Journal of Econometrics 186 (2015) 367–387

Contents lists available at ScienceDirect

Journal of Econometrics

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

Risks of large portfolios

Jianqing Fan [a,b,c,*], Yuan Liao [d], Xiaofeng Shi [a]

[a] Department of Operations Research and Financial Engineering, Princeton University, United States
[b] International School of Economics and Management, Capital Univeristy of Economics and Business, Beijing, China
[c] Bendheim Center for Finance, Princeton University, United States
[d] Department of Mathematics, University of Maryland, United States

ARTICLE INFO

Article history:
Available online 7 March 2015

JEL classification:
C58
C38

Keywords:
High dimensionality
Factor models
Principal components
Sparse matrix
Volatility

ABSTRACT

The risk of a large portfolio is often estimated by substituting a good estimator of the volatility matrix. However, the accuracy of such a risk estimator is largely unknown. We study factor-based risk estimators under a large amount of assets, and introduce a high-confidence level upper bound (H-CLUB) to assess the estimation. The H-CLUB is constructed using the confidence interval of risk estimators with either known or unknown factors. We derive the limiting distribution of the estimated risks in high dimensionality. We find that when the dimension is large, the factor-based risk estimators have the same asymptotic variance no matter whether the factors are known or not, which is slightly smaller than that of the sample covariance-based estimator. Numerically, H-CLUB outperforms the traditional crude bounds, and provides an insightful risk assessment. In addition, our simulated results quantify the relative error in the risk estimation, which is usually negligible using 3-month daily data.

1. Introduction

The potential of a portfolio's loss is termed as the portfolio risk. There are two types of portfolio risks. The *systematic risk* is the risk inherent to the entire market, such as risk associated with interest rates, currencies, recession, war and political instability, etc. The systematic risk cannot be diversified away, even with a well-diversified portfolio. In contrast, *specific risk* (or *idiosyncratic risk*) refers to the risk that affects a very specific group of securities or even an individual security. For example, it can be the risk of price changes due to the unique circumstances of a specific stock. Unlike systematic risk, specific risk can be reduced through diversification.

Estimating and assessing the risk of a large portfolio is an important topic in financial econometrics and risk management. The risk of a given portfolio allocation vector $\mathbf{w}_T$ is conveniently measured by $(\mathbf{w}_T'\boldsymbol{\Sigma}\mathbf{w}_T)^{1/2}$, in which $\boldsymbol{\Sigma}$ is a volatility (covariance) matrix of the assets' returns. Often multiple portfolio risks are at interests and hence it is essential to estimate the volatility matrix $\boldsymbol{\Sigma}$. The problem becomes challenging when the portfolio size is large. Suppose we have created a portfolio from two thousand assets and invested in a part of selected assets. The covariance matrix $\boldsymbol{\Sigma}$ involved then contains over two million unknown parameters. Yet, the sample size based on one year's daily data is around 252. It is hard to assess the estimation accuracy when the estimation errors from more than two million parameters are aggregated. Hence some regularization method is recommended to estimate and assess risks.

We estimate and assess the risks of a given portfolio vector $\mathbf{w}_T$ based on factor models. Two factor-based methods are compared, previously proposed by Fan et al. (2011, 2013). The first estimator assumes the factors to be known and observable. The second method deals with the case of unknown factors. In both cases, the factor model imposes a *conditionally sparse* structure, in that the idiosyncratic covariance is a large sparse matrix. This yields to an *approximate factor model* as in Chamberlain and Rothschild (1983), with a non-diagonal error covariance matrix.

We provide a new and practical method to assess the accuracy of risk estimation $\mathbf{w}_T'(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\mathbf{w}_T$. In the literature (e.g. Fan et al., 2012), this term has been bounded by

$$\widehat{\xi}_T=\|\mathbf{w}_T\|_1^2\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_{\max}$$

where $\|\mathbf{w}_T\|_1$ is the gross exposure of the portfolio, and is bounded when there are no extreme positions in the portfolio. However, this upper bound depends on the unknown $\boldsymbol{\Sigma}$, hence is not applicable in practice. In addition, numerical studies in this paper demonstrate

* Correspondence to: Department of Operations Research and Financial Engineering, Sherrerd Hall, Princeton University, Princeton, NJ 08544, United States. Tel.: +1 609 258 3255.
E-mail address: jqfan@princeton.edu (J. Fan).

http://dx.doi.org/10.1016/j.jeconom.2015.02.015
0304-4076/

that this upper bound is too crude: it is often of the same or even larger scale than the estimated risk. In contrast, we provide a high-confidence level upper bound (H-CLUB) for $\mathbf{w}_T'(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\mathbf{w}_T$, which is of much smaller scale and easy to compute in practice. H-CLUB is constructed based on the confidence interval for the true risk. For the risk estimator $\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T$ and a given $\epsilon \in (0, 1)$, we find an H-CLUB $\widehat{U}(\epsilon)$ such that

$$P(|\mathbf{w}_T'(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\mathbf{w}_T| \le \widehat{U}(\epsilon)) \to 1 - \epsilon.$$

In contrast, $P(|\mathbf{w}_T'(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\mathbf{w}_T| \le \widehat{\xi}_T) = 1$. Hence H-CLUB is an upper bound for the risk estimation error with high confidence while the traditional bound $\widehat{\xi}_T$ is of full confidence.

For the inferential theory of the risk estimators with diversified portfolios, we prove that the effects of estimating the factor loadings and unknown factors are asymptotically negligible. Interestingly, it is found that when the dimensionality is larger than the sample size, the factor-based risk estimators have the same asymptotic variances no matter whether the factors are known or not. Hence the high dimensionality is in fact a bless for risk estimation instead of a curse from this point of view. In addition, the asymptotic variance of factor-based estimators is slightly smaller than that of the sample covariance-based estimator, but the difference is small. This demonstrates that the benefit of using a factor model is not in terms of a much smaller asymptotic variance, because the systematic risk cannot be diversified. Rather, factor models give a strictly positive definite covariance estimator, which is essential to estimate the optimal portfolio allocation vector, and also interprets the structure of the portfolio risks.

Using our simulated results based on the model calibrated from the US equity market data, we are able to quantify the relative error of the estimation error or coefficient of variation, defined as $\mathrm{STD}(\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T)/\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T$, where $\mathrm{STD}(\cdot)$ denotes the standard error of the estimated risk. Interestingly, this ratio is just a few percent and is approximately independent of the gross exposure $\|\mathbf{w}_T\|_1$ but sensitive to the length of the time series. On the other hand, we also quantify the relation between the crude bound and the practical H-CLUB. We find that $\widehat{\xi}_T$ is many times larger than $\widehat{U}(\epsilon)$, and the ratio $\widehat{\xi}_T/\widehat{U}(\epsilon)$ increases as the gross exposure increases. A sampling technique that picks a random portfolio with a given gross exposure level is introduced, which can be useful for portfolio optimization and understanding the overall risks within a given level of gross exposure.

The interest on large portfolios surges recently. Pesaran and Zaffaroni (2008) examined the asymptotic behavior of the portfolio weights. Brodie et al. (2009) and Fan et al. (2012) addressed the problem of portfolio selection using a regularization penalty. Gomez and Gallon (2011) numerically compared several methods of covariance matrix estimation for portfolio management. In particular, the optimal portfolio selection involves inverting an estimated $\mathbf{\Sigma}$, which is a challenging problem under a large number of assets. Gagliardini et al. (2010) considered a random coefficient model for an unbalanced panel, and focused on the observable factors, while we also study the inferential theory of the unobservable factor case. The recent works by Fan et al. (2011, 2013) are only concerned about covariance estimations and no inferential theories were studied. The literature is also found in Jacquier and Polson (2010), Antoine (2011), Chang and Tsay (2010), DeMiguel et al. (2009a,b), Ledoit and Wolf (2003), El Karoui (2010), Lai et al. (2011), Bannouh et al. (2012), Gandy and Veraart (2012), Bianchi and Carvalho (2011), among others.

The rest of the paper is organized as follows. Section 2 introduces risk estimators based on factor models under both known and unknown factors. Section 3 constructs the H-CLUB for each risk estimator based on the confidence interval for risks. Section 4 derives the limiting distributions of the risk estimators and compares their asymptotic variances. Section 5 presents simulation results. An empirical study is considered in Section 6. Finally, Section 7 concludes. All the proofs are given in the Appendix.

Throughout the paper, $\|\mathbf{w}_T\|_1 = \sum_{i=1}^N |w_i|$ is used to denote the gross exposure of a given portfolio allocation vector. For a square matrix $\mathbf{A}$, $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ represent its minimum and maximum eigenvalues; $\|\mathbf{A}\|_1 = \max_i \sum_j |A_{ij}|$. Let $\|\mathbf{A}\|_{\max}$ and $\|\mathbf{A}\|$ denote its element-wise sup-norm and operator norm, given by $\|\mathbf{A}\|_{\max} = \max_{i,j} |A_{ij}|$ and $\|\mathbf{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$ respectively.

2. Estimation of portfolio's risks

Let $\{\mathbf{R}_t\}_{t=1}^T$ be a strictly stationary time series of an $N \times 1$ vector of observed excess returns and $\mathbf{\Sigma} = \mathrm{cov}(\mathbf{R}_t)$, often known as the volatility matrix. The portfolio risk of a given allocation vector $\mathbf{w}_T$ is given by $\sqrt{\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T}$. With a covariance estimator $\widehat{\mathbf{\Sigma}}$, a straightforward estimator of the portfolio risk is $\sqrt{\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T}$. But how good such a substitution estimator is and how to assess its estimation accuracy when the dimension N is large relative to T are the questions addressed here.

The problem of estimating the risk of a given portfolio is challenging due to the high dimensionality of $\mathbf{\Sigma}$. In most cases N can be much larger than T. We assume $\mathbf{\Sigma}$ to be time-invariant within a short period, which holds approximately for locally stationary time series. Recently, Chang and Tsay (2010) proposed a Cholesky decomposition approach to estimating the large covariance matrix, and used simulation to assess its performance. A natural alternative approach is through the factor model (e.g. Stock and Watson, 2002; Bai, 2003), because the assets' returns are usually driven by a few market factors. Estimating $\mathbf{\Sigma}$ is possible when both the factor and the idiosyncratic components can be estimated well. We thus consider three estimators of $\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T$ for a given $\mathbf{w}_T$, based on three different estimators $\widehat{\mathbf{\Sigma}}$: sample covariance estimator, and factor model estimators with either observed or unobserved factors.

2.1. Sample-covariance-based estimator

The first estimator $\widehat{\mathbf{\Sigma}} = \mathbf{S}$ is the conventional sample covariance matrix based on $\{\mathbf{R}_t\}_{t=1}^T$. Because we are mainly concerned about the variance, for simplicity and exposition, let us assume that the returns have mean zero and $\mathbf{S} = T^{-1}\sum_{t=1}^T \mathbf{R}_t\mathbf{R}_t'$. The asymptotic impact of using $\mathbf{S}$ on the risk management has been studied by Fan et al. (2008, 2012) when N is much larger than T. The sample covariance estimator does not require any structural assumption on the assets' returns. It was shown by the aforementioned authors that for a given portfolio $\mathbf{w}_T$ with a bounded gross exposure (that is, $\|\mathbf{w}_T\|_1$ is bounded),

$$\mathbf{w}_T'(\mathbf{S} - \mathbf{\Sigma})\mathbf{w}_T \le \|\mathbf{w}_T\|_1^2\|\mathbf{S} - \mathbf{\Sigma}\|_{\max} = O_p\left(\sqrt{\frac{\log N}{T}}\right).$$

However, when $N > T$, it is well known that $\mathbf{S}$ is singular, and therefore may result in an estimated risk close to zero for certain portfolios.

2.2. Estimating risks based on factor models

We assume the true data generating process (DGP) of $\mathbf{R}_t$ to be an "approximate factor model" (Chamberlain and Rothschild, 1983):

$$\mathbf{R}_t = \mathbf{B}\mathbf{f}_t + \mathbf{u}_t, \quad t \le T, \tag{2.1}$$

where $\mathbf{B}$ is an $N \times K$ matrix of factor loadings; $\mathbf{f}_t$ is a $K \times 1$ vector of common factors, and $\mathbf{u}_t$ is an $N \times 1$ vector of idiosyncratic error components. In contrast to N and T, here K is assumed to be fixed. The common factors may or may not be observable. For example, Fama and French (1992) identified three known factors that have successfully described the US stock market. On the other hand, in an empirical study, Bai and Ng (2002) determined two unobservable factors for stocks traded on the New York Stock Exchange during 1994–1998.

Let $\mathrm{cov}(\mathbf{f}_t)$ and $\mathbf{\Sigma}_u = \mathrm{cov}(\mathbf{u}_t)$ denote the covariance matrices of $\mathbf{f}_t$ and $\mathbf{u}_t$, $K \times K$ and $N \times N$ respectively. Suppose $\mathbf{f}_t$ and $\mathbf{u}_t$ are uncorrelated. The factor model then implies the following decomposition of $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \mathbf{B}\,\mathrm{cov}(\mathbf{f}_t)\mathbf{B}' + \mathbf{\Sigma}_u. \tag{2.2}$$

We shall assume that $\mathbf{\Sigma}_u$ is a sparse matrix in the sense that many off-diagonal elements of the covariance are either zero or nearly so. The rationale of the sparsity here is, after the common factors are taken out, the remaining idiosyncratic components should be mostly weakly correlated. The decomposition (2.2) also implies that the first K eigenvalues of $\mathbf{\Sigma}$ grow at the rate $O(N)$, due to those of $\mathbf{B}\,\mathrm{cov}(\mathbf{f}_t)\mathbf{B}'$, while the remaining eigenvalues are bounded away from both zero and infinity. See, e.g., Onatski (2010) and Ahn and Horenstein (2013).

2.2.1. Known-factor-based estimator

We first consider the case of observable factors, and construct an estimator of $\mathbf{\Sigma}$ based on thresholding the covariance matrix of idiosyncratic errors. Suppose $\widehat{\mathbf{B}}$ is the least squares estimator of $\mathbf{B}$. The residual sample covariance matrix of $\mathbf{u}_t$ is then given by

$$\mathbf{S}_u = T^{-1}\sum_{t=1}^{T}(\widehat{\mathbf{u}}_t - \bar{\mathbf{u}})(\widehat{\mathbf{u}}_t - \bar{\mathbf{u}})' = (S_{u,ij})_{N\times N}, \quad \widehat{\mathbf{u}}_t = \mathbf{R}_t - \widehat{\mathbf{B}}\mathbf{f}_t,$$
$$\bar{\mathbf{u}} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t.$$

Let $s_{ij}(\cdot) : \mathbb{R} \to \mathbb{R}$ be an entry-dependent adaptive thresholding function and for some thresholding parameter $\tau_{ij}^f > 0$,

$$s_{ij}(z) = 0 \quad \text{when } |z| \le \tau_{ij}^f, \quad \text{and} \quad |s_{ij}(z) - z| \le \tau_{ij}^f. \tag{2.3}$$

The thresholding parameter is taken to be, for some positive constant $C > 0$,

$$\tau_{ij}^f = C\sqrt{S_{u,ii}S_{u,jj}}\sqrt{\frac{\log N}{T}}.$$

A simple example is the hard-thresholding $s_{ij}(z) = zI(|z| \ge \tau_{ij}^f)$, namely, setting all correlation coefficients smaller than $C\sqrt{\log N/T}$ to zero. The soft-thresholding rule is given by $s_{ij}(z) = (z - \tau_{ij}^f)_+$. See Antoniadis and Fan (2001), Rothman et al. (2009) and Cai and Liu (2011) for detailed discussions of various thresholding functions.

Let

$$\widehat{\Sigma}_{u,ij} = \begin{cases} S_{u,ii}, & i = j \\ s_{ij}(S_{u,ij}), & i \ne j. \end{cases}$$

Let $\widehat{\mathrm{cov}}(\mathbf{f}_t)$ denote the sample covariance of the common factors. Define the estimated covariance matrix as

$$\widehat{\mathbf{\Sigma}}_f = \widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}' + \widehat{\mathbf{\Sigma}}_u, \qquad \widehat{\mathbf{\Sigma}}_u = (\widehat{\Sigma}_{u,ij})_{N\times N}. \tag{2.4}$$

2.2.2. Covariance matrix estimator with unknown factors

When the common factors are unobservable, we estimate $\mathbf{\Sigma}$ by "principal orthogonal complements thresholding" (POET), recently proposed by Fan et al. (2013). Because K, the number of factors, might also be unknown, this estimator uses a data-driven number of factors $\widehat{K}$.

The POET works as follows: let $\widehat{\lambda}_1 \ge \cdots \ge \widehat{\lambda}_N$ be the ordered eigenvalues of the sample covariance $\mathbf{S}$, whose corresponding eigenvectors are denoted by $\{\widehat{\boldsymbol{\xi}}_j\}_{j=1}^N$. We then estimate $\mathbf{\Sigma}$ by

$$\widehat{\mathbf{\Sigma}}_{P,\widehat{K}} = \sum_{j=1}^{\widehat{K}}\widehat{\lambda}_j\widehat{\boldsymbol{\xi}}_j\widehat{\boldsymbol{\xi}}_j' + \widehat{\mathbf{\Omega}}, \qquad \widehat{\mathbf{\Omega}} = (\widehat{\Omega}_{ij})_{N\times N},$$
$$\widehat{\Omega}_{ij} = \begin{cases} \displaystyle\sum_{k=\widehat{K}+1}^{N}\widehat{\lambda}_k\widehat{\xi}_{k,i}^2, & i = j \\ s_{ij}\left(\displaystyle\sum_{k=\widehat{K}+1}^{N}\widehat{\lambda}_k\widehat{\xi}_{k,i}\widehat{\xi}_{k,j}\right), & i \ne j \end{cases}$$

where $s_{ij}(\cdot)$ is the same adaptive thresholding function as before, based on an entry-dependent threshold τ_{ij}^P:

$$\tau_{ij}^P = C\sqrt{\widehat{\Omega}_{ii}\widehat{\Omega}_{jj}}\left(\sqrt{\frac{\log N}{T}} + \frac{1}{\sqrt{N}}\right).$$

Here C is a user-specified constant to maintain the finite sample positive definiteness. Even when $N/T \to \infty$, there is $C^* > 0$ such that for any $C > C^*$, both $\widehat{\mathbf{\Sigma}}_f$ and $\widehat{\mathbf{\Sigma}}_{P,\widehat{K}}$ are strictly positive definite for a given finite sample. Previous simulated and empirical studies suggested that $C = 0.5$ is a good choice when s_{ij} is the soft thresholding (see Fan et al., 2013; Fryzlewicz, 2012).

The number of factors can be estimated by the information criterion as in Bai and Ng (2002):

$$\widehat{K} = \arg\min_{0\le k\le M}\frac{1}{N}\mathrm{tr}\left(\sum_{j=k+1}^{N}\widehat{\lambda}_j\widehat{\boldsymbol{\xi}}_j\widehat{\boldsymbol{\xi}}_j'\right) + \frac{k(N+T)}{NT}\log\left(\frac{NT}{N+T}\right), \tag{2.5}$$

where M is a prescribed upper bound. The accuracy of the risk estimation $\mathbf{w}_T'(\widehat{\mathbf{\Sigma}}_{P,\widehat{K}} - \mathbf{\Sigma})\mathbf{w}_T$ is robust to over-estimating K, as the variance introduced by unnecessary factors are usually small. This is verified in our simulated studies.

Based on the factor model, our proposed risk estimator is either $\sqrt{\mathbf{w}_T'\widehat{\mathbf{\Sigma}}_f\mathbf{w}_T}$ or $\sqrt{\mathbf{w}_T'\widehat{\mathbf{\Sigma}}_{P,\widehat{K}}\mathbf{w}_T}$ for a given portfolio allocation vector $\mathbf{w}_T$, depending on whether $\mathbf{f}_t$ is observable. This paper focuses on the limiting distributions of these risk estimators and their assessment for a given diversified $\mathbf{w}_T$. We will see that under high dimensionality, the factor-based estimators have the same asymptotic variance, and is smaller than that of the sample covariance-based estimator.

3. Assessment of the risk estimation

This section proposes a new method to assess the estimated risks for a given portfolio allocation vector $\mathbf{w}_T$. We will assume $\|\mathbf{w}_T\|_1 \le c$ for some $c \ge 1$, where $\|\mathbf{w}_T\|_1$ is the gross exposure of the portfolio. This prevents extreme positions. For simplicity, we shall also assume the number of factors to be fixed, which is easy to be relaxed to allow for slowly-growing K.

3.1. Data-dependent portfolio vector

One of the simplest examples of $\mathbf{w}_T$ is the equally weighted $(1/N)$ portfolios. DeMiguel et al. (2009b) empirically showed how such simple strategies can outperform more sophisticated strategies. There are also many methods proposed to choose data-dependent portfolios when the number of assets becomes large. Often one can estimate $\mathbf{w}_T$ by solving constrained data-dependent optimization problems, or directly estimate the unknown parameters when the ideal portfolio $\mathbf{w}_T$ depends on some unknown quantities (e.g., Brandt et al. (2009), El Karoui (2010), Yen (2013), Lai et al. (2011)).

While this paper focuses on assessing the risk of a given portfolio vector $\mathbf{w}_T$, we allow to use an estimated vector $\widehat{\mathbf{w}}_T$, which consistently estimates $\mathbf{w}_T$ in the L_1-norm:

$$\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1 = o_p(1). \tag{3.1}$$

The L_1-convergence is the right notion for consistency in the current context, because the portfolio vector itself has finite gross exposure $\|\mathbf{w}_T\|_1$. This paper will study the statistical inference about $\widehat{\mathbf{w}}_T'\boldsymbol{\Sigma}\widehat{\mathbf{w}}_T$, which is the true risk of such types of data-dependent portfolios and strategies.

Example 3.1. The *global minimum variance* portfolio is the solution to the problem:

$$\mathbf{w}_T^{gmv} = \arg\min_{\mathbf{w}}(\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}), \quad \text{such that } \mathbf{w}'\mathbf{e} = 1$$

where $\mathbf{e} = (1, \ldots, 1)$, yielding $\mathbf{w}_T^{gmv} = \boldsymbol{\Sigma}^{-1}\mathbf{e}/(\mathbf{e}'\boldsymbol{\Sigma}^{-1}\mathbf{e})$. Although this portfolio does not belong to the efficient frontier, Jagannathan and Ma (2003) showed that its performance is comparable with those of other tangency portfolios. Suppose $\|\boldsymbol{\Sigma}_u^{-1}\|_1 = O(1)$ and $\mathbf{e}'\boldsymbol{\Sigma}^{-1}\mathbf{e} \geq CN$ for some $C > 0$ and all large N (in the factor model, a sufficient condition is $\|\mathbf{e}'\boldsymbol{\Sigma}_u^{-1}\mathbf{B}\| = o(N)$). Then $\|\boldsymbol{\Sigma}^{-1}\|_1 = O(1)$ and $\|\mathbf{w}_T^{gmv}\|_1 \leq (CN)^{-1}\|\boldsymbol{\Sigma}^{-1}\|_1 N = O(1)$, so $\mathbf{w}_T^{gmv}$ has a finite gross exposure.

