

# Estimation of the large covariance matrix with two-step monotone missing data

Masashi Hyodo<sup>1\*</sup>, Nobumichi Shutoh<sup>2</sup>,  
Takashi Seo<sup>1</sup>, and Tatjana Pavlenko<sup>3</sup>

<sup>1</sup> Department of Mathematical Information Science,  
Tokyo University of Science

<sup>2</sup> Department of Health Sciences,  
Oita University of Nursing and Health Sciences

<sup>3</sup> Department of Mathematics, KTH Royal Institute of Technology

*We suggest a shrinkage based technique for estimating covariance matrix in the high-dimensional normal model with missing data. Our asymptotic framework allows the dimensionality  $p$  grow to infinity together with the sample size,  $N$ , and extends the methodology of Ledoit and Wolf (2004) to the case of 2-step monotone missing data. Two new shrinkage type estimators are derived and their dominance properties over the Ledoit and Wolf (2004) estimator are shown under the expected quadratic loss. We perform a simulation study and conclude that the proposed estimators are successful for a range of missing data scenarios.*

**Key Words** Monotone missing data; High-dimensional estimation; Shrinkage type estimation.

**Mathematics Subject Classification** 62H12; 62F12.

## 1 Introduction

Estimation of the covariance matrix  $\Sigma$  in high dimensions has been important for many statistical procedures such as principal component analysis, linear regression and discriminant analysis. Some of these procedures require estimation of the covariance structure and eigenvalues of  $\Sigma$  while the others need the inverse. It has long been known that standard estimators such as maximum likelihood or least squares developed for large sample sizes  $N$  and relatively small number of variables  $p$  do not provide a satisfactory solution in high dimensions. Therefore modification of these estimators is of crucial importance for high-dimensional problems.

During recent years, many methods for improving the estimators of  $\Sigma$  and its inverse in the context of high dimensional data, i.e. if the number of variables  $p$  is much larger than  $N$ , have been developed; see e.g., Kubokawa and Srivastava (2008) and Ledoit and Wolf (2004). However, all of these methods have been focused on the case where all data is observed. Since in practice data sets often contain missing patterns, it is important to develop approaches for high dimensional covariance estimation in presence of missing data.

In this study, we focus on a particular type of statistical missing data, a *monotone missing scheme* with missing values patterns occurring completely at random. In other words, we assume that the underlying missing-data mechanism is ignorable under parametric estimation. This type of missing data problem has been extensively studied in the statistical literature; see e.g. Anderson (1957), Bhargava (1975). The sample scheme treated in them is called a  $k$ -step monotone missing data. In particular, in Anderson and Olkin (1985), the MLE of  $\Sigma$  for the case of 2-step monotone missing data was derived explicitly. Further, Kanda and Fujikoshi (1998) studied asymptotic properties of MLE of  $\Sigma$  constructed by 2-step monotone missing data and extend these results to the general case of  $k$ -step monotone missing data. However, in high dimensional situations, the MLE of  $\Sigma$  is known to perform poorly even for the case of complete data, since it is usually impossible to collect

enough observations to fulfill  $p \ll n$  condition.

In this paper, we present a solution to this problem by extending the idea of Ledoit and Wolf (2004) to the case of 2-step monotone missing data and suggest a new estimator of the covariance matrix that is adjusted to high dimensionality in combination with missing data.

The rest of the paper is organized as follows. In Section 2, we set up 2-step monotone missing data scheme and derive some auxiliary unbiased estimators, which then will be used in Section 3 where we present our new shrinkage estimators. Section 4 reports numerical studies on the risk performance of the suggested estimators and comparison with results by Ledoit and Wolf (2004). We conclude in Section 5 and present the proof of key results in Appendix (A.2, A.3 and A.4).

## 2 Auxiliary unbiased estimators and their asymptotic property under the 2-step monotone missing data

To specify the problem considered here, let  $\mathbf{x}$  be the  $p$ -dimensional random vector distributed as  $N_p(\boldsymbol{\mu}, \Sigma)$ , where  $\boldsymbol{\mu}$  is mean vector and  $\Sigma$  is covariance matrix. Assume that  $\mathbf{x}$  can be decomposed as  $(\mathbf{x}'_1, \mathbf{x}'_2)'$ , where  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are  $p_1$  and  $p_2$ -dimensional vectors, respectively. Suppose now that we have  $N_1$  independent observations on the full set of variables,  $\mathbf{x}$ , and  $N_2$  independent observations on  $\mathbf{x}_1$ . We introduce some notation for the sample means and covariance matrices. Let  $\bar{\mathbf{x}}^{(1)}$  denote the sample mean of  $\mathbf{x}$  based on the  $N_1$  observations, and  $\bar{\mathbf{x}}^{(1)} = (\bar{\mathbf{x}}_1^{(1)'}, \bar{\mathbf{x}}_2^{(1)'})'$ ,  $\bar{\mathbf{x}}_i^{(1)}: p_i \times 1$ ,  $i = 1, 2$ . Let further  $\bar{\mathbf{x}}^{(2)}$  denote the  $p_1$ -dimensional sample mean vector for based on the  $N_2$  observations. Throughout this paper, we use the letter  $j$  only as running suffix for sample observations.

With above notations, the 2-step monotone missing data scheme, described in Shutoh et al. (2011) can be presented by

$$\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \dots, \mathbf{x}_{N_1}^{(1)}, \mathbf{x}_1^{(2)}, \mathbf{x}_2^{(2)}, \dots, \mathbf{x}_{N_2}^{(2)},$$

where  $\mathbf{x}_j^{(1)}(j = 1, \dots, N_1)$  denotes the  $p$ -dimensional sample vector and  $\mathbf{x}_j^{(2)}(j =$

$1, \dots, N_2)$  denotes the  $p_1$ -dimensional sample vector. The graphical representation of this scheme is given in Figure 1.1.

Please insert Figure 1.1 around here.

Further the sample covariance matrices based on the  $N_1$  and  $N_2$  observations are expressed as

$$S^{(1)} = \frac{1}{n_1} \sum_{j=1}^{N_1} (\mathbf{x}_j^{(1)} - \bar{\mathbf{x}}^{(1)})(\mathbf{x}_j^{(1)} - \bar{\mathbf{x}}^{(1)})', \quad S^{(2)} = \frac{1}{n_2} \sum_{j=1}^{N_2} (\mathbf{x}_j^{(2)} - \bar{\mathbf{x}}^{(2)})(\mathbf{x}_j^{(2)} - \bar{\mathbf{x}}^{(2)})',$$

respectively, where  $n_i = N_i - 1$ ,  $i = 1, 2$ .

