



Robust inference of risks of large portfolios



Jianqing Fan^{a,*}, Fang Han^b, Han Liu^a, Byron Vickers^a

^a Department of Operations Research and Financial Engineering, Princeton University, USA

^b Department of Statistics, University of Washington, WA 98195, USA

ARTICLE INFO

Article history:

Available online 2 June 2016

JEL classification:

C58

Keywords:

High dimensionality
Robust inference
Rank statistics
Quantile statistics
Risk management

ABSTRACT

We propose a bootstrap-based robust high-confidence level upper bound (Robust H-CLUB) for assessing the risks of large portfolios. The proposed approach exploits rank-based and quantile-based estimators, and can be viewed as a robust extension of the H-CLUB procedure (Fan et al., 2015). Such an extension allows us to handle possibly misspecified models and heavy-tailed data, which are stylized features in financial returns. Under mixing conditions, we analyze the proposed approach and demonstrate its advantage over H-CLUB. We further provide thorough numerical results to back up the developed theory, and also apply the proposed method to analyze a stock market dataset.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Let $\mathbf{R}_1, \dots, \mathbf{R}_T$ be a stationary multivariate time series with $\mathbf{R}_t \in \mathbb{R}^d$ representing asset returns at time t . Letting $\mathbf{w} \in \mathbb{R}^d$ be a portfolio allocation vector, we define the risk of \mathbf{w} as

$$\text{Risk}(\mathbf{w}) := (\text{Var}(\mathbf{w}^T \mathbf{R}_t))^{1/2} = (\mathbf{w}^T \Sigma \mathbf{w})^{1/2},$$

where Σ denotes the unknown volatility (or covariance) matrix of \mathbf{R}_t . Assessing the risk of a portfolio consists of two steps: construction of a covariance matrix estimator $\widehat{\Sigma}_{\text{est}}$, and construction of a confidence interval for $\mathbf{w}^T \Sigma \mathbf{w}$ based on $\widehat{\Sigma}_{\text{est}}$.

Risk estimation is particularly challenging when d is large. For example, given a pool of 2000 candidate assets, the volatility matrix Σ involves more than 2 million parameters. However, the volatility matrix is local to the time and there are only around 252 daily returns even when the time horizon is taken as one year. This is a typical “small n , large d ” problem which leads to the accumulation of estimation errors (Pesaran and Zaffaroni, 2008; Fan et al., 2012). To handle the curse of dimensionality, more structural regularization is imposed in estimating Σ . For example, Fan et al. (2008) and Fan et al. (2013) impose the factor model structure on the covariance matrix. The assumed factor structure reduces the effective number of parameters that have to be estimated. In addition, Ledoit and Wolf (2003) propose a shrinkage estimator of Σ . Moreover, Barndorff-Nielsen (2002),

Zhang et al. (2005), and Fan et al. (2012) consider estimating Σ based on high-frequency data. Other literature includes Chang and Tsay (2010), Gómez and Gallón (2011), Lai et al. (2011), Fan et al. (2011), Bai and Liao (2015), and Fryzlewicz (2013).

However, most of these papers focus on risk estimation instead of uncertainty assessment. To construct a confidence interval for $\mathbf{w}^T \Sigma \mathbf{w}$, Fan et al. (2012) propose to use $\|\mathbf{w}\|_1^2 \|\widehat{\Sigma}_{\text{est}} - \Sigma\|_{\max}$ ¹ as an upper bound of $|\mathbf{w}^T (\widehat{\Sigma}_{\text{est}} - \Sigma) \mathbf{w}|$. However, this bound depends on the unknown Σ and has proven to be overly conservative and useless widely in numerical studies. To handle this problem, Fan et al. (2015) propose a high-confidence level upper bound (H-CLUB) of $|\mathbf{w}^T (\widehat{\Sigma}_{\text{est}} - \Sigma) \mathbf{w}|$: for a given confidence level $1 - \gamma$, under certain moment and dependence assumptions on the time series, the derived H-CLUB estimate proves to dominate $|\mathbf{w}^T (\widehat{\Sigma}_{\text{est}} - \Sigma) \mathbf{w}|$ with probability approximating $1 - \gamma$ as both T and d increase to infinity. They show that the resulting confidence bands are informative enough for most practical applications.

This paper proposes new methods for uncertainty assessment of risks of large portfolios for high dimensional heavy-tailed data. In particular, we derive confidence intervals for $\mathbf{w}^T \Sigma \mathbf{w}$ when the asset returns $\mathbf{R}_1, \dots, \mathbf{R}_T$ are elliptically distributed. This setting has been commonly adopted in financial econometrics (Cont, 2001). To handle heavy-tailed data, we propose a new risk uncertainty assessment method named robust high-confidence level upper bound (Robust H-CLUB). Robust H-CLUB exploits a

* Correspondence to: Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA.

E-mail address: jqfan@princeton.edu (J. Fan).

¹ We will provide the definitions of the vector ℓ_1 norm ($\|\cdot\|_1$) and matrix ℓ_{\max} norm $\|\cdot\|_{\max}$ later.

new block-bootstrap-based approach for uncertainty assessment of Risk(\mathbf{w}). More specifically, we decompose the problem of assessing the risk $\mathbf{w}^T \Sigma \mathbf{w}$ and quantifying its uncertainty into two parts: (i) we propose a robust estimator $\widehat{\Sigma}_{\text{est}}$ of Σ ; and (ii) we derive the variance of $\mathbf{w}^T (\widehat{\Sigma}_{\text{est}} - \Sigma) \mathbf{w}$. To estimate Σ , we exploit rank-based Kendall's tau estimators and quantile-based median absolute deviation estimators. To estimate the variance of $\mathbf{w}^T (\widehat{\Sigma}_{\text{est}} - \Sigma) \mathbf{w}$, we employ the circular block bootstrap method (Politis and Romano, 1992).

