AStA Advances in Statistical Analysis (2021) 105:601–627
https://doi.org/10.1007/s10182-021-00396-7

ORIGINAL PAPER

Positive-definite modification of a covariance matrix by minimizing the matrix ℓ_∞ norm with applications to portfolio optimization

Seonghun Cho[1] · Shota Katayama[2] · Johan Lim[1] · Young-Geun Choi[3]

Received: 27 May 2020 / Accepted: 17 February 2021 / Published online: 13 March 2021

Abstract

The covariance matrix, which should be estimated from the data, plays an important role in many multivariate procedures, and its positive definiteness (PDness) is essential for the validity of the procedures. Recently, many regularized estimators have been proposed and shown to be consistent in estimating the true matrix and its support under various structural assumptions on the true covariance matrix. However, they are often not PD. In this paper, we propose a simple modification to make a regularized covariance matrix be PD while preserving its support and the convergence rate. We focus on the matrix ℓ_∞ norm error in covariance matrix estimation because it could allow us to bound the error in the downstream multivariate procedure relying on it. Our proposal in this paper is an extension of the fixed support positive-definite (FSPD) modification by Choi et al. (2019) from spectral and Frobenius norms to the matrix ℓ_∞ norm. Like the original FSPD, we consider a convex combination between the initial estimator (the regularized covariance matrix without PDness) and a given form of the diagonal matrix minimize the ℓ_∞ distance between the initial estimator and the convex combination, and find a closed-form expression for the modification. We apply the procedure to the minimum variance portfolio (MVP) optimization problem and show that the vector ℓ_∞ error in the estimation of the optimal portfolio weight is bounded by the matrix ℓ_∞ error of the plug-in covariance matrix estimator. We illustrate the MVP results with S&P 500 daily returns data from January 1978 to December 2014.

Keywords High-dimensional covariance matrix · Linear shrinkage · Matrix ℓ_∞ norm · Minimum variance portfolio · Positive definiteness · Regularized covariance matrix estimator

✉ Johan Lim
johanlim@snu.ac.kr

[1] Department of Statistics, Seoul National University, Seoul, Republic of Korea

[2] Faculty of Economics, Keio University, Tokyo, Japan

[3] Department of Statistics, Sookmyung Women's University, Seoul, Republic of Korea

1 Introduction

The covariance matrix plays an important role in many multivariate procedures, and its positive definiteness (PDness) is essential for the validity of those methods. Recently, as high-dimensional data have become more commonplace, many regularized covariance matrix estimators have been proposed under various structural assumptions on the true matrix (Bickel and Levina 2008b, a; Cai et al. 2010; Cai and Liu 2011; Cai and Zhou 2012b; Won et al. 2013; Khare et al. 2015). The regularized estimators are well understood in their asymptotic theory and, in particular, are known to be consistent in estimating the true covariance matrix and its support. However, these estimators are often not PD in finite samples; for example, the banding and thresholding methods in the literature elementwisely regularize the sample covariance matrix, ignoring its positive-semidefinite eigenstructure.

Some efforts have been made to resolve the difficulty of non-PDness. Rothman (2012), Xue et al. (2012), and Liu et al. (2014) find an estimator that attains both sparsity and PDness by incorporating these requirements into a single optimization problem. The authors rewrite the soft-thresholding of the sample covariance matrix as a solution to an ℓ_1 regularized convex minimization problem and add an additional penalty (or constraint) to ensure the PDness of the solution, which is computationally demanding. Instead, Choi et al. (2019b) propose a two-stage approach to update the non-PD regularized estimators to recover their PDness while retaining the same supports and asymptotic properties. Specifically, for a given covariance matrix estimator $\hat{\boldsymbol{\Sigma}}$, they consider a distance minimization program

$$\begin{aligned}
&\underset{\hat{\boldsymbol{\Sigma}}^*}{\text{minimize}}\left\{ \|\hat{\boldsymbol{\Sigma}}^* - \hat{\boldsymbol{\Sigma}}\| \,:\, \gamma_1(\hat{\boldsymbol{\Sigma}}^*) \geq \epsilon,\ \mathrm{supp}(\hat{\boldsymbol{\Sigma}}^*) = \mathrm{supp}(\hat{\boldsymbol{\Sigma}}),\ \hat{\boldsymbol{\Sigma}}^* = (\hat{\boldsymbol{\Sigma}}^*)^\top \right\}, \\
&\text{subject to } \hat{\boldsymbol{\Sigma}}^* \in \left\{ \boldsymbol{\Phi}_{\mu,\alpha}(\hat{\boldsymbol{\Sigma}}) \equiv \alpha\hat{\boldsymbol{\Sigma}} + (1-\alpha)\mu\,\mathbf{I} \,:\, \alpha \in [0,1],\ \mu \in \mathbb{R} \right\},
\end{aligned} \tag{1}$$

where $\|\cdot\|$ is a matrix norm, $\epsilon > 0$ is a predetermined small constant and $\gamma_1(\hat{\boldsymbol{\Sigma}}^*)$ denotes the smallest eigenvalue of $\hat{\boldsymbol{\Sigma}}^*$. They solve (1) for two matrix norms: the spectral norm and Frobenius norm. However, the solvability of (1) for the matrix ℓ_∞ norm is still unknown in the literature. It is particularly important to retaining the matrix ℓ_∞ norm (equivalently, the matrix ℓ_1 norm for symmetric matrices) convergence rate of the initial estimator, because it could allow us to bound the error in the downstream multivariate procedure relying on it (as shown in this paper).

In this paper, we solve the PD modification problem (1) for the matrix ℓ_∞ norm, where we find a closed-form expression. Like the original FSPD, the modified PD estimator, denoted by 'ℓ_∞-FSPD', exhibits several advantages. First, the computation of the ℓ_∞-FSPD is optimization-free since the choices of μ and α are explicitly expressible. Second, suitable choices of μ and α in (1) make the ℓ_∞-FSPD estimator and the initial estimator have the same rate of convergence to the true covariance matrix under some mild conditions. Finally, the ℓ_∞-FSPD procedure is generic in the sense that it is applicable to any (possibly) non-PD estimator of covariance matrix, as well as precision matrix (the inverse of the covariance matrix).

The use of the matrix ℓ_∞ norm in the estimation of the covariance matrix often provides a better understanding of the performance of the downstream multivariate procedures using it. In this paper, we apply the proposed procedure to the minimum variance portfolio (MVP) optimization problem. We show that, under some conditions on the true covariance matrix, the vector ℓ_∞ error in the estimation of the optimal portfolio weight is bounded by the matrix ℓ_∞ error of the plug-in covariance matrix estimator. We consider the universal thresholding estimator by Bickel and Levina (2008a) or Cai and Liu (2011) as an initial estimator of ℓ_∞-FSPD and have an estimator of the optimal portfolio weight whose vector ℓ_∞ error is well understood. We illustrate the MVP results with S&P 500 daily returns data from January 1978 to December 2014.

The paper is organized as follows. In Sect. 2, we solve the distance minimization problem (1) for the matrix ℓ_∞ norm and find an analytical solution to it. In Sect. 3, we show that the proposed ℓ_∞-FSPD estimator preserves the ℓ_∞ norm convergence rate of the original estimator. In Sect. 4, we apply the ℓ_∞-FSPD estimator to the MVP optimization problem. We show that the rate of the plug-in estimator of the optimal portfolio is bounded by the ℓ_∞ norm error of the covariance matrix estimator. In Sect. 5, we apply the MVP results to S&P 500 data and compare the MVP with the ℓ_∞ norm-based solution to existing approaches. Finally, in Sect. 6, we conclude the paper with discussion.

2 FSPD with matrix ℓ_∞ norm (ℓ_∞-FSPD)

2.1 Distance minimization

For a predetermined small positive number $\epsilon > 0$, if the smallest eigenvalue $\widehat{\gamma}_1$ of the estimate $\widehat{\mathbf{\Sigma}}$ of the covariance matrix $\mathbf{\Sigma}$ is not less than ϵ, we do not need to modify $\widehat{\mathbf{\Sigma}}$. Hence, we only consider the case in which $\widehat{\gamma}_1$ is less than ϵ. Our purpose is to modify $\widehat{\mathbf{\Sigma}}$ so that its smallest eigenvalue is larger than or equal to ϵ while minimizing the distance between the modified estimate and the original one. Here, we consider a class of linear shrinkage as a modification of $\widehat{\mathbf{\Sigma}}$:

$$\left\{ \mathbf{\Phi}_{\mu,\alpha}(\widehat{\mathbf{\Sigma}}) \equiv \alpha\widehat{\mathbf{\Sigma}} + (1-\alpha)\mu\mathbf{I} \,:\, \mu \in \mathbb{R}, \alpha \in [0,1] \right\}.$$

Now, we must solve the following minimization problem.

$$\begin{aligned} \underset{\mu\in\mathbb{R},\alpha\in[0,1]}{\text{minimize}} \quad & \|\mathbf{\Phi}_{\mu,\alpha}(\widehat{\mathbf{\Sigma}}) - \widehat{\mathbf{\Sigma}}\|_\infty \\ \text{subject to} \quad & \alpha\widehat{\gamma}_1 + (1-\alpha)\mu \geq \epsilon \end{aligned} \tag{2}$$

where for a symmetric $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{p\times p}$, $\|\mathbf{A}\|_\infty = \max_{1\leq i\leq p}\sum_{j=1}^{p}|a_{ij}|$. The assumption $\widehat{\gamma}_1 < \epsilon$ and the constraint $\alpha\widehat{\gamma}_1 + (1-\alpha)\mu \geq \epsilon$ imply that μ should be larger than ϵ.

2.2 The choice of α

As in Choi et al. (2019b), we observe that

$$\|\boldsymbol{\Phi}_{\mu,\alpha}(\widehat{\boldsymbol{\Sigma}}) - \widehat{\boldsymbol{\Sigma}}\|_\infty = (1-\alpha)\|\mu\mathbf{I} - \widehat{\boldsymbol{\Sigma}}\|_\infty.$$

Thus, we have the following lemma.

Lemma 1 *Assume that $\widehat{\boldsymbol{\Sigma}}$ is given and $\widehat{\gamma}_1 < \epsilon$. Then, for any fixed $\mu \in (\epsilon, \infty)$,*

$$\alpha^*(\mu) = \frac{\mu - \epsilon}{\mu - \widehat{\gamma}_1} \tag{3}$$

minimizes the problem (2).

2.3 The choice of μ

By substituting (3) into (2), we obtain a reduced minimization problem, which depends only on μ,

$$\underset{\mu:\mu>\epsilon}{\text{minimize}}\ \|\boldsymbol{\Phi}_{\mu,\alpha^*(\mu)}(\widehat{\boldsymbol{\Sigma}}) - \widehat{\boldsymbol{\Sigma}}\|_\infty. \tag{4}$$

Let $S(\mu)$ be the objective function of the minimization problem (4). Then, we observe

$$\begin{aligned} S(\mu) &= \|\boldsymbol{\Phi}_{\mu,\alpha^*(\mu)}(\widehat{\boldsymbol{\Sigma}}) - \widehat{\boldsymbol{\Sigma}}\|_\infty \\ &= \frac{\epsilon - \widehat{\gamma}_1}{\mu - \widehat{\gamma}_1}\|\mu\mathbf{I} - \widehat{\boldsymbol{\Sigma}}\|_\infty = \frac{\epsilon - \widehat{\gamma}_1}{\mu - \widehat{\gamma}_1} \max_{1\le j\le p} \sum_{i=1}^{p} \left|\mu\delta_{ij} - \widehat{\sigma}_{ij}\right| \\ &= \frac{\epsilon - \widehat{\gamma}_1}{\mu - \widehat{\gamma}_1} \max_{1\le j\le p} \left\{ \left|\mu - \widehat{\sigma}_{jj}\right| + \sum_{i:i\ne j} \left|\widehat{\sigma}_{ij}\right| \right\}, \end{aligned} \tag{5}$$

where $\delta_{ij} = 1$ if $i = j$, and is 0 otherwise. To simplify the objective function (5), we introduce the following lemma.