Let the partitions of  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$  and  $S^{(1)}$  corresponding to ones of  $\mathbf{x}$  be

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad S^{(1)} = \begin{pmatrix} S_{11}^{(1)} & S_{12}^{(1)} \\ S_{21}^{(1)} & S_{22}^{(1)} \end{pmatrix}.$$

respectively. Let  $\hat{\boldsymbol{\mu}}$  and  $\hat{\boldsymbol{\Sigma}}$  denote the MLE of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ , respectively, and assume that  $\hat{\boldsymbol{\mu}}$  and  $\hat{\boldsymbol{\Sigma}}$  can be partitioned in the same way as  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ . From Anderson and Olkin (1985), we can express the MLE's  $\hat{\boldsymbol{\mu}}$  and  $\hat{\boldsymbol{\Sigma}}$  as follows

$$\begin{aligned} \hat{\boldsymbol{\mu}}_1 &= \frac{1}{N} (N_1 \bar{\mathbf{x}}_1^{(1)} + N_2 \bar{\mathbf{x}}^{(2)}), \quad \hat{\boldsymbol{\mu}}_2 = \bar{\mathbf{x}}_2^{(1)} - \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} (\bar{\mathbf{x}}_1^{(1)} - \hat{\boldsymbol{\mu}}_1), \\ \hat{\Sigma}_{11} &= \frac{1}{N} (W_{11}^{(1)} + W^{(2)}), \quad \hat{\Sigma}_{12} = \hat{\Sigma}_{11} (W_{11}^{(1)})^{-1} W_{12}^{(1)}, \\ \hat{\Sigma}_{22} &= \frac{1}{N_1} W_{22,1}^{(1)} + \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12}, \end{aligned}$$

where

$$\begin{aligned} N &= N_1 + N_2, \\ W^{(1)} &= n_1 S^{(1)}, \quad W^{(2)} = n_2 S^{(2)} + \frac{N_1 N_2}{N} (\bar{\mathbf{x}}_1^{(1)} - \bar{\mathbf{x}}^{(2)})(\bar{\mathbf{x}}_1^{(1)} - \bar{\mathbf{x}}^{(2)})', \\ W^{(1)} &= \begin{pmatrix} W_{11}^{(1)} & W_{12}^{(1)} \\ W_{21}^{(1)} & W_{22}^{(1)} \end{pmatrix}, \quad W_{22,1}^{(1)} = W_{22}^{(1)} - W_{21}^{(1)} (W_{11}^{(1)})^{-1} W_{12}^{(1)}. \end{aligned}$$

In Lemma 2.1 stated below, we propose two auxiliary unbiased estimators of  $\boldsymbol{\Sigma}$  under 2-step monotone missing data.

**Lemma 2.1.** *Let  $\hat{\Sigma}_1$  and  $\hat{\Sigma}_2$  be defined as*

$$\hat{\Sigma}_1 = \begin{pmatrix} \hat{\Sigma}_{1,11} & \hat{\Sigma}_{1,12} \\ \hat{\Sigma}_{1,21} & \hat{\Sigma}_{1,22} \end{pmatrix}, \quad \hat{\Sigma}_2 = \begin{pmatrix} \hat{\Sigma}_{2,11} & \hat{\Sigma}_{2,12} \\ \hat{\Sigma}_{2,21} & \hat{\Sigma}_{2,22} \end{pmatrix},$$

where

$$\begin{aligned}\hat{\Sigma}_{1,11} &= \frac{1}{N-1}(W_{11}^{(1)} + W^{(2)}), \quad \hat{\Sigma}_{1,12} = \hat{\Sigma}_{11}(W_{11}^{(1)})^{-1}W_{12}^{(1)}, \\ \hat{\Sigma}_{1,21} &= \hat{\Sigma}'_{1,12}, \quad \hat{\Sigma}_{1,22} = b_1 W_{22.1}^{(1)} + \hat{\Sigma}_{21}\hat{\Sigma}^{-1}\hat{\Sigma}_{12}, \\ \hat{\Sigma}_{2,11} &= \frac{1}{N-1}(W_{11}^{(1)} + W^{(2)}), \quad \hat{\Sigma}_{2,12} = \frac{1}{N_1-1}W_{12}^{(1)}, \\ \hat{\Sigma}_{2,21} &= \hat{\Sigma}'_{2,12}, \quad \hat{\Sigma}_{2,22} = \frac{1}{N_1-1}W_{22}^{(1)},\end{aligned}$$

and

$$b_1 = \frac{1}{N_1 - p_1 - 1} \left( 1 - \frac{p_1}{N-1} - \frac{N_2 p_1}{(N-1)(N_1 - p_1 - 2)} \right).$$

Then

$$\mathbb{E}[\hat{\Sigma}_1] = \mathbb{E}[\hat{\Sigma}_2] = \Sigma.$$

**Proof :** See Appendix (A.2).

Due to the unbiasedness property, both estimators,  $\hat{\Sigma}_1$  and  $\hat{\Sigma}_2$  are expected to perform better than  $S^{(1)}$  which is based on the complete data part only. The following lemma provides exact results on the relative performance accuracy of these three estimators in terms of expected quadratic loss.

**Lemma 2.2.** *Let*

$$\beta_0 = \frac{p}{n_1}a_1^2 + \frac{1}{n_1}a_2,$$

where  $a_i = \text{tr } \Sigma^i / p$  and  $\Sigma^i$  denotes  $i$ -th power of the matrix  $\Sigma$  for  $i = 1, 2$ . Then,

- (i)  $\mathbb{E}[\|S^{(1)} - \Sigma\|_F^2] = \beta_0,$
- (ii)  $\mathbb{E}[\|\hat{\Sigma}_1 - \Sigma\|_F^2] = \beta_0 - m_1,$
- (iii)  $\mathbb{E}[\|\hat{\Sigma}_2 - \Sigma\|_F^2] = \beta_0 - m_2,$

where  $\|\cdot\|_F^2$  denotes the normalized Frobenious norm,

$$\begin{aligned}m_1 &= \frac{N_2}{nn_1} \left\{ pa_1^2 + a_2 - k_1 \frac{\text{tr } \Sigma_{22.1}^2}{p} - k_2 \frac{(\text{tr } \Sigma_{22.1})^2}{p} - 2k_3 \left( \frac{\text{tr } \Sigma_{11} \text{tr } \Sigma_{22.1}}{p} \right. \right. \\ &\quad \left. \left. + \frac{\text{tr } \Sigma_{22.1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}}{p} + \frac{\text{tr } \Sigma_{22.1} \text{tr } \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}}{p} \right) \right\}, \\ m_2 &= \frac{N_2}{nn_1} \left( pa_1^{(11)^2} + a_2^{(11)} \right),\end{aligned}$$