When $N > T$, we can estimate $\mathbf{w}_T^{gmv}$ using a factor model. This yields a data-dependent portfolio:

$$\widehat{\mathbf{w}}_T^{gmv} = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{e}}{\mathbf{e}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{e}}, \qquad \widehat{\boldsymbol{\Sigma}}^{-1} = \begin{cases} \widehat{\boldsymbol{\Sigma}}_f^{-1} & \text{known factors;} \\ \widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}^{-1}, & \text{unknown factors.} \end{cases}$$

To investigate the consistency of $\widehat{\mathbf{w}}_T^{gmv}$, note that under the condition that the factors are *pervasive* (Assumption 4.3(ii) below), the factor-based inverse covariance estimators of Fan et al. (2011, 2013) are L_1-consistent, that is,

$$\|\widehat{\boldsymbol{\Sigma}}_f^{-1} - \boldsymbol{\Sigma}^{-1}\|_1 = o_p(1), \qquad \|\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}^{-1} - \boldsymbol{\Sigma}^{-1}\|_1 = o_p(1).$$

This implies, for $\widehat{\boldsymbol{\Sigma}}^{-1} = \widehat{\boldsymbol{\Sigma}}_f^{-1}$ or $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}^{-1}$,

$$\begin{aligned}\|\widehat{\mathbf{w}}_T^{gmv} - \mathbf{w}_T^{gmv}\|_1 \leq{}& \frac{1}{\mathbf{e}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{e}}\|(\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1})\mathbf{e}\|_1 \\ &+ \frac{|\mathbf{e}'(\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{e})|}{(\mathbf{e}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{e})(\mathbf{e}'\boldsymbol{\Sigma}^{-1}\mathbf{e})}\|\boldsymbol{\Sigma}^{-1}\mathbf{e}\|_1.\end{aligned}$$

Note that there is $C > 0$ so that $\mathbf{e}'\boldsymbol{\Sigma}^{-1}\mathbf{e} \geq CN$ and $\mathbf{e}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{e} \geq CN$ with probability approaching one. The first term on the right-hand-side is bounded by $O_p(N^{-1})\|\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\|_1 N = o_p(1)$; the second term is bounded by $O_p(N^{-2}\|\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\|N)\|\boldsymbol{\Sigma}^{-1}\|_1 N = o_p(1)$. Consequently, $\widehat{\mathbf{w}}_T^{gmv}$ is L_1-consistent. ■

3.2. Measuring risks using full confidence bound

One of the main problems to be addressed is to assess the risk error $\Delta = |\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T|$ when $N > T$. A commonly used upper bound for Δ is based on the following inequality:

$$\Delta \leq \|\widehat{\mathbf{w}}_T\|_1^2\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\max} \equiv \widehat{\xi}_T, \tag{3.2}$$

which is asymptotically tight for risk assessment in the sense that $\widehat{\xi}_T$ stochastically converges to zero. However, for the purpose of statistical inference, $\widehat{\xi}_T$ is infeasible as it depends on the true $\boldsymbol{\Sigma}$. In addition, our simulation results have shown that the upper bound $\widehat{\xi}_T$ is actually too crude to be useful. Let us consider the following toy example.

Example 3.2. Consider three stocks with annualized returns that jointly follow a multivariate Gaussian distribution $\mathcal{N}_3(0, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = 0.04 \times \mathbf{I}_3$. The equal-weight portfolio $\mathbf{w}_T = (1/3, 1/3, 1/3)'$ is constructed. Our task is to estimate the portfolio risk using the sample covariance matrix $\mathbf{S}$ based on the simulated 21-day (one month) returns.

The true value of portfolio variance is $\mathbf{w}_T'\boldsymbol{\Sigma}\mathbf{w}_T = 0.0133$, which corresponds to a true risk, by definition, $(\mathbf{w}_T'\boldsymbol{\Sigma}\mathbf{w}_T)^{1/2} = 11.55\%$ per annum. Based on a simulated data set, the estimated portfolio variance $\mathbf{w}_T'\mathbf{S}\mathbf{w}_T = 0.0131$, equivalent to a perceived risk $(\mathbf{w}_T'\mathbf{S}\mathbf{w}_T)^{1/2} = 11.43\%$ per annum. In addition, for this realization, we also get $\widehat{\xi}_T = \|\mathbf{w}_T\|_1^2\|\boldsymbol{\Sigma} - \mathbf{S}\|_{\max} = 0.0248$. Based on this upper bound, a simple calculation shows that $\mathbf{w}_T'\boldsymbol{\Sigma}\mathbf{w}_T \in [0, 0.0379]$. In other words, the true risk $(\mathbf{w}_T'\boldsymbol{\Sigma}\mathbf{w}_T)^{1/2}$ lies in $[0, 19.46\%]$, an interval that is too wide to be meaningful. ■

Note that the inequality (3.2) holds for every sampling sequence $\{\mathbf{R}_t\}_{t=1}^T$. Hence $\widehat{\xi}_T$ is in fact an upper bound of full confidence, that is,

$$P(|\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T| \leq \widehat{\xi}_T) = 1.$$

The toy example is typical in the sense that $\widehat{\xi}_T$ is already too crude for small portfolios. In statistical inference, often people use bounds of high confidence levels instead, e.g., quantities that bound Δ with a high probability. This paper pursues such a high-confidence-level upper bound (H-CLUB) based on the confidence interval.

3.3. High-confidence-level upper bound

We propose a confidence upper bound for $\Delta = |\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T|$ to assess the estimation error of the portfolio risks. More specifically, for each proposed matrix estimator $\widehat{\boldsymbol{\Sigma}}$ and any given $\epsilon > 0$, we find a quantity $\widehat{U}(\epsilon)$ such that for all large N and T,

$$P(|\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T| \leq \widehat{U}(\epsilon)) \geq 1 - \epsilon.$$

Therefore, $\widehat{U}(\epsilon)$ is an asymptotic $(1 - \epsilon)100\%$ confidence upper bound for Δ, which is obtained based on the limiting distribution of $\widehat{\mathbf{w}}_T'\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{w}}_T$. In addition, it is data-driven (up to user-specified tuning parameters), hence can be easily calculated in practice and used to construct confidence intervals for the true risks.

4. Limiting distributions of risk estimators

4.1. Regularity conditions

We will treat N as an increasing function of T. Hence N grows via a fixed trajectory, e.g., $N = N_T = T^\alpha$ for some $\alpha > 0$, and can be faster than T, namely, $\alpha > 1$.

We shall show that for a well behaved diversified $\widehat{\mathbf{w}}_T$,

$$\sqrt{T}\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T = \frac{1}{\sqrt{T}}\sum_{t=1}^T Z_{T,t} + o_p(1),$$

where in the sample covariance based risk estimator, $Z_{T,t} = (\mathbf{w}_T'\mathbf{R}_t)^2 - E(\mathbf{w}_T'\mathbf{R}_t)^2$; in the factor based risk estimator, $Z_{T,t} = (\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2 - E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2$. Hence the limiting distribution of the estimated risk depends on that of $\sum_{t=1}^T Z_{T,t}$. Note that $\{Z_{T,t}\}_{t=1}^T$ is a triangular array of weakly dependent random variables, and a triangular array central limit theorem for weakly dependent time series data (e.g., Peligrad, 1996) can be applied. For this purpose, some technical assumptions are in order.

4.1.1. Data generating process

Assumption 4.1. (i) Assume that (2.1) holds. In addition, $\{\mathbf{R}_t, \mathbf{u}_t, \mathbf{f}_t\}_{t=1}^T$ is strictly stationary, $\{\mathbf{u}_t\}_{t=1}^T$ and $\{\mathbf{f}_t\}_{t=1}^T$ are independent, and $Eu_{it} = Ef_{jt} = 0$ for all i, j.
(ii) There exist $r_1, r_2 \in (0, 2]$ and $b_1, b_2 > 0$, such that for any $s > 0$,

$$P(|u_{it}| > s) \leq \exp(-(s/b_1)^{r_1}), \qquad P(|f_{jt}| > s) \leq \exp(-(s/b_2)^{r_2}).$$

(iii) There is $C > 0$ such that $C^{-1} < \lambda_{\min}(\mathbf{\Sigma}_u) \leq \lambda_{\max}(\mathbf{\Sigma}_u) < C$, $\|\mathbf{B}\|_{\max} < C$ and $\lambda_{\min}(\mathrm{cov}(\mathbf{f}_t)) > C^{-1}$.

For simplicity, we assume the true DGP to be a factor model for all the three risk estimators. In fact, when estimating $\mathbf{\Sigma}$ using the sample covariance only, assuming the factor structure on $\mathbf{R}_t$ is not necessary, as long as the exponential-tail condition is assumed directly on $\mathbf{R}_t$. The exponential-tail condition (ii) enables us to apply the large deviation theory to control the uniform convergence of $\max_{i\leq N} \|\frac{1}{T}\sum_{t=1}^T u_{it}\mathbf{f}_t\|$ and $\max_{i,j\leq N} |\frac{1}{T}\sum_{t=1}^T R_{it}R_{jt} - ER_{it}R_{jt}|$, which are needed for the consistent estimation of high-dimensional covariance matrices. On the other hand, when financial noises or factors have heavier tails, the analysis will be much more technically involved for non-i.i.d. data. Research along that line will be left for the future. Note that in the factor model being considered: $\mathbf{R}_t = \mathbf{B}\mathbf{f}_t + \mathbf{u}_t$, the exponential-tail condition also ensures that both the error term and the factors have bounded finite moments, which also implies that $\mathbf{w}'\mathbf{\Sigma}\mathbf{w}$ indeed exists. Moreover, it is standard in the literature to assume that the eigenvalues of covariance matrices for $\mathbf{f}_t$ and $\mathbf{u}_t$ are bounded away from both zero and infinity.

The factor model is assumed to be conditionally sparse as follows:

Assumption 4.2. There is $q \in [0, 1)$ such that

$$s_N \equiv \max_{i\leq N}\sum_{j=1}^N |\Sigma_{u,ij}|^q = o(\min\{(T/\log N)^{(1-q)/2}, N^{(1-q)/2}\}). \quad (4.1)$$

When $q = 0$ we define $s_N = \max_{i\leq N}\sum_{j=1}^N I(\Sigma_{u,ij} \neq 0)$ as the maximum number of non-vanishing elements in each row. Assumption 4.2, though slightly stronger than those in Chamberlain and Rothschild (1983), is meaningful in practice. For example, when the idiosyncratic components represent firms' individual shocks, they are either uncorrelated or weakly correlated among the firms across different industries, because the industry specific components are not pervasive for the whole economy (Connor and Korajczyk, 1993).

Technically, this assumption is easy to satisfy by many sparse covariances as long as $T/\log N \to \infty$. For instance, when $\mathbf{\Sigma}_u$ is a diagonal matrix, $s_N = 1$; when $\mathbf{\Sigma}_u$ is a block-diagonal matrix with uniformly bounded block sizes, s_N is the maximum block size; when $\mathbf{\Sigma}_u$ is a banded matrix, s_N is the width of the bands. In all these cases $q = 0$, and s_N is bounded, while the right hand side of (4.1) diverges fast. Recently, Gagliardini et al. (2010) discussed a detailed example of block dependence and its relation to the sparsity assumption and the approximate factor structure.

The following assumption is standard for high-dimensional factor models (e.g., Bai, 2003; Bai and Ng, 2002; Stock and Watson, 2002). Under these assumptions, the unknown factors and loadings can be consistently estimated.

Assumption 4.3. (i) There is $M > 0$ such that,
$E[N^{-1/2}(\mathbf{u}_s'\mathbf{u}_t - E\mathbf{u}_s'\mathbf{u}_t)]^4 < M$ and $E\|N^{-1/2}\sum_{i=1}^N \mathbf{b}_i u_{it}\|^4 < M$.
(ii) As $N \to \infty$, the eigenvalues of $\mathbf{B}'\mathbf{B}/N$ are bounded away from both zero and infinity.

4.1.2. Weakly dependent data

The autoregressive function of $Z_{T,t}$ plays a central role in the asymptotic variance of $\Delta = \widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\widehat{\mathbf{w}}_T$. For the sample covariance based risk estimator, the relevant object is $Z_{T,t} = (\mathbf{w}_T'\mathbf{R}_t)^2 - E(\mathbf{w}_T'\mathbf{R}_t)^2$. Hence we define the autoregressive function

$$\gamma_T(h) = \mathrm{cov}((\mathbf{w}_T'\mathbf{R}_t)^2, (\mathbf{w}_T'\mathbf{R}_{t+h})^2), \quad h \in \mathbb{Z}$$

which depends on T through $\dim(\mathbf{w}_T) = N = N_T$. For factor-based risk estimator, the relevant autoregressive function is defined for $Z_{T,t} = (\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2 - E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2$:

$$\gamma_f(h) = \mathrm{cov}((\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2, (\mathbf{w}_T'\mathbf{B}\mathbf{f}_{t+h})^2) \quad h \in \mathbb{Z}.$$

Accordingly, the confidence interval of $\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T$ depends on

$$\sigma_T^2 = \gamma_T(0) + 2\sum_{h=1}^\infty \gamma_T(h), \qquad \sigma_f^2 = \gamma_f(0) + 2\sum_{h=1}^\infty \gamma_f(h). \quad (4.2)$$

We make the following assumption on the strength of the serial dependence.

Assumption 4.4. As $T, N \to \infty$,
(i) $\frac{1}{T}\sum_{h=1}^T h\gamma_T(h) = o(\sigma_T^2)$, $\frac{1}{T}\sum_{h=1}^T h\gamma_f(h) = o(\sigma_f^2)$,
(ii) When the factors are unknown, $\sigma_f^2 N/T \to \infty$.

Condition (i) is standard for stationary weakly-dependent time series data. Let $\rho(h)$ denote the autocorrelation function of $Z_{T,t}$, corresponding to either $\gamma_T(h)$ or $\gamma_f(h)$. Then by the dominated convergence theorem, a sufficient condition is that $\sum_{h=1}^\infty |\rho(h)| < \infty$ and $1 + 2\sum_{h\geq 1}\rho(h)$ is bounded away from zero, which is satisfied by many standard time series models such as ARMA(1,1). Condition (ii) is needed only when the common factors are not observable, so that a large dimension N helps to estimate the unknown factors accurately. This condition requires at least $N/T \to \infty$ in order for the effect of estimating the common factors to be negligible in the asymptotic expansion of $\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\widehat{\mathbf{w}}_T$. See more explanations on this effect in Remark 4.4.

Next, we state the α-mixing condition. This condition enables us to apply the central limit theorem for the triangular array weakly dependent data. Let $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_T^\infty$ denote the σ-algebras generated by $\{(\mathbf{R}_t, \mathbf{f}_t, \mathbf{u}_t) : -\infty < t \leq 0\}$ and $\{(\mathbf{R}_t, \mathbf{f}_t, \mathbf{u}_t) : T \leq t < \infty\}$ respectively. Define the mixing coefficient

$$\alpha(T) = \sup_{A\in\mathcal{F}_{-\infty}^0, B\in\mathcal{F}_T^\infty} |P(A)P(B) - P(AB)|.$$

Assumption 4.5. There exist $r_3 > 0$, $M_1 > 0$ and $M_2 \in (0, 1)$ such that: for all $T \in \mathbb{Z}^+$,

$$\alpha(T) \leq \exp(-M_1 T^{r_3}), \quad \text{and} \quad \alpha(T) \leq M_2 \min\{\gamma_T(0), \gamma_f(0)\}.$$

The stated strong-mixing condition is also standard in the literature (e.g., Merlevède et al. (2011)). Note that it also follows from the α-mixing condition that $\sum_{h=1}^\infty |\gamma_T(h)| = O(1)$ and $\sum_{h=1}^\infty |\gamma_f(h)| = O(1)$ (see Lemma B.6 in the Appendix) and hence Assumption 4.4 holds easily as noted before. Moreover, we also show in Lemma B.5 that $\gamma_T(0) = O(1)$ and $\gamma_f(0) = O(1)$. Hence by the definition of (4.2),

$$\sigma_f^2 \leq \gamma_f(0) + 2\sum_{h=1}^\infty |\gamma_f(h)| = O(1).$$

Similarly, $\sigma_T^2 = O(1)$.

4.1.3. Diversified portfolio

We are particularly interested in the risks of diversified portfolios. These portfolios diversify away the idiosyncratic risk $\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T$, and hence the risks are mainly contributed by the factor risks $\mathbf{w}_T'\mathbf{B}\operatorname{cov}(\mathbf{f}_t)\mathbf{B}'\mathbf{w}_T$. A diversified portfolio vector $\mathbf{w}_T$ should not be dominated by a few of the cross-sectional units, and should satisfy the following technical condition:

Assumption 4.6. $\|\mathbf{w}_T\|_1 = O(1)$, $\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T = o(\sigma_f s_N^{-1}N^{1/2-q/2}T^{-1/2})$, and $\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T = o(\sigma_f^4 + \sigma_f s_N^{-1}T^{-q/2}(\log N)^{-(1-q)/2})$.

Assuming $\|\mathbf{w}_T\|_1 = O(1)$ ensures that the portfolio has a bounded exposure, which implies $\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T = O(1)$ and $\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T = O_p(1)$ because $\|\widehat{\mathbf{\Sigma}}\|_{\max} = O_p(1)$. This is formally proved by Lemma A.1 in the Appendix.

The required upper bounds on $\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T$ is a technical condition for our asymptotic analysis. Since the eigenvalues of $\mathbf{\Sigma}_u$ are bounded away from both zero and infinity, this condition can also be understood as requiring the same upper bounds on $\|\mathbf{w}_T\|_2^2$. While the required upper bounds seem complicated and technical, the intuition is clear: $\|\mathbf{w}_T\|_2$ should decay as $N \to \infty$, so that $\mathbf{w}_T$ is diversified. Conditions of a similar spirit can be found, for instance, in Chudik et al. (2011, Assumption 2.2). Moreover, recall that q and $s_N \geq 1$ are defined in Assumption 4.2. In the usual case of $q = 0$ and $s_N = O(1)$, the required upper bounds in the above assumption are simplified to $\sigma_f N^{1/2}T^{-1/2}$ and $\sigma_f^4 + \sigma_f(\log N)^{-1/2}$ respectively.

The above condition is also related to the "risk ratio":

$$\frac{\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T}{\mathbf{w}_T'\mathbf{B}\operatorname{cov}(\mathbf{f}_t)\mathbf{B}'\mathbf{w}_T},$$

the ratio of the idiosyncratic risk relative to the factor risk. Assumption 4.6 then requires the idiosyncratic risk to be dominated by the factor risk. To illustrate the intuition, consider the following example.

Example 4.1. Consider a one-factor model on the asset returns with $\operatorname{var}(f_t^2) > 0$. Suppose $\{\mathbf{f}_t\}_{t=1}^T$ are independent across t, and thus $\sigma_f^2 = \gamma_f(0) = (\mathbf{w}_T'\mathbf{B})^4\operatorname{var}(f_t^2)$. It is straightforward to show that Assumption 4.6 is satisfied as long as:

$$\frac{\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T}{\mathbf{w}_T'\mathbf{B}\mathbf{B}'\mathbf{w}_T} = o(\min\{1/\sqrt{\log N}, \sqrt{N/T}\}). \tag{4.3}$$

This condition is often satisfied by a diversified portfolio $\mathbf{w}_T$. For example, the equal-weight allocation $\mathbf{w}_T = (1/N, \ldots, 1/N)$ gives $\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T = O(N^{-1})$. For $C_N = |N^{-1}\sum_{i=1}^N b_i| = \|\mathbf{w}_T'\mathbf{B}\|$, (4.3) holds as long as $(\log N)/N^2 = o(C_N^4)$ and $T/N^3 = o(C_N^4)$. This is true since C_N is often bounded away from zero and $T/N^3 \to 0$. ■

Finally, the following condition controls the error of using $\widehat{\mathbf{w}}_T$ to estimate $\mathbf{w}_T$.

Assumption 4.7. For r_1, r_2 defined in Assumption 4.1,

$$\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1 = o_p(\min\{\sigma_T^2, \sigma_f^2\} \times \min\{(\log NT)^{-4/r_1}, (\log T)^{-4/r_2}\}).$$

This condition is only slightly stronger than the L_1-consistency $\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1 = o_p(1)$, because often σ_T^2 and σ_f^2 are bounded away from zero, and $\log(NT)$ is growing slowly. As an example, for the global minimum variance portfolio in Example 3.1, suppose $\sigma_T^2 > 0$ and $\sigma_f^2 > 0$,

$$\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1 = O_p(\|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_1) = O_p\left(s_N\left(\frac{\log N}{T}\right)^{(1-q)/2}\right).$$

Hence Assumption 4.7 is satisfied so long as s_N does not grow too fast (in fact, for many examples of sparse matrices, s_N is often bounded).

4.2. Sample covariance based risk estimator

Let us start with the risk estimator based on the sample covariance matrix $\mathbf{S}$. The estimation error has an asymptotic expansion $\widehat{\mathbf{w}}_T'(\mathbf{S}-\mathbf{\Sigma})\widehat{\mathbf{w}}_T = \frac{1}{T}\sum_{t=1}^T Z_{T,t} + R$, where $Z_{T,t} = (\mathbf{w}_T'\mathbf{R}_t)^2 - E(\mathbf{w}_T'\mathbf{R}_t)^2$. The remaining term R arises from the use of estimated portfolio $\widehat{\mathbf{w}}_T$, which is asymptotically negligible under Assumption 4.7. To construct the confidence interval for the risk $\widehat{\mathbf{w}}_T'\mathbf{\Sigma}\widehat{\mathbf{w}}_T$, let us estimate the autocovariance function $\gamma_T(h)$ by

$$\widehat{\gamma}(h) = T^{-1}\sum_{t=1}^{T-h}((\widehat{\mathbf{w}}_T'\mathbf{R}_t)^2 - \widehat{\mathbf{w}}_T'\mathbf{S}\widehat{\mathbf{w}}_T)((\widehat{\mathbf{w}}_T'\mathbf{R}_{t+h})^2 - \widehat{\mathbf{w}}_T'\mathbf{S}\widehat{\mathbf{w}}_T).$$

In particular, $\widehat{\gamma}(0) = T^{-1}\sum_{t=1}^T(\widehat{\mathbf{w}}_T'\mathbf{R}_t)^4 - (\widehat{\mathbf{w}}_T'\mathbf{S}\widehat{\mathbf{w}}_T)^2$. We employ the Newey–West estimator with Bartlett kernel to estimate the variance $\sigma_T^2 = \gamma_T(0) + 2\sum_{h=1}^\infty \gamma_T(h)$:

$$\widehat{\sigma}^2 = \widehat{\gamma}(0) + 2\sum_{h=1}^L\left(1 - \frac{h}{L}\right)\widehat{\gamma}(h), \tag{4.4}$$

where $L = L(T) \to \infty$ is a truncation parameter (see Newey and West, 1987). This estimator is always nonnegative in finite sample.

We are now ready to define the H-CLUB $\widehat{U}_S(\epsilon)$ under the confidence level $(1-\epsilon)100\%$, which is data-driven once a user-specified L is determined. Let $z_{\epsilon/2}$ denote the upper $\epsilon/2$ quantile of the standard normal distribution. Let

$\widehat{U}_S(\epsilon) = z_{\epsilon/2}\widehat{\sigma}/\sqrt{T}$.

Lemma 4.1. *Under Assumptions 4.1, 4.4 and 4.5, suppose $L^3 = o(T\sigma_T^4)$, $L\sigma_T^2 \to \infty$, and $\sum_{h>L}\gamma(h) = o(\sigma_T^2)$, then*

$$|\widehat{\sigma}^2 - \sigma_T^2| = o_p(\sigma_T^2) \quad \text{and} \quad \widehat{U}_S(\epsilon) = o\left(\sqrt{\frac{\log N}{T}}\right).$$

Remark 4.1. The conditions $L^3 = o(T\sigma_T^4)$, $L\sigma_T^2 \to \infty$, and $\sum_{h>L}\gamma(h) = o(\sigma_T^2)$ are called to control the aggregated errors of estimating $\gamma_T(h)$ and the estimation bias when we use a truncated sum to approximate $\sum_{h=1}^\infty \gamma_T(h)$. These estimation errors often arise in estimating the covariances with serially correlated data (see Newey and West, 1987; Andrews, 1991), and the required conditions are satisfied with a slowly growing L, e.g., $L^3 = o(T)$ when σ_T^2 is bounded away from zero. ■

The following theorem gives the limiting distribution of the estimated risk. It also demonstrates that $\widehat{U}_S(\epsilon)$ is a valid H-CLUB for $|\widehat{\mathbf{w}}_T'(\mathbf{S}-\mathbf{\Sigma})\widehat{\mathbf{w}}_T|$.

