Journal of Econometrics 147 (2008) 186–197

Contents lists available at ScienceDirect

Journal of Econometrics

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

High dimensional covariance matrix estimation using a factor model[☆]

Jianqing Fan[a], Yingying Fan[b], Jinchi Lv[b,*]

[a] Department of Operations Research and Financial Engineering, Princeton University, United States

[b] Information and Operations Management Department, Marshall School of Business, University of Southern California, United States

ARTICLE INFO

Article history:
Available online 18 September 2008

JEL classification:
C13
C51

Keywords:
Factor model
Diverging dimensionality
Covariance matrix estimation
Asymptotic properties
Portfolio management

ABSTRACT

High dimensionality comparable to sample size is common in many statistical problems. We examine covariance matrix estimation in the asymptotic framework that the dimensionality p tends to ∞ as the sample size n increases. Motivated by the Arbitrage Pricing Theory in finance, a multi-factor model is employed to reduce dimensionality and to estimate the covariance matrix. The factors are observable and the number of factors K is allowed to grow with p. We investigate the impact of p and K on the performance of the model-based covariance matrix estimator. Under mild assumptions, we have established convergence rates and asymptotic normality of the model-based estimator. Its performance is compared with that of the sample covariance matrix. We identify situations under which the factor approach increases performance substantially or marginally. The impacts of covariance matrix estimation on optimal portfolio allocation and portfolio risk assessment are studied. The asymptotic results are supported by a thorough simulation study.

1. Introduction

Covariance matrix estimation is fundamental for almost all areas of multivariate analysis and many other applied problems. In particular, covariance matrices and their inverses play a central role in portfolio risk assessment and optimal portfolio allocation. For example, the smallest and largest eigenvalues of a covariance matrix are related to the minimum and maximum variances of the selected portfolio, respectively, and the eigenvectors are related to optimal portfolio allocation. Therefore, we need a good covariance matrix estimator inverting which does not excessively amplify the estimation error. See Goldfarb and Iyengar (2003) for applications of covariance matrices to portfolio selections and Johnstone (2001) for their statistical implications.

Estimating high-dimensional covariance matrices is intrinsically challenging. For example, in optimal portfolio allocation and portfolio risk assessment, the number of stocks p, which is typically of the same order as the sample size n, can well be in the order of hundreds. In particular, when $p = 200$ there are more than 20,000 parameters in the covariance matrix. Yet, the available sample size is usually in the order of hundreds or a few thousand because longer time series (larger n) increases modeling bias. For instance, by taking daily data of the past three years we have only roughly $n = 750$. So it is hard or even unrealistic to estimate covariance matrices without imposing any structure (see the rejoinder in Fan (2005)).

Factor models have been widely used both theoretically and empirically in economics and finance. Derived by Ross (1976, 1977) using the Arbitrage Pricing Theory (APT) and by Chamberlain and Rothschild (1983) in a large economy, the multi-factor model states that the excessive return of any asset Y_i over the risk-free interest rate satisfies

$$Y_i = b_{i1}f_1 + \cdots + b_{iK}f_K + \varepsilon_i, \quad i = 1, \ldots, p, \tag{1}$$

where $f_1, \ldots, f_K$ are the excessive returns of K factors, b_{ij}, $i = 1, \ldots, p, j = 1, \ldots, K$, are unknown factor loadings, and $\varepsilon_1, \ldots, \varepsilon_p$ are p idiosyncratic errors uncorrelated given $f_1, \ldots, f_K$. The factor models have been widely applied and studied in economics and finance. See, for example, Ross (1976, 1977), Engle and Watson (1981), Chamberlain (1983), Chamberlain and Rothschild (1983), Diebold and Nerlove (1989), Fama and French (1992, 1993), Aguilar and West (2000), Bai (2003), Ledoit and Wolf (2003), Stock and Watson (2005) and references therein. These are extensions of the famous Capital Asset Pricing Model (CAPM) and can be regarded as efforts to approximate the market portfolio in the CAPM.

Thanks to the multi-factor model (1), if a few factors can completely capture the cross-sectional risks, the number of

☆ Financial support from the NSF under grant DMS-0532370 is gratefully acknowledged. Lv's research was partially supported by National Science Foundation grant DMS-0806030 and the 2008 Zumberge Individual Award from the James H. Zumberge Faculty Research and Innovation Fund at the University of Southern California. We are grateful to the guest editor, Jiti Gao and two anonymous referees for their constructive and helpful comments.

* Corresponding address: Information and Operations Management Department, Marshall School of Business, University of Southern California, 90089 Los Angeles, CA, United States. Tel.: +1 213 740 6603.

E-mail address: jinchilv@marshall.usc.edu (J. Lv).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.09.017

parameters in covariance matrix estimation can be significantly reduced. For example, using the Fama-French three-factor model (Fama and French, 1992, 1993), there are $4p$ instead of $p(p+1)/2$ parameters to be estimated. Despite the popularity of factor models in the literature, the impact of dimensionality on the estimation errors of covariance matrices and its applications to optimal portfolio allocation and portfolio risk assessment are poorly understood so, in this paper, determined efforts are made on such an investigation. To make the multi-factor model more realistic, we allow K to grow with the number of assets p and hence with the sample size n. As a result, we also investigate the impact of the number of factors on the estimation of covariance matrices, as well as its applications to optimal portfolio allocation and portfolio risk assessment. To appreciate the derived rates of convergence, we compare them with those without using the factor structure. One natural candidate is the sample covariance matrix. This also allows us to examine the impact of dimensionality on the performance of the sample covariance matrix. We will assume that the factors are observable as in Fama and French (1992, 1993). Our results also provide an important milestone for understanding the performance of factor models with unobservable factors.

The traditional covariance matrix estimator, the sample covariance matrix, is known to be unbiased, and it is invertible when the dimensionality is no larger than the sample size. See, for example, Eaton and Tyler (1994) for the asymptotic spectral distributions of random matrices including sample covariance matrices and their statistical implications. In the absence of prior information about the population covariance matrix, the sample covariance matrix is certainly a natural candidate in the case of small dimensionality, but no longer performs very well for moderate or large dimensionality [see, e.g. Lin and Perlman (1985) and Johnstone (2001)]. Many approaches were proposed in the literature to construct good covariance matrix estimators. Among them, two main directions were taken. One is to remedy the sample covariance matrix and construct a better one by using approaches such as shrinkage and the eigen-method, etc. See, for example, Ledoit and Wolf (2004) and Stein (1975). The other one is to reduce dimensionality by imposing some structure on the data. Many structures, such as sparsity, compound symmetry, and the autoregressive model, are widely used. Various approaches were taken to seek a balance between the bias and variance of covariance matrix estimators. See, for example, Wong et al. (2003), Huang et al. (2006), and Bickel and Levina (2008).

The rest of the paper is organized as follows. Section 2 introduces the estimators of the covariance matrix. In Section 3 we give some basic assumptions and present sampling properties of the estimators. We study the impacts of covariance matrix estimation on optimal portfolio allocation and portfolio risk assessment in Section 4. A simulation study is presented in Section 5, which augments our theoretical study. Section 6 contains some concluding remarks. The proofs of our results are given in the Appendix.

2. Covariance matrix estimation

We always denote by n the sample size, by p the dimensionality, and by $f_1, \ldots, f_K$ the K observable factors, where p grows with sample size n and K increases with dimensionality p. For ease of presentation, we rewrite the factor model (1) in matrix form

$$\mathbf{y} = \mathbf{B}_n\mathbf{f} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{y} = (Y_1, \ldots, Y_p)'$, $\mathbf{B}_n = (\mathbf{b}_1, \ldots, \mathbf{b}_p)'$ with $\mathbf{b}_i = (b_{n,i1}, \ldots, b_{n,iK})'$, $i = 1, \ldots, p$, $\mathbf{f} = (f_1, \ldots, f_K)'$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)'$. Throughout we assume that $E(\boldsymbol{\varepsilon}|\mathbf{f}) = \mathbf{0}$ and $\mathrm{cov}(\boldsymbol{\varepsilon}|\mathbf{f}) = \boldsymbol{\Sigma}_{n,0}$ is diagonal. For brevity of notation, we suppress the first subscript n in some situations where the dependence on n is self-evident.

