymbol{I}+d\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}| = d\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}] - \frac{d^2}{2n^2}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] + O_p(d^2n^{-5/2+(3/2)\delta}) + O(d^3n^{-3+\delta}). \tag{A.21}$$

Then from (A.15), we get

$$\Delta_2 = dE[n^{-1}\mathrm{tr}\,[\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}]] - E[\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}]] + \frac{d^2}{2n^2}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2]$$
$$+ O_p(d^2n^{-5/2+(3/2)\delta}) + O(d^3n^{-3+\delta}). \tag{A.22}$$

Finally, we evaluate $dn^{-1}E[\mathrm{tr}\,[(n\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Lambda}}]]$ in (A.22), which is rewritten as

$$dn^{-1}E[\mathrm{tr}\,[(n\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Lambda}}]] = dn^{-1}E[\mathrm{tr}\,[(n\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}]]$$
$$+ dn^{-1}E[\mathrm{tr}\,[(n\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})]].$$

It can be easy to see that

$$\frac{d}{n}E[\mathrm{tr}\,[(n\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}]]=\frac{(p+1)d}{n(n-p-1)}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}].$$

Also, from (2.10), condition (A3) and the same argument as in (3.1), it follows that

$$\begin{aligned}dn^{-1}E[\mathrm{tr}\,[(n\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})]]\le&\Big[\mathrm{tr}\,[(n\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})^2]\mathrm{tr}\,[\{(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}\}^2]\Big]^{1/2}\\ \le& dn^{-1}\Big[O_p(n^{-1+2\delta})O_p(n^{-1+\delta})\Big]^{1/2}=O_p(dn^{-2+3\delta/2}),\end{aligned}$$

so that

$$dn^{-1}E[\mathrm{tr}\,[(n\boldsymbol{V}^{-1}\boldsymbol{\Sigma}-\boldsymbol{I})\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Lambda}}]]=\frac{pd}{n^2}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]+O(dn^{-2+3\delta/2}). \tag{A.23}$$

Combining (A.22) and (A.23) gives the risk evaluation given in (4.1) in Theorem 4. ■

Proof of Theorem 5. The risk of the estimator $c\boldsymbol{V}^{-1}$ is $R_2^*(\boldsymbol{\Sigma},c\boldsymbol{V}^{-1})=E[c\mathrm{tr}\,[\boldsymbol{V}^{-1}\boldsymbol{\Sigma}]-p\log c-\log|\boldsymbol{V}^{-1}\boldsymbol{\Sigma}|-p$, which means that the best constant c is $c=p/E[\mathrm{tr}\,[\boldsymbol{V}^{-1}\boldsymbol{\Sigma}]]=m-1=c_4$. The risk of the estimator $\widehat{\boldsymbol{\Sigma}}_\Lambda^{-1}=c_4(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}$ is expressed as

$$\begin{aligned}R_2^*(\boldsymbol{\Sigma},\widehat{\boldsymbol{\Sigma}}_\Lambda^{-1})=&E[c_4\mathrm{tr}\,[(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}\boldsymbol{\Sigma}]-\log|c_4(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}\boldsymbol{\Sigma}|-p]\\=&R_2^*(\boldsymbol{\Sigma},\widehat{\boldsymbol{\Sigma}}_0)+c_4E[\mathrm{tr}\,[\{(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}-\boldsymbol{V}^{-1}\}\boldsymbol{\Sigma}]+E[\log|\boldsymbol{I}+d\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}|].\end{aligned} \tag{A.24}$$

It can be seen from (A.21) that

$$\log|\boldsymbol{I}+d\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}|=d\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}]-\frac{d^2}{2n^2}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2]+O_p(d^2n^{-5/2+3\delta/2})+O_p(d^3n^{-3+\delta}).$$

Also, from (A.23), it follows that

$$dE[\mathrm{tr}\,[\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Lambda}}]]=\frac{d}{n}E[\mathrm{tr}\,[\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}]]+\frac{pd}{n^2}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]+O(dn^{-2+3\delta/2}).$$