Theorem 4.1. *Under the assumptions of Lemma 4.1, as $T \to \infty$ and $N = N_T \to \infty$,*

$$\left[\operatorname{var}\left(\sum_{t=1}^T(\mathbf{w}_T'\mathbf{R}_t)^2\right)\right]^{-1/2} T\widehat{\mathbf{w}}_T'(\mathbf{S}-\mathbf{\Sigma})\widehat{\mathbf{w}}_T \to^d \mathcal{N}(0,1),$$

and for any $\epsilon > 0$,

$$P\left(|\widehat{\mathbf{w}}_T'(\mathbf{S}-\mathbf{\Sigma})\widehat{\mathbf{w}}_T| \leq \widehat{U}_S(\epsilon)\right) \to 1-\epsilon.$$

By the delta-method, we have the following corollary for the risk estimation. Define

$$\hat{R}(\widehat{\mathbf{w}}_T) = \sqrt{\widehat{\mathbf{w}}_T'\mathbf{S}\widehat{\mathbf{w}}_T}, \qquad R(\widehat{\mathbf{w}}_T) = \sqrt{\widehat{\mathbf{w}}_T'\mathbf{\Sigma}\widehat{\mathbf{w}}_T}.$$

Corollary 4.1. *Under the assumptions of Lemma 4.1, for any $\epsilon > 0$, as $T, N \to \infty$,*

$$P\left(|\hat{R}(\widehat{\mathbf{w}}_T) - R(\widehat{\mathbf{w}}_T)| \leq \widehat{U}_S(\epsilon)/\sqrt{4\widehat{\mathbf{w}}_T'\mathbf{S}\widehat{\mathbf{w}}_T}\right) \to 1-\epsilon.$$

4.3. Factor-based risk estimator

Let us now approach the problem in the factor model. We assume $\mathbf{R}_t = \mathbf{B}\mathbf{f}_t + \mathbf{u}_t$, where in this section, $\{\mathbf{f}_t\}_{t=1}^T$ are observed common factors. We show that the dominating term in the asymptotic expansion of $\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\widehat{\mathbf{w}}_T$ is

$$\frac{1}{T}\sum_{t=1}^{T}(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2 - E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2.$$

Hence the estimation error of the risk only comes from the systematic error brought by the common factors, and the risk component $\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\widehat{\mathbf{w}}_T$ introduced by the idiosyncratic error can be diversified away by a selected portfolio allocation vector.

To construct H-CLUB, we need to first estimate $\gamma_f(h)$, the autocovariance function of $(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2$. For $\widehat{\mathrm{cov}}(\mathbf{f}_t) = T^{-1}\sum_{t=1}^T \mathbf{f}_t\mathbf{f}_t'$, define

$$\widehat{\gamma}_f(h) = T^{-1}\sum_{t=1}^{T-h}[(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\mathbf{f}_{t+h})^2 - \widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\widehat{\mathbf{w}}_T] \times [(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\mathbf{f}_t)^2 - \widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\widehat{\mathbf{w}}_T],$$

where $\widehat{\mathbf{B}}$ is the least squares estimator of $\mathbf{B}$. For some $L = L(T) \to \infty$, apply the Newey–West estimator to estimate $\sigma_f^2 = \gamma_f(0) + 2\sum_{h=1}^{\infty}\gamma_f(h)$:

$$\widehat{\sigma}_f^2 = \widehat{\gamma}_f(0) + 2\sum_{l=1}^{L}\left(1 - \frac{h}{L}\right)\widehat{\gamma}_f(h), \tag{4.5}$$

which is always nonnegative. Define

$\widehat{U}_f(\epsilon) = z_{\epsilon/2}\widehat{\sigma}_f/\sqrt{T}$.

Let $\beta = 3(r_1^{-1} + r_2^{-1} + r_3^{-1})$, where r_1, r_2, r_3 are defined as in Assumptions 4.1 and 4.5.

Lemma 4.2. *Suppose* $(\log N)^{2\beta+2} = o(T)$, *and the truncation satisfies* $L\sqrt{(L + \log N)/T} + \sum_{h>L}\gamma_f(h) = o(\sigma_f^2)$, *and* $L\sigma_f^2 \to \infty$. *Under Assumptions* 4.1, 4.2 *and* 4.4–4.6,

$$|\widehat{\sigma}_f^2 - \sigma_f^2| = o_p(\sigma_f^2), \quad \text{and} \quad \widehat{U}_f(\epsilon) = o\left(\sqrt{\frac{\log N}{T}}\right).$$

Remark 4.2. The condition $L\sqrt{(L + \log N)/T} + \sum_{h>L}\gamma_f(h) = o(\sigma_f^2)$ and $L\sigma_f^2 \to \infty$ ensure that the effect of estimating the limiting distribution is asymptotically negligible. The first term $L\sqrt{(L + \log N)/T}$ represents the error of estimating $\gamma_f(0) + 2\sum_{l=1}^{L}(1-h/L)\gamma_f(h)$, while $\sum_{h>L}\gamma_f(h)$ arises from approximating $\sum_{h=1}^{\infty}\gamma_f(h)$ by a truncated sum. These conditions are easily satisfied with a slowly-growing L, for instance, when $L^3 = o(T)$ and σ_f^2 is bounded away from zero.

The following theorem shows that $\widehat{U}_f(\epsilon)$ is a valid H-CLUB for the risk estimation error. Technically, the estimation error for the factor loadings is asymptotically negligible even under high dimensionality.

Theorem 4.2. *Suppose that the common factors are observable, and that the thresholded* $\widehat{\mathbf{\Sigma}}_f$ (2.4) *is used as the covariance estimator. Under the assumptions of Lemma* 4.2,

$$\left[\mathrm{var}\left(\sum_{t=1}^{T}(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2\right)\right]^{-1/2} T\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\widehat{\mathbf{w}}_T \to^d \mathcal{N}(0,1),$$

and for any $\epsilon > 0$,

$$P\left(|\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\widehat{\mathbf{w}}_T| \le \widehat{U}_f(\epsilon)\right) \to 1 - \epsilon.$$

Remark 4.3. Similar to Corollary 4.1, if we use $\hat{R}_f(\widehat{\mathbf{w}}_T) = \sqrt{\widehat{\mathbf{w}}_T'\widehat{\mathbf{\Sigma}}_f\widehat{\mathbf{w}}_T}$ to estimate $R(\widehat{\mathbf{w}}_T) = \sqrt{\widehat{\mathbf{w}}_T'\mathbf{\Sigma}\widehat{\mathbf{w}}_T}$, then applying the delta-method yields

$$P\left(|\hat{R}_f(\widehat{\mathbf{w}}_T) - R(\widehat{\mathbf{w}}_T)| \le \widehat{U}_f(\epsilon)/\sqrt{4\widehat{\mathbf{w}}_T'\widehat{\mathbf{\Sigma}}_f\widehat{\mathbf{w}}_T}\right) \to 1 - \epsilon.$$

Hence $\widehat{U}_f(\epsilon)/\sqrt{4\widehat{\mathbf{w}}_T'\widehat{\mathbf{\Sigma}}_f\widehat{\mathbf{w}}_T}$ is a valid H-CLUB for $|\hat{R}_f(\widehat{\mathbf{w}}_T) - R(\widehat{\mathbf{w}}_T)|$.

It is also interesting to compare $\widehat{U}_f(\epsilon)$ with $\widehat{U}_S(\epsilon)$ and see if knowing the factor structure results in a reduced upper bound. This is equivalent to comparing the asymptotic variances of the estimated risks between a pure nonparametric estimator (sample covariance) and an estimator based on factor models. We will see in the following subsection that the factor-based risk estimator indeed gives a slightly smaller asymptotic variance.

Recently Gagliardini et al. (2010) considered a similar covariance thresholding problem for the *risk premium* in a random coefficient panel model, where the factors are assumed to be observable. They studied the inferential theory for estimating the *risk premium*. While we are based on similar frameworks of the conditionally sparse factor model, we focus on the inferential theory of the estimated risks for a given portfolio vector, and the impacts on risks from estimating large covariance matrices. In the following subsection, we also study the unobservable factor case.

4.4. Risk estimation with unknown factors

When the market assets' returns are driven by a few unknown factors, one needs to handle the difficulty of not knowing the common factors in estimating the risk covariance matrix. In this case, the covariance estimator of Fan et al. (2013) is defined by

$$\widehat{\mathbf{\Sigma}}_{P,\widehat{K}} = \sum_{j=1}^{\widehat{K}}\widehat{\lambda}_j\widehat{\boldsymbol{\xi}}_j\widehat{\boldsymbol{\xi}}_j' + \widehat{\mathbf{\Omega}} \tag{4.6}$$

as described in Section 2.2. The risk estimator for a given portfolio $\widehat{\mathbf{w}}_T$ is then $\widehat{\mathbf{w}}_T'\widehat{\mathbf{\Sigma}}_{P,\widehat{K}}\widehat{\mathbf{w}}_T$, which incorporates the case of unknown number of factors K. As we shall demonstrate below, using the consistent estimator of K does not affect the asymptotic behavior of the covariance estimator. In addition to the systematic risk $\sum_{t=1}^{T}Z_{T,t}$ as before, there are three other components in the asymptotic expansion of the risk: effects of estimating the unknown loadings, factors, as well as the idiosyncratic risk. All these three components are asymptotically negligible.

Under the conditional sparsity condition, Fan et al. (2011, 2013) showed that, when the common factors are observable,

$$\|\widehat{\mathbf{\Sigma}}_f^{-1} - \mathbf{\Sigma}^{-1}\| = O_p\left(s_N\left(\frac{\log N}{T}\right)^{1/2-q/2}\right). \tag{4.7}$$

When the common factors are unobservable,

$$\|\widehat{\mathbf{\Sigma}}_{P,\widehat{K}}^{-1} - \mathbf{\Sigma}^{-1}\| = O_p\left(s_N\left(\frac{\log N}{T} + \frac{1}{N}\right)^{1/2-q/2}\right) \tag{4.8}$$

where q and s_N are defined in Assumption 4.2. The term $1/N$ in (4.8) is the price for not knowing $\mathbf{f}_t$. When $T = o(N\log N)$, the above convergence rates are the same. This also explains the condition (iii) in Assumption 4.4, which requires $N/T \to \infty$ in the case when $\mathbf{f}_t$ is unobservable. Intuitively, as the dimensionality increases, more information about the common factors is collected, and eventually the common factors can be treated as though they are known. Therefore, the effect of estimating the unknown factors on the estimated risk is negligible, and $\widehat{\mathbf{w}}_T'\widehat{\mathbf{\Sigma}}_{P,\widehat{K}}\widehat{\mathbf{w}}_T$ and $\widehat{\mathbf{w}}_T'\widehat{\mathbf{\Sigma}}_f\widehat{\mathbf{w}}_T$

have the same limiting distribution. This is often true for the asset returns' time series data. The number of assets can be in thousands while the sample size on the monthly returns over ten years is slightly larger than a hundred.

To define an H-CLUB for a factor model with unknown factors, we first apply the principal components method (e.g., Stock and Watson, 2002) to estimate σ_f^2. Let $\widehat{\mathbf{F}} = (\widehat{\mathbf{f}}_1, \ldots, \widehat{\mathbf{f}}_T)$ be a $\widehat{K} \times T$ matrix such that the rows of $\widehat{\mathbf{F}}/\sqrt{T}$ are the eigenvectors corresponding to the $\widehat{K}$ largest eigenvalues of the $T \times T$ matrix $\mathbf{R}'\mathbf{R}$, where $\mathbf{R} = (\mathbf{R}_1, \ldots, \mathbf{R}_T)$. Let $\widehat{\mathbf{B}} = \mathbf{R}\widehat{\mathbf{F}}'/T$. Define

$$\widehat{\gamma}_P(h) = T^{-1}\sum_{t=1}^{T-h}[(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\,\widehat{\mathbf{f}}_{t+h})^2 - \widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\,\widehat{\mathbf{B}}'\widehat{\mathbf{w}}_T] \times [(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\,\widehat{\mathbf{f}}_t)^2 - \widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\,\widehat{\mathbf{B}}'\widehat{\mathbf{w}}_T].$$

For some $L = L(T) \to \infty$, let

$$\widehat{\sigma}_P^2 = \widehat{\gamma}_P(0) + 2\sum_{h=1}^{L}\left(1 - \frac{h}{L}\right)\widehat{\gamma}_P(h),$$

$$\widehat{U}_P(\epsilon) = z_{\epsilon/2}\sqrt{\widehat{\sigma}_P^2/T}. \tag{4.9}$$

The following lemma characterizes the estimation error of $\widehat{\sigma}_P^2$.

Lemma 4.3. *Suppose conditions in Lemma 4.2 are satisfied and $L = o(\sqrt{N}\sigma_f^2)$. Under Assumptions 4.1–4.6,*

$$|\widehat{\sigma}_P^2 - \sigma_f^2| = o_p(\sigma_f^2), \qquad \widehat{U}_P(\epsilon) = o\left(\sqrt{\frac{\log N}{T}}\right).$$

Remark 4.4. Compared to the conditions in Lemma 4.2, the extra requirement $L = o(\sqrt{N}\sigma_f^2)$ controls the effect of estimating the unknown factors. So the rates of convergence of estimating $\boldsymbol{\Sigma}_u$ and σ_f^2 are the same when $N(\log N)/T \to \infty$. Intuitively, as the number of unknown factors is $O(T)$, we require more cross sectional units (N) to accurately estimate them. This is also seen in the results in Bai (2003) and Fan et al. (2013).

The following theorem shows that $\widehat{U}_P(\epsilon)$ is an H-CLUB for $\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T$. Interestingly, $\widehat{\mathbf{w}}_T'\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}\widehat{\mathbf{w}}_T$ and $\widehat{\mathbf{w}}_T'\widehat{\boldsymbol{\Sigma}}_f\widehat{\mathbf{w}}_T$ have the same asymptotic limiting distribution. The price paid for not knowing the factors is asymptotically negligible.

Theorem 4.3. *Suppose the common factors are unobservable, and $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}$ (4.6) is used as the covariance estimator. Under the assumptions of Lemma 4.3,*

$$\left[\mathrm{var}\left(\sum_{t=1}^{T}(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2\right)\right]^{-1/2} T\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T \to^d \mathcal{N}(0,1),$$

and for any $\epsilon > 0$,

$$P\left(|\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T| \le \widehat{U}_P(\epsilon)\right) \to 1 - \epsilon.$$

Remark 4.5. Similarly, if we define $\hat{R}_P(\widehat{\mathbf{w}}_T) = \sqrt{\widehat{\mathbf{w}}_T'\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}\widehat{\mathbf{w}}_T}$, then $\widehat{U}_P(\epsilon)/\sqrt{4\widehat{\mathbf{w}}_T'\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}\widehat{\mathbf{w}}_T}$ is a valid H-CLUB for $|\hat{R}_P(\widehat{\mathbf{w}}_T) - R(\widehat{\mathbf{w}}_T)|$.

Knowing the factor-structure of the return $\mathbf{R}_t$ improves the estimation efficiency relative to the sample covariance estimator. This is demonstrated by the following theorem.

Theorem 4.4. *Under the assumptions of Theorem 4.3,*

$$\mathrm{var}\left[\sum_{t=1}^{T}(\mathbf{w}_T'\mathbf{R}_t)^2\right] > \mathrm{var}\left[\sum_{t=1}^{T}(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2\right].$$

Table 1
Mean and covariance used to generate $\mathbf{b}_i$.

$\boldsymbol{\mu}_B$	$\boldsymbol{\Sigma}_B$		
0.9833	0.0921	−0.0178	0.0436
−0.1233	−0.0178	0.0862	−0.0211
0.0839	0.0436	−0.0211	0.7624

In fact, the difference of the above two variances is small when $\mathbf{w}_T$ is diversified enough, and this fact is further verified by our simulation results (see Tables 3 and 5 in Section 5). The reason is that the systematic risk cannot be diversified, and dominates the idiosyncratic risk. On the other hand, factor models give a strictly positive definite covariance estimator, whereas the sample covariance may produce a risk estimator being zero for certain portfolio allocation vectors. The positive definiteness is particularly important to estimate the optimal portfolio allocation vector. Furthermore, factor models interpret the structure of portfolio's risks. It is clearly seen in both Theorems 4.2 and 4.3 that the idiosyncratic risks are diversified away by the portfolio allocation.

5. Monte Carlo examples

In this section, we examine the finite-sample performance of both the full confidence upper bound $\widehat{\xi}_T$ defined in (3.2) and H-CLUB, based on three covariance estimators $\widehat{\boldsymbol{\Sigma}}$, using portfolios $\mathbf{w}_T$ with different gross exposure constraints.

Excess returns of the ith stock of a portfolio over the risk-free interest rate is assumed to follow the Fama–French three-factor model [Fama and French (1992)]:

$$R_{it} = \lambda_{i1}f_{1t} + \lambda_{i2}f_{2t} + \lambda_{i3}f_{3t} + u_{it}.$$

The first factor is the excess return of the whole equity market, while the second and third factors are SMB ("small minus big" cap) and HML ("high minus low" book/price) respectively. Using US equity market data, we calibrate a sub-model to generate the loadings $\mathbf{b}_i = (\lambda_{i1}, \lambda_{i2}, \lambda_{i3})'$, the idiosyncratic noises $\mathbf{u}_t$ and the factors $\mathbf{f}_t = (f_{1t}, f_{2t}, f_{3t})'$.

5.1. Calibration

To calibrate parameters in the model, we use the data on daily returns of S&P 500's top 100 constituents ranked by market capitalization (on June 29th 2012), the data on 3-month Treasury bill rates, and daily return data of the Fama–French factors. They are obtained from COMPUSTAT database, the data library of Kenneth French's website, and CRSP database respectively. The excess returns $(\tilde{\mathbf{R}}_t, \tilde{\mathbf{f}}_t)$ are analyzed for the period from July 1st, 2008 to June 29th 2012, approximately 1000 trading days.

(1) Calculate the least square estimator $\tilde{\mathbf{B}}$ of $\tilde{\mathbf{R}}_t = \mathbf{B}\tilde{\mathbf{f}}_t + \mathbf{u}_t$, and compute the sample mean vector $\boldsymbol{\mu}_B$ and sample covariance matrix $\boldsymbol{\Sigma}_B$ of all the row vectors of $\tilde{\mathbf{B}}$. These parameters are reported in Table 1. The factor loadings $\{\mathbf{b}_i\}_{i=1}^N$ of the simulated models are then generated from a trivariate Gaussian distribution $\mathcal{N}_3(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B)$.

(2) Assume that the factors follow the stationary vector autoregressive VAR(1) model $\mathbf{f}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{f}_{t-1} + \boldsymbol{\varepsilon}_t$ for some 3×3 matrix $\boldsymbol{\Phi}$, where $\boldsymbol{\varepsilon}_t$ follows i.i.d $\mathcal{N}_3(0, \boldsymbol{\Sigma}_\epsilon)$. The model parameters $\boldsymbol{\Phi}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_\epsilon$ are calibrated using the daily excess returns of the Fama–French factors $\tilde{\mathbf{f}}_t$. The covariance matrix $\mathrm{cov}(\mathbf{f}_t)$ is then obtained by solving the linear equation $\mathrm{cov}(\mathbf{f}_t) = \boldsymbol{\Phi}\,\mathrm{cov}(\mathbf{f}_t)\boldsymbol{\Phi}' + \boldsymbol{\Sigma}_\epsilon$. Results are summarized in Table 2.

(3) The error covariance matrix is sparse in our setting. For each fixed N, it is created by $\boldsymbol{\Sigma}_u = \mathbf{D}\boldsymbol{\Sigma}_0\mathbf{D}$, where $\mathbf{D} = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$. To be more specific, $\sigma_1, \ldots, \sigma_p$ are generated independently from a Gamma distribution $G(\alpha, \beta)$, in which α and β are selected to

Table 2
Parameters used to generate $\mathbf{f}_t$.

$\boldsymbol{\mu}$	$\boldsymbol{\Phi}$			$\mathrm{cov}(\mathbf{f}_t)$		
0.0260	−0.1006	0.2803	−0.0365	3.2351	0.1783	0.7783
0.0211	−0.0191	−0.0944	0.0186	0.1783	0.5069	0.0102
−0.0043	0.0116	−0.0272	0.0272	0.7783	0.0102	0.6586

Table 3
Averages and standard deviations of RE_1 over 500 replications.

	$c = 1$	$c = 1.2$	$c = 1.4$	$c = 1.6$	$c = 1.8$	$c = 2$
RE_1	4.2552	6.0757	8.3303	10.9470	14.0553	17.2724
$\mathbf{S}$	(1.4595)	(2.0517)	(2.8590)	(3.6818)	(5.0410)	(6.4417)
RE_1	4.2462	6.0469	8.3545	10.9765	14.0610	17.3298
$\widehat{\boldsymbol{\Sigma}}_f$	(1.4838)	(2.0638)	(2.9640)	(3.7347)	(5.1142)	(6.5919)
RE_1	4.1813	5.9698	8.2153	10.8263	13.8465	17.0570
$\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}$	(1.4888)	(2.0866)	(2.9936)	(3.7560)	(5.1758)	(6.6062)

match the sample mean and sample standard deviation of the 100 standard deviations of the errors $\tilde{\mathbf{u}}_t = \tilde{\mathbf{R}}_t - \tilde{\mathbf{B}}\tilde{\mathbf{f}}_t$ (recall that each $\tilde{\mathbf{u}}_t$ is 100 dimensional; see also Fan et al., 2008). An additional restriction is imposed on σ_i that only values in between the minimum and maximum of the standard deviation of $\tilde{\mathbf{u}}_t$ are accepted. We then generate the off-diagonal entries of the correlation matrix $\boldsymbol{\Sigma}_0$ independently from a Gaussian distribution, with mean and standard deviation equal to those of the sample correlations of the estimated residuals. Moreover, absolute values of the off-diagonal entries are set to no greater than 0.95. Finally the hard-thresholding is applied to make $\boldsymbol{\Sigma}_0$ sparse, where the threshold is set to be the smallest constant that makes $\boldsymbol{\Sigma}_0$ positive definite.

5.2. Representative portfolios

We examine the performance of H-CLUB based on $\mathbf{w}_T$ with a couple of different gross exposures. For a given exposure c and given number of assets N, we randomly generate portfolios $\mathbf{w}_T$ that satisfy $\sum_{i=1}^N w_i = 1$ and $\sum_{i=1}^N |w_i| = c$. This task, which generates uniformly from the above set in $\mathbb{R}^N$, is of independent interest for portfolio optimization and research. We propose the following method.