and

$$k_1 = \left( 1 - 3r_1 + 4r_1^2 - 2r_1^3r_2 - \frac{4 - 11r_1 + 6r_1^2r_2}{n_1} + \frac{3 - 4r_1r_2}{n_1^2} \right) \left\{ (1 - r_1) \left( 1 - r_1 - \frac{1}{n_1} \right) \left( 1 - r_1 - \frac{3}{n_1} \right) \right\}^{-1}, \quad (2.1)$$

$$k_2 = \left( 1 - 4r_1 + 7r_1^2 - 4r_1^3 - 2r_1^3r_2 + 2r_1^4r_2 + \frac{8r_1^3r_2 - 4r_1^2r_2 - 15r_1^2 + 16r_1 - 5}{n_1} + \frac{8r_1^2r_2 - 2r_1r_2 - 12r_1 + 7}{n_1^2} + \frac{2r_1r_2 - 3}{n_1^3} \right) \left\{ (1 - r_1) \left( 1 - r_1 - \frac{1}{n_1} \right)^2 \left( 1 - r_1 - \frac{3}{n_1} \right) \right\}^{-1}, \quad (2.2)$$

$$k_3 = \left( 1 - r_1 - \frac{1}{n_1} \right)^{-1}, \quad (2.3)$$

$$r_1 = \frac{p_1}{n_1}, \quad r_2 = \frac{n_1}{n}.$$

**Proof :** See Appendix (A.3).

It follows from (2.1)-(2.3) that  $k_i = 1 + O(n^{-1})$  for all  $i$ , assuming that  $p$  is fixed and  $n$  is relatively large. Then the following relationship between  $m_1$  and  $m_2$  holds

$$m_2 = m_1 + \frac{N_2}{nn_1} \left( \frac{\text{tr}(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^2}{p} + \frac{\Sigma_{21}\Sigma_{12}}{p} + \frac{(\text{tr} \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^2}{p} + \frac{(\text{tr} \Sigma_{11} \text{tr} \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^2}{p} \right) + o\left(\frac{1}{n}\right),$$

which implies that  $\hat{\Sigma}_1$  is more accurate estimator than  $\hat{\Sigma}_2$  in a large sample case. However,  $\hat{\Sigma}_1$  will be worse than  $S^{(1)}$  since  $k_1$ ,  $k_2$  and  $k_3$  grow when  $p_1/N_1 \rightarrow 1$ . On the other hand, since  $\hat{\Sigma}_2$  does not depend on the relationship between  $p_1$  and  $N_1$ , its accuracy is better than  $S^{(1)}$ . By this reason, it can be said that  $\hat{\Sigma}_2$  is a *flexibly* accurate estimator.

### 3 The shrinkage estimator for high-dimensional data with 2-step monotone missing values

On the complete samples, Ledoit and Wolf (2004) provide a shrinkage estimator for  $\Sigma$  given by

$$\hat{\Sigma}_f^* = \frac{\hat{\beta}_f}{\hat{\delta}_f} \hat{\mu}_f I + \frac{\hat{\alpha}_f}{\hat{\delta}_f} S^{(1)},$$

where

$$\hat{\alpha}_f = \hat{a}_2 - \hat{a}_1^2, \quad \hat{\beta}_f = \left(\frac{p}{n_1}\right) \hat{a}_1^2, \quad \hat{\delta}_f = \hat{a}_2 + \left(\frac{p}{n_1} - 1\right) \hat{a}_1^2, \quad \hat{\mu}_f = \hat{a}_1.$$

In this study, we extend this approach to the case of 2-step monotone missing data using auxiliary estimators  $\hat{\Sigma}_1$  and  $\hat{\Sigma}_2$  derived in Section 2.

Ledoit and Wolf type estimator is derived by the linear combination  $\Sigma^* = \rho_1 I + \rho_2 \hat{\Sigma}_b$  of the identity matrix  $I$  and the unbiased estimator  $\hat{\Sigma}_b = (\hat{\Sigma}_1 \text{ or } \hat{\Sigma}_2)$  whose expected quadratic loss is minimum. Observed that  $\hat{\Sigma}^* = \hat{\Sigma}_f^*$  if  $\hat{\Sigma}_b = S^{(1)}$ . Following the technique of Ledoit and Wolf (2004), consider now the optimization problem

$$\min_{\rho_1, \rho_2} \mathbb{E}[\|\Sigma^* - \Sigma\|_F^2] \text{ s.t. } \Sigma^* = \rho_1 I + \rho_2 \hat{\Sigma}_b, \quad (3.1)$$

where the coefficients  $\rho_1$  and  $\rho_2$  are non-random. The solution of (3.1) is given by

$$\Sigma^* = \frac{\beta}{\delta} \mu I + \frac{\alpha}{\delta} \hat{\Sigma}_b,$$

where  $\mu = \text{tr} \Sigma / p$ ,  $\alpha = \|\Sigma - \mu I\|_F^2$ ,  $\beta = \mathbb{E}[\|\hat{\Sigma}_b - \Sigma\|_F^2]$ ,  $\delta = \mathbb{E}[\|\hat{\Sigma}_b - \mu I\|_F^2]$ . With this notations we obtain  $\mu = a_1$ ,  $\alpha = a_2 - a_1^2$ .

### 3.1 The suggested estimator based on $\hat{\Sigma}_1$

A crucial step of our estimation is to get the explicit expressions of  $\mathbb{E}[\|\hat{\Sigma}_1 - \Sigma\|_F^2]$  and  $\mathbb{E}[\|\hat{\Sigma}_1 - \mu I\|_F^2]$  in terms of  $a_i$  and  $a_i^{(22.1)}$  which are functions of  $\text{tr} \Sigma_{22.1}$ . We consider the following conditions

$$\text{C0: } n_1, n_2, p_1, p \rightarrow \infty, \frac{p_1}{n_1} \rightarrow c_0 \in [0, 1), \frac{p}{n_i} \rightarrow c_i \in (0, \infty), i = 1, 2.$$

$$\text{C1: } 0 < \lim_{p \rightarrow \infty} a_i = \lim_{p \rightarrow \infty} \text{tr} \Sigma^i / p < \infty, i = 1, \dots, 4.$$

