High Dimensional Covariance Matrix Estimation Using a Factor Model *

By Jianqing Fan, Yingying Fan and Jinchi Lv

Princeton University

August 12, 2006

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 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 portfolio allocation and risk management are studied. The asymptotic results are supported by a thorough simulation study.

Short Title: Large Covariance Matrix Estimation.

AMS 2000 subject classifications. Primary 62F12, 62H12; secondary 62J05, 62E20.

Key words and phrases. Factor model, diverging dimensionality, covariance matrix estimation, consistency, asymptotic normality, optimal portfolio, risk management.

---

*Financial support from the NSF under grant DMS-0532370 is gratefully acknowledged.

1. Introduction.

1.1. *Background.* 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 risk management and 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 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 portfolio allocation and risk management, 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 thousands 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

$$Y_i = b_{i1} f_1 + \cdots + b_{iK} f_K + \varepsilon_i, \quad i = 1, \cdots, p, \tag{1.1}$$

where $f_1, \cdots, f_K$ are the excessive returns of K factors, b_{ij}, $i = 1, \cdots, p$, $j = 1, \cdots, K$, are unknown factor loadings, and $\varepsilon_1, \cdots, \varepsilon_p$ are p idiosyncratic errors uncorrelated given $f_1, \cdots, f_K$. In economics and finance literature, factors are implicitly assumed to be observable and there is a large literature contributed to construction of factors (e.g. Fama and French, 1992, 1993). The factor models have been widely applied 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), and 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 (1.1), if a few factors can completely capture the cross-sectional risks, the number of 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 portfolio allocation and risk management 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 portfolio allocation and risk management. 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. Our results can also be regarded as an important step to understand the performance of factor models with unobservable factors.

The factor model has been extensively studied in the literature [see, e.g. Scott (1966) and (1969), Browne (1987), Browne and Shapiro (1987), and Yuan and Bentler (1997)], but traditional work assumes the sample size n tends to infinity while the dimensionality p and the number of factors K are fixed. There is a relatively small literature on studies of models with a diverging number of parameters. See, for example, Huber (1973), Yohai and Maronna (1979), Portnoy (1984, 1985), and Bai (2003). In particular, Fan and Peng (2004) establish some asymptotic properties, as well as an oracle property, for nonconcave penalized likelihood estimators in the presence of a diverging number of parameters. One can further refer to seminal reviews by Donoho (2000) and Fan and Li (2006) for challenges of high dimensionality. But it still remains open to examine factor models with diverging dimensionality and growing number of factors for the purpose of covariance matrix estimation.

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 (1991, 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, Dempster (1972), Leonard and Hsu (1992), Chiu, Leonard and Tsui (1996), Diggle and Verbyla (1998), Pourahmadi (2000), Boik (2002), Smith and Kohn (2002), Wong, Carter and Kohn (2003), Wu and Pourahmadi (2003), Huang, Liu and Pourahmadi (2004), and Li and Gui (2005).

1.2. *Covariance matrix estimation.* We always denote by n the sample size, by p the dimensionality, and by $f_1, \cdots, f_K$ the K observable factors, where p grows with sample size n and K increases with dimensionality p. For ease of presentation, we rewrite factor model (1.1) in matrix form

$$\mathbf{y} = \mathbf{B}_n \mathbf{f} + \boldsymbol{\varepsilon}, \tag{1.2}$$

where $\mathbf{y} = (Y_1, \cdots, Y_p)'$, $\mathbf{B}_n = (\mathbf{b}_1, \cdots, \mathbf{b}_p)'$ with $\mathbf{b}_i = (b_{n,i1}, \cdots, b_{n,iK})'$, $i = 1, \cdots, p$, $\mathbf{f} = (f_1, \cdots, f_K)'$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_p)'$. 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 $(\mathbf{f}_1, \mathbf{y}_1), \cdots, (\mathbf{f}_n, \mathbf{y}_n)$ 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

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

Under model (1.2), we have

$$\boldsymbol{\Sigma}_n = \mathrm{cov}(\mathbf{B}_n \mathbf{f}) + \mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{B}_n \mathrm{cov}(\mathbf{f}) \mathbf{B}_n' + \boldsymbol{\Sigma}_{n,0}. \tag{1.3}$$

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

$$\widehat{\boldsymbol{\Sigma}}_n = \widehat{\mathbf{B}}_n \widehat{\mathrm{cov}}(\mathbf{f}) \widehat{\mathbf{B}}_n' + \widehat{\boldsymbol{\Sigma}}_{n,0}, \tag{1.4}$$

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

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

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

$$\widehat{\mathbf{\Sigma}}_{\mathrm{sam}} = (n-1)^{-1}\,\mathbf{Y}\mathbf{Y}' - \{n\,(n-1)\}^{-1}\,\mathbf{Y}\mathbf{1}\mathbf{1}'\mathbf{Y}'. \tag{1.5}$$

This paper mainly provides a theoretical understanding of the factor model with a diverging dimensionality and growing number of factors for the purpose of covariance matrix estimation; it does not aim to compare with other popular estimators. Throughout the paper, we always contrast the performance of the covariance matrix estimator $\widehat{\mathbf{\Sigma}}$ in (1.4) with that of the sample covariance matrix $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}$ in (1.5). With prior information of the true factor structure, the substitution estimator $\widehat{\mathbf{\Sigma}}$ is expected to perform better than $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}$. However, this has not formally been 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 regard to this. Our conclusion can be summarized as follows.

- $\widehat{\mathbf{\Sigma}}$ is always invertible, even if $p > n$, while $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}$ suffers from the problem of possibly being singular when dimensionality p is close to or larger than sample size n.

- The advantage of the factor model lies in the estimation of the inverse of the covariance matrix, not the estimation of the covariance matrix itself. When the parameters involve the inverse of the covariance matrix, the factor model shows substantial gains, whereas when the parameters involved the covariance matrix directly, the factor model does not have much advantage. The latter is a surprise to the conventional wisdom.

- Portfolio allocations involve the inverse of the covariance matrix and the factor-model based estimates gain substantially, whereas the risk management involves directly the covariance matrix and the gain is only marginally.
- $\widehat{\mathbf{\Sigma}}$ has asymptotic normality, while in general $\widehat{\mathbf{\Sigma}}_{\text{sam}}$ may not have asymptotic normality of the same kind.

These properties will be demonstrated in our paper as follows.

1.3. *Outline of the paper.* In section 2 we discuss some basic assumptions and present the sampling properties of the estimator $\widehat{\mathbf{\Sigma}}$, as well as those of $\widehat{\mathbf{\Sigma}}_{\text{sam}}$. We study the impacts of the covariance matrix estimation on portfolio allocation and risk management in Section 3. A simulation study is presented in Section 4, which augments our theoretical study. Section 5 contains some concluding remarks. The proofs of our results are given in Section 6. All the technical lemmas are relegated to the Appendix.

2. Sampling properties. In this section we study the sampling properties of $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Sigma}}_{\text{sam}}$ with growing dimensionality and number of factors. We discuss some basic assumptions in Section 2.1. The sampling properties are presented in Section 2.2.

In the presence of diverging dimensionality, we should carefully choose appropriate norms for high dimensional matrices in different situations. We first introduce some notation. We always denote by $\lambda_1(\mathbf{A}), \cdots, \lambda_q(\mathbf{A})$ 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

$$\|\mathbf{A}\| = \left\{\mathrm{tr}(\mathbf{A}\mathbf{A}')\right\}^{1/2}. \tag{2.1}$$

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 (1985)] 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 $\operatorname{cov}(\boldsymbol{\varepsilon}|f) = I_p$ in model (1.2) with a single factor f. Then we have a substitution covariance matrix estimator $\widehat{\boldsymbol{\Sigma}} = \mathbf{1}\widehat{\operatorname{var}}(f)\mathbf{1}' + I_p$ as in (1.4). It is a classical result that

$$E\left|\widehat{\operatorname{var}}(f) - \operatorname{var}(f)\right|^2 = O(n^{-1}).$$

Thus by (1.3), we have

$$\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} = \mathbf{1}\left[\widehat{\operatorname{var}}(f) - \operatorname{var}(f)\right]\mathbf{1}'$$

and the Frobenius norm $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\| = \left|\widehat{\operatorname{var}}(f) - \operatorname{var}(f)\right| p$ picks up and amplifies the estimation error from $\widehat{\operatorname{var}}(f)$. Consequently,

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

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

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

This shows that under Frobenius norm, $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Sigma}}_{\text{sam}}$ have the same convergence rate and perform roughly the same. Thus we should seek other norms that fully employ the factor structure. By assuming the eigenvalues of $\boldsymbol{\Sigma}$ are bounded away from 0 and $\operatorname{var}(f) > 0$, routine calculations show that

$$\left\|\boldsymbol{\Sigma}^{-1/2}\left(\widehat{\boldsymbol{\Sigma}}_{\text{sam}} - \boldsymbol{\Sigma}\right)\boldsymbol{\Sigma}^{-1/2}\right\| = O_P(n^{-1/2}p^{3/2}),$$

whereas $\|\boldsymbol{\Sigma}^{-1/2}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1/2}\| = O_P(n^{-1/2})$. Therefore, with prior information of the true factor structure, $\widehat{\boldsymbol{\Sigma}}$ performs much better than $\widehat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ from this point of view.

Motivated by the above example, we first fix a sequence of positive definite covariance matrices $\boldsymbol{\Sigma}_n$ of dimensionality p_n, $n = 1, 2, \cdots$, and define a new norm

$$\|\mathbf{A}\|_{\boldsymbol{\Sigma}_n} = p_n^{-1/2} \left\| \boldsymbol{\Sigma}_n^{-1/2} \mathbf{A} \boldsymbol{\Sigma}_n^{-1/2} \right\| \tag{2.2}$$

for any $p_n \times p_n$ matrix $\mathbf{A}$. In particular, we have $\|\boldsymbol{\Sigma}_n\|_{\boldsymbol{\Sigma}_n} = p^{-1/2}\|I_p\| = 1$. The inclusion of a normalization factor $p^{-1/2}$ above is not essential and we incorporate it to take into account the diverging dimensionality. As seen below, under this new norm $\|\cdot\|_{\boldsymbol{\Sigma}}$, the consistency rate in the factor approach is better than that in the sample approach. Equivalently, we are investigating convergence rates under the loss function

$$L(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = p^{1/2} \left\| \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} \right\|_{\boldsymbol{\Sigma}} = \left\{ \mathrm{tr}[\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - I]^2 \right\}^{1/2}. \tag{2.3}$$

The above definition of the norm $\|\cdot\|_{\boldsymbol{\Sigma}}$ seems a bit artificial and involves the inverse of the true covariance matrix, but it is very similar to the entropy loss function proposed by James and Stein (1961). See Section 4 for further details. Intrinsically, this norm takes into account and fully employs the factor structure. In fact, as shown in the above example, the advantage of the factor structure lies in better performance of the inverse $\widehat{\boldsymbol{\Sigma}}^{-1}$. We will see later in this section that $\widehat{\boldsymbol{\Sigma}}^{-1}$ is a much better estimator of $\boldsymbol{\Sigma}^{-1}$ than $\widehat{\boldsymbol{\Sigma}}_{\mathrm{sam}}^{-1}$, and this advantage is carried further in portfolio allocation.

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

The first and second parts are usual conditions, and it is realistic to put $K \leq p$. The assumption that $\mathbf{f}$ has a continuous distribution is made to ensure that the $K \times K$ matrix $\mathbf{XX}'$ is invertible with probability one when $n \geq K$. Clearly, the covariance matrix estimator $\widehat{\mathbf{\Sigma}}$ is positive definite with probability one whenever $n \geq K$. By the assumption that the K factors capture the cross-sectional risks, the idiosyncratic noises are uncorrelated, so $\mathbf{\Sigma}_{n,0}$ is diagonal.

(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 $\lambda_K(\mathrm{cov}(\mathbf{f})) \geq \sigma_1$ for all n.

This is a technical assumption. In view of $E\|\mathbf{y}\|^2 = \sum_{i=1}^p Ey_i^2$, $b_n = O(p)$ is a reasonable condition. The assumption $c_n = O(1)$ shows that the fourth moments of $\mathbf{f}$ are bounded across n, which facilitates the study of the sample covariance matrix of $\mathbf{f}$. The uniform lower bound imposed on the eigenvalues of $\mathrm{cov}(\mathbf{f})$ helps the study of the inverse of the sample covariance matrix of $\mathbf{f}$ since $K \to \infty$, and it along with $b_n = O(p)$ entails that $\|\mathbf{B}_n\| = O(p^{1/2})$. It is evident from our theoretical analysis that $\lambda_K(\mathrm{cov}(\mathbf{f}))$ can be allowed to tend to zero at some rate, which results in slower convergence rates of the estimators. But we do not pursue in this direction here.

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

This is a reasonable assumption and ensures that all the eigenvalues of $\mathbf{\Sigma}_n$'s are bounded away from 0 in view of (1.3). In particular, we have $\|\mathbf{\Sigma}_n^{-1}\| = O(p^{1/2})$. Our theoretical analysis applies to the case where $\lambda_p(\mathbf{\Sigma}_{n,0})$ tends to zero at some rate, but we do not pursue along this direction for simplicity.

(D) The K factors $f_1, \cdots, f_K$ 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 $\mathbf{A}$.

This assumption is used *only* to establish asymptotic normality of the estimator $\widehat{\mathbf{\Sigma}}$, which facilitates statistical inferences. In view of $p^{-1}\mathbf{B}_n'\mathbf{B}_n = p^{-1}(\mathbf{b}_1\mathbf{b}_1' + \cdots + \mathbf{b}_p\mathbf{b}_p')$, this assumption is reasonable when K is fixed.