Let w^+ be the total long position and w^- be the total short position: $w^+ = (c+1)/2$ and $w^- = (c-1)/2$, where $c = \|\mathbf{w}_T\|_1$. For $c = 1$, there are no-short positions. For $c > 1$, there are both long and short positions (positive and negative components of $\mathbf{w}_T$). The identities (or indices) of long and short positions are hard to identify, but the following sampling scheme is a reasonable approximation: The positive positions are determined by a Bernoulli trial (N times) with probability of success $w^+/(w^+ + w^-) = (c+1)/(2c)$. Once the identities are determined, we can normalize them and the problem reduces to the case with $c = 1$. For the case with $c = 1$, the uniform distribution on the set $\{w_i : \sum_{i=1}^N w_i = 1, w_i \geq 0\}$ can be generated from a normalized exponential distribution:

$$w_i = \zeta_i / \sum_{i=1}^N \zeta_i, \quad \zeta_i \sim_{i.i.d.} \text{standard exponential}.$$

Combining the above two steps, we can generate a randomly selected portfolio $\mathbf{w}_T$ of size N with a gross exposure c as follows.

1. Generate a positive integer k, the number of stocks with positive weights in $\mathbf{w}_T$, from a binomial distribution $\mathrm{Bin}(N, \frac{c+1}{2c})$.
2. Generate independently $\{\zeta_i\}_{i=1}^k$ from the standard exponential distribution and set each $w_i^+ = (c+1)\zeta_i/(2\sum_{j=1}^k \zeta_j)$, for $i = 1, \ldots, k$.
3. Analogously compute for $i = 1, \ldots, N-k$, $w_i^- = (1-c)\zeta_i/(2\sum_{j=1}^{N-k} \zeta_j)$, where $\{\zeta_j\}_{j=1}^{N-k}$ are obtained independently from the standard exponential distribution.
4. Take the portfolio weights $\mathbf{w}_T$ as a random permutation of the numbers $\{w_i^+\}_{i=1}^k$ and $\{-w_i^-\}_{i=1}^{N-k}$.

5.3. Choosing the time lag L

The time lag can be chosen using the plug-in method, previously suggested by Newey and West (1994) and Andrews (1991). The plug-in method chooses L by minimizing the estimated mean squared error of the variance estimator $\hat{\sigma}^2(L)$ of σ_T^2 in (4.2). As suggested by Newey and West (1994), we find L^* by

$$L^* = 1.447T^{1/3}(\hat{s}_1/\hat{s}_0)^{2/3}$$

where

$$\hat{s}_1 = 2\sum_{h=1}^n h\hat{\gamma}(h), \qquad \hat{s}_0 = \hat{\gamma}(0) + 2\sum_{h=1}^n \hat{\gamma}(h),$$

$$n = 4(T/100)^{2/9}.$$

We recommend initially setting $L = L^*$, and then exercise some judgment about sensitivity of results to the choice of n and L. Some evidence from the literature suggests that the final result is less sensitive to n than to L (Silverman, 1986). In the simulation and empirical studies below, L is chosen by this method.

5.4. Simulation

5.4.1. Data generation

For each given gross exposure c, number of assets N and length of time series T, we generate 50 different models and 200 testing portfolios for each model, so a total of 10,000 portfolios are actually used. To be more specific, in any simulation with fixed c, N and T, we repeat the following steps for 50 times:

1. Generate $\{\mathbf{b}_i\}_{i=1}^N$ independently from $\mathcal{N}_3(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B)$. Set $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_p)'$.
2. Generate $\{\mathbf{u}_t\}_{t=1}^T$ independently from $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_u)$.
3. Generate $\{\mathbf{f}_t\}_{t=1}^T$ from a VAR(1) model $\mathbf{f}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{f}_{t-1} + \boldsymbol{\varepsilon}_t$ with parameters specified in the calibration part.
4. Calculate $\mathbf{R}_t = \mathbf{B}\mathbf{f}_t + \mathbf{u}_t$ for $t = 1, \ldots, T$.
5. Calculate the sample covariance matrix $\mathbf{S} = T^{-1}\sum_{t=1}^T(\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})'$; obtain the factor-based covariance estimators by soft-thresholding the sample correlation matrices.
6. Generate 200 $\mathbf{w}_T$ according to the method described in Section 5.2.
7. (i) Compute the true risk $R(\mathbf{w}_T) = \sqrt{\mathbf{w}_T'\boldsymbol{\Sigma}\mathbf{w}_T}$; (ii) For $\widehat{\boldsymbol{\Sigma}} = \mathbf{S}$, $\widehat{\boldsymbol{\Sigma}}_f$ and $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}$, calculate $\Delta = |\mathbf{w}_T'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\mathbf{w}_T|$, $\widehat{\xi}_T = \|\mathbf{w}_T\|_1^2\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\max}$ and $\widehat{U}(0.05) = 1.96\widehat{\sigma}/\sqrt{T}$; (iii) Calculate empirical coverage probabilities based on the 200 generated portfolios $\mathbf{w}_T$, i.e., the proportions of times that $\Delta < \widehat{U}(0.05)$. Note that time lags L are determined by the plug-in method described in Section 5.3.

5.4.2. Output

We first produce the graph of risk domain by plotting averages of $R(\mathbf{w}_T)$ as a function of c and N, followed by plots of Δ, $\widehat{\xi}_T$ and $\widehat{U}(\epsilon)$ against N for sample-based, factor-based and POET-based estimators. Average absolute distance between the empirical coverage probability and the nominal level (95%) are also plotted against N, for all three kinds of estimators. Finally, fix the dimensionality N and increase the number of model generation to 500, we look at two ratio quantities, namely

$$\mathrm{RE}_1 = \frac{\widehat{\xi}_T}{\widehat{U}(\epsilon)} = \frac{\|\mathbf{w}_T\|_1^2\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\max}}{1.96\sqrt{\widehat{\mathrm{var}}(\mathbf{w}_T'\boldsymbol{\Sigma}\mathbf{w}_T)}} \quad \text{and}$$

$$\mathrm{RE}_2 = \frac{\sqrt{\widehat{\mathrm{var}}(\mathbf{w}_T'\widehat{\boldsymbol{\Sigma}}\mathbf{w}_T)}}{2\mathbf{w}_T'\boldsymbol{\Sigma}\mathbf{w}_T}.$$

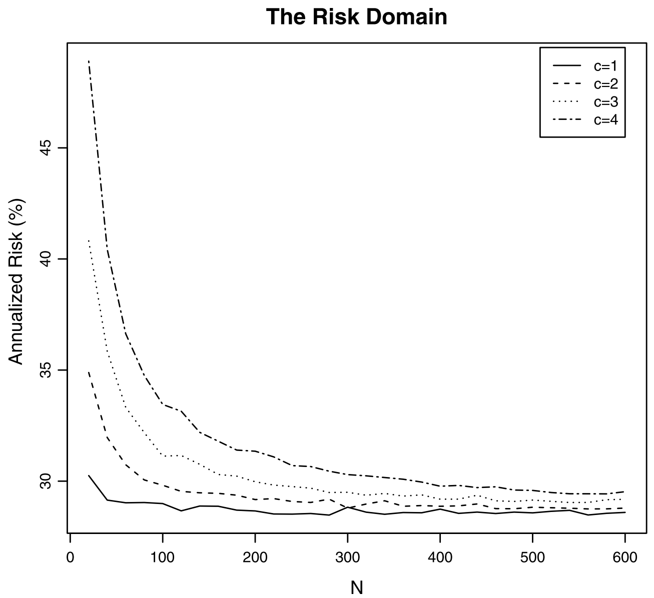


Fig. 1. Averages of annualized risks $R(\mathbf{w}_T)$ over 10,000 portfolios.

Means and standard deviations of RE_1 and RE_2 under a selection of c and T are reported, for all three kinds of estimators. Respective plots and tables are presented and analyzed in Section 5.5.

5.5. Results

In Fig. 1, we gradually increase N from 20 to 600 in increments of 20. Averages of $R(\mathbf{w}_T)$ over 10,000 portfolios are plotted against N. We produce the curves associated with four choices of c, all with $T = 300$. For the sake of comparison, we overlay four curves together on the same plot. The following observations can be made from Fig. 1.

(1) The average risk ranges from less than 30% to around 50% per annum.
(2) The average risk is higher for larger exposure c. This is consistent with the fact that portfolios with greater gross exposure are more volatile, and hence incur higher risks.
(3) Given a gross exposure c, as the portfolio size N increases, the average risk decreases. The rate of decline is very fast until N is around 150. This is consistent with the theory that as N increases, the portfolio becomes more diversified and the idiosyncratic risk is reduced through diversification.

Let N grow from 100 to 600 in grid of 20. In Fig. 2, their respective averages of $\Delta = |\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T|$, $\widehat{\xi}_T$ and $\widehat{U}(\epsilon)$ over 10,000 portfolios are plotted using estimators $\widehat{\boldsymbol{\Sigma}} = \mathbf{S}$, $\widehat{\boldsymbol{\Sigma}}_f$ and $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}$. In particular, $c = 1.6$ results in 130% long positions and 30% short positions (130/30 strategy). The 130/30 structure is popular in long-short funds. In each subfigure, the dashed curve corresponds to Δ, the solid curve corresponds to $\widehat{U}(\epsilon)$, $\epsilon = 0.05$, and the two-dash curve corresponds to $\widehat{\xi}_T$. Based on these plots, we can observe the following features:

(1) Solid curves lie entirely above dashed curves, ensuring $\widehat{U}(\epsilon)$ as a valid H-CLUB. This feature maybe difficult to observe on Fig. 2(c) and (d) with the presence of two-dash curves ($\widehat{\xi}_T$). To make the comparison clearer, we also produce plots without the presence of $\widehat{\xi}_T$, as shown in Fig. 3 (a zoomed-in version of Fig. 2).
(2) The full confidence upper bound $\widehat{\xi}_T$ is indeed a very crude bound and is much larger than $\widehat{U}(\epsilon)$. The larger c is, the larger the difference is. This feature is further demonstrated in Table 3.

Table 4
Averages and standard deviations of RE_1 over 500 replications using POET-based estimator.

	$\epsilon = 0.05$	$\epsilon = 0.01$	$\epsilon = 0.005$	$\epsilon = 0.001$
RE_1	4.1813	3.2413	2.4030	2.1498
$c = 1$	(1.4888)	(1.1541)	(0.8556)	(0.7654)
RE_1	10.8263	8.3925	6.2220	5.5662
$c = 1.6$	(2.9936)	(2.3206)	(1.7205)	(1.5391)

Table 5
Averages and standard deviations of RE_2 over 500 replications, with $T = 200$.

	$c = 1$	$c = 1.2$	$c = 1.4$	$c = 1.6$	$c = 1.8$	$c = 2$
RE_2	4.8968%	4.8187%	4.9171%	4.8137%	4.8808%	4.8217%
$\mathbf{S}$	(0.8440%)	(0.8211%)	(0.8694%)	(0.8113%)	(0.8586%)	(0.8085%)
RE_2	4.8888%	4.8139%	4.8959%	4.7921%	4.8610%	4.7995%
$\widehat{\boldsymbol{\Sigma}}_f$	(0.8428%)	(0.8152%)	(0.8740%)	(0.8043%)	(0.8570%)	(0.8021%)
RE_2	4.8918%	4.8158%	4.9015%	4.7967%	4.8636%	4.8032%
$\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}$	(0.8443%)	(0.8177%)	(0.8746%)	(0.8039%)	(0.8583%)	(0.8015%)

(3) H-CLUB slightly increases with larger N, but its degree of increase is much smaller than the crude bound $\widehat{\xi}_T$.

In order to further justify the validity of $\widehat{U}(\epsilon)$ as an H-CLUB, we investigate the empirical coverage probabilities for all three estimators. We employ same settings as in Figs. 2 and 3, plot average absolute distances between the empirical coverage probability and the nominal level 95% against dimensionality N in Fig. 4. The average absolute distances vary from 1% to 5%, which is in an acceptable range. This is in line with our expectation which ensures $\widehat{U}(\epsilon)$ as a valid H-CLUB of Δ.

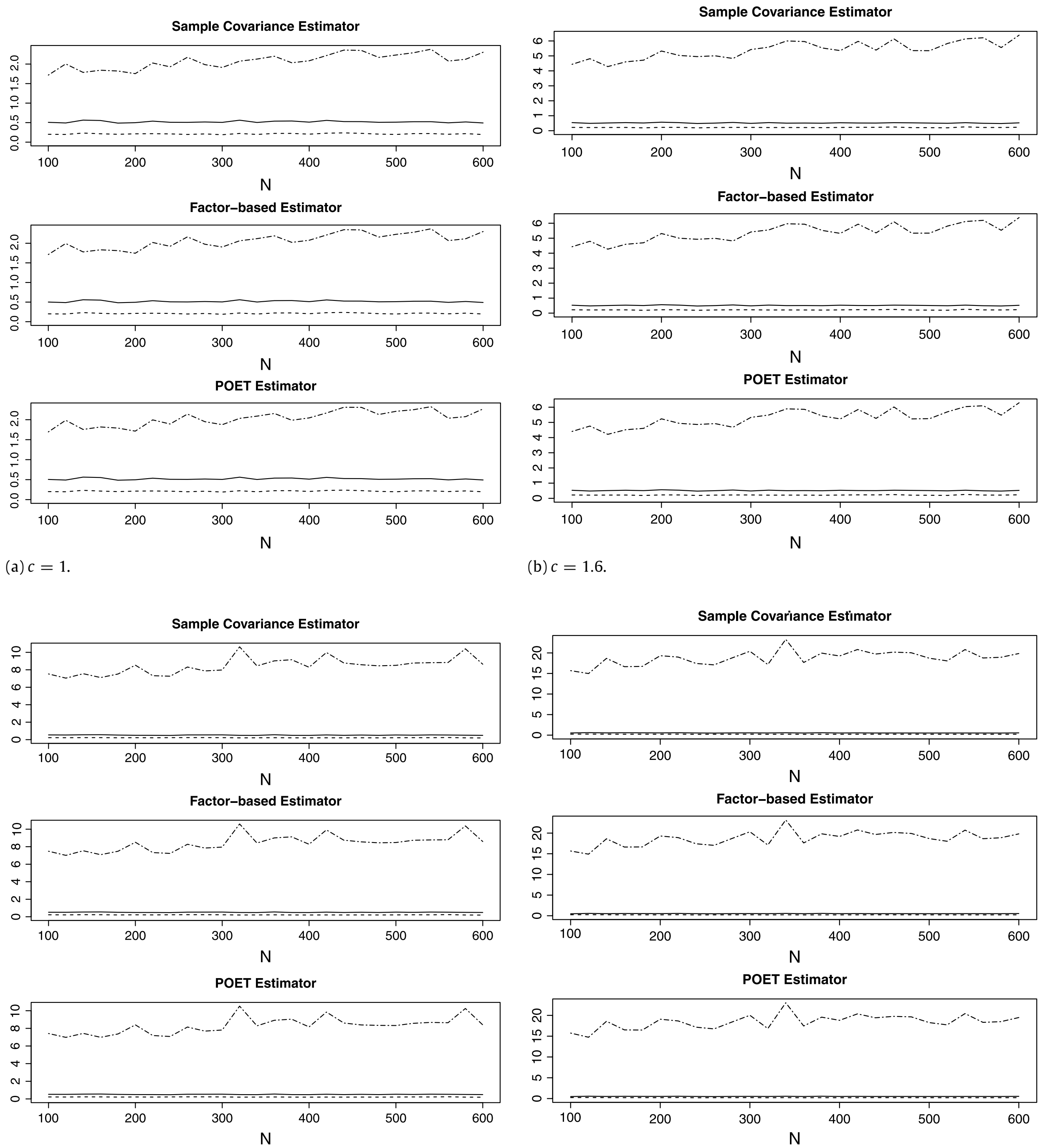
Means and standard deviations (in parentheses) of RE_1 for all three kinds of estimators are summarized in Table 3. Here we fix $N = 600$ and $T = 300$ with 500 replications. The ratio RE_1 quantifies the relation between the full confidence bound and the H-CLUB. Numerical results justify our observations in Fig. 2 in the sense that $\widehat{\xi}_T$ is in general many times greater than $\widehat{U}(\epsilon)$. Moreover, RE_1 increases dramatically as the exposure c increases.

Under the same setting, we also look at values of RE_1 under multiple choices of ϵ. For illustrating purposes, only POET-based estimator with two gross exposures are considered. Means and standard deviations are summarized in Table 4. As ϵ decreases, it is not difficult to identify simultaneous decline trends of RE_1. This is due to the fact that $\widehat{U}(\epsilon)$ grows as the confidence level increases.

Averages and standard deviations of relative error (see e.g. Corollary 4.1)

$$\mathrm{RE}_2 = \sqrt{\widehat{\mathrm{var}}(\mathbf{w}_T'\widehat{\boldsymbol{\Sigma}}\mathbf{w}_T)}/(2\mathbf{w}_T'\boldsymbol{\Sigma}\mathbf{w}_T)$$

with two choices of T are summarized in Tables 5 and 6, respectively. RE_2 measures the accuracy of the perceived risk $\widehat{R}(\mathbf{w}_T)^{\frac{1}{2}}$ with respect to the true risk $R(\mathbf{w}_T)^{\frac{1}{2}}$. Indeed by delta's method, $\mathrm{RE}_2 \approx \mathrm{ASD}(\widehat{R}(\mathbf{w}_T)^{\frac{1}{2}})/R(\mathbf{w}_T)^{\frac{1}{2}}$, where "ASD" stands for asymptotic standard deviation. From both tables, it is not difficult to observe that standard deviations are small when compared to their corresponding means. The results also show that the relative error are negligible, at around 3%–5%, ensuring the estimate of $R(\mathbf{w}_T)$ a high level of accuracy. More interestingly, we realize that this ratio is approximately independent of the gross exposure c but sensitive to the length of the time series. RE_2 steadily decreases as T grows. We also observe from Tables 3–6 that the asymptotic variances (reflected by $\widehat{U}(\epsilon)$) of the estimators based on known and unknown factors are almost the same, and slightly smaller than that of the sample covariance estimator.



(a) $c = 1$.

(b) $c = 1.6$.

(c) $c = 2$.

(d) $c = 3$.

Fig. 2. Averages of Δ (dashed curve), $\widehat{U}(\epsilon)$ with $\epsilon = 0.05$ (solid curve) and $\widehat{\xi}_T$ (two-dash curve) over 10,000 portfolios for $c = 1$, 1.6, 2 and 3, based on three estimated covariance estimators.

Table 6
Averages and standard deviations of RE_2 over 500 replications, with $T = 400$.

	$c = 1$	$c = 1.2$	$c = 1.4$	$c = 1.6$	$c = 1.8$	$c = 2$
RE_2 $\mathbf{S}$	3.4783% (0.4438%)	3.4684% (0.4215%)	3.4831% (0.4611%)	3.4807% (0.4691%)	3.4979% (0.4392%)	3.4793% (0.4479%)
RE_2 $\widehat{\mathbf{\Sigma}}_f$	3.4708% (0.4425%)	3.4607% (0.4183%)	3.4742% (0.4626%)	3.4659% (0.4649%)	3.4772% (0.4370%)	3.4553% (0.4438%)
RE_2 $\widehat{\mathbf{\Sigma}}_{P,\widehat{K}}$	3.4745% (0.4431%)	3.4634% (0.4170%)	3.4775% (0.4617%)	3.4708% (0.4665%)	3.4811% (0.4380%)	3.4586% (0.4431%)

In addition, in order to examine the sensitivity to over-estimating K in the POET-based estimator, we compute RE_2 with multiple choices of $\widehat{K}$ (the true K equals 3). Averages and standard deviations of RE_2 are summarized in Tables 7 and 8.

In both tables, the average RE_2 for $\widehat{K} = 3 \ldots 7$ are very close to each other, and hence the POET-based risk estimator is robust to over-estimating K.

6. Empirical studies

We assess the performance of H-CLUB in a portfolio allocation. We use the daily excess returns of 100 industrial portfolios formed

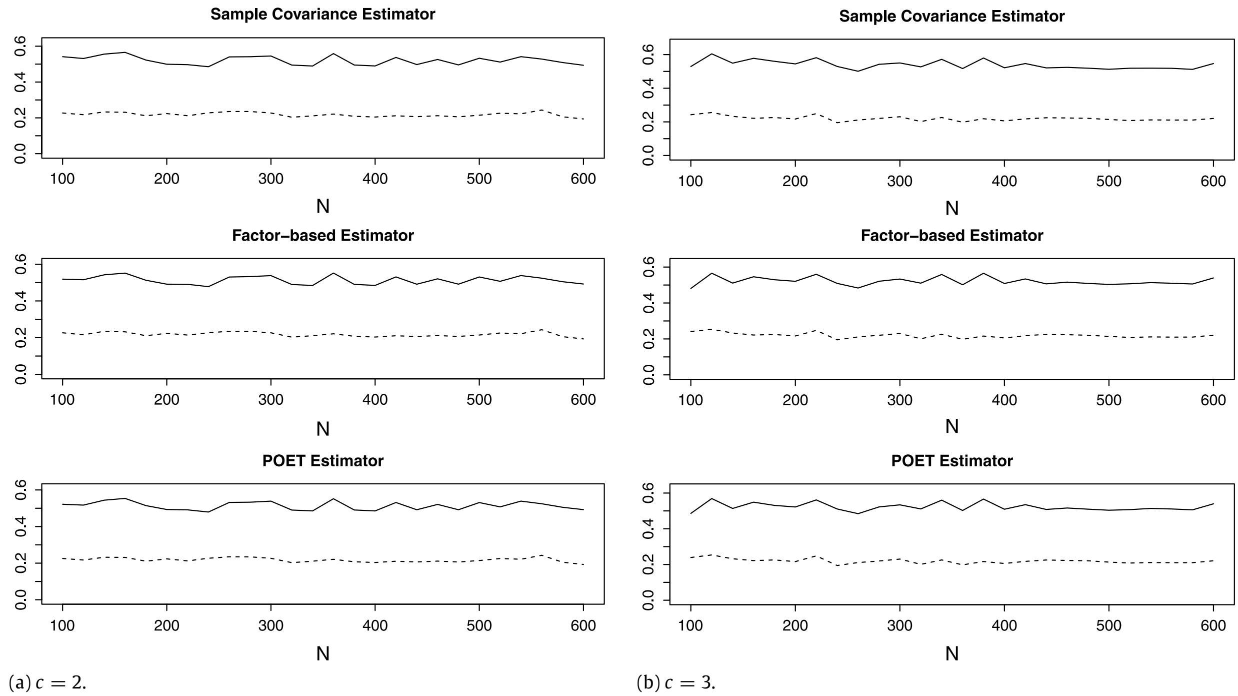


Fig. 3. Averages of Δ (dashed curve), $\widehat{U}(\epsilon)$ with $\epsilon = 0.05$ (solid curve) over 10,000 portfolios. This is a zoom version of Fig. 2.

Table 7
Averages and standard deviations of RE_2 over 500 replications, with $T = 200$.

	$c = 1$	$c = 1.2$
RE_2 $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}(\widehat{K} = 3)$	4.8918% (0.8443%)	4.8158% (0.8177%)
RE_2 $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}(\widehat{K} = 4)$	4.8918% (0.8439%)	4.8154% (0.8178%)
RE_2 $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}(\widehat{K} = 5)$	4.8921% (0.8439%)	4.8155% (0.8178%)
RE_2 $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}(\widehat{K} = 6)$	4.8925% (0.8436%)	4.8144% (0.8185%)
RE_2 $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}(\widehat{K} = 7)$	4.8922% (0.8434%)	4.8145% (0.8192%)

Table 8
Averages and standard deviations of RE_2 over 500 replications, with $T = 400$.