Lemma 2 *For $a_i, b_i \in \mathbb{R}$ $(i = 1, \dots, n)$, define a function $f : \mathbb{R} \to \mathbb{R}$ by*

$$f(x) = \max_{1\le i\le n} \left\{ \left|x - a_i\right| + b_i \right\}.$$

Let $M_1 = \max_{1\le i\le n}(a_i + b_i)$ and $M_2 = \max_{1\le i\le n}(-a_i + b_i)$. Then, f can be rewritten as

$$f(x) = |x - A| + B$$

where $A = (M_1 - M_2)/2$ and $B = (M_1 + M_2)/2$.

Proof We use mathematical induction. When $n = 1$, it is trivial. We show that the conclusion holds when $n = 2$. Let $f(x) = \max\left\{\left|x - a_1\right| + b_1, \left|x - a_2\right| + b_2\right\}$, and let $m_1 = \max\{a_1 + b_1, a_2 + b_2\}$ and $m_2 = \max\{-a_1 + b_1, -a_2 + b_2\}$. We note that there are only four cases in which a combination of indices achieves the maximum values. If $m_1 = a_1 + b_1$ and $m_2 = -a_1 + b_1$, then $\left|x - a_1\right| + b_1$ is larger than $\left|x - a_2\right| + b_2$ for any x. Hence, $f(x) = \left|x - a_1\right| + b_1 = |x - a| + b$, where $a = (m_1 - m_2)/2$ and $b = (m_1 + m_2)/2$. Similarly, if $m_1 = a_2 + b_2$ and $m_2 = -a_2 + b_2$, then the result holds. We need to show that the result holds when $m_1 = a_1 + b_1$ and $m_2 = -a_2 + b_2$. (The other case in which $m_1 = a_2 + b_2$ and $m_2 = -a_1 + b_1$ is symmetrically identical.) In this case, we can easily verify that for $x \le a = (m_1 - m_2)/2$, $f(x) = -x + a + b$ and for $x > a$, $f(x) = x - a + b$, which implies $f(x) = |x - a| + b$. Therefore, the conclusion holds when $n = 2$. We assume that the conclusion holds when $n = m$. Now, we consider the case $n = m + 1$. By the assumption, we observe

$$f(x) = \max_{1 \le i \le m+1}\left\{\left|x - a_i\right| + b_i\right\} = \max\{|x - a| + b, |x - a_{m+1}| + b_{m+1}\}$$

where $m_1 = \max_{1 \le i \le m}(a_i + b_i)$, $m_2 = \max_{1 \le i \le m}(-a_i + b_i)$, $a = (m_1 - m_2)/2$, and $b = (m_1 + m_2)/2$. Using the conclusion for $n = 2$, we can see that the conclusion also holds when $n = m + 1$. Inductively, the conclusion holds for all integers n. □

By Lemma 2, we can write the objective function $S(\mu)$ as

$$S(\mu) = (\epsilon - \widehat{\gamma}_1)\frac{|\mu - A| + B}{\mu - \widehat{\gamma}_1} \tag{6}$$

where $A = (M_1 - M_2)/2$ and $B = (M_1 + M_2)/2$ with

$$M_1 = \max_j \left\{\widehat{\sigma}_{jj} + \sum_{i:i\neq j}\left|\widehat{\sigma}_{ij}\right|\right\} = \|\widehat{\mathbf{\Sigma}}\|_\infty,$$
$$M_2 = \max_j \left\{-\widehat{\sigma}_{jj} + \sum_{i:i\neq j}\left|\widehat{\sigma}_{ij}\right|\right\}.$$

Note that

$$\begin{aligned} A &= \frac{M_1 - M_2}{2} = \frac{\max_j\left\{\widehat{\sigma}_{jj} + \sum_{i:i\neq j}\left|\widehat{\sigma}_{ij}\right|\right\} - \max_j\left\{-\widehat{\sigma}_{jj} + \sum_{i:i\neq j}\left|\widehat{\sigma}_{ij}\right|\right\}}{2} \\ &\ge \min_j \widehat{\sigma}_{jj} = \min_j e_j^\top \widehat{\mathbf{\Sigma}} e_j \\ &\ge \min_{v:\|v\|_2=1} v^\top \widehat{\mathbf{\Sigma}} v = \widehat{\gamma}_1 \end{aligned}$$

where e_j is a p-dimensional vector with its jth entry 1 and the remaining entries of zero. If $\widehat{\gamma}_1 = A$, then $\widehat{\mathbf{\Sigma}}$ is positive definite, which is not the case we are interested in. Hence, we exclude this case. In other words, we assume that A is strictly larger than $\widehat{\gamma}_1$. The following lemma shows the solution to the minimization problem (4).

Lemma 3 *For A and B defined above, assume that $\widehat{\boldsymbol{\Sigma}}$ is given, and $\widehat{\gamma}_1 < \min\{\epsilon, A\}$. Then, for any $\mu \in (\epsilon, \infty)$*

$$\|\boldsymbol{\Phi}_{\mu,\alpha^*(\mu)}(\widehat{\boldsymbol{\Sigma}}) - \widehat{\boldsymbol{\Sigma}}\|_\infty \geq (\epsilon - \widehat{\gamma}_1) \min\left\{1, \frac{B}{A - \widehat{\gamma}_1}\right\}.$$

In detail, if $\widehat{\gamma}_1 + M_2 > 0$, then $\|\boldsymbol{\Phi}_{\mu,\alpha^*(\mu)}(\widehat{\boldsymbol{\Sigma}}) - \widehat{\boldsymbol{\Sigma}}\|_\infty$ is a decreasing function of μ and converges to $(\epsilon - \widehat{\gamma}_1)$ as $\mu \to \infty$. In particular, if $\mu \geq \frac{(1+\delta)\widehat{\gamma}_1 + M_2}{\delta}$ for some $\delta > 0$, then $\|\boldsymbol{\Phi}_{\mu,\alpha^*(\mu)}(\widehat{\boldsymbol{\Sigma}}) - \widehat{\boldsymbol{\Sigma}}\|_\infty \leq (1+\delta)(\epsilon - \widehat{\gamma}_1)$. If $\widehat{\gamma}_1 + M_2 = 0$, then $\|\boldsymbol{\Phi}_{\mu,\alpha^*(\mu)}(\widehat{\boldsymbol{\Sigma}}) - \widehat{\boldsymbol{\Sigma}}\|_\infty = \epsilon - \widehat{\gamma}_1$ for any $\mu \geq A$. If $\widehat{\gamma}_1 + M_2 < 0$, then $\|\boldsymbol{\Phi}_{\mu,\alpha^*(\mu)}(\widehat{\boldsymbol{\Sigma}}) - \widehat{\boldsymbol{\Sigma}}\|_\infty$ achieves its minimum at $\mu = A$.

Proof Below, we solve the minimization problem (4) for the three cases, (a) $\widehat{\gamma}_1 + M_2 > 0$, (b) $\widehat{\gamma}_1 + M_2 = 0$, and (c) $\widehat{\gamma}_1 + M_2 < 0$, respectively. The graph of $S(\mu)$ for each case is presented in Fig. 1. First, if (a) $\widehat{\gamma}_1 + M_2 > 0$, then $S(\mu) > \epsilon - \widehat{\gamma}_1$ for any $\mu \in (\epsilon, \infty)$. Since $S(\mu)$ decreases as μ increases, there is no solution to the minimization problem. However, instead of finding the best solution, we can make $S(\mu) = \|\boldsymbol{\Phi}_{\mu,\alpha^*}(\widehat{\boldsymbol{\Sigma}}) - \widehat{\boldsymbol{\Sigma}}\|_\infty$ close to $(\epsilon - \widehat{\gamma}_1)$ by setting μ to a large value. Specifically, we can easily determine that if $\mu \geq \frac{(1+\delta)\widehat{\gamma}_1 + M_2}{\delta}$ for some $\delta > 0$, then $\|\boldsymbol{\Phi}_{\mu,\alpha^*(\mu)}(\widehat{\boldsymbol{\Sigma}}) - \widehat{\boldsymbol{\Sigma}}\|_\infty \leq (1+\delta)(\epsilon - \widehat{\gamma}_1)$. Here, a large choice of μ makes α^* close to 1 and slightly attenuates $\widehat{\boldsymbol{\Sigma}}$ and its eigenvalues.

In addition, note that $\widehat{\gamma}_1 + M_2 > 0$ is equivalent to

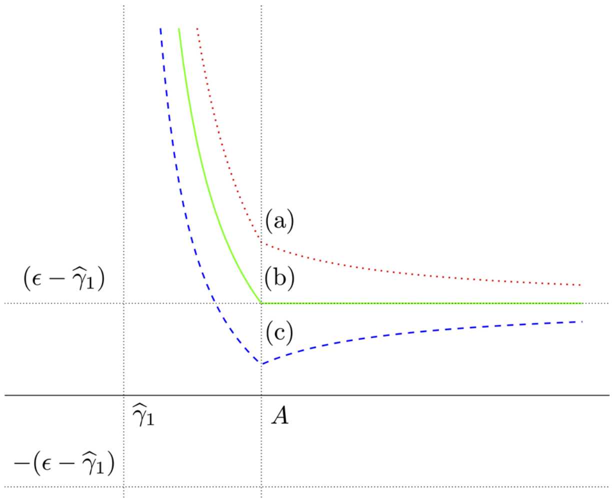


Fig. 1 Graphs of $S(\mu)$ for each case: **a** $\widehat{\gamma}_1 + M_2 > 0$ (red dotted line), **b** $\widehat{\gamma}_1 + M_2 = 0$ (green solid line), and **c** $\widehat{\gamma}_1 + M_2 < 0$ (blue dashed line) (color figure online)

$$\sum_{i:i\neq j} \left|\widehat{\sigma}_{ij}\right| > \widehat{\sigma}_{jj} - \widehat{\gamma}_1 \text{ for some } j. \tag{7}$$

Thus, the condition (7) holds easily when $\widehat{\gamma}_1$ is large and the initial estimate $\widehat{\boldsymbol{\Sigma}}$ is well posed. In this case, we do not have to shrink $\widehat{\boldsymbol{\Sigma}}$ much. Second, if (b) $\widehat{\gamma}_1 + M_2 = 0$, $S(\mu)$ has a minimum of $\epsilon - \widehat{\gamma}_1$ at any μ larger than A. Taking μ as any value larger than A gives us the desired result. Third, if (c) $\widehat{\gamma}_1 + M_2 < 0$, then $S(\mu)$ has a minimum of $\frac{B}{A-\widehat{\gamma}_1}$ at $\mu = A$. Again, note that case (c) is equivalent to

$$\sum_{i:i\neq j} \left|\widehat{\sigma}_{ij}\right| < \widehat{\sigma}_{jj} - \widehat{\gamma}_1 \text{ for all } j. \tag{8}$$

The condition (8) holds easily for a small value of $\widehat{\gamma}_1$. In other words, if the initial estimate $\widehat{\boldsymbol{\Sigma}}$ is non-PD or close to non-PD, there is a unique optimal solution μ, which makes the smallest eigenvalue of $\boldsymbol{\Phi}_{\mu,\alpha^*}(\widehat{\boldsymbol{\Sigma}})$ larger than ϵ, and $\boldsymbol{\Phi}_{\mu,\alpha^*}(\widehat{\boldsymbol{\Sigma}})$ is close to $\widehat{\boldsymbol{\Sigma}}$ with respect to $\|\cdot\|_\infty$. □

3 Statistical properties of the ℓ_∞–FSPD estimator

To establish the convergence rate of the ℓ_∞–FSPD estimator, we make the assumption, which Choi et al. (2019b) used for the spectral norm, and show that the convergence rate of the ℓ_∞–FSPD estimator is at least equivalent to that of the initial estimator in terms of the ℓ_∞ norm. The assumption we make is

(A1) ϵ is less than the smallest eigenvalue γ_1 of $\boldsymbol{\Sigma}$.

