CIRJE-F-958

Linear Shrinkage Estimation of Large Covariance Matrices with Use of Factor Models

Yuri Ikeda
Graduate School of Economics, The University of Tokyo

Tatsuya Kubokawa
The University of Tokyo

February 2015

Linear Shrinkage Estimation of Large Covariance Matrices with Use of Factor Models

Yuki Ikeda*and Tatsuya Kubokawa†
University of Tokyo

Abstract

The problem of estimating large covariance matrices with use of factor models is addressed when both the sample size and the dimension of covariance matrix tend to infinity. In this paper, we consider a general class of weighted estimators which includes (i) linear combinations of the sample covariance matrix and the model-based estimator under the factor model and (ii) ridge-type estimators without factors as special cases. The optimal weights in the class are derived, and the plug-in weighted estimators are suggested since the optimal weights depend on unknown parameters. Numerical results show our methods perform well. Finally, an application to portfolio managements is given.

Key words and phrases: Covariance matrix, factor model, high dimension, large sample, non-normal distribution, normal distribution, portfolio management, ridge-type estimator, risk function.

1 Introduction

Estimation of a large covariance matrices is a fundamental issue in economics, financial engineering, biologics, signal processing and other literatures and has been widely studied. In the estimation of the $p \times p$ covariance matrix $\mathbf{\Sigma}_{11}$, the classical large sample theory assumes that sample size N is allowed to grow, but dimension p is fixed. In this setting, we can estimate the covariance matrix $\mathbf{\Sigma}_{11}$ by its sample covariance matrix, denoted here by $\widehat{\mathbf{\Sigma}}_{11}$, which is a consistent estimator. However, in applications, we often encounter very large data sets which contain variables in high dimension. In this case, using $\widehat{\mathbf{\Sigma}}_{11}$ is inappropriate since $\widehat{\mathbf{\Sigma}}_{11}$ becomes singular when p is larger than N. Even if $p < N$, $\widehat{\mathbf{\Sigma}}_{11}$ is instable as pointed out by Fan, Fan and Lv (2008).

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

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

Various methods have been proposed to estimate $\mathbf{\Sigma}_{11}$ in high dimension. Ledoit and Wolf (2004), Schafer and Strimmer (2005), Chen, Wiesel, Eldar and Hero (2010), Fisher and Sun (2011) and others suggested well-conditioned estimators combining $\widehat{\mathbf{\Sigma}}_{11}$ and more stable statistics, which are called weighted or ridge-type estimators. When covariates, called *factors*, are available, Ledoit and Wolf (2003), Ren and Shimotsu (2009) and Fan, *et al.* (2008) suggested more refined linear shrinkage estimators of $\mathbf{\Sigma}_{11}$ by incorporating the common factor structure in the factor models. Ledoit and Wolf (2003), Ren and Shimotsu (2009) suggested weighted estimators but not considered the high dimensional settings. In this paper, we also propose factor-model-based weighted estimators but in high dimension. Also many papers studied regularization and thresholding techniques such as Bickel and Levina (2008a, b), Rothman, Levina and Zhu (2009), Lam and Fan (2009), Cai and Zhou (2010), Cai and Liu (2011) and others. Recently, Fan, Liao and Mincheva (2011) suggested estimators of $\mathbf{\Sigma}_{11}$ by thresholding the sample covariance matrix of estimated residuals in the factor model and derived the favorable convergence rates under the sparcity of covariance of idiosyncratic components.

Factor models have been widely used to relate variables of interest, $\boldsymbol{y}_i$'s, to some factors, $\boldsymbol{x}_i$'s and has been used in many applications. Among others, Fama and French (1992) found out that excess asset returns are well explained by the three factors of sensitivity to the market excess return, the market capitalization and the book-to-price ratio. Factor models often assumes the independence among the idiosyncratic components, so that the error covariance matrix becomes diagonal. The cross-sectional independence, however, is restrictive in many applications as pointed out in Chamberlain and Rothschild (1983).

Fan, *et al.* (2011) relaxed the assumption of the cross-sectional independence in the factor models and suggest invertible estimators under cross-sectional correlations of idiosyncratic noises, when both p and N are allowed to diverge. In this paper, we also permit the cross-sectional independence in the factor models and suggest new invertible estimators based on all the data $\boldsymbol{y}_i$'s and $\boldsymbol{x}_i$'s for any (p, N).

To explain more specifically the problem addressed here, we describe the underlying model and the factor model. Assume that observations $(\boldsymbol{y}_1, \boldsymbol{x}_1), \ldots, (\boldsymbol{y}_N, \boldsymbol{x}_N)$ are available where $\boldsymbol{y}_i$ and $\boldsymbol{x}_i$ are, respectively, p- and q-dimensional vectors. Consider estimation of $p \times p$ covariance matrix $\mathbf{\Sigma}_{11}$ of $\boldsymbol{y}_i$ where $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ are mutually independently distributed as $E[\boldsymbol{y}_i] = \boldsymbol{\mu}_1$ and $\mathbf{Cov}(\boldsymbol{y}_i) = \mathbf{\Sigma}_{11}$, namely

$$\boldsymbol{y}_i \sim \text{ i.i.d.}(\boldsymbol{\mu}_1, \mathbf{\Sigma}_{11}). \tag{1.1}$$

The sample covariance matrix of $\mathbf{\Sigma}_{11}$ is $\widehat{\mathbf{\Sigma}}_{11} = n^{-1} \sum_{i=1}^{N} (\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^T$ for $n = N-1$ and $\overline{\boldsymbol{y}} = N^{-1} \sum_{i=1}^{N} \boldsymbol{y}_i$. It may be possible to improve on $\widehat{\mathbf{\Sigma}}_{11}$ by using additional observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ when the following factor model is suspected:

$$\begin{aligned} \boldsymbol{y}_i =& \boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{x}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, N, \\ & \boldsymbol{\epsilon}_i \sim \text{ i.i.d.}(\mathbf{0}, \boldsymbol{D}), \quad \boldsymbol{x}_i \sim \text{ i.i.d.}(\boldsymbol{\mu}_2, \mathbf{\Sigma}_{22}), \end{aligned} \tag{1.2}$$

where $\boldsymbol{\epsilon}_i$'s are idiosyncratic error components which are not correlated with $\boldsymbol{x}_i$'s, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are, respectively, p- and q-variate unknown vectors, and $\boldsymbol{D}$ and $\boldsymbol{\Sigma}_{22}$ are, respectively, $p \times p$ and $q \times q$ unknown positive definite symmetric matrices.

It is convenient to treat two models (1.1) and (1.2) in a unified expression. Let $\boldsymbol{\Sigma}_{12} = \mathbf{Cov}\,(\boldsymbol{y}_i, \boldsymbol{x}_i)$. Then from (1.2), it is seen that $\boldsymbol{\mu}_1 = \boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_{12} = E[(\boldsymbol{y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}\boldsymbol{\mu}_2)(\boldsymbol{x}_i - \boldsymbol{\mu}_2)^T] = \boldsymbol{\beta}\boldsymbol{\Sigma}_{22}$. This implies that

$$\begin{aligned}\boldsymbol{\alpha} =& \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2,\\ \boldsymbol{\beta} =& \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1},\end{aligned}$$

and factor model (1.2) can be rewritten as

$$\begin{aligned}\boldsymbol{y}_i =& (\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2) + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{x}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, N,\\ & \boldsymbol{\epsilon}_i \sim \text{ i.i.d.}(\mathbf{0}, \boldsymbol{D}), \quad \boldsymbol{x}_i \sim \text{ i.i.d.}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}),\end{aligned} \tag{1.3}$$

which yields that

$$\mathbf{Cov}\,(\boldsymbol{y}_i) = \boldsymbol{D} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{1.4}$$

Letting $\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, we can rewrite (1.4) as $\mathbf{Cov}\,(\boldsymbol{y}_i) = \boldsymbol{\Sigma}_{11} + \boldsymbol{D} - \boldsymbol{\Sigma}_{11.2}$. As setups of $\boldsymbol{D}$, the following three cases are considered:

(C0) Fully unknown case: $\boldsymbol{D} = \boldsymbol{\Sigma}_{11.2}$.

(C1) Sphericity case: $\boldsymbol{D} = p^{-1}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2})\boldsymbol{I}$.

(C2) Diagonality case: $\boldsymbol{D} = \mathrm{diag}\,(\boldsymbol{\Sigma}_{11.2})$.

In fully unknown case (C0), it follows from (1.4) that $\mathbf{Cov}\,(\boldsymbol{y}_i) = \boldsymbol{\Sigma}_{11.2} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{11}$, which corresponds to model (1.1), and $\boldsymbol{\Sigma}_{11}$ is estimated based on the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}_{11}$ without any information on $\boldsymbol{x}_i$'s. In restricted cases (C1) and (C2), the covariance matrix $\boldsymbol{\Sigma}_{11}$ can be estimated with more refined estimators since $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{22}$ and $\boldsymbol{\Sigma}_{11.2}$ are estimated based on all the data $\boldsymbol{y}_i$'s and $\boldsymbol{x}_i$'s. The restricted cases mean that the idiosyncratic components in $\boldsymbol{\epsilon}_i$ are uncorrelated, and model (1.2) or (1.3) is called the strict factor model, while it is called the approximate factor model in the unrestricted case. Thus, model (1.3) gives a unified expression of strict and approximate factor models.

Using the unified expression of factor models (1.3), we consider estimation of the covariance matrix $\boldsymbol{\Sigma}_{11}$ in model (1.1) when the strict factor models with restrictions (C1) and (C2) are suspected. Under the restrictions, $\boldsymbol{\Sigma}_{11}$ is estimated by

$$\widehat{\boldsymbol{\Sigma}}_f = \widehat{\boldsymbol{D}} + \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21},$$

with appropriate estimators $\widehat{\boldsymbol{D}}$, $\widehat{\boldsymbol{\Sigma}}_{12}$ and $\widehat{\boldsymbol{\Sigma}}_{22}$, so that it is reasonable to estimate $\boldsymbol{\Sigma}_{11}$ with the weighted estimators

$$\widehat{\boldsymbol{\Sigma}}_\alpha = \alpha\widehat{\boldsymbol{\Sigma}}_{11} + (1-\alpha)\widehat{\boldsymbol{\Sigma}}_f = \widehat{\boldsymbol{\Sigma}}_{11} + \alpha(\widehat{\boldsymbol{\Sigma}}_{11} - \widehat{\boldsymbol{\Sigma}}_f)$$

since $\widehat{\boldsymbol{\Sigma}}_\alpha$ shrinks $\widehat{\boldsymbol{\Sigma}}_{11}$ towards $\widehat{\boldsymbol{\Sigma}}_f$ which is the estimator when the strict factor model is suspected. As the true error covariance matrix goes far away from the strict factor model,

however, the weighted estimator $\widehat{\boldsymbol{\Sigma}}_\alpha$ becomes ill-conditioned because it converges to the sample covariance matrix. It is here noted that $\widehat{\boldsymbol{\Sigma}}_\alpha$ is rewritten as $\widehat{\boldsymbol{\Sigma}}_\alpha = \alpha\widehat{\boldsymbol{\Sigma}}_{11.2} + (1-\alpha)\widehat{\boldsymbol{D}} + \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}$. Thus, we consider to shrink further the term $\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}$ toward the restricted estimators, which results in the doubly weighted estimators

$$\widehat{\boldsymbol{\Sigma}}(\gamma,\beta) = \gamma\widehat{\boldsymbol{\Sigma}}_{11.2} + (1-\gamma)\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2}) + \beta\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21} + (1-\beta)\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}),$$

where $\boldsymbol{\Lambda}(\cdot)$ is a function satisfying $\widehat{\boldsymbol{D}} = \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2})$, which corresponds to the restriction. This doubly weighted estimators are well-conditioned when the true error covariance matrix goes far away from the strict factor model.

The goal of this paper is to derive the optimal weights on γ and β in terms of minimizing the mean squared error $Risk(\widehat{\boldsymbol{\Sigma}}(\gamma,\beta)) = p^{-1}E[\mathrm{tr}\,[\{\widehat{\boldsymbol{\Sigma}}(\gamma,\beta) - \boldsymbol{\Sigma}_{11}\}^2]]$, and to suggest the consistent estimators with the optimal weights. As we mentioned above, we permit the cross-sectional correlations among the idiosyncratic components while restricting the density of elements of $\boldsymbol{\Sigma}_{11.2}$ as $\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}^2) = O(p)$. This corresponds to the sparcity condition of $\boldsymbol{\Sigma}_{11.2}$ assumed in Fan, *et al*. al. (2011) where many entries of the off-diagonal elements are zero, and the number of nonzero off-diagonal entries is restricted to grow slowly. On the other hand, for our final interest $\boldsymbol{\Sigma}_{11}$, most weighted (linear shrinkage) estimators in the literature assume that $\mathrm{tr}\,(\boldsymbol{\Sigma}_{11}^2) = O(p)$. However, this can be inappropriate in covariance estimation with use of factor model, since if each element of the factor loadings $\boldsymbol{\Sigma}_{12}$ is of order O(1), $\mathrm{tr}\,(\boldsymbol{\Sigma}_{11}^2) = \mathrm{tr}\,((\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} + \boldsymbol{\Sigma}_{11.2})^2) = O(p^2)$, though it is true that if $\boldsymbol{\Sigma}_{12}$ is sparse or not dense, assuming $\mathrm{tr}\,(\boldsymbol{\Sigma}_{11}^2) = O(p)$ is reasonable. Thus, we treat the two cases, (i) $\mathrm{tr}\,(\boldsymbol{\Sigma}_{11}^2) = O(p^2)$ and (ii) $\mathrm{tr}\,(\boldsymbol{\Sigma}_{11}^2) = O(p)$ depending on the density of the factor loadings $\boldsymbol{\Sigma}_{12}$. We mainly treat the former case but the results in the latter is also shown.

The rest of this paper is organized as follows: In Section 2, we introduce the weighted estimators and derive the optimal weights under appropriate assumptions on non-sparcity and sparcity of factor loadings. The mean squared errors of the estimators with the optimal weights are provided. In Section 3, we give several estimators of unknown parameters included in the optimal estimators and suggest the plug-in estimators by substituing the estimators into the optimal estimators. Section 4 conducts numerical studies, and Section 5 gives an application to portfolio management. Concluding remarks are given in Section 6.

