Journal of Multivariate Analysis 152 (2016) 61–81

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Linear shrinkage estimation of large covariance matrices using factor models

Yuki Ikeda, Tatsuya Kubokawa *

Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

ARTICLE INFO

Article history:
Received 12 February 2015
Available online 21 August 2016

AMS subject classifications:
62F12
62H12
62J07

Keywords:
Covariance matrix
Factor model
High dimension
Large sample
Non-normal distribution
Normal distribution
Portfolio management
Ridge-type estimator
Risk function

ABSTRACT

The problem of estimating a large covariance matrix using a factor model is addressed when both the sample size and the dimension of the covariance matrix tend to infinity. 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) linear shrinkage estimators without factors as special cases. The optimal weights in the class are derived, and plug-in weighted estimators are proposed, given that the optimal weights depend on unknown parameters. Numerical results show that our method performs well. Finally, we provide an application to portfolio management.

1. Introduction

The estimation of a large covariance matrix is a fundamental subject in economics, financial engineering, biology, signal processing, and other fields. Accordingly, the problem has been widely studied. In the estimation of a $p \times p$ covariance matrix, the classical large-sample theory assumes that p is fixed and the sample size N tends to infinity. In this setting, the covariance matrix can be estimated from its sample version, which is a consistent estimator. However, in applications, we often encounter large data sets that contain high-dimensional variables. In this case, using the sample covariance matrix is inappropriate because it becomes singular when p is larger than N. Even if $p < N$, the sample covariance matrix is unstable as noted by Fan et al. [9].

Various methods have been proposed to estimate the covariance matrix in high dimension. Ledoit and Wolf [18], Schafer and Strimmer [23], Chen et al. [6], Fisher and Sun [12], Touloumis [27], and others suggested well-conditioned estimators that combine the sample covariance matrix and more stable statistics, which are called linear shrinkage or weighted estimators. Many other well-conditioned estimators such as regularization and thresholding techniques and non-linear shrinkage methods have been studied by Bickel and Levina [1,2], Rothman et al. [22], Cai and Liu [3], Cai and Zhou [4], Ledoit and Wolf [19], Ledoit and Wolf [20], Fan et al. [9], Fan et al. [10], Fan et al. [11], and others.

* Corresponding author.
E-mail addresses: pt2y1003@gmail.com (Y. Ikeda), tatsuya@e.u-tokyo.ac.jp (T. Kubokawa).

http://dx.doi.org/10.1016/j.jmva.2016.08.001
0047-259X/

When additional information on the covariance matrix is available, we can improve the sample covariance matrix. Factor models use correlations between variables of interest and several covariates, called factors. Factor models have been used in many applications, particularly, to explain stock returns in financial economics. Fama and French [8] found that excess asset returns are well explained by three factors, namely sensitivity to market excess return, market capitalization and book-to-price ratio. Ledoit and Wolf [17], Ren and Shimotsu [21] and Fan et al. [9] considered shrinking the sample covariance matrix toward the statistics constructed from the factor models. Ledoit and Wolf [17] and Ren and Shimotsu [21] suggested linear shrinkage estimators but did not address high-dimensional settings. Fan et al. [9] proposed a covariance estimator based on a strict factor model in high-dimensional settings. The strict factor model assumes independence among idiosyncratic components; hence, the error covariance matrix becomes diagonal. However, cross-sectional independence is restrictive in many applications, as noted by Chamberlain and Rothschild [5]. Recently, Fan et al. [10] applied a thresholding technique and suggested an invertible covariance estimator based on an approximate factor model, which allows cross-sectional correlations. Fan et al. [10] derived favorable convergence rates under sparsity of covariance of idiosyncratic components when both p and N tend to infinity.

To describe the results in the present paper, let $\boldsymbol{y}$ and $\boldsymbol{x}$ be a variable of interest and a factor variable, respectively. The variance and covariance matrices of $\boldsymbol{y}$ and $\boldsymbol{x}$ are denoted by $\boldsymbol{\Sigma}_y$, $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_{yx}$, and their sample variance and covariance matrices are denoted by $\boldsymbol{S}_y$, $\boldsymbol{S}_x$ and $\boldsymbol{S}_{yx}$. The problem is the estimation of the $p \times p$ matrix $\boldsymbol{\Sigma}_y$, but the sample covariance matrix $\boldsymbol{S}_y$ is ill-conditioned in high dimensions. When $\boldsymbol{S}_y$ is expressed as

$$\boldsymbol{S}_y = \boldsymbol{S}_{y|x} + \boldsymbol{S}_{yx}\boldsymbol{S}_x^{-1}\boldsymbol{S}_{xy} \quad \text{for } \boldsymbol{S}_{y|x} = \boldsymbol{S}_y - \boldsymbol{S}_{yx}\boldsymbol{S}_x^{-1}\boldsymbol{S}_{xy},$$

the linear shrinkage estimator given in the literature suggests shrinking the term $\boldsymbol{S}_{y|x}$. In this paper, we consider shrinking the term $\boldsymbol{S}_{yx}\boldsymbol{S}_x^{-1}\boldsymbol{S}_{xy}$ as well as $\boldsymbol{S}_{y|x}$. Specifically, we suggest a double shrinkage estimator denoted by $\boldsymbol{\delta}_{\alpha,\beta}$ with weights α and β. We obtain the optimal weights and derive approximations and estimators thereof, which lead to the proposed plug-in double shrinkage estimator of $\boldsymbol{\Sigma}_y$.