Let $(\mathbf{f}_1, \mathbf{y}_1), \ldots, (\mathbf{f}_n, \mathbf{y}_n)$ be n independent and identically distributed (i.i.d.) samples of $(\mathbf{f}, \mathbf{y})$. We introduce here some notation used throughout the paper. Let

$$\boldsymbol{\Sigma}_n = \mathrm{cov}(\mathbf{y}), \quad \mathbf{X} = (\mathbf{f}_1, \ldots, \mathbf{f}_n),$$
$$\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n) \quad \text{and} \quad \mathbf{E} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n).$$

Under model (2), we have

$$\boldsymbol{\Sigma}_n = \mathrm{cov}(\mathbf{B}_n\mathbf{f}) + \mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{B}_n\mathrm{cov}(\mathbf{f})\mathbf{B}_n' + \boldsymbol{\Sigma}_{n,0}. \tag{3}$$

A natural idea for estimating $\boldsymbol{\Sigma}_n$ is to plug in the least-squares estimators of $\mathbf{B}_n$, $\mathrm{cov}(\mathbf{f})$, and $\boldsymbol{\Sigma}_{n,0}$. Therefore, we have a substitution estimator

$$\hat{\boldsymbol{\Sigma}}_n = \widehat{\mathbf{B}}_n\widehat{\mathrm{cov}}(\mathbf{f})\widehat{\mathbf{B}}_n' + \hat{\boldsymbol{\Sigma}}_{n,0}, \tag{4}$$

where $\widehat{\mathbf{B}}_n = \mathbf{YX}'(\mathbf{XX}')^{-1}$ is the matrix of estimated regression coefficients, $\widehat{\mathrm{cov}}(\mathbf{f}) = (n-1)^{-1}\mathbf{XX}' - \{n(n-1)\}^{-1}\mathbf{X11}'\mathbf{X}'$ is the sample covariance matrix of the factors $\mathbf{f}$, and

$$\hat{\boldsymbol{\Sigma}}_{n,0} = \mathrm{diag}\left(n^{-1}\widehat{\mathbf{E}}\widehat{\mathbf{E}}'\right)$$

is the diagonal matrix of $n^{-1}\widehat{\mathbf{E}}\widehat{\mathbf{E}}'$ with $\widehat{\mathbf{E}} = \mathbf{Y} - \widehat{\mathbf{B}}\mathbf{X}$ the matrix of residuals. If the factor model is not employed, then we have the sample covariance matrix estimator

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}} = (n-1)^{-1}\,\mathbf{YY}' - \{n\,(n-1)\}^{-1}\,\mathbf{Y11}'\mathbf{Y}'. \tag{5}$$

Ledoit and Wolf (2003) propose an interesting idea of combining the single-index ($K = 1$, CAPM) model based estimation of the covariance matrix with the sample covariance matrix to improve the estimate of the covariance matrix. It aims at a trade-off between the bias and variance of the two estimated covariance matrices for practical applications.

In the paper we mainly aim to provide a theoretical understanding of the factor model with a diverging dimensionality and growing number of factors for the purpose of covariance matrix estimation, but not to compare with other popular estimators. Throughout the paper, we always contrast the performance of the covariance matrix estimator $\hat{\boldsymbol{\Sigma}}$ in (4) with that of the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ in (5). The paper also provides a theoretical study on the two estimators used in the procedure of Ledoit and Wolf (2003). With prior information of the true factor structure, the substitution estimator $\hat{\boldsymbol{\Sigma}}$ is expected to perform better than $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$. However, this has not been formally shown, especially when $p \to \infty$ and $K \to \infty$, and this is not always true. In addition, exact properties of this kind are not well understood. As the problem is important for portfolio management, determined efforts are devoted in this regard.

3. Sampling properties

In this section we study the sampling properties of $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ with growing dimensionality and number of factors. We give some basic assumptions in Section 3.1. The sampling properties are presented in Section 3.2.

In the presence of diverging dimensionality, one should carefully choose appropriate norms for large matrices in different situations. We first introduce some notation. We always denote by $\lambda_1(\mathbf{A}), \ldots, \lambda_q(\mathbf{A})$ the q eigenvalues of a $q \times q$ symmetric matrix $\mathbf{A}$ in decreasing order. For any matrix $\mathbf{A} = (a_{ij})$, its Frobenius norm is given by

$$\|\mathbf{A}\| = \left\{\mathrm{tr}(\mathbf{AA}')\right\}^{1/2}. \tag{6}$$

In particular, if $\mathbf{A}$ is a $q \times q$ symmetric matrix, then $\|\mathbf{A}\| = \left\{\sum_{i=1}^{q} \lambda_i(\mathbf{A})^2\right\}^{1/2}$. The Frobenius norm as well as many other matrix norms [see Horn and Johnson (1990)] is intrinsically related to the eigenvalues or singular values of matrices.

Despite its popularity, the Frobenius norm is not appropriate for understanding the performance of the factor-model based estimation of the covariance matrix. To see this, let us consider a simple example. Suppose we know ideally that $\mathbf{B} = \mathbf{1}$ and $\text{cov}(\boldsymbol{\varepsilon}|f) = I_p$ in model (2) with a single factor f. Then we have a substitution covariance matrix estimator $\hat{\boldsymbol{\Sigma}} = \mathbf{1}\widehat{\text{var}}(f)\mathbf{1}' + I_p$ as in (4). It is a classical result that

$$E\,|\widehat{\text{var}}(f) - \text{var}(f)|^2 = O(n^{-1}).$$

Thus by (3), we have

$$\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} = \mathbf{1}\,[\widehat{\text{var}}(f) - \text{var}(f)]\,\mathbf{1}'$$

and the Frobenius norm $\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\| = |\widehat{\text{var}}(f) - \text{var}(f)|\,p$ picks up every element from the matrix and amplifies the estimation error from $\widehat{\text{var}}(f)$. Consequently,

$$E\left\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|^2 = O(n^{-1}p^2).$$

On the other hand, by assuming boundedness of the fourth moments of $\mathbf{y}$ across n, a routine calculation reveals that

$$E\left\|\hat{\boldsymbol{\Sigma}}_{\text{sam}} - \boldsymbol{\Sigma}\right\|^2 = O(n^{-1}p^2).$$

This shows that under Frobenius norm, $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\text{sam}}$ have the same convergence rate and perform roughly the same. Hence, the Frobenius norm does not help us understand the factor structure. Thus we should seek other norms that fully employ the factor structure.

James and Stein (1961) introduce the entropy (or Stein) loss function

$$L_1(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \text{tr}\left(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right) - \log\left|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right| - p.$$

Inspired by this and motivated by the above example, we introduce below a new norm $\|\cdot\|_{\boldsymbol{\Sigma}}$ that is closely related to the entropy loss and quadratic loss $L_2(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \text{tr}[\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - I]^2$. Fix a sequence of positive definite covariance matrices $\boldsymbol{\Sigma}_n$ of dimensionality p_n, $n = 1, 2, \ldots,$ and define

$$\|\mathbf{A}\|_{\boldsymbol{\Sigma}_n} = p_n^{-1/2}\left\|\boldsymbol{\Sigma}_n^{-1/2}\mathbf{A}\boldsymbol{\Sigma}_n^{-1/2}\right\| \quad (7)$$

for any $p_n \times p_n$ matrix $\mathbf{A}$. In particular, we have $\|\boldsymbol{\Sigma}_n\|_{\boldsymbol{\Sigma}_n} = p^{-1/2}\|I_p\| = 1$. Note that

$$\left\|\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n\right\|_{\boldsymbol{\Sigma}_n} = p_n^{-1/2}L_2(\hat{\boldsymbol{\Sigma}}_n, \boldsymbol{\Sigma}_n),$$

so it is really only a rescale of the quadratic loss. The inclusion of a normalization factor $p^{-1/2}$ above is not essential and we incorporate it to take into account the diverging dimensionality. Intrinsically, this norm takes into account and fully employs the factor structure since it involves the inverse of the covariance matrix. We will see later in this section that under norm $\|\cdot\|_{\boldsymbol{\Sigma}}$, the consistency rate in the factor approach is better than that in the sample approach, and the factor approach has much advantage in estimating the inverse of covariance matrix.