2 Weighted Estimators and Mean Squared Errors

2.1 Weighted estimators

Let us assume that the observations $(\boldsymbol{y}_1,\boldsymbol{x}_1),\ldots,(\boldsymbol{y}_N,\boldsymbol{x}_N)$ have model (1.3). Let $\overline{\boldsymbol{y}} = N^{-1}\sum_{i=1}^N \boldsymbol{y}_i$, $\overline{\boldsymbol{x}} = N^{-1}\sum_{i=1}^N \boldsymbol{x}_i$, $\boldsymbol{V}_{11} = \sum_{i=1}^N(\boldsymbol{y}_i-\overline{\boldsymbol{y}})(\boldsymbol{y}_i-\overline{\boldsymbol{y}})^T$, $\boldsymbol{V}_{12} = \boldsymbol{V}_{21}^T = \sum_{i=1}^N(\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^T$ and $\boldsymbol{V}_{22} = \sum_{i=1}^N(\boldsymbol{x}_i-\overline{\boldsymbol{x}})(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^T$. Let $\widehat{\boldsymbol{\Sigma}}_{ij} = n^{-1}\boldsymbol{V}_{ij}$ $(i,j=1,2)$ for

$n = N - 1$. Also, let

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{pmatrix}. \tag{2.1}$$

We consider the estimation of the covariance matrix $\mathbf{Cov}(\boldsymbol{y}_i) = \boldsymbol{\Sigma}_{11}$. An estimator $\boldsymbol{\delta}$ of $\boldsymbol{\Sigma}_{11}$ is evaluated in terms of the mean squared error (MSE) $Risk(\boldsymbol{\delta}) = p^{-1}E[\mathrm{tr}\,[(\boldsymbol{\delta} - \boldsymbol{\Sigma}_{11})^2]]$. A standard estimator of $\boldsymbol{\Sigma}_{11}$ is the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}_{11} = n^{-1}\boldsymbol{V}_{11}$. This estimator may be used if no information on the model is available. In the financial economics, the factor models are used for explaining stock returns. When factor model (1.3) is suspected, from (1.4), we can suggest the estimator

$$\widehat{\boldsymbol{\Sigma}}_f = \widehat{\boldsymbol{D}} + \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}, \tag{2.2}$$

where $\widehat{\boldsymbol{D}} = \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2})$ for $\widehat{\boldsymbol{\Sigma}}_{11.2} = \widehat{\boldsymbol{\Sigma}}_{11} - \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}$. As examples of $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2})$, we consider the two cases;

(C1) Case of sphericity: $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2}) = p^{-1}\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11.2})\boldsymbol{I}$.

(C2) Case of diagonality: $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2}) = \mathrm{diag}(\widehat{\boldsymbol{\Sigma}}_{11.2})$.

It is noted that this estimator is available in the case of $n \geq q$ due to existence of $\widehat{\boldsymbol{\Sigma}}_{22}^{-1}$. Otherwise, the estimator $\widehat{\boldsymbol{\Sigma}}_{22}^{-1}$ should be replaced with the generalized inverse $\widehat{\boldsymbol{\Sigma}}_{22}^{-}$.

In this paper, we consider to combine the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}_{11}$ and the estimator $\widehat{\boldsymbol{\Sigma}}_f$ in the factor model. For constant α satisfying $0 \leq \alpha \leq 1$, the weighted estimator is

$$\widehat{\boldsymbol{\Sigma}}_\alpha = \alpha\widehat{\boldsymbol{\Sigma}}_{11} + (1-\alpha)\widehat{\boldsymbol{\Sigma}}_f,$$

which is rewritten as

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}_\alpha =& \alpha(\widehat{\boldsymbol{\Sigma}}_{11} - \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}) + \alpha\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21} + (1-\alpha)\widehat{\boldsymbol{D}} + (1-\alpha)\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21} \\ =& \alpha\widehat{\boldsymbol{\Sigma}}_{11.2} + (1-\alpha)\widehat{\boldsymbol{D}} + \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}. \end{aligned} \tag{2.3}$$

Since $\alpha\widehat{\boldsymbol{\Sigma}}_{11.2} + (1-\alpha)\widehat{\boldsymbol{D}} = \widehat{\boldsymbol{\Sigma}}_{11.2} - (1-\alpha)(\widehat{\boldsymbol{\Sigma}}_{11.2} - \widehat{\boldsymbol{D}})$, it shrinks $\widehat{\boldsymbol{\Sigma}}_{11.2}$ toward $\widehat{\boldsymbol{D}}$. The estimator $\widehat{\boldsymbol{\Sigma}}_{11}$ is decomposed as $\widehat{\boldsymbol{\Sigma}}_{11} = \widehat{\boldsymbol{\Sigma}}_{11.2} + \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}$. Thus, the weighted estimator $\widehat{\boldsymbol{\Sigma}}_\alpha$ shrinks the part $\widehat{\boldsymbol{\Sigma}}_{11.2}$ in $\widehat{\boldsymbol{\Sigma}}_{11}$ toward $\widehat{\boldsymbol{D}}$.

It is noted that if $\boldsymbol{D}$ in model (1.3) is less restrictive or if the true error covariance matrix is far away from the strict factor model, then $\widehat{\boldsymbol{\Sigma}}_\alpha$ approches to the sample covariance matrix which is ill-conditioned in the large dimensional case. To fix this problem, we consider to shrink further the term $\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}$ toward the restricted statistics. As examples of the restriced cases, we consider the two cases;

(C1) Case of sphericity: $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}) = p^{-1}\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21})\boldsymbol{I}$.

(C2) Case of diagonality: $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}) = \mathrm{diag}(\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21})$.

We use the notation $\mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11})$ defined by $\mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11}) = \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2}) + \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21})$. Thus, we suggest a class of the doubly weighted estimators

$$\widehat{\mathbf{\Sigma}}(\gamma,\beta) = \gamma\widehat{\mathbf{\Sigma}}_{11.2} + (1-\gamma)\mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2}) + \beta\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21} + (1-\beta)\mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21}). \tag{2.4}$$

where $\mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2})$ is given above.

The class of the weighted estimators includes not only $\widehat{\mathbf{\Sigma}}_{\alpha}$ but also other specific estimators. For example, $\widehat{\mathbf{\Sigma}}(\gamma,\beta)$ reduces to $\widehat{\mathbf{\Sigma}}_{\alpha}$ when $\beta = 1$ and $\gamma = \alpha$. Also, putting $\beta = \gamma = w$ in the case of (C1) reduces to

$$\widehat{\mathbf{\Sigma}}_T^s = w\widehat{\mathbf{\Sigma}}_{11} + (1-w)\frac{1}{p}\mathrm{tr}\,(\widehat{\mathbf{\Sigma}}_{11})\boldsymbol{I}, \tag{2.5}$$

which is a ridge-type estimator given for high dimensional covariance matrix with no covariates. One can see that $\widehat{\mathbf{\Sigma}}_T^s$ is obtain by shrinking the sample covariance matrix to the direction of the spherical matrix $p^{-1}\mathrm{tr}\,(\widehat{\mathbf{\Sigma}}_{11})$. Finally, the ridge-type estimator in the case of (C2) is given by

$$\widehat{\mathbf{\Sigma}}_T^d = w\widehat{\mathbf{\Sigma}}_{11} + (1-w)\mathrm{diag}\,(\widehat{\mathbf{\Sigma}}_{11}). \tag{2.6}$$

Throughout the paper, we use the following notations for population matrices:
In the case of sphericity (C1), $\mathbf{\Lambda}(\mathbf{\Sigma}_{11}) = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11})\boldsymbol{I}$, $\mathbf{\Lambda}(\mathbf{\Sigma}_{11.2}) = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11.2})\boldsymbol{I}$ and $\mathbf{\Lambda}(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}) = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})\boldsymbol{I}$.
In the case of diagonality (C2), $\mathbf{\Lambda}(\mathbf{\Sigma}_{11}) = \mathrm{diag}\,(\mathbf{\Sigma}_{11})$, $\mathbf{\Lambda}(\mathbf{\Sigma}_{11.2}) = \mathrm{diag}\,(\mathbf{\Sigma}_{11.2})$ and $\mathbf{\Lambda}(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}) = \mathrm{diag}\,(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})$.

2.2 Approximations of the risk function under normality

We now approximate the risk function of the estimator $\widehat{\mathbf{\Sigma}}(\gamma,\beta)$ and derive the optimal weights of γ and β, where the risk function of an estimator $\boldsymbol{\delta}$ of $\mathbf{\Sigma}_{11}$ is given by

$$Risk(\boldsymbol{\delta}) = p^{-1}E[\mathrm{tr}\,[\{\boldsymbol{\delta} - \mathbf{\Sigma}_{11}\}^2]].$$

To this end, we need to assume normality of the distribution. Thus, it is assumed that $(\boldsymbol{y}_1,\boldsymbol{x}_1),\ldots,(\boldsymbol{y}_N,\boldsymbol{x}_N)$ are mutually independently and identically distributed as

$$\begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{x}_i \end{pmatrix} \sim \mathcal{N}_{p+q}\Big(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}\Big). \tag{2.7}$$

Then, the marginal distribution of $\boldsymbol{y}_i$ and the conditional distribution of $\boldsymbol{y}_i$ given $\boldsymbol{x}_i$ are

$$\boldsymbol{y}_i \sim \mathcal{N}_p(\boldsymbol{\mu}_1, \mathbf{\Sigma}_{11}), \tag{2.8}$$

$$\boldsymbol{y}_i|\boldsymbol{x}_i \sim \mathcal{N}_p(\boldsymbol{\mu}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_2), \mathbf{\Sigma}_{11.2}). \tag{2.9}$$

Conditional distribution (2.9) is expressed as

$$\boldsymbol{y}_i = (\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2) + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{x}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{11.2}), \tag{2.10}$$

which is identical to (1.3) if the covariance matrix $\boldsymbol{\Sigma}_{11.2}$ of $\boldsymbol{\epsilon}_i$ is replaced with $\boldsymbol{D} = \boldsymbol{\Sigma}_{11.2}$ or $\boldsymbol{D} = \boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{11.2})$.

Under assumption of normality, $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}})$ is distributed as

$$\begin{pmatrix} \overline{\boldsymbol{y}} \\ \overline{\boldsymbol{x}} \end{pmatrix} \sim \mathcal{N}_{p+q}\Big(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \frac{1}{N} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \Big),$$

which is independent of $\boldsymbol{V}_{11}$, $\boldsymbol{V}_{12}$ and $\boldsymbol{V}_{22}$. When $n \geq p+q$, $\boldsymbol{V}$ has the Wishart distribution

$$\boldsymbol{V} \sim \mathcal{W}_{p+q}\Big(n, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \Big).$$

However, in the case of $n < p+q$, it is noted that $\boldsymbol{V}$ does not have a Wishart distribution, and we cannot use distributional properties of the Wishart distribution.

To approximate the risk functions of the estimators, we assume the following conditions:

(A1) n, p and q satisfy that $(n,p) \to \infty$, $n \geq q$ and q is bounded.

(A2) $a_1 = p^{-1}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11}) = O(1)$, $a_{20} = p^{-1}\sum_{j=1}^{p}(\boldsymbol{\Sigma}_{11.2})_{jj}^2 = O(1)$, $b_1 = p^{-1}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}) = O(1)$, $b_2 = p^{-1}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}^2) = O(1)$, $b_{20} = p^{-1}\sum_{j=1}^{p}(\boldsymbol{\Sigma}_{11.2})_{jj}^2 = O(1)$, $\phi_{11} = p^{-1}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{11.2}) = O(1)$ and $\phi_{110} = p^{-1}\sum_{j=1}^{p}(\boldsymbol{\Sigma}_{11})_{jj}(\boldsymbol{\Sigma}_{11.2})_{jj} = O(1)$, where $(\boldsymbol{M})_{jj}$ stands for the (j,j)-element of $\boldsymbol{M}$.

(A3) $a_2 = p^{-1}\mathrm{tr}\,[\boldsymbol{\Sigma}_{11}^2] = O(p)$.

(A4) $a_2 = O(1)$.

Assumption (A1) means that we consider the high dimensional case that both n and p tend to infinity, but q is bounded. This restriction enables us to use standard theories of the Wishart distributions for $\boldsymbol{V}_{22}$. Assumption (A2) permits non-zero off-diagonal elements of $\boldsymbol{\Sigma}_{11.2}$, i.e., the cross sectional correlations of idiosyncratic components, while these elements are not so dense since $b_2 = p^{-1}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}^2) = O(1)$, $\phi_{110} = p^{-1}\sum_{j=1}^{p}(\boldsymbol{\Sigma}_{11})_{jj}(\boldsymbol{\Sigma}_{11.2})_{jj} = O(1)$, etc. The corresponding assumption in Fan, *et al.* al. (2011) is sparcity of $\boldsymbol{\Sigma}_{11.2}$, which restricts the maximal number of non-zero elements over the rows of $\boldsymbol{\Sigma}_{11.2}$ to order of $o(\sqrt{N/\log p})$. Note that, however, these two assumptions are not equivalent.

The condition (A3) $a_2 = O(p)$ implies that $\boldsymbol{\Sigma}_{12}$ is dense while (A4) $a_2 = O(1)$ implies that factor loadings $\boldsymbol{\Sigma}_{12}$ in factor model (1.3) is sparce or not dense because of $\mathrm{tr}\,(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^2 = O(p)$. The former is supposed to be more natural since we consider factors of bounded number q. As we mentioned in section 1, many literature of linear shrinkage estimation of large covariance matrices consider the case of $a_2 = O(1)$ whether

employing factor structure or not, however we mainly treat the case (A3) in which factor loadings are dense whereas the case (A4) is argued in the end of this subsection.

We begin by handling the risk of the general type of estimators $\widehat{\mathbf{\Sigma}}(\gamma, \beta)$ given in (2.4). The straightforward calculation leads to the following lemma.

Lemma 2.1. *In the case of* (C1) *or* (C2)*, the risk of the estimator* $\widehat{\mathbf{\Sigma}}(\gamma, \beta)$ *is written as*

$$Risk(\widehat{\mathbf{\Sigma}}(\gamma, \beta)) = \begin{pmatrix} \gamma & \beta \end{pmatrix} \begin{pmatrix} J_{11} & J_{12} \\ J_{12} & J_{22} \end{pmatrix} \begin{pmatrix} \gamma \\ \beta \end{pmatrix} - 2 \begin{pmatrix} J_{10} & J_{20} \end{pmatrix} \begin{pmatrix} \gamma \\ \beta \end{pmatrix} + R_0, \tag{2.11}$$

so that the optimal weights γ^* *and* β^* *are given by*

$$\begin{pmatrix} \gamma^* \\ \beta^* \end{pmatrix} = \begin{pmatrix} J_{11} & J_{12} \\ J_{12} & J_{22} \end{pmatrix}^{-1} \begin{pmatrix} J_{10} \\ J_{20} \end{pmatrix}, \tag{2.12}$$

and the risk of the estimator $\widehat{\mathbf{\Sigma}}(\gamma, \beta)$ *with the optimal weights* γ^* *and* β^* *is*

$$Risk(\widehat{\mathbf{\Sigma}}(\gamma^*, \beta^*)) = R_0 - \begin{pmatrix} J_{10} & J_{20} \end{pmatrix} \begin{pmatrix} J_{11} & J_{12} \\ J_{12} & J_{22} \end{pmatrix}^{-1} \begin{pmatrix} J_{10} \\ J_{20} \end{pmatrix},$$

where $R_0 = p^{-1} E\mathrm{tr}(\mathbf{\Sigma}_{11} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11}))^2$, $J_{11} = p^{-1} E\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{11.2} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2}))^2$,

$$\begin{aligned}
J_{22} =& \frac{1}{p} E\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21}))^2, \\
J_{12} =& \frac{1}{p} E\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{11.2} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2}))(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21})), \\
J_{10} =& \frac{1}{p} E\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{11.2} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2}))(\mathbf{\Sigma}_{11} - \mathbf{\Lambda}(\mathbf{\Sigma}_{11})), \\
J_{20} =& \frac{1}{p} E\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21}))(\mathbf{\Sigma}_{11} - \mathbf{\Lambda}(\mathbf{\Sigma}_{11})).
\end{aligned}$$

The moments given in Lemma 2.1 can be evaluated under normality assumption (2.7). We first assume non-sparcity condition (A3) to approximate the moments.