Theorem 1 *Let $\widehat{\boldsymbol{\Sigma}}$ be any estimator of $\boldsymbol{\Sigma}$. Suppose that ϵ satisfies the assumption* (**A1**). *Supposing that we take μ as in Lemma* 3, *then we have*

$$\|\boldsymbol{\Phi}_{\mu,\alpha^*}(\widehat{\boldsymbol{\Sigma}}) - \boldsymbol{\Sigma}\|_\infty \leq (1 + C)\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty,$$

for some constant $C > 0$.

Proof For ϵ satisfying (**A1**), we observe

$$(\epsilon - \widehat{\gamma}_1)_+ \leq (\gamma_1 - \widehat{\gamma}_1)_+ \leq \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_2 \leq \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty. \tag{9}$$

The second inequality follows from the Weyl's perturbation inequality, and the third inequality follows from Hölder's inequality, which states that for a symmetric matrix $\mathbf{A}$,

$$\|\mathbf{A}\|_2 \leq \sqrt{\|\mathbf{A}\|_1 \|\mathbf{A}\|_\infty} = \|\mathbf{A}\|_\infty.$$

Combining the results of Lemma 3 and the inequality (9) completes the proof. □

We next discuss some regularized covariance matrix estimators and their convergence rates in the ℓ_∞ norm. To our knowledge, the only estimator whose ℓ_∞ norm rate is predominantly studied is the universal thresholding estimator $\widehat{\boldsymbol{\Sigma}}^{\mathrm{BL}} = T_\lambda(\mathbf{S}) = \left(\widehat{\sigma}_{ij}\right)_{1\le i,j\le p}$ by Bickel and Levina (2008a), where $\lambda = c\sqrt{\frac{\log p}{n}}$ and

$$\widehat{\sigma}_{ij} = s_{ij} I\left(\left|s_{ij}\right| > \lambda\right) \text{ for all } i \neq j.$$

However, many other popular sparse covariance matrix estimators, which include the banding estimator by Bickel and Levina (2008b), the tapering estimator by Cai et al. (2010), and the adaptive thresholding estimator by Cai and Liu (2011), are also known to achieve the minimax optimal rate of the convergence in the matrix ℓ_∞ norm. In Bickel and Levina (2008b) and Cai and Liu (2011), the authors focus on the spectral norm error but compute the matrix ℓ_∞ norm error as its upper bound (see their proofs). Thus, their convergence rates in the matrix ℓ_∞ norm are the same as those in the spectral norm (reported in the papers) and are minimax optimal over different classes of covariance matrices.

The minimax rate of $\widehat{\boldsymbol{\Sigma}}^{\mathrm{BL}}$ in the ℓ_∞ norm is computed by Cai and Zhou (2012a) and Cai and Zhou (2012b). Two papers derive the minimax rate of convergence for estimating a class of large covariance matrices under the matrix operator norm. Cai and Zhou (2012b) consider the following class of sparse matrices:

$$\mathcal{G}_q(c_{n,p}) = \left\{\boldsymbol{\Sigma} = (\sigma_{ij})_{1\le i,j\le p} : \sigma_{-j,j} \in B_q^{p-1}(c_{n,p}) \text{ for } 1 \le j \le p\right\} \tag{10}$$

for $0 \le q < 1$, where $\sigma_{-j,j}$ denotes the jth column of $\boldsymbol{\Sigma}$ with σ_{jj} removed,

$$B_q^{p-1}(c) = \left\{\xi \in \mathbb{R}^{p-1} : |\xi|_{(k)}^q \le \frac{c}{k} \text{ for all } k = 1, 2, \dots, p-1\right\}, \tag{11}$$

and $|\xi|_{(k)}$ denotes the kth largest element in magnitude of the vector ξ. They show that:

Theorem A (Cai and Zhou 2012b, Theorem 5) *Assume that* $c_{n,p} \le M n^{\frac{1-q}{2}} (\log p)^{-\frac{3-q}{2}}$ *for* $0 \le q < 1$ *and* $1 < n^\beta \le p$ *for some constants* $\beta > 1$. *Then, the minimax risk of estimating the covariance matrix* $\boldsymbol{\Sigma}$ *under the matrix* ℓ_w*-norm for* $1 \le w \le \infty$ *over the class* $\mathcal{P}_q(\tau, c_{n,p})$ *satisfies*

$$\inf_{\widehat{\boldsymbol{\Sigma}}} \sup_{\mathcal{P}_q(\tau, c_{n,p})} \mathbb{E}\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_w^2 \asymp c_{n,p}^2 \left(\frac{\log p}{n}\right)^{1-q} + \frac{\log p}{n} \tag{12}$$

where $a_n \asymp b_n$ *implies that there exist positive constants* c *and* C *such that* $c \le a_n/b_n \le C$, $\mathcal{P}_q(\tau, c_{n,p})$ *denotes the class of distributions of* p*-dimensional random vector* X *satisfying* $\mathrm{Cov}(X) = \boldsymbol{\Sigma} \in \mathcal{G}_q(c_{n,p})$ *and*

$$\Pr(|v^{\mathsf{T}}(X - \mathbb{E}X)| > t) \le e^{-\frac{t^2}{2\tau}} \quad \text{for all } t > 0 \text{ and } \|v\|_2 = 1. \tag{13}$$

Moreover, the universal thresholding estimator $\widehat{\boldsymbol{\Sigma}}^{\mathrm{BL}}$ *by Bickel and Levina (2008a) is rate-optimal.*

4 Application to MVP

Let $\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Phi}_{\mu^*,\alpha^*}(\widehat{\boldsymbol{\Sigma}})$ be the ℓ_∞-FSPD modification of an estimator $\widehat{\boldsymbol{\Sigma}}$ and assume that

$$\Pr\left(\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty \leq \alpha_n \right) \to 1$$

for some sequence $\alpha_n \to 0$. Clearly, we also have $\Pr\left(\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_2 \leq \alpha_n \right) \to 1$. From Theorem 1 in Sect. 3 and Theorem 2 of Choi et al. (2019b), the FSPD modification $\widetilde{\boldsymbol{\Sigma}}$ of $\widehat{\boldsymbol{\Sigma}}$ preserves the convergence rate in ℓ_∞ of the initial estimator $\widehat{\boldsymbol{\Sigma}}$. Suppose that their rates in the matrix ℓ_∞ norm are the same as α_n, i.e. as $n \to \infty$, we have

$$\Pr\left(\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty \leq \alpha_n \right) \to 1 \quad \text{and} \quad \Pr\left(\|\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty \leq 2\alpha_n \right) \to 1.$$

For the MVP application, we make an additional assumption regarding the initial estimator and the precision matrix $\boldsymbol{\Sigma}^{-1}$, which is

(A2) $\gamma_p \alpha_n \|\boldsymbol{\Sigma}^{-1}\|_\infty^2 \to 0, \quad$ as $n \to \infty$,

where γ_p denotes the maximum eigenvalue of $\boldsymbol{\Sigma}$.

4.1 Universal thresholding estimator as an initial estimator

The proposition below shows that the universal thresholding estimator $\widehat{\boldsymbol{\Sigma}}^{\mathrm{BL}}$ by Bickel and Levina (2008a) satisfies the assumption **(A2)** under the weak assumption on the true covariance matrix.

Proposition 1 *Suppose that* $\boldsymbol{\Sigma}$ *satisfies* $\boldsymbol{\Sigma} = (\sigma_{ij}) \in \mathcal{G}_q(c_{1,n,p})$, $\boldsymbol{\Sigma}^{-1} = (\omega_{ij}) \in \mathcal{G}_q(c_{2,n,p})$, $\sigma^2_{\max}(\equiv \max_i \sigma_{ii}) = O(1)$ *and* $\omega^2_{\max}(\equiv \max_i \omega_{ii}) = O(1)$. *Then, for* $\widehat{\boldsymbol{\Sigma}}^{\mathrm{BL}}$, *the assumption* **(A2)** *holds if*

$$\alpha_n = \max\left(c_{1,n,p}\left(\frac{\log p}{n}\right)^{\frac{1-q}{2}}, \left(\frac{\log p}{n}\right)^{1/2} \right) \to 0. \tag{14}$$

Proof First, we observe that the matrix ℓ_∞ norm of precision matrix $\boldsymbol{\Sigma}^{-1} = (\omega_{ij})_{1\leq i,j\leq p}$ is bounded as

$$\begin{aligned}
\|\boldsymbol{\Sigma}^{-1}\|_\infty &= \max_i \sum_j |\omega_{ij}| \\
&\leq \max_i \left\{ \omega_{ii} + \sum_{j=1}^{p-1} \left(\frac{c_{2,n,p}}{j}\right)^{1/q} \right\} = O(1).
\end{aligned}$$

Similarly, we can also obtain $\|\boldsymbol{\Sigma}\|_\infty = O(1)$. This implies that $\gamma_p = \|\boldsymbol{\Sigma}\|_2 = O(\sqrt{1})$. According to the result of Cai and Zhou (2012b), the rate of convergence of the universal thresholding estimator $\widehat{\boldsymbol{\Sigma}}^{\mathrm{BL}}$ is $\alpha_n = c_{1,n,p}\left(\frac{\log p}{n}\right)^{\frac{1-q}{2}} + \left(\frac{\log p}{n}\right)^{\frac{1}{2}}$. Hence, we obtain the following convergence of the rate:

$$\gamma_p \alpha_n \|\boldsymbol{\Sigma}^{-1}\|_\infty^2 = O\big(\alpha_n\big) = o(1),$$

which completes the proof. □

Now, let us discuss the class of true covariance matrix $\boldsymbol{\Sigma}$ in which $\boldsymbol{\Sigma} \in \mathcal{G}_q(c_{1,n,p})$ and $\boldsymbol{\Sigma}^{-1} \in \mathcal{G}_q(c_{2,n,p})$. We claim that the class is not small, with the following example. Suppose that we consider the following model for $X = (X_1, \cdots, X_p)^\top$:

$$\begin{aligned}
X_1 &= \epsilon_1 \\
X_2 &= a_{21}X_2 + \epsilon_2 \\
X_3 &= a_{31}X_1 + a_{32}X_2 + \epsilon_3 \\
&\vdots \\
X_p &= a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{p(p-1)}X_{p-1} + \epsilon_p
\end{aligned}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_p)^\top$ has variance $\mathrm{Var}(\epsilon) = \mathrm{diag}(d_i) = D$. Then, we have $\boldsymbol{\Sigma} = \mathrm{Var}(X) = (I - A)^{-1}D(I - A^\top)^{-1}$, where $A = (a_{ij})$ with $a_{ij} = 0$ for $i \le j$. Now, we assume $|a_{ij}| \le c_1\rho_1^{i-j}$ for some $c_1 > 0$ and $0 < \rho_1 < 1$ with $(1 + c_1)\rho_1 < 1$. Since A is a lower triangular matrix with zero diagonal entries, the power matrix A^r vanishes as r increases ($A^r = O$ for $r \ge p$). Hence, $(I - A)^{-1} = I + A + R$ where $R = A^2 + \cdots + A^{p-1}$, and

$$\boldsymbol{\Sigma} = (I + A)D(I + A)^\top + \mathrm{RD} + \mathrm{DR}^\top + \mathrm{RDA}^\top + \mathrm{ADR}^\top + \mathrm{RDR}^\top.$$