	$c = 1$	$c = 1.2$
RE_2 $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}(\widehat{K} = 3)$	3.4745% (0.4431%)	3.4634% (0.4170%)
RE_2 $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}(\widehat{K} = 4)$	3.4750% (0.4430%)	3.4635% (0.4170%)
RE_2 $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}(\widehat{K} = 5)$	3.4750% (0.4431%)	3.4639% (0.4171%)
RE_2 $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}(\widehat{K} = 6)$	3.4752% (0.4433%)	3.4636% (0.4172%)
RE_2 $\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}}(\widehat{K} = 7)$	3.4750% (0.4434%)	3.4635% (0.4175%)

on the size and book to market ratio from the website of Kenneth French. The study period is from July 1st 2008 to June 29th 2012, which spans a total of 1000 trading days. At the end of each month the covariance matrix is estimated by three estimators, the sample covariance, the factor-based estimator, and the POET estimator, using daily returns of the preceding 12 months ($T = 252$). In particular, we employ the Fama–French three-factor model to construct the factor-based estimator. As in Section 5, we adopt plug-in method to estimate the time lags L, refer to Section 5.3 for more details. Two types of strategies are tested, namely the equally weighted portfolio, and the minimum variance portfolio. The optimal portfolios are constructed under two exposure constraints ($c = 1$ and $c = 1.6$). The equally weighted portfolio is given by $\widehat{\mathbf{w}}_T = (1/N, \ldots, 1/N)$. The data-dependent minimum variance portfolio is given by

$$\widehat{\mathbf{w}}_T = \operatorname*{argmin}_{\mathbf{w}_T'\mathbf{1}=1, \|\mathbf{w}_T\|_1=c} \mathbf{w}_T'\widehat{\boldsymbol{\Sigma}}\mathbf{w}_T.$$

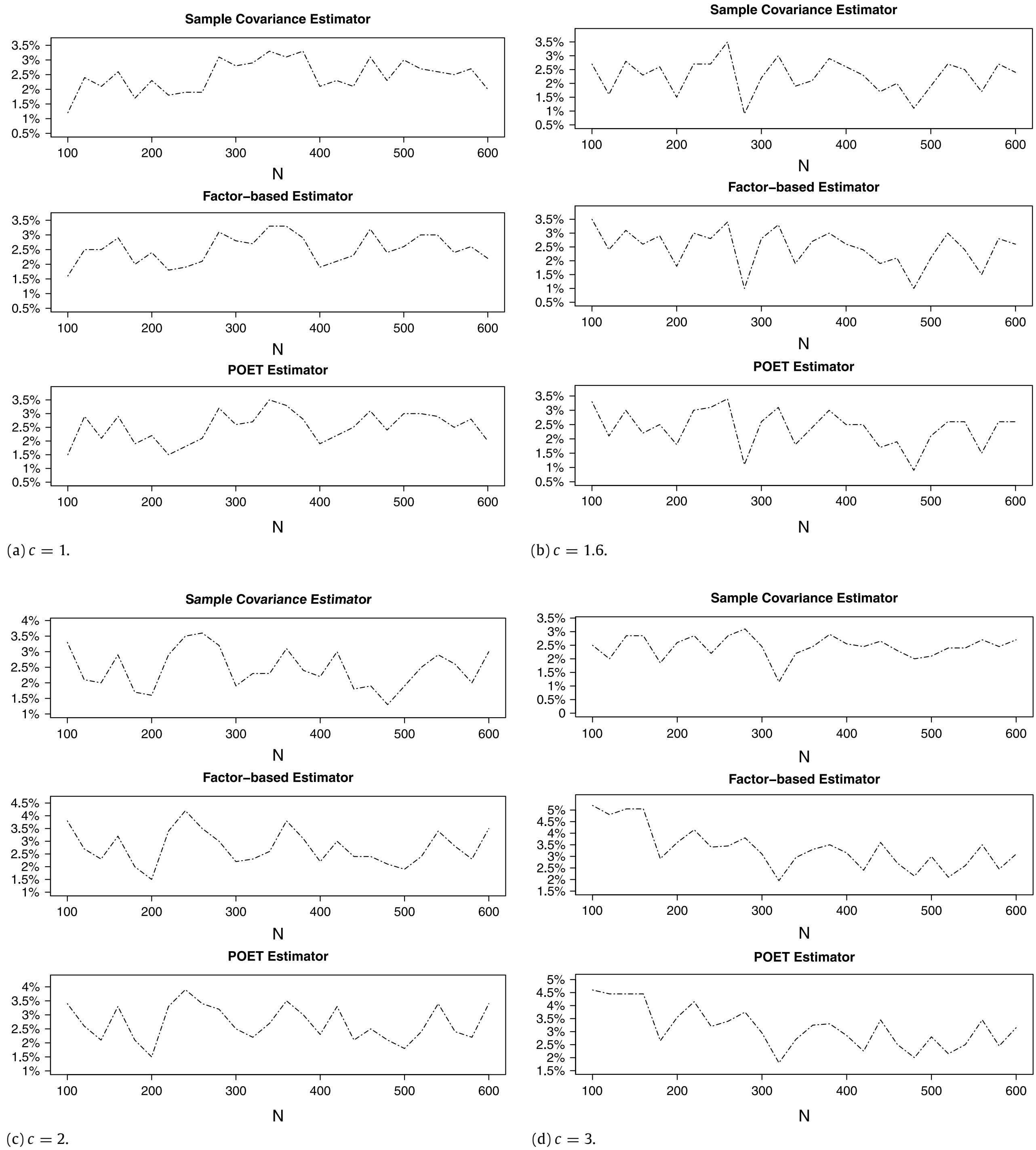
Portfolios are held for one month and rebalanced at the beginning of the next month.

Because the "true" volatility matrix $\boldsymbol{\Sigma}$ is unknown, to illustrate our method on the real data, at any time point, we look one month ahead and define the true $\boldsymbol{\Sigma}$ to be the sample covariance of daily excess returns over the future month. For example, at day 10, the corresponding $\boldsymbol{\Sigma}$ and $R(\widehat{\mathbf{w}}_T)$ are defined to be

$$\boldsymbol{\Sigma} = \frac{1}{21}\sum_{t=11}^{31}\mathbf{R}_t\mathbf{R}_t', \qquad R(\widehat{\mathbf{w}}_T) = \left(\widehat{\mathbf{w}}_T'\boldsymbol{\Sigma}\widehat{\mathbf{w}}_T\right)^{1/2},$$

where $\widehat{\mathbf{w}}_T$ is the portfolio defined above. On the other hand, the "sample covariance matrix" $\mathbf{S}$ in Table 9 below is the sample covariance of daily excess returns over the past twelve months. This is aggregated over the entire testing period. We choose $\epsilon = 0.01$, i.e. $\widehat{U}(\epsilon)$ is an empirical 99% upper bound. For each covariance matrix estimator and strategy, we study five quantities, whose respective averages over the whole study period are summarized in Table 9. Recall that $\widehat{U}(\epsilon)/\sqrt{4\widehat{\mathbf{w}}_T'\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{w}}_T}$ is the H-CLUB for the true risk error $|\widehat{R}(\widehat{\mathbf{w}}_T) - R(\widehat{\mathbf{w}}_T)|$, see, for example, Corollary 4.1. All risks are annualized.

By comparing the first two columns in Table 9, we observe that $\widehat{U}(\epsilon)$ is uniformly greater that Δ, regardless of the strategies and the covariance matrix estimators. Moreover, the true risk errors are very large, around 40%–50% of the true risk. There are two possible causes: The sample covariance based on one month data (holding period) can be different from the true covariance matrix. It incurs



(a) $c = 1$.

(b) $c = 1.6$.

(c) $c = 2$.

(d) $c = 3$.

Fig. 4. Average absolute distances between empirical coverage probabilities and the nominal level (95%).

substantial sampling variability. The second cause can be the non-stationarity of the financial returns. However, as shown in the two rightmost columns, results are still satisfactory in the sense that the U-CLUB's are uniformly larger but quite close ($<1\%$ per annum) to the true risk error.

7. Conclusions

In this paper we address the estimation and assessment for the risk of a large portfolio. The risk is estimated by a substitution of a good estimator of the volatility matrix. We study factor-based risk estimators, based on the approximate factor model with known factors and unknown factors. We derive the limiting distribution of the estimated risks under high dimensionality.

Given that the existing upper bound for the risk estimation error is too crude and not applicable in practice, we introduce a new method, H-CLUB, to assess the accuracy of the risk estimation based on the confidence intervals. Our numerical results demonstrate that the proposed upper bounds significantly outperform the traditional crude bounds, and provide insightful assessment of the estimation of the true portfolio risks.

Table 9
True risk errors and estimated risk errors based on the 100 Fama–French Industrial Portfolios.

Strategy	Average of $\Delta(\times 10^{-4})$	Average of $\widehat{U}(0.01)(\times 10^{-4})$	Average of true risk	True risk error	H-CLUB
Sample-based covariance estimator					
Equal weighted	2.356	2.757	20.81%	11.18%	11.37%
Min variance ($c = 1$)	1.006	1.233	14.38%	7.00%	7.44%
Min variance ($c = 1.6$)	0.497	0.621	11.58%	4.69%	5.17%
Factor-based covariance estimator					
Equal weighted	2.352	2.768	20.81%	11.16%	11.39%
Min variance ($c = 1$)	0.999	1.226	14.45%	6.95%	7.41%
Min variance ($c = 1.6$)	0.475	0.594	11.79%	4.52%	4.98%
POET-based covariance estimator					
Equal weighted	2.353	2.768	20.81%	11.17%	11.38%
Min variance ($c = 1$)	1.005	1.231	14.38%	6.99%	7.43%
Min variance ($c = 1.6$)	0.490	0.626	11.59%	4.61%	5.22%

Here $\Delta = |\widehat{\mathbf{w}}_T'(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}})\widehat{\mathbf{w}}_T|$ and $\widehat{U}(0.01) = 2.58(\widehat{\text{var}}(\widehat{\mathbf{w}}_T'\widehat{\mathbf{\Sigma}}\widehat{\mathbf{w}}_T))^{1/2}$. True risk is $R(\widehat{\mathbf{w}}_T)$. The True Risk Error is $|(\widehat{\mathbf{w}}_T'\mathbf{\Sigma}\widehat{\mathbf{w}}_T)^{1/2} - (\widehat{\mathbf{w}}_T'\widehat{\mathbf{\Sigma}}\widehat{\mathbf{w}}_T)^{1/2}|$ and H-CLUB is $\widehat{U}(0.01)/\sqrt{4\widehat{\mathbf{w}}_T'\widehat{\mathbf{\Sigma}}\widehat{\mathbf{w}}_T}$ respectively.

The empirical study suggests that the financial excess returns may not be globally stationary. Our method also allows for locally stationary time series, as well as slowly time-varying covariance matrices through the localization in time (time-domain smoothing). In fact, the empirical study analyzes a kind of time-varying model, in which we estimate covariance matrices based on the data in rolling windows.

Acknowledgments

The research was partially supported by DMS-1206464 and NIH R01GM100474-01, NIH R01-GM072611.

Appendix A. Proofs for the sample covariance

In this section, $Z_{T,t} = \mathbf{w}_T'\mathbf{R}_t\mathbf{R}_t'\mathbf{w}_T - E\mathbf{w}_T'\mathbf{R}_t\mathbf{R}_t'\mathbf{w}_T$, and $\gamma_T(h) = EZ_{T,t}Z_{T,t+h}$. In particular, $\gamma_T(0) = \text{var}(Z_{T,t})$.

We first prove that for $\widehat{\mathbf{\Sigma}} = \mathbf{S}, \widehat{\mathbf{\Sigma}}_f, \widehat{\mathbf{\Sigma}}_{P,\widehat{K}}$, the risk estimator $\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T = O_p(1)$.

Lemma A.1. $\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T = O_p(1)$ *and* $\widehat{\mathbf{w}}_T'\widehat{\mathbf{\Sigma}}\widehat{\mathbf{w}}_T = O_p(1)$.

Proof. It is true that for $\widehat{\mathbf{\Sigma}} = \mathbf{S}, \widehat{\mathbf{\Sigma}}_f$, or $\widehat{\mathbf{\Sigma}}_{P,\widehat{K}}$, we have $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} = o_p(1)$ (see Theorem 3.2 of Fan et al., 2013). Hence $\|\widehat{\mathbf{\Sigma}}\|_{\max} \le \|\mathbf{\Sigma}\|_{\max} + \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} = O_p(1)$. In addition, since $\|\mathbf{w}_T\|_1 = O(1)$, we have $\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T \le \|\mathbf{w}_T\|_1^2\|\widehat{\mathbf{\Sigma}}\|_{\max} = O_p(1)$. Finally, because $\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1 = o_p(1)$, $\|\widehat{\mathbf{w}}_T\|_1 = O_p(1)$. The result then also follows from the inequality

$$\widehat{\mathbf{w}}_T'\widehat{\mathbf{\Sigma}}\widehat{\mathbf{w}}_T \le \|\widehat{\mathbf{w}}_T\|_1^2\|\widehat{\mathbf{\Sigma}}\|_{\max} = O_p(1). \quad \blacksquare$$

A.1. Proof of Lemma 4.1

Lemma A.2. $\max_{t\le T}\|\mathbf{f}_t\| = O_p((\log T)^{1/r_2})$, $\max_{i\le N,t\le T}|u_{it}| = O_p((\log NT)^{1/r_1})$, $\max_{i\le N,t\le T}|R_{it}| = O_p((\log T)^{1/r_2} + (\log NT)^{1/r_1})$.

Proof. Let $s = b_1(2\log NT)^{1/r_1}$. Then

$$P(\max_{i\le N,t\le T}|u_{it}| > s) \le NT\exp(-(s/b_1)^{r_1})$$
$$\le \exp(\log NT - (s/b_1)^{r_1}) \to 0.$$

Hence $\max_{i\le N,t\le T}|u_{it}| = O_p((\log NT)^{1/r_1})$. Similarly, $\max_{t\le T}\|\mathbf{f}_t\| = O_p((\log T)^{1/r_2})$.

$$\max_{i,t}|R_{it}| \le \max_i\|\mathbf{b}_i\|\max_t\|\mathbf{f}_t\| + \max_{i\le N,t\le T}|u_{it}|$$
$$= O_p((\log T)^{1/r_2} + (\log NT)^{1/r_1}). \quad \blacksquare$$

Lemma A.3. (i) $|(\mathbf{w}_T'\mathbf{S}\mathbf{w}_T)^2 - (\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T)^2| = O_p(T^{-1/2}\sigma_T)$.

(ii) $\max_{h\le L}|T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\mathbf{R}_t)^2(\mathbf{w}_T'\mathbf{R}_{t+h})^2 - E(\mathbf{w}_T'\mathbf{R}_t)^2(\mathbf{w}_T'\mathbf{R}_{t+h})^2| = O_p(\sqrt{L/T})$.

(iii) $\max_{h\le L}|\mathbf{w}_T'\mathbf{S}\mathbf{w}_T - T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\mathbf{R}_t)^2| = O_p(L^2\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T/T)$.

(iv) $\max_{h\le L}|\mathbf{w}_T'\mathbf{S}\mathbf{w}_T - T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\mathbf{R}_{t+h})^2| = O_p(L^2\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T/T)$.

Proof. Note that for any $N\times N$ matrix $\mathbf{A} = (a_{ij})$, $|\mathbf{w}_T'\mathbf{A}\mathbf{w}_T| \le \|\mathbf{A}\|_{\max}\|\mathbf{w}_T\|_1^2$. Thus

$$|(\mathbf{w}_T'\mathbf{S}\mathbf{w}_T)^2 - (\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T)^2| \le \|\mathbf{S} + \mathbf{\Sigma}\|_{\max}\|\mathbf{w}_T\|_1^2|\mathbf{w}_T'(\mathbf{S} - \mathbf{\Sigma})\mathbf{w}_T|$$
$$= O_p(|\mathbf{w}_T'(\mathbf{S} - \mathbf{\Sigma})\mathbf{w}_T|) = O_p\left(\left|T^{-1}\sum_{t=1}^{T}Z_{T,t}\right|\right).$$

The Chebyshev inequality implies $|T^{-1}\sum_{t=1}^{T}Z_{T,t}| = O_p(T^{-1/2}\sqrt{\sigma_T^2})$.

(ii) Let $X_{t,h} = (\mathbf{w}_T'\mathbf{R}_t)^2(\mathbf{w}_T'\mathbf{R}_{t+h})^2$. By the Chebyshev inequality, for any $s > 0$,

$$P\left(\max_{h\le L}\left|\frac{1}{T}\sum_{t=1}^{T}X_{t,h} - EX_{t,h}\right| > s\right)$$
$$\le L\max_{h\le L}P\left(\left|\frac{1}{T}\sum_{t=1}^{T}X_{t,h} - EX_{t,h}\right| > s\right) \le \frac{L\max_{h\le L}\text{var}\left(\sum_{t=1}^{T}X_{t,h}\right)}{T^2s^2}.$$

Note that $\max_{h\le L}\text{var}(\sum_{t=1}^{T}X_{t,h}) = O(T)$ since $\max_{h\le L}\text{var}(X_{t,h}) = O(1)$ and $\max_{h\le L}\sum_{t=1}^{T}\text{cov}(X_{1,h}, X_{t+1,h}) = O(1)$. Therefore, for arbitrarily small $\epsilon > 0$, by choosing $s > \sqrt{LM/(\epsilon T)}$, $P(\max_{h\le L}|\frac{1}{T}\sum_{t=1}^{T}X_{t,h} - EX_{t,h}| > s) < \epsilon$, which implies $\max_{h\le L}|\frac{1}{T}\sum_{t=1}^{T}X_{t,h} - EX_{t,h}| = O_p(\sqrt{L/T})$. In addition,

$$P(\max_{h\le L}|X_{t,h} - EX_{t,h}| > (L\max_{h\le L}\text{var}(X_{t,h})/\epsilon)^{1/2})$$
$$\le \frac{L\max_{h\le L}\text{var}(X_{t,h})}{L\max_{h\le L}\text{var}(X_{t,h})/\epsilon} = \epsilon.$$

Thus $\max_{h\le L}|X_{t,h} - EX_{t,h}| = O_p(\sqrt{L})$, which implies

$$\max_{h\le L}\left|\frac{1}{T}\sum_{t=1}^{T-h}X_{t,h} - EX_{t,h}\right| \le O_p(\sqrt{L/T}) + \max_{h\le L}|X_{t,h} - EX_{t,h}|L/T$$
$$= O_p\left(\sqrt{L/T}\right).$$

(iii) The left hand side is $\max_{h\le L}T^{-1}\sum_{t=T-h+1}^{T}(\mathbf{w}_T'\mathbf{R}_t)^2 = \max_{T-L+1\le t\le T}(\mathbf{w}_T'\mathbf{R}_t)^2L/T$. For any $s > 0$, $P(\max_{T-L+1\le t\le T}(\mathbf{w}_T'\mathbf{R}_t)^2$

$> s) \leq LP((\mathbf{w}_T'\mathbf{R}_t)^2 > s) \leq L\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T/s$, which then implies $\max_{T-L+1\leq t\leq T}(\mathbf{w}_T'\mathbf{R}_t)^2 = O_p(L\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T)$. The desired result then follows.

(iv) A similar argument as above shows $\max_{T+1\leq t\leq T+L}(\mathbf{w}_T'\mathbf{R}_t)^2 = O_p(L\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T)$. Hence $\max_{h\leq L} T^{-1}\sum_{t=T-h+1}^{T}(\mathbf{w}_T'\mathbf{R}_{t+h})^2 \leq \max_{T+1\leq t\leq T+L}(\mathbf{w}_T'\mathbf{R}_t)^2 L/T = O_p(L^2\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T/T)$. This implies that the desired quantity is bounded by $a + O_p(L^2\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T/T)$ where

$$a = \max_{h\leq L}\left|\frac{1}{T}\sum_{t=1}^{T}[(\mathbf{w}_T\mathbf{R}_t)^2 - (\mathbf{w}_T'\mathbf{R}_{t+h})^2]\right| \leq \left|\frac{1}{T}\sum_{t=1}^{L}(\mathbf{w}_T'\mathbf{R}_t)^2\right| + \left|\frac{1}{T}\sum_{t=1}^{L}(\mathbf{w}_T'\mathbf{R}_{T+t})^2\right|.$$

Note that $|\frac{1}{T}\sum_{t=1}^{L}(\mathbf{w}_T'\mathbf{R}_t)^2| \leq \max_{1\leq t\leq L}(\mathbf{w}_T'\mathbf{R}_t)^2 L/T = O_p(L^2\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T/T)$. Similarly we have $|\frac{1}{T}\sum_{t=1}^{L}(\mathbf{w}_T'\mathbf{R}_{T+t})^2| = O_p(L^2\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T/T)$. ■

Lemma A.4. $\max_{h\leq L}|\widehat{\gamma}(h) - \gamma_T(h)| = O_p(\sqrt{L/T} + \|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1((\log T)^{4/r_2} + (\log NT)^{4/r_1}))$.