Theorem 2.1. *Assume* (A1), (A2) *and non-sparcity condition* (A3) *and that* $(\boldsymbol{y}_i, \boldsymbol{x}_i)$*'s are normally distributed as* (2.7)*. Then,* $R_0 = a_2 - a_1^2 + O(n^{-1})$ *in the case of* (C1) *and* $R_0 = a_2 - a_{20} + O(n^{-1})$ *in the case of* (C2)*. Also, in the case of* (C1)*,*

$$\begin{aligned}
J_{11} =& b_2 - b_1^2 + \frac{p}{n} b_1^2 + O(n^{-1}) + O(n^{-2}p), \\
J_{22} =& (1 + n^{-1})a_2 - 2\phi_{11} + b_2 - (a_1 - b_1)^2 + \frac{p}{n}(a_1^2 - b_1^2) + O(n^{-1}) + O(n^{-2}p), \\
J_{12} =& \phi_{11} - a_1 b_1 - b_2 + b_1^2 + O(n^{-1}), \\
J_{10} =& \phi_{11} - a_1 b_1 + O(n^{-1}), \\
J_{20} =& a_2 - a_1^2 - \phi_{11} + a_1 b_1 + O(n^{-1}).
\end{aligned} \tag{2.13}$$

On the other hand, in the case of (C2)*,*

$$\begin{aligned}
J_{11} =& b_2 - b_{20} + \frac{p}{n} b_1^2 + O(n^{-1}) + O(n^{-2}p), \\
J_{22} =& (1+n^{-1})a_2 - 2\phi_{11} + b_2 - a_{20} + 2\phi_{110} - b_{20} + \frac{p}{n}(a_1^2 - b_1^2) + O(n^{-1}) + O(n^{-2}p), \\
J_{12} =& \phi_{11} - \phi_{110} - b_2 + b_{20} + O(n^{-1}), \\
J_{10} =& \phi_{11} - \phi_{110} + O(n^{-1}), \\
J_{20} =& a_2 - a_{20} - \phi_{11} + \phi_{110} + O(n^{-1}).
\end{aligned} \tag{2.14}$$

Here, we analyze two special cases and suggest an partially optimal estimator in the latter.

[**Risk of $\widehat{\mathbf{\Sigma}}_{11}$**] We can easily see from (A.13) that

$$Risk(\widehat{\mathbf{\Sigma}}_{11}) = \frac{a_2}{n} + \frac{p}{n} a_1^2.$$

[**Risk of $\widehat{\mathbf{\Sigma}}_{\alpha}$**] In the case of sphericity, putting $\gamma = \alpha,\ \beta = 1$ and using (2.13) yields that

$$Risk(\widehat{\mathbf{\Sigma}}_{\alpha}) = (b_2 - b_1^2 + \frac{p}{n} b_1^2)\alpha^2 - 2(b_2 - b_1^2)\alpha + \frac{a_2}{n} + \frac{p}{n}(a_1^2 - b_1^2) + b_2 - b_1^2 + O(n^{-1}),$$

which is minimized at

$$\alpha^* = \frac{b_2 - b_1^2}{b_2 - b_1^2 + (p/n) b_1^2} \tag{2.15}$$

and the risk of $\widehat{\mathbf{\Sigma}}_{\alpha^*}$ becomes

$$Risk(\widehat{\mathbf{\Sigma}}_{\alpha^*}) = \frac{a_2}{n} + \frac{p}{n} a_1^2 - b_1^2 \left(\frac{p}{n} - \frac{p(b_2/b_1^2 - 1)}{n(b_2/b_1^2 - 1) + p} \right) + O(n^{-1}),$$

which is smaller than the risk of $\widehat{\mathbf{\Sigma}}_{11}$ in the leading terms.

In the case of diagonality, putting $\gamma = \alpha,\ \beta = 1$ and using (2.14) leads that

$$Risk(\widehat{\mathbf{\Sigma}}_{\alpha}) = (b_2 - b_{20} + \frac{p}{n} b_1^2)\alpha^2 - 2(b_2 - b_{20})\alpha + \frac{a_2}{n} + \frac{p}{n}(a_1^2 - b_1^2) + b_2 - b_{20} + O(n^{-1}),$$

which is minimized at

$$\alpha^* = \frac{b_2 - b_{20}}{b_2 - b_{20} + (p/n) b_1^2} \tag{2.16}$$

and the risk of $\widehat{\mathbf{\Sigma}}_{\alpha^*}$ becomes

$$Risk(\widehat{\mathbf{\Sigma}}_{\alpha^*}) = \frac{a_2}{n} + \frac{p}{n} a_1^2 - b_1^2 \left(\frac{p}{n} - \frac{p(b_2 - b_{20})/b_1^2}{n(b_2 - b_{20})/b_1^2 + p} \right) + O(n^{-1}),$$

which is smaller than the risk of $\widehat{\mathbf{\Sigma}}_{11}$ in the leading terms.

We next provide approximations of the moments under sparcity condition (A3). The proof is omitted.

Theorem 2.2. *Assume sparcity condition* (A4) *instead of* (A3) *together with* (A1) *and* (A2). *Then,* R_0 *and* J_{22} *in* (2.13) *should be replaced with* $R_0 = a_2 - a_1^2 + O(n^{-1}p^{-1})$ *and*

$$J_{22} = a_2 - 2\phi_{11} + b_2 - (a_1 - b_1)^2 + \frac{p}{n}(a_1^2 - b_1^2) + O(n^{-1}) + O(n^{-2}p),$$

in the case of (C1). *Similarly,* $R_0 = a_2 - a_1^2 + O(n^{-1}p^{-1})$ *and*

$$J_{22} = a_2 - 2\phi_{11} + b_2 - a_{20} + 2\phi_{110} - b_{20} + \frac{p}{n}(a_1^2 - b_1^2) + O(n^{-1}) + O(n^{-2}p),$$

in the case of (C2). *For the other terms* J_{ij}*'s, their evaluations under sparcity of factor loadings are the same as in Theorem 2.1 in both of* (C1) *and* (C2).

3 Construction of Plug-In Estimators

Since the estimator $\widehat{\mathbf{\Sigma}}_{\gamma,\beta}$ with the optimal weights depends on the unknown parameters, we need to estimate them. For a_1, a_2, b_1 and b_2, we use the estimators given in Srivastava (2005) as

$$\begin{aligned}
&\hat{a}_1 = \frac{1}{p}\mathrm{tr}\,(\widehat{\mathbf{\Sigma}}_{11}), \quad \hat{a}_2 = \frac{n^2}{p(n-1)(n+2)}\Big(\mathrm{tr}\,(\widehat{\mathbf{\Sigma}}_{11}^2) - (\mathrm{tr}\,\widehat{\mathbf{\Sigma}}_{11})^2/n\Big),\\
&\hat{b}_1 = \frac{n}{p(n-q)}\mathrm{tr}\,(\widehat{\mathbf{\Sigma}}_{11.2}), \quad \hat{b}_2 = \frac{n^2}{p(n-q-1)(n-q+2)}\Big(\mathrm{tr}\,(\widehat{\mathbf{\Sigma}}_{11.2}^2) - (\mathrm{tr}\,\widehat{\mathbf{\Sigma}}_{11.2})^2/(n-q)\Big).
\end{aligned} \tag{3.1}$$

Also, for a_{20}, b_{20}, ϕ_{11} and ϕ_{110}, we show that the followings are unbiased and consistent:

$$\begin{aligned}
\hat{a}_{20} =& \frac{n}{p(n+2)}\mathrm{tr}\,((\mathrm{diag}\widehat{\mathbf{\Sigma}}_{11})^2), \quad \hat{b}_{20} = \frac{n^2}{p(n-q)(n-q+2)}\mathrm{tr}\,((\mathrm{diag}\widehat{\mathbf{\Sigma}}_{11.2})^2),\\
\widehat{\phi}_{11} =& \frac{n}{p(n-q)}\Big(\mathrm{tr}\,(\widehat{\mathbf{\Sigma}}_{11}\widehat{\mathbf{\Sigma}}_{11.2}) - \frac{n-q-2}{(n-q-1)(n-q+2)}\mathrm{tr}\,(\widehat{\mathbf{\Sigma}}_{11.2}^2)\\
&- \frac{n-q}{(n-q-1)(n-q+2)}(\mathrm{tr}\,\widehat{\mathbf{\Sigma}}_{11.2})^2\Big),\\
\widehat{\phi}_{110} =& \frac{n}{p(n-q)}\mathrm{tr}\,\Big(\mathrm{diag}(\widehat{\mathbf{\Sigma}}_{11})\mathrm{diag}(\widehat{\mathbf{\Sigma}}_{11.2})\Big) - \frac{2n}{p(n-q)(n-q+2)}\mathrm{tr}\,\Big(\mathrm{diag}(\widehat{\mathbf{\Sigma}}_{11.2})^2\Big).
\end{aligned} \tag{3.2}$$

To evaluate these estimators, we need to assume two more conditions:

(A5) $a_4 = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11}^4) = O(p^3)$, $b_3 = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11.2}^3) = O(1)$, $b_4 = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11.2}^4) = O(1)$, $\mathrm{tr}\,(\mathbf{\Sigma}_{11.2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})^2 = O(p^2)$ and $\mathrm{tr}\,(\mathbf{\Sigma}_{11.2}^2\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}) = O(p)$.

(A6) $a_{40} = p^{-1}\sum_{j=1}^p(\mathbf{\Sigma}_{11})_{jj}^4 = O(1)$, $b_{40} = p^{-1}\sum_{j=1}^p(\mathbf{\Sigma}_{11.2})_{jj}^4 = O(1)$, $\mathrm{tr}\,(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\circ\mathbf{\Sigma}_{11.2})^2 = O(p)$, $\mathrm{tr}\,(\mathbf{\Sigma}_{11.2}\circ\mathbf{\Sigma}_{11.2})^2 = O(p)$, $\mathrm{tr}\,(\mathbf{\Sigma}_{11.2}\mathrm{diag}\,(\mathbf{\Sigma}_{11.2}))^2 = O(p)$, $\mathrm{tr}\,(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathrm{diag}\,(\mathbf{\Sigma}_{11.2}))^2 = O(p^2)$, $\mathrm{tr}\,(\mathrm{diag}\,(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})\mathrm{diag}\,(\mathbf{\Sigma}_{11.2}))^2 = O(p)$

and $\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}\mathrm{diag}\,(\boldsymbol{\Sigma}_{11.2})\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\mathrm{diag}\,(\boldsymbol{\Sigma}_{11.2})) = O(p)$,

where $(\boldsymbol{M})_{jj}$ stands for the (j,j) element of $\boldsymbol{M}$ and $M \circ L = (M_{ij}L_{ij})_{ij}$ is the Hadamard product.

We first evaluate the above estimators under the non-sparcity condition (A3).

Theorem 3.1. *Assume* (A1), (A2), (A3) *and* (A5), *then under normality,* $\hat{a}_1$, $\hat{a}_2$, $\hat{b}_1$ *and* $\hat{b}_2$ *are all unbiased and*

$$\begin{aligned} &\hat{a}_1 = a_1 + O_p(n^{-1/2}),\ \hat{a}_2 = a_2 + O_p(n^{-1/2}p), \\ &\hat{b}_1 = b_1 + O_p((np)^{-1/2}),\ \hat{b}_2 = b_2 + O_p((np)^{-1/2}) + O_p(n^{-1}). \end{aligned} \tag{3.3}$$

Also, (A1), (A2), (A3), (A5) *and* (A6) *with normality implies that* $\hat{a}_{20}$, $\hat{a}_{20}$, $\widehat{\phi}_{11}$ *and* $\widehat{\phi}_{110}$ *are all unbiased and*

$$\begin{aligned} &\hat{a}_{20} = a_{20} + O_p(n^{-1/2}),\ \hat{b}_{20} = b_{20} + O_p((np)^{-1/2}), \\ &\widehat{\phi}_{11} = \phi_{11} + O_p(n^{-1/2}),\ \widehat{\phi}_{110} = \phi_{110} + O_p(n^{-1/2}). \end{aligned} \tag{3.4}$$

Thus, all of the aboves are reasonable estimators of the corresponding ones except $\hat{a}_2$.

A problem is the convergence rate of $\hat{a}_2$. If sparcity condition (A4) holds, namely, $a_2 = O(1)$, then $\hat{a}_2 = a_2 + O_p((np)^{-1/2}) + O_p(n^{-1})$, so that even $p \geq n$, $\hat{a}_2$ is a consistent estimator of a_2. If $a_2 = O(p)$, however, we see that $\hat{a}_2 = a_2 + O_p(n^{-1/2}p)$. Therefore, $\hat{a}_2$ is unbiased but unstable in the sense of $\hat{a}_2$ itself is not a consistent estimator of a_2 unless $n > p$ and $p = O(n^{\delta})$ with $\delta < 1/2$, which is a strong condition.

Substituting (3.3) and (3.4) into (2.13) or (2.14) and further substituting (2.13) or (2.14) into (2.12), we get estimated optimal weights $(\hat{\gamma}^*, \hat{\beta}^*)$ of estimator (2.4). We suggest the plug-in estimator $\widehat{\boldsymbol{\Sigma}}_{\hat{\gamma}^*,\hat{\beta}^*}$. Also, substituting (3.3) and (3.4) into (2.15) or (2.16), we get estimated optimal weights $\hat{\alpha}^*$ of estimator (2.3), so that we suggest the plug-in estimator $\widehat{\boldsymbol{\Sigma}}_{\hat{\alpha}^*}$.

If sparcity condition (A4) holds instead of (A3), we can show the consistency of all the estimators in the following theorem, where the proof is omitted.

