数理解析研究所講究録
第1954巻2015年32-43

Improved linear shrinkage estimators of large structured covariance matrices

池田祐樹*, 久保川達也†
東京大学
Yuki Ikeda and Tatsuya Kubokawa
University of Tokyo

Abstract

The problem of estimating large covariance matrices with use of factor models is addressed. In this article, 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. Some asymptotic arguments are given. Numerical results show our methods perform well under both normal and non-normal distributions.

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 研究の背景と概要

1.1 研究の背景 (1); 高次元共分散行列

- 高次元ベクトル $\boldsymbol{y}_i$ の分散共分散行列 $\boldsymbol{\Sigma}_{11} := \mathbf{Cov}\,(\boldsymbol{y}_i)$ の推定問題:
  次元 p がサンプル数 N を超えるとき逆行列存在せず.
  $p < N$ でも p が相対的に大きいと不安定.

- 既存研究：多数の文献 (ここでは特に２つ列挙)

*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

- **weighted estimators**(ridge-type estimators);
  Ledoit and Wolf (2004), Schafer and Strimmer (2005), Chen, Wiesel, Eldar and Hero (2010), Kubokawa and Srivastava (2013),...

$$\widehat{\boldsymbol{\Sigma}}_T^s := w\widehat{\boldsymbol{\Sigma}}_{11} + (1-w)\frac{1}{p}\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11})\boldsymbol{I}, \tag{1.1}$$

- **regularization/thresholding estimators**;
  Bickel and Levina (2008a, 2008b), Rothman, Levina and Zhu (2009), Cai and Liu (2011), ...
- 他にも Bayse 推定量, 修正コレスキー分解, 行列指数関数,...

1.2 研究の背景 (2); ファクターモデル

- その分散共分散行列に興味のある変数 $\boldsymbol{y}_i$ 以外に共変量 $\boldsymbol{x}_i$ が利用可能な場合 : (線形) ファクターモデル

- 3-ファクターモデル(Fama and French (1993)):

$$r_{it} = b_{i1}f_{1t} + b_{i2}f_{2t} + b_{i3}f_{3t} + u_{it},$$

  r_{it}:excess asset return, f_{1t}:sensitivity to the market excess return, f_{2t}:market capitalization, f_{3t}:book-to-price ratio.

- ファクターモデルにおける高次元共分散行列の推定:
  Fan, Fan and Lv (2008)

- 近似ファクターモデルにおける高次元共分散行列の推定:
  Fan, Liao and Mincheva (2011, 2013)

- これら Fan, *et al.* (2011, 2013) は, Error covariance matrix に thresholding を適用 (Error covariance matrix にスパース性の仮定).

1.3 研究の背景 (3); weighted estimators とファクターモデル

- Ledoit and Wolf (2003), Ren and Shimotsu (2009):

  - p:固定, $N \to \infty$(非高次元) で, (1) 標本分散共分散行列と (2) ファクターモデルから示唆される推定量の線形結合を提案.

- 本発表 : $p, N \to \infty$ の高次元の枠組みで, ファクターモデルを利用した線形縮小推定量 (linear shrinkage, weighted or ridge-type estimators) を提案.

- 基本的なアイデアは, Ledoit and Wolf (2003), Ren and Shimotsu (2009) と同様:
  標本分散共分散行列とファクターモデルが成り立つもとで示唆される推定量の線形結合全体からなるクラスを考え, 最適な weight を求め, そのもとでのリスクを与えた.

- 数値実験の結果, 正規・非正規分布下で, 提案推定量は分散共分散行列だけでなく逆行列の推定でも, 標本共分散行列や Fan, *et al.* (2011) を含む他の推定量に比べ, 良い結果を与えた.

2 Weighted Estimators とその MSE

2.1 Unrestricted model and Factor model

- $\boldsymbol{y}_i \in \mathbb{R}^p$ および $\boldsymbol{y}_i$ と相関のある共変量 $\boldsymbol{x}_i \in \mathbb{R}^q$ が次のように生成されているとする:

$$\boldsymbol{y}_i \sim \text{ i.i.d.}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}), \quad \boldsymbol{x}_i \sim \text{ i.i.d.}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}). \tag{2.1}$$

ここで $\boldsymbol{y}_i$ に関して $\boldsymbol{x}_i$ によるファクターモデルが成り立っていると仮定する:

$$\begin{aligned} \boldsymbol{y}_i =& \boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{x}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, N, \\ & \boldsymbol{\epsilon}_i \sim \text{ i.i.d.}(\boldsymbol{0}, \boldsymbol{D}), \quad \boldsymbol{x}_i \sim \text{ i.i.d.}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}), \end{aligned} \tag{2.2}$$