2.2. *Sampling properties.*

THEOREM 1 (Rates of convergence under Frobenius norm). *Under conditions (A) and (B), we have* $\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\| = O_P(n^{-1/2}pK)$ *and* $\left\|\widehat{\mathbf{\Sigma}}_{sam} - \mathbf{\Sigma}\right\| = O_P(n^{-1/2}pK)$. *In addition, we have*

$$\max_{1\le k\le p}\left|\lambda_k(\widehat{\mathbf{\Sigma}}_n) - \lambda_k(\mathbf{\Sigma}_n)\right| = o_P\{(p^2K^2\log n/n)^{1/2}\}$$

and

$$\max_{1\le k\le p}\left|\lambda_k(\widehat{\mathbf{\Sigma}}_{sam}) - \lambda_k(\mathbf{\Sigma}_n)\right| = o_P\{(p^2K^2\log n/n)^{1/2}\}.$$

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 $\widehat{\mathbf{\Sigma}}$ is optimal. To see it, let us extend the previous example by including K factors $f_1, \cdots, f_K$ and setting $\mathbf{B} = (\mathbf{1}, \cdots, \mathbf{1})_{p\times K}$. Further suppose we know ideally that $\mathrm{cov}(\mathbf{f}) = \mathrm{var}(f_1)I_K$. Then we have

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

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

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

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

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

Theorem 2 (Rates of convergence under norm $\|\cdot\|_{\mathbf{\Sigma}}$). *Suppose that* $K = O(n^{\alpha_1})$ *and* $p = O(n^{\alpha})$. *Under conditions (A)–(C), we have* $\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\mathbf{\Sigma}} = O_P(n^{-\beta/2})$ *with* $\beta = \min(1 - 2\alpha_1, 2 - \alpha - \alpha_1)$ *and* $\left\|\widehat{\mathbf{\Sigma}}_{sam} - \mathbf{\Sigma}\right\|_{\mathbf{\Sigma}} = O_P(n^{-\beta_1/2})$ *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 $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}$ has slower convergence. An interesting case is $K = O(1)$. In this case, under the norm $\|\cdot\|_{\mathbf{\Sigma}}$, $\widehat{\mathbf{\Sigma}}$ has convergence rate $n^{-\beta/2}$ with $\beta = \min(1, 2 - \alpha)$, whereas $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}$ has slower convergence rate $n^{-\beta_1/2}$ with $\beta_1 = 1 - \alpha$. In particular, when $\alpha \leq 1$, $\widehat{\mathbf{\Sigma}}$ is root-n-consistent under $\|\cdot\|_{\mathbf{\Sigma}}$. 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*

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

whereas

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

From this theorem, we see that when $K = o(p)$, $\widehat{\mathbf{\Sigma}}^{-1}$ performs much better than $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}^{-1}$. 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), $\widehat{\mathbf{\Sigma}}^{-1}$ has a consistency rate slightly slower than $\widehat{\mathbf{\Sigma}}$ under the Frobenius norm, since $\widehat{\mathbf{\Sigma}}^{-1}$ involves the inverse of the $K \times K$ sample covariance matrix of $\mathbf{f}$. The consistency result of $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}^{-1}$ is implied by that of $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}$, thanks to a simple inequality in matrix theory on inverses under perturbation. However, the consistency result of $\widehat{\mathbf{\Sigma}}^{-1}$ needs a very delicate analysis of inverse matrices. This theorem will be used in Section 3.1 to examine the variance 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 $\mathrm{vec}(\mathbf{A})$ 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 $\mathrm{vech}(\mathbf{A})$ the $d(d+1)/2 \times 1$ vector obtained from $\mathrm{vec}(\mathbf{A})$ 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

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

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

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

where $P_D = (D'D)^{-1} D'$. 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 $\widehat{\mathbf{\Sigma}}$ satisfies*

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

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

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

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

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

$$\operatorname{cov}(u_{ij}, u_{kl}) = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}.$$

The diverging dimensionality takes care of a trouble term in establishing asymptotic normality. However, in the finite dimensional setting, one can only show asymptotic normality when $\mathbf{f}$ has mean $\mathbf{0}$, where $\operatorname{cov}(\mathbf{f})$ can be estimated as $\widehat{\operatorname{cov}}(\mathbf{f}) = n^{-1}\mathbf{X}\mathbf{X}'$, and in general, $\widehat{\mathbf{\Sigma}}$ may have no asymptotic normality because the term $\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{X}'\left(\mathbf{X}\mathbf{X}'\right)^{-1}\mathbf{X}$ may not have a limiting behavior as $n \to \infty$ (at least it is not clear now). This is an interesting phenomenon in the presence of diverging dimensionality.

3. Impacts on portfolio allocation and risk management. In this section we examine the impacts of covariance matrix estimation on portfolio allocation and risk management, respectively.

3.1. *Impact on portfolio allocation.* For practical use in portfolio allocation, one would expect that the optimal portfolio constructed from the covariance matrix estimated from the history should not deviate too much from the true one. So we examine the behavior of the optimal portfolio constructed using $\widehat{\mathbf{\Sigma}}$ estimated from historical data.

Markowitz (1952) defines the mean-variance optimal portfolio as the solution $\boldsymbol{\xi}_n \in \mathbb{R}^p$ to the following minimization problem

$$\min_{\boldsymbol{\xi}} \ \boldsymbol{\xi}'\mathbf{\Sigma}_n\boldsymbol{\xi} \qquad (3.1)$$
$$\text{Subject to } \boldsymbol{\xi}'\mathbf{1} = 1 \text{ and } \boldsymbol{\xi}'\boldsymbol{\mu}_n = \gamma_n,$$

where $\mathbf{1}$ is a $p \times 1$ vector of ones, $\boldsymbol{\mu}_n = E(\mathbf{y})$, 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, Lo and MacKinlay (1997)] is

$$\boldsymbol{\xi}_n = \frac{\phi_n - \gamma_n \psi_n}{\varphi_n \phi_n - \psi_n^2} \boldsymbol{\Sigma}_n^{-1} \mathbf{1} + \frac{\gamma_n \varphi_n - \psi_n}{\varphi_n \phi_n - \psi_n^2} \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n \tag{3.2}$$

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

$$\boldsymbol{\xi}_n' \boldsymbol{\Sigma}_n \boldsymbol{\xi}_n = \frac{\varphi_n \gamma_n^2 - 2\psi_n \gamma_n + \phi_n}{\varphi_n \phi_n - \psi_n^2}. \tag{3.3}$$

Denote by $\boldsymbol{\xi}_{ng}$ the $\boldsymbol{\xi}_n$ in (3.2) with γ_n replaced by ψ_n/φ_n. The global minimum variance without constraint on the expected return is

$$\boldsymbol{\xi}_{ng}' \boldsymbol{\Sigma}_n \boldsymbol{\xi}_{ng} = \varphi_n^{-1}, \tag{3.4}$$

which is attained in (3.3) when $\gamma_n = \psi_n/\varphi_n$.

Based on the history, we can construct $\widehat{\boldsymbol{\Sigma}}_n$ as before. Also, 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 $\boldsymbol{\Sigma}_n$ and $\boldsymbol{\mu}_n$ replaced by $\widehat{\boldsymbol{\Sigma}}_n$ and $\widehat{\boldsymbol{\mu}}_n$, respectively.

It is interesting to study the deviation of the constructed optimal portfolio $\widehat{\boldsymbol{\xi}}_n$ and the globally optimal portfolio $\widehat{\boldsymbol{\xi}}_{ng}$ from the theoretical ones, say, $\boldsymbol{\xi}_n$ and $\boldsymbol{\xi}_{ng}$. But here we do not pursue in this direction because it is more valuable to study the risk associated with them. Therefore, we only examine the behavior of the minimum variance $\widehat{\boldsymbol{\xi}}_n' \widehat{\boldsymbol{\Sigma}}_n \widehat{\boldsymbol{\xi}}_n$ and global minimum variance $\widehat{\boldsymbol{\xi}}_{ng}' \widehat{\boldsymbol{\Sigma}}_n \widehat{\boldsymbol{\xi}}_{ng}$ in this section.

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*

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

whereas

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

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*

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

whereas

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

The assumptions on φ_n, ψ_n and ϕ_n in Theorems 5 and 6 are technical and reasonable. In view of (3.4), 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 (3.3). From the above two theorems, we see that when $K = o(p)$, $\widehat{\boldsymbol{\Sigma}}$ performs much better than $\widehat{\boldsymbol{\Sigma}}_{\text{sam}}$ from the point of view of 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 $\widehat{\boldsymbol{\Sigma}}$ or $\widehat{\boldsymbol{\Sigma}}_{\text{sam}}$ behave similarly to theoretical ones. So high dimensionality does impose a great challenge on 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.

3.2. *Impact on risk management.* Risk management is a different story from portfolio allocation. As mentioned in Section 1.1, 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 $\boldsymbol{\xi}_n'\widehat{\boldsymbol{\Sigma}}_n\boldsymbol{\xi}_n$. For practical use in risk management, we need to examine the behavior of portfolio variance based on $\widehat{\boldsymbol{\Sigma}}_n$ estimated from historical data.

THEOREM 7 (Weak convergence of variance). *Under conditions (A) and (B), we have*

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

and

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

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

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

and

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

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

4. 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 $\widehat{\boldsymbol{\Sigma}}$ as well as $\widehat{\boldsymbol{\Sigma}}^{-1}$. To this end, we fix sample size $n = 756$, which is the practical 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 $\widehat{\boldsymbol{\Sigma}}$ only to that of sample covariance matrix $\widehat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$. To contrast with $\widehat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$, we examine the covariance matrix estimation errors of $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ under the Frobenius norm, the norm $\|\cdot\|_{\boldsymbol{\Sigma}}$ introduced in Section 2, and the Stein (or entropy) loss function

$$L(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right) - \log\left|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right| - p,$$

which was proposed by James and Stein (1961). Meanwhile, we compare estimation errors of $\widehat{\boldsymbol{\Sigma}}^{-1}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{sam}}^{-1}$ under the Frobenius norm. Furthermore, we evaluate estimated variances of optimal portfolios with expected rate of return $\gamma_n = 10\%$ based on $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$ 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 $\boldsymbol{\xi}_p = (1/p, \cdots, 1/p)$, based on $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{sam}}$, respectively.

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

$$Y_{pi} = b_{pi1}f_1 + b_{pi2}f_2 + b_{pi3}f_3 + \varepsilon_i, \quad i = 1, \cdots, p. \tag{4.1}$$

Here, we use the first subscript p to stress that the three-factor model varies across dimensionality p. As before, we let $\mathbf{y} = (Y_1, \cdots, Y_p)'$ and $\mathbf{f} = (f_1, f_2, f_3)'$. The Fama-French three-factor model [Fama and French (1993)] is a practical example of model

(4.1). 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, $i = 1, \cdots, p$. 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. Specifically, the second factor f_2, `SMB` (Small Minus Big),

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

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),

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

is the average return on the two value portfolios minus the average return on the two growth portfolios. See their 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 (4.1) 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 (1.4), we get 30 estimated factor loading vectors $\widehat{\mathbf{b}}_1 = (b_{11}, b_{12}, b_{13}), \cdots, \widehat{\mathbf{b}}_{30} = (b_{30,1}, b_{30,2}, b_{30,3})$ and 30 estimated standard deviations $\widehat{\sigma}_1, \cdots, \widehat{\sigma}_{30}$ of the errors, where $\widehat{\mathbf{b}}_i$ and $\widehat{\sigma}_i$ correspond to the i-th portfolio, $i = 1, \cdots, 30$.

The sample average of $\widehat{\sigma}_1, \cdots, \widehat{\sigma}_{30}$ 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.

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

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 $\mathcal{N}\left(\mu_{\mathbf{f}}, \mathrm{cov}_{\mathbf{f}}\right)$.

- Then, for each dimensionality p increasing from 16 to 1000 with increment 20, we do the following.

- Generate p factor loading vectors $\mathbf{b}_1, \cdots, \mathbf{b}_p$ as a random sample of size p from the trivariate normal distribution $\mathcal{N}\left(\mu_{\mathbf{b}}, \mathrm{cov}_{\mathbf{b}}\right)$.

- Generate p standard deviations $\sigma_1, \cdots, \sigma_p$ 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 2.1, 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

$$\left(\alpha\beta - \frac{0.1950}{2}p\right) / (1-p) \quad \text{and} \quad \left(\alpha\beta^2 + \alpha^2\beta^2 - \frac{0.1950^2}{2}p\right) / (1-p)\,,$$

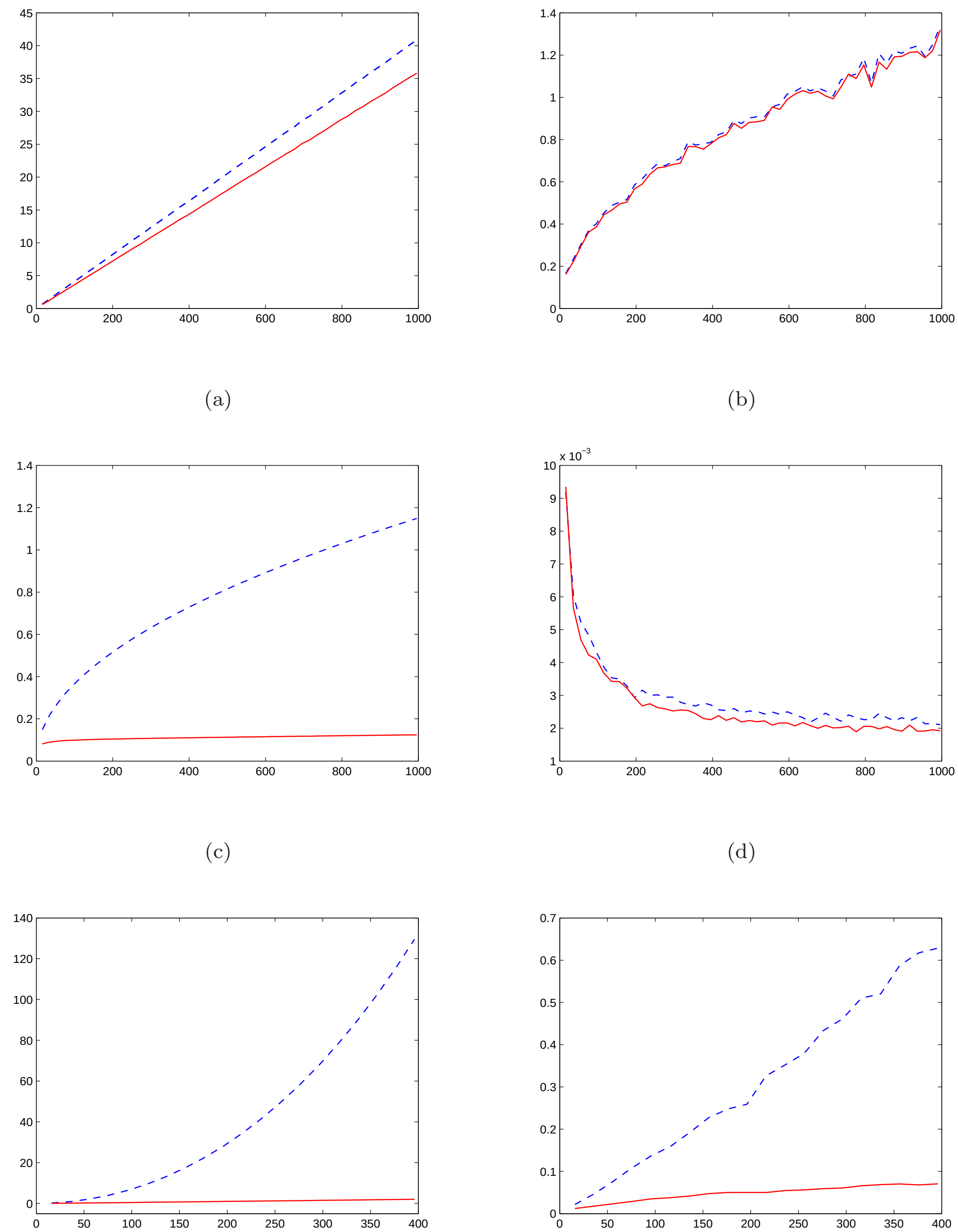
where p is the probability of falling below 0.1950 under $G(\alpha,\beta)$. By matching the mean 0.66081 and standard deviation 0.3275 for $G(\alpha_0,\beta_0)$, we obtain $\alpha_0 = 4.0713$ and $\beta_0 = 0.1623$. Therefore, following the above approximations, by recursively matching the conditional 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 $\sigma_1,\cdots,\sigma_p$ of the errors, we generate a random sample of $\boldsymbol{\varepsilon} = (\varepsilon_1,\cdots,\varepsilon_p)'$ with size $n = 756$ from the p-variate normal distribution $\mathcal{N}\left(0, \operatorname{diag}\left(\sigma_1^2,\cdots,\sigma_p^2\right)\right)$.

- Then from model (4.1), we get a random sample of $\mathbf{y} = (Y_1,\cdots,Y_p)'$ with size $n = 756$.