Theorem 3.2. *Assume sparcity condition* (A4) *together with* (A1), (A2), (A5) *and* (A6), *where the conditions on* a_4, $\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^2$ *and* $\mathrm{tr}\,(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})^2$ *in* (A5) *and* (A6) *are replaced with* $a_4 = p^{-1}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11}^4) = O(1)$,

$$\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^2 = O(p) \text{ and } \mathrm{tr}\,(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})^2 = O(p).$$

Then, under normality, one gets

$$\begin{aligned} &\hat{a}_1 = a_1 + O_p((np)^{-1/2}),\ \hat{a}_2 = a_2 + O_p((np)^{-1/2}) + O_p(n^{-1}), \\ &\hat{b}_1 = b_1 + O_p((np)^{-1/2}),\ \hat{b}_2 = b_2 + O_p((np)^{-1/2}) + O_p(n^{-1}), \\ &\hat{a}_{20} = a_{20} + O_p((np)^{-1/2}),\ \hat{b}_{20} = b_{20} + O_p((np)^{-1/2}), \\ &\widehat{\phi}_{11} = \phi_{11} + O_p((np)^{-1/2}),\ \widehat{\phi}_{110} = \phi_{110} + O_p((np)^{-1/2}). \end{aligned} \tag{3.5}$$

4 Simulation Studies

We now investigate numerical performance of MSEs or risks of the proposed estimators through simulations.

As a structure of the covariance matrices, we follow model (1.3) where $\boldsymbol{\epsilon}_i$'s and $\boldsymbol{x}_i$'s are given by

$$\boldsymbol{\epsilon}_i = \boldsymbol{\Sigma}_{11.2}^{1/2}\boldsymbol{u}_i, \qquad \boldsymbol{x}_i = \boldsymbol{\Sigma}_{22}^{1/2}\boldsymbol{v}_i$$

with $\boldsymbol{u}_i = (u_{ij})_{1\le j\le p}$ and $\boldsymbol{v}_i = (v_{ij})_{1\le j\le q}$ mutually independent. Here, we set $\boldsymbol{\mu}_2 = \boldsymbol{0}$, $\boldsymbol{\Sigma}_{22} = \boldsymbol{I}_q$ and $(\boldsymbol{\Sigma}_{12})_{ij} \sim$ i.i.d. $\mathcal{N}(0.5, 1)$ through the studies, and the following two cases are treated for $\boldsymbol{\Sigma}_{11.2}$:

(M1) the strict factor model(sphericity) with $\boldsymbol{\Sigma}_{11.2} = 3\boldsymbol{I}$,

(M2) the approximate factor model with

$$\boldsymbol{\Sigma}_{11.2} = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_p \end{pmatrix} \begin{pmatrix} \rho^{|1-1|/7} & \rho^{|1-2|/7} & \dots & \rho^{|1-p|/7} \\ \rho^{|2-1|/7} & \rho^{|2-2|/7} & \dots & \rho^{|2-p|/7} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{|p-1|/7} & \rho^{|p-2|/7} & \dots & \rho^{|p-p|/7} \end{pmatrix} \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_p \end{pmatrix},$$

for $\sigma_j = 3 + 0.2(-1)^{j-1}(p-j+1)/p$ and $\rho = 0.1$.

Note that the above settings are consistent with assumptions (A2), (A3), (A5) and (A6). Concerning the distributions of $\boldsymbol{u}_i = (u_{ij})_{1\le j\le p}$ and $\boldsymbol{v}_i = (v_{ij})_{1\le j\le q}$, we treat the two cases:

(D1) $u_{ij}, v_{ij} \sim \mathcal{N}(0,1)$

(D2) $u_{ij}, v_{ij} = (w_{ij} - \nu)/\sqrt{2\nu}, \quad w_{ij} \sim \chi^2_\nu$ for $\nu = 2$.

Let $\widehat{\boldsymbol{\Sigma}}_T^s$ and $\widehat{\boldsymbol{\Sigma}}_T^d$ be given in (2.5) and (2.6), and let $\widehat{\boldsymbol{\Sigma}}_\alpha^s$, $\widehat{\boldsymbol{\Sigma}}_\alpha^d$, $\widehat{\boldsymbol{\Sigma}}_{\gamma,\beta}^s$ and $\widehat{\boldsymbol{\Sigma}}_{\gamma,\beta}^d$ be given in Sections 2 and 3, where the corresponding optimal weights are estimated by the consistent estimators. Then we compare the performances of the seven estimators: the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}_{11}$, $\widehat{\boldsymbol{\Sigma}}_T^s$, $\widehat{\boldsymbol{\Sigma}}_T^d$, $\widehat{\boldsymbol{\Sigma}}_\alpha^s$, $\widehat{\boldsymbol{\Sigma}}_\alpha^d$, $\widehat{\boldsymbol{\Sigma}}_{\gamma,\beta}^s$ and $\widehat{\boldsymbol{\Sigma}}_{\gamma,\beta}^d$.

Some simulation experiments are carried out in the cases of $N = 100$, $q = 3$ and $p = 50$, 100 and 200. Based on 1,000 replications, we calculate averages of the risks given by

$$Risk(\boldsymbol{\delta}) = E[\mathrm{tr}\,(\boldsymbol{\delta} - \boldsymbol{\Sigma}_{11})^2]/p,$$

where $\boldsymbol{\delta}$ is an estimator of $\boldsymbol{\Sigma}_{11}$. We also investigate the performances of the risks when we estimate the precision matrix $\boldsymbol{\Sigma}_{11}^{-1}$ with the inverse of the above estimators, where the risk is evaluated in lights of

$$Risk(\boldsymbol{\delta}^{-1}) = E[\mathrm{tr}\,(\boldsymbol{\delta}^{-1}\boldsymbol{\Sigma}_{11} - \boldsymbol{I})^2]/p,$$

where $\boldsymbol{\delta}^{-1}$ is an estimator of $\boldsymbol{\Sigma}_{11}^{-1}$.

Tables 1 and 2 report the simulation results in estimation of $\mathbf{\Sigma}_{11}$ ane $\mathbf{\Sigma}_{11}^{-1}$, respectively, under strict factor model (M1) with underlying distribution (D1) or (D2), where the values in the parenthesis denote the corresponding standard deviations. Under the assumption of strict factor model (M1), the four estimators $\widehat{\mathbf{\Sigma}}_{\alpha}^{s}$, $\widehat{\mathbf{\Sigma}}_{\alpha}^{d}$, $\widehat{\mathbf{\Sigma}}_{\gamma,\beta}^{s}$ and $\widehat{\mathbf{\Sigma}}_{\gamma,\beta}^{d}$ perform well. Especially, the estimator $\widehat{\mathbf{\Sigma}}_{\gamma,\beta}^{s}$ is the best of these. Tables 3 and 4 report the simulation results under approximate factor model (M2). Since (M2) does not assume the strict factor model, the estimators $\widehat{\mathbf{\Sigma}}_{T}^{s}$ and $\widehat{\mathbf{\Sigma}}_{T}^{d}$ are better than $\widehat{\mathbf{\Sigma}}_{\alpha}^{s}$ and $\widehat{\mathbf{\Sigma}}_{\alpha}^{d}$ in relatively small p, however $\widehat{\mathbf{\Sigma}}_{T}^{s}$ and $\widehat{\mathbf{\Sigma}}_{T}^{d}$ are ill-conditioned in the estimation of precision $\mathbf{\Sigma}_{11}^{-1}$ in large p as indicated in Table 4. The estimator $\widehat{\mathbf{\Sigma}}_{\gamma,\beta}^{s}$ is the best in most of these kinds of simulation experiments. Though we do not show the results, we also compared with the estimators in Fan *et al.* al. (2011) and found out that $\widehat{\mathbf{\Sigma}}_{\gamma,\beta}$'s perform better. Especially, in approximate factor model (M2), the estimator of the inverse $\mathbf{\Sigma}_{11}^{-1}$ in Fan *et al.* al. (2011) is quite unstable probably since $\mathbf{\Sigma}_{11.2}$ is not sparse. Moreover, Fan *et al.* al. (2011) is based on adaptive thresholding proposed in Cai and Liu (2011) which suggests a computationally hard data-driven choice of a tuning parameter that controls the threshold, while our estimators can be implemented easily. We thus recommend the estimator $\widehat{\mathbf{\Sigma}}_{\gamma,\beta}^{s}$ in estimation of $\mathbf{\Sigma}_{11}$ and $\mathbf{\Sigma}_{11}^{-1}$.

Table 1: Comparison of Estimators of $\mathbf{\Sigma}_{11}$ under Strict Factor Model (M1)

N	p	q	dist.	$\widehat{\mathbf{\Sigma}}_{11}$	$\widehat{\mathbf{\Sigma}}_{T}^{s}$	$\widehat{\mathbf{\Sigma}}_{T}^{d}$	$\widehat{\mathbf{\Sigma}}_{\alpha}^{s}$	$\widehat{\mathbf{\Sigma}}_{\alpha}^{d}$	$\widehat{\mathbf{\Sigma}}^{s}(\gamma,\beta)$	$\widehat{\mathbf{\Sigma}}^{d}(\gamma,\beta)$
100	50	3	(D1)	24.5	22.6	22.7	20.0	20.2	18.7	18.9
				(7.1)	(6.6)	(6.6)	(7.1)	(7.1)	(6.4)	(6.4)
100	100	3	(D1)	47.3	42.9	43.0	38.4	38.6	35.4	35.7
				(12.3)	(9.9)	(9.9)	(12.3)	(12.3)	(9.8)	(9.8)
100	200	3	(D1)	93.9	85.9	86.0	76.2	76.4	70.8	71.0
				(23.1)	(19.2)	(19.2)	(23.1)	(23.1)	(18.8)	(18.8)
100	50	3	(D2)	38.2	34.7	34.9	33.2	33.9	30.5	31.2
				(27.2)	(21.7)	(21.7)	(27.3)	(27.2)	(22.0)	(22.3)
100	100	3	(D2)	75.2	68.4	68.6	65.8	66.5	60.5	61.2
				(49.6)	(38.9)	(38.9)	(49.7)	(49.7)	(39.6)	(39.9)
100	200	3	(D2)	158.3	142.7	142.9	140.1	140.8	127.5	128.3
				(93.8)	(71.7)	(71.7)	(94.0)	(93.9)	(73.4)	(73.6)

Table 2: Comparison of Estimators of $\mathbf{\Sigma}_{11}^{-1}$ under Strict Factor Model (M1)

N	p	q	dist.	$\widehat{\mathbf{\Sigma}}_{11}$	$\widehat{\mathbf{\Sigma}}_T^s$	$\widehat{\mathbf{\Sigma}}_T^d$	$\widehat{\mathbf{\Sigma}}_\alpha^s$	$\widehat{\mathbf{\Sigma}}_\alpha^d$	$\widehat{\mathbf{\Sigma}}^s(\gamma,\beta)$	$\widehat{\mathbf{\Sigma}}^d(\gamma,\beta)$
100	50	3	(D1)	10.57	2.07	2.24	0.58	0.68	0.48	0.57
				(2.50)	(0.36)	(0.39)	(0.10)	(0.12)	(0.08)	(0.10)
100	100	3	(D1)	NA	8.29	9.90	1.16	1.34	0.96	1.11
					(1.04)	(1.29)	(0.18)	(0.15)	(0.13)	(0.15)
100	200	3	(D1)	NA	33.06	44.47	2.37	2.71	1.97	2.24
					(3.47)	(5.11)	(0.25)	(0.28)	(0.22)	(0.24)
100	50	3	(D2)	12.16	2.26	2.63	0.61	0.82	0.51	0.67
				(3.24)	(0.44)	(0.52)	(0.14)	(0.18)	(0.12)	(0.15)
100	100	3	(D2)	NA	9.57	11.89	1.28	1.64	1.07	1.32
					(1.65)	(2.07)	(0.24)	(0.30)	(0.22)	(0.26)
100	200	3	(D2)	NA	38.06	54.52	2.63	3.27	2.19	2.64
					(6.46)	(10.2)	(0.45)	(0.54)	(0.40)	(0.48)

Table 3: Comparison of Estimators of $\mathbf{\Sigma}_{11}$ under Approximate Factor Model (M2)

N	p	q	dist.	$\widehat{\mathbf{\Sigma}}_{11}$	$\widehat{\mathbf{\Sigma}}_T^s$	$\widehat{\mathbf{\Sigma}}_T^d$	$\widehat{\mathbf{\Sigma}}_\alpha^s$	$\widehat{\mathbf{\Sigma}}_\alpha^d$	$\widehat{\mathbf{\Sigma}}^s(\gamma,\beta)$	$\widehat{\mathbf{\Sigma}}^d(\gamma,\beta)$
100	50	3	(D1)	102.1	88.6	90.0	92.6	94.0	88.1	89.6
				(19.0)	(15.6)	(15.7)	(17.1)	(17.2)	(15.6)	(15.6)
100	100	3	(D1)	175.3	144.2	145.4	147.0	148.4	138.5	140.1
				(26.7)	(19.2)	(19.2)	(23.1)	(23.0)	(18.9)	(18.9)
100	200	3	(D1)	338.5	262.6	263.7	257.1	258.6	235.2	236.9
				(37.1)	(28.2)	(28.2)	(34.0)	(33.9)	(25.6)	(25.6)
100	50	3	(D2)	108.0	90.5	92.2	96.6	98.4	90.8	92.7
				(29.8)	(20.5)	(20.6)	(26.2)	(26.2)	(20.9)	(21.2)
100	100	3	(D2)	212.6	171.3	173.1	181.3	183.5	166.2	168.8
				(65.3)	(40.3)	(40.4)	(62.2)	(62.1)	(43.0)	(43.5)
100	200	3	(D2)	410.8	325.5	327.3	325.7	328.3	296.5	299.4
				(125.7)	(79.6)	(79.6)	(120.1)	(120.2)	(84.8)	(85.3)

5 Applications to Portfolio Managements

In this section, we consider Markowitz's problem (return maximization) under the constraint of the portfolio variance:

$$R = \max \boldsymbol{c}^T \boldsymbol{\mu} \quad \text{s.t.} \quad \boldsymbol{c}^T \mathbf{1} \leq 1 \quad \text{and} \quad \boldsymbol{c}^T \mathbf{\Sigma} \boldsymbol{c} \leq \sigma_0^2,$$

Table 4: Comparison of Estimators of $\mathbf{\Sigma}_{11}^{-1}$ under Approximate Factor Model (M2)