Using Lemma A.1, under the conditions C0 and C1,

$$\mathbb{E}[\|\hat{\Sigma}_1 - \Sigma\|_F^2] = \beta_1 + o(1), \quad \mathbb{E}[\|\hat{\Sigma}_1 - \mu I\|_F^2] = \delta_1 + o(1),$$

where

$$\begin{aligned}\beta_1 &= \frac{p}{n_1} a_1^2 - m_1^*, \\ \delta_1 &= a_2 + \left( \frac{p}{n_1} - 1 \right) a_1^2 - m_1^*, \\ m_1^* &= m_1^*(a_1, a_1^{(22.1)}) = \frac{N_2 p}{n n_1} \left\{ a_1^2 - 2k_3 a_1^{(22.1)} a_1 + (2k_3 - k_2) a_1^{(22.1)^2} \right\}\end{aligned}$$

and  $a_1^{(22.1)} = \text{tr } \Sigma_{22.1}/p$ . Thus, the asymptotically optimal linear combination of the identity matrix  $I$  and  $\hat{\Sigma}_1$  is given by

$$\Sigma_1^* = \frac{\beta_1}{\delta_1} \mu I + \frac{\alpha}{\delta_1} \hat{\Sigma}_1.$$

Observe however that  $\Sigma_1^*$  depends on unknown parameters. Therefore, the next step is to replace these parameters their consistent estimators, and show that the asymptotic properties of the resulting estimator are unchanged.

Under the conditions C0 and C1, the consistent estimators of  $\alpha$ ,  $\beta_1$  and  $\delta_1$  are given by

$$\hat{\alpha} = \hat{a}_2 - \hat{a}_1^2, \quad \hat{\beta}_1 = \frac{p}{n_1} \hat{a}_1^2 - \hat{m}_1^*, \quad \hat{\delta}_1 = (p/n_1 - 1) \hat{a}_1^2 + \hat{a}_2 - \hat{m}_1^*, \quad (3.2)$$

where  $\hat{m}_1^* = m_1^*(\hat{a}_1, \hat{a}_1^{(22.1)})$ , and  $\hat{a}_1$ ,  $\hat{a}_2$  and  $\hat{a}_1^{(22.1)}$  are consistent estimators of the corresponding parameters

$$\begin{aligned}\hat{a}_1 &= \frac{\text{tr } S^{(1)}}{p}, \quad \hat{a}_2 = \frac{n_1^2}{(n_1 - 1)(n_1 + 2)} \left( \frac{\text{tr } S^{(1)^2}}{p} - \frac{(\text{tr } S^{(1)})^2}{n_1 p} \right), \\ \hat{a}_1^{(22.1)} &= \frac{n_1 \text{tr } S_{22.1}^{(1)}}{(n_1 - p_1)p}.\end{aligned} \quad (3.3)$$

Now, by combining (3.2) and (3.3), we define the our estimator as

$$\hat{\Sigma}_1^* = \frac{\hat{\beta}_1}{\hat{\delta}_1} \hat{\mu} I + \frac{\hat{\alpha}}{\hat{\delta}_1} \hat{\Sigma}_1.$$

One can see that  $\hat{\Sigma}_1^*$  has the same asymptotic properties as  $\Sigma_1^*$ :  $\|\hat{\Sigma}_1^* - \Sigma_1^*\|_F^2 \xrightarrow{P} 0$  and  $\text{E}[\|\hat{\Sigma}_1^* - \Sigma\|_F^2] \rightarrow \text{E}[\|\Sigma^* - \Sigma\|_F^2]$  under conditions C0 and C1.

Consider further the relationship between  $\hat{\Sigma}_1^*$  and  $\hat{\Sigma}_f^*$  by Ledoit and Wolf (2004) when using complete data part only. The following proposition provides conditions for the estimator  $\hat{\Sigma}_f^*$  to dominate  $\hat{\Sigma}_1^*$  under the expected quadratic loss.

**Proposition 3.1.** *Under the conditions C1 and  $n_1, n_2, p_1, p \rightarrow \infty$  with  $p_1/n_1 \rightarrow 0$ ,  $p/n_i \rightarrow c_i \in (0, \infty)$ ,  $i = 1, 2$  and  $n_i/(n_1 + n_2) \rightarrow c_i \in (0, 1)$ ,  $i = 3, 4$*

$$\mathbb{E}[\|\hat{\Sigma}_f^* - \Sigma\|_F^2] - \mathbb{E}[\|\hat{\Sigma}_1^* - \Sigma\|_F^2] \rightarrow \frac{c_1 c_4 \alpha}{\delta_0 \delta_1} (a_1^2 - a_1^{(22 \cdot 1)})^2.$$

**Proof :** *See Appendix (A.4).*

### 3.2 The suggested estimator based on $\hat{\Sigma}_2$

The dominance property of  $\hat{\Sigma}_1^*$  over  $\hat{\Sigma}_f^*$  follows from the condition  $p_1/n_1 \rightarrow 0$ , which is very strong. Furthermore, if  $p_1 > n_1$ ,  $\hat{\Sigma}_1^*$  is not applicable due to singularity of  $\hat{\Sigma}_1$ . In this subsection, we get another condition for the estimator of  $\Sigma$  to dominate  $\hat{\Sigma}_f^*$  in the case of  $p_1 > n_1$ . Using the technique of subsection 3.1 and by considering

$$\begin{aligned} \text{C3 : } n_1, n_2, p_1, p \rightarrow \infty, \frac{p_1}{n_1} \rightarrow c_0 \in [0, \infty), \frac{p}{n_i} \rightarrow c_i \in (0, \infty), i = 1, 2, \\ n_i/(n_1 + n_2) \rightarrow c_i \in (0, 1), i = 3, 4, \end{aligned}$$

we obtain

$$\mathbb{E}[\|\hat{\Sigma}_2 - \Sigma\|_F^2] = \beta_2 + o(1), \quad \mathbb{E}[\|\hat{\Sigma}_2 - \mu I\|_F^2] = \delta_2 + o(1),$$

where

$$\beta_2 = \frac{p}{n_1} a_1^2 - m_2^*, \quad \delta_2 = a_2 + \left( \frac{p}{n_1} - 1 \right) a_1^2 - m_2^*, \quad m_2^* = m_2^*(a_1^{(11)}) = \frac{N_2 p}{n n_1} a_1^{(11)^2}$$

and  $a_1^{(11)} = \text{tr } \Sigma_{11}/p$ . Observe that C3 presents more flexible asymptotic framework than C0. Now, the asymptotically optimal linear combination of the identity matrix  $I$  and  $\hat{\Sigma}_2$  is given by

$$\Sigma_2^* = \frac{\beta_2}{\delta_2} \mu I + \frac{\alpha}{\delta_2} \hat{\Sigma}_2.$$

Of this, we propose our estimator of the form

$$\hat{\Sigma}_2^* = \frac{\hat{\beta}_2}{\hat{\delta}_2} \hat{\mu} I + \frac{\hat{\alpha}}{\hat{\delta}_2} \hat{\Sigma}_2,$$

where

$$\hat{\alpha} = \hat{a}_2 - \hat{a}_1^2, \quad \hat{\beta}_2 = \frac{p}{n_1} \hat{a}_1^2 - m_2^*(\hat{a}_1^{(11)}), \quad \hat{\delta}_2 = (p/n_1 - 1) \hat{a}_1^2 + \hat{a}_2 - m_2^*(\hat{a}_1^{(11)})$$

and  $\hat{a}_1^{(11)} = \text{tr } S_{11}^{(1)}/p$ .