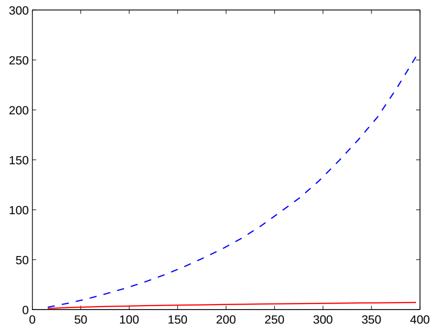
- Finally, we compute estimated covariance matrices $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Sigma}}_{\text{sam}}$, as well as $\widehat{\mathbf{\Sigma}}^{-1}$ and $\widehat{\mathbf{\Sigma}}_{\text{sam}}^{-1}$, and record the errors in the aforementioned measures. Meanwhile, we calculate MSEs of estimated variances of the optimal portfolios with $\gamma_n = 10\%$ as well as MSEs of estimated global minimum variances based on $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Sigma}}_{\text{sam}}$, respectively. Also, we record MSEs of estimated variances of the equally weighted portfolio based on $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Sigma}}_{\text{sam}}$, respectively.

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

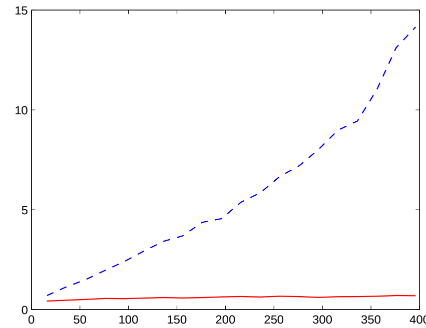


(a) (b) (c) (d) (e) (f)

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



(a)



(b)

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

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

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

- By comparing corresponding averages and standard deviations of the errors shown in Figures 1 and 2, we see that the Monte-Carlo errors are negligible.

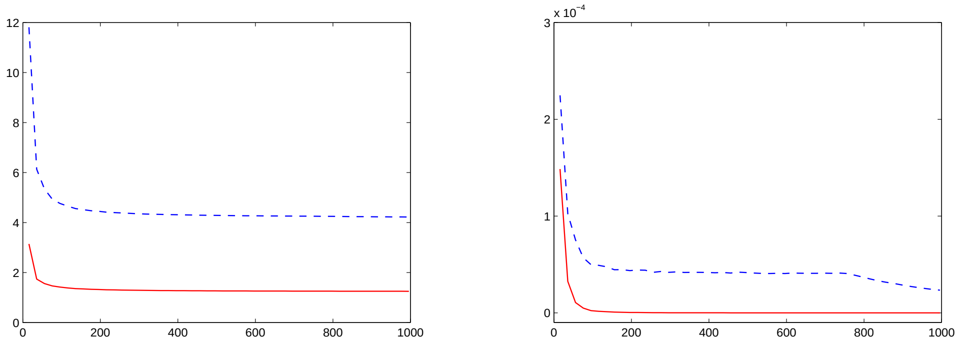
- Figure 1(a) shows that under the Frobenius norm, $\widehat{\mathbf{\Sigma}}$ performs roughly the same as (slightly better than) $\widehat{\mathbf{\Sigma}}_{\text{sam}}$, which is consistent with the results in Theorem 1. Nevertheless, this is a surprise and is against the conventional wisdom.

- Figure 1(c) reveals that under norm $\|\cdot\|_{\mathbf{\Sigma}}$, $\widehat{\mathbf{\Sigma}}$ performs much better than $\widehat{\mathbf{\Sigma}}_{\text{sam}}$, which is consistent with the results in Theorem 2. In particular, we see that the estimation errors of $\widehat{\mathbf{\Sigma}}$ under norm $\|\cdot\|_{\mathbf{\Sigma}}$ 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 $\widehat{\mathbf{\Sigma}}$ under norm $\|\cdot\|_{\mathbf{\Sigma}}$ when $p = O(n)$ shown in Theorem 2. Also, the apparent growth pattern of estimation errors in $\widehat{\mathbf{\Sigma}}_{\text{sam}}$ with p is in accordance with its $(n/p)^{1/2}$-consistency under norm $\|\cdot\|_{\mathbf{\Sigma}}$ shown in Theorem 2.

- Figure 1(e) shows that under entropy loss, $\widehat{\mathbf{\Sigma}}$ significantly outperforms $\widehat{\mathbf{\Sigma}}_{\text{sam}}$, which strongly supports the factor-model based estimator $\widehat{\mathbf{\Sigma}}$ over the sample one $\widehat{\mathbf{\Sigma}}_{\text{sam}}$. We only report the results for p truncated at 400. This is because for larger p, sample covariance matrices $\widehat{\mathbf{\Sigma}}_{\text{sam}}$ are nearly singular with a big chance in the simulation, which results in extremely large entropy losses.

- From Figure 2(a), we see that under the Frobenius norm, the estimator $\widehat{\mathbf{\Sigma}}^{-1}$ significantly outperforms $\widehat{\mathbf{\Sigma}}_{\text{sam}}^{-1}$, which is in line with the results in Theorem 3.

- Figures 3(a) and 3(b) demonstrate convincingly that $\widehat{\mathbf{\Sigma}}$ outperforms $\widehat{\mathbf{\Sigma}}_{\text{sam}}$ in portfolio allocation. These results are in accordance with Theorems 5 and 6. One may notice that in Figure 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. Figures 3(b) can be interpreted in the same way.



(a) (b)

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

- Figure 4 reveals that the factor-model based approach and the sample approach have almost the same performance in risk management, which is consistent with Theorem 7. The high-dimensionality behavior is essentially due to self-averaging as in Figure 3(a).

5. 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 can not be improved much in the factor approach. This is somewhat surprising and is against the conventional wisdom.

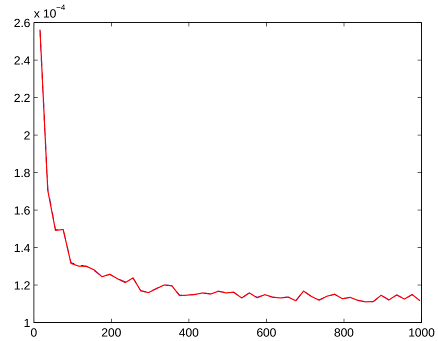


Figure 4: The MSEs of estimated variances of the equally weighted portfolio over 500 simulations based on $\widehat{\mathbf{\Sigma}}$ (solid curve) and $\widehat{\mathbf{\Sigma}}_{\text{sam}}$ (dashed curve) against p.

Optimal portfolio allocation and minimum variance portfolio involve the inverse of the covariance matrix. Hence, it is advantageous to employ the factor structure in portfolio allocation. On the other hand, intrinsically the risk management does not depend on the covariance structure and hence there is no advantage to appeal to the factor model in risk management.

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 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.

6. Proofs of theorems. In this section, we give rigorous proofs of Theorems 1–7.

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

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

and

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

where $\mathbf{H} \mathrel{\widehat{=}} \mathbf{X}'\,(\mathbf{X}\mathbf{X}')^{-1}\,\mathbf{X}$ 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 $\widehat{\mathbf{B}} = \mathbf{Y}\mathbf{X}'\,(\mathbf{X}\mathbf{X}')^{-1} = \mathbf{B} + \mathbf{C}_n$, $\widehat{\mathrm{cov}}(\mathbf{f}) = (n-1)^{-1}\,\mathbf{X}\mathbf{X}' - \{n\,(n-1)\}^{-1}\,\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{X}' = \mathrm{cov}(\mathbf{f}) + \mathbf{D}_n$, $\widehat{\mathbf{\Sigma}}_0 = \mathrm{diag}\left(n^{-1}\widehat{\mathbf{E}}\widehat{\mathbf{E}}'\right) = \mathbf{\Sigma}_0 + \mathbf{F}_n$ and

$$\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma} + \mathbf{B}\mathbf{D}_n\mathbf{B}' + \left[\mathbf{B}\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n' + \mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{B}'\right] + \mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n' + \mathbf{F}_n, \tag{6.1}$$

This shows that $\widehat{\mathbf{\Sigma}}$ 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 (6.1) that

$$\begin{aligned} E\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|^2 \le 4\, &\left[E\ \mathrm{tr}\left\{\left(\mathbf{B}\mathbf{D}_n\mathbf{B}'\right)^2\right\} + E\ \mathrm{tr}\left\{\left[\mathbf{B}\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n' + \mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{B}'\right]^2\right\}\right. \\ &\left. + E\ \mathrm{tr}\left\{\left[\mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n'\right]^2\right\} + E\ \mathrm{tr}\left(\mathbf{F}_n^2\right)\right]. \end{aligned}$$

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 $\mathrm{cov}(\mathbf{f}) \ge \sigma_1 I_K$, 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 (1.3) that

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

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

$$\|\mathbf{B}_n\| = O(p^{1/2}). \tag{6.2}$$

Clearly, $\|\mathbf{B}_n'\mathbf{B}_n\| = \|\mathbf{B}_n\mathbf{B}_n'\|$, and by (A.1) in Lemma 1 and (6.2) we have

$$\|\mathbf{B}_n'\mathbf{B}_n\| = \|\mathbf{B}_n\mathbf{B}_n'\| \leq \|\mathbf{B}_n\|\|\mathbf{B}_n'\| = \|\mathbf{B}_n\|^2 = O(p). \tag{6.3}$$

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 $E\ \operatorname{tr}\{(\mathbf{B}\mathbf{D}_n\mathbf{B}')^2\}$. 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

$$E\left(\|\mathbf{D}_n\|^2\right) = O(n^{-1}K^2), \tag{6.4}$$

which is an important fact that will be used very often and also helps study the inverse $\widehat{\operatorname{cov}}(\mathbf{f})^{-1}$ by keeping in mind that $K \to \infty$. By (A.2) in Lemma 1, (6.3), and (6.4), we have

$$E\ \operatorname{tr}\left[\left(\mathbf{B}\mathbf{D}_n\mathbf{B}'\right)^2\right] \leq \left\|\mathbf{B}'\mathbf{B}\right\|^2 E\left(\|\mathbf{D}_n\|^2\right) = O(n^{-1}(pK)^2). \tag{6.5}$$

The remaining three terms are taken care of by Lemmas 2 and 3. Therefore, in view of (6.3), combining (6.5) with (A.5)–(A.7) in Lemmas 2 and 3 gives

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

In particular, this implies that $\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\| = O_P(n^{-1/2}pK)$, which proves $(pK)^{-1}\,n^{1/2}$-consistency of the covariance matrix estimator $\widehat{\boldsymbol{\Sigma}}$ under Frobenius norm.

(2) Then, we show that $\widehat{\mathbf{\Sigma}}_{\text{sam}}$ is $(pK)^{-1}\, n^{1/2}$-consistent under the Frobenius norm. By (1.3) and (1.5), we have

$$
\begin{aligned}
(6.6) \qquad \widehat{\mathbf{\Sigma}}_{\text{sam}} = \mathbf{\Sigma} + \mathbf{B}\mathbf{D}_n\mathbf{B}' + \mathbf{G}_n + (n-1)^{-1}\left\{\mathbf{B}\mathbf{X}\mathbf{E}' + \mathbf{E}\mathbf{X}'\mathbf{B}\right\} \\
- \left[n\,(n-1)\right]^{-1}\left\{\mathbf{B}\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{E}' + \mathbf{E}\mathbf{1}\mathbf{1}'\mathbf{X}'\mathbf{B}'\right\},
\end{aligned}
$$

where $\mathbf{G}_n \mathrel{\widehat{=}} \left\{(n-1)^{-1}\,\mathbf{E}\mathbf{E}' - [n(n-1)]^{-1}\,\mathbf{E}\mathbf{1}\mathbf{1}'\mathbf{E}'\right\} - \mathbf{\Sigma}_0$. This shows that $\widehat{\mathbf{\Sigma}}_{\text{sam}}$ is also a four-term perturbation of the population covariance matrix. By the Cauchy-Schwarz inequality, it follows from (6.6) that

$$
\begin{aligned}
E\left\|\widehat{\mathbf{\Sigma}}_{\text{sam}} - \mathbf{\Sigma}\right\|^2 \le 4\,\Big[E\left\|\mathbf{B}\mathbf{D}_n\mathbf{B}'\right\|^2 + E\left\|\mathbf{G}_n\right\|^2 + 2\,(n-1)^{-2}\,E\left\|\mathbf{B}\mathbf{X}\mathbf{E}'\right\|^2 \\
+ 2\left[n\,(n-1)\right]^{-2}\,E\left\|\mathbf{B}\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{E}'\right\|^2\Big].
\end{aligned}
$$

As in part (1), we will examine each of the above four terms on the right hand side separately. The first term $E\left\|\mathbf{B}\mathbf{D}_n\mathbf{B}'\right\|^2$ has been bounded in (6.5). Using the same argument as in Lemma 6, we can show that $E\left\|\mathbf{G}_n\right\|^2 = O(n^{-1}p^2)$. In view of (6.3), it is shown that

$$E\left\|\mathbf{B}\mathbf{X}\mathbf{E}'\right\|^2 = O(np^2K)$$

in the proof of Lemma 2. Using the same argument as in Lemma 2 to bound $E\left\|\mathbf{B}\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}\mathbf{E}'\right\|^2$, we can easily get

$$E\left\|\mathbf{B}\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{E}'\right\|^2 = O(n^3p^2K),$$

which along with (6.5) and the above results yields

$$E\left\|\widehat{\mathbf{\Sigma}}_{\text{sam}} - \mathbf{\Sigma}\right\|^2 = O(n^{-1}(pK)^2).$$

This proves $(pK)^{-1}\, n^{1/2}$-consistency of $\widehat{\mathbf{\Sigma}}_{\text{sam}}$ under the Frobenius norm.

(3) Finally, we prove the uniform weak convergence of eigenvalues. It follows from

Corollary 6.3.8 of Horn and Johnson (1985) that

$$\max_{1\le k\le p}\left|\lambda_k(\widehat{\boldsymbol{\Sigma}}_n)-\lambda_k(\boldsymbol{\Sigma}_n)\right|\le\left\{\sum_{k=1}^{p}\left[\lambda_k(\widehat{\boldsymbol{\Sigma}}_n)-\lambda_k(\boldsymbol{\Sigma}_n)\right]^2\right\}^{1/2}\le\left\|\widehat{\boldsymbol{\Sigma}}_n-\boldsymbol{\Sigma}_n\right\|.$$

Therefore, the uniform weak convergence of the eigenvalues of the $\widehat{\boldsymbol{\Sigma}}_n$'s follows immediately from the $(pK)^{-1}\, n^{1/2}$-consistency of $\widehat{\boldsymbol{\Sigma}}$ under the Frobenius norm shown in part (1). Similarly, by the $(pK)^{-1}\, n^{1/2}$-consistency of $\widehat{\boldsymbol{\Sigma}}_{\text{sam}}$ under the Frobenius norm shown in part (2), the same conclusion holds for $\widehat{\boldsymbol{\Sigma}}_{\text{sam}}$. $\square$

Proof of Theorem 2. (1) First, we show that $\widehat{\boldsymbol{\Sigma}}$ is $n^{\beta/2}$-consistent under norm $\|\cdot\|_{\boldsymbol{\Sigma}}$. The main idea of the proof is similar to that of Theorem 1, but the proof is more tricky and involved here since the norm $\|\cdot\|_{\boldsymbol{\Sigma}}$ involves the inverse of the covariance matrix $\boldsymbol{\Sigma}$. By the Cauchy-Schwarz inequality, it follows from (6.1) that

$$\begin{aligned}E\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\right\|_{\boldsymbol{\Sigma}}^2\le 4\ \Big[&E\left\|\mathbf{B}\mathbf{D}_n\mathbf{B}'\right\|_{\boldsymbol{\Sigma}}^2+E\left\|\mathbf{B}\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n'+\mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{B}'\right\|_{\boldsymbol{\Sigma}}^2\\&+E\left\|\mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n'\right\|_{\boldsymbol{\Sigma}}^2\}+E\left\|\mathbf{F}_n\right\|_{\boldsymbol{\Sigma}}^2\Big].\end{aligned}$$

As in the proof of Theorem 1, we will study each of the above four terms on the right hand side separately.