N	p	q	dist.	$\widehat{\mathbf{\Sigma}}_{11}$	$\widehat{\mathbf{\Sigma}}_T^s$	$\widehat{\mathbf{\Sigma}}_T^d$	$\widehat{\mathbf{\Sigma}}_\alpha^s$	$\widehat{\mathbf{\Sigma}}_\alpha^d$	$\widehat{\mathbf{\Sigma}}^s(\gamma,\beta)$	$\widehat{\mathbf{\Sigma}}^d(\gamma,\beta)$
100	50	3	(D1)	21.92	1.29	1.42	1.51	1.71	1.15	1.30
				(4.97)	(0.17)	(0.19)	(0.19)	(0.23)	(0.14)	(0.17)
100	100	3	(D1)	NA	2.54	2.89	2.06	2.37	1.71	1.96
					(0.32)	(0.38)	(0.18)	(0.22)	(0.15)	(0.18)
100	200	3	(D1)	NA	5.5	6.41	2.89	3.32	2.42	2.76
					(0.58)	(0.73)	(0.21)	(0.25)	(0.18)	(0.21)
100	50	3	(D2)	24.01	1.33	1.49	1.55	1.82	1.26	1.44
				(5.42)	(0.20)	(0.23)	(0.22)	(0.28)	(0.18)	(0.22)
100	100	3	(D2)	NA	2.67	3.05	2.22	2.65	1.78	2.07
					(0.2)	(0.39)	(0.47)	(0.32)	(0.20)	(0.24)
100	200	3	(D2)	NA	6.87	8.37	3.07	3.66	2.62	3.08
					(1.41)	(1.92)	(0.31)	(0.38)	(0.28)	(0.33)

where $\boldsymbol{\mu}$ is a vector of expected returns of assets in a portfolio, $\boldsymbol{c}$ is a vector of weights of the corresponding assets, and R is the expected return of the portfolio. Also, $\mathbf{\Sigma}$ is a covariance matrix of assets, and $\sigma_0^2 > 0$ denotes an upper bound of the portfolio variance. The analytic solution of the above quadratic design problem is known as:

(1) if $\sigma_0^2 \mathbf{1}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu} / \sqrt{\boldsymbol{\mu}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu}} \leq 1$,

$$\boldsymbol{c} = \frac{\sigma_0}{\sqrt{\boldsymbol{\mu}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu}}} \mathbf{\Sigma}^{-1}\boldsymbol{\mu},$$
$$R = \sigma_0 \sqrt{\boldsymbol{\mu}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu}},$$

(2) if $\sigma_0^2 \mathbf{1}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu} / \sqrt{\boldsymbol{\mu}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu}} \geq 1$,

$$\boldsymbol{c} = \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T \mathbf{\Sigma}^{-1}\mathbf{1}} + \Delta \left(\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - \frac{\mathbf{1}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T \mathbf{\Sigma}^{-1}\mathbf{1}} \mathbf{\Sigma}^{-1}\mathbf{1} \right),$$
$$R = \frac{\mathbf{1}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T \mathbf{\Sigma}^{-1}\mathbf{1}} + \Delta \left(\boldsymbol{\mu}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu})^2}{\mathbf{1}^T \mathbf{\Sigma}^{-1}\mathbf{1}} \right),$$

where

$$\Delta = \sqrt{\frac{\sigma_0^2 \mathbf{1}^T \mathbf{\Sigma}^{-1}\mathbf{1} - 1}{(\boldsymbol{\mu}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu})(\mathbf{1}^T \mathbf{\Sigma}^{-1}\mathbf{1}) - (\mathbf{1}^T \mathbf{\Sigma}^{-1}\boldsymbol{\mu})^2}}.$$

In practice, portfolio managers are usually willing to maximize expected returns subject to given risk upper bounds which are typically determined by financial circumstances

of their firms or requirement of their customers. Since they do not know the true covariance matrix $\mathbf{\Sigma}$ and the true vector of the expected returns $\boldsymbol{\mu}$, they need to estimate them somehow by $\widehat{\mathbf{\Sigma}}$ and $\hat{\boldsymbol{\mu}}$ respectively and estimate the optimal weights as $\hat{\boldsymbol{c}}$ based on these estimators. Hence, the estimated variance of the portfolio would be $\hat{\boldsymbol{c}}^T\widehat{\mathbf{\Sigma}}\hat{\boldsymbol{c}}$ and the estimated expected return of it would be $\hat{\boldsymbol{c}}^T\hat{\boldsymbol{\mu}}$. However, it is known that these estimators are typically upper biased (e.g. Bai, Huixia, and Wing-Keung (2009)). Thus, if portfolio managers compose portfolio satisfying that $\hat{\boldsymbol{c}}^T\widehat{\mathbf{\Sigma}}\hat{\boldsymbol{c}}$ is lower than the predetermined risk (variance) upper bound σ_0^2, the actual risk may exceed σ_0^2.

Therefore it is of interest to find out the actual risk of the portfolio which is composed based on estimators of $\mathbf{\Sigma}$. We simulate the actual portfolio variances in the same settings as in Section 4. Table 5 reports actual variances of portfolio based on covariance estimators in Section 4, which indicates that the estimators $\widehat{\mathbf{\Sigma}}^s_{\gamma,\beta}$ and $\widehat{\mathbf{\Sigma}}^d_{\gamma,\beta}$ preserve the constraints of the actual portfolio variance. Here, we set $N = p = 100$, $q = 3$ and calculate variances based on 1000 times replications under model (M2) and underlying distribution (D1) in Section 4. For other settings of distributions of $\boldsymbol{x}_i$'s and $\boldsymbol{\epsilon}_i$'s, we get similar performances and omit the details here.

Table 5: Comparison of Actual Variances of Portfolio based on Covariance Estimators under (M2) and (D1) (ρ denotes the parameter associate with correlation in case (M2))

ρ	σ_0^2	$\widehat{\mathbf{\Sigma}}_{11}$	$\widehat{\mathbf{\Sigma}}^s_\alpha$	$\widehat{\mathbf{\Sigma}}^d_\alpha$	$\widehat{\mathbf{\Sigma}}^s_T$	$\widehat{\mathbf{\Sigma}}^d_T$	$\widehat{\mathbf{\Sigma}}^s(\gamma,\beta)$	$\widehat{\mathbf{\Sigma}}^d(\gamma,\beta)$
0	0.098	NA	0.102	0.109	0.286	0.314	0.092	0.097
0.2	0.111	NA	0.140	0.148	0.142	0.159	0.113	0.116
0.4	0.366	NA	0.427	0.457	0.327	0.373	0.298	0.306

6 Concluding Remarks

In estimation of large covariance matrices, we have considered the general class of weighted estimators which includes (i) linear combinations of the sample covariance matrix and the specific estimators suggested under the strict factor models and (ii) the ridge-type estimators suggested in high dimensional situations as special cases. Under the assumptions of non-sparcity and sparcity of factor loadings, we have derived the optimal weights and provided their consistent estimators. The resulting plug-in estimators are invertible and well-conditioned. They are also useful not only when the strict factor models are suspected, but also when the approximate factor models hold or the factor models do not hold. Numerical results have shown that the suggested estimators perform well under

both normal and non-normal distributions. In the application to the portfolio managements, we have shown that the procedures based on the suggested estimators preserve the predetermined risk upper bounds robustly.

As pointed out below Theorem 3.1, the sparcity and non-sparcity of factor loadings influences the consistency of $\hat{a}_2$, namely, in the case of sparcity $a_2 = O(1)$, $\hat{a}_2$ is consistent when $(np)^{-1/2} \to 0$, while in the case of non-sparcity, $\hat{a}_2$ is consistent when $pn^{-1/2} \to 0$. This condition in the non-sparcity case is restrictive, and an improved estimator of a_2 will be desired.

Acknowledgments.

The second author acknowledges support from Grant-in-Aid for Scientific Research (23243039 and 26330036), Japan.

A Proofs

A.1 Proof of Theorem 2.1

Denote $\boldsymbol{V}_{11.2} := \boldsymbol{V}_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}$. If $n \geq p+q$, then from the standard theory of the Wishart distribution, $\boldsymbol{V}_{11.2}$ is independent of $(\boldsymbol{V}_{12}, \boldsymbol{V}_{22})$, and $\boldsymbol{V}_{11.2} \sim \mathcal{W}_p(n-q, \boldsymbol{\Sigma}_{11.2})$, $\boldsymbol{V}_{12}|\boldsymbol{V}_{22} \sim \mathcal{N}_{p,q}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{V}_{22}, \boldsymbol{\Sigma}_{11.2}, \boldsymbol{V}_{22})$ and $\boldsymbol{V}_{22} \sim \mathcal{W}_q(n, \boldsymbol{\Sigma}_{22})$. These properties will be helpful for evaluating the risk in the case of $n \geq p+q$. When $p > n$, however, we could not use these properties. To evaluate the risk in this case, we prepare the following lemmas which hold for any order of n and p.

Lemma A.1. *Let $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{X}^T = \sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^T$ for $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$, where $\boldsymbol{X}_i$'s are mutually independently and identically distributed as $\boldsymbol{X}_i \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. Then, $E[\boldsymbol{W}] = n\boldsymbol{\Sigma}$ and $E[\boldsymbol{W}^2] = n(n+1)\boldsymbol{\Sigma}^2 + n(\mathrm{tr}\,[\boldsymbol{\Sigma}])\boldsymbol{\Sigma}$. For $p \times p$ symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, it holds that $E[\mathrm{tr}\,\boldsymbol{A}\boldsymbol{W}^2] = n(n+1)(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{\Sigma}^2) + n(\mathrm{tr}\,(\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{\Sigma})$ and $E[(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{B}\boldsymbol{W})] = n^2(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{B}\boldsymbol{\Sigma}) + 2n\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma})$. Also,*

$$\begin{aligned} E[(\mathrm{tr}\,\boldsymbol{W}^2)^2] =& 4n^2(2n^2+5n+5)\mathrm{tr}\,\boldsymbol{\Sigma}^4 + 16n(n+1)\mathrm{tr}\,\boldsymbol{\Sigma}^3\mathrm{tr}\,\boldsymbol{\Sigma} \\ &+ n(n^3+2n^2+5n+4)(\mathrm{tr}\,\boldsymbol{\Sigma}^2)^2 \\ &+ 2n(n^2+n+4)(\mathrm{tr}\,\boldsymbol{\Sigma}^2)(\mathrm{tr}\,\boldsymbol{\Sigma})^2 + n^2(\mathrm{tr}\,\boldsymbol{\Sigma})^4, \end{aligned} \tag{A.1}$$

and

$$\begin{aligned} E[(\mathrm{tr}\,\boldsymbol{W})^4] =& 48n\mathrm{tr}\,\boldsymbol{\Sigma}^4 + 32n^2\mathrm{tr}\,\boldsymbol{\Sigma}^3\mathrm{tr}\,\boldsymbol{\Sigma} \\ &+ 12n^2(\mathrm{tr}\,\boldsymbol{\Sigma}^2)^2 + 12n^3(\mathrm{tr}\,\boldsymbol{\Sigma}^2)(\mathrm{tr}\,\boldsymbol{\Sigma})^2 + n^4(\mathrm{tr}\,\boldsymbol{\Sigma})^4. \end{aligned} \tag{A.2}$$

Watamori (1990) derived the equalities in (A.1) and (A.2) using the properties of the Wishart distribution in the case of $n \geq p$. Lemma A.1 comfirms that the same equalities hold for any (n, p) without assuming the Wishart distribution.

Proof. It is easy to see that $E[\boldsymbol{W}] = \sum_{i=1}^{n} E[\boldsymbol{X}_i \boldsymbol{X}_i^T] = n\boldsymbol{\Sigma}$. For evaluating the other terms, we use the Stein-Haff or Konno identity and the Stein identity, respectively given by

$$\begin{aligned} E[\boldsymbol{W}\boldsymbol{G}_1] =& E[n\boldsymbol{\Sigma}\boldsymbol{G}_1 + \boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\nabla}^T)^T\boldsymbol{G}_1], && \text{(A.3)} \\ E[\boldsymbol{X}\boldsymbol{G}_2] =& E[\boldsymbol{\Sigma}\boldsymbol{\nabla}\boldsymbol{G}_2], && \text{(A.4)} \end{aligned}$$

where $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ are, respectively, $p \times p$ and $n \times p$ matrices of functions of $\boldsymbol{W}$, and $\boldsymbol{\nabla} = (\partial/\partial x_{ij})$ is the $p \times n$ matrix of differential operators. See Konno (2009) and Stein (1973, 1981) for (A.3) and (A.4), respectively. It follows from (A.3) that

$$E[\mathrm{tr}\,(\boldsymbol{W}\boldsymbol{G}_1)] = E[n\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{G}_1) + \mathrm{tr}\,(\boldsymbol{X}\boldsymbol{\nabla}^T(\boldsymbol{\Sigma}\boldsymbol{G}_1^T))]. \tag{A.5}$$

To carry out the calculus $\boldsymbol{\nabla}^T(\boldsymbol{\Sigma}\boldsymbol{G}_1^T)$, the following equalities due to Haff (1979, 82) are useful:

$$\begin{aligned} \boldsymbol{\nabla}(\boldsymbol{U}\boldsymbol{V}) =& (\boldsymbol{\nabla}\boldsymbol{U})\boldsymbol{V} + (\boldsymbol{U}^T\boldsymbol{\nabla}^T)^T\boldsymbol{V}, \\ \mathrm{tr}\,[\boldsymbol{\nabla}(\boldsymbol{U}\boldsymbol{V})] =& \mathrm{tr}\,[(\boldsymbol{\nabla}\boldsymbol{U})\boldsymbol{V}] + \mathrm{tr}\,[\boldsymbol{U}^T(\boldsymbol{\nabla}^T\boldsymbol{V}^T)], \\ \boldsymbol{\nabla}^T(\boldsymbol{A}_1\boldsymbol{X}) =& (\mathrm{tr}\,\boldsymbol{A}_1)\boldsymbol{I}_n,\ \boldsymbol{\nabla}(\boldsymbol{A}_2\boldsymbol{X}) = \boldsymbol{A}_2^T. \end{aligned} \tag{A.6}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are matrices of functions of $\boldsymbol{W}$ such that the product $\boldsymbol{\nabla}\boldsymbol{U}\boldsymbol{V}$ is defined, and $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are $p \times p$ and $n \times p$ matrices, respectively, of constants.

For $E[\boldsymbol{W}^2]$, it is seen that $E[\boldsymbol{W}^2] = E[n\boldsymbol{\Sigma}\boldsymbol{W} + \boldsymbol{\Sigma}(\boldsymbol{X}\boldsymbol{\nabla}^T)^T\boldsymbol{W}]$. It can be shown that $(\boldsymbol{X}\boldsymbol{\nabla}^T)^T\boldsymbol{W} = \boldsymbol{W} + (\mathrm{tr}\,\boldsymbol{W})\boldsymbol{I}$, so that

$$E[\boldsymbol{W}^2] = E[n\boldsymbol{\Sigma}\boldsymbol{W} + \boldsymbol{\Sigma}\boldsymbol{W} + (\mathrm{tr}\,\boldsymbol{W})\boldsymbol{\Sigma}] = n(n+1)\boldsymbol{\Sigma}^2 + n(\mathrm{tr}\,\boldsymbol{\Sigma})\boldsymbol{\Sigma}.$$

This implies that $E[\mathrm{tr}\,\boldsymbol{A}\boldsymbol{W}^2] = n(n+1)(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{\Sigma}^2) + n(\mathrm{tr}\,\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{\Sigma})$.