By the same arguments as in subsection 3.1,  $\Sigma_2^*$  and  $\hat{\Sigma}_2^*$  have the same asymptotic properties because  $\|\hat{\Sigma}_2^* - \Sigma_2^*\|_F^2 \xrightarrow{P} 0$  and  $\text{E}[\|\hat{\Sigma}_2^* - \Sigma\|_F^2] \rightarrow \text{E}[\|\Sigma_2^* - \Sigma\|_F^2]$  under conditions C1 and C3.

The dominance property of  $\hat{\Sigma}_2^*$  over  $\hat{\Sigma}_f^*$  follows from the following proposition.

**Proposition 3.2.** *Under the conditions C1 and C3,*

$$\text{E}[\|\hat{\Sigma}_f^* - \Sigma\|_F^2] - \text{E}[\|\hat{\Sigma}_2^* - \Sigma\|_F^2] \rightarrow \frac{c_1 c_4 \alpha}{\delta_0 \delta_2} a_1^{(11)^2}.$$

**Proof :** The proof of this proposition which we omit, is straightforward application of the technique derived in Appendix (A.4).

## 4 Numerical results

Using Monte Carlo simulations, we compare performance of our new estimators  $\hat{\Sigma}_1^*$  and  $\hat{\Sigma}_2^*$  to the shrinkage estimator by Ledoit and Wolf (2004).

The simulation experiment is planned as follows. The data are generated from the normal model with zero mean vector and covariance matrix

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)(\rho^{|i-j|})\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p),$$

where  $\sigma_i$  is a random variable following continuous uniform distribution on the interval  $(0, 5)$ . For all simulations we assume  $N_1 = N_2$  for convenience. We set  $p = 100, 150, 200$ , and for each one consider three different values of  $\rho = 0.2, 0.5, 0.8$  to vary the covariance parameters.

The first setting represents the case when missing part of the data asymptotically exceeds the complete part. We fix  $p_1 = 10$ , the size of complete part, and assume that  $N_1 = N_2 = 50$ . This implies that  $p_2$ , the dimension of the missing part grows with  $p$ .

In the second setting, we vary the proportion of the data that are missing. To put it more precisely, for each  $p = 100, 150, 200$ , we set  $p_1 = [q \times p]$ , for  $q = 0.5, 0.7, 0.9$ , and

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)(0.8^{|i-j|})\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p),$$

where  $\sigma_i$  is a random variable having continuous uniform distribution on the interval  $(0, 5)$ . Here,  $[a]$  denotes the integer part of  $a$ .

Within the first setting, three following estimators are evaluated:  $\hat{\Sigma}_f^*$ ,  $\hat{\Sigma}_1^*$ ,  $\hat{\Sigma}_2^*$ , whereas in the second setting we only focus on  $\hat{\Sigma}_f^*$  and  $\hat{\Sigma}_2^*$ .

To evaluate performance accuracy we use the loss function based on normalized Frobenius norm,  $\mathcal{L}(\Sigma, \Delta) = \|\Delta - \Sigma\|_F^2$ , and define the risk function as the expected loss,  $R(\Sigma, \Delta) = E[\mathcal{L}(\Sigma, \Delta)]$ , when estimating  $\Sigma$  by  $\Delta$ . We assess the relative risk rate as follows

$$RR(\Delta) = R(\Sigma, \Delta)/R(\Sigma, S^{(1)}),$$

where  $S^{(1)}$  is used reference estimator for all comparisons. For all the settings, the value of  $RR$  is computed by 100,000 replications.

Tables 4.1-4.3 report the risk rates for all three estimators in the for the first setting, and figures 4.1-4.3 plot the risk rates for  $\hat{\Sigma}_f^*$  and  $\hat{\Sigma}_2^*$  as functions of  $p$  for the second setting.

Overview through the simulation results reveals that our new estimators provide significant improvements over  $\hat{\Sigma}_f^*$  for both types of missing scenarios. Although  $RR$ s of  $\hat{\Sigma}_1^*$  and  $\hat{\Sigma}_2^*$  are very close to each other for all  $\rho$ ,  $\hat{\Sigma}_2^*$  dominates  $\hat{\Sigma}_1^*$  in the case when missing part grows. As expected,  $\hat{\Sigma}_2^*$  yields substantial improvement over  $\hat{\Sigma}_f^*$  even in the case of fixed missing part. It could be observed that both estimators are sensitive to the size of  $\rho \in (0, 1)$  which is naturally expected in a AR(1) model.

In conclusion, both suggested estimators,  $\hat{\Sigma}_1^*$  and  $\hat{\Sigma}_2^*$  are shown to have dominance properties over  $\hat{\Sigma}_f^*$  in high-dimensional normal model with 2-step monotone missing data.

Please insert Tables 4.1–4.3 and Figures 4.1–4.3 around here.

## 5 Conclusion

This paper provided the estimators for  $\Sigma$  in the case of 2-step monotone missing data with high-dimensional setting by extending Ledoit and Wolf's (2004) estimator. The result provided a solution to the problem such that the existing MLE on the basis of 2-step monotone missing data is not also applicable in the case of high-dimensional setting. Further, it could be observed that our estimators were superior to Ledoit and Wolf's (2004) estimator in view of the relative risk rate for the performed simulations. In the future, we may also derive the theoretical results of the procedures with our estimators which are applicable to the high-dimensional and missing data.

## Appendix

To prove propositions 2.1 and 2.2 we need two additional lemmas stated below. For these proofs we need the some preliminary results formulated in Lemma A.1 and Lemma A.2.

### A.1 Preliminary results

**Lemma 5.1.** *Suppose that  $A$  follows Wishart distribution  $W_p(\Sigma, n)$ , where  $A$  admits the similar partition as that of  $\Sigma$  in Section 2, and let  $A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$ . Then*

- (i)  $A_{22.1} \sim W_p(\Sigma_{22.1}, n - p_1)$ , and  $A_{22.1}$  is independent of  $A_{11}$  and  $A_{12}$ ,
- (ii)  $A_{12}|A_{11} \sim N_{p_1 \times p_2}(\text{Vec}(A_{11}\Sigma_{11}^{-1}\Sigma_{12}), \Sigma_{22.1} \otimes A_{11})$

**Proof.** For the proof, see Muirhead (1982) and Siotani et al. (1985).

Now we state Lemma A.2.