Further, it is noted that

$$\begin{aligned}\frac{d}{n}E[\mathrm{tr}\,[\widehat{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}]]=&\frac{d}{n}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]+\frac{d}{n}E[\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}]]\\=&\frac{d}{n}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]+O(dn^{-2+\delta}),\end{aligned}$$

so that, from (A.24),

$$\begin{aligned}\Delta_2^*=&c_4E\big[\mathrm{tr}\,[\{(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}-\boldsymbol{V}^{-1}\}\boldsymbol{\Sigma}]\big]+\frac{d(n+p)}{n^2}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}]-\frac{d^2}{2n^2}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2]\\&+O(d^2n^{-5/2+3\delta/2})+O(d^3n^{-3+\delta})+O(dn^{-2+\delta}).\end{aligned} \tag{A.25}$$

We next estimate the term $I_0=c_4E[\mathrm{tr}\,[\{(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}-\boldsymbol{V}^{-1}\}\boldsymbol{\Sigma}]]$. By (3.7), I_0 is rewritten as

$$\begin{aligned}I_0=&-c_4E\big[\mathrm{tr}\,[(\boldsymbol{V}+d\widehat{\boldsymbol{\Lambda}})^{-1}d\widehat{\boldsymbol{\Lambda}}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}]\big]=-c_4E\big[\mathrm{tr}\,[(\boldsymbol{W}+d\widehat{\boldsymbol{\Gamma}})^{-1}d\widehat{\boldsymbol{\Gamma}}\boldsymbol{W}^{-1}]\big]\\=&-c_4dE\big[\mathrm{tr}\,[\boldsymbol{W}^{-1}\widehat{\boldsymbol{\Gamma}}\boldsymbol{W}^{-1}]\big]+c_4d^2E\big[\mathrm{tr}\,[(\boldsymbol{W}+d\widehat{\boldsymbol{\Gamma}})^{-1}\widehat{\boldsymbol{\Gamma}}\boldsymbol{W}^{-1}\widehat{\boldsymbol{\Gamma}}\boldsymbol{W}^{-1}]\big]\\=&-c_4dE\big[\mathrm{tr}\,[\boldsymbol{W}^{-1}\widehat{\boldsymbol{\Gamma}}\boldsymbol{W}^{-1}]\big]+c_4d^2E\big[\mathrm{tr}\,[\boldsymbol{W}^{-1}\widehat{\boldsymbol{\Gamma}}\boldsymbol{W}^{-1}\widehat{\boldsymbol{\Gamma}}\boldsymbol{W}^{-1}]\big]\\&-c_4d^3E\big[\mathrm{tr}\,[(\boldsymbol{W}+d\widehat{\boldsymbol{\Gamma}})^{-1}(\widehat{\boldsymbol{\Gamma}}\boldsymbol{W}^{-1})^3]\big].\end{aligned}$$

Here, from (A.20), it follows that under condition (A7),

$$c_4 d^3 E\big[\mathrm{tr}\,[(\boldsymbol{W}+d\widehat{\boldsymbol{\Gamma}})^{-1}(\widehat{\boldsymbol{\Gamma}}\boldsymbol{W}^{-1})^3]\big] \le nd^3 \mathrm{tr}\,[\boldsymbol{W}^{-1}(\widehat{\boldsymbol{\Gamma}}\boldsymbol{W}^{-1})^3] = O_p(d^3 n^{-3+\delta}).$$

Thus, I_0 can be evaluated as

$$\begin{aligned} I_0 =& - c_4 dE\big[\mathrm{tr}\,[\boldsymbol{W}^{-2}\boldsymbol{\Gamma}]\big] - c_4 dE\big[\mathrm{tr}\,[\boldsymbol{W}^{-2}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})]\big] + c_4 d^2 E\big[\mathrm{tr}\,[\boldsymbol{W}^{-2}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]\big] \\ &+ 2c_4 d^2 E\big[\mathrm{tr}\,[\boldsymbol{W}^{-2}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})]\big] + c_4 d^2 E\big[\mathrm{tr}\,[\boldsymbol{W}^{-2}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{W}^{-1}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})]\big] \\ &+ O_p(d^3 n^{-3+\delta}). \end{aligned}$$