Using (A.5), one gets

$$E[(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{B}\boldsymbol{W})] = E[n(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{B}\boldsymbol{W}) + \mathrm{tr}\,(\boldsymbol{X}\boldsymbol{\nabla}^T(\boldsymbol{\Sigma}\boldsymbol{A}(\mathrm{tr}\,\boldsymbol{B}\boldsymbol{W})))].$$

Since $\boldsymbol{\nabla}^T\mathrm{tr}\,(\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}^T) = 2\boldsymbol{X}^T\boldsymbol{B}$, we have

$$\begin{aligned} E[(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{B}\boldsymbol{W})] =& E[n(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{B}\boldsymbol{W}) + 2\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{W}\boldsymbol{B})] \\ =& n^2(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{B}\boldsymbol{\Sigma}) + 2n\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma}). \end{aligned}$$

The calculation of $E[\mathrm{tr}\,(\boldsymbol{W}^2)^2]$ is not easy to show. The skech of the proof is given as follows: It follows from (A.5) that

$$E[(\mathrm{tr}\,\boldsymbol{W}^2)^2] = E[n(\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{W}^2) + \mathrm{tr}\,\boldsymbol{X}\boldsymbol{\nabla}^T(\boldsymbol{\Sigma}\boldsymbol{W}(\mathrm{tr}\,\boldsymbol{W}^2))].$$

It can be here demonstrated that

$$\begin{aligned} \mathrm{tr}\,\boldsymbol{X}\boldsymbol{\nabla}^T(\boldsymbol{\Sigma}\boldsymbol{W}(\mathrm{tr}\,\boldsymbol{W}^2)) =& (\mathrm{tr}\,\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{W}^2) + \mathrm{tr}\,(\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{\Sigma}\boldsymbol{\nabla})^T\boldsymbol{X}^T(\mathrm{tr}\,\boldsymbol{W}^2)), \\ \mathrm{tr}\,(\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{\Sigma}\boldsymbol{\nabla})^T\boldsymbol{X}^T(\mathrm{tr}\,\boldsymbol{W}^2)) =& (\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{W}^2) + \mathrm{tr}\,(\boldsymbol{X}^T\boldsymbol{\Sigma}(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\nabla}^T)^T(\mathrm{tr}\,\boldsymbol{W}^2)), \\ \mathrm{tr}\,(\boldsymbol{X}^T\boldsymbol{\Sigma}(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\nabla}^T)^T(\mathrm{tr}\,\boldsymbol{W}^2)) =& 4\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{W}^3). \end{aligned}$$

Combining these terms gives the expression

$$E[(\mathrm{tr}\,\boldsymbol{W}^2)^2] = E[(n+1)(\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{W}^2) + (\mathrm{tr}\,\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{W}^2) + 4(\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W}^3)]. \quad \text{(A.7)}$$

Similarly, it can be seen that

$$\begin{aligned} E[(\mathrm{tr}\,\boldsymbol{W}^2)(\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W})] =& E[(n+1)(\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W})^2 + (\mathrm{tr}\,\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W}) + 2\mathrm{tr}\,\boldsymbol{\Sigma}^2\boldsymbol{W}^2], \\ E[(\mathrm{tr}\,\boldsymbol{W}^2)(\mathrm{tr}\,\boldsymbol{W})] =& E[(n+1)(\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{W}) + (\mathrm{tr}\,\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{W})^2 + 2\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W}^2], \\ E[\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W}^3] =& E[(n+2)\mathrm{tr}\,\boldsymbol{\Sigma}^2\boldsymbol{W}^2 + (\mathrm{tr}\,\boldsymbol{\Sigma}^2)(\mathrm{tr}\,\boldsymbol{W}^2) + (\mathrm{tr}\,\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{\Sigma}^2\boldsymbol{W})]. \end{aligned} \quad \text{(A.8)}$$

Combining (A.7) and (A.8), and using the second moments of $\boldsymbol{W}$, we can show the equality (A.1).

For $E[(\mathrm{tr}\,\boldsymbol{W})^4]$, the identity in (A.5) gives

$$E[(\mathrm{tr}\,\boldsymbol{W})^4] = E[n(\mathrm{tr}\,\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{W})^3 + \mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{X}\boldsymbol{\nabla}^T(\mathrm{tr}\,\boldsymbol{W})^3].$$

Since $\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{X}\boldsymbol{\nabla}^T(\mathrm{tr}\,\boldsymbol{W})^3 = 6(\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{W})^2$, we have

$$E[(\mathrm{tr}\,\boldsymbol{W})^4] = E[n(\mathrm{tr}\,\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{W})^3 + 6(\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{W})^2]. \quad \text{(A.9)}$$

Similarly, it can be demonstrated that

$$\begin{aligned} E[(\mathrm{tr}\,\boldsymbol{W})^3] =& E[n(\mathrm{tr}\,\boldsymbol{\Sigma})(\mathrm{tr}\,\boldsymbol{W})^2 + 4(\mathrm{tr}\,\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W})], \\ E[(\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{W})^2] =& E[n(\mathrm{tr}\,\boldsymbol{\Sigma}^2)(\mathrm{tr}\,\boldsymbol{W})^2 + 4(\mathrm{tr}\,\boldsymbol{\Sigma}^2\boldsymbol{W})(\mathrm{tr}\,\boldsymbol{W})]. \end{aligned} \quad \text{(A.10)}$$

Hence, combining (A.9) and (A.10), and using the second moments of $\boldsymbol{W}$, we can show the equality in (A.2). □

Lemma A.2. *For the random matrix $\boldsymbol{V}$ given in* (2.1)*, the following properties hold for any positive integers n, p and q with $n \geq q$:*

(1) *$\boldsymbol{V}_{11.2} = \boldsymbol{V}_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}$ is expressed as $\boldsymbol{V}_{11.2} = \boldsymbol{U}\boldsymbol{U}^T$ for $\boldsymbol{U} = (\boldsymbol{U}_1, \ldots, \boldsymbol{U}_{n-q})$, where $\boldsymbol{U}_i$'s are mutually independently and identicall distributed as $\boldsymbol{U}_i \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{11.2})$.*

(2) *$\boldsymbol{V}_{11.2}$ is independent of $(\boldsymbol{V}_{12}, \boldsymbol{V}_{22})$.*

(3) *$\boldsymbol{V}_{12}|\boldsymbol{V}_{22} \sim \mathcal{N}_{p,q}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{V}_{22}, \boldsymbol{\Sigma}_{11.2}, \boldsymbol{V}_{22})$ and $\boldsymbol{V}_{22} \sim \mathcal{W}_q(n, \boldsymbol{\Sigma}_{22})$.*

Proof. We first note that these exist $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$ and $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ such that $\boldsymbol{V}_{11} = \boldsymbol{Y}\boldsymbol{Y}^T$, $\boldsymbol{V}_{12} = \boldsymbol{Y}\boldsymbol{X}^T$, $\boldsymbol{V}_{22} = \boldsymbol{X}\boldsymbol{X}^T$, and $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$'s are mutually independently and identically distributed as

$$\begin{pmatrix} \boldsymbol{Y}_i \\ \boldsymbol{X}_i \end{pmatrix} \sim \mathcal{N}_{p+q}\Big(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\Big).$$

Let $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n) = \boldsymbol{Y} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}$. It is seen that $\boldsymbol{Z}$ is independent of $\boldsymbol{X}$ and $\boldsymbol{Z}_i$'s are mutually independently and identically distributed as $\boldsymbol{Z}_i \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{11.2})$. Then, $\boldsymbol{V}_{11}$ and $\boldsymbol{V}_{12}$ are rewritten as

$$\begin{aligned} \boldsymbol{V}_{11} =& (\boldsymbol{Z} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X})(\boldsymbol{Z}^T + \boldsymbol{X}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}) \\ =& \boldsymbol{Z}\boldsymbol{Z}^T + \boldsymbol{Z}\boldsymbol{X}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}\boldsymbol{Z}^T + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}, \\ \boldsymbol{V}_{12} =& (\boldsymbol{Z} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X})\boldsymbol{X}^T = \boldsymbol{Z}\boldsymbol{X}^T + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}\boldsymbol{X}^T. \end{aligned} \quad \text{(A.11)}$$

Also, $\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}$ is expressed as

$$\begin{aligned}\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21} =&(\boldsymbol{Z}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X})\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}(\boldsymbol{Z}^T+\boldsymbol{X}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})\\ =&\boldsymbol{Z}\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\boldsymbol{Z}^T+\boldsymbol{Z}\boldsymbol{X}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}\boldsymbol{Z}^T\\ &+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.\end{aligned}$$

Thus, one gets

$$\boldsymbol{V}_{11.2}=\boldsymbol{Z}\{\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\}\boldsymbol{Z}^T. \tag{A.12}$$

Since $\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}$ is idempotent and of rank $n-q$, it can be seen that there exists a $p\times(n-q)$ random matrix $\boldsymbol{U}$ such that $\boldsymbol{V}_{11.2}=\boldsymbol{U}\boldsymbol{U}^T$ and $\boldsymbol{U}\sim\mathcal{N}_{p,n-q}(\boldsymbol{0},\boldsymbol{\Sigma}_{11.2},\boldsymbol{I}_{n-q})$. This shows part (1) of Lemma A.2.

For part (2), note that $\boldsymbol{X}\boldsymbol{X}^T$ is complete and sufficient for $\boldsymbol{\Sigma}_{22}$ and $\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}$ is ancillary, since $\boldsymbol{X}\boldsymbol{X}^T\sim\mathcal{W}_q(n,\boldsymbol{\Sigma}_{22})$. It follows from Basu's theorem that $\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}$ is independent of $\boldsymbol{X}\boldsymbol{X}^T$. Recall expression (A.12). Then, it is easy to see that $\boldsymbol{V}_{11.2}$ is independent of $\boldsymbol{V}_{22}=\boldsymbol{X}\boldsymbol{X}^T$.

To check the independence between $\boldsymbol{V}_{11.2}$ and $\boldsymbol{V}_{12}$, from (A.11) and (A.12), it is sufficient to show that $\boldsymbol{Z}\boldsymbol{X}^T$ is independent of $\boldsymbol{Z}\{\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\}$. Since the two are conditionally mutually independent given $\boldsymbol{X}$, it is seen that for measurable sets $A\subset\mathbb{R}^{p\times q}$ and $B\subset\mathbb{R}^{p\times n}$,

$$\begin{aligned}&P(\boldsymbol{Z}\boldsymbol{X}^T\in A,\boldsymbol{Z}\{\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\}\in B)\\ &=E[P(\boldsymbol{Z}\boldsymbol{X}^T\in A,\boldsymbol{Z}\{\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\}\in B|\boldsymbol{X})]\\ &=E[P(\boldsymbol{Z}\boldsymbol{X}^T\in A|\boldsymbol{X})P(\boldsymbol{Z}\{\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\}\in B|\boldsymbol{X})].\end{aligned}$$

Since $\boldsymbol{Z}\sim\mathcal{N}_{p,n}(\boldsymbol{0},\boldsymbol{\Sigma}_{11.2},\boldsymbol{I}_n)$, one can see that $\boldsymbol{Z}\boldsymbol{X}^T|\boldsymbol{X}\sim\mathcal{N}_{p,q}(\boldsymbol{0},\boldsymbol{\Sigma}_{11.2},\boldsymbol{X}\boldsymbol{X}^T)$ conditionally, namely, the conditional distribution of $\boldsymbol{Z}\boldsymbol{X}^T$ depends on $\boldsymbol{X}$ through $\boldsymbol{X}\boldsymbol{X}^T$. Using the fact that $\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}$ is independent of $\boldsymbol{X}\boldsymbol{X}^T$ again, one can see that

$$\begin{aligned}&E[P(\boldsymbol{Z}\boldsymbol{X}^T\in A|\boldsymbol{X})P(\boldsymbol{Z}\{\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\}\in B|\boldsymbol{X})]\\ &\quad=E[P(\boldsymbol{Z}\boldsymbol{X}^T\in A|\boldsymbol{X})]E[P(\boldsymbol{Z}\{\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\}\in B|\boldsymbol{X})],\end{aligned}$$

which equal to $P(\boldsymbol{Z}\boldsymbol{X}^T\in A)P(\boldsymbol{Z}\{\boldsymbol{I}-\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\}\in B)$.

For part (3), it follows that $\boldsymbol{V}_{12}|\boldsymbol{X}\sim\mathcal{N}_{p,q}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}\boldsymbol{X}^T,\boldsymbol{\Sigma}_{11.2},\boldsymbol{X}\boldsymbol{X}^T)$, since $\boldsymbol{V}_{12}=\boldsymbol{Z}\boldsymbol{X}^T+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}\boldsymbol{X}^T$. Since this conditional distribution depends on $\boldsymbol{X}$ through $\boldsymbol{V}_{22}=\boldsymbol{X}\boldsymbol{X}^T$, the conditional distribution of $\boldsymbol{V}_{12}$ given $\boldsymbol{V}_{22}$ has the same distribution. Therefore, the proof of Lemma A.2 is complete. □

Proof of Theorem 2.1. In this proof, we use the notation $m=n-q$ for simplicity. We only consider the case of sphericity since the case of diagonalization is similar.