Before going further, let us bound $\left\|\mathbf{B}'\boldsymbol{\Sigma}^{-1}\mathbf{B}\right\|$. From (1.3), we know that $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0+\mathbf{B}\mathrm{cov}(\mathbf{f})\mathbf{B}'$, which along with the Sherman-Morrison-Woodbury formula shows that

$$\boldsymbol{\Sigma}^{-1}=\boldsymbol{\Sigma}_0^{-1}-\boldsymbol{\Sigma}_0^{-1}\mathbf{B}\left[\mathrm{cov}(\mathbf{f})^{-1}+\mathbf{B}'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}\right]^{-1}\mathbf{B}'\boldsymbol{\Sigma}_0^{-1}. \tag{6.7}$$

Thus it follows that

$$\begin{aligned}\mathbf{B}'\boldsymbol{\Sigma}^{-1}\mathbf{B}&=\mathbf{B}'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}-\mathbf{B}'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}\left[\mathrm{cov}(\mathbf{f})^{-1}+\mathbf{B}'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}\right]^{-1}\mathbf{B}'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}\\&=\mathbf{B}'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}\left[\mathrm{cov}(\mathbf{f})^{-1}+\mathbf{B}'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}\right]^{-1}\mathrm{cov}(\mathbf{f})^{-1}\\&=\mathrm{cov}(\mathbf{f})^{-1}-\mathrm{cov}(\mathbf{f})^{-1}\left[\mathrm{cov}(\mathbf{f})^{-1}+\mathbf{B}'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}\right]^{-1}\mathrm{cov}(\mathbf{f})^{-1},\end{aligned}$$

which implies that

$$\left\|\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right\| \leq \left\|\mathrm{cov}(\mathbf{f})^{-1}\right\| + \left\|\mathrm{cov}(\mathbf{f})^{-1}\left[\mathrm{cov}(\mathbf{f})^{-1} + \mathbf{B}'\mathbf{\Sigma}_0^{-1}\mathbf{B}\right]^{-1}\mathrm{cov}(\mathbf{f})^{-1}\right\|.$$

Note that $\mathrm{cov}(\mathbf{f})^{-1}$ is symmetric positive definite and $\mathbf{B}'\mathbf{\Sigma}_0^{-1}\mathbf{B}$ is symmetric positive semidefinite. Thus, $\mathrm{cov}(\mathbf{f})^{-1}+\mathbf{B}'\mathbf{\Sigma}_0^{-1}\mathbf{B} \geq \mathrm{cov}(\mathbf{f})^{-1}$, which in turn implies that $\left[\mathrm{cov}(\mathbf{f})^{-1} + \mathbf{B}'\mathbf{\Sigma}_0^{-1}\mathbf{B}\right]^{-1} \leq \mathrm{cov}(\mathbf{f})$ and

$$\mathrm{cov}(\mathbf{f})^{-1}\left[\mathrm{cov}(\mathbf{f})^{-1} + \mathbf{B}'\mathbf{\Sigma}_0^{-1}\mathbf{B}\right]^{-1}\mathrm{cov}(\mathbf{f})^{-1} \leq \mathrm{cov}(\mathbf{f})^{-1}\mathrm{cov}(\mathbf{f})\mathrm{cov}(\mathbf{f})^{-1} = \mathrm{cov}(\mathbf{f})^{-1}.$$

In particular, this entails that

$$\left\|\mathrm{cov}(\mathbf{f})^{-1}\left[\mathrm{cov}(\mathbf{f})^{-1} + \mathbf{B}'\mathbf{\Sigma}_0^{-1}\mathbf{B}\right]^{-1}\mathrm{cov}(\mathbf{f})^{-1}\right\| \leq \left\|\mathrm{cov}(\mathbf{f})^{-1}\right\|,$$

so now the problem of bounding $\left\|\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right\|$ reduces to bounding $\left\|\mathrm{cov}(\mathbf{f})^{-1}\right\|$. By assumption (B), $\lambda_K(\mathrm{cov}(\mathbf{f})) \geq \sigma_1$ for some constant $\sigma_1 > 0$. Thus the largest eigenvalues of $\mathrm{cov}(\mathbf{f})^{-1}$ are bounded across n, which easily implies that $\left\|\mathrm{cov}(\mathbf{f})^{-1}\right\| = O(K^{1/2})$. This together with the above results shows that

$$\left\|\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right\| = O(K^{1/2}). \tag{6.8}$$

Now we are ready to examine the first term, say $E\left\|\mathbf{B}\mathbf{D}_n\mathbf{B}'\right\|_{\mathbf{\Sigma}}^2$. By (A.1) in Lemma 1, we have

$$\left\|\mathbf{B}\mathbf{D}_n\mathbf{B}'\right\|_{\mathbf{\Sigma}}^2 = p^{-1}\mathrm{tr}\left[\left(\mathbf{D}_n\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right)^2\right] \leq p^{-1}\left\|\mathbf{D}_n\right\|^2\left\|\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right\|^2.$$

Therefore, it follows from (6.4) and (6.8) that

$$E\left\|\mathbf{B}\mathbf{D}_n\mathbf{B}'\right\|_{\mathbf{\Sigma}}^2 = O(n^{-1}p^{-1}K^3). \tag{6.9}$$

Then, we consider the second term $E\left\|\mathbf{B}\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n' + \mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{B}'\right\|_{\mathbf{\Sigma}}^2$. Note that

$$
\begin{aligned}
(6.10)\quad E\left\|\mathbf{B}\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n' + \mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{B}'\right\|_{\mathbf{\Sigma}}^2 &\le 2\left[E\left\|\mathbf{B}\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n'\right\|_{\mathbf{\Sigma}}^2 + E\left\|\mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{B}'\right\|_{\mathbf{\Sigma}}^2\right] \\
&= 4\, E\left\|\mathbf{B}\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n'\right\|_{\mathbf{\Sigma}}^2 \le 8\Big[(n-1)^{-2}\, E\left\|\mathbf{B}\mathbf{X}\mathbf{X}'\mathbf{C}_n'\right\|_{\mathbf{\Sigma}}^2 \\
&\quad + n^{-2}(n-1)^{-2}\, E\left\|\mathbf{B}\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{X}'\mathbf{C}_n'\right\|_{\mathbf{\Sigma}}^2\Big] \\
&\;\widehat{=}\; 8(n-1)^{-2}\mathcal{L}_1 + 8n^{-2}(n-1)^{-2}\mathcal{L}_2.
\end{aligned}
$$

Since $E(\boldsymbol{\varepsilon}|\mathbf{f}) = \mathbf{0}$, conditioning on $\mathbf{X}$ gives

$$
\begin{aligned}
\mathcal{L}_1 &= p^{-1}E\ \mathrm{tr}\left[\mathbf{X}E\left(\mathbf{E}'\mathbf{\Sigma}^{-1}\mathbf{E}|\mathbf{X}\right)\mathbf{X}'\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right] \\
&= p^{-1}E\ \mathrm{tr}\left[\mathbf{X}\ \mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0\right) I_n\ \mathbf{X}'\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right] \\
&\le p^{-1}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0\right) E\left(\|\mathbf{X}\mathbf{X}'\|\right)\left\|\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right\|.
\end{aligned}
$$

In the proof of Lemma 2, it is shown that $E\left(\|\mathbf{X}\mathbf{X}'\|^2\right) = O(n^2K^2)$, which implies that

$$
E\left(\|\mathbf{X}\mathbf{X}'\|\right) \le \left[E\left(\|\mathbf{X}\mathbf{X}'\|^2\right)\right]^{1/2} = O(nK).
$$

By (1.3) and assumptions (B) and (C), we can easily get

$$
\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0\right) \le \mathrm{tr}\left(\mathbf{\Sigma}^{-1}\right) O(1) = O(p),
$$

which along with (6.8) and the above results shows that

$$
\mathcal{L}_1 = O(nK^{3/2}).
$$

Similarly, by conditioning on $\mathbf{X}$ we have

$$
\begin{aligned}
\mathcal{L}_2 &= p^{-1}E\ \mathrm{tr}\left[\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}E\left(\mathbf{E}'\mathbf{\Sigma}^{-1}\mathbf{E}|\mathbf{X}\right)\mathbf{H}\mathbf{1}\mathbf{1}'\mathbf{X}'\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right] \\
&= p^{-1}E\ \mathrm{tr}\left[\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}\ \mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0\right) I_n\ \mathbf{H}\mathbf{1}\mathbf{1}'\mathbf{X}'\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right].
\end{aligned}
$$

Then, applying (A.1)–(A.3) in Lemma 1 gives

$$\begin{aligned}\mathcal{L}_2 &\le p^{-1}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0\right) E\left\|\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}\mathbf{1}\mathbf{1}'\mathbf{X}'\right\|\left\|\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right\| \\ &\le p^{-1}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0\right) E\left\|\mathbf{H}\right\|\left\|\mathbf{X}'\mathbf{X}\right\|\left\|\mathbf{1}\mathbf{1}'\mathbf{1}\mathbf{1}'\right\|\left\|\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right\| \\ &= n^2p^{-1}K^{1/2}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0\right) E\left\|\mathbf{X}'\mathbf{X}\right\|\left\|\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\right\|,\end{aligned}$$

which together with the above results shows that

$$\mathcal{L}_2 = O(n^3K^2).$$

Thus, in view of (6.10) we have

$$E\left\|\mathbf{B}\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n' + \mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{B}'\right\|_{\mathbf{\Sigma}}^2 = O(n^{-1}K^2). \tag{6.11}$$

The third and fourth terms are examined in Lemmas 4 and 5, respectively. Since $K \le p$ by assumption (A), combining (6.9) and (6.11) with (A.8) and (A.11) in Lemmas 4 and 5 results in

$$E\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\mathbf{\Sigma}}^2 = O(n^{-1}K^2) + O(n^{-2}pK).$$

In particular, when $K = O(n^{\alpha_1})$ and $p = O(n^{\alpha})$ for some $0 \le \alpha_1 < 1/2$ and $0 \le \alpha < 2 - \alpha_1$, we have

$$\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\mathbf{\Sigma}} = O_P(n^{-\beta/2})$$

with $\beta = \min(1 - 2\alpha_1, 2 - \alpha - \alpha_1)$, which proves $n^{\beta/2}$-consistency of covariance matrix estimator $\widehat{\mathbf{\Sigma}}$ under norm $\|\cdot\|_{\mathbf{\Sigma}}$.

(2) Then, we prove the $n^{\beta_1/2}$-consistency of $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}$ under norm $\|\cdot\|_{\mathbf{\Sigma}}$. By the Cauchy-Schwarz inequality, it follows from (6.6) that

$$\begin{aligned}E\left\|\widehat{\mathbf{\Sigma}}_{\mathrm{sam}} - \mathbf{\Sigma}\right\|_{\mathbf{\Sigma}}^2 \le 4\ \Big[&E\left\|\mathbf{B}\mathbf{D}_n\mathbf{B}'\right\|_{\mathbf{\Sigma}}^2 + E\left\|\mathbf{G}_n\right\|_{\mathbf{\Sigma}}^2 + 2\,(n-1)^{-2}\,E\left\|\mathbf{B}\mathbf{X}\mathbf{E}'\right\|_{\mathbf{\Sigma}}^2 \\ &+ 2\,[n\,(n-1)]^{-2}\,E\left\|\mathbf{B}\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{E}'\right\|_{\mathbf{\Sigma}}^2\Big].\end{aligned}$$

As in part (1), we will examine each of the above four terms on the right hand side separately. The first term $E\left\|\mathbf{B}\mathbf{D}_n\mathbf{B}'\right\|_{\mathbf{\Sigma}}^2$ has been bounded in (6.9), and the second term $E\left\|\mathbf{G}_n\right\|_{\mathbf{\Sigma}}^2$ is considered in Lemma 6. The third term $E\left\|\mathbf{B}\mathbf{X}\mathbf{E}'\right\|_{\mathbf{\Sigma}}^2$ is exactly $\mathcal{L}_1$ in part (1) above. Using the same argument that was used in part (1) to prove $\mathcal{L}_2$, we can easily get

$$E\left\|\mathbf{B}\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{E}'\right\|_{\mathbf{\Sigma}}^2 = O(n^3K^{3/2}).$$

Thus, by (6.9) and (A.12) in Lemma 6 along with the above results, we have

$$E\left\|\widehat{\mathbf{\Sigma}}_{\mathrm{sam}} - \mathbf{\Sigma}\right\|_{\mathbf{\Sigma}}^2 = O(n^{-1}p^{-1}K^3) + O(n^{-1}p) + O(n^{-1}K^{3/2}).$$

In particular, when $K = O(n^{\alpha_1})$ and $p = O(n^{\alpha})$ for some $0 \leq \alpha < 1$ and $0 \leq \alpha_1 < (1+\alpha)/3$, we have

$$\left\|\widehat{\mathbf{\Sigma}}_{\mathrm{sam}} - \mathbf{\Sigma}\right\|_{\mathbf{\Sigma}} = O_P(n^{-\beta_1/2})$$

with $\beta_1 = 1 - \max(\alpha, 3\alpha_1/2, 3\alpha_1 - \alpha)$, which shows $n^{\beta_1/2}$-consistency of $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}$ under norm $\|\cdot\|_{\mathbf{\Sigma}}$. $\square$

Proof of Theorem 3. (1) First, we prove the weak convergence of $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}^{-1}$ under the Frobenius norm. Note that $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}$ involves sample covariance matrix estimation of $\mathbf{\Sigma}_0$, so the technique in part (2) below does not help. In general, the only available way is as follows. We define $\mathbf{Q}_n = \widehat{\mathbf{\Sigma}}_{\mathrm{sam}} - \mathbf{\Sigma}_n$. It is a basic fact in matrix theory that

$$\left\|\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}^{-1} - \mathbf{\Sigma}_n^{-1}\right\| \leq \left\|\mathbf{\Sigma}_n^{-1}\right\| \frac{\left\|\mathbf{\Sigma}_n^{-1}\mathbf{Q}_n\right\|}{1 - \left\|\mathbf{\Sigma}_n^{-1}\mathbf{Q}_n\right\|} \leq \frac{\left\|\mathbf{\Sigma}_n^{-1}\right\|^2\left\|\mathbf{Q}_n\right\|}{1 - \left\|\mathbf{\Sigma}_n^{-1}\right\|\left\|\mathbf{Q}_n\right\|} \tag{6.12}$$

whenever $\left\|\mathbf{\Sigma}_n^{-1}\right\|\left\|\mathbf{Q}_n\right\| < 1$. From Theorem 1, we know that

$$\left\|\mathbf{Q}_n\right\| = O_P(n^{-1/2}pK).$$

By (A.9), we have $\left\|\mathbf{\Sigma}_n^{-1}\right\| = O(p^{1/2})$. Since $pK^{1/2} = o((n/\log n)^{1/4})$ we see that

$$\left\|\mathbf{\Sigma}_n^{-1}\right\|\left\|\mathbf{Q}_n\right\| \xrightarrow{\mathrm{P}} 0 \quad \text{and} \quad \sqrt{np^{-4}K^{-2}/\log n}\;\left\|\mathbf{\Sigma}_n^{-1}\right\|^2\left\|\mathbf{Q}_n\right\| \xrightarrow{\mathrm{P}} 0.$$