We observe that for i, j with $i - j > 1$

$$\begin{aligned}
|R_{ij}| &= \left| \sum_{d=1}^{i-j-1} \sum_{j<k_d<\cdots<k_1<i} a_{ik_1}a_{k_1k_2}\cdots a_{k_dj} \right| \\
&\le \sum_{d=1}^{i-j-1} \sum_{j<k_d<\cdots<k_1<i} c_1^{d+1}\rho_1^{i-j} = \sum_{d=1}^{i-j-1} \binom{i-j-1}{d} c_1^{d+1}\rho_1^{i-j} \\
&\le \frac{c_1}{1+c_1}\big\{(1+c_1)\rho_1\big\}^{i-j} = c_2\rho_2^{i-j}.
\end{aligned}$$

Hence, for $i > j$, $|(\mathrm{RD})_{ij}| \le c_2 d_{\max}\rho_2^{i-j}$. For $i \ge j - 1$,

$$
\begin{aligned}
|(\mathrm{ADR}^\top)_{ij}| = &\left|\sum_{k=1}^{p} a_{ik} d_k R_{jk}\right| \\
&\leq c_1 c_2 d_{\max} \sum_{k=1}^{j-2} \rho_1^{i-k} \rho_2^{j-k} = c_1 c_2 d_{\max} \rho_1^{i-j} \sum_{k=2}^{j-1} \rho_1^k \rho_2^k \\
&\leq \frac{c_1 c_2 d_{\max} \rho_1 \rho_2^2}{1-\rho_1\rho_2} \rho_1^{i-(j-1)},
\end{aligned}
$$

and for $i < j-1$,

$$
\begin{aligned}
|(\mathrm{ADR}^\top)_{ij}| &\leq c_1 c_2 d_{\max} \sum_{k=1}^{i-1} \rho_1^{i-k} \rho_2^{j-k} = c_1 c_2 d_{\max} \rho_2^{j-i} \sum_{k=1}^{i-1} \rho_1^k \rho_2^k \\
&\leq \frac{c_1 c_2 d_{\max} \rho_1 \rho_2^2}{1-\rho_1\rho_2} \rho_2^{(j-1)-i}.
\end{aligned}
$$

Similarly, we can show that $|(\mathrm{RDR}^\top)_{ij}| \leq \frac{c_2 d_{\max} \rho_2^4}{1-\rho_2^2} \rho_2^{|i-j|}$. Additionally, we observe for $i > j$

$$
\begin{aligned}
\left|\left((I+A)D(I+A)^\top\right)_{ij}\right| = \left|\sum_{k=1}^{p} d_k(\delta_{ik}+a_{ik})(\delta_{jk}+a_{jk})\right| &= \left|\sum_{k=1}^{j-1} d_k a_{ik} a_{jk} + d_j a_{ij}\right| \\
&\leq c_1^2 d_{\max} \sum_{k=1}^{j-1} \rho_1^{i+j-2k} + c_1 d_{\max} \rho_1^{i-j} \\
&\leq \left(\frac{c_1^2 d_{\max} \rho_1^2}{1-\rho_1^2} + c_1 d_{\max}\right) \rho_1^{i-j}.
\end{aligned}
$$

To summarize, the entries of $\boldsymbol{\Sigma}$ vanish exponentially farther away from the diagonal, which implies that $\boldsymbol{\Sigma} \in \mathcal{G}_q(c_{1,n,p})$ for some $c_{1,n,p} > 0$. In the same way, we can show that $\Omega = \boldsymbol{\Sigma}^{-1} = (I-A)^\top D^{-1}(I-A) \in \mathcal{G}_q(c_{2,n,p})$ for some $c_{2,n,p} > 0$.

4.2 MVP allocation

In finance, a portfolio refers to a family of (risky) assets held by an institution or private individual. If there are multiple assets to invest in, a combination of assets is considered and it becomes an important issue to select an optimal portfolio allocation. The MVP allocation is one of the most well-established strategies for portfolio allocation. Chan et al. (2015) propose to choose a portfolio that minimizes risk, or standard deviation of return. To be specific, let $\mathbf{r} = (r_1, \ldots, r_p)^\top$ be a p-variate random vector in which, for each $j \in \{1, \ldots, p\}$, r_j represents the return of the jth asset constituting the portfolio. Denote by $\boldsymbol{\Sigma} = \mathrm{Var}(\mathbf{r})$ the unknown covariance matrix of assets. A $p \times 1$ vector $\mathbf{w}$ represents an allocation of the investor's wealth in such a way that each w_j indicates the weight of the jth asset and $w_1 + \cdots + w_p = 1$. Then, the MVP optimization by Chan et al. (2015) is formulated as

$$\underset{\mathbf{w}}{\text{minimize}}\ \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^\top \mathbf{1} = 1.$$

In applying MVP to the practice, the covariance matrix $\boldsymbol{\Sigma}$ is unknown and should be replaced by a suitable PD estimator. A sample covariance matrix is typically used, but it often fails at various points when the number of assets in the universe increases (p is large). To resolve this difficulty, researchers propose to use regularized estimators of high-dimensional covariance in order to make some additional constraints for the optimization problem (Ledoit and Wolf 2003, 2004; DeMiguel et al. 2009; Fan et al. 2013; Glasserman and Kang 2014; Ledoit and Wolf 2017a, b; Dai and Wen 2018; Dai et al. 2020). In particular, recently, many regularized high-dimensional covariance estimators have been proposed by many authors, and their theoretical properties are understood well. However, as pointed out by Choi et al. (2019b), they frequently become non-PD. Here, we propose to modify the initial regularized estimator, say $\widehat{\boldsymbol{\Sigma}}$, with ℓ_∞-FSPD. Let the resulting estimator be $\widetilde{\boldsymbol{\Sigma}}$. Thus, we solve

$$\underset{\mathbf{w}}{\text{minimize}}\ \mathbf{w}^\top \widetilde{\boldsymbol{\Sigma}} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^\top \mathbf{1} = 1. \tag{15}$$

and let the solution be $\widetilde{\mathbf{w}}$.

We claim that the ℓ_∞ error of $\widetilde{\mathbf{w}}$ can be bounded through $\|\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty$ in Theorem 2.

Theorem 2 *For* $\widetilde{\mathbf{w}}$, *the solution to* (15), *under the assumption* **(A2)**,

$$\|\widetilde{\mathbf{w}} - \mathbf{w}\|_\infty \le \frac{C\gamma_p \alpha_n \left\|\boldsymbol{\Sigma}^{-1}\right\|_\infty^2 \left(1 + 2\gamma_p \left\|\boldsymbol{\Sigma}^{-1}\right\|_\infty\right)}{p}, \tag{16}$$

which converges to 0 *with probability approaching* 1.

Proof Note that $\widetilde{\mathbf{w}} = \widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}/\mathbf{1}^\top\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$ and $\mathbf{w} = \boldsymbol{\Sigma}^{-1}\mathbf{1}/\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}$. We have

$$\begin{aligned}
\|\widetilde{\mathbf{w}} - \mathbf{w}\|_\infty &= \left\| \frac{\left(\widetilde{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\right)\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}} + \frac{\mathbf{1}^\top\left(\widetilde{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\right)\mathbf{1}}{\left(\mathbf{1}^\top\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}\right)\left(\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}\right)}\boldsymbol{\Sigma}^{-1}\mathbf{1} \right\|_\infty \\
&\le \frac{\left\|\widetilde{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\right\|_\infty}{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}} + \frac{\left|\mathbf{1}^\top\left(\widetilde{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\right)\mathbf{1}\right|}{\left(\mathbf{1}^\top\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}\right)\left(\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}\right)}\left\|\boldsymbol{\Sigma}^{-1}\right\|_\infty.
\end{aligned} \tag{17}$$

To evaluate the first term in (17), say E, note that

$$
\begin{aligned}
\left\|\widetilde{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right\|_{\infty} &\leq \|\boldsymbol{\Sigma}^{-1}\|_{\infty}\left\|\widetilde{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\right\|_{\infty}\left\|\widetilde{\boldsymbol{\Sigma}}^{-1}\right\|_{\infty} \\
&\leq 2\alpha_n\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty}\left(\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty}+\left\|\widetilde{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right\|_{\infty}\right),
\end{aligned}
$$

which is equivalent to

$$
\left\|\widetilde{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right\|_{\infty} \leq \frac{2\alpha_n\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty}^{2}}{1-2\alpha_n\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty}} \leq C\alpha_n\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty}^{2},
$$

for $C>2$, where the assumption (A2) is used for the second inequality. The assumption (A2) implies that $\alpha_n\|\boldsymbol{\Sigma}^{-1}\|_{\infty}\to 0$ as $n\to\infty$. Thus,

$$
\mathrm{E} \leq \frac{1}{p}C\gamma_p\alpha_n\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty}^{2}
$$

with probability approaching 1.

Next, we evaluate the second term in (17), say F. Since

$$
\begin{aligned}
\mathbf{1}^{\top}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1} \geq & \mathbf{1}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}-\mathbf{1}^{\top}\left(\boldsymbol{\Sigma}^{-1}-\widetilde{\boldsymbol{\Sigma}}^{-1}\right)\mathbf{1} \geq p\gamma_p^{-1}-p\left\|\boldsymbol{\Sigma}^{-1}-\widetilde{\boldsymbol{\Sigma}}^{-1}\right\|_{\infty}, \\
\mathrm{F} = & \frac{\left|\mathbf{1}^{\top}\left(\widetilde{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right)\mathbf{1}\right|}{\left(\mathbf{1}^{\top}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}\right)\left(\mathbf{1}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}\right)}\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty} \\
\leq & \frac{\left|\mathbf{1}^{\top}\left(\widetilde{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right)\mathbf{1}\right|\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty}}{p^2\gamma_p^{-2}\left(1-\gamma_p\left\|\widetilde{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right\|_{\infty}\right)} \\
\leq & \frac{\left\|\widetilde{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right\|_{\infty}\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty}}{p\gamma_p^{-2}\left(1-\gamma_p\left\|\widetilde{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right\|_{\infty}\right)}
\end{aligned}
$$

Using assumption (A2), it follows that

$$
\mathrm{F} \leq \frac{2C\gamma_p^2\alpha_n\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty}^{3}}{p}
$$

with probability approaching 1. Finally, we obtain the following inequality.

$$
\|\widetilde{\mathbf{w}}-\mathbf{w}\|_{\infty} \leq \frac{C\gamma_p\alpha_n\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty}^{2}\left(1+2\gamma_p\left\|\boldsymbol{\Sigma}^{-1}\right\|_{\infty}\right)}{p}
$$

with probability approaching 1. □

5 Numerical study

We numerically compare the performance of the ℓ_∞-FSPD estimator to other FSPD estimators in regard to the estimation error of the covariance matrix and the risk (variance) of the optimal portfolio. We denote our method by 'FSPD (μ_∞)', and other FSPD estimators by 'FSPD (μ_{S})', 'FSPD (μ_{F})', and 'FSPD (∞)' as in Choi et al. (2019b). We recall that 'FSPD (μ_{S})' and 'FSPD (μ_{F})' minimize the spectral norm distance and Frobenius norm distance between the initial estimator and the class of linear shrinkage estimators, respectively, and 'FSPD (∞)' sets $\mu = \infty$. In the comparison, we use the adaptive soft thresholding estimator created by Cai and Liu (2011) as the initial covariance matrix estimator.