$\boldsymbol{\Sigma}_{12} = \mathbf{Cov}\,(\boldsymbol{y}_i, \boldsymbol{x}_i)$ とおくと, 上式は次のように書き直すことができる.

$$\begin{aligned} \boldsymbol{y}_i =& (\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2) + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{x}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, N, \\ & \boldsymbol{\epsilon}_i \sim \text{ i.i.d.}(\boldsymbol{0}, \boldsymbol{D}), \quad \boldsymbol{x}_i \sim \text{ i.i.d.}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}), \end{aligned} \tag{2.3}$$

- 厳密なファクターモデル (strict factor model) では $\boldsymbol{D} = d\boldsymbol{I}_p$ (sphericity) または $\boldsymbol{D} = \mathrm{diag}\,(d_1, \dots, d_p)$(diagonality) を想定.
- すると上のファクターモデルが成り立つならば,

$$\boldsymbol{\Sigma}_{11} := \mathbf{Cov}\,(\boldsymbol{y}_i) = \boldsymbol{D} + \boldsymbol{\beta}^T\boldsymbol{\Sigma}_{22}\boldsymbol{\beta} = \boldsymbol{D} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

2.2 Weighted Estimators

- $n = N - 1$, $\overline{\boldsymbol{y}} = N^{-1}\sum_{i=1}^N \boldsymbol{y}_i$, $\overline{\boldsymbol{x}} = N^{-1}\sum_{i=1}^N \boldsymbol{x}_i$, $\widehat{\boldsymbol{\Sigma}}_{11} = \sum_{i=1}^N (\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^T/n$, $\widehat{\boldsymbol{\Sigma}}_{12} = \widehat{\boldsymbol{\Sigma}}_{21}^T = \sum_{i=1}^N (\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T/n$, $\widehat{\boldsymbol{\Sigma}}_{22} = \sum_{i=1}^N (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T/n$, $\widehat{\boldsymbol{\Sigma}}_{11.2} = \widehat{\boldsymbol{\Sigma}}_{11} - \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}$.
- $\boldsymbol{\Sigma}_{11}$ の推定量 $\boldsymbol{\delta}$ は, 以下の mean squared error (MSE) のもとで評価する:

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

ただし ω は未知のパラメータの組.

- $\boldsymbol{x}_i$ をもちいない $\boldsymbol{\Sigma}_{11}$ の最も標準的な推定量は $\widehat{\boldsymbol{\Sigma}}_{11}$.

- 一方 $\boldsymbol{x}_i$ による厳密なファクターモデルが推測されるときには, 以下の推定量が自然:

$$\widehat{\boldsymbol{\Sigma}}_f = \boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2}) + \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}. \tag{2.4}$$

ここで $\boldsymbol{\Lambda}$ は $\boldsymbol{D}$ の推定量であり,

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

- 近似ファクターモデルまで考慮するため, 分散共分散行列 $\widehat{\boldsymbol{\Sigma}}_{11}$ を, 厳密なファクターモデルにおける自然な推定量 $\widehat{\boldsymbol{\Sigma}}_f$ の方向に縮小した以下のような推定量を考える.

$$\widehat{\boldsymbol{\Sigma}}_\alpha = \alpha\widehat{\boldsymbol{\Sigma}}_{11} + (1-\alpha)\widehat{\boldsymbol{\Sigma}}_f, \quad \alpha \in [0,1]. \tag{2.5}$$

すると以下のように書き直せる:

$$\widehat{\boldsymbol{\Sigma}}_\alpha = \alpha\widehat{\boldsymbol{\Sigma}}_{11.2} + (1-\alpha)\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2}) + \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}. \tag{2.6}$$

- $\boldsymbol{\Sigma}_{11.2}$ の sphericity あるいは daigonality の度合いが小さくなるほど α は 1 に近づき, $\widehat{\boldsymbol{\Sigma}}_\alpha$ は単なる標本共分散行列に退化していってしまう。そこで:以下のようなより広いクラスを考える

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}(\gamma,\beta) = \gamma\widehat{\boldsymbol{\Sigma}}_{11.2} &+ (1-\gamma)\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11.2})\\ &+ \beta\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21} + (1-\beta)\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}).\end{aligned} \tag{2.7}$$

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

- すると $\widehat{\boldsymbol{\Sigma}}_\alpha$ は, $\gamma = \alpha$ かつ $\beta = 1$ に対応する.
- とくに (C1)(sphericity) のケースでは, $\gamma = \beta = w$ とおくと, Kubokawa and Srivastava(2013) で扱われている共変量をもちいない weighted estimator

$$\widehat{\boldsymbol{\Sigma}}_T = w\widehat{\boldsymbol{\Sigma}}_{11} + (1-w)\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11})\boldsymbol{I} \tag{2.8}$$

になる.