3.1. Some basic assumptions

Let $b_n = E\|\mathbf{y}\|^2$, $c_n = \max_{1\le i\le K} E(f_i^4)$, and $d_n = \max_{1\le i\le p} E(\varepsilon_i^4)$.

(A) $(\mathbf{f}_1, \mathbf{y}_1), \ldots, (\mathbf{f}_n, \mathbf{y}_n)$ are n i.i.d. samples of $(\mathbf{f}, \mathbf{y})$. $E(\boldsymbol{\varepsilon}|\mathbf{f}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\varepsilon}|\mathbf{f}) = \boldsymbol{\Sigma}_{n,0}$ is diagonal. Also, the distribution of $\mathbf{f}$ is continuous and $K \le p$.

(B) $b_n = O(p)$ and the sequences c_n and d_n are bounded. Also, there exists a constant $\sigma_1 > 0$ such that $\lambda_K(\text{cov}(\mathbf{f})) \ge \sigma_1$ for all n.

(C) There exists a constant $\sigma_2 > 0$ such that $\lambda_p(\boldsymbol{\Sigma}_{n,0}) \ge \sigma_2$ for all n.

(D) The K factors $f_1, \ldots, f_K$ are fixed across n, and $p^{-1}\mathbf{B}_n'\mathbf{B}_n \to \mathbf{A}$ as $n \to \infty$ for some $K \times K$ symmetric positive semidefinite matrix $\mathbf{A}$.

3.2. Sampling properties

Theorem 1 (*Rates of Convergence Under Frobenius Norm*). *Under conditions* (A) *and* (B), *we have* $\left\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\| = O_P(n^{-1/2}pK)$ *and* $\left\|\hat{\boldsymbol{\Sigma}}_{\text{sam}} - \boldsymbol{\Sigma}\right\| = O_P(n^{-1/2}pK)$. *In addition, we have*

$$\max_{1\le k\le p}\left|\lambda_k(\hat{\boldsymbol{\Sigma}}_n) - \lambda_k(\boldsymbol{\Sigma}_n)\right| = O_P(n^{-1/2}pK)$$

and

$$\max_{1\le k\le p}\left|\lambda_k(\hat{\boldsymbol{\Sigma}}_{\text{sam}}) - \lambda_k(\boldsymbol{\Sigma}_n)\right| = O_P(n^{-1/2}pK).$$

From this theorem, we see that under the Frobenius norm, the dimensionality reduces rates of convergence by an order of pK, which is the order of the number of parameters. The above rate of eigenvalues of $\hat{\boldsymbol{\Sigma}}$ is optimal. To see it, let us extend the previous example by including K factors $f_1, \ldots, f_K$ and setting $\mathbf{B} = (\mathbf{1}, \ldots, \mathbf{1})_{p\times K}$. Further suppose we know ideally that $\text{cov}(\mathbf{f}) = \text{var}(f_1)I_K$. Then we have

$$\boldsymbol{\Sigma}_n = I_p + \text{var}(f_1)K\mathbf{1}\mathbf{1}' \quad \text{and}$$
$$\hat{\boldsymbol{\Sigma}}_n = I_p + \widehat{\text{var}}(f_1)K\mathbf{1}\mathbf{1}'.$$

It is easy to see that $\lambda_1(\boldsymbol{\Sigma}_n) = \text{var}(f_1)pK + 1$, $\lambda_k(\boldsymbol{\Sigma}_n) = 1$, $k = 2, \ldots, p$ and $\lambda_1(\hat{\boldsymbol{\Sigma}}_n) = \widehat{\text{var}}(f_1)pK + 1$, $\lambda_k(\hat{\boldsymbol{\Sigma}}_n) = 1$, $k = 2, \ldots, p$. Thus,

$$\max_{1\le k\le p}\left|\lambda_k(\hat{\boldsymbol{\Sigma}}_n) - \lambda_k(\boldsymbol{\Sigma}_n)\right| = |\widehat{\text{var}}(f_1) - \text{var}(f_1)|\,pK$$
$$= O_P(n^{-1/2}pK).$$

Therefore, $\hat{\boldsymbol{\Sigma}}$ here attains the optimal uniform weak convergence rate of eigenvalues.

Theorem 1 shows that the factor structure does not give much advantage in estimating $\boldsymbol{\Sigma}$. The next theorem shows that when $\boldsymbol{\Sigma}^{-1}$ is involved, the rate of convergence is improved.

Theorem 2 (*Rates of Convergence Under Norm* $\|\cdot\|_{\boldsymbol{\Sigma}}$). *Suppose that* $K = O(n^{\alpha_1})$ *and* $p = O(n^{\alpha})$. *Under conditions* (A)–(C), *we have* $\left\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\boldsymbol{\Sigma}} = O_P(n^{-\beta/2})$ *with* $\beta = \min(1 - 2\alpha_1, 2 - \alpha - \alpha_1)$ *and* $\left\|\hat{\boldsymbol{\Sigma}}_{\text{sam}} - \boldsymbol{\Sigma}\right\|_{\boldsymbol{\Sigma}} = O_P(n^{-\beta_1/2})$ *with* $\beta_1 = 1 - \max(\alpha, 3\alpha_1/2, 3\alpha_1 - \alpha)$.

It is easy to show that $\beta > \beta_1$ whenever $\alpha > 2\alpha_1$ and $\alpha_1 < 1$. Hence, the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{\text{sam}}$ has slower convergence. An interesting case is $K = O(1)$. In this case, under the norm $\|\cdot\|_{\boldsymbol{\Sigma}}$, $\hat{\boldsymbol{\Sigma}}$ has convergence rate $n^{-\beta/2}$ with $\beta = \min(1, 2 - \alpha)$, whereas $\hat{\boldsymbol{\Sigma}}_{\text{sam}}$ has slower convergence rate $n^{-\beta_1/2}$ with $\beta_1 = 1 - \alpha$. In particular, when $\alpha \le 1$, $\hat{\boldsymbol{\Sigma}}$ is root-n-consistent under $\|\cdot\|_{\boldsymbol{\Sigma}}$. This can be shown to be optimal by some calculations using a specific factor model mentioned above.

Theorem 3 (*Rates of Convergence of Inverse Under Frobenius Norm*). *Under conditions* (A)–(C), *we have*

$$\left\|\hat{\mathbf{\Sigma}}_n^{-1} - \mathbf{\Sigma}_n^{-1}\right\| = o_P\{(p^2K^4 \log n/n)^{1/2}\},$$

whereas

$$\left\|\hat{\mathbf{\Sigma}}_{\text{sam}}^{-1} - \mathbf{\Sigma}_n^{-1}\right\| = o_P\{(p^4K^2 \log n/n)^{1/2}\}.$$

From this theorem, we see that when $K = o(p)$, $\hat{\mathbf{\Sigma}}^{-1}$ performs much better than $\hat{\mathbf{\Sigma}}_{\text{sam}}^{-1}$. As expected, they perform roughly the same in the extreme case where K is proportional to p. It is very pleasing that under an additional assumption (C), $\hat{\mathbf{\Sigma}}^{-1}$ has a consistency rate slightly slower than $\hat{\mathbf{\Sigma}}$ under the Frobenius norm, since $\hat{\mathbf{\Sigma}}^{-1}$ involves the inverse of the $K \times K$ sample covariance matrix of $\mathbf{f}$. The consistency result of $\hat{\mathbf{\Sigma}}_{\text{sam}}^{-1}$ is implied by that of $\hat{\mathbf{\Sigma}}_{\text{sam}}$, thanks to a simple inequality in matrix theory on inverses under perturbation. However, the consistency result of $\hat{\mathbf{\Sigma}}^{-1}$ needs a very delicate analysis of inverse matrices. This theorem will be used in Section 4.1 to examine the volatility of a mean–variance optimal portfolio.