Proof. The triangular inequality implies $\max_{h\leq L}|\widehat{\gamma}(h) - \gamma_T(h)| \leq \sum_{i=1}^{8} a_i$, where

$$a_1 = \max_{h\leq L}\left|T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\mathbf{R}_t)^2(\mathbf{w}_T'\mathbf{R}_{t+h})^2 - E(\mathbf{w}_T'\mathbf{R}_t)^2(\mathbf{w}_T'\mathbf{R}_{t+h})^2\right|,$$

$$a_2 = |(\mathbf{w}_T'\mathbf{S}\mathbf{w}_T)^2 - (\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T)^2|,$$

$$a_3 = \mathbf{w}_T'\mathbf{S}\mathbf{w}_T \max_{h\leq L}\left|\mathbf{w}_T'\mathbf{S}\mathbf{w}_T - T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\mathbf{R}_t)^2\right|,$$

$$a_4 = \mathbf{w}_T'\mathbf{S}\mathbf{w}_T \max_{h\leq L}\left|\mathbf{w}_T'\mathbf{S}\mathbf{w}_T - T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\mathbf{R}_{t+h})^2\right|,$$

$$a_5 = \max_{h\leq L}\left|\frac{1}{T}\sum_{t=1}^{T-h}(\mathbf{w}_T'\mathbf{R}_t)^2(\mathbf{w}_T'\mathbf{R}_{t+h})^2 - (\widehat{\mathbf{w}}_T'\mathbf{R}_t)^2(\widehat{\mathbf{w}}_T'\mathbf{R}_{t+h})^2\right|$$

$$a_6 = \max_{h\leq L}\left|\frac{1}{T}\sum_{t=1}^{T-h}(\mathbf{w}_T'\mathbf{R}_t)^2\mathbf{w}_T'\mathbf{S}\mathbf{w}_T - (\widehat{\mathbf{w}}_T'\mathbf{R}_t)^2\widehat{\mathbf{w}}_T'\mathbf{S}\widehat{\mathbf{w}}_T\right|$$

$$a_7 = \max_{h\leq L}\left|\frac{1}{T}\sum_{t=1}^{T-h}(\mathbf{w}_T'\mathbf{R}_{t+h})^2\mathbf{w}_T'\mathbf{S}\mathbf{w}_T - (\widehat{\mathbf{w}}_T'\mathbf{R}_{t+h})^2\widehat{\mathbf{w}}_T'\mathbf{S}\widehat{\mathbf{w}}_T\right|,$$

$$a_8 = |(\mathbf{w}_T'\mathbf{S}\mathbf{w}_T)^2 - (\widehat{\mathbf{w}}_T'\mathbf{S}\widehat{\mathbf{w}}_T)^2|.$$

We have, $\mathbf{w}_T'\mathbf{S}\mathbf{w}_T \leq |\mathbf{w}_T'(\mathbf{S}-\mathbf{\Sigma})\mathbf{w}_T| + \mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T = O_p(\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T + T^{-1/2}\sigma_T^2)$. It then follows from Lemma A.3 and $\sigma_T^2 = O(1)$, $L^3 = O(T)$, $\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T = O(1)$ that $a_i = O_p(\sqrt{L/T})$ for $i = 1\ldots 4$. In addition,

$$\begin{aligned}
a_5 &\leq \|\mathbf{w}_T + \widehat{\mathbf{w}}_T\|_1(\|\widehat{\mathbf{w}}_T\|_1^2 + \|\mathbf{w}_T\|_1^2)\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1(\max_{i\leq N, t\leq T}|R_{it}|)^4 \\
&= O_p(\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1((\log T)^{4/r_2} + (\log NT)^{4/r_1})), \\
a_6 &\leq \|\mathbf{w}_T - \widehat{\mathbf{w}}_T\|_1 \max_{it}|R_{it}|^2\|\mathbf{S}\|_{\max}(\|\mathbf{w}_T\|_1^3 + \|\widehat{\mathbf{w}}_T\|_1^3 \\
&\quad + \|\mathbf{w}\|_1^2\|\widehat{\mathbf{w}}_T\| + \|\widehat{\mathbf{w}}_T\|_1^2\|\mathbf{w}_T\|_1) \\
&= O_p(\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1((\log T)^{2/r_2} + (\log NT)^{2/r_1})).
\end{aligned}$$

Term a_7 is bounded in the same way as a_6. Finally,

$$\begin{aligned}
a_8 &\leq (\|\mathbf{w}_T\|_1^2 + \|\widehat{\mathbf{w}}_T\|_1^2)(\|\mathbf{w}_T\|_1 + \|\widehat{\mathbf{w}}_T\|_1)\|\mathbf{S}\|_{\max}^2\|\mathbf{w}_T - \widehat{\mathbf{w}}_T\|_1 \\
&= O_p(\|\mathbf{w}_T - \widehat{\mathbf{w}}_T\|_1),
\end{aligned}$$

which implies $\max_{h\leq L}|\widehat{\gamma}(h) - \gamma_T(h)| = O_p(\sqrt{L/T})$. ■

Proof of Lemma 4.1. By the triangular inequality, $|\widehat{\sigma}^2 - \sigma_T^2| \leq \sum_{i=1}^{3} b_i$, where

$$b_1 = |\widehat{\gamma}(0) - \gamma_T(0)|, \qquad b_2 = 2\sum_{h=1}^{L}(1 - h/L)|\widehat{\gamma}(h) - \gamma_T(h)|,$$

$$b_3 = 2\sum_{h>L}\gamma_T(h), \qquad b_4 = \frac{2}{L}\sum_{h=1}^{L}h\gamma_T(h)$$

$b_2 \leq 2L\max_{h\leq L}|\widehat{\gamma}(h) - \gamma_T(h)| = O_p(L\sqrt{L/T} + L\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1((\log T)^{4/r_2} + (\log NT)^{4/r_1}))$. Then

$$|\widehat{\sigma}^2 - \sigma_T^2| = O_p\Bigg(L^{3/2}T^{-1/2} + \frac{1}{L}\sum_{h=1}^{L}h\gamma(h) + \sum_{h>L}\gamma_T(h) + L\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1((\log T)^{4/r_2} + (\log NT)^{4/r_1})\Bigg).$$

By the lemma's conditions and Assumption 4.7, the first, third and the fourth terms are all stochastically dominated by σ_T^2. As for the second term $\frac{1}{L}\sum_{h=1}^{L}h\gamma(h)$, note that $\sum_{h=1}^{\infty}|\gamma_T(h)|/\sigma_T^2 < \infty$. Then by dominated convergence theorem, $\lim_T\frac{1}{L}\sum_{h=1}^{L}\frac{h}{\sigma_T^2}\gamma_T(h) = \sum_{h=1}^{\infty}\lim_T h\gamma_T(h)\frac{1}{\sigma_T^2 L}I(h\leq L) = 0$ given that $L\sigma_T^2 \to \infty$.

The second part $\widehat{U_S}(\epsilon) = o(\sqrt{\log N/T})$ is due to $\widehat{\sigma}^2 = O_p(\sigma_T^2)$ as $|\sigma_T^2 - \widehat{\sigma}^2| = o_p(\sigma_T^2)$ and $\sigma_T^2 = O(1) = o(\log N)$, as $N\to\infty$. ■

A.2. Proof of Theorem 4.1

Lemma A.5. (i) $EZ_{T,1}^2 = O(1)$ *and* $\max_{l\leq T}|\gamma_T(l)| = O(1)$.

(ii) *For any* $K \in [m, T]$, $\mathrm{var}(\sum_{t=1}^{K}Z_{T,t}) = K\gamma_T(0) + 2K\sum_{h=1}^{K}(1 - h/K)\gamma_T(h) = O(K)$.

Proof. (i) It suffices to show $E(\mathbf{w}_T'\mathbf{R}_t)^4 = O(1)$. In fact by $\max_{i\leq N} ER_{it}^4 = O(1)$, $E(\mathbf{w}_T'\mathbf{R}_t)^4 = \sum_{ijkl=1}^{N}w_iw_jw_kw_lER_{it}R_{jt}R_{kt}R_{lt} \leq \max_{i\leq N} ER_{it}^4\|\mathbf{w}_T\|_1^4 = O(1)$. The second part follows immediately.

(ii) It is well known that for a stationary process with zero mean, $\mathrm{var}(K^{-1}\sum_{t=1}^{K}Z_{T,t}) = K^{-1}\gamma_T(0) + 2K^{-1}\sum_{h=1}^{K}(1 - h/K)\gamma_T(h)$, which implies the result. ■

Lemma A.6. *Under the assumptions of Theorem 4.1,*

$$\left[\mathrm{var}\left(\sum_{t=1}^{T}(\mathbf{w}_T'\mathbf{R}_t)^2\right)\right]^{-1/2} T\mathbf{w}_T'(\mathbf{S} - \mathbf{\Sigma})\mathbf{w}_T \to^d \mathcal{N}(0, 1). \qquad \text{(A.1)}$$

Proof. The proof is based on Theorem 2.1 of Peligrad (1996). We have $\sqrt{T}\mathbf{w}_T'(\mathbf{S} - \mathbf{\Sigma})\mathbf{w}_T = T^{-1/2}\sum_{t=1}^{T}Z_{T,t}$. Define $B_{T,K}^2 = \mathrm{var}(\sum_{t=1}^{K}Z_{T,t})$ and $B_T^2 = \mathrm{var}(\sum_{t=1}^{T}Z_{T,t}) = O(T)$. By Davydov's inequality (Proposition 2.5 of Fan and Yao, 2003 with $p = 1/2$ and $q = 1/4$), there are constants $M, M_1, M_2 > 0$ such that for any integer $h \geq 0$,

$$\begin{aligned}
|\gamma_T(h)| &\leq 8\alpha(|h|)^{1/4}(E(\mathbf{w}_T'\mathbf{R}_t)^2)^{1/2}(E(\mathbf{w}_T'\mathbf{R}_t)^4)^{1/4} \\
&= M_2\exp(-M|h|^{r_3}/4)
\end{aligned}$$

where the last equality follows from the α-mixing condition and that $E(\mathbf{w}_T'\mathbf{R}_t)^4 = O(1)$. By the assumption that $\alpha_R(T) = o(\gamma_T(0))$, the correlation $\rho(T) = |\,\mathrm{Corr}(Z_{T,t}, Z_{T,t+T})| \leq |\gamma_T(T)|/\gamma_T(0) \leq \exp(-MT^{r_3}/4)$. To apply Theorem 2.1 of Peligrad (1996), we also need to check

$$\limsup_T \frac{1}{B_T^2}\sum_{t=1}^{T}EZ_{T,t}^2 < \infty \qquad \text{(A.2)}$$

and the Lindeberg condition: $\forall\epsilon > 0$, $\frac{1}{B_T^2}\sum_{t=1}^T EZ_{T,t}^2 I(|Z_{T,t}| > \epsilon B_T) \to 0$.

Note that $\frac{1}{T}B_T^2/\gamma_T(0) = [1+2\sum_{h=1}^T \rho(h)](1+o(1))$. Hence (A.2) holds because $1+2\sum_{h=1}^T \rho_T(h)$ is bounded away from zero. To verify the Lindeberg condition, note that there are $c_1, c_2 > 0$, for all large T, $c_1 < \frac{1}{T}B_T^2/\gamma_T(0) < c_2$. Then for any $\epsilon > 0$, for some $c > 0$ and all large T,

$$\frac{1}{B_T^2}\sum_{t=1}^T EZ_{T,t}^2 I(|Z_{T,t}| > \epsilon B_T) \le \frac{c}{\gamma(0)}EZ_{T,t}^2 I(Z_{T,t}^2 > T\gamma(0)\epsilon^2 c_1).$$

Because $\frac{EZ_{T,t}^2}{\gamma(0)} < \infty$, hence by the dominated convergence theorem,

$$\lim_T \frac{c}{\gamma(0)}EZ_{T,t}^2 I(Z_{T,t}^2 > T\gamma(0)\epsilon^2 c_1)$$
$$= cE\lim_T \frac{1}{\gamma(0)}Z_{T,t}^2 I(Z_{T,t}^2 > T\gamma(0)\epsilon^2 c_1) = 0.$$

Hence the conditions of Theorem 2.1 of Peligrad (1996) are satisfied, which implies $B_T^{-1}\sum_{t=1}^T Z_{T,t} \to^d \mathcal{N}(0,1)$, equivalent to (A.1). ■

Proofs of Theorem 4.1 and Corollary 4.1. Note that $T\sigma_T^2/B_T^2 \to 1$, we have

$$\frac{T}{B_T}[\widehat{\mathbf{w}}_T'(\mathbf{S}-\boldsymbol{\Sigma})\widehat{\mathbf{w}}_T - \mathbf{w}_T'(\mathbf{S}-\boldsymbol{\Sigma})\mathbf{w}_T]$$
$$\le \frac{T}{B_T}\|\mathbf{S}-\boldsymbol{\Sigma}\|_{\max}\|\widehat{\mathbf{w}}_T - \mathbf{w}\|_1(\|\widehat{\mathbf{w}}\|_1 + \|\mathbf{w}\|_1)$$
$$\le O_p\left(\frac{\sqrt{T}}{\sigma_T}\sqrt{\frac{\log N}{T}}\right)o_p(\sigma_T(\log N)^{-4/r_1}) = o_p(1).$$

Hence by (A.1), $\frac{T}{B_T}[\widehat{\mathbf{w}}_T'(\mathbf{S}-\boldsymbol{\Sigma})\widehat{\mathbf{w}}_T] \to \mathcal{N}(0,1)$. This also implies

$$\sqrt{\frac{T}{\sigma_T^2}}\widehat{\mathbf{w}}_T'(\mathbf{S}-\boldsymbol{\Sigma})\widehat{\mathbf{w}}_T \to^d \mathcal{N}(0,1) \tag{A.3}$$

which also implies $\widehat{\mathbf{w}}_T'(\mathbf{S}-\boldsymbol{\Sigma})\widehat{\mathbf{w}}_T = O_p(T^{-1/2}\sqrt{\sigma_T^2})$. Moreover, since $|\sigma_T^2 - \widehat{\sigma}^2| = o_p(\sigma_T^2)$,

$$\sqrt{T}|\widehat{\mathbf{w}}_T'(\mathbf{S}-\boldsymbol{\Sigma})\widehat{\mathbf{w}}_T|\left|\frac{1}{\sqrt{\sigma_T^2}} - \frac{1}{\sqrt{\widehat{\sigma}^2}}\right|$$
$$= \sqrt{T}|\widehat{\mathbf{w}}_T'(\mathbf{S}-\boldsymbol{\Sigma})\widehat{\mathbf{w}}_T|\frac{|\widehat{\sigma}^2 - \sigma_T^2|}{\sigma_T^2(1+o_p(1))} = o_p(1).$$

It then follows from (A.3) that $\sqrt{T/\widehat{\sigma}^2}\widehat{\mathbf{w}}_T'(\mathbf{S}-\boldsymbol{\Sigma})\widehat{\mathbf{w}}_T \to^d \mathcal{N}(0,1)$, which gives the H-CLUB. Corollary 4.1 follows straightforward from applying the delta-method. ■

Appendix B. Proofs for the factor-based estimation

B.1. Proof of Lemma 4.2

Lemma B.1. $\max_{h\le L}|\widehat{\gamma}_f(h) - \gamma_f(h)| = O_p(\sqrt{(L+\log N)/T} + (\log T)^{4/r_2}\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1)$.

Proof. The triangular inequality implies $\max_{h\le L}|\widehat{\gamma}_f(h) - \gamma_f(h)| \le \sum_{i=1}^4 a_i$, where

$$a_1 = \max_{h\le L}\left|T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_{t+h})^2(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_t)^2 - E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2(\mathbf{w}_T'\mathbf{B}\mathbf{f}_{t+h})^2\right|,$$

$$a_2 = |(\mathbf{w}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\mathbf{w}_T)^2 - (\mathbf{w}_T'\mathbf{B}\,\mathrm{cov}(\mathbf{f}_t)\mathbf{B}'\mathbf{w}_T)^2|$$

$$a_3 = \mathbf{w}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\mathbf{w}_T \max_{h\le L}\left|\mathbf{w}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\mathbf{w}_T - T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\widehat{\mathbf{B}}\mathbf{f}_t)^2\right|,$$

$$a_4 = \mathbf{w}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\mathbf{w}_T \times \max_{h\le L}\left|\mathbf{w}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\mathbf{w}_T - T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\widehat{\mathbf{B}}\mathbf{f}_{t+h})^2\right|,$$

$$a_5 = \max_{h\le L}\left|T^{-1}\sum_{t=1}^{T-h}(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\mathbf{f}_{t+h})^2(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\mathbf{f}_t)^2 - (\mathbf{w}_T'\widehat{\mathbf{B}}\mathbf{f}_{t+h})^2(\mathbf{w}_T'\widehat{\mathbf{B}}\mathbf{f}_t)^2\right|,$$

$$a_6 = \max_{h\le L}\left|T^{-1}\sum_{t=1}^{T-h}(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\mathbf{f}_{t+h})^2(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\widehat{\mathbf{w}}_T) - (\mathbf{w}_T'\widehat{\mathbf{B}}\mathbf{f}_{t+h})^2(\mathbf{w}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\mathbf{w}_T)\right|$$

$$a_7 = \max_{h\le L}\left|T^{-1}\sum_{t=1}^{T-h}(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\mathbf{f}_t)^2(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\widehat{\mathbf{w}}_T) - (\mathbf{w}_T'\widehat{\mathbf{B}}\mathbf{f}_t)^2(\mathbf{w}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\mathbf{w}_T)\right|$$

$$a_8 = \max_{h\le L}\left|T^{-1}\sum_{t=1}^{T-h}(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\widehat{\mathbf{w}}_T)^2 - (\mathbf{w}_T'\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\mathbf{w}_T)^2\right|$$

a_1 is bounded by $a_{11} + a_{12}$, where

$a_{11} = \max_{h\le L}|T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\mathbf{B}\mathbf{f}_{t+h})^2(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2 - E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2(\mathbf{w}_T'\mathbf{B}\mathbf{f}_{t+h})^2|$, and $a_{12} = \max_{h\le L}|T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_{t+h})^2(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_t)^2 - (\mathbf{w}_T'\mathbf{B}\mathbf{f}_{t+h})^2(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2|$.

Given the assumption that $\max_{h\le L}\sum_{t=1}^T \mathrm{cov}[(\mathbf{w}_T'\mathbf{B}\mathbf{f}_1)^2(\mathbf{w}_T'\mathbf{B}\mathbf{f}_{1+h})^2, (\mathbf{w}_T'\mathbf{B}\mathbf{f}_{1+t})^2(\mathbf{w}_T'\mathbf{B}\mathbf{f}_{1+t+h})^2] = O(1)$, the same argument of the proof of Lemma A.3(ii) implies $a_{11} = O_p(\sqrt{L/T})$. On the other hand, by (B.14) of Fan et al. (2011), $\|\widehat{\mathbf{B}} - \mathbf{B}\|_{\max} = O_p(\sqrt{\log N/T})$, which implies $\|\mathbf{w}_T'(\widehat{\mathbf{B}} - \mathbf{B})\| = O_p(\sqrt{\log N/T})$. It is then easy to show that $a_{12} = O_p(\sqrt{\log N/T})$. It follows that $a_1 = O_p(\sqrt{(L+\log N)/T})$. By the triangular inequality, $a_2 = O_p(\sqrt{\log N/T})$. By the same argument of the proof of Lemma A.3, we have $a_3 = O_p(L^2/T) = a_4$. Moreover, because $\max_{i\le N}\|\widehat{\mathbf{b}}_i\| = O_p(1)$, $\max_{i\le N,t\le T}|\widehat{\mathbf{b}}_i'\mathbf{f}_t| = O_p((\log T)^{1/r_2})$,

$$a_5 \le \|\widehat{\mathbf{w}}_T + \mathbf{w}\|_1\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1 \max_{i\le N,t\le T}|\widehat{\mathbf{b}}_i'\mathbf{f}_t|^4(\|\mathbf{w}_T\|_1^2 + \|\widehat{\mathbf{w}}_T\|_1^2)$$
$$= O_p((\log T)^{4/r_2}\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1)$$
$$a_6 \le \|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1 \max_{i\le N,t\le T}|\widehat{\mathbf{b}}_i'\mathbf{f}_t|^2\|\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\|_{\max} \times (\|\mathbf{w}_T\|_1^2\|\widehat{\mathbf{w}}_T\|_1 + \|\mathbf{w}_T\|_1^3 + \|\widehat{\mathbf{w}}_T\|_1^2\|\widehat{\mathbf{w}}_T + \mathbf{w}_T\|_1)$$
$$= O_p((\log T)^{2/r_2}\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1).$$

Term a_7 is bounded as the same as a_6. Finally,

$$a_8 \le \|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1\|\widehat{\mathbf{B}}\,\widehat{\mathrm{cov}}(\mathbf{f}_t)\widehat{\mathbf{B}}'\|_{\max}^2(\|\widehat{\mathbf{w}}_T\|_1^2 + \|\mathbf{w}_T\|_1^2) \times (\|\widehat{\mathbf{w}}_T\|_1 + \|\mathbf{w}_T\|_1) = O_p(\|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1|). \quad ■$$

Proof of Lemma 4.2. We have $|\widehat{\sigma}_f^2 - \sigma_f^2| \le \sum_{i=1}^3 b_i$, where $b_1 = |\widehat{\gamma}_f(0) - \gamma_f(0)|$,

$$b_2 = 2\sum_{h=1}^L(1-h/L)|\widehat{\gamma}_f(h) - \gamma_f(h)|, \qquad b_3 = 2\sum_{h>L}\gamma_f(h),$$

$$b_4 = \frac{2}{L}\sum_{h=1}^L h\gamma_f(h).$$

By the lemma's condition, $b_3 = o_p(\sigma_f^2)$ By Lemma B.1,

$$b_1 + b_2 \le 2L \max_{h \le L} |\widehat{\gamma}_f(h) - \gamma_f(h)|$$
$$= O_p(L\sqrt{(L + \log N)/T} + L(\log T)^{4/r_2} \|\widehat{\mathbf{w}}_T - \mathbf{w}_T\|_1) = o_p(\sigma_f^2).$$

As for b_4, note that $\sum_{h=1}^{\infty} |\gamma_f(h)|/\sigma_f^2 < \infty$. By dominated convergence theorem,

$$\lim_T \frac{1}{L} \sum_{h=1}^{L} \frac{h}{\sigma_f^2} \gamma_f(h) = \sum_{h=1}^{\infty} \lim_T h\gamma_f(h) \frac{1}{\sigma_f^2 L} I(h \le L) = 0$$

given that $L\sigma_f^2 \to \infty$. The second statement is due to $\widehat{\sigma}_f^2 = o_p(\log N)$. ■

B.2. Proof of Theorem 4.2

Write $\mathbf{R} = (\mathbf{R}_1, \ldots, \mathbf{R}_T)$ be $N \times T$; $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_T)$ be $r \times T$, and $\widehat{\text{cov}}(\mathbf{f}_t) = \mathbf{F}\mathbf{F}'/T$. We have $\widehat{\mathbf{B}} = \mathbf{R}\mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}$. Define $\mathbf{C}_T = \widehat{\mathbf{B}} - \mathbf{B}$ and $\mathbf{D}_T = \widehat{\text{cov}}(\mathbf{f}_t) - \text{cov}(\mathbf{f}_t)$. Then we have the following decomposition: $\mathbf{w}_T'(\widehat{\boldsymbol{\Sigma}}_f - \boldsymbol{\Sigma})\mathbf{w}_T = \sum_{i=1}^{4} d_i$, where

$$d_1 = \mathbf{w}_T'\mathbf{B}\mathbf{D}_T\mathbf{B}'\mathbf{w}_T; \qquad d_2 = 2\mathbf{w}_T'\mathbf{C}_T\,\widehat{\text{cov}}(\mathbf{f}_t)\mathbf{B}'\mathbf{w}_T;$$
$$d_3 = \mathbf{w}_T'\mathbf{C}_T\,\widehat{\text{cov}}(\mathbf{f}_t)\mathbf{C}_T'\mathbf{w}_T, \qquad d_4 = \mathbf{w}_T'(\widehat{\boldsymbol{\Sigma}}_u - \boldsymbol{\Sigma}_u)\mathbf{w}_T.$$

We now study each of the above four terms separately. Let $\mathbf{E} = (\mathbf{u}_1, \ldots, \mathbf{u}_T)$ be $N \times T$. Then $\mathbf{C}_T = \mathbf{E}\mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}$. ■

Lemma B.2. (i) $\|\mathbf{F}\mathbf{E}'\mathbf{w}_T\| = O_p(T^{1/2}(\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T)^{1/4}(E|\mathbf{w}_T'\mathbf{u}_t|^4)^{1/8})$.
(ii) $|d_2| = O_p(T^{-1/2}(\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T)^{1/4}(E|\mathbf{w}_T'\mathbf{u}_t|^4)^{1/8})$.