It follows easily that

$$\sqrt{np^{-4}K^{-2}/\log n}\ \frac{\left\|\boldsymbol{\Sigma}_n^{-1}\right\|^2 \|\mathbf{Q}_n\|}{1-\left\|\boldsymbol{\Sigma}_n^{-1}\right\| \|\mathbf{Q}_n\|} \xrightarrow{\mathrm{P}} 0,$$

which along with (6.12) shows that

$$\sqrt{np^{-4}K^{-2}/\log n}\ \left\|\widehat{\boldsymbol{\Sigma}}_{\mathrm{sam}}^{-1}-\boldsymbol{\Sigma}_n^{-1}\right\| \xrightarrow{\mathrm{P}} 0 \quad \text{as } n\to\infty.$$

(2) Then, we show the weak convergence of $\widehat{\boldsymbol{\Sigma}}^{-1}$ under the Frobenius norm. The basic idea is to examine the estimation error for each term of $\widehat{\boldsymbol{\Sigma}}^{-1}$, which has an explicit form thanks to the factor structure. From (1.4), we know that $\widehat{\boldsymbol{\Sigma}}=\widehat{\mathbf{B}}\widehat{\mathrm{cov}}(\mathbf{f})\widehat{\mathbf{B}}'+\widehat{\boldsymbol{\Sigma}}_0$, which along with the Sherman-Morrison-Woodbury formula shows that

$$\widehat{\boldsymbol{\Sigma}}^{-1}=\widehat{\boldsymbol{\Sigma}}_0^{-1}-\widehat{\boldsymbol{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1}+\widehat{\mathbf{B}}'\widehat{\boldsymbol{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1}\widehat{\mathbf{B}}'\widehat{\boldsymbol{\Sigma}}_0^{-1}. \tag{6.13}$$

Thus by (6.7), we have

(6.14)

$$\begin{aligned}
\left\|\widehat{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right\| \le{}& \left\|\widehat{\boldsymbol{\Sigma}}_0^{-1}-\boldsymbol{\Sigma}_0^{-1}\right\|+\left\|\left(\widehat{\boldsymbol{\Sigma}}_0^{-1}-\boldsymbol{\Sigma}_0^{-1}\right)\widehat{\mathbf{B}}\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1}+\widehat{\mathbf{B}}'\widehat{\boldsymbol{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1}\widehat{\mathbf{B}}'\widehat{\boldsymbol{\Sigma}}_0^{-1}\right\| \\
&+\left\|\boldsymbol{\Sigma}_0^{-1}\widehat{\mathbf{B}}\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1}+\widehat{\mathbf{B}}'\widehat{\boldsymbol{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1}\widehat{\mathbf{B}}'\left(\widehat{\boldsymbol{\Sigma}}_0^{-1}-\boldsymbol{\Sigma}_0^{-1}\right)\right\| \\
&+\left\|\boldsymbol{\Sigma}_0^{-1}\left(\widehat{\mathbf{B}}-\mathbf{B}\right)\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1}+\widehat{\mathbf{B}}'\widehat{\boldsymbol{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1}\widehat{\mathbf{B}}'\boldsymbol{\Sigma}_0^{-1}\right\| \\
&+\left\|\boldsymbol{\Sigma}_0^{-1}\mathbf{B}\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1}+\widehat{\mathbf{B}}'\widehat{\boldsymbol{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1}\left(\widehat{\mathbf{B}}'-\mathbf{B}'\right)\boldsymbol{\Sigma}_0^{-1}\right\| \\
&+\left\|\boldsymbol{\Sigma}_0^{-1}\mathbf{B}\left\{\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1}+\widehat{\mathbf{B}}'\widehat{\boldsymbol{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1}-\left[\mathrm{cov}(\mathbf{f})^{-1}+\mathbf{B}'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}\right]^{-1}\right\}\mathbf{B}'\boldsymbol{\Sigma}_0^{-1}\right\| \\
\widehat{=}{}& \mathcal{K}_1+\mathcal{K}_2+\mathcal{K}_3+\mathcal{K}_4+\mathcal{K}_5+\mathcal{K}_6.
\end{aligned}$$

To study $\left\|\widehat{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right\|$, we need to examine each of the above six terms $\mathcal{K}_1,\cdots,\mathcal{K}_6$ separately, so it would be lengthy work to check all the details here. Therefore, we only sketch the idea of the proof and leave the details to the reader.

From assumption (C), we know that the diagonal entries of $\mathbf{\Sigma}_0$ are bounded away from 0. Note that $\widehat{\mathbf{\Sigma}}_0$ and $\mathbf{\Sigma}_0$ are both diagonal, and thus, by the same argument as in Lemma 5, we can easily show that

$$\mathcal{K}_1 = \left\| \widehat{\mathbf{\Sigma}}_0^{-1} - \mathbf{\Sigma}_0^{-1} \right\| = O_P(n^{-1/2}p^{1/2}) + O_P(n^{-1}pK^{1/2}) = O_P(n^{-1/2}p^{1/2}), \tag{6.15}$$

since $pK^{1/2} = o((n/\log n)^{1/2})$. Now we consider the second term $\mathcal{K}_2$. By (A.1) in Lemma 1, we have

$$\begin{aligned}\mathcal{K}_2 &\le \left\| \left(\widehat{\mathbf{\Sigma}}_0^{-1} - \mathbf{\Sigma}_0^{-1} \right) \widehat{\mathbf{\Sigma}}_0^{1/2} \right\| \left\| \widehat{\mathbf{\Sigma}}_0^{-1/2} \widehat{\mathbf{B}} \left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1} + \widehat{\mathbf{B}}' \widehat{\mathbf{\Sigma}}_0^{-1} \widehat{\mathbf{B}} \right]^{-1} \widehat{\mathbf{B}}' \widehat{\mathbf{\Sigma}}_0^{-1/2} \right\| \left\| \widehat{\mathbf{\Sigma}}_0^{-1/2} \right\| \\ &\mathrel{\widehat{=}} \mathcal{L}_1 \mathcal{L}_2 \left\| \widehat{\mathbf{\Sigma}}_0^{-1/2} \right\|,\end{aligned}$$

and we will examine each of the above two terms $\mathcal{L}_1$ and $\mathcal{L}_2$, as well as $\left\| \widehat{\mathbf{\Sigma}}_0^{-1/2} \right\|$. Since $\widehat{\mathbf{\Sigma}}_0$ and $\mathbf{\Sigma}_0$ are diagonal, a similar argument to that bounding $\mathcal{K}_1$ above applies to show that

$$\left\| \widehat{\mathbf{\Sigma}}_0^{-1/2} \right\| = O_P(p^{1/2}) \quad \text{and} \quad \mathcal{L}_1 = O_P(n^{-1/2}p^{1/2}).$$

Clearly, $\widehat{\mathbf{\Sigma}}_0^{-1/2} \widehat{\mathbf{B}} \left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1} + \widehat{\mathbf{B}}' \widehat{\mathbf{\Sigma}}_0^{-1} \widehat{\mathbf{B}} \right]^{-1} \widehat{\mathbf{B}}' \widehat{\mathbf{\Sigma}}_0^{-1/2}$ is symmetric positive semidefinite with rank at most K and $\widehat{\mathbf{\Sigma}}_0^{1/2} \widehat{\mathbf{\Sigma}}^{-1} \widehat{\mathbf{\Sigma}}_0^{1/2} \ge 0$. Thus it follows from (6.13) that

$$\widehat{\mathbf{\Sigma}}_0^{-1/2} \widehat{\mathbf{B}} \left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1} + \widehat{\mathbf{B}}' \widehat{\mathbf{\Sigma}}_0^{-1} \widehat{\mathbf{B}} \right]^{-1} \widehat{\mathbf{B}}' \widehat{\mathbf{\Sigma}}_0^{-1/2} = I_p - \widehat{\mathbf{\Sigma}}_0^{1/2} \widehat{\mathbf{\Sigma}}^{-1} \widehat{\mathbf{\Sigma}}_0^{1/2} \le I_p,$$

which implies that $\widehat{\mathbf{\Sigma}}_0^{-1/2} \widehat{\mathbf{B}} \left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1} + \widehat{\mathbf{B}}' \widehat{\mathbf{\Sigma}}_0^{-1} \widehat{\mathbf{B}} \right]^{-1} \widehat{\mathbf{B}}' \widehat{\mathbf{\Sigma}}_0^{-1/2}$ has at most K positive eigenvalues and all of them are bounded by one. This shows that $\mathcal{L}_2 \le K^{1/2}$, which along with the above results gives

$$\mathcal{K}_2 = O_P(n^{-1/2}pK^{1/2}). \tag{6.16}$$

Similarly, we can also show that

$$\mathcal{K}_3 = O_P(n^{-1/2}pK^{1/2}). \tag{6.17}$$

Then we consider terms $\mathcal{K}_4$ and $\mathcal{K}_5$. Clearly, $\widehat{\mathrm{cov}}(\mathbf{f})^{-1} + \widehat{\mathbf{B}}'\widehat{\mathbf{\Sigma}}_0^{-1}\widehat{\mathbf{B}} \geq \widehat{\mathrm{cov}}(\mathbf{f})^{-1}$, which in turn entails that $\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1} + \widehat{\mathbf{B}}'\widehat{\mathbf{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1} \leq \widehat{\mathrm{cov}}(\mathbf{f})$ and

$$\left\|\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1} + \widehat{\mathbf{B}}'\widehat{\mathbf{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1}\right\| \leq \|\widehat{\mathrm{cov}}(\mathbf{f})\|.$$

It is easy to show that $\|\widehat{\mathrm{cov}}(\mathbf{f})\| = O_P(K)$. Thus we have

$$\begin{aligned}\mathcal{K}_4 &\leq \left\|\mathbf{\Sigma}_0^{-1}\left(\widehat{\mathbf{B}} - \mathbf{B}\right)\right\| \left\|\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1} + \widehat{\mathbf{B}}'\widehat{\mathbf{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1}\right\| \left\|\widehat{\mathbf{B}}'\mathbf{\Sigma}_0^{-1}\right\| \\ &= O_P(n^{-1}p^{1/2})O_P(K)O_P(p^{1/2}) = O_P(n^{-1/2}pK)\end{aligned} \tag{6.18}$$

and

$$\begin{aligned}\mathcal{K}_5 &\leq \|\mathbf{\Sigma}_0^{-1}\mathbf{B}\| \left\|\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1} + \widehat{\mathbf{B}}'\widehat{\mathbf{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1}\right\| \left\|\left(\widehat{\mathbf{B}}' - \mathbf{B}'\right)\mathbf{\Sigma}_0^{-1}\right\| \\ &= O_P(p^{1/2})O_P(K)O_P(n^{-1}p^{1/2}K) = O_P(n^{-1/2}pK).\end{aligned} \tag{6.19}$$

Finally, by the same argument as in part (1) above, we can show that

$$\left\|\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1} + \widehat{\mathbf{B}}'\widehat{\mathbf{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1} - \left[\mathrm{cov}(\mathbf{f})^{-1} + \mathbf{B}'\mathbf{\Sigma}_0^{-1}\mathbf{B}\right]^{-1}\right\| = o_P((n/\log n)^{-1/2}K^2).$$

Thus by (A.2) in Lemma 1, we have

$$\begin{aligned}\mathcal{K}_6 &\leq \left\|\left[\widehat{\mathrm{cov}}(\mathbf{f})^{-1} + \widehat{\mathbf{B}}'\widehat{\mathbf{\Sigma}}_0^{-1}\widehat{\mathbf{B}}\right]^{-1} - \left[\mathrm{cov}(\mathbf{f})^{-1} + \mathbf{B}'\mathbf{\Sigma}_0^{-1}\mathbf{B}\right]^{-1}\right\| \left\|\mathbf{B}'\mathbf{\Sigma}_0^{-2}\mathbf{B}\right\| \\ &= o_P((n/\log n)^{-1/2}K^2)O(p) = o_P((n/\log n)^{-1/2}pK^2).\end{aligned} \tag{6.20}$$

Therefore, it follows from (6.14)–(6.20) that

$$\sqrt{np^{-2}K^{-4}/\log n}\ \left\|\widehat{\mathbf{\Sigma}}_n^{-1} - \mathbf{\Sigma}_n^{-1}\right\| \xrightarrow{\mathrm{P}} 0 \quad \text{as } n \to \infty,$$

which completes the proof. $\square$

Proof of Theorem 4. We aim at establishing asymptotic normality of the $K \times K$ matrix $\sqrt{n}p^{-2}\mathbf{B}'\left(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right)\mathbf{B}$, and only here are the K factors $f_1, \cdots, f_K$ assumed fixed across n. The basic idea is to use its four-term decomposition below and to show that the first term has asymptotic normality by the classical central limit theorem, while the remaining three terms are all negligible, say $o_P(1)$, which along with Slutsky's theorem leads to the desired conclusion. In view of (6.1), we have

$$\begin{aligned}\sqrt{n}p^{-2}\mathbf{B}'\left(\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\right)\mathbf{B} &= \sqrt{n}p^{-2}\mathbf{B}'\mathbf{B}\mathbf{D}_n\mathbf{B}'\mathbf{B}+\sqrt{n}p^{-2}\mathbf{B}'\left\{\mathbf{B}\widehat{\text{cov}}(\mathbf{f})\mathbf{C}_n'+\mathbf{C}_n\widehat{\text{cov}}(\mathbf{f})\mathbf{B}'\right\}\mathbf{B}\\ &\quad+\sqrt{n}p^{-2}\mathbf{B}'\mathbf{C}_n\widehat{\text{cov}}(\mathbf{f})\mathbf{C}_n'\mathbf{B}+\sqrt{n}p^{-2}\mathbf{B}'\mathbf{F}_n\mathbf{B}\\ &\;\widehat{=}\;\mathcal{A}_1+\mathcal{A}_2+\mathcal{A}_3+\mathcal{A}_4. \qquad (6.21)\end{aligned}$$

We will study each of the above four terms $\mathcal{A}_1,\cdots,\mathcal{A}_4$ separately.