(C1) Case of sphericity: $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11}) = p^{-1}\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11})\boldsymbol{I}$,

(C2) Case of diagonality: $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\Sigma}}_{11}) = \mathrm{diag}(\widehat{\boldsymbol{\Sigma}}_{11})$.

- 以下では上の推定量のクラス (2.7) において, MSE を最小にする β と γ を求める.

2.3 risk function の近似

- ケース (C1) または (C2) のもとで, $\widehat{\mathbf{\Sigma}}(\gamma,\beta)$ のリスクは以下のように表される:

$$R(\omega,\widehat{\mathbf{\Sigma}}(\gamma,\beta)) = \begin{pmatrix}\gamma & \beta\end{pmatrix}\begin{pmatrix}J_{11} & J_{12}\\ J_{12} & J_{22}\end{pmatrix}\begin{pmatrix}\gamma\\ \beta\end{pmatrix} - 2\begin{pmatrix}J_{10} & J_{20}\end{pmatrix}\begin{pmatrix}\gamma\\ \beta\end{pmatrix} + R_0, \tag{2.9}$$

したがって最適なウエイト γ^*,β^* は

$$\begin{pmatrix}\gamma^*\\ \beta^*\end{pmatrix} = \begin{pmatrix}J_{11} & J_{12}\\ J_{12} & J_{22}\end{pmatrix}^{-1}\begin{pmatrix}J_{10}\\ J_{20}\end{pmatrix}. \tag{2.10}$$

で与えられ, これら γ^*,β^* をウエイトに持つ推定量 $\widehat{\mathbf{\Sigma}}(\gamma,\beta)$ のリスクは

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

- ただし,

(C1) Case of sphericity: $\mathbf{\Lambda}(\mathbf{\Sigma}_{11}) = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11})\boldsymbol{I}$,

(C2) Case of diagonality: $\mathbf{\Lambda}(\mathbf{\Sigma}_{11}) = \mathrm{diag}(\mathbf{\Sigma}_{11})$,

として,

$$\begin{aligned}
R_0 =& \frac{1}{p}E\mathrm{tr}(\mathbf{\Sigma}_{11} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11}))^2, \quad J_{11} = \frac{1}{p}E\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{11.2} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2}))^2,\\
J_{22} =& \frac{1}{p}E\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21}))^2,\\
J_{12} =& \frac{1}{p}E\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{11.2} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2}))(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21})),\\
J_{10} =& \frac{1}{p}E\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{11.2} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{11.2}))(\mathbf{\Sigma}_{11} - \mathbf{\Lambda}(\mathbf{\Sigma}_{11})),\\
J_{20} =& \frac{1}{p}E\mathrm{tr}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21} - \mathbf{\Lambda}(\widehat{\mathbf{\Sigma}}_{12}\widehat{\mathbf{\Sigma}}_{22}^{-1}\widehat{\mathbf{\Sigma}}_{21}))(\mathbf{\Sigma}_{11} - \mathbf{\Lambda}(\mathbf{\Sigma}_{11})).
\end{aligned} \tag{2.11}$$

- 以下, $\mathbf{\Sigma}_{11.2} := \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$ とする.
- 以下では簡単のため, ケース **(C1)**:sphericity のときのみ考える.
- 推定量のリスクを近似するにあたって, 以下の **(A1)(A2)** を仮定し, さらに **(A3)** または (A4) のどちらかを仮定する:

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

(A2) $a_1 = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11}) = O(1)$, $b_1 = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11.2}) = O(1)$, $b_2 = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11.2}^2) = O(1)$ and $\phi_{11} = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11}\mathbf{\Sigma}_{11.2}) = O(1)$.

(A3) $a_2 = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11}^2) = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11.2} + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})^2 = O(p)$.