For the study, we follow the simulation settings of Choi et al. (2019b). We generate p-dimensional random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. We consider two types of covariance matrix as true $\boldsymbol{\Sigma}$:

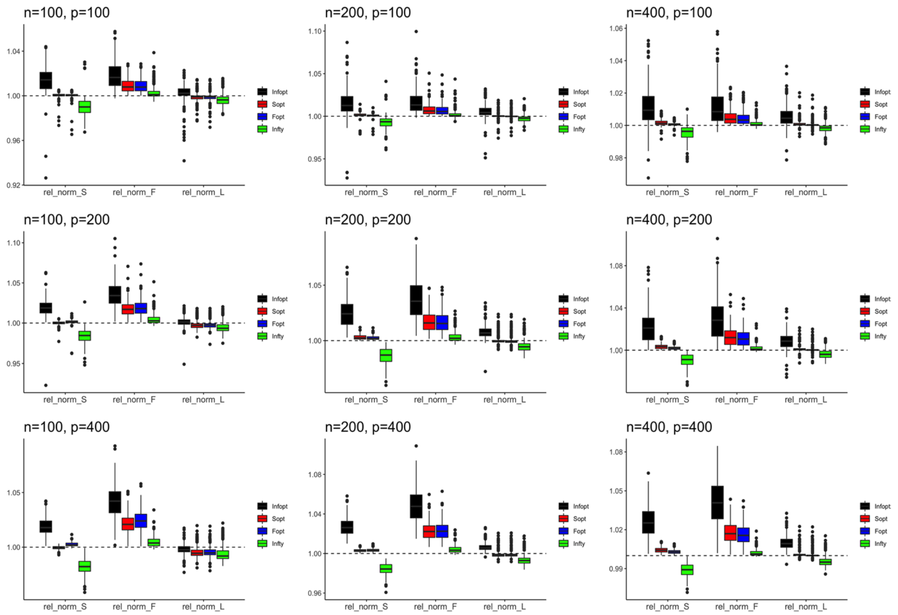
1. "Linearly tapered Toeplitz" matrix $\mathbf{M}_1$ with $[\mathbf{M}_1]_{ij} = \left(1 - \frac{|i-j|}{L}\right)_+$, where $L \in \{10, 20\}$;
2. "Overlapped block-diagonal" matrix $\mathbf{M}_2$ with $[\mathbf{M}_2]_{ij} = I(i=j) + \rho I\big((i,j) \in (I_k \cup \{i_k+1\}) \times (I_k \cup \{i_k+1\}) \text{ for some } k\big)$, where $I_k = \{20(k-1)+1, \ldots, 20k\}$, $i_k = 20k$, and $\rho \in \{0.6, 0.9\}$.

We set $n \in \{100, 200, 400\}$, $p \in \{100, 200, 400\}$ and $\epsilon = 10^{-4}$. We generate 200 datasets for each of the 36 possible combinations of parameters. We note that for some combinations of parameters, the minimum eigenvalue of the initial covariance matrix estimate is larger than the prespecified level ϵ, which makes FSPD modification unnecessary. Hence, for each repetition, we generate a dataset until the minimum eigenvalue of its initial covariance matrix estimate becomes non-PD. For every simulated dataset, we compute an adaptive soft thresholding estimate, where we use 5-fold cross-validation to select the regularization parameter of the estimator. We then apply four FSPD methods to the initial estimate.

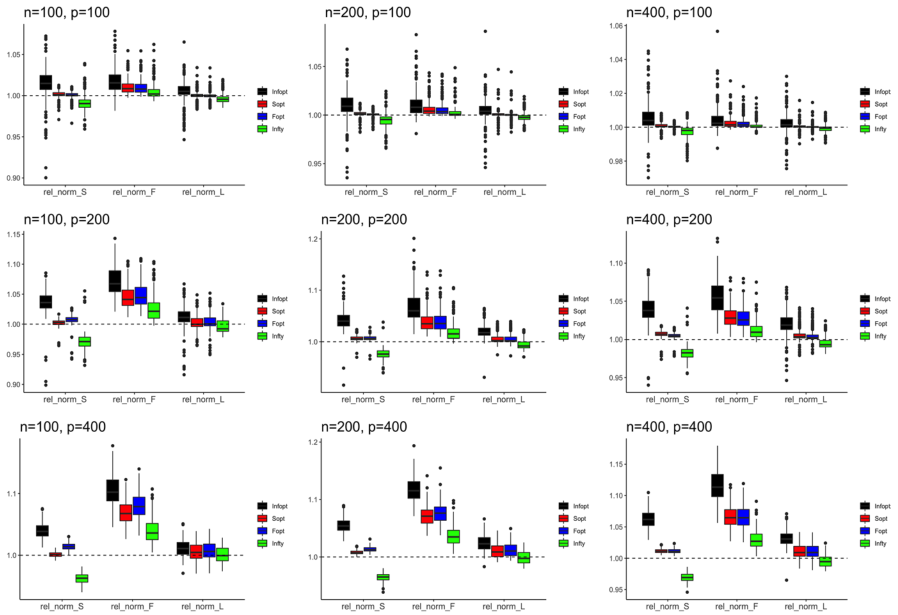
We first compare the relative error of the FSPD estimator with the initial estimator, which is defined as

$$\text{relative error} = \frac{\|\Phi_{\mu,\alpha^*(\mu)}(\widehat{\boldsymbol{\Sigma}}) - \boldsymbol{\Sigma}\|}{\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|}.$$

We consider three matrix norms to calculate relative errors: the spectral norm, the Frobenius norm, and the matrix ℓ_∞ norm. The box plots of relative errors are shown in Figs. 2 and 3. The figures show that the relative error of FSPD (μ_∞) is larger than others in most cases, except the cases with the matrix ℓ_∞ error. Additionally,

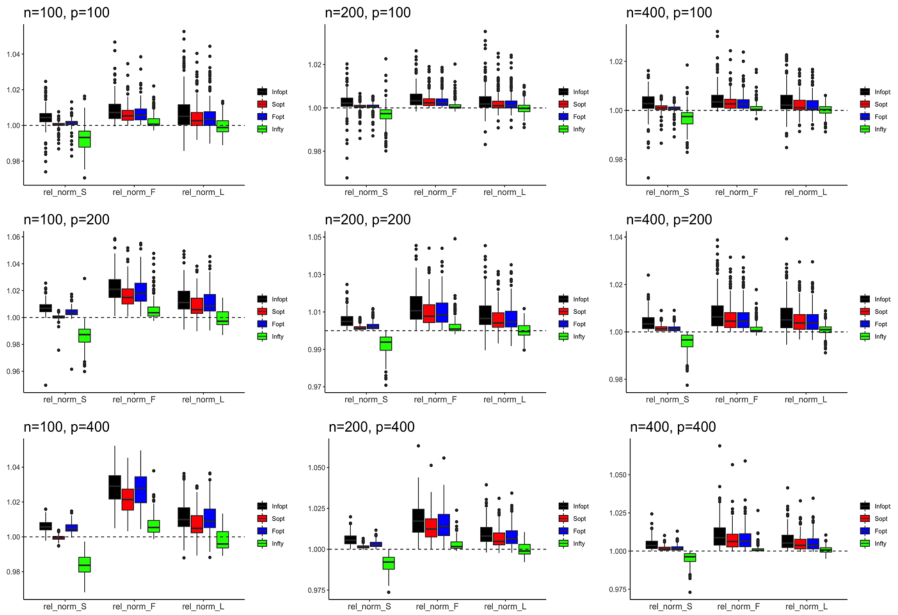


(a) $L = 10$

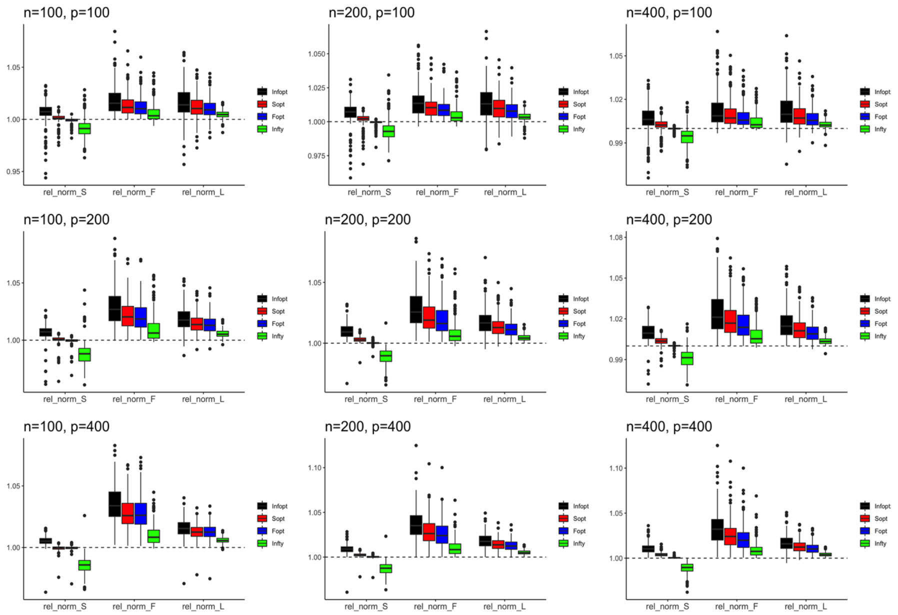


(b) $L = 20$

Fig. 2 Box plots of relative errors of modified estimates against initial estimates when $\mathbf{\Sigma} = \mathbf{M}_1$. 'FSPD ($\mu_{\mathrm{S}}$)', 'FSPD ($\mu_{\mathrm{F}}$)', and 'FSPD ($\mu_\infty$)' choose optimal μ, which minimizes spectral norm error, Frobenius norm error, and ℓ_∞ norm error, respectively, and 'FSPD (∞)' sets $\mu = \infty$. Corresponding n and p values are represented in the upper left corner of each figure



(a) $L = 10$



(b) $L = 20$

Fig. 3 Box plots of relative errors of modified estimates against initial estimates when $\boldsymbol{\Sigma} = \mathbf{M}_2$. 'FSPD ($\mu_{\mathrm{S}}$)', 'FSPD ($\mu_{\mathrm{F}}$)', and 'FSPD ($\mu_\infty$)' choose optimal μ, which minimizes spectral norm error, Frobenius norm error, and ℓ_∞ norm error, respectively, and 'FSPD (∞)' sets $\mu = \infty$. Corresponding n and p values are represented in the upper left corner of each figure interestingly, we find that FSPD (∞) performs best in all cases. The reason is that fixing $\mu = \infty$ reduces the variability of the FSPD estimates, while the choice of μ in other FSPD estimates depends on the samples.

We next compare the performance of the FSPD methods in terms of the MVP allocation. The risk of the optimal portfolio with a covariance matrix estimate $\widehat{\boldsymbol{\Sigma}}$ is defined as the minimum variance $\widehat{\mathbf{w}}\boldsymbol{\Sigma}\widehat{\mathbf{w}}$, where $\boldsymbol{\Sigma}$ is true covariance matrix and $\widehat{\mathbf{w}} = \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}/\mathbf{1}^{\top}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$ is the optimal allocation vector. Here, the initial adaptive thresholding covariance matrix is frequently not PD, and its risk is not available. For this reason, we compute the relative risks of FSPD (μ_{S}), FSPD (μ_{F}), and FSPD (∞) versus the FSPD (μ_{∞}) method. The box plots of relative risks are shown in Figs. 4 and 5. The figures show that the optimal portfolio with FSPD (μ_{∞}) has the smallest variance compared with the other three FSPD estimates in all cases. We find the same phenomenon from the real data example followed in the next section.