**Lemma 5.2.** For  $\hat{\Sigma}_1$  defined in Section 2, it holds that

$$\begin{aligned} \mathbb{E}[\text{tr } \hat{\Sigma}_1^2] &= \frac{n_1 + 1}{n_1} \text{tr } \Sigma^2 + \frac{1}{n_1} (\text{tr } \Sigma)^2 - \frac{N_2}{nn_1} \{ \text{tr } \Sigma^2 + (\text{tr } \Sigma)^2 - k_1 \text{tr } \Sigma_{22 \cdot 1}^2 \\ &\quad - k_2 (\text{tr } \Sigma_{22 \cdot 1})^2 - 2k_3 (\text{tr } \Sigma_{22 \cdot 1} \text{tr } \Sigma_{11} + \text{tr } \Sigma_{22} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \\ &\quad + \text{tr } \Sigma_{22} \text{tr } \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \}, \end{aligned}$$

where  $k_i$ ,  $i = 1, 2, 3$  is defined in Lemma 2.2.

**Proof.** Using the following expression

$$\mathbb{E}[\text{tr } \hat{\Sigma}_1^2] = \mathbb{E}[\text{tr}(\hat{\Sigma}_{1,11}^2 + \hat{\Sigma}_{1,12} \hat{\Sigma}_{1,21})] + \mathbb{E}[\text{tr}(\hat{\Sigma}_{1,22}^2 + \hat{\Sigma}_{1,21} \hat{\Sigma}_{1,12})] \quad (\text{A.1})$$

and the results of Lemma A.1, we obtain

$$\begin{aligned} \mathbb{E}[\text{tr}(\hat{\Sigma}_{1,11}^2 + \hat{\Sigma}_{1,12} \hat{\Sigma}_{1,21})] &= \frac{n+1}{n} \text{tr}(\Sigma_{11}^2 + \Sigma_{12} \Sigma_{21}) + \frac{1}{n} \text{tr } \Sigma \text{tr } \Sigma_{11} \\ &\quad + \frac{N_2}{n(n_1 - p_1 - 1)} \text{tr } \Sigma_{22 \cdot 1} \text{tr } \Sigma_{11}, \end{aligned} \quad (\text{A.2})$$

For the second term in (A.1) we get

$$\begin{aligned} \mathbb{E}[\text{tr}(\hat{\Sigma}_{1,22}^2 + \hat{\Sigma}_{1,21} \hat{\Sigma}_{1,12})] &= \frac{1}{n^2} \mathbb{E}[\text{tr } \hat{\Psi}_{21} W_{11}^2 \hat{\Psi}_{12}] + b_1^2 \mathbb{E}[\text{tr } W_{22 \cdot 1}^{(1)2}] \\ &\quad + \frac{b_1}{n} \mathbb{E}[\text{tr } \hat{\Psi}_{21} W_{11} \hat{\Psi}_{12} W_{22 \cdot 1}^{(1)}] \\ &\quad + \frac{b_1}{n} \mathbb{E}[\text{tr } W_{22 \cdot 1}^{(1)} \hat{\Psi}_{21} W_{11} \hat{\Psi}_{12}] \\ &\quad + \frac{1}{n^2} \mathbb{E}[\text{tr } \hat{\Psi}_{21} W_{11} \hat{\Psi}_{12} \hat{\Psi}_{21} W_{11} \hat{\Psi}_{12}], \end{aligned} \quad (\text{A.3})$$

where  $\hat{\Psi}_{21} = W_{21}^{(1)} \left( W_{11}^{(1)} \right)^{-1}$ . From Lemma A.1,

$$\begin{aligned} \mathbb{E}[\text{tr } \hat{\Psi}_{21} W_{11}^2 \hat{\Psi}_{12}] &= n(n+1) \text{tr } \Sigma_{21} \Sigma_{12} + n(\text{tr } \Sigma_{11} \text{tr } \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \\ &\quad + \text{tr } \Sigma_{22 \cdot 1} \text{tr } \Sigma_{11}) + \frac{nN_2}{(n-p_1-1)} \text{tr } \Sigma_{11} \text{tr } \Sigma_{22 \cdot 1}, \\ \mathbb{E}[\text{tr } W_{22 \cdot 1}^{(1)2}] &= (n_1 - p_1)(n_1 - p_1 - 1) \text{tr } \Sigma_{22 \cdot 1}^2 \\ &\quad + (n_1 - p_1)(\text{tr } \Sigma_{22 \cdot 1})^2, \end{aligned} \tag{A.4}$$

$$\begin{aligned} \mathbb{E}[\text{tr } \hat{\Psi}_{21} W_{11} \hat{\Psi}_{12} W_{22 \cdot 1}^{(1)}] &= (n_1 - p_1)p_1 \left( 1 + \frac{N_2}{n_1 - p_1 - 1} \right) \text{tr } \Sigma_{22 \cdot 1}^2 \\ &\quad + n(n_1 - p_1) \text{tr } \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22 \cdot 1} \}, \end{aligned} \tag{A.5}$$

$$\begin{aligned} \mathbb{E}[\text{tr}(\hat{\Psi}_{21} W_{11} \hat{\Psi}_{12})^2] &= \left\{ p_1^2 + \frac{2p_1^2 N_2}{n_1 - p_1 - 1} + c_1(2N_2 p_1 + N_2^2 p_1^2) \right. \\ &\quad \left. + 2c_2(N_2 p_1 + N_2 p_1^2 + N_2^2 p_1) \right\} \text{tr } \Sigma_{22 \cdot 1}^2 \\ &\quad + \left\{ p_1 + \frac{2p_1 N_2}{n_1 - p_1 - 1} + c_2(2N_2 p_1 + N_2^2 p_1^2) \right. \\ &\quad \left. + c_2(n_1 - p_1 - 1)(N_2 p_1 + N_2^2 p_1 + N_2 p_1^2) \right\} \\ &\quad + \{ \text{tr } \Sigma_{22 \cdot 1}^2 + (\text{tr } \Sigma_{22 \cdot 1})^2 \} + 2 \left( n + N_2 \right. \\ &\quad \left. + \frac{N_2^2 + N_2 p_1 + N_2}{n_1 - p_1 - 1} \right) \text{tr } \Sigma_{22 \cdot 1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \\ &\quad + 2n p_1 \left( 1 + \frac{N_2}{n_1 - p_1 - 1} \right) \text{tr } \Sigma_{22 \cdot 1} \text{tr } \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \\ &\quad + (n^2 + n) \text{tr}(\Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})^2 \\ &\quad + n \left( \text{tr } \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right)^2. \end{aligned} \tag{A.6}$$