Proof. We have,

$$E\|\mathbf{F}\mathbf{E}'\mathbf{w}_T\|^2 = E[\text{tr}(\mathbf{w}_T'\mathbf{E}\mathbf{F}'\mathbf{F}\mathbf{E}'\mathbf{w}_T)] = \text{tr}[E(\mathbf{F}\mathbf{E}'\mathbf{w}_T\mathbf{w}_T'\mathbf{E}\mathbf{F}')]$$
$$= \text{tr}[E(\mathbf{F}E(\mathbf{E}'\mathbf{w}_T\mathbf{w}_T'\mathbf{E}|\mathbf{F})\mathbf{F}')] = \text{tr}[E(\mathbf{F}E(\mathbf{E}'\mathbf{w}_T\mathbf{w}_T'\mathbf{E})\mathbf{F}')].$$

Note that $E(\mathbf{E}'\mathbf{w}_T\mathbf{w}_T'\mathbf{E}) = (E[\mathbf{u}_t'\mathbf{w}_T\mathbf{w}_T'\mathbf{u}_s])_{t \le t, s \le T} = (\text{cov}(\mathbf{w}_T'\mathbf{u}_t, \mathbf{w}_T'\mathbf{u}_s))_{t \le t, s \le T}$. By Davydov's inequality, (see, e.g., Proposition 2.5 of Fan and Yao, 2003, with $p = 1/2$ and $q = 1/4$), $|\text{cov}(\mathbf{w}_T'\mathbf{u}_t, \mathbf{w}_T'\mathbf{u}_s)| \le 8\alpha_f(|t-s|)^{1/4}(\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T)^{1/2}(E|\mathbf{w}_T'\mathbf{u}_t|^4)^{1/4}$, where $\alpha(\cdot)$ denotes the α-mixing coefficient. By $\sum_{t=1}^{\infty} \alpha_f(t)^{1/4} < \infty$, we have

$$E\|\mathbf{F}\mathbf{E}'\mathbf{w}_T\|^2 = \sum_{k=1}^{r}\sum_{t=1}^{T}\sum_{s=1}^{T} \text{cov}(\mathbf{w}_T'\mathbf{u}_t, \mathbf{w}_T'\mathbf{u}_s)E(f_{kt}f_{ks})$$
$$= O(1)(\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T)^{1/2}(E|\mathbf{w}_T'\mathbf{u}_t|^4)^{1/4}\sum_{t=1}^{T}\sum_{s=1}^{T}\alpha_f(|t-s|)^{1/4}$$
$$= O(T(\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T)^{1/2}(E|\mathbf{w}_T'\mathbf{u}_t|^4)^{1/4}),$$

which then implies (i). For part (ii), we have

$$|d_2| = 2|\mathbf{w}_T'\mathbf{B}\,\widehat{\text{cov}}(\mathbf{f}_t)(\mathbf{F}\mathbf{F}')^{-1}\mathbf{F}\mathbf{E}'\mathbf{w}_T| = \frac{2}{T}|\mathbf{w}_T'\mathbf{B}\mathbf{F}\mathbf{E}'\mathbf{w}_T|$$
$$\le \frac{2}{T}\|\mathbf{w}_T'\mathbf{B}\|\,\|\mathbf{F}\mathbf{E}'\mathbf{w}_T\|.$$

Now write $\mathbf{B} = (b_{ij})_{i \le N, j \le K}$, then $\|\mathbf{w}_T'\mathbf{B}\|^2 = \sum_{j=1}^{K}(\sum_{i=1}^{N} w_i b_{ij})^2 \le \max_{i,j}|b_{ij}|K\|\mathbf{w}_T\|_1^2 = O(1)$. ■

Lemma B.3. *For the factor-based thresholded error covariance matrix,*

$$\|\widehat{\boldsymbol{\Sigma}}_u - \boldsymbol{\Sigma}_u\| = O_p\left(s_N\left(\frac{\log N}{T}\right)^{1/2-q/2}\right).$$

Proof. By Lemma 3.1 in Fan et al. (2011), we have, $\max_{i \le N} T^{-1}\sum_{t=1}^{T}(\hat{u}_{it} - u_{it})^2 = O_p(\log N/T)$. The result then follows from Theorem A.1 in Fan et al. (2013). ■

Lemma B.4. (i) $|d_3| = O_p(T^{-1}(\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T)^{1/2}(E|\mathbf{w}_T'\mathbf{u}_t|^4)^{1/4})$.
(ii) $|d_4| = O_p(s_N(\log N/T)^{1/2-q/2}\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T)$.

Proof. (i) Because $\|(\mathbf{F}\mathbf{F}')^{-1}\| = O_p(T^{-1})$, $|d_3| = T^{-1}\mathbf{w}_T'\mathbf{E}\mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}\mathbf{F}\mathbf{E}'\mathbf{w}_T = O_p(T^{-2}\|\mathbf{F}\mathbf{E}'\mathbf{w}_T\|^2)$. It then follows from Lemma B.2.

(ii) It follows from $|d_4| \le \|\widehat{\boldsymbol{\Sigma}}_u - \boldsymbol{\Sigma}_u\|\|\mathbf{w}_T\|^2 \le \lambda_{\min}^{-1}(\boldsymbol{\Sigma}_u)\|\widehat{\boldsymbol{\Sigma}}_u - \boldsymbol{\Sigma}_u\|\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T$ and Lemma B.3. ■

Lemma B.5. (i) $E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^4 = O(1)$ *and* $E(\mathbf{w}_T'\mathbf{R}_t)^4 = O(1)$;
(ii) $\gamma_T(0) = O(1)$ *and* $\gamma_f(0) = O(1)$.

Proof. (i) For all $s > 0$, $P(|f_{jt}| > s) \le \exp(-(s/b_2)^{r_2})$ implies

$$\max_{j \le K} Ef_{jt}^4 \le \max_{j \le K}\int_x^{\infty} P(|f_{jt}|^4 > x)dx \le \int \exp(-(s/b_2)^{r_2})ds < \infty.$$

We have, because the dimension of $\mathbf{f}_t$ is fixed,

$$E(\mathbf{w}_T\mathbf{B}\mathbf{f}_t)^4 \le \|\mathbf{w}_T\mathbf{B}\|^4 E\|\mathbf{f}_t\|^4 \le \|\mathbf{w}_T\|_1^4 \max_{i,j}|b_{ij}|^4 O(1) = O(1).$$

In addition,

$$E(\mathbf{w}_T'\mathbf{R}_t)^4 = E\left(\sum_{j=1}^{N} w_{T,j}R_{jt}\right)^4$$
$$= \sum_{i,j,k,l \le N} w_{T,i}w_{T,j}w_{T,k}w_{T,l}ER_{it}R_{jt}R_{kt}R_{lt}$$
$$\le \max_{i,j,k,l}|ER_{it}R_{jt}R_{kt}R_{lt}|\|\mathbf{w}_T\|_1^4 \le \max_{j \le N} ER_{jt}^4\|\mathbf{w}_T\|_1^4 = O(1).$$

(ii) The result follows from (i) and by observing that $\gamma_T(0) = \text{var}((\mathbf{w}_T'\mathbf{R}_t)^2) \le E(\mathbf{w}_T'\mathbf{R}_t)^4$ and that $\gamma_f(0) = \text{var}((\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2) \le E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^4$. ■

Lemma B.6. $\sum_{h=1}^{\infty}|\gamma_f(h)| < \infty$, *and* $\sum_{h=1}^{\infty}|\gamma_T(h)| < \infty$.

Proof. By Davydov's inequality (Proposition 2.5 of Fan and Yao, 2003, with $p = 1/2$ and $q = 1/4$), there are constants $M_1, M_2 > 0$ such that for any integer h,

$$|\gamma_f(|h|)| \le 8\alpha_f(|h|)^{1/4}(E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2)^{1/2}(E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^4)^{1/4}$$
$$= M_2\exp(-M|h|^{r_3}/4)$$

where the last equality follows from the α-mixing condition as well as the fact that $E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^4 = O(1)$ due to $\|\mathbf{w}_T\|_1 = O(1)$ (Lemma B.5). The first result the follows from $\sum_{h=1}^{\infty}\exp(-Ch^{r_3}) < \infty$ for any $C, r_3 > 0$. The proof of $\sum_{h=1}^{\infty}|\gamma_T(h)| < \infty$ follows from the same arguments. ■

Lemma B.7. $\sqrt{T/\sigma_f^2}d_1 \to^d \mathcal{N}(0, 1)$.

Proof. Let $Z_{T,t} = \mathbf{w}_T'\mathbf{B}(\mathbf{f}_t\mathbf{f}_t' - E\mathbf{f}_t\mathbf{f}_t')\mathbf{B}'\mathbf{w}_T$, which depends on T through $\dim(\mathbf{w}_T) = N_T$. Hence $d_1 = T^{-1}\sum_{t=1}^{T} Z_{T,t}$. Note that $\|\mathbf{w}_T'\mathbf{B}\|^2 \le K\|\mathbf{B}\|_{\max}^2\|\mathbf{w}_T\|_1^2 = O(1)$. Hence $EZ_{T,1}^2 = O(1)$. We define $B_{f,K}^2 = \text{var}(\sum_{t=1}^{K} Z_{T,t})$ and $B_f^2 = \text{var}(\sum_{t=1}^{T} Z_{T,t}) = O(T)$. By the assumption that $\alpha_f(T) = o(\gamma_f(0))$, the correlation $|\text{Corr}(Z_{T,t}, Z_{T,t+T})| \le |\gamma_f(T)|/\gamma_f(0) = o(1)$. Moreover, the Lindeberg condition can be verified by the same lines as in the proof of Lemma A.6.

Hence the conditions of Theorem 2.1 of Peligrad (1996) are satisfied, which implies

$$B_f^{-1}\sum_{t=1}^{T} Z_{T,t} \to^d \mathcal{N}(0, 1). \tag{B.1}$$

Now $B_f^2 = T\gamma_f(0) + 2T\sum_{h=1}^T \gamma_f(h) - 2T\sum_{h=1}^T h\gamma_f(h)/T$. Because $\sum_{h=1}^T h\gamma_f(h)/T = o(\gamma_f(0)+2\sum_{h=1}^\infty \gamma_f(h))$, we have $T^{-1/2}(\sigma_f^2)^{-1/2}\sum_{t=1}^T Z_{T,t} \to^d \mathcal{N}(0,1)$. ■

Lemma B.8.

$$\sqrt{\frac{T}{\sigma_f^2}}\mathbf{w}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\mathbf{w}_T \to^d \mathcal{N}(0,1). \tag{B.2}$$

Proof. In fact, $\sqrt{\frac{T}{\sigma_f^2}}\mathbf{w}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\mathbf{w}_T = \sqrt{\frac{T}{\sigma_f^2}}d_1 + \sqrt{\frac{T}{\sigma_f^2}}(d_2+d_3+d_4)$. By Lemma B.7, it suffices to show that $\sqrt{T/\sigma_f^2}(d_2+d_3+d_4) = o_p(1)$. By Lemma B.2,

$$\begin{aligned}\sqrt{T/\sigma_f^2}|d_2| &= O_p((\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T)^{1/4}(E|\mathbf{w}_T'\mathbf{u}_t|^4)^{1/8})/\sqrt{\sigma_f^2}\\ &= O_p((\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T)^{1/4}/\sqrt{\sigma_f^2}) = o_p(1)\end{aligned}$$

since $E|\mathbf{w}_T'\mathbf{u}_t|^4 = O(1)$. Lemma B.4 implies $\sqrt{T/\sigma_f^2}|d_3| = O_p((\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T)^{1/2}(T\sigma_f^2)^{-1/2}) = o_p(1)$ since $\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T = o(\sigma_f^4) = O(1)$. It also follows from Lemma B.4 that $\sqrt{T/\sigma_f^2}|d_4| = O_p(\mathbf{w}_T'\mathbf{\Sigma}_u\mathbf{w}_T s_N(\sigma_f^2)^{-1/2}(\log N)^{1/2-q/2}T^{q/2}) = o_p(1)$. This implies the desired result. ■

Proof of Theorem 4.2. By Theorem 3.2 of Fan et al. (2011), $\|\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma}\|_{\max} = O_p(\sqrt{\frac{\log NT}{T}})$. Note that $T\sigma_f^2/B_f^2 \to 1$, we have

$$\begin{aligned}&\frac{T}{B_f}[\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\widehat{\mathbf{w}}_T - \mathbf{w}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\mathbf{w}_T]\\ &\le \frac{T}{B_f}\|\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma}\|_{\max}\|\widehat{\mathbf{w}}_T - \mathbf{w}\|_1(\|\widehat{\mathbf{w}}\|_1 + \|\mathbf{w}\|_1)\\ &\le O_p\left(\frac{\sqrt{T}}{\sigma_f}\sqrt{\frac{\log NT}{T}}\right)o_p(\sigma_f(\log NT)^{-4/r_1}) = o_p(1).\end{aligned} \tag{B.3}$$

The first statement $\left[\mathrm{var}\left(\sum_{t=1}^T(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2\right)\right]^{-1/2}T\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\widehat{\mathbf{w}}_T \to^d \mathcal{N}(0,1)$ then follows from (B.2) and (B.3). They also yield

$$\sqrt{\frac{T}{\sigma_f^2}}\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\widehat{\mathbf{w}}_T \to^d \mathcal{N}(0,1), \tag{B.4}$$

and $\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\widehat{\mathbf{w}}_T = O_p(T^{-1/2}\sqrt{\sigma_f^2})$. Moreover, since $|\sigma_f^2 - \widehat{\sigma}_f^2| = o_p(\sigma_f^2)$,

$$\begin{aligned}&\sqrt{T}|\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\widehat{\mathbf{w}}_T||\sigma_f^{-1} - \widehat{\sigma}_f^{-1}|\\ &= \sqrt{T}|\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\widehat{\mathbf{w}}_T|\frac{|\widehat{\sigma}_f^2 - \sigma_f^2|}{\sigma_f^2(1+o_p(1))} = o_p(1).\end{aligned}$$

It then follows from (B.4) that $\sqrt{T/\widehat{\sigma}_f^2}\widehat{\mathbf{w}}_T'(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma})\widehat{\mathbf{w}}_T \to^d \mathcal{N}(0,1)$, which gives the H-CLUB. ■

Appendix C. Proofs for the POET-based estimation

Let $\mathbf{V}$ denote the $K\times K$ diagonal matrix of the first K largest eigenvalues of $\mathbf{S}$ in decreasing order. Let $\widehat{\mathbf{F}} = (\widehat{\mathbf{f}}_1,\ldots,\widehat{\mathbf{f}}_T)$ be a $K\times T$ matrix such that the rows of $\widehat{\mathbf{F}}/\sqrt{T}$ are the eigenvectors corresponding to the r largest eigenvalues of the $T\times T$ matrix $\mathbf{R}'\mathbf{R}$. Let $\widehat{\mathbf{B}} = \mathbf{R}\widehat{\mathbf{F}}'/T$. Define a $K\times K$ matrix

$$\mathbf{H} = \frac{1}{T}\mathbf{V}^{-1}\widehat{\mathbf{F}}\mathbf{F}'\mathbf{B}'\mathbf{B}.$$

According to Stock and Watson (2002), $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{f}}_t$ can be treated as estimators of $\mathbf{B}\mathbf{H}^{-1}$ and $\mathbf{H}\mathbf{f}_t$ respectively.

C.1. Proof of Lemma 4.3

Lemma C.1. (i) $\|\mathbf{w}_T'\widehat{\mathbf{B}}\| = O_p(1)$, *and* $\|\mathbf{w}_T'(\widehat{\mathbf{B}} - \mathbf{B}\mathbf{H}^{-1})\| = O_p(N^{-1/2} + (\log N/T)^{1/2})$

(ii) $\|\widehat{\mathbf{F}} - \mathbf{H}\mathbf{F}\|^2/T = T^{-1}\sum_{t=1}^T\|\widehat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t\|^2 = O_p(N^{-1}+T^{-2})$.

(iii) $\|\mathbf{w}_T'\mathbf{E}\|^2 = O_p(T)$.

(iv) $\|T^{-1}\sum_{t=1}^T[\widehat{\mathbf{f}}_t\widehat{\mathbf{f}}_t' - \mathbf{H}\mathbf{f}_t(\mathbf{H}\mathbf{f}_t)']\| = O_p(N^{-1/2}+T^{-1})$.

Proof. (i) By Lemma B.16 in an earlier version of Fan et al. (2013),[1] $\|\widehat{\mathbf{B}}\|_{\max} \le \|\widehat{\mathbf{B}} - \mathbf{B}\mathbf{H}^{-1}\|_{\max} + \|\mathbf{B}\|_{\max} = O_p(1)$. Thus $\|\mathbf{w}_T'\widehat{\mathbf{B}}\|^2 \le r\|\widehat{\mathbf{B}}\|_{\max}^2\|\mathbf{w}_T\|_1^2 = O_p(1)$. On the other hand, $\|\mathbf{w}_T'(\widehat{\mathbf{B}} - \mathbf{B}\mathbf{H}^{-1})\|^2 \le r\|\widehat{\mathbf{B}} - \mathbf{B}\|_{\max}^2\|\mathbf{w}_T\|_1^2 = O_p(1/N + \log N/T)$.

(ii) By (A.1) in Bai (2003), the following identity holds:

$$\begin{aligned}\widehat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t = (\mathbf{V}/N)^{-1}\Bigg(&\frac{1}{T}\sum_{s=1}^T\widehat{\mathbf{f}}_sE(\mathbf{u}_s'\mathbf{u}_t)/N + \frac{1}{T}\sum_{s=1}^T\widehat{\mathbf{f}}_s\zeta_{st}\\ &+ \frac{1}{T}\sum_{s=1}^T\widehat{\mathbf{f}}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^T\widehat{\mathbf{f}}_s\xi_{st}\Bigg)\end{aligned} \tag{C.1}$$

where $\zeta_{st} = \mathbf{u}_s'\mathbf{u}_t/N - E(\mathbf{u}_s'\mathbf{u}_t)/N$, $\eta_{st} = \mathbf{f}_s'\sum_{i=1}^N\mathbf{b}_iu_{it}/N$, and $\xi_{st} = \mathbf{f}_t'\sum_{i=1}^p\mathbf{b}_iu_{is}/N$. It follows from Lemma C.7 in Fan et al. (2013) that

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^T\left(\frac{1}{T}\sum_{s=1}^T\hat{f}_{is}\zeta_{st}\right)^2 + \frac{1}{T}\sum_{t=1}^T\left(\frac{1}{T}\sum_{s=1}^T\hat{f}_{is}\eta_{st}\right)^2\\ &+ \frac{1}{T}\sum_{t=1}^T\left(\frac{1}{T}\sum_{s=1}^T\hat{f}_{is}\xi_{st}\right)^2 = O_p\left(\frac{1}{N}\right).\end{aligned}$$

Moreover, by Lemma C.9 of Fan et al. (2013), $\max_{i\le r}\frac{1}{T}\sum_{t=1}^T(\widehat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t)_i^2 = O_p(1/T + 1/N)$. Applying the inequality $(a+b)^2 \le 2a^2 + 2b^2$ gives,

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^T\left(\frac{1}{T}\sum_{s=1}^T\hat{f}_{is}E(\mathbf{u}_s'\mathbf{u}_t)/N\right)^2\\ &\le \frac{1}{T}\sum_{t=1}^T\left(\frac{1}{T}\sum_{s=1}^T[|(\widehat{\mathbf{f}}_s - \mathbf{H}\mathbf{f}_s)_i| + |(\mathbf{H}\mathbf{f}_s)_i|]|E(\mathbf{u}_s'\mathbf{u}_t)|/N\right)^2\\ &\le \frac{2}{T}\sum_{t=1}^T\left(\frac{1}{T}\sum_{s=1}^T|(\widehat{\mathbf{f}}_s - \mathbf{H}\mathbf{f}_s)_i||E(\mathbf{u}_s'\mathbf{u}_t)|/N\right)^2\\ &+ \frac{2}{T}\sum_{t=1}^T\left(\frac{1}{T}\sum_{s=1}^T|(\mathbf{H}\mathbf{f}_s)_i||E(\mathbf{u}_s'\mathbf{u}_t)|/N\right)^2.\end{aligned}$$

[1] Downloadable from http://terpconnect.umd.edu/~yuanliao/factor2/factor2.html.

By the Cauchy–Schwarz inequality and that $\max_{t\le T}\sum_{s=1}^T|E(\mathbf{u}_s'\mathbf{u}_t)/N|^2=O(1)$,

$$\frac{2}{T}\sum_{t=1}^T\left(\frac{1}{T}\sum_{s=1}^T|(\widehat{\mathbf{f}}_s-\mathbf{H}\mathbf{f}_s)_i||E(\mathbf{u}_s'\mathbf{u}_t)|/N\right)^2 \le \max_{i\le r}\frac{2}{T}\sum_{s=1}^T(\widehat{\mathbf{f}}_s-\mathbf{H}\mathbf{f}_s)_i^2\frac{1}{T}\sum_{s=1}^T(|E\mathbf{u}_s'\mathbf{u}_t|/N)^2=O_p\left(\frac{1}{T^2}+\frac{1}{NT}\right).$$

Also, $\frac{2}{T}\sum_{t=1}^T(\frac{1}{T}\sum_{s=1}^T|(\mathbf{H}\mathbf{f}_s)_i||E(\mathbf{u}_s'\mathbf{u}_t)|/N)^2 \le O_p(T^{-1})\sum_{t=1}^T(\frac{1}{T}\sum_{s=1}^T\|\mathbf{f}_s\||E(\mathbf{u}_s'\mathbf{u}_t)|/N)^2$. We have $T^{-1}\sum_{t=1}^T(\frac{1}{T}\sum_{s=1}^T\|\mathbf{f}_s\||E(\mathbf{u}_s'\mathbf{u}_t)|/N)^2=O_p(T^{-2})$ since

$$E\frac{1}{T}\sum_{t=1}^T\left(\frac{1}{T}\sum_{s=1}^T\|\mathbf{f}_s\||E(\mathbf{u}_s'\mathbf{u}_t)|/N\right)^2 =\frac{1}{T^2}\sum_{s=1}^T\sum_{l=1}^T E\|\mathbf{f}_s\|\|\mathbf{f}_l\|\frac{|E\mathbf{u}_s'\mathbf{u}_t|}{N}\frac{|E\mathbf{u}_l'\mathbf{u}_t|}{N} \le \max_{s\le T}E\|\mathbf{f}_s\|^2\max_{s\le T}\left(\frac{1}{T}\sum_{s=1}^T|E\mathbf{u}_s'\mathbf{u}_t|/N\right)^2=O(T^{-2}).$$

This implies $\max_{i\le r}T^{-1}\sum_{t=1}^T(\widehat{\mathbf{f}}_t-\mathbf{H}\mathbf{f}_t)_i^2=O_p(N^{-1}+T^{-2})$, and thus $T^{-1}\sum_{t=1}^T\|\widehat{\mathbf{f}}_t-\mathbf{H}\mathbf{f}_t\|^2=O_p(N^{-1}+T^{-2})$.

(iii) $E\|\mathbf{w}_T'\mathbf{E}\|^2=E\sum_{t=1}^T(\sum_{i=1}^N w_iu_{it})^2=T\max_{i,j}|Eu_{it}u_{jt}|\|\mathbf{w}_T\|_1^2=O(T)$. Thus $\|\mathbf{w}_T'\mathbf{E}\|^2=O_p(T)$. Finally, (iv) follows from the Cauchy–Schwarz inequality and part (ii). ■

Lemma C.2. $\max_{h\le L}|\widehat{\gamma}_P(h)-\gamma_f(h)|=O_p(\sqrt{\frac{L+\log N}{T}}+\frac{1}{\sqrt{N}}+(\log T)^{2/r_2}\|\widehat{\mathbf{w}}_T-\mathbf{w}_T\|_1)$.