6 Data example: S&P 500 data

6.1 Data and rebalancing strategy

We use daily returns of the constituents of the S&P 500 stock market index, which are used in Choi et al. (2019a). The dataset contains the daily returns from 3 January 1978 to 31 December 2014. There are 9334 trading days during this period, and we only consider $p = 91$ stocks that are among the constituents of the index during the entire period.

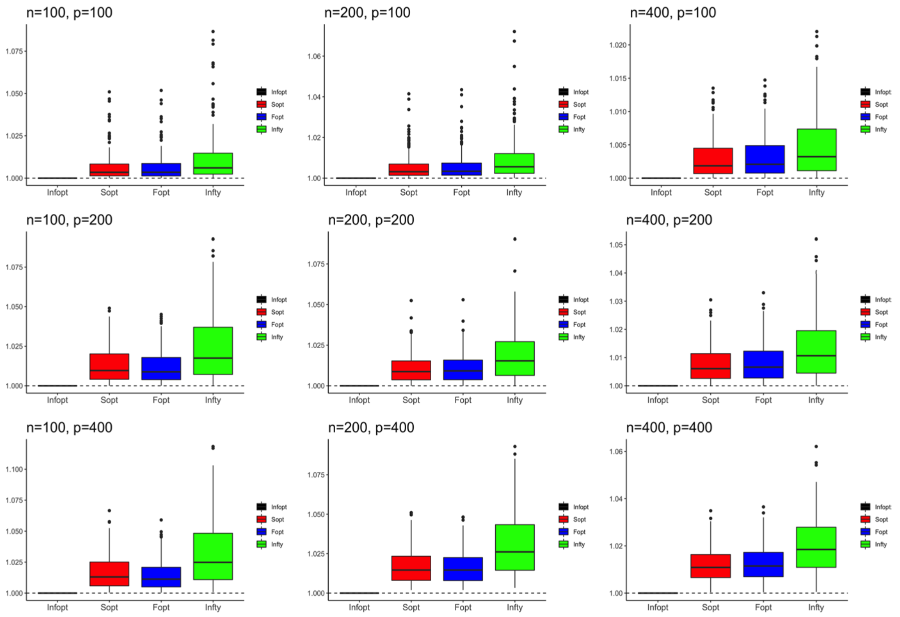
Since stock return is not stationary over a long-term period, we use the rebalancing strategy to construct a portfolio. This strategy renews the portfolio periodically and holds it until the next update. Specifically, let $\mathbf{r}_t = (r_{t1}, \dots, r_{tp})^{\top}$ be a vector of returns on the tth date. We estimate the covariance matrix $\boldsymbol{\Sigma} = \mathrm{Var}(\mathbf{r}_t)$ based on the stock returns of the past M days. Let $\widehat{\boldsymbol{\Sigma}}(M, t)$ be an estimator of $\boldsymbol{\Sigma}$ based on the data $\{\mathbf{r}_s : t - M \le s \le t - 1\}$. We renew the estimate of $\boldsymbol{\Sigma}$ periodically at every L dates. Denote the rebalancing dates by $t_k = L(k-1) + 1$ for $k = 1, 2, \dots, K$. Using the covariance matrix estimate $\widehat{\boldsymbol{\Sigma}}(M, t_k)$, we obtain the optimal weight for the kth period (more precisely, for time t_k) by

$$\mathbf{w}^{(k)} = \frac{\widehat{\boldsymbol{\Sigma}}(M, t_k)^{-1}\mathbf{1}}{\mathbf{1}^{\top}\widehat{\boldsymbol{\Sigma}}(M, t_k)^{-1}\mathbf{1}}.$$

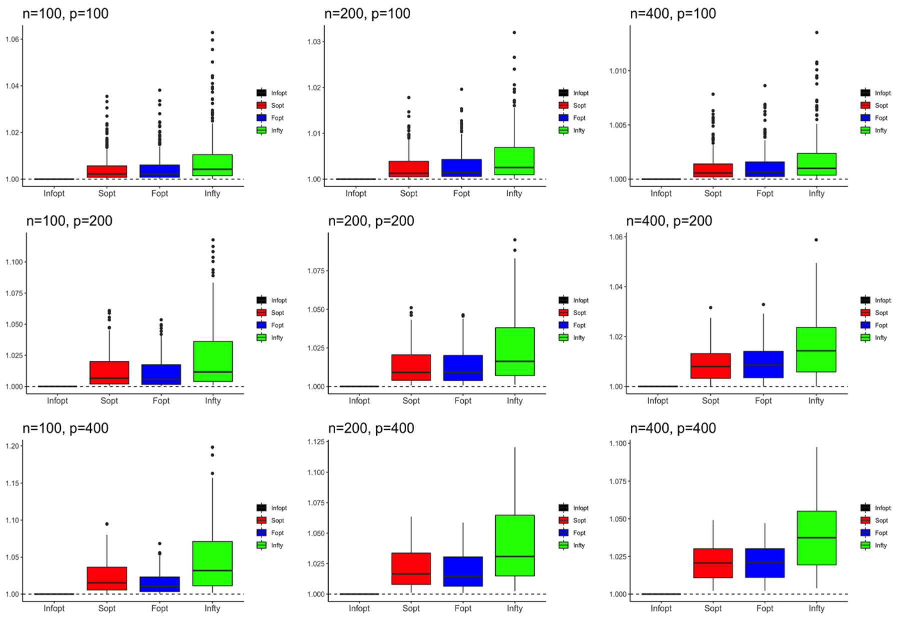
The portfolio weights for other time points in the kth period, $\mathbf{w}_t$ ($t = t_k + 1, t_k + 2, \dots, t_{k+1} - 1$), depend on the asset values at the time, and simple algebra shows that, for $t = t_k, t_k + 1, \dots, t_{k+1} - 1$, $\mathbf{w}_t = (w_{t1}, w_{t2}, \dots, w_{tp})^{\top}$ with

$$w_{tj} = \frac{(1 + r_{(t-1)j})w_{(t-1)j}}{\sum_{i=1}^{p}(1 + r_{(t-1)i})w_{(t-1)i}},$$

where $\mathbf{w}_{t_k} = \mathbf{w}^{(k)}$. We construct the portfolios for the period from 2 January 1980 to 31 December 2014 for $T = 8829$ trading days.

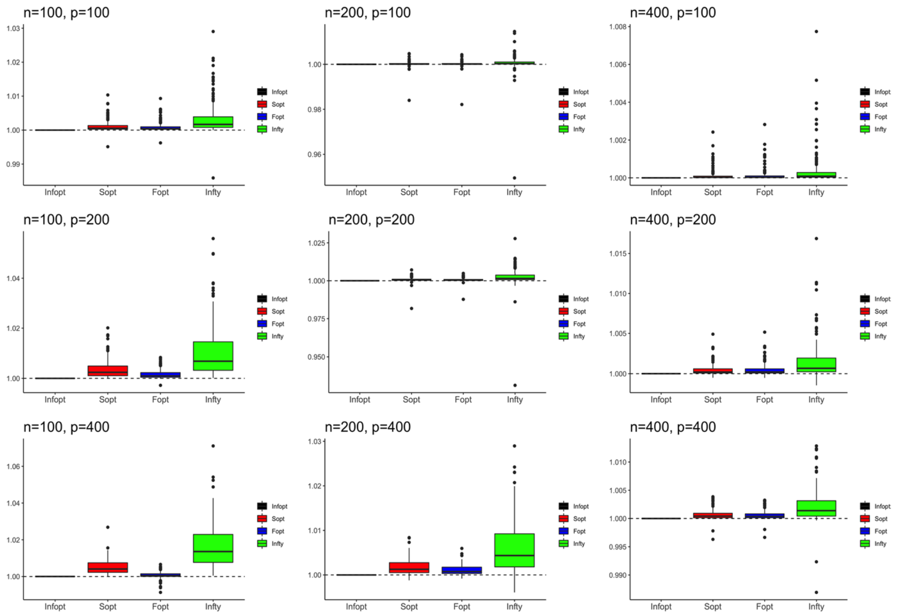


(a) $L = 10$

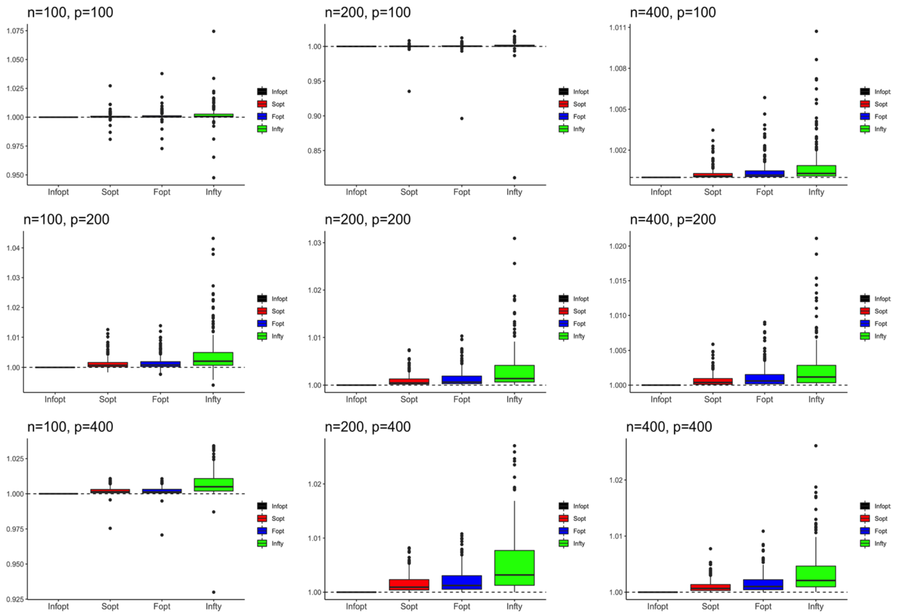


(b) $L = 20$

Fig. 4 Box plots of relative risk of 'FSPD (μ_S)', 'FSPD (μ_F)', and 'FSPD (∞)' estimates against 'FSPD (μ_∞)' estimate when $\mathbf{\Sigma} = \mathbf{M}_1$. 'FSPD ($\mu_S$)', 'FSPD ($\mu_F$)', and 'FSPD ($\mu_\infty$)' choose optimal μ, which minimizes spectral norm error, Frobenius norm error, and ℓ_∞ norm error, respectively, and 'FSPD (∞)' sets $\mu = \infty$. Corresponding n and p values are represented in the upper left corner of each figure



(a) $L = 10$



(b) $L = 20$

Fig. 5 Box plots of relative risk of 'FSPD (μ_S)', 'FSPD (μ_F)', and 'FSPD (∞)' estimates against 'FSPD (μ_∞)' estimate when $\mathbf{\Sigma} = \mathbf{M}_2$. 'FSPD ($\mu_S$)', 'FSPD ($\mu_F$)', and 'FSPD ($\mu_\infty$)' choose optimal μ, which minimizes spectral norm error, Frobenius norm error, and ℓ_∞ norm error, respectively, and 'FSPD (∞)' sets $\mu = \infty$. Corresponding n and p values are represented in the upper left corner of each figure

We set the formation period, or the length of the estimation window, as $M = 120$ and the holding period as $L = 120$. From $t_1 =$ (2 January 1980), we have $K = 73$ periods with L trading days. On the first day of each period, we rebalance the portfolio using the return data of the past M days. We consider two initial estimators, the universal thresholding estimator ('Univ') by Bickel and Levina (2008a) and the adaptive soft thresholding estimator ('Adap') by Cai and Liu (2011). As discussed in Sect. 3, both the universal and adaptive thresholding estimators are minimax optimal in the ℓ_∞ norm. Along with two initial estimators, we consider eight estimators: (1) 'Univ'; (2) 'Univ + FSPD (μ_{SF})'; (3) 'Univ + FSPD (∞)'; (4) 'Univ + FSPD (μ_∞)'; (5) 'Adap'; (6) 'Adap + FSPD (μ_{SF})'; (7) 'Adap + FSPD (∞)'; and (8) 'Adap + FSPD (μ_∞)'. When constructing initial thresholding estimates, we use 5-fold cross-validation to select the regularization parameters. 'FSPD (μ_{SF})' and 'FSPD (∞)' are the FSPD procedures by Choi et al. (2019b). 'FSPD (μ_∞)' is the ℓ_∞-FSPD modification.