From Lemmas A.1 and A.2, it follows that $E[\widehat{\boldsymbol{\Sigma}}_{11}]=\boldsymbol{\Sigma}_{11}$, $E[\widehat{\boldsymbol{\Sigma}}_{11}^2]=n^{-1}(n+1)\boldsymbol{\Sigma}_{11}^2+n^{-1}(\mathrm{tr}\,[\boldsymbol{\Sigma}_{11}])\boldsymbol{\Sigma}_{11}$, $E[(\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}_{11})^2]=2n^{-1}\mathrm{tr}\,\boldsymbol{\Sigma}_{11}^2+(\mathrm{tr}\,\boldsymbol{\Sigma}_{11})^2$, $E[\widehat{\boldsymbol{\Sigma}}_{11.2}]=n^{-1}m\boldsymbol{\Sigma}_{11.2}$, $E[\widehat{\boldsymbol{\Sigma}}_{11.2}^2]=$

$n^{-2}m(m+1)\mathbf{\Sigma}_{11.2}^2+n^{-2}m(\mathrm{tr}\,[\mathbf{\Sigma}_{11.2}])\mathbf{\Sigma}_{11.2}$ and $E[(\mathrm{tr}\,\widehat{\mathbf{\Sigma}}_{11.2})^2] = 2n^{-2}m\mathrm{tr}\,\mathbf{\Sigma}_{11.2}^2+n^{-2}m^2(\mathrm{tr}\,\mathbf{\Sigma}_{11.2})^2$. Thus, it is observed that

$$\begin{aligned}
p^{-1}E[\mathrm{tr}\,[\widehat{\mathbf{\Sigma}}_{11}]] =& a_1,\\
p^{-1}E[\mathrm{tr}\,[\widehat{\mathbf{\Sigma}}_{11}^2]] =& (1+1/n)a_2 + (p/n)a_1^2,\\
E[(p^{-1}\mathrm{tr}\,[\widehat{\mathbf{\Sigma}}_{11}])^2] =& a_1^2 + O(n^{-1}),\\
p^{-1}E[\mathrm{tr}\,[\widehat{\mathbf{\Sigma}}_{11.2}]] =& b_1 + O(n^{-1}),\\
p^{-1}E[\mathrm{tr}\,[\widehat{\mathbf{\Sigma}}_{11.2}^2]] =& b_2 + (p/n)b_1^2 + O(n^{-1}) + O(pn^{-2}),\\
E[(p^{-1}\mathrm{tr}\,[\widehat{\mathbf{\Sigma}}_{11.2}])^2] =& b_1^2 + O(n^{-1}).
\end{aligned} \tag{A.13}$$

These evaluations are used to approximate the terms R_0, J_{11}, J_{12}, J_{22}, J_{10} and J_{20} given in Lemma 2.1. For R_0 and J_{11},

$$\begin{aligned}
R_0 =& \frac{1}{p}E\mathrm{tr}\,(\mathbf{\Sigma}_{11} - \widehat{\mathbf{\Lambda}}(\widehat{\mathbf{\Sigma}}_{11}))^2 = a_2 - 2a_1^2 + E[(p^{-1}\mathrm{tr}\,[\widehat{\mathbf{\Sigma}}_{11}])^2] = a_2 - a_1^2 + O(n^{-1}).\\
J_{11} =& \frac{1}{p}E\mathrm{tr}\,(\widehat{\mathbf{\Sigma}}_{11.2} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2}))^2 = p^{-1}E[\mathrm{tr}\,[\widehat{\mathbf{\Sigma}}_{11.2}^2]] - E[(p^{-1}\mathrm{tr}\,[\widehat{\mathbf{\Sigma}}_{11.2}])^2]\\
=& b_2 - b_1^2 + (p/n)b_1^2 + O(n^{-1}) + O(pn^{-2}).
\end{aligned}$$

Since $\widehat{\mathbf{\Sigma}}_{11.2}$ and $\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21}$ are independent,

$$\begin{aligned}
J_{12} =& \frac{1}{p}E\mathrm{tr}\,(\widehat{\mathbf{\Sigma}}_{11.2} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2}))(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21}))\\
=& \frac{1}{p}\mathrm{tr}\,E(\widehat{\mathbf{\Sigma}}_{11.2} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2}))E(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21}))\\
=& \frac{m}{pn}\mathrm{tr}\,\Big(\mathbf{\Sigma}_{11.2} - \mathbf{\Lambda}(\mathbf{\Sigma}_{11.2})\Big)\Big((\mathbf{\Sigma}_{11} - \mathbf{\Lambda}(\mathbf{\Sigma}_{11})) - \frac{n}{m}\big(\mathbf{\Sigma}_{11.2} - \mathbf{\Lambda}(\mathbf{\Sigma}_{11.2})\big)\Big)\\
=& \frac{m}{n}\Big(\phi_{11} - a_1b_1\Big) - \Big(\frac{m}{n}\Big)^2(b_2 - b_1^2)\\
=& \phi_{11} - a_1b_1 - b_2 + b_1^2 + O(n^{-1}).
\end{aligned}$$

Similarly, it is seen that

$$\begin{aligned}
J_{22} =& \frac{1}{p} E\mathrm{tr}\,((\widehat{\boldsymbol{\Sigma}}_{11} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11})) - (\widehat{\boldsymbol{\Sigma}}_{11.2} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2})))^2 \\
=& \frac{1}{p} E\mathrm{tr}\,((\widehat{\boldsymbol{\Sigma}}_{11} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11}))^2 - \frac{2}{p}\mathrm{tr}\,((\widehat{\boldsymbol{\Sigma}}_{11} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11}))(\widehat{\boldsymbol{\Sigma}}_{11.2} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2}))) + J_{11} \\
=& \frac{1}{p} E\mathrm{tr}\,((\widehat{\boldsymbol{\Sigma}}_{11} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11}))^2 - 2(J_{12} + J_{11}) + J_{11} \\
=& \frac{1}{p} E\mathrm{tr}\,((\widehat{\boldsymbol{\Sigma}}_{11} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11}))^2 - 2J_{12} - J_{11} \\
=& ((1+n^{-1})a_2 + (p/n)a_1^2 - a_1^2 + O(n^{-1})) - 2(\phi_{11} - a_1 b_1 - b_2 + b_1^2 + O(n^{-1})) \\
& - (b_2 - b_1^2 + (p/n)b_1^2 + O(n^{-1}) + O(pn^{-2})) \\
=& (1+n^{-1})a_2 - 2\phi_{11} + b_2 + \frac{p}{n}(a_1^2 - b_1^2) - (a_1 - b_1)^2 + O(n^{-1}) + O(n^{-2}p),
\end{aligned}$$

and

$$\begin{aligned}
J_{10} =& \frac{1}{p} E\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11.2} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2}))(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{11})) = \frac{1}{p}\mathrm{tr}\,(\frac{m}{n}(\boldsymbol{\Sigma}_{11.2} - \boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{11.2})))(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{11})) \\
=& \frac{m}{n}(\phi_{11} - a_1 b_1) = \phi_{11} - a_1 b_1 + O(n^{-1}).
\end{aligned}$$

Finally, it is observed that

$$\begin{aligned}
J_{20} =& \frac{1}{p} E\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}))(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{11})) \\
=& \frac{1}{p} E\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11}))(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{11})) - \frac{1}{p} E\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11.2} - \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2}))(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{11})) \\
=& \frac{1}{p}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{11}))^2 - \frac{m}{pn}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2} - \boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{11.2}))(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{11})) \\
=& a_2 - a_1^2 - \frac{m}{n}(\phi_{11} - a_1 b_1) = a_2 - a_1^2 - \phi_{11} + a_1 b_1 + O(n^{-1}),
\end{aligned}$$

which proves Theorem 2.1. □

A.2 Proof of Theorem 3.1

In this proof, we use the notation $m = n - q$ for simplicity. Since the unbiasedness of the estimators is easy to show, we only prove their consistency. For this purpose, we shall evaluate the variances of the estimators.

For $\hat{a}_1$ and $\hat{a}_2$, Srivastava (2005) shows that $\mathrm{Var}(\hat{a}_1) = 2a_2/(np)$ and

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

so that $\hat{a}_1 = a_1 + O_p(n^{-1/2})$ and $\hat{a}_2 = a_2 + O_p(n^{-1/2}p)$, since $a_1 = O(1)$ but $a_2 = O(p)$ and $a_4 = O(p^3)$. Similarly, Srivastava (2005) with Lemmas A.1 and A.2 imply that $\mathrm{Var}(\hat{b}_1) = 2b_2/(np)$ and

$$\mathrm{Var}(\hat{b}_2) = \frac{8(m+2)(m+3)(m-1)^2}{pn^5}b_4 + \frac{4(m+2)(m-1)}{n^4}\left(b_2^2 - \frac{b_4}{p}\right),$$

so that $\hat{b}_1 = b_1 + O_p((np)^{-1/2})$ and $\hat{b}_2 = b_2 + O_p(n^{-1}) + O_p((np)^{-1/2})$, since $b_1 = O(1)$, $b_2 = O(1)$ and $b_4 = O(1)$.

For $\hat{a}_{20}$,

$$\hat{a}_{20} - a_{20} = \frac{n}{p(n+2)}\left(\mathrm{tr}\,(\mathrm{diag}\,\widehat{\boldsymbol{\Sigma}}_{11})^2 - \frac{n+2}{n}\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{11})^2\right),$$

so that

$$\mathrm{Var}(\hat{a}_{20}) = \frac{n^2}{p^2(n+2)^2}\left((\mathrm{tr}\,(\mathrm{diag}\,\widehat{\boldsymbol{\Sigma}}_{11})^2)^2 - \frac{(n+2)^2}{n^2}(\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{11})^2)^2\right).$$

We denote $v_{ii} = n(\hat{\sigma}_{11})_{ii}/(\sigma_{11})_{ii}$ for $\widehat{\boldsymbol{\Sigma}}_{11} = ((\hat{\sigma}_{11})_{ij})$ and $\boldsymbol{\Sigma}_{11} = ((\sigma_{11})_{ij})$. Then one gets

$$n^4(\mathrm{tr}\,(\mathrm{diag}\,\widehat{\boldsymbol{\Sigma}}_{11})^2)^2 = (\sum_{i=1}^{p}(\sigma_{11})_{ii}^2 v_{ii}^2)^2 = \sum_{i=1}^{p}(\sigma_{11})_{ii}^4 v_{ii}^4 + \sum_{i\neq j}(\sigma_{11})_{ii}^2(\sigma_{11})_{jj}^2 v_{ii}^2 v_{jj}^2.$$

Moments of the Wishart distribution imply that

$$E[(\mathrm{tr}\,(\mathrm{diag}\,\widehat{\boldsymbol{\Sigma}}_{11})^2)^2] = n^{-3}(n+2)(n+4)(n+6)\sum_{i=1}^{p}(\sigma_{11})_{ii}^4 + n^{-2}(n+2)^2\sum_{i\neq j}(\sigma_{11})_{ii}^2(\sigma_{11})_{jj}^2.$$

From condition (A6) and the fact that $\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{11})^2 = \sum_{i=1}^{p}(\sigma_{11})_{ii}^4 + \sum_{i\neq j}(\sigma_{11})_{ii}^2(\sigma_{11})_{jj}^2$, it follows that

$$\mathrm{Var}(\hat{a}_{20}) = \frac{n^2}{p^2(n+2)^2}\left(n^{-3}(n+2)(n+4)(n+6)\sum_{i=1}^{p}(\sigma_{11})_{ii}^4 - \sum_{i=1}^{p}(\sigma_{11})_{ii}^4\right) = O(n^{-1}p^{-1}).$$

Therefore, $\hat{a}_{20} = a_{20} + O_p(n^{-1/2}p^{-1/2})$. Similarly, one easily gets $\hat{b}_{20} = b_{20} + O_p(n^{-1/2}p^{-1/2})$.

For $\widehat{\phi}_{11}$, one writes it as

$$\begin{aligned}\widehat{\phi}_{11} - \phi_{11} =& \frac{1}{pnm(m-1)(m+2)}\Big(m^2\Big(\mathrm{tr}\,\boldsymbol{V}_{11.2}^2 - m((m+1)\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^2 + (\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2})^2)\Big)\\ &- m\Big((\mathrm{tr}\,\boldsymbol{V}_{11.2})^2 - m(m(\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2})^2 + 2\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^2)\Big)\\ &+ (m-1)(m+2)\Big(\mathrm{tr}\,\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}\boldsymbol{V}_{11.2} - (nm\mathrm{tr}\,\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{11.2} - m^2\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^2)\Big)\Big)\\ =:& \frac{1}{pnm(m-1)(m+2)}\Big(m^2 I_1 - mI_2 + (m-1)(m+2)I_3\Big)),\end{aligned} \tag{A.14}$$

where $E[I_1] = E[I_2] = E[I_3] = 0$. Hence, it is sufficient to evaluate variances of I_1, I_2 and I_3, since the cross product terms are bounded by Cauchy-Schwartz's inequality. For I_1, one sees that

$$E(I_1^2) = E[(\mathrm{tr}\, \boldsymbol{V}_{11.2}^2)^2] - m^2((m+1)\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^2 + (\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2})^2)^2.$$

Since from Lemma A.1,

$$\begin{aligned} E[(\mathrm{tr}\, \boldsymbol{V}_{11.2}^2)^2] =& 4m^2(2m^2+5m+5)\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^4 + 16m(m+1)\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^3 \mathrm{tr}\, \boldsymbol{\Sigma}_{11.2} \\ &+ m(m^3+2m^2+5m+4)(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^2)^2 \\ &+ 2m(m^2+m+4)(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^2)(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2})^2 + m^2(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2})^4, \end{aligned}$$

we have

$$\begin{aligned} E(I_1^2) =& 4m(2m^2+5m+5)\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^4 + 16m(m+1)\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^3 \mathrm{tr}\, \boldsymbol{\Sigma}_{11.2} \\ &+ 4m(m+1)(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^2)^2 + 8m\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^2(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2})^2. \end{aligned}$$

Hence, from conditions (A2), (A3) and (A5),

$$I_1 = O_p(n^{3/2}p^{1/2}) + O_p(np) + O_p(n^{1/2}p^{3/2}). \tag{A.15}$$

Also, for I_2, one sees that

$$E(I_2^2) = E[(\mathrm{tr}\, \boldsymbol{V}_{11.2})^4] - m^2(m(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2})^2 + 2\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^2)^2.$$

From Lemma A.1,

$$\begin{aligned} E[(\mathrm{tr}\, \boldsymbol{V}_{11.2})^4] =& 48m\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^4 + 32m^2\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^3 \mathrm{tr}\, \boldsymbol{\Sigma}_{11.2} \\ &+ 12m^2(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^2)^2 + 12m^3(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^2)(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2})^2 + m^4(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2})^4, \end{aligned}$$

so that

$$E(I_2^2) = 48m\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^4 + 32m^2\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^3 \mathrm{tr}\, \boldsymbol{\Sigma}_{11.2} + 8m^2(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^2)^2 + 8m^3(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2}^2)(\mathrm{tr}\, \boldsymbol{\Sigma}_{11.2})^2.$$