In evaluating the optimal weights α and β, their estimators and large-sample properties, we consider two cases: (i) $\mathrm{tr}(\boldsymbol{\Sigma}_y^2) = O(p^2)$ and (ii) $\mathrm{tr}(\boldsymbol{\Sigma}_y^2) = O(p)$. $\boldsymbol{\Sigma}_y$ is expressed as $\boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}_{y|x} + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy} = \boldsymbol{\Sigma}_{y|x} + \boldsymbol{B}\boldsymbol{\Sigma}_x\boldsymbol{B}^\top$ for $\boldsymbol{\Sigma}_{y|x} = \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}$ because the factor loading $\boldsymbol{B}$ corresponds to $\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}$ as seen in Section 2. For $\boldsymbol{\Sigma}_{y|x}$, we allow cross-sectional correlations among the idiosyncratic components and assume that $\mathrm{tr}(\boldsymbol{\Sigma}_{y|x}^2) = O(p)$, which corresponds to the boundedness of eigenvalues of $\boldsymbol{\Sigma}_{y|x}$ given in [10]. Thus, case (i) holds if the factor loadings are dense, specifically, $\boldsymbol{B}^\top\boldsymbol{B} = O(p)$. On the other hand, when the maximal number of non-zero elements over the columns of the factor loadings $\boldsymbol{B}$ is uniformly bounded with respect to N and p, we have case (ii). Hence, when the factor loadings are dense, assumption (i) is reasonable, and when the factor loadings are sparse or not dense, assumption (ii) is appropriate. In this paper, we treat the two cases, depending on the density of the factor loadings.

Concerning the choice of loss function, we use the unnormalized loss $\mathrm{tr}(\boldsymbol{\delta} - \boldsymbol{\Sigma}_y)^2$ for the estimator $\boldsymbol{\delta}$ of $\boldsymbol{\Sigma}_y$ while Fan et al. [9], Fan et al. [10] mainly work with the normalized quadratic loss function $\mathrm{tr}(\boldsymbol{\delta}\boldsymbol{\Sigma}_y^{-1} - \boldsymbol{I})^2$. Fan et al. [11] noted that the unnormalized loss functions of many estimators diverge even if the growth rate of p is moderate (pp. 614–615). Because the unnormalized loss function is widely used, however, we treat it in this paper.

The remainder of this paper is organized as follows. In Section 2, we introduce the multivariate model and the factor model; we then express the factor model in the framework of the multivariate model. The double shrinkage estimators are introduced and the optimal weights are derived under appropriate assumptions for non-sparsity and sparsity of factor loadings. The mean squared errors of the estimators with optimal weights are provided. In Section 3, we give several estimators of the unknown parameters included in the optimal weights and suggest plug-in estimators by substituting the estimators for the optimal weights. Section 4 reports numerical studies, and Section 5 presents an application to portfolio management. Concluding remarks are given in Section 6, and technical proofs are provided in the Appendix.

2. Linear shrinkage estimators

2.1. Multivariate and factor models

Consider the following multivariate model: N random pairs $(\boldsymbol{y}_1, \boldsymbol{x}_1), \ldots, (\boldsymbol{y}_N, \boldsymbol{x}_N)$ are mutually independently and identically distributed as $\mathrm{E}(\boldsymbol{y}_i) = \boldsymbol{\mu}_y$, $\mathrm{E}(\boldsymbol{x}_i) = \boldsymbol{\mu}_x$, $\mathrm{E}\{(\boldsymbol{y}_i - \boldsymbol{\mu}_y)(\boldsymbol{y}_i - \boldsymbol{\mu}_y)^\top\} = \boldsymbol{\Sigma}_y$, $\mathrm{E}\{(\boldsymbol{x}_i - \boldsymbol{\mu}_x)(\boldsymbol{x}_i - \boldsymbol{\mu}_x)^\top\} = \boldsymbol{\Sigma}_x$ and $\mathrm{E}\{(\boldsymbol{y}_i - \boldsymbol{\mu}_y)(\boldsymbol{x}_i - \boldsymbol{\mu}_x)^\top\} = \boldsymbol{\Sigma}_{yx} = \boldsymbol{\Sigma}_{xy}^\top$, where $\boldsymbol{y}_i$ and $\boldsymbol{x}_i$ are, respectively, p- and q-dimensional vectors, namely

$$\begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{x}_i \end{pmatrix} \sim \text{ i.i.d.} \left[\begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_x \end{pmatrix} \right]. \tag{2.1}$$

Let

$$\boldsymbol{S}_y = \frac{1}{n}\sum_{i=1}^{N}(\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^\top, \qquad \boldsymbol{S}_x = \frac{1}{n}\sum_{i=1}^{N}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top$$