6.2 Performance measures

To assess the performance of a portfolio, we use mean return, risk (standard deviation), Sharpe ratio, and (normalized) wealth growth. When calculating wealth growth, we take the transaction costs incurred in practical trading into consideration.

The return of the portfolio at the tth date with weight $\mathbf{w}_t = (w_{t1}, \dots, w_{tp})^\intercal$, R_t, is defined by the weighted average of the returns $\mathbf{r}_t = (r_{t1}, \dots, r_{tp})^\intercal$ on invested assets:

$$\text{RET}_t = R_t := \sum_{j=1}^{p} w_{tj} r_{tj}.$$

The annualized mean return M_RET, the annualized standard deviation SD_RET of returns, and the Sharpe ratio SR are defined as follows:

$$\begin{aligned}
\text{M_RET} &:= \left(\prod_{t=1}^{T} (1 + R_t) \right)^{252/T} - 1 \\
\text{SD_RET} &:= \sqrt{252} \cdot sd(R_t) \\
\text{SR} &:= \frac{\text{M_RET} - \text{RF}}{\text{SD_RET}}
\end{aligned}$$

where 252 is the usual number of trading days in a year and RF is the annualized geometric average of risk-free rates. According to Wharton Research Data Services (WRDS), during the period from 2 January 1980 to 31 December 2014, the annualized average of risk-free rate was approximately 4.7%.

The turnover ratio TO_k at the start of the kth period is defined by the sum of the absolute values of the portfolio adjustment:

$$\text{TO}_k := \sum_{j=1}^{p} \left| w_{t_k j} - w_{(t_k - 1)j} \right|.$$

The wealth growth W_t at date t is defined as follows:

$$W_t := \begin{cases} W_{t-1}(1+R_t)(1-\eta \mathrm{TO}_k) & \text{if } t = t_k \text{ for some } k, \\ W_{t-1}(1+R_t) & \text{otherwise}, \end{cases}$$

where η is the cost of buying or selling a unit share of stock and the initial wealth W_0 is set to 1.

6.3 Results

6.3.1 Whole period

We first compare the performances of the portfolios for the entire period from January 1978 to December 2014. The annualized returns by eight methods are reported in Table 1. Regarding the Sharpe ratio, 'Adap+FSPD (μ_∞)' performs best among all methods, and 'Univ+FSPD (μ_∞)' also performs better than other FSPD modifications based on 'Univ'. In addition, the initial estimator 'Adap' leads to better performance than 'Univ' after the PD modification. Figure 6 plots the normalized wealth growth and turnover of the eight methods. First, we clearly see the advantage of the PD modifications compared to the original with respect to the turnover rate and normalized wealth. As we read from the formula of the optimal MVP allocation, the small (or negative) eigenvalues are influential to the estimated optimal allocation. The PD modification makes it stable and, as a consequence, significantly reduces the rate of turnover and enables greater wealth growth. Second, again, the initial estimator 'Adap' leads to better performance than 'Univ' after the PD modification. Third, although the differences are not large, FSPD with $\mu = \mu_\infty$ (ℓ_∞-FSPD) performs well compared to other FSPDs. Figure 7 shows the eigenvalues of covariance matrix estimates $\widehat{\boldsymbol{\Sigma}}(M, t^*)$, where t^* is 1 May 2010. We can see that both 'Univ' and 'Adap' have negative eigenvalues, whereas the FSPD estimates are positive definite. We find, in both 'Univ' and 'Adap', large eigenvalues (eigenvalues away from 0) of FSPD with $\mu = \mu_\infty$, which are closer to those of the initial estimators than other FSPDs ($\mu = \mu_{\mathrm{SF}}, \infty$). We conjecture that this results in good performance of the FSPD with $\mu = \mu_\infty$.

Table 1 Whole period: annualized mean return, annualized standard deviation of returns, and Sharpe ratio

Estimator	Mean	SD	Sharpe ratio
Adap	12.579	15.040	0.5239
Adap + FSPD (μ_{SF})	13.163	12.405	0.6822
Adap + FSPD (∞)	13.159	12.409	0.6817
Adap + FSPD (μ_∞)	13.210	12.396	0.6865
Univ	11.618	23.773	0.2910
Univ + FSPD (μ_{SF})	13.045	12.629	0.6608
Univ + FSPD (∞)	13.039	12.638	0.6598
Univ + FSPD (μ_∞)	13.074	12.605	0.6643

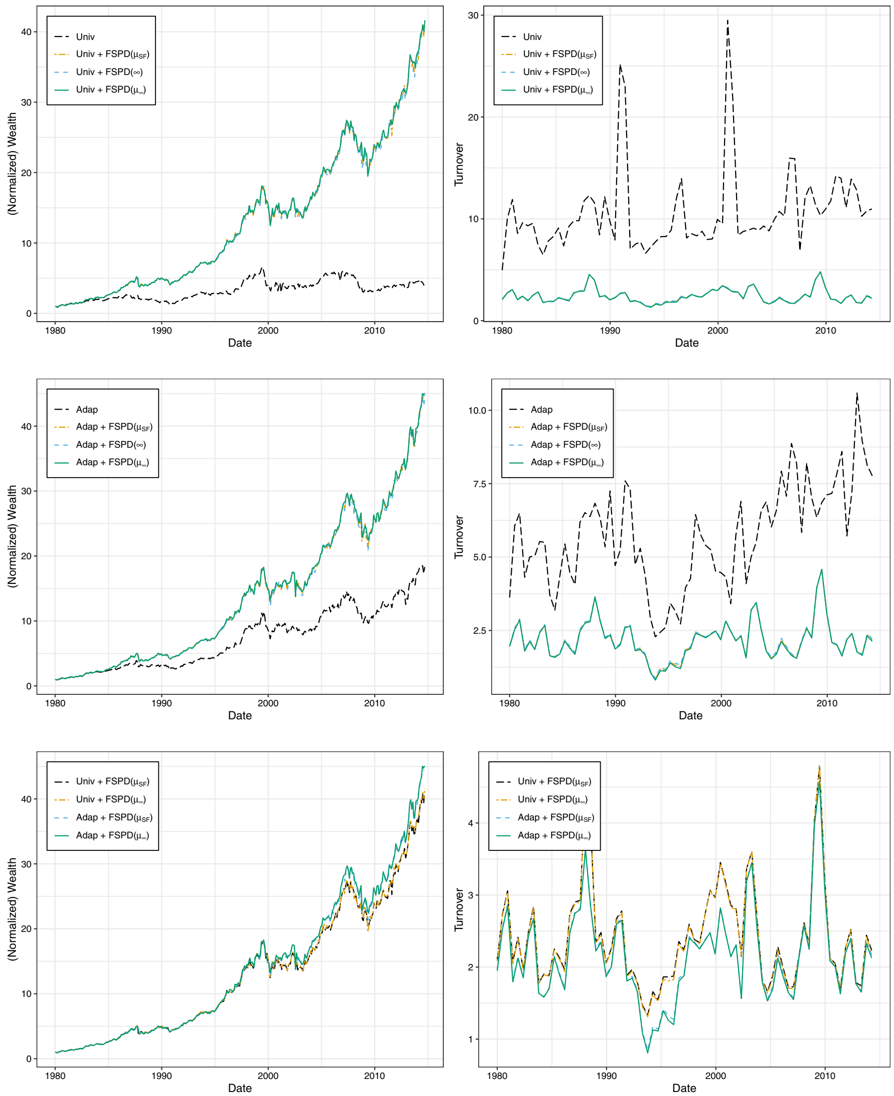


Fig. 6 Normalized wealth growth and turnover

6.3.2 Subperiods: two market crises and one bull market

We next apply the methods to three subperiods. The three subperiods that we focus on are (i) the period from 1987-01-02 to 1988-12-30, which contains the crisis named Black Monday, (ii) the period from 2007-01-03 to 2008-12-31, which contains the global financial crisis (GFC), and (iii) the bull run period from 2010-01-04 to 2014-12-31 due to the rise of the tech industry. As in the whole period analysis, we compare the annualized mean return, annualized standard deviation of returns,

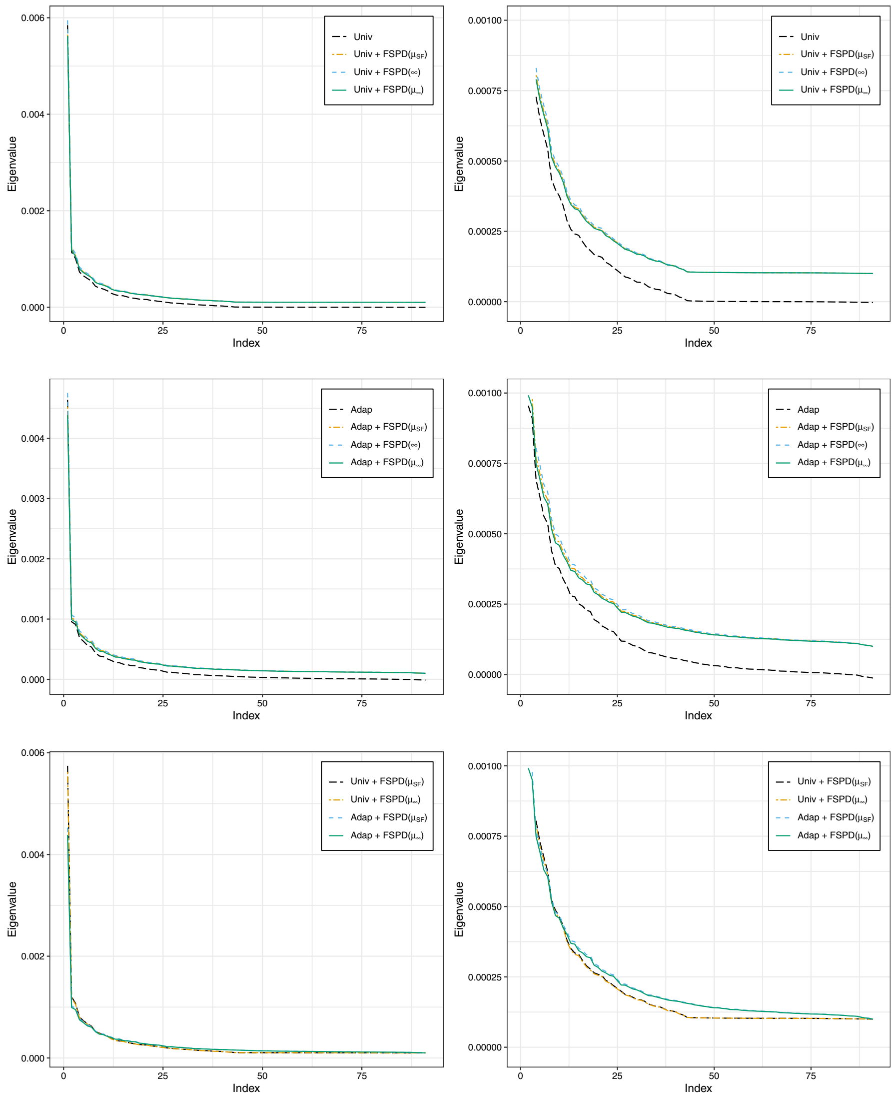


Fig. 7 Eigenvalues of the covariance matrix estimates $\widehat{\mathbf{\Sigma}}(120, t^*)$, where t^* is 1 May 2010. The right plot is a magnification of the left plot

and Sharpe ratio by eight methods for each subperiod, as reported in Table 2. The table shows that the use of FSPD introduces a great increase in Sharpe ratio. However, the Sharpe ratios of three FSPDs are almost identical to each other. We would expect that this result is because the lengths of subperiods are not large enough to differentiate them.