Before going further, we first introduce some standard notation. Let $\mathbf{A} = (a_{ij})$ be a $q \times r$ matrix and denote by $\text{vec}(\mathbf{A})$ the $qr \times 1$ vector formed by stacking the r columns of $\mathbf{A}$ underneath each other in the order from left to right. In particular, for any $d \times d$ symmetric matrix $\mathbf{A}$, we denote by $\text{vech}(\mathbf{A})$ the $d(d+1)/2 \times 1$ vector obtained from $\text{vec}(\mathbf{A})$ by removing the above-diagonal entries of $\mathbf{A}$. It is not difficult to see that there exists a unique $d^2 \times d(d+1)/2$ matrix D_d of zeros and ones such that

$$D_d \,\text{vech}(\mathbf{A}) = \text{vec}(\mathbf{A})$$

for any $d \times d$ symmetric matrix $\mathbf{A}$. D_d is called the duplication matrix of order d. Clearly, for any $d \times d$ symmetric matrix $\mathbf{A}$, we have

$$P_D \text{vec}(\mathbf{A}) = \text{vech}(\mathbf{A}),$$

where $P_D = \left(D'D\right)^{-1} D'$. For any $q \times r$ matrix $\mathbf{A}_1 = (a_{ij})$ and $s \times t$ matrix $\mathbf{A}_2$, we define their Kronecker product $\mathbf{A}_1 \otimes \mathbf{A}_2$ as the $qs \times rt$ matrix $(a_{ij}\mathbf{A}_2)$.

Theorem 4 (*Asymptotic Normality*). *Under conditions* (A), (B), *and* (D), *if* $p \to \infty$ *as* $n \to \infty$, *then the estimator* $\hat{\mathbf{\Sigma}}$ *satisfies*

$$\sqrt{n}\,\text{vech}\left[p^{-2}\mathbf{B}_n'\left(\hat{\mathbf{\Sigma}}_n - \mathbf{\Sigma}_n\right)\mathbf{B}_n\right] \xrightarrow{\text{D}} \mathcal{N}(0, G),$$

where $G = P_D(\mathbf{A} \otimes \mathbf{A})DHD'(\mathbf{A} \otimes \mathbf{A})P_D'$, $H = \text{cov}[\text{vech}(U)]$ *with* $U = (u_{ij})_{K \times K}$ *and*

$$\text{cov}\left(u_{ij}, u_{kl}\right) = \kappa^{ijkl} + \kappa^{ik}\kappa^{jl} + \kappa^{il}\kappa^{jk},$$

$\kappa^{i_1 \cdots i_r}$ *is the central moment* $E\left[(f_{i_1} - Ef_{i_1}) \cdots (f_{i_r} - Ef_{i_r})\right]$ *of* $\mathbf{f} = (f_1, \ldots, f_K)'$, D *is the duplication matrix of order* K, *and* $P_D = \left(D'D\right)^{-1} D'$.

When $\mathbf{f}$ has a K-variate normal distribution with covariance matrix $(\sigma_{ij})_{K \times K}$, the matrix H in Theorem 4 is determined by $\text{cov}\left(u_{ij}, u_{kl}\right) = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}$.

4. Impacts on optimal portfolio allocation and portfolio risk assessment

In this section we examine the impacts of covariance matrix estimation on optimal portfolio allocation and portfolio risk assessment, respectively.

4.1. Impact on optimal portfolio allocation

Markowitz (1952) defines the mean–variance optimal portfolio of p risky assets with expected returns $\boldsymbol{\mu}_n$ and covariance matrix $\mathbf{\Sigma}_n$ as the solution of the allocation vector $\boldsymbol{\xi}_n \in \mathbb{R}^p$ to the following minimization problem

$$\min_{\boldsymbol{\xi}} \ \boldsymbol{\xi}'\mathbf{\Sigma}_n\boldsymbol{\xi} \tag{8}$$

Subject to $\boldsymbol{\xi}'\mathbf{1} = 1$ and $\boldsymbol{\xi}'\boldsymbol{\mu}_n = \gamma_n$,

where $\mathbf{1}$ is a $p \times 1$ vector of ones and γ_n is the expected rate of return imposed on the portfolio. It is well known that Markowitz's optimal portfolio [see Markowitz (1959), Cochrane (2001), or Campbell et al. (1997)] is

$$\boldsymbol{\xi}_n = \frac{\phi_n - \gamma_n\psi_n}{\varphi_n\phi_n - \psi_n^2}\mathbf{\Sigma}_n^{-1}\mathbf{1} + \frac{\gamma_n\varphi_n - \psi_n}{\varphi_n\phi_n - \psi_n^2}\mathbf{\Sigma}_n^{-1}\boldsymbol{\mu}_n \tag{9}$$

with $\varphi_n = \mathbf{1}'\mathbf{\Sigma}_n^{-1}\mathbf{1}$, $\psi_n = \mathbf{1}'\mathbf{\Sigma}_n^{-1}\boldsymbol{\mu}_n$, and $\phi_n = \boldsymbol{\mu}_n'\mathbf{\Sigma}_n^{-1}\boldsymbol{\mu}_n$, and its variance is

$$\boldsymbol{\xi}_n'\mathbf{\Sigma}_n\boldsymbol{\xi}_n = \frac{\varphi_n\gamma_n^2 - 2\psi_n\gamma_n + \phi_n}{\varphi_n\phi_n - \psi_n^2}. \tag{10}$$

Denote by $\boldsymbol{\xi}_{ng}$ the $\boldsymbol{\xi}_n$ in (9) with γ_n replaced by ψ_n/φ_n. The global minimum variance without constraint on the expected return is

$$\boldsymbol{\xi}_{ng}'\mathbf{\Sigma}_n\boldsymbol{\xi}_{ng} = \varphi_n^{-1}, \tag{11}$$

which is attained in (10) when $\gamma_n = \psi_n/\varphi_n$.

Based on the history, we can construct $\hat{\mathbf{\Sigma}}_n$ as before. Also, by the factor model (2), we have a substitution estimator $\widehat{\boldsymbol{\mu}}_n = \widehat{\mathbf{B}}_n n^{-1}(\mathbf{f}_1 + \cdots + \mathbf{f}_n)$ of the mean vector $\boldsymbol{\mu}_n$. As above, we can define estimators $\widehat{\boldsymbol{\xi}}_n$, $\widehat{\boldsymbol{\xi}}_{ng}$ and $\widehat{\varphi}_n$, $\widehat{\psi}_n$, $\widehat{\phi}_n$ with $\mathbf{\Sigma}_n$ and $\boldsymbol{\mu}_n$ replaced by $\hat{\mathbf{\Sigma}}_n$ and $\widehat{\boldsymbol{\mu}}_n$, respectively.

Theorem 5 (*Weak Convergence of Global Minimum Variance*). *Suppose that all the* φ_n*'s are bounded away from zero. Under conditions* (A)–(C), *we have*

$$\widehat{\boldsymbol{\xi}}_{ng}'\hat{\mathbf{\Sigma}}_n\widehat{\boldsymbol{\xi}}_{ng} - \boldsymbol{\xi}_{ng}'\mathbf{\Sigma}_n\boldsymbol{\xi}_{ng} = o_P\{(p^4K^4 \log n/n)^{1/2}\},$$

whereas

$$\widehat{\boldsymbol{\xi}}_{ng}'\hat{\mathbf{\Sigma}}_{\text{sam}}\widehat{\boldsymbol{\xi}}_{ng} - \boldsymbol{\xi}_{ng}'\mathbf{\Sigma}_n\boldsymbol{\xi}_{ng} = o_P\{(p^6K^2 \log n/n)^{1/2}\}.$$

Theorem 6 (*Weak Convergence to Optimal Portfolio*). *Suppose that* $\varphi_n\phi_n - \psi_n^2$ *are bounded away from zero and* $\varphi_n/(\varphi_n\phi_n - \psi_n^2)$, $\psi_n/(\varphi_n\phi_n - \psi_n^2)$, $\phi_n/(\varphi_n\phi_n - \psi_n^2)$, γ_n *are bounded. Under conditions* (A)–(C), *we have*

$$\widehat{\boldsymbol{\xi}}_n'\hat{\mathbf{\Sigma}}_n\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_n'\mathbf{\Sigma}_n\boldsymbol{\xi}_n = o_P\{(p^4K^4 \log n/n)^{1/2}\},$$

whereas

$$\widehat{\boldsymbol{\xi}}_n'\hat{\mathbf{\Sigma}}_{\text{sam}}\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_n'\mathbf{\Sigma}_n\boldsymbol{\xi}_n = o_P\{(p^6K^2 \log n/n)^{1/2}\}.$$

The assumptions on φ_n, ψ_n and ϕ_n in Theorems 5 and 6 are technical and reasonable. In view of (11), the assumption on φ_n in Theorem 5 amounts to saying that the global minimum variances are bounded across n. The additional assumptions in Theorem 6 can be understood in a similar way in light of (10). From the above two theorems, we see that when $K = o(p)$, $\hat{\mathbf{\Sigma}}$ performs much better than $\hat{\mathbf{\Sigma}}_{\text{sam}}$ from the point of view of optimal portfolio allocation. On the other hand, we also see that dimensionality as well as number of factors can only grow slowly with sample size so that the globally optimal portfolio and the mean–variance optimal portfolio constructed using estimated covariance matrix $\hat{\mathbf{\Sigma}}$ or $\hat{\mathbf{\Sigma}}_{\text{sam}}$

behave similarly to theoretical ones. So high dimensionality does impose a great challenge on optimal portfolio allocation.