First, we consider the term $\mathcal{A}_1$. Define

$$\mathcal{H}_n=\frac{n}{n-1}\left(n^{-1}\sum_{i=1}^n\mathbf{f}_i-E\mathbf{f}\right)\left(n^{-1}\sum_{i=1}^n\mathbf{f}_i'-E\mathbf{f}'\right).$$

Then we have

$$\widehat{\text{cov}}(\mathbf{f})=(n-1)^{-1}\sum_{i=1}^n\left(\mathbf{f}_i-E\mathbf{f}\right)\left(\mathbf{f}_i'-E\mathbf{f}'\right)-\mathcal{H}_n. \qquad (6.22)$$

By the classical central limit theorem, we know that

$$\sqrt{n}\left(n^{-1}\sum_{i=1}^n\mathbf{f}_i-E\mathbf{f}\right)\xrightarrow{\text{D}}\mathcal{N}\left(\mathbf{0},\text{cov}(\mathbf{f})\right).$$

It follows from the law of large numbers that $n^{-1}\sum_{i=1}^n\mathbf{f}_i-E\mathbf{f}\xrightarrow{\text{P}}\mathbf{0}$. Thus, by Slutsky's theorem we have $\sqrt{n}\mathcal{H}_n\xrightarrow{\text{D}}\mathbf{0}$, which in turn implies that

$$\sqrt{n}\mathcal{H}_n\xrightarrow{\text{P}}\mathbf{0};$$

that is, $\mathcal{H}_n=o_P(n^{-1/2})$. So in view of (6.22), we have

$$\widehat{\text{cov}}(\mathbf{f})=n^{-1}\sum_{i=1}^n\left(\mathbf{f}_i-E\mathbf{f}\right)\left(\mathbf{f}_i'-E\mathbf{f}'\right)+o_P(n^{-1/2}). \qquad (6.23)$$

Therefore, it follows easily from $p^{-1}\mathbf{B}_n'\mathbf{B}_n \to \mathbf{A}$ and (6.23) that

$$\mathcal{A}_1 = \mathbf{A}\left\{n^{-1/2}\sum_{i=1}^{n}\left[(\mathbf{f}_i - E\mathbf{f})\left(\mathbf{f}_i' - E\mathbf{f}'\right) - \operatorname{cov}(\mathbf{f})\right]\right\}\mathbf{A} + o_P(1). \tag{6.24}$$

We define

$$n^{-1/2}\sum_{i=1}^{n}\left[(\mathbf{f}_i - E\mathbf{f})\left(\mathbf{f}_i' - E\mathbf{f}'\right) - \operatorname{cov}(\mathbf{f})\right] \;\widehat{=}\; U_n = (u_{ij})_{K\times K}.$$

By the classical central limit theorem, we know that [see, e.g. Muirhead (1982)]

$$\operatorname{vech}(U_n) \xrightarrow{\mathrm{D}} \mathcal{N}(0, H), \tag{6.25}$$

where H is determined in an obvious way by

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

with $\kappa^{i_1\cdots i_r}$ the central moment $E\left[(f_{i_1} - Ef_{i_1})\cdots(f_{i_r} - Ef_{i_r})\right]$ of $\mathbf{f} = (f_1, \cdots, f_K)'$. It follows easily from (6.24) and (6.25) that

$$\operatorname{vech}(\mathcal{A}_1) \xrightarrow{\mathrm{D}} \mathcal{N}(0, G), \tag{6.26}$$

where $G = P_D(\mathbf{A}\otimes\mathbf{A})DHD'(\mathbf{A}\otimes\mathbf{A})P_D'$, D is the duplication matrix of order K, and $P_D = (D'D)^{-1}D'$.

Then, we examine the second term $\mathcal{A}_2$. From $p^{-1}\mathbf{B}_n'\mathbf{B}_n \to \mathbf{A}$, we know that

$$\left\|\mathbf{B}_n'\mathbf{B}_n\right\| = \left\|\mathbf{B}_n\mathbf{B}_n'\right\| = O(p), \tag{6.27}$$

which is in line with (6.3). It follows that

$$\begin{aligned}\|\mathcal{A}_2\| &\le 2\left\|\sqrt{n}p^{-2}\mathbf{B}'\mathbf{B}\widehat{\operatorname{cov}}(\mathbf{f})\mathbf{C}_n'\mathbf{B}\right\| \le 2n^{1/2}p^{-2}\left\|\mathbf{B}'\mathbf{B}\right\|\left\|\widehat{\operatorname{cov}}(\mathbf{f})\mathbf{C}_n'\mathbf{B}\right\| \\ &\le 2n^{1/2}p^{-2}\left\|\mathbf{B}'\mathbf{B}\right\|\left\{(n-1)^{-1}\left\|\mathbf{X}\mathbf{E}'\mathbf{B}\right\| + n^{-1}(n-1)^{-1}\left\|\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}\mathbf{E}'\mathbf{B}\right\|\right\} \\ &= O(n^{-1/2}p^{-1})\left\|\mathbf{X}\mathbf{E}'\mathbf{B}\right\| + O(n^{-3/2}p^{-1})\left\|\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}\mathbf{E}'\mathbf{B}\right\|.\end{aligned} \tag{6.28}$$

Since $E(\boldsymbol{\varepsilon}|\mathbf{f}) = \mathbf{0}$ and $\boldsymbol{\Sigma}_0$ is diagonal, conditioning on $\mathbf{X}$ gives

$$E\left\|\mathbf{X}\mathbf{E}'\mathbf{B}\right\|^2 = E \text{ tr}\left[\mathbf{X}E\left(\mathbf{E}'\mathbf{B}\mathbf{B}'\mathbf{E}|\mathbf{X}\right)\mathbf{X}'\right] = E \text{ tr}\left[\mathbf{X} \text{ tr}\left(\mathbf{B}\mathbf{B}'\boldsymbol{\Sigma}_0\right) I_n \mathbf{X}'\right]$$
$$= \text{tr}\left(\mathbf{B}\mathbf{B}'\boldsymbol{\Sigma}_0\right) E\left\|\mathbf{X}\right\|^2 = O(p)O(n) = O(np).$$

Similarly, by conditioning on $\mathbf{X}$ we have

$$E\left\|\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}\mathbf{E}'\mathbf{B}\right\|^2 = E \text{ tr}\Big[\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}E\left(\mathbf{E}'\mathbf{B}\mathbf{B}'\mathbf{E}'|\mathbf{X}\right)\mathbf{H}\mathbf{1}\mathbf{1}'\mathbf{X}'\Big]$$
$$= E \text{ tr}\left[\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H} \text{ tr}\left(\mathbf{B}\mathbf{B}'\boldsymbol{\Sigma}_0\right) I_n \mathbf{H}\mathbf{1}\mathbf{1}'\mathbf{X}'\right]$$

and then applying (A.2) and (A.3) in Lemma 1 yields

$$E\left\|\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}\boldsymbol{\varepsilon}'\mathbf{B}\right\|^2 \le \text{tr}\left(\mathbf{B}\mathbf{B}'\boldsymbol{\Sigma}_0\right) E\Big\{\left\|\mathbf{X}'\mathbf{X}\right\|\left\|\mathbf{1}\mathbf{1}'\mathbf{1}\mathbf{1}'\right\|\left\|\mathbf{H}\right\|\Big\}$$
$$\le O(p)n^2K^{1/2}\left\{E\left(\|\mathbf{X}'\mathbf{X}\|^2\right)\right\}^{1/2} = O(n^3p).$$

It follows that $\|\mathbf{X}\mathbf{E}'\mathbf{B}\| = O_P(n^{1/2}p^{1/2})$ and $\|\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}\mathbf{E}'\mathbf{B}\| = O_P(n^{3/2}p^{1/2})$, which together with (6.28) shows that

$$\mathcal{A}_2 = o_P(1); \tag{6.29}$$

that is, $\mathcal{A}_2$ is a negligible term.

Finally, the third and fourth terms $\mathcal{A}_3$ and $\mathcal{A}_4$ can also be shown to be negligible by invoking Lemma 3. By (6.27) and (A.6) and (A.7) in Lemma 3, we have

$$E\left\|\mathbf{B}'\mathbf{C}_n\widehat{\text{cov}}(\mathbf{f})\mathbf{C}_n'\mathbf{B}\right\|^2 \le \left\|\mathbf{B}\mathbf{B}'\right\|^2 E\left\|\mathbf{C}_n\widehat{\text{cov}}(\mathbf{f})\mathbf{C}_n'\right\|^2$$
$$= O(p^2)O(n^{-2}p^2) = O(n^{-2}p^4)$$

and

$$E\left\|\mathbf{B}'\mathbf{F}_n\mathbf{B}\right\|^2 \le \left\|\mathbf{B}\mathbf{B}'\right\|^2 E\left\|\mathbf{F}_n\right\|^2 = O(p^2)O(n^{-1}p) = O(n^{-1}p^3).$$

It follows that $\|\mathbf{B}'\mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n'\mathbf{B}\| = O_P(n^{-1}p^2)$ and $\|\mathbf{B}'\mathbf{F}_n\mathbf{B}\| = O_P(n^{-1/2}p^{3/2})$, which implies that

$$\mathcal{A}_3 = o_P(1) \quad \text{and} \quad \mathcal{A}_4 = o_P(1). \tag{6.30}$$

Therefore, in view of (6.26), (6.29), and (6.30), applying Slutsky's theorem gives

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

which proves the asymptotic normality of covariance matrix estimator $\widehat{\mathbf{\Sigma}}$. $\square$

Proof of Theorem 5. (1) First, we prove the weak convergence of the estimated global minimum variance based on $\widehat{\mathbf{\Sigma}}$. From Theorem 3, we know that

$$\sqrt{np^{-2}K^{-4}/\log n}\ \left\|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\right\| \xrightarrow{\mathrm{P}} 0.$$

Note that

$$\begin{aligned}|\widehat{\varphi}_n - \varphi_n| &= \left|\mathbf{1}'\left(\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\right)\mathbf{1}\right| = \left|\mathrm{tr}\left[\left(\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\right)\mathbf{1}\mathbf{1}'\right]\right| \\ &\le \left\|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\right\|\|\mathbf{1}\mathbf{1}'\| = p\left\|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\right\|.\end{aligned}$$

Thus we have

$$\sqrt{n\left(pK\right)^{-4}/\log n}\ |\widehat{\varphi}_n - \varphi_n| \xrightarrow{\mathrm{P}} 0.$$

Since all the φ_n's are bounded away from zero, it follows easily that

$$\sqrt{n\left(pK\right)^{-4}/\log n}\ \left|\widehat{\boldsymbol{\xi}}_{ng}'\widehat{\mathbf{\Sigma}}_n\widehat{\boldsymbol{\xi}}_{ng} - \boldsymbol{\xi}_{ng}'\mathbf{\Sigma}_n\boldsymbol{\xi}_{ng}\right| = \sqrt{n\left(pK\right)^{-4}/\log n}\ \left|\widehat{\varphi}_n^{-1} - \varphi_n^{-1}\right| \xrightarrow{\mathrm{P}} 0.$$

(2) Then, we prove the conclusion for $\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}$. From Theorem 3, we know that

$$\sqrt{np^{-4}K^{-2}/\log n}\ \left\|\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}^{-1} - \mathbf{\Sigma}^{-1}\right\| \xrightarrow{\mathrm{P}} 0.$$

Therefore, the above argument in part (1) applies to show that

$$\sqrt{np^{-6}K^{-2}/\log n}\ \left|\widehat{\boldsymbol{\xi}}_{ng}'\widehat{\mathbf{\Sigma}}_{\mathrm{sam}}\widehat{\boldsymbol{\xi}}_{ng} - \boldsymbol{\xi}_{ng}'\mathbf{\Sigma}_n\boldsymbol{\xi}_{ng}\right| = \sqrt{np^{-6}K^{-2}/\log n}\ \left|\widehat{\varphi}_n^{-1} - \varphi_n^{-1}\right| \xrightarrow{\mathrm{P}} 0. \quad \square$$

Proof of Theorem 6. (1) First, we prove the weak convergence of the estimated variance of the optimal portfolio based on $\widehat{\mathbf{\Sigma}}$. From Theorem 3, we know that

$$\sqrt{np^{-2}K^{-4}/\log n}\ \left\|\widehat{\mathbf{\Sigma}}^{-1}-\mathbf{\Sigma}^{-1}\right\| \xrightarrow{\mathrm{P}} 0, \tag{6.31}$$

and from part (1) in the proof of Theorem 5, we see that

$$\sqrt{n\,(pK)^{-4}/\log n}\ |\widehat{\varphi}_n-\varphi_n| \xrightarrow{\mathrm{P}} 0. \tag{6.32}$$

Now we show the same rate for $\left|\widehat{\psi}_n-\psi_n\right|$, say

$$\sqrt{n\,(pK)^{-4}/\log n}\ \left|\widehat{\psi}_n-\psi_n\right| \xrightarrow{\mathrm{P}} 0. \tag{6.33}$$

By $b_n = O(p)$ in assumption (B), a routine calculation yields $\|\boldsymbol{\mu}_n\| = O(p^{1/2})$ and $E\left\|\widehat{\boldsymbol{\mu}}_n-\boldsymbol{\mu}_n\right\|^2 = O(n^{-1}p)$, and thus

$$\|\widehat{\boldsymbol{\mu}}_n-\boldsymbol{\mu}_n\| = O_P(n^{-1/2}p^{1/2}).$$

It follows that

$$\begin{aligned}\left|\widehat{\psi}_n-\psi_n\right| &\le \left|\mathbf{1}'\left(\widehat{\mathbf{\Sigma}}^{-1}-\mathbf{\Sigma}^{-1}\right)\widehat{\boldsymbol{\mu}}\right| + \left|\mathbf{1}'\mathbf{\Sigma}^{-1}\left(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\right)\right| \le \|\mathbf{1}'\|\left\|\widehat{\mathbf{\Sigma}}^{-1}-\mathbf{\Sigma}^{-1}\right\| \\ &\quad \cdot \left(\|\boldsymbol{\mu}\|+\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\|\right) + \|\mathbf{1}'\|\left\|\mathbf{\Sigma}^{-1}\right\|\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\|.\end{aligned}$$

Then we have

$$\begin{aligned}\left|\widehat{\psi}_n-\psi_n\right| &\le p^{1/2}\left\|\widehat{\mathbf{\Sigma}}^{-1}-\mathbf{\Sigma}^{-1}\right\|\left[O(p^{1/2})+O_P(n^{-1/2}p^{1/2})\right] \\ &\quad + p^{1/2}O(p^{1/2})O_P(n^{-1/2}p^{1/2}) \\ &= \left\|\widehat{\mathbf{\Sigma}}^{-1}-\mathbf{\Sigma}^{-1}\right\|O(p)+O_P(n^{-1/2}p^{3/2}) = \left\|\widehat{\mathbf{\Sigma}}^{-1}-\mathbf{\Sigma}^{-1}\right\|O(p),\end{aligned}$$

which together with (6.31) proves (6.32). Similarly, we can also show that

$$\sqrt{n\,(pK)^{-4}/\log n}\ \left|\widehat{\phi}_n-\phi_n\right| \xrightarrow{\mathrm{P}} 0. \tag{6.34}$$

Since $\varphi_n\phi_n - \psi_n^2$ 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, the conclusion follows from (3.3) and (6.32)–(6.34).

(2) Now we prove the conclusion for $\widehat{\mathbf{\Sigma}}_{\text{sam}}$. From Theorem 3, we know that

$$\sqrt{np^{-4}K^{-2}/\log n}\ \left\|\widehat{\mathbf{\Sigma}}_{\text{sam}}^{-1} - \mathbf{\Sigma}^{-1}\right\| \xrightarrow{\text{P}} 0,$$

and from part (2) in the proof of Theorem 5, we see that

$$\sqrt{np^{-6}K^{-2}/\log n}\ |\widehat{\varphi}_n - \varphi_n| \xrightarrow{\text{P}} 0.$$

Since $b_n = O(p)$ by assumption (B), a routine calculation shows that

$$\|\widehat{\boldsymbol{\mu}}_{\text{sam}} - \boldsymbol{\mu}_n\| = O_P(n^{-1/2}p^{1/2}),$$

where $\widehat{\boldsymbol{\mu}}_{\text{sam}}$ is the sample mean of $\boldsymbol{\mu}_n$. Therefore, the argument in part (1) above applies to show that

$$\sqrt{np^{-6}K^{-2}/\log n}\ \left|\widehat{\boldsymbol{\xi}}_n'\widehat{\mathbf{\Sigma}}_{\text{sam}}\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_n'\mathbf{\Sigma}_n\boldsymbol{\xi}_n\right| \xrightarrow{\text{P}} 0 \quad \text{as } n \to \infty. \quad \square$$

Proof of Theorem 7. Since $\boldsymbol{\xi}_n = O(1)\mathbf{1}$, the conclusion follows easily from consistency results of $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Sigma}}_{\text{sam}}$ under the Frobenius norm in Theorem 1. In particular, when the portfolios $\boldsymbol{\xi}_n = (\xi_1, \cdots, \xi_p)'$ have no short positions, we have

$$\|\boldsymbol{\xi}_n\| = \sqrt{\xi_1^2 + \cdots + \xi_p^2} \leq \sqrt{\xi_1 + \cdots + \xi_p} = 1.$$

It therefore follows easily that

$$\sqrt{n\,(pK)^{-2}\,/\log n}\ \left|\boldsymbol{\xi}_n'\widehat{\mathbf{\Sigma}}_n\boldsymbol{\xi}_n - \boldsymbol{\xi}_n'\mathbf{\Sigma}_n\boldsymbol{\xi}_n\right| \xrightarrow{\text{P}} 0 \quad \text{as } n \to \infty$$

and

$$\sqrt{n\,(pK)^{-2}\,/\log n}\ \left|\boldsymbol{\xi}_n'\widehat{\mathbf{\Sigma}}_{\text{sam}}\boldsymbol{\xi}_n - \boldsymbol{\xi}_n'\mathbf{\Sigma}_n\boldsymbol{\xi}_n\right| \xrightarrow{\text{P}} 0 \quad \text{as } n \to \infty. \quad \square$$

APPENDIX

Throughout the paper, we denote by $\mathbf{H}$ the $n \times n$ hat matrix $\mathbf{X}'\left(\mathbf{X}\mathbf{X}'\right)^{-1}\mathbf{X}$, which is symmetric and positive semidefinite with probability one by assumption (A).