We next plot the normalized wealth growth by each method in Fig. 8. From the figure, we find that the portfolios with FSPD perform similar to those without FSPD

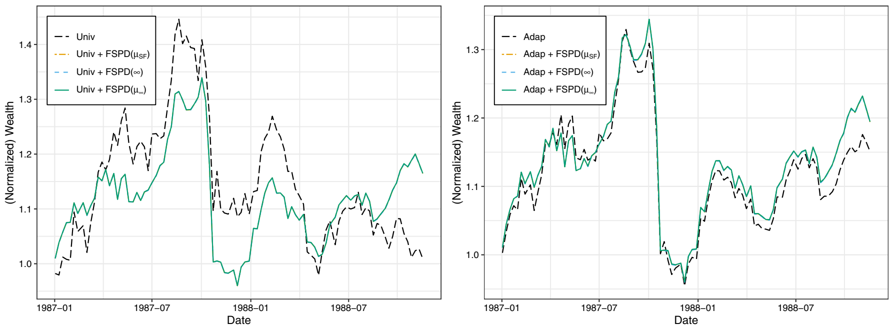
Table 2 Subperiods: annualized mean return, annualized standard deviation of returns, and Sharpe ratio

	Estimator	Mean	SD	Sharpe ratio
Sub period 1	Adap	10.570	23.156	0.2535
	Adap + FSPD (μ_{SF})	11.352	21.880	0.3040
	Adap + FSPD (∞)	11.329	21.884	0.3029
	Adap + FSPD (μ_∞)	11.368	21.878	0.3048
	Univ	6.734	29.977	0.0679
	Univ + FSPD (μ_{SF})	10.319	22.308	0.2519
	Univ + FSPD (∞)	10.303	22.315	0.2511
	Univ + FSPD (μ_∞)	10.331	22.303	0.2525
Sub period 2	Adap	−2.765	18.950	−0.3939
	Adap + FSPD (μ_{SF})	−0.276	17.797	−0.2796
	Adap + FSPD (∞)	−0.264	17.781	-0.2792
	Adap + FSPD (μ_∞)	−0.287	17.810	-0.2800
	Univ	-8.936	21.715	-0.6279
	Univ + FSPD (μ_{SF})	−1.267	17.753	−0.3361
	Univ + FSPD (∞)	−1.273	17.738	−0.3368
	Univ + FSPD (μ_∞)	−1.272	17.764	−0.3362
Sub period 3	Adap	16.386	13.227	0.8835
	Adap + FSPD (μ_{SF})	15.269	9.900	1.0676
	Adap + FSPD (∞)	15.300	9.895	1.0712
	Adap + FSPD (μ_∞)	15.250	9.903	1.0653
	Univ	15.734	15.740	0.7010
	Univ + FSPD (μ_{SF})	15.478	9.911	1.0875
	Univ + FSPD (∞)	15.511	9.907	1.0913
	Univ + FSPD (μ_∞)	15.460	9.913	1.0854

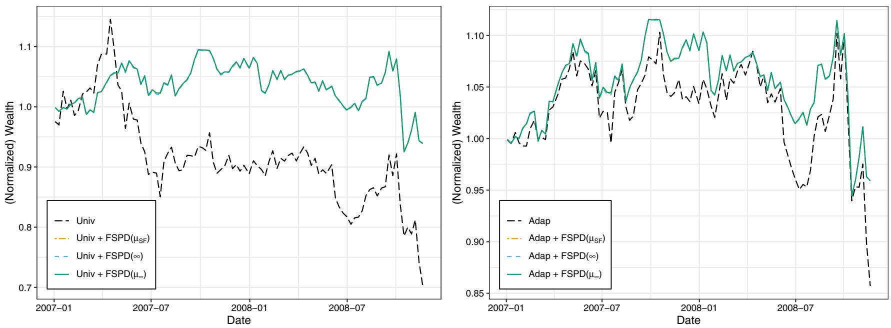
at the beginning but become better at the end of the period. Again, the portfolios with different FSPDs are very close to each other, and it is difficult to differentiate their performance.

7 Conclusion

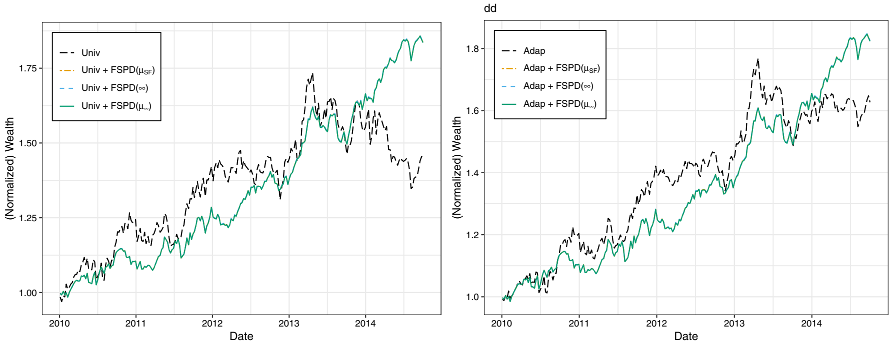
In this paper, in view of methodology, we propose a new FSPD modification for the high-dimensional regularized covariance matrix estimator, ℓ_∞–FSPD, which preserves the support and matrix ℓ_∞ convergence rate of the initial estimator. It is an extension of the FSPD modification by Choi et al. (2019b) from the spectral and Frobenius norms to matrix ℓ_∞ (equivalently, matrix ℓ_1 norm). Like the original FSPD, ℓ_∞–FSPD is optimization-free, preserves the support and matrix ℓ_∞ convergence rate of the initial estimator, and is generic in the sense that it can be applicable to any non-PD symmetric matrix including non-PD precision or correlation matrix estimators. Further, the matrix ℓ_∞ error of the initial covariance estimator often bounds the error of the results of the multivariate procedure using it. In this



(a) 1987~1988



(b) 2007~2008



(c) 2010~2014

Fig. 8 Normalized wealth growth and turnover for each subperiod. **a** From 2 January 1987 to 30 December 1988; **b** from 3 January 2007 to 31 December 2008; **c** from 4 January 2010 to 31 December 2014

paper, we study the MVP optimization as an application of ℓ_∞-FSPD. The ℓ_∞ error of the estimated optimal portfolio allocation is bounded by the matrix ℓ_∞ error of the plug-in FSPD estimator.

Second, in view of portfolio optimization, this paper says the choice of covariance matrix estimator is important in selecting portfolio from a large set of stocks. We find that the FSPD estimators could be a good choice for the high-dimensional minimum variance portfolio optimization problem. In addition, the portfolio weight with the ℓ_∞–FSPD estimator shows the best performance in both numerical study and real data example.

Finally, in this paper, we discuss MVP as an example whose error of the resulting estimator (the estimate of optimal portfolio allocation) is bounded by the matrix ℓ_∞ error of the plug-in covariance matrix estimator. However, such examples could be numerous and are not limited to this one in particular. Another example that we have in mind is the high-dimensional regression wherein the sup-norm of the estimator of the regression coefficient is bounded by the matrix ℓ_∞–norm of the Gram matrix of the covariate vectors.

Acknowledgements We are grateful to the associate editor and two reviewers for many valuable comments. S. Katayama's research is supported by JSPS KAKENHI Grant No. 18K18009, and J. Lim and Y-G. Choi's research is supported by the National Research Foundation of Korea (NRF-2017R1A2B2012264 and NRF-2020R1G1A1A01006229).

References

Bickel, P., Levina, E.: Covariance regularization by thresholding. Ann. Stat. **36**(6), 2577–2604 (2008a)

Bickel, P., Levina, E.: Regularized estimation of large covariance matrices. Ann. Stat. **36**(1), 199–227 (2008b)

Cai, T.T., Liu, W.: Adaptive thresholding for sparse covariance matrix estimation. J. Am. Stat. Assoc. **106**(494), 672–684 (2011)

Cai, T.T., Zhang, C.-H., Zhou, H.H.: Optimal rates of convergence for covariance matrix estimation. Ann. Stat. **38**(4), 2118–2144 (2010)

Cai, T.T., Zhou, H.H.: Minimax estimation of large covariance matrices under ℓ_1-norm. Stat. Sin. **22**(4), 1319–1349 (2012a)

Cai, T.T., Zhou, H.H.: Optimal rates of convergence for sparse covariance matrix estimation. Ann. Stat. **40**(5), 2389–2420 (2012b)

Chan, L.K.C., Karceski, J., Lakonishok, J.: On portfolio optimization: forecasting covariances and choosing the risk model. Rev. Financ. Stud. **12**(5), 937–974 (2015)

Choi, Y.-G., Lim, J., Choi, S.: High-dimensional Markowitz portfolio optimization problem: empirical comparison of covariance matrix estimators. J. Stat. Comput. Simul. **89**(7), 1278–1300 (2019a)

Choi, Y.-G., Lim, J., Roy, A., Park, J.: Fixed support positive-definite modification of covariance matrix estimators via linear shrinkage. J. Multivar. Anal. **171**, 234–249 (2019b)

Dai, Z., Dong, X., Kang, J., Hong, L.: Forecasting stock market returns: new technical indicators and two-step economic constraint method. N. Am. J. Econ. Finance **53**, 101216 (2020)

Dai, Z., Wen, F.: Some improved sparse and stable portfolio optimization problems. Finance Res. Lett. **27**, 46–52 (2018)

DeMiguel, V., Garlappi, L., Nogales, F.J., Uppal, R.: A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. Manag.Sci. **55**(5), 798–812 (2009)

Fan, J., Liao, Y., Mincheva, M.: Large covariance estimation by thresholding principal orthogonal complements. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **75**(4), 603–680 (2013)

Glasserman, P., Kang, W.: OR forum-design of risk weights. Oper. Res. **62**(6), 1204–1220 (2014)

Khare, K., Oh, S.-Y., Rajaratnam, B.: A convex pseudo-likelihood framework for high dimensional partial correlation estimation with convergence guarantees. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **77**(4), 803–825 (2015)

Ledoit, O., Wolf, M.: Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empir. Finance **10**(5), 603–621 (2003)

Ledoit, O., Wolf, M.: Honey, I shrunk the sample covariance matrix. J. Portfolio Manag. **30**(4), 110–119 (2004)

Ledoit, O., Wolf, M.: Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. Rev. Financ. Stud. **30**(12), 4349–4388 (2017a)

Ledoit, O., Wolf, M.: Numerical implementation of the QuEST function. Comput. Stat. Data Anal. **115**, 199–223 (2017b)

Liu, H., Wang, L., Zhao, T.: Sparse covariance matrix estimation with Eigenvalue constraints. J. Comput. Gr. Stat. **23**(2), 439–459 (2014)

Rothman, A.J.: Positive definite estimators of large covariance matrices. Biometrika **99**(3), 733–740 (2012)

Won, J.-H., Lim, J., Kim, S.-J., Rajaratnam, B.: Condition-number-regularized covariance estimation. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **75**(3), 427–450 (2013)

Xue, L., Ma, S., Zou, H.: Positive-definite ℓ_1-penalized estimation of large covariance matrices. J. Am. Stat. Assoc. **107**(500), 1480–1491 (2012)