Substituting (A.4)-(A.6) into (A.3), the proof of Lemma A.1 is complete.  $\square$

## A.2 Proof of Lemma 2.1

Using the moment properties of the Wishart distribution and applying Lemma A.1, we get

$$\begin{aligned}
\mathbb{E}[\hat{\Sigma}_{1,11}] &= \mathbb{E} \left[ \frac{1}{N-1} (W_{11}^{(1)} + W^{(2)}) \right] \\
&= \frac{1}{N-1} \{ (N_1 - 1) \Sigma_{11} + N_2 \Sigma_{11} \} \\
&= \Sigma_{11},
\end{aligned} \tag{A.7}$$

$$\begin{aligned}
\mathbb{E}[\hat{\Sigma}_{1,12}] &= \frac{1}{N-1} \mathbb{E} \left[ (W_{11}^{(1)} + W^{(2)}) \left( W_{11}^{(1)} \right)^{-1} W_{12}^{(1)} \right] \\
&= \frac{1}{N-1} \mathbb{E} \left[ W_{12}^{(1)} + W^{(2)} \left( W_{11}^{(1)} \right)^{-1} W_{12}^{(1)} \right] \\
&= \frac{1}{N-1} \{ (N_1 - 1) \Sigma_{12} + N_2 \Sigma_{12} \} \\
&= \Sigma_{12},
\end{aligned} \tag{A.8}$$

$$\begin{aligned}
\mathbb{E}[\hat{\Sigma}_{1,22}] &= \frac{1}{N_1 - p_1 - 1} \left( 1 - \frac{p_1}{N-1} - \frac{N_2 p_1}{(N-1)(N_1 - p_1 - 2)} \right) \mathbb{E}[W_{22 \cdot 1}^{(1)}] \\
&\quad + \frac{1}{N-1} \mathbb{E} \left[ W_{21}^{(1)} \left( W_{11}^{(1)} \right)^{-1} \left( W_{11}^{(1)} + W^{(2)} \right) \left( W_{11}^{(1)} \right)^{-1} W_{12}^{(1)} \right].
\end{aligned} \tag{A.9}$$

Then, for the first term of (A.9) we have

$$\begin{aligned}
&\frac{1}{N_1 - p_1 - 1} \left( 1 - \frac{p_1}{N-1} - \frac{N_2 p_1}{(N-1)(N_1 - p_1 - 2)} \right) \mathbb{E}[W_{22 \cdot 1}^{(1)}] \\
&= \left( 1 - \frac{p_1}{N-1} - \frac{N_2 p_1}{(N-1)(N_1 - p_1 - 2)} \right) \Sigma_{22 \cdot 1}.
\end{aligned} \tag{A.10}$$

For the second term of (A.9) we obtain

$$\begin{aligned}
&\frac{1}{N-1} \mathbb{E} \left[ W_{21}^{(1)} \left( W_{11}^{(1)} \right)^{-1} \left( W_{11}^{(1)} + W^{(2)} \right) \left( W_{11}^{(1)} \right)^{-1} W_{12}^{(1)} \right] \\
&= \Sigma_{22} - \left( 1 - \frac{p_1}{N-1} - \frac{N_2 p_1}{(N-1)(N_1 - p_1 - 2)} \right) \Sigma_{22 \cdot 1}.
\end{aligned} \tag{A.11}$$

Substituting (A.10) and (A.11) into (A.9), we have

$$\mathbb{E}[\hat{\Sigma}_{1,22}] = \Sigma_{22}. \tag{A.12}$$

Combining (A.7), (A.8) and (A.12), we get  $\mathbb{E}[\hat{\Sigma}_1] = \Sigma$ .

The unbiasedness of  $\hat{\Sigma}_2$  is easily shown by calculating the first moment of the Wishart matrices  $W^{(1)}$  and  $W^{(2)}$ .  $\square$

### A.3 Proof of Lemma 2.2

The result in (i) is established in (see, e.g., Ledoit and Wolf (2004)). To show (ii) we first express it as

$$\mathbb{E}[\|\hat{\Sigma}_1 - \Sigma\|_F^2] = \mathbb{E}[\text{tr} \hat{\Sigma}_1^2]/p - a_2.$$

By using Lemma A.2, we can calculate  $\mathbb{E}[\text{tr} \hat{\Sigma}_1^2]$ , which gives the statement (ii). It should be noted that (i) is shown assuming that  $N_2 = 0$ . Finally, we give the proof of (iii), by first expressing

$$\mathbb{E}[\|\hat{\Sigma}_2 - \Sigma\|_F^2] = \mathbb{E}[\text{tr} \hat{\Sigma}_2^2]/p - a_2 \quad (\text{A.13})$$

and using the decomposition of  $\hat{\Sigma}_2$  as

$$\begin{pmatrix} \frac{n_1}{n} S_{11}^{(1)} + \frac{N_2}{n} S^{(2)} & S_{12}^{(1)} \\ S_{21}^{(1)} & S_{22}^{(1)} \end{pmatrix}.$$

Then

$$\begin{aligned} \frac{1}{p} \mathbb{E}[\text{tr} \hat{\Sigma}_2^2] &= \frac{1}{p} \left\{ \mathbb{E} \left[ \text{tr} \left( \frac{n_1}{n} S_{11}^{(1)} + \frac{N_2}{n} S^{(2)} \right)^2 \right] + 2 \mathbb{E}[\text{tr} S_{12}^{(1)} S_{21}^{(1)}] \right. \\ &\quad \left. + \mathbb{E}[\text{tr} S_{22}^{(1)2}] \right\} \\ &= \frac{1}{p} \left\{ \frac{n_1 + 1}{n_1} \text{tr} \Sigma^2 + \frac{1}{n_1} (\text{tr} \Sigma)^2 \right\} - \frac{N_2 p}{n n_1} \{ \text{tr} \Sigma_{11}^2 + (\text{tr} \Sigma_{11})^2 \}. \end{aligned} \quad (\text{A.14})$$

Substituting (A.14) into (A.13), we obtain (iii). It should be noted that (i) is shown assuming that  $N_2 = 0$ .  $\square$