LEMMA 1 (Basic facts).

(i) For any $q \times r$ matrix $\mathbf{A}_1$ and $r \times q$ matrix $\mathbf{A}_2$, we have

$$|\mathrm{tr}\left(\mathbf{A}_1\mathbf{A}_2\right)| \leq \|\mathbf{A}_1\| \|\mathbf{A}_2\| \;\; \textit{and} \;\; \|\mathbf{A}_1\mathbf{A}_2\| \leq \|\mathbf{A}_1\| \|\mathbf{A}_2\|. \tag{A.1}$$

In particular, for any $q \times r$ matrix $\mathbf{A}_1$ and $r \times r$ symmetric matrix $\mathbf{A}_2$, we have

$$\left|\mathrm{tr}\left(\mathbf{A}_1\mathbf{A}_2\mathbf{A}_1'\right)\right| \leq \left\|\mathbf{A}_1'\mathbf{A}_1\right\| \|\mathbf{A}_2\| \;\; \textit{and} \;\; \left\|\mathbf{A}_1\mathbf{A}_2\mathbf{A}_1'\right\| \leq \left\|\mathbf{A}_1'\mathbf{A}_1\right\| \|\mathbf{A}_2\|. \tag{A.2}$$

(ii) With probability one, the hat matrix $\mathbf{H}$ is idempotent with

$$\mathrm{tr}\left(\mathbf{H}^2\right) = \mathrm{tr}\left(\mathbf{H}\right) = K, \tag{A.3}$$

and it satisfies

$$0 \leq \mathrm{tr}\left(\mathbf{H}\mathbf{1}\mathbf{1}'\mathbf{H}\right) \leq K^{1/2}n \quad \textit{and} \quad 0 \leq \mathrm{tr}\left[\left(\mathbf{H}\mathbf{1}\mathbf{1}'\mathbf{H}\right)^2\right] \leq Kn^2. \tag{A.4}$$

PROOF. One can refer to Horn and Johnson (1990) for standard proofs of (A.1) and (A.2). The fact that the hat matrix $\mathbf{H}$ is idempotent with (A.3) is known in multivariate statistical analysis. Clearly, $\mathrm{tr}\left(\mathbf{H}\mathbf{1}\mathbf{1}'\mathbf{H}\right) = \mathbf{1}'\mathbf{H}\mathbf{1} \geq 0$. Thus by (A.1) and (A.3), we have

$$\mathrm{tr}\left(\mathbf{H}\mathbf{1}\mathbf{1}'\mathbf{H}\right) = \mathrm{tr}\left(\mathbf{H}\mathbf{1}\mathbf{1}'\right) \leq \|\mathbf{H}\| \left\|\mathbf{1}\mathbf{1}'\right\| = K^{1/2}n$$

and

$$\mathrm{tr}\left[\left(\mathbf{H}\mathbf{1}\mathbf{1}'\mathbf{H}\right)^2\right] = \mathrm{tr}\left[\left(\mathbf{H}\mathbf{1}\mathbf{1}'\right)^2\right] \leq \left\|\mathbf{H}\mathbf{1}\mathbf{1}'\right\|^2 \leq \|\mathbf{H}\|^2 \left\|\mathbf{1}\mathbf{1}'\right\|^2 = Kn^2.$$

This completes the proof. $\square$

The main trick in the proofs of the technical lemmas below is conditioning on $\mathbf{X}$ and resorting to the basic facts from Lemma 1.

Lemma 2. *Under conditions (A) and (B), we have*

$$E\ \mathrm{tr}\left\{\left[\mathbf{B}\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n' + \mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{B}'\right]^2\right\} \le \left\|\mathbf{B}'\mathbf{B}\right\| O(n^{-1}pK^{3/2}). \tag{A.5}$$

Proof. It follows from (A.1) that

$$\begin{aligned}
E\ \mathrm{tr}\left\{\left[\mathbf{B}\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{C}_n' + \mathbf{C}_n\widehat{\mathrm{cov}}(\mathbf{f})\mathbf{B}'\right]^2\right\} &\le 2\,(n-1)^{-2}\, E\ \mathrm{tr}\left\{\left[\mathbf{BXE}' + \mathbf{EX}'\mathbf{B}'\right]^2\right\} \\
&\quad + 2n^{-2}\,(n-1)^{-2}\, E\ \mathrm{tr}\left\{\left[\mathbf{BX11}'\mathbf{HE}' + \mathbf{EH11}'\mathbf{X}'\mathbf{B}'\right]^2\right\} \\
&\widehat{=}\ 2\,(n-1)^{-2}\,\mathcal{A}_1 + 2n^{-2}\,(n-1)^{-2}\,\mathcal{A}_2.
\end{aligned}$$

We will consider the above two terms $\mathcal{A}_1$ and $\mathcal{A}_2$ separately. By $c_n = O(1)$ in assumption (B), we can easily get $\left\|E\left(\mathbf{ff}'\right)\right\| = O(K)$ and $E\left(\|\mathbf{f}\|^4\right) = O(K^2)$.

Since $E(\mathbf{E}|\mathbf{f}) = \mathbf{0}$, by (A.1) and (A.2) conditioning on $\mathbf{X}$ results in

$$\begin{aligned}
E\left\|\mathbf{BXE}'\right\|^2 &= E\ \mathrm{tr}\left[\mathbf{BX}E\left(\mathbf{E}'\mathbf{E}|\mathbf{X}\right)\mathbf{X}'\mathbf{B}'\right] = E\ \mathrm{tr}\left[\mathbf{BX}\ \mathrm{tr}\left(\mathbf{\Sigma}_0\right) I_n\ \mathbf{X}'\mathbf{B}'\right] \\
&= n\ \mathrm{tr}\left(\mathbf{\Sigma}_0\right) E\ \mathrm{tr}\left[\mathbf{Bff}'\mathbf{B}'\right] = n\ \mathrm{tr}\left(\mathbf{\Sigma}_0\right)\ \mathrm{tr}\left[\mathbf{B}E\left(\mathbf{ff}'\right)\mathbf{B}'\right] \\
&\le n\ \mathrm{tr}\left(\mathbf{\Sigma}_0\right)\left\|\mathbf{B}'\mathbf{B}\right\|\left\|E\left(\mathbf{ff}'\right)\right\| = \left\|\mathbf{B}'\mathbf{B}\right\|\mathrm{tr}\left(\mathbf{\Sigma}_0\right) O(nK).
\end{aligned}$$

Similarly, by conditioning on $\mathbf{X}$ we have

$$\begin{aligned}
E\left\|\mathbf{BX11}'\mathbf{HE}'\right\|^2 &= E\ \mathrm{tr}\left[\mathbf{BX11}'\mathbf{H}E\left(\mathbf{E}'\mathbf{E}|\mathbf{X}\right)\mathbf{H11}'\mathbf{X}'\mathbf{B}'\right] \\
&= E\ \mathrm{tr}\left[\mathbf{BX11}'\mathbf{H}\ \mathrm{tr}\left(\mathbf{\Sigma}_0\right) I_n\ \mathbf{H11}'\mathbf{X}'\mathbf{B}'\right],
\end{aligned}$$

and then applying (A.1) and (A.2) in Lemma 1 gives

$$\begin{aligned}
E\left\|\mathbf{BX11}'\mathbf{HE}'\right\|^2 &\le \mathrm{tr}\left(\mathbf{\Sigma}_0\right)\ E\left\{\left\|\mathbf{B}'\mathbf{B}\right\|\left\|\mathbf{X}'\mathbf{X}\right\|\left\|\mathbf{11}'\mathbf{11}'\right\|\left\|\mathbf{H}\right\|\right\} \\
&\le K^{1/2}n^2\left\|\mathbf{B}'\mathbf{B}\right\|\mathrm{tr}\left(\mathbf{\Sigma}_0\right)\left\{E\left(\left\|\mathbf{X}'\mathbf{X}\right\|^2\right)\right\}^{1/2}.
\end{aligned}$$

Note that $E\left(\|\mathbf{X}'\mathbf{X}\|^2\right) = nE\left(\|\mathbf{f}\|^4\right) + n\,(n-1)\left\|E\left(\mathbf{f}\mathbf{f}'\right)\right\|^2 = O(n^2K^2)$. Thus,

$$E\left\|\mathbf{B}\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}\mathbf{E}'\right\|^2 \le \left\|\mathbf{B}'\mathbf{B}\right\| \operatorname{tr}\left(\boldsymbol{\Sigma}_0\right) O(n^3K^{3/2}).$$

Therefore, by (A.1) we have

$$\mathcal{A}_1 \le 4E\left\|\mathbf{B}\mathbf{X}\mathbf{E}'\right\|^2 \le \left\|\mathbf{B}'\mathbf{B}\right\| \operatorname{tr}\left(\boldsymbol{\Sigma}_0\right) O(nK)$$

and

$$\mathcal{A}_2 \le 4E\left\|\mathbf{B}\mathbf{X}\mathbf{1}\mathbf{1}'\mathbf{H}\mathbf{E}'\right\|^2 \le \left\|\mathbf{B}'\mathbf{B}\right\| \operatorname{tr}\left(\boldsymbol{\Sigma}_0\right) O(n^3K^{3/2}),$$

which together yield (A.5) since clearly $\operatorname{tr}\left(\boldsymbol{\Sigma}_0\right) = O(p)$. $\square$

LEMMA 3. *Under conditions (A) and (B), we have*

$$E\ \operatorname{tr}\left\{\left[\mathbf{C}_n\widehat{\operatorname{cov}}(\mathbf{f})\mathbf{C}_n'\right]^2\right\} = O(n^{-2}p^2K) \tag{A.6}$$

and

$$E\ \operatorname{tr}\left(\mathbf{F}_n^2\right) = O(n^{-1}pK) + O(n^{-2}p^2K). \tag{A.7}$$

PROOF. The proofs of (A.6) and (A.7) are similar to those in Lemmas 4 and 5 below, respectively. For brevity, we omit them here. $\square$

LEMMA 4. *Under conditions (A)–(C), we have*

$$E\left\|\mathbf{C}_n\widehat{\operatorname{cov}}(\mathbf{f})\mathbf{C}_n'\right\|_{\boldsymbol{\Sigma}}^2 = O(n^{-2}pK). \tag{A.8}$$

PROOF. Note that

$$\begin{aligned} E\left\|\mathbf{C}_n\widehat{\operatorname{cov}}(\mathbf{f})\mathbf{C}_n'\right\|_{\boldsymbol{\Sigma}}^2 &\le 2\,(n-1)^{-2}\,E\left\|\mathbf{E}\mathbf{H}\mathbf{E}'\right\|_{\boldsymbol{\Sigma}}^2 + 2n^{-2}\,(n-1)^{-2}\,E\left\|\mathbf{E}\mathbf{H}\mathbf{1}\mathbf{1}'\mathbf{H}\mathbf{E}'\right\|_{\boldsymbol{\Sigma}}^2 \\ &\mathrel{\widehat{=}} 2\,(n-1)^{-2}\,\mathcal{K}_1 + 2n^{-2}\,(n-1)^{-2}\,\mathcal{K}_2. \end{aligned}$$

We will consider the above two terms $\mathcal{K}_1$ and $\mathcal{K}_2$ separately. First, we study the term $\mathcal{K}_1$, which can further be decomposed into four terms. Since $E(\boldsymbol{\varepsilon}|\mathbf{f}) = \mathbf{0}$, by conditioning on $\mathbf{X}$ we have

$$\begin{aligned}
\mathcal{K}_1 &= p^{-1} E \text{ tr} \left[E \left(\sum_{i,j=1}^{n} \mathbf{H}_{ij} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j' \boldsymbol{\Sigma}^{-1} \sum_{k,l=1}^{n} \mathbf{H}_{kl} \boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_l' \boldsymbol{\Sigma}^{-1} | \mathbf{X} \right) \right] \\
&= p^{-1} \mathcal{L}_1 + p^{-1} \mathcal{L}_2 + p^{-1} \mathcal{L}_3 + p^{-1} \mathcal{L}_4,
\end{aligned}$$

where

$$\mathcal{L}_1 = E \text{ tr} \left\{ \sum_{i=1}^{n} (\mathbf{H}_{ii})^2 E \left[\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' \boldsymbol{\Sigma}^{-1} \right)^2 \right] \right\}, \; \mathcal{L}_2 = E \text{ tr} \left\{ \sum_{i \neq j} \mathbf{H}_{ii} \mathbf{H}_{jj} E \left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j' \boldsymbol{\Sigma}^{-1} \right) \right\},$$

$$\mathcal{L}_3 = E \text{ tr} \left\{ \sum_{i \neq j} (\mathbf{H}_{ij})^2 E \left[\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j' \boldsymbol{\Sigma}^{-1} \right)^2 \right] \right\}, \; \mathcal{L}_4 = E \text{ tr} \left[\sum_{i \neq j} \mathbf{H}_{ij} \mathbf{H}_{ji} E \left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_i' \boldsymbol{\Sigma}^{-1} \right) \right],$$

and $\mathbf{H}_{ij}$ is the (i,j)-entry of the $n \times n$ hat matrix $\mathbf{H}$. Then we consider each of these four terms separately. By (1.3) and assumptions (C) and (B), it is easy to see that

$$\text{tr}\left(\boldsymbol{\Sigma}^{-1}\right) = O(p), \; \left\|\boldsymbol{\Sigma}^{-1}\right\| = O(p^{1/2}), \text{ and tr} \left[\left(\boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}^{-1} \right)^2 \right] = O(p). \tag{A.9}$$