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

- $\boldsymbol{x}_i$, ϵ_i(したがって $\boldsymbol{y}_i$) に正規性を仮定すれば R_0, J_{kl} が解析的に求まる.
- ケース **(C1)** のときには, 仮定 **(A3)** または (A4) の下で,

$$\begin{aligned}J_{11} =&b_2 - b_1^2 + \frac{p}{n}b_1^2 + O(n^{-1}) + O(n^{-2}p),\\ J_{12} =&\phi_{11} - a_1b_1 - b_2 + b_1^2 + O(n^{-1}),\\ J_{10} =&\phi_{11} - a_1b_1 + O(n^{-1}),\\ J_{20} =&a_2 - a_1^2 - \phi_{11} + a_1b_1 + O(n^{-1}).\end{aligned}$$

であり, 仮定 **(A3)** のときは,

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

仮定 (A4) のときには,

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

- **Risk of $\widehat{\boldsymbol{\Sigma}}_{11}$:** 簡単な計算から $Risk(\widehat{\boldsymbol{\Sigma}}_{11}) = a_2/n + (p/n)a_1^2$.
- **Risk of $\widehat{\boldsymbol{\Sigma}}_{\alpha}$:** ケース **(C1)**:sphericity のとき, (2.9) 式で $\gamma = \alpha$, $\beta = 1$ とおいて前ページの式をもちいると

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

となりこれを最小化する α は

$$\alpha^* = \frac{b_2 - b_1^2}{b_2 - b_1^2 + (p/n)b_1^2}. \tag{2.12}$$

$\widehat{\boldsymbol{\Sigma}}_{\alpha^*}$ のリスクは

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

これは leading terms において $\widehat{\Sigma}_{11}$ より小さい.

3 Plug-In 推定量の構成

3.1 Plug-In 推定量

- ここまで $\boldsymbol{\Sigma}_\alpha$, $\widehat{\boldsymbol{\Sigma}}(\gamma, \beta)$ の最適なウエイトを評価してきたが, 実際に推定を行うにはこれらのウエイトを推定する必要がある.

- a_1, a_2, b_1, b_2 に関しては以下の各推定量をもちいる (Srivastava (2005)):

$$\begin{aligned}
\hat{a}_1 &= \frac{1}{p}\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11}), \quad \hat{a}_{20} = \frac{n}{p(n+2)}\mathrm{tr}\,((\mathrm{diag}\widehat{\boldsymbol{\Sigma}}_{11})^2), \\
\hat{a}_2 &= \frac{n^2}{p(n-1)(n+2)}\Big(\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11}^2) - (\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}_{11})^2/n\Big), \\
\hat{b}_1 &= \frac{n}{p(n-q)}\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11}), \quad \hat{b}_{20} = \frac{n^2}{p(n-q)(n-q+2)}\mathrm{tr}\,((\mathrm{diag}\widehat{\boldsymbol{\Sigma}}_{11.2})^2), \\
\hat{b}_2 &= \frac{n^2}{p(n-q-1)(n-q+2)}\Big(\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11.2}^2) - (\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}_{11.2})^2/(n-q)\Big),
\end{aligned} \tag{3.1}$$

- また ϕ_{11} に関しては以下をもちいる:

$$\begin{aligned}
\widehat{\phi}_{11} = \frac{n}{p(n-q)}\Big(\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11}\widehat{\boldsymbol{\Sigma}}_{11.2}) &- \frac{n-q-2}{(n-q-1)(n-q+2)}\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_{11.2}^2) \\
&- \frac{n-q}{(n-q-1)(n-q+2)}(\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}_{11.2})^2\Big).
\end{aligned}$$

- これらを (2.12) に代入することにより, Plug-In 推定量 $\boldsymbol{\Sigma}_{\hat{\alpha}^*}$ を提案する. またこれらを 15 ページの各 J の leading terms に代入し、さらに (2.10) に代入することにより, $\widehat{\boldsymbol{\Sigma}}(\hat{\gamma}^*, \hat{\beta}^*)$ を提案する.

3.2 一致性

- さらに以下の条件を仮定し, 上述の推定量の一致性に関して議論する:

(A5) $a_4 = p^{-1}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11}^4) = O(p^3)$, $b_3 = p^{-1}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}^3) = O(1)$, $b_4 = p^{-1}\mathrm{tr}\,(\boldsymbol{\Sigma}_{11.2}^4) = O(1)$,

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

Theorem 3.1. *Assume* (A1), (A2), (A3) *and* (A5), *then under normality,* $\hat{a}_1$, $\hat{a}_2$, $\hat{b}_1$, $\hat{b}_2$ *and* $\widehat{\phi}_{11}$ *are all unbiased and*

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

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

4 数値実験

4.1 設定

- データ生成過程:

$$\boldsymbol{y}_i = \boldsymbol{\alpha} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{x}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i = \boldsymbol{\Sigma}_{11.2}^{1/2}\boldsymbol{u}_i, \quad \boldsymbol{x}_i = \boldsymbol{\Sigma}_{22}^{1/2}\boldsymbol{v}_i$$

  ここで $\boldsymbol{u}_i = (u_{ij})_{1\le j\le p}$, $\boldsymbol{v}_i = (v_{ij})_{1\le j\le q}$ は互いに独立とする.