Proof. The triangular inequality implies $\max_{h\le L}|\widehat{\gamma}_P(h)-\gamma_f(h)|\le\sum_{i=1}^8 a_i$, where

$$a_1=\max_{h\le L}\left|T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_{t+h})^2(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_t)^2-E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2(\mathbf{w}_T'\mathbf{B}\mathbf{f}_{t+h})^2\right|,$$
$$a_2=|(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{B}}'\mathbf{w}_T)^2-(\mathbf{w}_T'\mathbf{B}\mathbf{B}'\mathbf{w}_T)^2|,$$
$$a_3=\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{B}}'\mathbf{w}_T\max_{h\le L}\left|\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{B}}'\mathbf{w}_T-T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_t)^2\right|,$$
$$a_4=\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{B}}'\mathbf{w}_T\max_{h\le L}\left|\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{B}}'\mathbf{w}_T-T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_{t+h})^2\right|$$
$$a_5=\max_{h\le L}\left|T^{-1}\sum_{t=1}^{T-h}(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_{t+h})^2(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_t)^2-(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_{t+h})^2(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_t)^2\right|,$$
$$a_6=\max_{h\le L}\left|T^{-1}\sum_{t=1}^{T-h}(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_{t+h})^2(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}\mathbf{B}'}\widehat{\mathbf{w}}_T)-(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_{t+h})^2(\mathbf{w}_T'\widehat{\mathbf{B}\mathbf{B}'}\mathbf{w}_T)\right|$$
$$a_7=\max_{h\le L}\left|T^{-1}\sum_{t=1}^{T-h}(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_t)^2(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}\mathbf{B}'}\widehat{\mathbf{w}}_T)-(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_t)^2(\mathbf{w}_T'\widehat{\mathbf{B}\mathbf{B}'}\mathbf{w}_T)\right|$$
$$a_8=\max_{h\le L}\left|T^{-1}\sum_{t=1}^{T-h}(\widehat{\mathbf{w}}_T'\widehat{\mathbf{B}\mathbf{B}'}\widehat{\mathbf{w}}_T)^2-(\mathbf{w}_T'\widehat{\mathbf{B}\mathbf{B}'}\mathbf{w}_T)^2\right|.$$

Here a_1 is bounded by $a_{11}+a_{12}$, where $a_{11}=\max_{h\le L}|T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\mathbf{B}\mathbf{f}_{t+h})^2(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2-E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2(\mathbf{w}_T'\mathbf{B}\mathbf{f}_{t+h})^2|$, and $a_{12}=\max_{h\le L}|T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_{t+h})^2(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_t)^2-(\mathbf{w}_T'\mathbf{B}\mathbf{f}_{t+h})^2(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2|$.

As in the proof of Lemma B.1, $a_{11}=O_p(\sqrt{L/T})$. It follows from the Cauchy–Schwarz inequality that

$$\begin{aligned}a_{12}&=O_p\left(\max_{h\le L}\left|T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_{t+h})(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_t)-(\mathbf{w}_T'\mathbf{B}\mathbf{f}_{t+h})(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)\right|\right)\\&=O_p\left(\max_{h\le L}\left(T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_{t+h}-\mathbf{w}_T'\mathbf{B}\mathbf{f}_{t+h})^2\right)^{1/2}+\left(T^{-1}\sum_{t=1}^{T-h}(\mathbf{w}_T'\widehat{\mathbf{B}}\widehat{\mathbf{f}}_t-\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2\right)^{1/2}\right)\\&=O_p\left(\|\mathbf{w}_T'(\widehat{\mathbf{B}}-\mathbf{B}\mathbf{H}^{-1})\|+\|\mathbf{w}_T'\mathbf{B}\mathbf{H}^{-1}\|\times\left(T^{-1}\sum_{t=1}^T\|\widehat{\mathbf{f}}_t-\mathbf{H}\mathbf{f}_t\|^2\right)^{1/2}\right).\end{aligned}$$

It follows from Lemma C.1 that $a_{12}=O_p(\sqrt{\log N/T}+N^{-1/2})$, thus $a_1=O_p(\sqrt{(L+\log N)/T}+N^{-1/2})$. On the other hand, for $g_1,\ldots,g_5$ defined in (C.2),

$$a_2=O_p(|\mathbf{w}_T'(\widehat{\mathbf{B}}\widehat{\mathbf{B}}'-\mathbf{B}\mathbf{B}')\mathbf{w}_T|)=O_p\left(\sum_{i=1}^5|g_i|\right)=O_p\left(\sqrt{\sigma_f^2/T}\right)=o_p\left(\sqrt{L/T}\right).$$

$$a_3=O_p(1)\max_{h\le L}\left|\mathbf{w}_T'\widehat{\mathbf{B}}\left(\frac{1}{T}\sum_{t=T-h+1}^T\widehat{\mathbf{f}}_t\widehat{\mathbf{f}}_t'\right)\widehat{\mathbf{B}}'\mathbf{w}_T\right|=O_p\left(\frac{1}{T}\sum_{t=T-L+1}^T\|\widehat{\mathbf{f}}_t\|^2\right).$$

We have $\frac{1}{T}\sum_{t=T-L+1}^T\|\widehat{\mathbf{f}}_t\|^2\le\frac{2}{T}\sum_{t=T-L+1}^T\|\widehat{\mathbf{f}}_t-\mathbf{H}\mathbf{f}_t\|^2+\frac{2}{T}\sum_{t=T-L+1}^T\|\mathbf{H}\mathbf{f}_t\|^2$. On one hand, $E\frac{2}{T}\sum_{t=T-L+1}^T\|\mathbf{f}_t\|^2=O(L/T)$. On the other hand, by Theorem 3.3 in Fan et al. (2013), $\max_t\|\widehat{\mathbf{f}}_t-\mathbf{H}\mathbf{f}_t\|=o_p(1)$, hence $\frac{2}{T}\sum_{t=T-L+1}^T\|\widehat{\mathbf{f}}_t-\mathbf{H}\mathbf{f}_t\|^2=O_p(L/T)$. Thus $a_3=O_p(L/T)$. Similarly, we have $a_4=O_p(L/T)$.

Moreover, by Corollary 3.1 of Fan et al. (2013), $\max_{i,t}|\widehat{\mathbf{b}}_i'\widehat{\mathbf{f}}_t|\le o_p(1)+\max_{i,t}|\mathbf{b}_i'\mathbf{f}_t|=O_p((\log T)^{1/r_2})$, and $\|\widehat{\mathbf{B}}\widehat{\mathbf{B}}'\|_{\max}=O_p(1)$. Thus it follows from the same lines of proofs as those in Lemma B.1 that $a_5+\cdots+a_8=O_p((\log T)^{2/r_2}\|\widehat{\mathbf{w}}_T-\mathbf{w}_T\|_1)$. ■

Proof of Lemma 4.3. We have $|\widehat{\sigma}_P^2-\sigma_f^2|\le\sum_{i=1}^3 b_i$, where $b_1=|\widehat{\gamma}_P(0)-\gamma_f(0)|$,

$$b_2=2\sum_{h=1}^L(1-h/L)|\widehat{\gamma}_P(h)-\gamma_f(h)|,\qquad b_3=2\sum_{h>L}\gamma_f(h),$$
$$b_4=\frac{2}{L}\sum_{h=1}^L h\gamma_f(h).$$

By Lemma C.2, $b_1+b_2\le 2L\max_{h\le L}|\widehat{\gamma}_P(h)-\gamma_f(h)|$, which is $O_p(L\sqrt{(L+\log N)/T}+L(\log T)^{4/r_2}\|\widehat{\mathbf{w}}_T-\mathbf{w}_T\|_1+L/\sqrt{N})=o_p(\sigma_f^2)$. Hence the proof is the same as that of Lemma 4.2. ■

C.2. Proof of Theorem 4.3

If we write $\widetilde{\mathbf{C}}_T = \widehat{\mathbf{B}} - \mathbf{B}\mathbf{H}^{-1}$ and $\widetilde{\mathbf{D}}_T = T^{-1}\sum_{t=1}^T \mathbf{f}_t\mathbf{f}_t' - \mathrm{cov}(\mathbf{f}_t)$, then $\mathbf{w}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\mathbf{w}_T = \sum_{i=1}^5 g_i$, where

$$g_1 = \mathbf{w}_T'\mathbf{B}\widetilde{\mathbf{D}}_T\mathbf{B}'\mathbf{w}_T, \qquad g_2 = \mathbf{w}_T'\widetilde{\mathbf{C}}_T\widehat{\mathbf{B}}'\mathbf{w}_T,$$
$$g_3 = \mathbf{w}_T'\mathbf{B}\mathbf{H}^{-1}\widetilde{\mathbf{C}}_T'\mathbf{w}_T, \qquad g_4 = \mathbf{w}_T'(\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Sigma}_u)\mathbf{w}_T,$$
$$g_5 = \mathbf{w}_T'\mathbf{B}\mathbf{H}^{-1}\frac{1}{T}\sum_{t=1}^T[\widehat{\mathbf{f}}_t\widehat{\mathbf{f}}_t' - \mathbf{H}\mathbf{f}_t(\mathbf{H}\mathbf{f}_t)']\mathbf{H}^{-1'}\mathbf{B}'\mathbf{w}_T. \tag{C.2}$$

Recall the definition of d_1 in Appendix B.2, $g_1 = d_1$. Thus it follows from Lemma B.7 that $\sqrt{T/\sigma_f^2}g_1 \to^d \mathcal{N}(0,1)$. We first show that $\sqrt{T/\sigma_f^2}g_i$ are asymptotically negligible for $i = 2, \ldots, 5$. These results are given in the following lemma.

Lemma C.3. (i) $|g_2| = O_p(T^{-1/2}(\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T)^{1/4} + N^{-1/2} + T^{-1})$,
(ii) $|g_3| = O_p(T^{-1/2}(\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T)^{1/4} + N^{-1/2} + T^{-1})$.
(iii) $|g_5| = O_p(N^{-1/2} + T^{-1})$.

Proof. Using the facts that $\mathbf{R} = \mathbf{B}\mathbf{F} + \mathbf{E}$, $\widehat{\mathbf{B}} = \mathbf{R}\widehat{\mathbf{F}}'/T$ and $\widehat{\mathbf{F}}\widehat{\mathbf{F}}'/T = \mathbf{I}_K$, we have

$$\widehat{\mathbf{B}} - \mathbf{B}\mathbf{H}^{-1} = \mathbf{B}\mathbf{H}^{-1}(\mathbf{H}\mathbf{F} - \widehat{\mathbf{F}})\widehat{\mathbf{F}}'/T + \mathbf{E}(\widehat{\mathbf{F}} - \mathbf{H}\mathbf{F})'/T + \mathbf{E}\mathbf{F}'\mathbf{H}'/T.$$

Thus $g_2 = g_{21} + g_{22} + g_{23}$, where $g_{21} = \mathbf{w}_T'\mathbf{B}\mathbf{H}^{-1}(\mathbf{H}\mathbf{F} - \widehat{\mathbf{F}})\widehat{\mathbf{F}}'/T\widehat{\mathbf{B}}'\mathbf{w}_T$, $g_{22} = \mathbf{w}_T'\mathbf{E}(\widehat{\mathbf{F}} - \mathbf{H}\mathbf{F})'/T\widehat{\mathbf{B}}'\mathbf{w}_T$, and $g_{23} = \mathbf{w}_T'\mathbf{E}\mathbf{F}'\mathbf{H}'/T\widehat{\mathbf{B}}'\mathbf{w}_T$. It was shown by Fan et al. (2013) that $\|\mathbf{H}\| = O_p(1) = \|\mathbf{H}^{-1}\|$. Thus by Lemma C.1, $|g_{21}| \le O_p(1)\|\mathbf{H}\mathbf{F} - \widehat{\mathbf{F}}\|\|\widehat{\mathbf{F}}\|/T = O_p(\sqrt{1/N + 1/T^2})$. In addition, $|g_{22}| \le O_p(\sqrt{T})\|\widehat{\mathbf{F}} - \mathbf{H}\mathbf{F}\|/T = O_p(\sqrt{1/N + 1/T^2})$. Finally, by Lemma B.2 , $|g_{23}| \le \|\mathbf{w}_T'\mathbf{E}\mathbf{F}'\|O_p(1/T) = O_p(T^{-1/2}(\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T)^{1/4})$. The proof for the convergence rate of $|g_3|$ is the same, so is omitted. Finally, the rate of convergence for $|g_5|$ follows from Lemma C.1. ■

Proof of Theorem 4.3. By Lemma C.3 and the assumption that $T\sigma_f^2 \to \infty$, $\sigma_f^2 N/T \to \infty$ and $\mathbf{w}_T'\boldsymbol{\Sigma}_u\mathbf{w}_T = o(\sigma_f^4)$, we have $\sqrt{T/\sigma_f^2}g_i = o_p(1)$ for $i = 2, 3, 5$. In addition, by Theorem 3.1 of Fan et al. (2013),

$$\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Sigma}_u\| = O_p\left(s_N\left(\frac{\log N}{T} + \frac{1}{N}\right)^{1/2-q/2}\right),$$

which then implies that $\sqrt{T/\sigma_f^2}g_4 = o_p(1)$. It then follows from Assumption 4.6 and Lemma B.7, $\sqrt{T/\sigma_f^2}\mathbf{w}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\mathbf{w}_T \to^d \mathcal{N}(0, 1)$. By Theorem 3.2 of Fan et al. (2013), $\|\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma}\|_{\max} = O_p(\sqrt{\frac{\log N}{T}} + \frac{1}{\sqrt{N}})$. Due to $N/T \to \infty$ is assumed in the case of unknown factors,

$$\begin{aligned}
&\frac{T}{B_f}[\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T - \mathbf{w}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\mathbf{w}_T] \\
&\le \frac{T}{B_f}\|\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma}\|_{\max}\|\widehat{\mathbf{w}}_T - \mathbf{w}\|_1(\|\widehat{\mathbf{w}}\|_1 + \|\mathbf{w}\|_1) \\
&\le O_p\left(\frac{\sqrt{T}}{\sigma_f}\sqrt{\frac{\log N}{T}}\right)o_p(\sigma_f(\log NT)^{-4/r_1}) \\
&\quad + O_p\left(\frac{\sqrt{T}}{\sigma_f\sqrt{N}}\right)o_p(\sigma_f(\log NT)^{-4/r_1}) = o_p(1).
\end{aligned} \tag{C.3}$$

Therefore, $\left[\mathrm{var}\left(\sum_{t=1}^T(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2\right)\right]^{-1/2} T\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\mathbf{w}_T \to^d \mathcal{N}(0, 1)$, and

$$\sqrt{\frac{T}{\sigma_f^2}}\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T \to^d \mathcal{N}(0, 1). \tag{C.4}$$

Moreover, since $|\sigma_f^2 - \widehat{\sigma}_P^2| = o_p(\sigma_f^2)$,

$$\begin{aligned}
&\sqrt{T}|\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T||\sigma_f^{-1} - \widehat{\sigma}_P^{-1}| \\
&= \sqrt{T}|\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T|\frac{|\widehat{\sigma}_P^2 - \sigma_f^2|}{\sigma_f^2(1 + o_p(1))} = o_p(1).
\end{aligned}$$

It then follows from (C.4) that $\sqrt{T/\widehat{\sigma}_P^2}\widehat{\mathbf{w}}_T'(\widehat{\boldsymbol{\Sigma}}_{P,\widehat{K}} - \boldsymbol{\Sigma})\widehat{\mathbf{w}}_T \to^d \mathcal{N}(0, 1)$, which validates the H-CLUB. ■

C.3. Proof of Theorem 4.4

The theorem follows from the following lemma.

Lemma C.4. *Suppose $\mathbf{f}_t$ and $\mathbf{u}_t$ are independent, and $E\mathbf{f}_t = E\mathbf{u}_t = 0$, then*

$$\begin{aligned}
\mathrm{var}\left[\sum_{t=1}^T(\mathbf{w}_T'\mathbf{R}_t)^2\right] &= \mathrm{var}\left[\sum_{t=1}^T(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2\right] + \mathrm{var}\left[\sum_{t=1}^T(\mathbf{w}_T'\mathbf{u}_t)^2\right] \\
&\quad + \mathrm{var}\left[2\sum_{t=1}^T\mathbf{w}_T'\mathbf{B}\mathbf{f}_t\mathbf{w}_T'\mathbf{u}_t\right].
\end{aligned}$$

Proof. Since $\mathbf{R}_t = \mathbf{B}\mathbf{f}_t + \mathbf{u}_t$, and $\mathbf{f}_t$ and $\mathbf{u}_t$ are independent,

$$\begin{aligned}
&\mathrm{var}\left[\sum_{t=1}^T(\mathbf{w}_T'\mathbf{R}_t)^2\right] \\
&= \mathrm{var}\left[\sum_{t=1}^T(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2 + (\mathbf{w}_T'\mathbf{u}_t)^2 + 2\mathbf{w}_T'\mathbf{B}\mathbf{f}_t\mathbf{w}_T'\mathbf{u}_t\right] \\
&= \mathrm{var}\left[\sum_{t=1}^T(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2\right] + \mathrm{var}\left[\sum_{t=1}^T(\mathbf{w}_T'\mathbf{u}_t)^2\right] \\
&\quad + \mathrm{var}\left[2\sum_{t=1}^T\mathbf{w}_T'\mathbf{B}\mathbf{f}_t\mathbf{w}_T'\mathbf{u}_t\right] \\
&\quad + 2\,\mathrm{cov}\left[\sum_{t=1}^T(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2 + (\mathbf{w}_T'\mathbf{u}_t)^2, 2\sum_{t=1}^T\mathbf{w}_T'\mathbf{B}\mathbf{f}_t\mathbf{w}_T'\mathbf{u}_t\right].
\end{aligned}$$

It suffices to show the covariance term is zero. In fact, since $E\mathbf{w}_T'\mathbf{B}\mathbf{f}_t\mathbf{w}_T'\mathbf{u}_t = 0$,

$$\begin{aligned}
&\mathrm{cov}\left[\sum_{t=1}^T(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t)^2, \sum_{t=1}^T\mathbf{w}_T'\mathbf{B}\mathbf{f}_t\mathbf{w}_T'\mathbf{u}_t\right] \\
&= \sum_{s\le T, t\le T}\mathrm{cov}[(\mathbf{w}_T'\mathbf{B}\mathbf{f}_s)^2, \mathbf{w}_T'\mathbf{B}\mathbf{f}_t\mathbf{w}_T'\mathbf{u}_t] \\
&= \sum_{s,t}E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_s)^2\mathbf{w}_T'\mathbf{B}\mathbf{f}_t\mathbf{w}_T'\mathbf{u}_t = \sum_{s,t}E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_s)^2\mathbf{w}_T'\mathbf{B}\mathbf{f}_tE\mathbf{w}_T'\mathbf{u}_t = 0.
\end{aligned}$$

Finally, as $E\mathbf{f}_t = 0$ implies $E\mathbf{w}_T'\mathbf{B}\mathbf{f}_t = 0$, we have

$$\begin{aligned}
&\mathrm{cov}\left[\sum_{t=1}^T(\mathbf{w}_T'\mathbf{u}_t)^2, \sum_{t=1}^T\mathbf{w}_T'\mathbf{B}\mathbf{f}_t\mathbf{w}_T'\mathbf{u}_t\right] \\
&= \sum_{s\le T, t\le T}\mathrm{cov}[(\mathbf{w}_T'\mathbf{u}_s)^2, \mathbf{w}_T'\mathbf{B}\mathbf{f}_t\mathbf{w}_T'\mathbf{u}_t]
\end{aligned}$$

$$= \sum_{s,t} E(\mathbf{w}_T'\mathbf{u}_s)^2 \mathbf{w}_T'\mathbf{B}\mathbf{f}_t \mathbf{w}_T'\mathbf{u}_t$$

$$= \sum_{s,t} E(\mathbf{w}_T'\mathbf{B}\mathbf{f}_t) E(\mathbf{w}_T'\mathbf{u}_s)^2 \mathbf{w}_T'\mathbf{u}_t = 0. \quad \blacksquare$$

References

Ahn, S., Horenstein, A., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203–1227.
Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.
Antoine, B., 2011. Portfolio Selection with estimation risk: a test-based approach. J. Financ. Econom. 10, 164–197.
Antoniadis, A., Fan, J., 2001. Regularized wavelet approximations. J. Amer. Statist. Assoc. 96, 939–967.
Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–171.
Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.
Bannouh, K., Martens, M., Oomen, R., Dijk, D., 2012, Realized mixed-frequency factor models for vast dimensional covariance estimation. Manuscript.
Bianchi, D., Carvalho, C., 2011, Risk assessment in large portfolios: why imposing the wrong constraints hurts. Manuscript.
Brandt, M., Santa-Clara, P., Valkanov, R., 2009. Covariance regularization by parametric portfolio policie: exploiting characteristics in the cross-section of equity returns. Rev. Financ. Stud. 22, 3411–3447.
Brodie, J., Daubechies, I., Mol, C., Giannone, D., Loris, I., 2009. Sparse and stable Markowitz portfolios. PNAS 106, 12267–12272.
Cai, T., Liu, W., 2011. Adaptive thresholding for sparse covariance matrix estimation. J. Amer. Statist. Assoc. 106, 672–684.
Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure and mean–variance analyssi in large asset markets. Econometrica 51, 1305–1324.
Chang, C., Tsay, R., 2010. Estimation of covariance matrix via the sparse Cholesky factor with lasso. J. Statist. Plann. Inference 140, 3858–3873.
Chudik, A., Pesaran, H., Tosetti, E., 2011. Weak and strong cross section dependence and estimation of large panels. Econom. J. 14, C45–C90.
Connor, G., Korajczyk, R., 1993. A Test for the number of factors in an approximate factor model. J. Finance 48, 1263–1291.
DeMiguel, V., Garlappi, L., Nogales, F., Uppal, R., 2009a. A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. Manag. Sci. 55, 798–812.
DeMiguel, V., Garlappi, L., Nogales, F., Uppal, R., 2009b. Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? Rev. Financ. Stud. 22, 1915–1953.
El Karoui, N., 2010. High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation. Ann. Statist. 38, 3487–3566.
Fama, E., French, K., 1992. The cross-section of expected stock returns. J. Finance 47, 427–465.
Fan, J., Fan, Y., Lv, J., 2008. High dimensional covariance matrix estimation using a factor model. J. Econometrics 147, 186–197.
Fan, J., Liao, Y., Mincheva, M., 2011. High dimensional covariance matrix estimation in approximate factor models. Ann. Statist 39, 3320–3356.
Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements (with discussion). J. R. Stat. Soc. Ser. B 75, 603–680.
Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer-Verlag, New York, p. 576.
Fan, J., Zhang, J., Yu, K., 2012. Asset Allocation and Risk Assessment with Gross Exposure Constraints for Vast Portfolios. J. Amer. Statist. Assoc. 107, 592–606.
Fryzlewicz, P., 2012. High-dimensional volatility matrix estimation via wavelets and thresholding. Biometrika 100, 921–938.
Gagliardini, P., Ossola, E., Scaillet, O., 2010. Time-varying risk premium in large cross-sectional equidity datasets. Swiss Financ. Inst. Occasional Paper Ser. 11–40.
Gandy, A., Veraart, L., 2012. The effect of estimation in high-dimensional portfolios. Math. Finance 23, 531–559.
Gomez, K., Gallon, S., 2011. Comparison among high dimensional covariance matrix estimation methods. Revista Columb. Estad. 34, 567–588.
Jacquier, E., Polson, N., 2010, Simulation-based-estimation in portfolio selection. Manuscript.
Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: why imposing the wrong constraints helps. J. Finance 54, 1651–1683.
Lai, T., Xing, H., Chen, Z., 2011. Mean–variance portfolio optimization when means and covariances are unknown. Ann. Appl. Statist. 5, 798–823.
Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empirical Financ. 10, 603–621.
Merlevède, F., Peligrad, M., Rio, E., 2011. A Bernstein type inequality and moderate deviations for weakly dependent sequences. Probab. Theory Related Fields 151, 435–474.
Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Newey, W., West, K., 1994. Automatic lag selection in covariance matrix estimation. Rev. Econom. Stud. 61, 631–654.
Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Rev. Econ. Statist. 92, 1004–1016.
Peligrad, M., 1996. On the asymptotic normality of sequences of weak dependent random variables. J. Theoret. Probab. 9, 703–715.
Pesaran, M.H., Zaffaroni, P., 2008, Optimal asset allocation with factor models for large portfolios, Cesifo working paper 2326.
Rothman, A., Levina, E., Zhu, J., 2009. Generalized thresholding of large covariance matrices. J. Amer. Statist. Assoc. 104, 177–186.
Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.
Stock, J., Watson, M., 2002. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97, 1167–1179.
Yen, Y., 2013, Sparse weighted norm minimum variance portfolio. Manuscript.