It follows from (A.3) and (A.9) that

$$\begin{aligned}
\mathcal{L}_1 &\leq K \; E \left\{ \text{tr} \left[\left(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' \boldsymbol{\Sigma}^{-1} \right)^2 \right] \right\} = K \; E \left[\sum_{i,j=1}^{p} \left(\boldsymbol{\Sigma}^{-1} \right)_{ij} \varepsilon_i \varepsilon_j \sum_{k,l=1}^{p} \left(\boldsymbol{\Sigma}^{-1} \right)_{kl} \varepsilon_k \varepsilon_l \right] \\
&= K \sum_{i=1}^{p} \left(\boldsymbol{\Sigma}^{-1} \right)_{ii}^2 E\left(\varepsilon_i^4\right) + K \sum_{i \neq j} \left(\boldsymbol{\Sigma}^{-1} \right)_{ii} \left(\boldsymbol{\Sigma}^{-1} \right)_{jj} E\left(\varepsilon_i^2\right) E\left(\varepsilon_j^2\right) \\
&\quad + 2K \sum_{i \neq j} E\left(\varepsilon_i^2\right) \left(\boldsymbol{\Sigma}^{-1} \right)_{ij} E\left(\varepsilon_j^2\right) \left(\boldsymbol{\Sigma}^{-1} \right)_{ji} \\
&\leq \left[\text{tr}\left(\boldsymbol{\Sigma}^{-1} \right) \right]^2 O(K) + \text{tr}\left(\boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}^{-1} \right) O(K) = O(p^2 K)
\end{aligned}$$

and

$$\mathcal{L}_2 = E\left\{\sum_{i\neq j}\mathbf{H}_{ii}\mathbf{H}_{jj}\ \mathrm{tr}\left[E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)\boldsymbol{\Sigma}^{-1}E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)\boldsymbol{\Sigma}^{-1}\right]\right\}$$
$$\leq \mathrm{tr}\left[\left(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\right)^2\right]E\left\{\left[\mathrm{tr}\left(\mathbf{H}\right)\right]^2\right\} = O(pK^2).$$

Similarly, we have

$$\mathcal{L}_3 \leq K\ \mathrm{tr}\left\{E\left[\left(\boldsymbol{\varepsilon}\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1}\right)^2\right]\right\} = K\ E\left[\sum_{i,j=1}^{p}(\boldsymbol{\Sigma}^{-1})_{ij}\ \eta_i\varepsilon_j\sum_{k,l=1}^{p}(\boldsymbol{\Sigma}^{-1})_{kl}\ \eta_k\varepsilon_l\right]$$
$$= K\sum_{i=1}^{p}\left(\boldsymbol{\Sigma}^{-1}\right)_{ii}^2 E\left(\eta_i^2\right)E\left(\varepsilon_i^2\right) + 2K\sum_{i\neq j}E\left(\eta_i^2\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{ij}E\left(\varepsilon_j^2\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{ji}$$
$$\leq \left\|\boldsymbol{\Sigma}^{-1}\right\|^2 O(K) + \mathrm{tr}\left(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\right)O(K) = O(pK)$$

and

$$\mathcal{L}_4 \leq K\ \mathrm{tr}\left[E\left(\boldsymbol{\varepsilon}\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\right)\right] = K\left[E\left(\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}\right)\right]^2 = K\left[\sum_{i=1}^{p}(\boldsymbol{\Sigma}^{-1})_{ii}E(\varepsilon_i^2)\right]^2$$
$$= K\left[\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\right)O(1)\right]^2 = O(p^2K),$$

where $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_p)'$ is an independent copy of $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_p)'$. Since $K \leq p$ by assumption (A), combining $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_3$, and $\mathcal{L}_4$ together gives

$$\mathcal{K}_1 = E\left\|\mathbf{EHE}'\right\|_{\boldsymbol{\Sigma}}^2 = O(pK). \tag{A.10}$$

Now we consider the second term $\mathcal{K}_2$. By (A.4), the same calculation as above applies to show that

$$\mathcal{K}_2 = E\left\|\mathbf{EH11'HE}'\right\|_{\boldsymbol{\Sigma}}^2 = O(n^2pK).$$

Therefore, combining the above results together yields (A.8). □

LEMMA 5. *Under conditions (A)–(C), we have*

$$E\|\mathbf{F}_n\|_{\boldsymbol{\Sigma}}^2 = O(n^{-1}) + O(n^{-2}pK). \tag{A.11}$$

PROOF. Note that

$$E\|\mathbf{F}_n\|_{\mathbf{\Sigma}}^2 \leq 2E\left\|I_p \circ n^{-1}\mathbf{E}\mathbf{E}' - \mathbf{\Sigma}_0\right\|_{\mathbf{\Sigma}}^2 + 2n^{-2}E\left\|I_p \circ \mathbf{E}\mathbf{H}\mathbf{E}'\right\|_{\mathbf{\Sigma}}^2.$$

Since $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\operatorname{cov}(\boldsymbol{\varepsilon}|\mathbf{f}) = \mathbf{\Sigma}_0$, we have

$$\begin{aligned}
E\big\|I_p \circ n^{-1}\mathbf{E}\mathbf{E}' - \mathbf{\Sigma}_0\big\|_{\mathbf{\Sigma}}^2 &= p^{-1}E\left\|n^{-1}\mathbf{\Sigma}^{-1/2}\left(I_p \circ \mathbf{E}\mathbf{E}'\right)\mathbf{\Sigma}^{-1/2} - \mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1/2}\right\|_{\mathbf{\Sigma}}^2 \\
&= p^{-1}n^{-1}\left[E\left\|\mathbf{\Sigma}^{-1/2}\operatorname{diag}\left(\varepsilon_1^2, \cdots, \varepsilon_p^2\right)\mathbf{\Sigma}^{-1/2}\right\|^2 - \left\|\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1/2}\right\|^2\right] \\
&\leq p^{-1}n^{-1}E\ \operatorname{tr}\left\{\left[\mathbf{\Sigma}^{-1/2}\operatorname{diag}\left(\varepsilon_1^2, \cdots, \varepsilon_p^2\right)\mathbf{\Sigma}^{-1/2}\right]^2\right\} \mathrel{\widehat{=}} p^{-1}n^{-1}\mathcal{L}.
\end{aligned}$$

It follows from (A.9) that

$$\begin{aligned}
\mathcal{L} &= \sum_{i,j=1}^{p} E\left[\varepsilon_i^2\left(\mathbf{\Sigma}^{-1}\right)_{ij}\varepsilon_j^2\left(\mathbf{\Sigma}^{-1}\right)_{ji}\right] = \sum_{i=1}^{p}\left(\mathbf{\Sigma}^{-1}\right)_{ii}^2 E\left(\varepsilon^4\right) + \sum_{i\neq j}\left(\mathbf{\Sigma}^{-1}\right)_{ij}^2\left[E\left(\varepsilon^2\right)\right]^2 \\
&= \left\|\mathbf{\Sigma}^{-1}\right\|^2 O(1) = O(p),
\end{aligned}$$

which shows that $E\big\|I_p \circ n^{-1}\mathbf{E}\mathbf{E}' - \mathbf{\Sigma}_0\big\|_{\mathbf{\Sigma}}^2 = O(n^{-1})$. The argument proving (A.10) in Lemma 4 applies to show that

$$E\left\|I_p \circ \mathbf{E}\mathbf{H}\mathbf{E}'\right\|_{\mathbf{\Sigma}}^2 = O(pK).$$

Hence, combining the above results together gives (A.11). $\square$

LEMMA 6. *Under conditions (A)–(C), we have*

$$E\|\mathbf{G}_n\|_{\mathbf{\Sigma}}^2 = O(n^{-1}p). \tag{A.12}$$

PROOF. Recall that $\mathbf{G}_n \mathrel{\widehat{=}} \left\{(n-1)^{-1}\mathbf{E}\mathbf{E}' - [n(n-1)]^{-1}\mathbf{E}\mathbf{1}\mathbf{1}'\mathbf{E}'\right\} - \mathbf{\Sigma}_0$, as defined in part (2) of the proof of Theorem 1. Note that

$$\begin{aligned}
E\|\mathbf{G}_n\|_{\mathbf{\Sigma}}^2 \leq 3E\left\|n^{-1}\mathbf{E}\mathbf{E}' - \mathbf{\Sigma}_0\right\|_{\mathbf{\Sigma}}^2 &+ 3n^{-2}(n-1)^{-2}E\left\|\mathbf{E}\mathbf{E}'\right\|_{\mathbf{\Sigma}}^2 \\
&+ 3n^{-2}(n-1)^{-2}E\left\|\mathbf{E}\mathbf{1}\mathbf{1}'\mathbf{E}'\right\|_{\mathbf{\Sigma}}^2.
\end{aligned}$$

From the proofs of $\mathcal{L}_1$ and $\mathcal{L}_4$ in Lemma 4, we know that

$$E\left\{\operatorname{tr}\left[\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\right)^2\right]\right\} = O(p^2) \quad \text{and} \quad E\left[\operatorname{tr}\left(\boldsymbol{\varepsilon}\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\right)\right] = O(p^2),$$

where $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_p)'$ is an independent copy of $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_p)'$. Thus, we have

$$\begin{aligned}
E\left\|n^{-1}\mathbf{E}\mathbf{E}' - \boldsymbol{\Sigma}_0\right\|_{\boldsymbol{\Sigma}}^2 &= p^{-1}E\left\|n^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{E}\mathbf{E}'\boldsymbol{\Sigma}^{-1/2} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1/2}\right\|^2 \\
&\le p^{-1}n^{-1}E\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1/2}\right\|^2 = p^{-1}n^{-1}E\left\{\operatorname{tr}\left[\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\right)^2\right]\right\} = O(n^{-1}p).
\end{aligned}$$

Similarly, it follows that

$$\begin{aligned}
E\left\|\mathbf{E}\mathbf{E}'\right\|_{\boldsymbol{\Sigma}}^2 &= E\left\|\left(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n\right)\left(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n\right)'\right\|_{\boldsymbol{\Sigma}}^2 \le np^{-1}E\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1/2}\right\|^2 \\
&= np^{-1}E\left\{\operatorname{tr}\left[\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\right)^2\right]\right\} = O(np)
\end{aligned}$$

and

$$\begin{aligned}
E\left\|\mathbf{E}\mathbf{1}\mathbf{1}'\mathbf{E}'\right\|_{\boldsymbol{\Sigma}}^2 &\le np^{-1}E\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1/2}\right\|^2 + n\left(n-1\right)p^{-1}E\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\varepsilon}\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1/2}\right\|^2 \\
&\le np^{-1}E\left\{\operatorname{tr}\left[\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\right)^2\right]\right\} + n\left(n-1\right)p^{-1}E\left[\operatorname{tr}\left(\boldsymbol{\varepsilon}\boldsymbol{\eta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\right)\right] = O(n^2p).
\end{aligned}$$

Therefore, combining the above results together proves (A.12). $\square$

REFERENCES

Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics* **18** 338–357.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135–171.

Boik, R. J. (2002). Spectral models for covariance matrices. *Biometrika* **89** 159–182.

Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika* **74** 375–384.

Browne, M. W. and Shapiro, A. (1987). Adjustments for kurtosis in factor analysis with elliptically distributed errors. *J. Roy. Statist. Soc. Ser. B* **49** 346–352.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. G. (1997). *The Econometrics of Financial Markets*. Princeton University Press, New Jersey.

Chamberlain, G. (1983). Funds, factors and diversification in Arbitrage Pricing Theory. *Econometrica* **51** 1305–1323.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51** 1281–1304.

Chiu, T.Y.M., Leonard, T. and Tsui, K.W. (1996). The matrix-logarithm covariance model. *J. Amer. Statist. Assoc.* **91** 198–210.

Cochrane, J. H. (2001). *Asset Pricing*. Princeton University Press, New Jersey.

Dempster, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175.

Diebold, F. X. and Nerlove, M. (1989). The dynamics of exchange rate volatility: a multivariate latent-factor ARCH model. *Journal of Applied Econometrics* **4** 1–22.

Diggle, P. J. and Verbyla, A. P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* **54** 401–415.

Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. Aide-Memoire of a Lecture at AMS Conference on Math

Challenges of the 21st Century.

Eaton, M. L. and Tyler, D. E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Ann. Statist.* **19** 260–271.

Eaton, M. L. and Tyler, D. E. (1994). The asymptotic distribution of singular values with applications to canonical correlations and corresponding analysis. *J. Multiv. Anal.* **50** 238–264.

Engle, R. F. and Watson, M. W. (1981). A one-factor multivariate time series model of metropolitan wage rates. *J. Amer. Statist. Assoc.* **76** 774–781.

Fama, E. and French, K. (1992). The cross-section of expected stock returns. *Jour. Fin.* **47** 427–465.

Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Jour. Fin. Econ.* **33** 3–56.

Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics (with discussion). *Statist. Science*, to appear.

Fan, J. and Li, R. (2006). Statistical challenges with high dimensionality: feature selection in knowledge discovery. *Proceedings of the Madrid International Congress of Mathematicians 2006*, to apepar.

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32** 928–961.

Goldfarb, D. and Iyengar, G. (2003). Robust portfolio selection problems. *Math. Oper. Res.* **28** 1–38.

HORN, R. A. and JOHNSON, C. R. (1990). *Matrix Analysis.* Cambridge University Press, Cambridge.

HUANG, J. Z., LIU, N. and POURAHMADI, M. (2004). Covariance selection and esimation via penalized normal likelihood. *Manuscript.*

HUBER, P. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic Loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California Press, Berkeley.

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327.

LEDOIT, O. and WOLF, M. (2004). A well conditioned estimator for large–dimensional covariance matrices. *J. Multiv. Anal.* **88** 365–411.

LEONARD, T. and HSU, J.S.J. (1992). Bayesian inference for a covariance matrix. *Ann. Statist.* **20** 1669–1696.

LI, H. and GUI, J. (2005). Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks. *Manuscript.*

LIN, S. P. and PERLMAN, M. D. (1985). A Monte Carlo comparison of four estimators of a covariance matrix. In *Multivariate Analysis* **6**, Ed. P. R. Krishnaiah, 411–429. Amsterdam, North-Holland.

MARKOWITZ, H. M. (1952). Portfolio selection. *Journal of Finance* **7** 77–91.

Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments.* John Wiley & Sons, New Jersey.

Muirhead, Robb J. (1982). *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, New York.

Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *Ann. Statist.* **12** 1298–1309.

Portnoy, S. (1985). Asymptotic behavior of M estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika* **87** 425–435.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory* **13** 341–360.

——— (1977). The Capital Asset Pricing Model (CAPM), short-sale restrictions and related issues. *Journal of Finance* **32** 177–183.

Scott, J. T. (1966). Factor analysis and regression. *Econometrica* **34** 552–562.

Scott, J. T. (1969). Factor analysis regression revisited. *Econometrica* **37** 719.

Smith, M. and Kohn, R. (2002). Parsimonious covariance matrix estimation for longitudinal data. *J. Amer. Statist. Assoc.* **97** 1141–1153.

Stein, C. (1975). Estimation of a covariance matrix. *Rietz Lecture*, 39th IMS Annual Meeting, Atlanta, Georgia.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for VAR analysis. *Manuscript.*

WONG, F., CARTER, C. K. and KOHN, R. (2003). Efficient estimation of covariance selection models. *Biometrika* **90** 809–830.

WU, W. B. and POURAHMADI, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90** 831–844.

YOHAI, V. J. and MARONNA, R. A. (1979). Asymptotic behavior of M-estimators for the linear model. *Ann. Statist.* **7** 258–268.

YUAN, K.-H. and BENTLER, P. M. (1997). Mean and covariance structure analysis: theoretical and practical improvements. *J. Amer. Statist. Assoc.* **438** 767–774.

JIANQING FAN
DEPARTMENT OF OPERATIONS RESEARCH
AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
USA
E-MAIL: jqfan@princeton.edu

YINGYING FAN
DEPARTMENT OF OPERATIONS RESEARCH
AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY
Princeton, New Jersey 08544
USA
E-MAIL: yingying@princeton.edu

JINCHI LV
DEPARTMENT OF MATHEMATICS
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
USA
E-MAIL: jlv@princeton.edu