- 正規分布 (D1) $u_{ij}, v_{ij} \sim N(0,1)$
  非正規分布 (D2) $u_{ij}, v_{ij} = (w_{ij} - \nu)/\sqrt{2\nu}, \quad w_{ij} \sim \chi^2(\nu)$ かつ $\nu = 2$.

- $\boldsymbol{\mu}_2 = \boldsymbol{0}$, $\boldsymbol{\Sigma}_{22} = \boldsymbol{I}_q$, $(\boldsymbol{\Sigma}_{12})_{ij} \sim$ i.i.d. $\mathcal{N}(0.5, 1)$.

- 厳密なファクターモデル (M1):$\boldsymbol{\Sigma}_{11.2} = 3\boldsymbol{I}$,
  近似ファクターモデル (M2):

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

  ただし $\sigma_j = 3 + 0.2(-1)^{j-1}(p - i + 1)/p$ かつ $\rho = 0.2$.

- 7つの推定量 $\widehat{\boldsymbol{\Sigma}}_{11}$, $\widehat{\boldsymbol{\Sigma}}_T^s$, $\widehat{\boldsymbol{\Sigma}}_T^d$, $\widehat{\boldsymbol{\Sigma}}_{\hat{\alpha}^*}^s$, $\widehat{\boldsymbol{\Sigma}}_{\hat{\alpha}^*}^d$, $\widehat{\boldsymbol{\Sigma}}^s(\hat{\gamma}^*, \hat{\beta}^*)$, $\widehat{\boldsymbol{\Sigma}}^d(\hat{\gamma}^*, \hat{\beta}^*)$ [1]を比較. 各々1000回繰り返し, 次のロスで計算される経験リスクと対応する標準偏差を計算した.

- $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{11}^{-1}$ の推定に関して, それぞれ以下で評価.

$$p^{-1}\mathrm{tr}\left[(\boldsymbol{\delta} - \boldsymbol{\Sigma}_{11})^2\right] \ for\ \boldsymbol{\Sigma}_{11},$$
$$p^{-1}\mathrm{tr}\left[(\boldsymbol{\delta}\boldsymbol{\Sigma}_{11} - \boldsymbol{I})^2\right] \ for\ \boldsymbol{\Sigma}_{11}^{-1}.$$

- $N = 100$, $q = 3$ と固定し、$p = 50$, 100, 200 とした。

4.2 Numerical Results

末尾の Table 1-4 に示している.

[1]添え字の s, d はそれぞれ (C1):sphericity, (C2):diagonality を示す. $\widehat{\boldsymbol{\Sigma}}_T$ は (2.8) 式で与えられている

5 結論

5.1 議論

- リスクの近似の計算で、$O_p(n^{-2}p)$ の項が出てくるため、この項が無視できないほど p が大きいときは、誤差が生じてくる。
  - ファクターローディング (factor loadings)$\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}$(したがって $a_2 = \mathrm{tr}\,\mathbf{\Sigma}_{11}^2/p$) の密度が $\hat{a}_2$ の収束レートに深くかかわっている.
  - (A4)$a_2 = O(1)$ のときは仮定 (A5) において $a_4 = p^{-1}\mathrm{tr}\,(\mathbf{\Sigma}_{11}^4) = O(1)$ と修正しても非整合的ではなく, $(np)^{-1/2} \to 0$ のとき $\hat{a}_2$ は一致性をもつ.
- (A3)$\mathrm{tr}\,\mathbf{\Sigma}^2 = O(p^2)$ のもとでは漸近的には $|\gamma^* - \alpha^*| \to 0$, $\beta^* \to 1$ となるので, 漸近的に $\mathbf{\Sigma}_{\gamma,\beta}$ と $\mathbf{\Sigma}_\alpha$ のリスクは等しくなる。
  (A4)$\mathrm{tr}\,\mathbf{\Sigma}^2 = O(p)$ のもとでは漸近的にも $|\gamma^* - \alpha^*| \nrightarrow 0$, $\beta^* \nrightarrow 1$ となり, 漸近的にも $\mathbf{\Sigma}_\alpha$ に対する $\mathbf{\Sigma}_{\gamma,\beta}$ のリスクの改善がみられる.
- 縮小ターゲット:$(\mathrm{tr}\,\widehat{\mathbf{\Sigma}}/p)\boldsymbol{I}_p$, $\mathrm{diag}\,\widehat{\mathbf{\Sigma}}$,...