Our study reveals that for a large number of stocks, additional structures are needed. For example, we may group assets according to sectors and assume that the sector correlations are weak and negligible. Hence, the covariance structure is block diagonal. Our factor model approach can be used to estimate the covariance matrix within a block, and our results continue to apply.

4.2. Impact on portfolio risk assessment

Risk management requires assessing the risk of a portfolio, which is different from optimal portfolio allocation. Throughout, risk is referred to as portfolio variance. As mentioned in Section 1, the smallest and largest eigenvalues of the covariance matrix are related to the minimum and maximum variances of the selected portfolio, respectively. Throughout this section, we fix a sequence of selected portfolios $\boldsymbol{\xi}_n \in \mathbb{R}^p$ with $\boldsymbol{\xi}_n'\mathbf{1} = 1$ and $\boldsymbol{\xi}_n = O(1)\mathbf{1}$. Here we impose the condition $\boldsymbol{\xi}_n = O(1)\mathbf{1}$ to avoid extreme short positions — that is, some large negative components in $\boldsymbol{\xi}_n$. Then, the variance of portfolio $\boldsymbol{\xi}_n$ is

$$\mathrm{var}(\boldsymbol{\xi}_n'\mathbf{y}) = \boldsymbol{\xi}_n'\mathrm{cov}(\mathbf{y})\boldsymbol{\xi}_n = \boldsymbol{\xi}_n'\boldsymbol{\Sigma}_n\boldsymbol{\xi}_n.$$

The estimated risk associated with portfolio $\boldsymbol{\xi}_n$ is $\boldsymbol{\xi}_n'\hat{\boldsymbol{\Sigma}}_n\boldsymbol{\xi}_n$. For practical use in portfolio risk assessment, we need to examine the behavior of portfolio variance based on $\hat{\boldsymbol{\Sigma}}_n$ estimated from historical data.

Theorem 7 (*Weak Convergence of Variance*). *Under conditions* (A) *and* (B), *we have*

$$\boldsymbol{\xi}_n'\hat{\boldsymbol{\Sigma}}_n\boldsymbol{\xi}_n - \boldsymbol{\xi}_n'\boldsymbol{\Sigma}_n\boldsymbol{\xi}_n = o_P\{(p^4K^2\log n/n)^{1/2}\}$$

and

$$\boldsymbol{\xi}_n'\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}\boldsymbol{\xi}_n - \boldsymbol{\xi}_n'\boldsymbol{\Sigma}_n\boldsymbol{\xi}_n = o_P\{(p^4K^2\log n/n)^{1/2}\}.$$

On the other hand, if the portfolios $\boldsymbol{\xi}_n$'s have no short positions, then we have

$$\boldsymbol{\xi}_n'\hat{\boldsymbol{\Sigma}}_n\boldsymbol{\xi}_n - \boldsymbol{\xi}_n'\boldsymbol{\Sigma}_n\boldsymbol{\xi}_n = o_P\{(p^2K^2\log n/n)^{1/2}\}$$

and

$$\boldsymbol{\xi}_n'\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}\boldsymbol{\xi}_n - \boldsymbol{\xi}_n'\boldsymbol{\Sigma}_n\boldsymbol{\xi}_n = o_P\{(p^2K^2\log n/n)^{1/2}\}.$$

From this theorem, we see that $\hat{\boldsymbol{\Sigma}}$ behaves roughly the same as the sample covariance matrix estimator $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ in portfolio risk assessment. This is essential for both covariance matrix estimators, since portfolio risk assessment does not involve inverse of the covariance matrix, but the covariance matrix itself. The above theorem is implied by consistency results of $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ under the Frobenius norm in Theorem 1.

5. A simulation study

In this section we use a simulation study to illustrate and augment our theoretical results and to verify finite-sample performance of the estimator $\hat{\boldsymbol{\Sigma}}$ as well as $\hat{\boldsymbol{\Sigma}}^{-1}$. To this end, we fix sample size $n = 756$, which is the sample size of three-year daily financial data, and we let dimensionality p grow from low to high and ultimately exceed sample size. As mentioned before, our primary concern is a theoretical understanding of factor models with a diverging number of variables and factors for the purpose of covariance matrix estimation, but not comparison with other popular estimators. So we compare performance of the estimator $\hat{\boldsymbol{\Sigma}}$ only to that of sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$. To contrast with $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$, we examine the covariance matrix estimation errors of $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ under the Frobenius norm, the norm $\|\cdot\|_{\boldsymbol{\Sigma}}$ introduced in Section 3, and the entropy loss. Meanwhile, we also compare estimation errors of $\hat{\boldsymbol{\Sigma}}^{-1}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}^{-1}$ under the Frobenius norm. Furthermore, we evaluate estimated variances of optimal portfolios with expected rate of return $\gamma_n = 10\%$ based on $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ by comparing their mean-squared errors (MSEs). For the estimated global minimum variances, we also compare their MSEs. Moveover, we examine MSEs of estimated variances of the equally weighted portfolio $\boldsymbol{\xi}_p = (1/p, \ldots, 1/p)$, based on $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$, respectively.

For simplicity, we fix $K = 3$ in our simulation and consider the three-factor model

$$Y_{pi} = b_{pi1}f_1 + b_{pi2}f_2 + b_{pi3}f_3 + \varepsilon_i, \quad i = 1, \ldots, p. \tag{12}$$

Here, we use the first subscript p to stress that the factor loadings in the three-factor model varies across dimensionality p. The Fama-French three-factor model (Fama and French, 1993) is a practical example of model (12). To make our simulation more realistic, we take the parameters from a fit of the Fama-French three-factor model.

In the Fama-French three-factor model, Y_i is the excess return of the i-th stock or portfolio, $i = 1, \ldots, p$. The first factor f_1 is the excess return of the proxy of the market portfolio, which is the value-weighted return on all NYSE, AMEX and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate (from Ibbotson Associates). The other two factors are constructed using six value-weighted portfolios formed on size and book-to-market ratio. Specifically, the second factor f_2, SMB (Small Minus Big),

$$\begin{aligned}\mathrm{SMB} = {}& 1/3\,(\text{Small Value} + \text{Small Neutral} + \text{Small Growth}) \\ & - 1/3\,(\text{Big Value} + \text{Big Neutral} + \text{Big Growth})\end{aligned}$$

is the average return on the three small portfolios minus the average return on the three big portfolios, and the third factor f_3, HML (High Minus Low),

$$\begin{aligned}\mathrm{HML} = {}& 1/2\,(\text{Small Value} + \text{Big Value}) \\ & - 1/2\,(\text{Small Growth} + \text{Big Growth})\end{aligned}$$

is the average return on the two value portfolios minus the average return on the two growth portfolios. See the website http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html for more details about their three factors and the data sets of the three factors, risk free interest rates, and returns of many constructed portfolios.