Since $E[\boldsymbol{W}^{-2}] = (n-3)/\{m(m-1)(m-3)\}\boldsymbol{I}$, it is observed that

$$\begin{aligned} c_4 dE\big[\mathrm{tr}\,[\boldsymbol{W}^{-2}\boldsymbol{\Gamma}]\big] =& \frac{(n-3)d}{m(m-3)}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}], \\ c_4 d^2 E\big[\mathrm{tr}\,[\boldsymbol{W}^{-2}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}\boldsymbol{\Gamma}]\big] =& \frac{(n-1)(m-1)d^2}{(m+1)m(m-3)(m-5)}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] + O(d^2 n^{-3+2\delta}), \end{aligned}$$

where the second equality follows from Lemma 7. Also, from (A.18),

$$\begin{aligned} c_4 d^2 \mathrm{tr}\,[\boldsymbol{W}^{-2}\boldsymbol{\Gamma}\boldsymbol{W}^{-1}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})] \le& c_4 d^2 \big[\mathrm{tr}\,[(\boldsymbol{W}^{-2}\boldsymbol{\Gamma})^2]\mathrm{tr}\,[\{\boldsymbol{W}^{-1}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\}^2]\big]^{1/2} \\ =& O_p(d^2 n^{1+(-4+\delta-3+2\delta)/2}) = O_p(d^2 n^{-5/2+3\delta/2}). \end{aligned}$$

Similarly to (A.17), it can be verified that

$$c_4 d^2 \mathrm{tr}\,[\boldsymbol{W}^{-2}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{W}^{-1}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})] = O_p(d^2 n^{-3+2\delta}).$$

Finally, we evaluate the term $c_4 dE\big[\mathrm{tr}\,[\boldsymbol{W}^{-2}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})]\big]$. It is noted that

$$\begin{aligned} c_4 dE\big[\mathrm{tr}\,[\boldsymbol{W}^{-2}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})]\big] =& \frac{(n-3)d}{m(m-3)}E\big[\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}]\big] \\ &+ \frac{c_4 d}{m(m-1)}E\big[\mathrm{tr}\,[\big\{m(m-1)\boldsymbol{W}^{-2} - \frac{n-3}{m-3}\boldsymbol{I}\big\}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})]\big]. \end{aligned}$$

Clearly, $(n-3)\{m(m-3)\}^{-1}dE\big[\mathrm{tr}\,[(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})\boldsymbol{\Sigma}^{-1}]\big] = O(dn^{-2+\delta})$. Using Lemma 8, we can see that $E[\mathrm{tr}\,[\{m(m-1)\boldsymbol{W}^{-2} - (n-3)(m-3)^{-1}\boldsymbol{I}\}^2]] = O(p^2/n)$, which implies that

$$\begin{aligned} &\frac{c_4 d}{m(m-1)}\mathrm{tr}\,[\big\{m(m-1)\boldsymbol{W}^{-2} - \frac{n-3}{m-3}\boldsymbol{I}\big\}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})] \\ &\qquad \le O_p(dn^{-1+(-1+2\delta-1+\delta)/2}) = O_p(dn^{-2+3\delta/2}), \end{aligned}$$

where the Cauchy-Shwartz' inequality is used. Thus,

$$\begin{aligned} I_0 =& -\frac{(n-3)d}{m(m-3)}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + \frac{d^2}{n^2}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] + O(d^3 n^{-3+\delta}) \\ &+ O(d^2 n^{-3+2\delta}) + O(d^2 n^{-5/2+3\delta/2}) + O(dn^{-2+3\delta/2}). \end{aligned} \tag{A.26}$$

Combining (A.25) and (A.26) gives

$$\begin{aligned} \Delta_2^* =& \Big\{-\frac{(n-3)d}{m(m-3)} + \frac{d(n+p)}{n^2}\Big\}\mathrm{tr}\,[\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}] + \frac{d^2}{2n^2}\mathrm{tr}\,[(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})^2] \\ &+ O(d^3 n^{-3+\delta}) + O(d^2 n^{-5/2+3\delta/2}) + O(d^2 n^{-3+2\delta}) + O(dn^{-2+3\delta/2}), \end{aligned}$$

which yields (4.2) in Theorem 5. ■