Therefore,

$$I_2 = O_p(n^{1/2}p^{1/2}) + O_P(np) + O_p(n^{3/2}p^{3/2}). \tag{A.16}$$

Finally, for I_3 it is noted that

$$\begin{aligned} &\mathrm{tr}\, \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}\boldsymbol{V}_{11.2} = \mathrm{tr}\, \boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{A} \\ =&\mathrm{tr}\, (\boldsymbol{Y}-\boldsymbol{\theta}+\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta}+\boldsymbol{\theta})^T\boldsymbol{A} \\ =&\mathrm{tr}\, (\boldsymbol{Y}-\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{A} + \mathrm{tr}\, \boldsymbol{\theta}\boldsymbol{\theta}^T\boldsymbol{A} + 2\mathrm{tr}\, (\boldsymbol{Y}-\boldsymbol{\theta})\boldsymbol{\theta}^T\boldsymbol{A}, \end{aligned}$$

where $\boldsymbol{Y} = \boldsymbol{\Sigma}_{11.2}^{-1/2}\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1/2}$, $\boldsymbol{\theta} = \boldsymbol{\Sigma}_{11.2}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}\boldsymbol{V}_{22}^{1/2}$ and $\boldsymbol{A} = \boldsymbol{\Sigma}_{11.2}^{1/2}\boldsymbol{V}_{11.2}\boldsymbol{\Sigma}_{11.2}^{1/2}$. Then, I_3 is decomposed as

$$\begin{aligned} I_3 =&(\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{A} - qm\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^2) \\ &+ (\mathrm{tr}\,\boldsymbol{\theta}\boldsymbol{\theta}^T\boldsymbol{A} - nm\mathrm{tr}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11.2}) + 2\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})\boldsymbol{\theta}^T\boldsymbol{A} \\ =&: I_{3,1} + I_{3,2} + 2I_{3,3}, \end{aligned}$$

where $E[I_{3,1}] = E[I_{3,2}] = E[I_{3,3}] = 0$. Noting that $\boldsymbol{Y}|\boldsymbol{V}_{22} \sim \mathcal{N}_{p,q}(\boldsymbol{0}, \boldsymbol{I}_p, \boldsymbol{I}_q)$, one sees that from Stein's lemma,

$$\begin{aligned} &E[(\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{A})^2] \\ &=E\mathrm{tr}\,[\boldsymbol{\nabla}_{\boldsymbol{Y}}((\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{A})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{A})] \\ &=qE[\mathrm{tr}\,\boldsymbol{A}\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{A}] + E[(\boldsymbol{\nabla}_{\boldsymbol{Y}}^T\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{A})\mathrm{tr}\,\boldsymbol{A}^T(\boldsymbol{Y}-\boldsymbol{\theta})] \\ &=q^2E(\mathrm{tr}\,\boldsymbol{A})^2 + 2E\mathrm{tr}\,\boldsymbol{A}^T(\boldsymbol{Y}-\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{A} \\ &=q^2E(\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}\boldsymbol{V}_{11.2})^2 + 2qE\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}\boldsymbol{V}_{11.2})^2 \\ &=q^2(2m\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^4 + m^2(\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^2)^2) + 2q(m(m+1)\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^4 + m(\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^2)^2) \\ &=qm(2+qm)(\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^2)^2 + 2qm(n+1)\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^4, \end{aligned}$$

so that by (A5),

$$\begin{aligned} E[I_{3,1}^2] =&E[(\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{A})^2] - q^2m^2(\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^2)^2 \\ =&2qm(\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^2)^2 + 2qm(n+1)\mathrm{tr}\,\boldsymbol{\Sigma}_{11.2}^4 \\ =&O(np^2) + O(n^2p). \end{aligned}$$

Hence, one gets

$$I_{3,1} = O_p(n^{1/2}p) + O_p(np^{1/2}) \tag{A.17}$$

For $I_{3,2}$, one sees that

$$\begin{aligned} E[I_{3,2}^2] =&E[(\mathrm{tr}\,\boldsymbol{\theta}\boldsymbol{\theta}^T\boldsymbol{A})^2] - n^2m^2(\mathrm{tr}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11.2})^2 \\ =&2nm(\mathrm{tr}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11.2})^2 + 2nm(2m+1)\mathrm{tr}\,(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11.2})^2 \\ =&O(n^2p) + O(n^3p^2), \end{aligned}$$

which implies that

$$I_{3,2} = O_p(np^{1/2}) + O_p(n^{3/2}p). \tag{A.18}$$

For $I_{3,3}$, using Stein's lemma again, one sees that

$$\begin{aligned}
E[I_{3,3}^2] =& E[(\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})\boldsymbol{\theta}^T\boldsymbol{A})^2] \\
=& E\mathrm{tr}\,[\boldsymbol{\nabla}_{\boldsymbol{Y}}^T\boldsymbol{A}\boldsymbol{\theta}(\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})\boldsymbol{\theta}^T\boldsymbol{A})] \\
=& E\mathrm{tr}\,[\boldsymbol{\theta}^T\boldsymbol{A}\boldsymbol{\nabla}_{\boldsymbol{Y}}(\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})\boldsymbol{\theta}^T\boldsymbol{A})] \\
=& E\mathrm{tr}\,[\boldsymbol{\theta}^T\boldsymbol{A}^2\boldsymbol{\theta}] \\
=& E\mathrm{tr}\,[\boldsymbol{V}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{V}_{11.2}^2\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}] \\
=& nm(m+1)\mathrm{tr}\,(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11.2}^2) + nm\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2})\mathrm{tr}\,(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11.2}) \\
=& O(n^3p) + O(n^2p^2),
\end{aligned}$$

which yields

$$I_{3,3} = O_p(n^{3/2}p^{1/2}) + O_p(np). \tag{A.19}$$

Thus, (A.14), (A.15), (A.16), (A.17), (A.18) and (A.19) lead to

$$\widehat{\phi}_{11} - \phi_{11} = O_p(n^{-1/2}).$$

Finally,

$$\widehat{\phi}_{110} - \phi_{110} =: \frac{p}{nm(m+2)}((m+2)I_4 + mI_5), \tag{A.20}$$

where

$$\begin{aligned}
I_4 &= \mathrm{tr}\,\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{B} - qm\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})^2 - nm\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}), \\
I_5 &= \mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{V}_{11.2})^2 - m(m+2)\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})^2,
\end{aligned}$$

for $\boldsymbol{Y} = \boldsymbol{\Sigma}_{11.2}^{-1/2}\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1/2}$, $\boldsymbol{\theta} = \boldsymbol{\Sigma}_{11.2}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}\boldsymbol{V}_{22}^{1/2}$ and $\boldsymbol{B} = \boldsymbol{\Sigma}_{11.2}^{1/2}\mathrm{diag}\,(\boldsymbol{V}_{11.2})\boldsymbol{\Sigma}_{11.2}^{1/2}$. Note that $E(I_4) = E(I_5) = 0$. One decomposes I_4 into the three terms as

$$I_4 =: I_{4,1} + I_{4,2} + 2I_{4,3}, \tag{A.21}$$

where

$$\begin{aligned}
I_{4,1} =& \mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{B} - qm\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})^2, \\
I_{4,2} =& \mathrm{tr}\,\boldsymbol{\theta}\boldsymbol{\theta}^T\boldsymbol{B} - nm\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}), \\
I_{4,3} =& \mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})\boldsymbol{\theta}^T\boldsymbol{B}.
\end{aligned}$$

Observe that $I_{4,1}$, $I_{4,2}$ and $I_{4,3}$ are centered. Then, the following lemma is useful for calculation:

Lemma A.3. *Let $\boldsymbol{G} = (g_{ij})$ and $\boldsymbol{H} = (h_{ij})$ be $p \times p$ constant matrices. Then,*

$$E\mathrm{tr}\,\boldsymbol{G}(\mathrm{diag}\,\boldsymbol{V}_{11.2})\boldsymbol{H}(\mathrm{diag}\,\boldsymbol{V}_{11.2}) = 2m\sum_{i,j=1}^{p} g_{ij}h_{ji}(\boldsymbol{\Sigma}_{11.2})_{ij}^2 + m^2\mathrm{tr}\,\boldsymbol{G}(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})\boldsymbol{H}(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}).$$

Proof. Direct calculation shows that

$$\begin{aligned}
E\mathrm{tr}\,\boldsymbol{G}(\mathrm{diag}\,\boldsymbol{V}_{11.2})\boldsymbol{H}(\mathrm{diag}\,\boldsymbol{V}_{11.2}) &= E\sum_{i,j=1}^{p} g_{ij}h_{ji}(\boldsymbol{V}_{11.2})_{ii}(\boldsymbol{V}_{11.2})_{jj} \\
&= m(m+2)\sum_{i=1}^{p} g_{ii}h_{ii}(\boldsymbol{\Sigma}_{11.2})_{ii}^2 + \sum_{i\neq j} g_{ij}h_{ji}(2m(\boldsymbol{\Sigma}_{11.2})_{ij}^2 + m^2(\boldsymbol{\Sigma}_{11.2})_{ii}(\boldsymbol{\Sigma}_{11.2})_{jj}) \\
&= 2m\sum_{i,j=1}^{p} g_{ij}h_{ji}(\boldsymbol{\Sigma}_{11.2})_{ij}^2 + m^2\sum_{i,j=1}^{p} g_{ij}h_{ji}(\boldsymbol{\Sigma}_{11.2})_{ii}(\boldsymbol{\Sigma}_{11.2})_{jj} \\
&= 2m\sum_{i,j=1}^{p} g_{ij}h_{ji}(\boldsymbol{\Sigma}_{11.2})_{ij}^2 + m^2\mathrm{tr}\,\boldsymbol{G}(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})\boldsymbol{H}(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}),
\end{aligned}$$

which shows Lemma A.3. □

Using Lemma A.3 and the Stein identity leads to

$$\begin{aligned}
E(\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{B})^2 &= q^2E(\mathrm{tr}\,\boldsymbol{B})^2 + 2q\mathrm{tr}\,\boldsymbol{B}^2 \\
&= q^2E(\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})\boldsymbol{\Sigma}_{11.2})^2 + 2qE\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}\mathrm{diag}\,\boldsymbol{V}_{11.2})^2 \\
&= q^2(2m\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})^4 + m^2(\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})^2)^2) \\
&\quad + 2q(2m\sum_{i,j=1}^{p}(\boldsymbol{\Sigma}_{11.2})_{ij}^4 + m^2\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}))^2),
\end{aligned}$$

so that, by condition (A5),

$$\begin{aligned}
\mathrm{Var}(I_{4,1}) &= E(\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})(\boldsymbol{Y}-\boldsymbol{\theta})^T\boldsymbol{B})^2 - q^2m^2(\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})^2)^2 \\
&= 2q^2m\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})^4 + 2q(2m\sum_{i,j=1}^{p}(\boldsymbol{\Sigma}_{11.2})_{ij}^4 + m^2\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}))^2) \\
&= O(n^2p),
\end{aligned}$$

which implies that

$$I_{4,1} = O_p(np^{1/2}). \tag{A.22}$$

Also,

$$\begin{aligned}
E(\mathrm{tr}\,\boldsymbol{\theta}\boldsymbol{\theta}^T\boldsymbol{B})^2 &= E(\mathrm{tr}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{V}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}(\mathrm{diag}\,\boldsymbol{V}_{11.2}))^2 \\
&= 2n\mathrm{tr}\,(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}(\mathrm{diag}\,\boldsymbol{V}_{11.2}))^2 + n^2(\mathrm{tr}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}(\mathrm{diag}\,\boldsymbol{V}_{11.2}))^2 \\
&=: I_{4,2,1} + I_{4,2,2}.
\end{aligned}$$

By Lemmas A.1 and A.3,

$$
\begin{aligned}
I_{4,2,1} =& 2n(2m\sum_{i,j=1}^{p}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})_{ij}^2(\boldsymbol{\Sigma}_{11.2})_{ij}^2 + m^2\mathrm{tr}\,(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}))^2),\\
I_{4,2,2} =& n^2(\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})\boldsymbol{V}_{11.2})^2\\
=& n^2(2m\mathrm{tr}\,((\mathrm{diag}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}))^2 + m^2(\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}))^2).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathrm{Var}(I_{4,2}) =& E(\mathrm{tr}\,\boldsymbol{\theta}\boldsymbol{\theta}^T\boldsymbol{B})^2 - n^2m^2(\mathrm{tr}\,(\mathrm{diag}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}))^2\\
=& 2n(2m\sum_{i,j=1}^{p}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})_{ij}^2(\boldsymbol{\Sigma}_{11.2})_{ij}^2 + m^2\mathrm{tr}\,(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}))^2)\\
&+ 2n^2m\mathrm{tr}\,((\mathrm{diag}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2}))^2,
\end{aligned}
$$

so that from conditions (A5) and (A6),

$$
I_{4,2} = O_p(n^{3/2}p). \tag{A.23}
$$

For $I_{4,3}$, using the Stein identity, we can see that

$$
\begin{aligned}
\mathrm{Var}(I_{4,3}) =& E(\mathrm{tr}\,(\boldsymbol{Y}-\boldsymbol{\theta})\boldsymbol{\theta}^T\boldsymbol{B})^2 = E\mathrm{tr}\,\boldsymbol{\theta}^T\boldsymbol{B}^2\boldsymbol{\theta}\\
=& nE\mathrm{tr}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}(\mathrm{diag}\,\boldsymbol{V}_{11.2})\boldsymbol{\Sigma}_{11.2}(\mathrm{diag}\,\boldsymbol{V}_{11.2})\\
=& 2nm\sum_{i,j=1}^{p}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})_{ij}(\boldsymbol{\Sigma}_{11.2})_{ij}^3\\
&+ nm^2\mathrm{tr}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})\boldsymbol{\Sigma}_{11.2}(\mathrm{diag}\,\boldsymbol{\Sigma}_{11.2})\\
=& O(n^2p),
\end{aligned}
$$

so that

$$
I_{4,3} = O_p(np^{1/2}). \tag{A.24}
$$

Finally, since $\hat{b}_{20} - b_{20} = O_p(n^{-1/2}p^{-1/2})$, one gets

$$
I_5 = O_p(n^{3/2}p^{1/2}). \tag{A.25}
$$

Thus, from (A.20), (A.21), (A.22), (A.23), (A.24), and (A.25),

$$
\widehat{\phi}_{110} = \phi_{110} + O_p(n^{-1/2}),
$$

which completes the proof of Theorem 3.1. □

References

[1] Bai, Z., Huixia, L., and Wing-Keung, W. (2009). Enhancement of the applicability of markowitz's portfolio optimization by utilizing random matrix theory. *Math. Finance*, **19**, 639-667.

[2] Bickel, P., and Levina, E. (2008a). Covariance regularization by thresholding. *Ann. Statist.*, **36**, 2577-2604.

[3] Bickel, P., and Levina, E. (2008b). Regularized estimation of large covariance matrices. *Ann. Statist.*, **36**, 199-227.

[4] Cai, T., and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.*, **106**, 672-684.

[5] Cai, T., and Zhou, H. (2010). Optimal rates of convergence for sparse covariance matrix estimation. *Manuscript. University of Pennsylvania.*

[6] Chamberlain, G., and Rothschild, M. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica*, **51**, 1305-1324.

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

[8] Fama, E., and French, K. (1993). Common risk factors in the returns on stocks and bonds. *J. Financial Economics* **33**, 3-56.

[9] Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics*, **147**, 186-197.

[10] Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor model. *Ann. Statist.*, **39**, 3320-3356.

[11] Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. Royal Statist. Soc.*, **75**, 603-680.

[12] 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.

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

[14] Haff, L.R. (1982). Solutions of the Euler-Lagrange equations for certain multivariate normal estimation problems. *Unpublished manuscript.*

[15] Lam, C., and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.*, **37**, 4254-4278.

[16] 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.

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

[18] Ren, Y., and Shimotsu, K. (2009). Improvement in finite sample properties of the Hansen-Jagannathan distance test. *J. Empirical Finance*, **16**, 483-506.

[19] Rothman, A., Levina, E., and Zhu, J. (2009). Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.*, **104**, 177-186.

[20] Schafer, J., and Strimmer, K. (2005). An empirical bayes approach to inferring large-scale a gene association networks. *Bioinformatics*, **21**, 754-764.

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

[22] Stein, C. (1973). Estimation of the mean of a multivariate normal distribution, In *Proc. Prague Symp. Asymptotic Statist.*, 345–381.

[23] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135–1151.

[24] Watamori, Y. (1990). On the moments of traces of Wishart and inverted Wishart matrices. *South African Statist. J.*, **24**, 153-176.