We first fit three-factor model (12) with $n = 756$ and $p = 30$ using the three-year daily data of 30 Industry Portfolios from May 1, 2002 to Aug 29, 2005, which are available at the above website. Then, as in (4), we get 30 estimated factor loading vectors $\widehat{\mathbf{b}}_1 = (b_{11}, b_{12}, b_{13}), \ldots, \widehat{\mathbf{b}}_{30} = (b_{30,1}, b_{30,2}, b_{30,3})$ and 30 estimated standard deviations $\widehat{\sigma}_1, \ldots, \widehat{\sigma}_{30}$ of the errors, where $\widehat{\mathbf{b}}_i$ and $\widehat{\sigma}_i$ correspond to the i-th portfolio, $i = 1, \ldots, 30$. The sample average of $\widehat{\sigma}_1, \ldots, \widehat{\sigma}_{30}$ is 0.66081 with a sample standard deviation 0.3275. We report in Table 1 the sample means and sample covariance matrices of $\mathbf{f}$ and $\widehat{\mathbf{b}}$ denoted by $\mu_{\mathbf{f}}$, $\mu_{\mathbf{b}}$ and $\mathrm{cov}_{\mathbf{f}}$, $\mathrm{cov}_{\mathbf{b}}$, respectively.

For each simulation, we carry out the following steps:

- We first generate a random sample of $\mathbf{f} = (f_1, f_2, f_3)'$ with size $n = 756$ from the trivariate normal distribution $\mathcal{N}(\mu_{\mathbf{f}}, \mathrm{cov}_{\mathbf{f}})$.
- Then, for each dimensionality p increasing from 16 to 1000 with increment 20, we do the following.
- Generate p factor loading vectors $\mathbf{b}_1, \ldots, \mathbf{b}_p$ as a random sample of size p from the trivariate normal distribution $\mathcal{N}(\mu_{\mathbf{b}}, \mathrm{cov}_{\mathbf{b}})$.

Table 1
Sample means and sample covariance matrices of $\mathbf{f}$ and $\widehat{\mathbf{b}}$

$\mu_{\mathbf{f}}$	$\mathrm{cov}_{\mathbf{f}}$		
0.023558	1.2507	−0.034999	−0.20419
0.012989	−0.034999	0.31564	−0.0022526
0.020714	−0.20419	−0.0022526	0.19303
$\mu_{\mathbf{b}}$	$\mathrm{cov}_{\mathbf{b}}$		
0.78282	0.029145	0.023873	0.010184
0.51803	0.023873	0.053951	−0.006967
0.41003	0.010184	−0.006967	0.086856

- Generate p standard deviations $\sigma_1, \ldots, \sigma_p$ of the errors as a random sample of size p from a gamma distribution $G(\alpha, \beta)$ conditional on being bounded below by a threshold value. The threshold for the standard deviations of errors is required in accordance with condition (C) in Section 3.1, and it is set to 0.1950 in our simulation because we find $\min_{1\le i\le 30} \widehat{\sigma_i} = 0.1950$. Note that $G(\alpha, \beta)$ has mean $\alpha\beta$ and standard deviation $\alpha^{1/2}\beta$, and its conditional mean and conditional second moment on falling above 0.1950 can be approximated respectively by
$$\left(\alpha\beta - \frac{0.1950}{2}q\right) \Big/ (1-q) \quad \text{and}$$
$$\left(\alpha\beta^2 + \alpha^2\beta^2 - \frac{0.1950^2}{2}q\right) \Big/ (1-q),$$
where q is the probability of falling below 0.1950 under $G(\alpha, \beta)$. By matching the mean 0.66081 and standard deviation 0.3275 for $G(\alpha_0, \beta_0)$, we obtain $\alpha_0 = 4.0713$ and $\beta_0 = 0.1623$. Therefore, following the above approximations, by recursively matching the conditional mean 0.66081 and conditional second moment $0.3275^2 + 0.66081^2 = 0.54393$ for $G(\alpha, \beta)$, we finally get $\alpha = 3.3586$ and $\beta = 0.1876$.
- After getting p standard deviations $\sigma_1, \ldots, \sigma_p$ of the errors, we generate a random sample of $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)'$ with size $n = 756$ from the p-variate normal distribution $\mathcal{N}\left(0, \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_p^2\right)\right)$.
- Then from model (12), we get a random sample of $\mathbf{y} = (Y_1, \ldots, Y_p)'$ with size $n = 756$.
- Finally, we compute estimated covariance matrices $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$, as well as $\hat{\boldsymbol{\Sigma}}^{-1}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}^{-1}$, and record the errors in the aforementioned measures. Meanwhile, we calculate MSEs of estimated variances of the optimal portfolios with $\gamma_n = 10\%$ as well as MSEs of estimated global minimum variances based on $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$, respectively. Also, we record MSEs of estimated variances of the equally weighted portfolio based on $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$, respectively.

We repeat the above simulation 500 times and report the mean-square errors as well as the standard deviations of those errors.

In Figs. 1–4, solid curves and dashed curves correspond to $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$, respectively. Fig. 1 presents the averages and the standard deviations of their estimation errors under the Frobenius norm, norm $\|\cdot\|_{\boldsymbol{\Sigma}}$, and entropy loss against dimensionality p, respectively. Fig. 2 depicts the averages and the standard deviations of estimation errors of $\hat{\boldsymbol{\Sigma}}^{-1}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}^{-1}$ under the Frobenius norm against p. We report in Fig. 3 MSEs of estimated variances of the optimal portfolios with $\gamma_n = 10\%$ as well as MSEs of estimated global minimum variances using $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ against p. Fig. 4 presents MSEs of estimated variances of the equally weighted portfolio based on $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ against p.

Recall that both the sample size n and the number of factors K are kept fixed across p in our simulation. From Figs. 1–4, we observe the following:

- By comparing corresponding averages and standard deviations of the errors shown in Figs. 1 and 2, we see that the Monte-Carlo errors are negligible.
- Fig. 1(a) shows that under the Frobenius norm, $\hat{\boldsymbol{\Sigma}}$ performs roughly the same as (slightly better than) $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$, which is consistent with the results in Theorem 1. Nevertheless, this is a surprise and is against the conventional wisdom.
- Fig. 1(c) reveals that under norm $\|\cdot\|_{\boldsymbol{\Sigma}}$, $\hat{\boldsymbol{\Sigma}}$ performs much better than $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$, which is consistent with the results in Theorem 2. In particular, we see that the estimation errors of $\hat{\boldsymbol{\Sigma}}$ under norm $\|\cdot\|_{\boldsymbol{\Sigma}}$ are roughly at the same level across p. Recall that sample size n is fixed as 756 here. Thus, this is in line with the root-n-consistency of $\hat{\boldsymbol{\Sigma}}$ under norm $\|\cdot\|_{\boldsymbol{\Sigma}}$ when $p = O(n)$ shown in Theorem 2. Also, the apparent growth pattern of estimation errors in $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ with p is in accordance with its $(n/p)^{1/2}$-consistency under norm $\|\cdot\|_{\boldsymbol{\Sigma}}$ shown in Theorem 2.
- Fig. 1(e) shows that under entropy loss, $\hat{\boldsymbol{\Sigma}}$ significantly outperforms $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$, which strongly supports the factor-model based estimator $\hat{\boldsymbol{\Sigma}}$ over the sample one $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$. We only report the results for p truncated at 400. This is because for larger p, sample covariance matrices $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ are nearly singular with a big chance in the simulation, which results in extremely large entropy losses.
- From Fig. 2(a), we see that under the Frobenius norm, the estimator $\hat{\boldsymbol{\Sigma}}^{-1}$ significantly outperforms $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}^{-1}$, which is in line with the results in Theorem 3.
- Fig. 3(a) and (b) demonstrate convincingly that $\hat{\boldsymbol{\Sigma}}$ outperforms $\hat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ in optimal portfolio allocation. These results are in accordance with Theorems 5 and 6. One may notice that in Fig. 3(a), the MSEs are relatively large in magnitude for small p and then tend to stabilize when p grows large. This is because in our settings for the simulation, for small p the term $\varphi_n\phi_n - \psi_n^2$ is relatively small compared to $\varphi_n\gamma_n^2 - 2\psi_n\gamma_n + \phi_n$, which results in large variance of the optimal portfolio. The behavior of the MSEs for large p is essentially due to self-averaging in the dimensionality. Fig. 3(b) can be interpreted in the same way.
- Fig. 4 reveals that the factor-model based approach and the sample approach have almost the same performance in portfolio risk assessment, which is consistent with Theorem 7. The high-dimensionality behavior is essentially due to self-averaging as in Fig. 3(a).