### A.4 Proof of Proposition 3.1

Under the conditions C1 and assuming that  $n_1, n_2, p_1, p \rightarrow \infty$  with  $p_1/n_1 \rightarrow 0$  and  $p/n_i \rightarrow c_i \in (0, \infty)$ ,  $i = 1, 2$ ,

$$\mathbb{E}[\|\hat{\Sigma}_1^* - \Sigma\|_F^2] \rightarrow \frac{\alpha\beta_1}{\delta_1} \text{ and } \mathbb{E}[\|\hat{\Sigma}_f^* - \Sigma\|_F^2] \rightarrow \frac{\alpha\beta_0}{\delta_0}, \quad (\text{A.15})$$

where  $\delta_0 = \alpha + \beta_0$ . By using the continuous mapping theorem and (A.15), we obtain

$$E[\|\hat{\Sigma}_f^* - \Sigma\|_F^2] - E[\|\hat{\Sigma}_1^* - \Sigma\|_F^2] \rightarrow \frac{\alpha\beta_0}{\delta_0} - \frac{\alpha\beta_1}{\delta_1}. \quad (\text{A.16})$$

On the other hand, under the same assumptions as above,

$$\frac{\alpha\beta_0}{\delta_0} - \frac{\alpha\beta_1}{\delta_1} = \frac{\alpha}{\delta_0\delta_1}(\beta_0 - \beta_1) = \frac{\alpha}{\delta_0\delta_1}m_1 \rightarrow \frac{N_2p\alpha}{nn_1\delta_0\delta_1}(a_1^2 - a_1^{(22\cdot 1)})^2. \quad (\text{A.17})$$

Combining (A.16) and (A.17), Proposition 3.1 follows.  $\square$

## Acknowledgments

The authors would like to extend their sincere gratitude to the referee who gave invaluable comments and suggestions, which have greatly enhanced this paper. Any remaining errors are the authors' responsibility. This research was in part supported by grants from the Forum for Asian Studies (4440101-2010), Stockholm University, Sweden. M.Hyodo's research was in part supported by Grant-in-Aid for JSPS Fellow (23-9731). T. Seo's research was in part supported by Grant-in-Aid for Scientific Research (C) (23500360). T. Pavlenko's research was in part supported by grants from Swedish Research Council (421-2008-1966).

## References

- Anderson, T. W. (1957). Maximum likelihood estimates for multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, **52**, 200-203.
- Anderson, T.W. and Olkin, I. (1985). Maximum likelihood estimation of the parameters of multivariate normal distribution. *Linear Algebra and Its Applications*, **70**, 147-171.
- Bhargava, R. P. (1975). Some one-sample hypothesis testing problems when there is monotone sample from a multivariate normal population. *Annals of the Institute of Statistical Mathematics*, **27**, 327-339.

- Fujikoshi, Y. and Seo, T. (1998). Asymptotic approximations for EPMC's of the linear and the quadratic discriminant function when the sample size and the dimension are large. *Random Operators and Stochastic Equations*, **6**, 269-280.
- Jinadasa, K. G. and Tracy, D. S. (1992). Maximum likelihood estimation for multivariate normal distribution with monotone sample. *Communications in Statistics A, Theory and Methods*, **21**, 41-50.
- Kanda, T. and Fujikoshi, Y. (1998). Some basic properties of the MLE's for a multivariate normal distribution with monotone missing data. *American Journal of Mathematical and Management Sciences*, **18**, 161-190.
- Kubokawa, T. and Srivastava, M. S. (2008). Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *Journal of Multivariate Analysis*, **99**, 1906-1928.
- Little, R. J. A. and Rubin, D. R. (1987). *Statistical Analysis with Missing Data*, Wiley, New York.
- Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large dimensional covariance matrices. *Journal of Multivariate Analysis*, **88**, 365-411.
- Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*, Wiley, New York.
- Shutoh, N., Hyodo, M. and Seo, T. (2011). An asymptotic approximation for EPMC in linear discriminant analysis based on two-step monotone missing samples. *Journal of Multivariate Analysis*, **102**, 252-263.
- Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis*, A Graduate Course and Handbook. Ohio:American Science Press.

Table 4.1: The risk rates when  $\rho = 0.2$

	$p = 100$	$p = 150$	$p = 200$
$RR(\hat{\Sigma}_f^*)$	0.1968	0.1347	0.1050
$RR(\hat{\Sigma}_1^*)$	0.1946	0.1336	0.1044
$RR(\hat{\Sigma}_2^*)$	0.1942	0.1336	0.1044

Table 4.2: The risk rates when  $\rho = 0.5$

	$p = 100$	$p = 150$	$p = 200$
$RR(\hat{\Sigma}_f^*)$	0.3392	0.2550	0.2161
$RR(\hat{\Sigma}_1^*)$	0.3319	0.2511	0.2135
$RR(\hat{\Sigma}_2^*)$	0.3309	0.2507	0.2134

Table 4.3: The risk rates when  $\rho = 0.8$

	$p = 100$	$p = 150$	$p = 200$
$RR(\hat{\Sigma}_f^*)$	0.6821	0.5826	0.5039
$RR(\hat{\Sigma}_1^*)$	0.6574	0.5634	0.4895
$RR(\hat{\Sigma}_2^*)$	0.6508	0.5625	0.4890

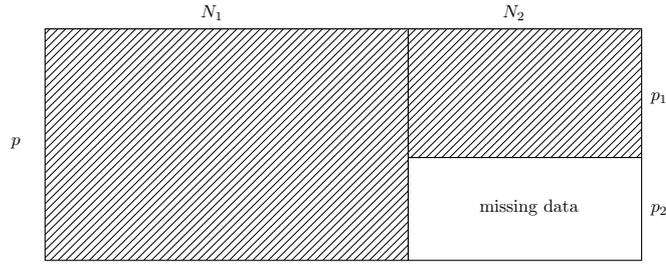


Figure 1.1: 2-step monotone missing data

$RR(\Delta)$  plotted as a function of the number of variables for  $\Delta = \hat{\Sigma}_f^*$ (solid) and  $\Delta = \hat{\Sigma}_2^*$ (dashed).

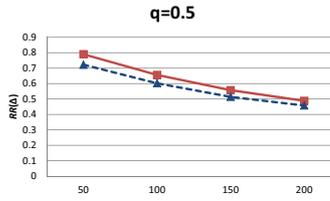


Fig.4.1.  $q = 0.5$ .

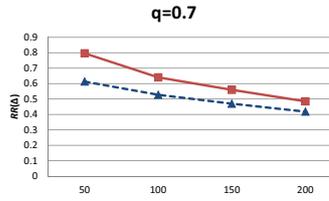


Fig.4.2.  $q = 0.7$ .

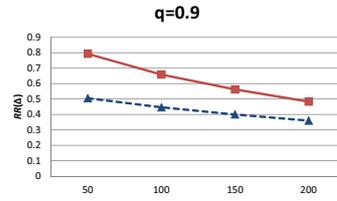


Fig.4.3.  $q = 0.9$ .