5.2 結論

- 本発表では, 高次元の枠組みにおいて $\boldsymbol{y}_i$ の共変量 $\boldsymbol{x}_i$ が得られるときに, $\boldsymbol{y}_i$ の標本共分散行列 $\widehat{\mathbf{\Sigma}}_{11}$ を $\boldsymbol{x}_i$ によるファクターモデルを仮定したもとで示唆される推定量 $\widehat{\mathbf{\Sigma}}_f$ の方向へ縮小したクラスの推定量を考え, さらにそれを含むより一般化した推定量 $\widehat{\mathbf{\Sigma}}(\gamma,\beta)$ のリスクを Mean Squared Error の下で調べた.
- 数値計算の結果, $\boldsymbol{y}_i$, $\boldsymbol{x}_i$ の正規性のあるなしに関わらず, 厳密な/近似ファクターモデルのもとで、$\widehat{\mathbf{\Sigma}}(\gamma,\beta)$ は分散共分散行列, 逆行列の推定で, ともに良い結果を与えている.
- 同じく数値実験により, ポートフォリオの分散上限制約下で期待リターンの最大化問題を考えたところ, 提案する推定量が制約を守りやすいことがわかった.
- 全体に正規性に依拠した枠組みであることから、非正規の場合の推定向上が課題である。

References

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

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

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

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

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

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

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

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

- Kubokawa, T., and Srivastava, M.S. (2013). Optimal ridge-type estimators of covariance matrix in high dimension. *Discussion Paper Series*, CIRJE-F-906.

- 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.

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

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

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

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

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

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

p	dist.	$\widehat{\mathbf{\Sigma}}_{11}$	$\widehat{\mathbf{\Sigma}}_T^s$	$\widehat{\mathbf{\Sigma}}_T^d$	$\widehat{\mathbf{\Sigma}}_\alpha^s$	$\widehat{\mathbf{\Sigma}}_\alpha^d$	$\widehat{\mathbf{\Sigma}}^s(\gamma,\beta)$	$\widehat{\mathbf{\Sigma}}^d(\gamma,\beta)$
50	(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	(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)
200	(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)
50	(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	(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)
200	(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)

p	dist.	$\widehat{\mathbf{\Sigma}}_{11}$	$\widehat{\mathbf{\Sigma}}_T^s$	$\widehat{\mathbf{\Sigma}}_T^d$	$\widehat{\mathbf{\Sigma}}_\alpha^s$	$\widehat{\mathbf{\Sigma}}_\alpha^d$	$\widehat{\mathbf{\Sigma}}^s(\gamma,\beta)$	$\widehat{\mathbf{\Sigma}}^d(\gamma,\beta)$
50	(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	(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)
200	(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)
50	(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	(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)
200	(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)

p	dist.	$\widehat{\mathbf{\Sigma}}_{11}$	$\widehat{\mathbf{\Sigma}}_T^s$	$\widehat{\mathbf{\Sigma}}_T^d$	$\widehat{\mathbf{\Sigma}}_\alpha^s$	$\widehat{\mathbf{\Sigma}}_\alpha^d$	$\widehat{\mathbf{\Sigma}}^s(\gamma,\beta)$	$\widehat{\mathbf{\Sigma}}^d(\gamma,\beta)$
50	(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	(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)
200	(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)
50	(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	(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)
200	(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)

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

p	dist.	$\widehat{\mathbf{\Sigma}}_{11}$	$\widehat{\mathbf{\Sigma}}_T^s$	$\widehat{\mathbf{\Sigma}}_T^d$	$\widehat{\mathbf{\Sigma}}_\alpha^s$	$\widehat{\mathbf{\Sigma}}_\alpha^d$	$\widehat{\mathbf{\Sigma}}^s(\gamma,\beta)$	$\widehat{\mathbf{\Sigma}}^d(\gamma,\beta)$
50	(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	(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)
200	(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)
50	(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	(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)
200	(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)