6. Concluding remarks

This paper investigates the impact of dimensionality on the estimation of covariance matrices. Two estimators are singled out for studies and comparisons: the sample covariance matrix and the factor-model based estimate. The inverse of the covariance matrix takes advantage of the factor structure and hence can be better estimated in the factor approach. As a result, when the parameters involve the inverse of the population covariance, substantial gain can be made. On the other hand, the covariance matrix itself does not take much advantage of the factor structure, and hence its estimate cannot be improved much in the factor approach. This is somewhat surprising and goes against the conventional wisdom.

Optimal portfolio allocation and minimum variance portfolio involve the inverse of the covariance matrix. Hence, it is advantageous to employ the factor structure in optimal portfolio allocation. On the other hand, portfolio risk assessment intrinsically depends only on the covariance structure and hence there is not much advantage to appeal to the factor model in portfolio risk assessment.

Our conclusion is also verified by an extensive simulation study, in which the parameters are taken in a neighborhood that is close to the reality. The choice of parameters relies on a fit to the famous

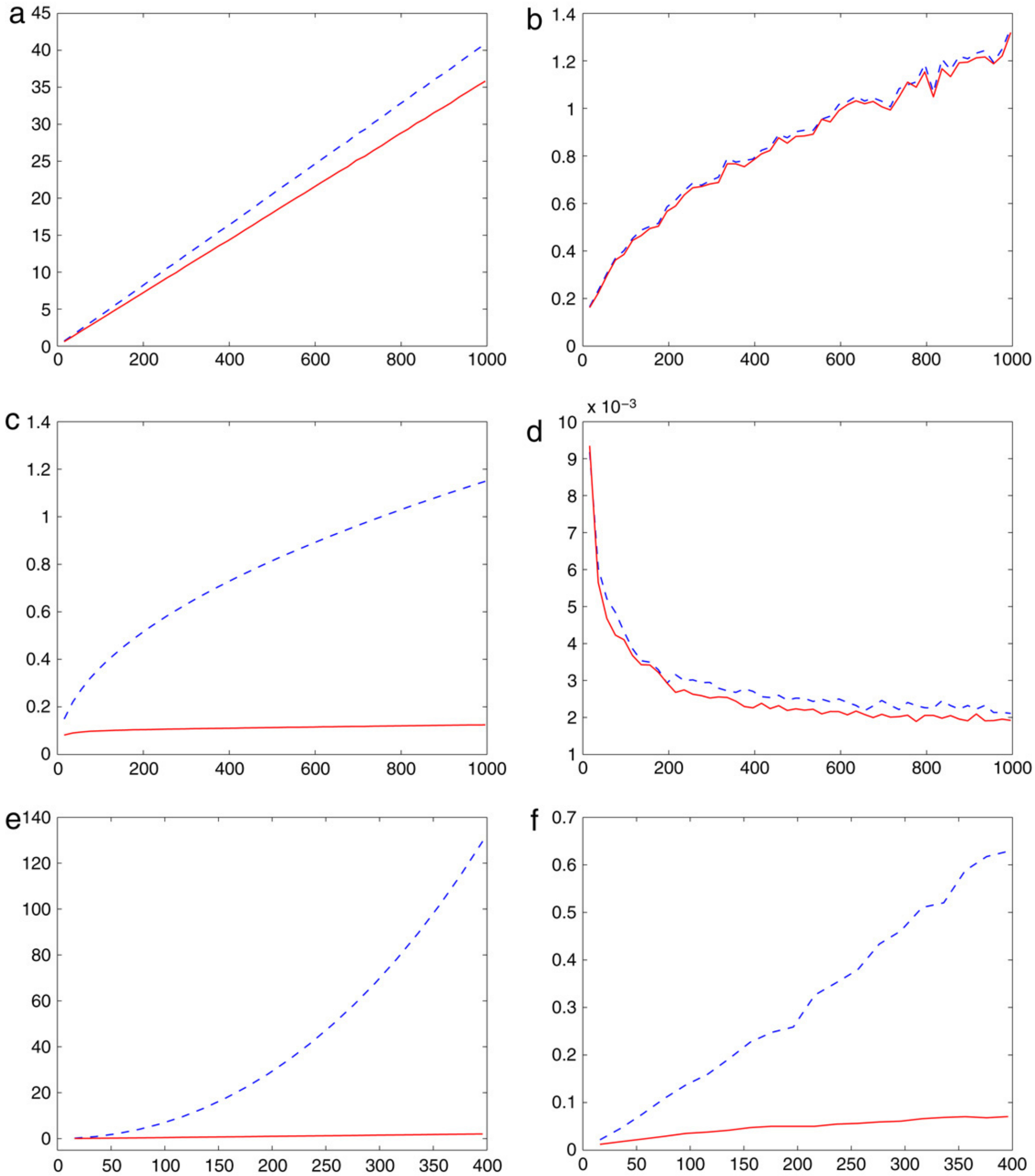

Fig. 1. (a), (c) and (e): The averages of errors over 500 simulations for $\hat{\mathbf{\Sigma}}$ (solid curve) and $\hat{\mathbf{\Sigma}}_{\text{sam}}$ (dashed curve) against p under Frobenius norm, norm $\|\cdot\|_{\mathbf{\Sigma}}$ and entropy losses, respectively. (b), (d) and (f): Corresponding standard deviations of errors over 500 simulations for $\hat{\mathbf{\Sigma}}$ (solid curve) and $\hat{\mathbf{\Sigma}}_{\text{sam}}$ (dashed curve).

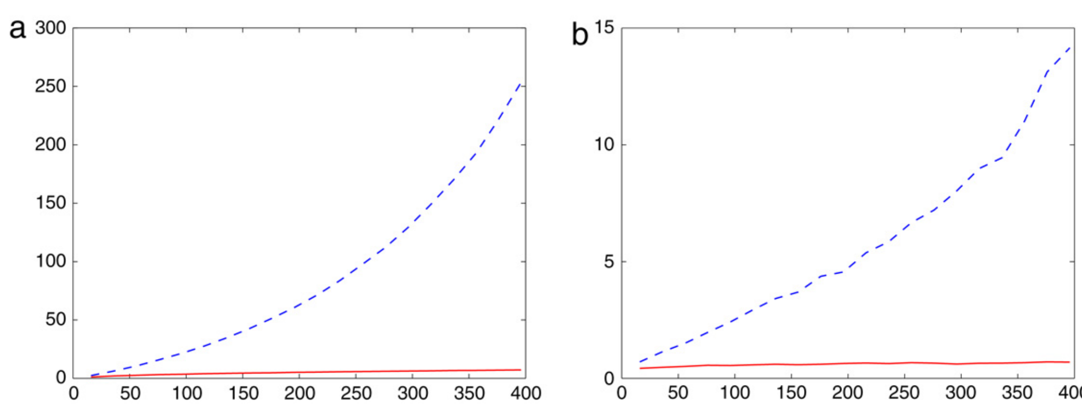

Fig. 2. (a) The averages of errors under Frobenius norm over 500 simulations for $\hat{\mathbf{\Sigma}}^{-1}$ (solid curve) and $\hat{\mathbf{\Sigma}}_{\text{sam}}^{-1}$ (dashed curve) against p. (b) Corresponding standard deviations of errors under Frobenius norm.

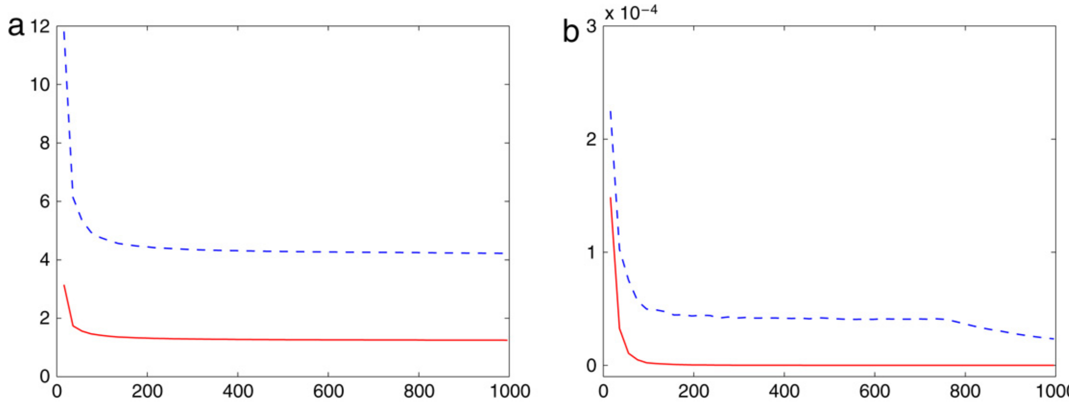

Fig. 3. (a) The MSEs of estimated variances of the optimal portfolios with $\gamma_n = 10\%$ over 500 simulations based on $\hat{\mathbf{\Sigma}}$ (solid curve) and $\hat{\mathbf{\Sigma}}_{\text{sam}}$ (dashed curve) against p. (b) The MSEs of estimated global minimum variances over 500 simulations based on $\hat{\mathbf{\Sigma}}$ (solid curve) and $\hat{\mathbf{\Sigma}}_{\text{sam}}$ (dashed curve) against p.

Fig. 4. The MSEs of estimated variances of the equally weighted portfolio over 500 simulations based on $\hat{\mathbf{\Sigma}}$ (solid curve) and $\hat{\mathbf{\Sigma}}_{\text{sam}}$ (dashed curve) against p.

Fama-French three-factor model to the portfolios traded in the market.

Our studies also reveal that the impact of dimensionality on the estimation of covariance matrices is severe. This should be taken into consideration in practical implementations.

Appendix. Proofs

Lemmas 1–6 mentioned throughout are referred to as those in technical report Fan et al. (2006).

Proof of Theorem 1. (1) First, we prove $(pK)^{-1} n^{1/2}$-consistency of $\hat{\mathbf{\Sigma}}$ under the Frobenius norm. To facilitate the presentation, we introduce here some notation used throughout the rest of the paper. Let $\mathbf{C}_n \widehat{=} \mathbf{E}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}$,

$$\mathbf{D}_n \widehat{=} \left\{(n-1)^{-1}\mathbf{X}\mathbf{X}' - [n(n-1)]^{-1}\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{X}'\right\} - \text{cov}(\mathbf{f})$$

and

$$\mathbf{F}_n \widehat{=} I_p \circ n^{-1}\mathbf{E}\left(I_n - \mathbf{H}\right)\mathbf{E}' - \mathbf{\Sigma}_0,$$

where $\mathbf{H} \widehat{=} \mathbf{X}'\left(\mathbf{X}\mathbf{X}'\right)^{-1}\mathbf{X}$ is the $n \times n$ hat matrix and $\mathbf{A}_1 \circ \mathbf{A}_2$ stands for the Hadamard product, i.e. the entrywise product, for any $q \times r$ matrices $\mathbf{A}_1$ and $\mathbf{A}_2$. Then we have $\widehat{\mathbf{B}} = \mathbf{Y}\mathbf{X}'\left(\mathbf{X}\mathbf{X}'\right)^{-1} = \mathbf{B} + \mathbf{C}_n$, $\widehat{\text{cov}}(\mathbf{f}) = (n-1)^{-1}\mathbf{X}\mathbf{X}' - \{n(n-1)\}^{-1}\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{X}' = \text{cov}(\mathbf{f}) + \mathbf{D}_n$, $\hat{\mathbf{\Sigma}}_0 = \text{diag}\left(n^{-1}\widehat{\mathbf{E}}\widehat{\mathbf{E}}'\right) = \mathbf{\Sigma}_0 + \mathbf{F}_n$ and

$$\hat{\mathbf{\Sigma}} = \mathbf{\Sigma} + \mathbf{B}\mathbf{D}_n\mathbf{B}' + \left[\mathbf{B}\widehat{\text{cov}}(\mathbf{f})\mathbf{C}_n' + \mathbf{C}_n\widehat{\text{cov}}(\mathbf{f})\mathbf{B}'\right] + \mathbf{C}_n\widehat{\text{cov}}(\mathbf{f})\mathbf{C}_n' + \mathbf{F}_n. \quad (13)$$

This shows that $\hat{\mathbf{\Sigma}}$ is a four-term perturbation of the population covariance matrix, and this representation is our key technical tool. By the Cauchy-Schwarz inequality, it follows from (13) that

$$E\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|^2 \le 4\Big[E\ \text{tr}\left\{\left(\mathbf{B}\mathbf{D}_n\mathbf{B}'\right)^2\right\} + E\ \text{tr}\left\{\left[\mathbf{B}\widehat{\text{cov}}(\mathbf{f})\mathbf{C}_n' + \mathbf{C}_n\widehat{\text{cov}}(\mathbf{f})\mathbf{B}'\right]^2\right\} + E\ \text{tr}\left\{\left[\mathbf{C}_n\widehat{\text{cov}}(\mathbf{f})\mathbf{C}_n'\right]^2\right\} + E\ \text{tr}\left(\mathbf{F}_n^2\right)\Big].$$

We will examine each of the above four terms on the right hand side separately. For brevity of notation, we suppress the first subscript n in some situations where the dependence on n is self-evident.

Before going further, let us bound $\|\mathbf{B}_n\|$. From assumption (B), we know that $\text{cov}(\mathbf{f}) \ge \sigma_1 I_K$, where for any symmetric positive semidefinite matrices $\mathbf{A}_1$ and $\mathbf{A}_2$, $\mathbf{A}_1 \ge \mathbf{A}_2$ means $\mathbf{A}_1 - \mathbf{A}_2$ is positive semidefinite. Thus it follows easily from (3) that

$$\sigma_1\mathbf{B}_n\mathbf{B}_n' = \mathbf{B}_n\left(\sigma_1 I_K\right)\mathbf{B}_n' \le \mathbf{B}_n\text{cov}(\mathbf{f})\mathbf{B}_n' \le \mathbf{\Sigma}_n,$$

which along with $b_n = O(p)$ in assumption (B) shows that $\|\mathbf{B}_n\|^2 = \text{tr}\left(\mathbf{B}_n\mathbf{B}_n'\right) \le \text{tr}\left(\mathbf{\Sigma}_n\right)/\sigma_1 \le \frac{b_n}{\sigma_1} = O(p)$, i.e.

$$\|\mathbf{B}_n\| = O(p^{1/2}). \quad (14)$$

Clearly, $\|\mathbf{B}_n'\mathbf{B}_n\| = \|\mathbf{B}_n\mathbf{B}_n'\|$, and by Lemma 1 and (14) we have

$$\|\mathbf{B}_n'\mathbf{B}_n\| = \|\mathbf{B}_n\mathbf{B}_n'\| \le \|\mathbf{B}_n\|\|\mathbf{B}_n'\| = \|\mathbf{B}_n\|^2 = O(p). \quad (15)$$

This fact is a key observation that will be used very often and, as shown above, it is entailed only by assumptions (A) and (B), which are valid throughout the paper.

Now we consider the first term, say $E\ \text{tr}\{\left(\mathbf{B}\mathbf{D}_n\mathbf{B}'\right)^2\}$. From $c_n = O(1)$ in assumption (B), we see that the fourth moments of $\mathbf{f}$ are bounded across n, thus a routine calculation reveals that

$$E\left(\|\mathbf{D}_n\|^2\right) = O(n^{-1}K^2), \quad (16)$$

which is an important fact that will be used very often and also helps study the inverse $\widehat{\text{cov}}(\mathbf{f})^{-1}$ by keeping in mind that $K \to \infty$. By Lemma 1, (15) and (16), we have

$$E\ \text{tr}\left[\left(\mathbf{B}\mathbf{D}_n\mathbf{B}'\right)^2\right] \le \left\|\mathbf{B}'\mathbf{B}\right\|^2 E\left(\|\mathbf{D}_n\|^2\right) = O(n^{-1}(pK)^2). \quad (17)$$