Theoretically, when $T, d \rightarrow \infty$ and d is possibly much larger than T , we develop an inferential theory of the robust risk estimators. In particular, we show that $\sqrt{T} \mathbf{w}^T (\widehat{\Sigma}_{\text{est}} - \Sigma) \mathbf{w}$ is asymptotically normal with variance σ^2 , and the block-bootstrap-based estimator $\widehat{\sigma}_{\text{est}}^2$ of σ^2 is consistent. The theory holds even when d is nearly exponentially larger than T . Moreover, it holds under any elliptical model. Thus we no longer need strong moment conditions (e.g., exponentially decaying rate on tails of distributions) on the asset returns. This is particularly important for applications to the financial returns, as heavy tails are stylized features.

Our proposed procedure is fundamentally different from H-CLUB, proposed in Fan et al. (2015). The main differences are threefold. Methodologically: (i) we employ rank- and quantile-based statistics to approximate the risk, in contrast to the moment-based statistics in the H-CLUB procedure. Both of these are introduced to reduce the moment conditions of the underlying data. Hence, our estimators are intrinsically more robust to heavy tails and outliers. (ii) We employ the circular block bootstrap procedure to obtain the robust high-confidence level upper bound. In comparison, H-CLUB directly calculates the main part of asymptotic variances of the risk estimators, with the residuals left un-explored. And theoretically: (iii) we develop novel theories for analyzing rank- and quantile-based statistics under mixing conditions. These theories have potential impact on analyzing many other problems, and are of independent interest.

1.1. Other related work

There is a vast literature on estimating large sparse/factor-based covariance matrices. Under the assumption that data points are mutually independent, many sample covariance based regularization methods, including banding (Bickel and Levina, 2008b), tapering (Cai et al., 2010), thresholding (Bickel and Levina, 2008a; Cai and Zhou, 2012), and factor structures (Fan et al., 2008; Agarwal et al., 2012; Hsu et al., 2011), have been proposed. They are further applied to study stationary time series data under vector autoregressive dependence (Loh and Wainwright, 2012; Han et al., 2015), mixing conditions (Pan and Yao, 2008; Fan et al., 2011, 2013; Han and Liu, 2013), and physical dependence (Xiao and Wu, 2012; Chen et al., 2013).

This paper is also related to the literature on estimating large correlation/covariance matrices under misspecified or heavy-tailed models. For example, Han and Liu (2014), Han and Liu (2016), Wegkamp and Zhao (2016), and Mitra and Zhang (2014) exploit rank statistics, while Qiu et al. (2015a) focus on quantile statistics. None of these works study the risk inference problem as in our paper.

1.2. Notation

Let $\mathbf{v} = (v_1, \dots, v_d)^T$ be a d dimensional real vector and $\mathbf{M} = [M_{jk}]$ be a d by d real matrix. For $0 < q < \infty$, let the vector ℓ_q norm be $\|\mathbf{v}\|_q := (\sum_{j=1}^d |v_j|^q)^{1/q}$ and the vector ℓ_∞ norm be $\|\mathbf{v}\|_\infty := \max_{j=1}^d |v_j|$. For two subsets $I, J \in \{1, \dots, d\}$, we denote \mathbf{v}_I and $\mathbf{M}_{I,J}$ as the sub-vector of \mathbf{v} with entries indexed by I and sub-matrix of \mathbf{M} with rows and columns indexed by I and J . We denote the matrix

ℓ_{\max} norm of \mathbf{M} as $\|\mathbf{M}\|_{\max} := \max_{jk} |M_{jk}|$. Letting $\mathbf{N} = [N_{jk}] \in \mathbb{R}^{d \times d}$ be another d by d real matrix, we denote by $\mathbf{M} \circ \mathbf{N} = [M_{jk} N_{jk}]$ the Hadamard product between \mathbf{M} and \mathbf{N} . Letting $f : \mathbb{R} \rightarrow \mathbb{R}$ be a real function, we denote by $f(\mathbf{M}) = [f(M_{jk})]$ the matrix with $f(M_{jk})$ as its (j, k) entry. We write $\mathbf{M} = \text{diag}(\mathbf{M}_1, \dots, \mathbf{M}_k)$ if \mathbf{M} is block diagonal with diagonal matrices $\mathbf{M}_1, \dots, \mathbf{M}_k$. For random vectors $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^d$, we write $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$ if \mathbf{X} and \mathbf{Y} are identically distributed. Throughout the paper, we use c, c_1, c_2, \dots , and C, C_1, C_2, \dots to represent generic absolute positive constants, for which the actual values may change from one line to another. For any real positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \gtrsim b_n$ if we have $a_n \geq c b_n$ for some absolute constant c and all large enough n . We write $a_n \lesssim b_n$ if we have $b_n \gtrsim a_n$, and $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $a_n \gtrsim b_n$. For $a \in \mathbb{R}$, we define $\lceil a \rceil$ and $\lfloor a \rfloor$ to be the smallest integer larger than a and the largest integer smaller than a respectively.

1.3. Paper organization

The rest of this paper is organized as follows. Section 2 introduces the Robust H-CLUB procedure for assessing the uncertainty of the portfolio risk. We consider three settings: (i) the marginal variances of the returns are known; (ii) the marginal variances are unknown, but additional information exists to help determine their values; and (iii) the marginal variances are unknown and there is no additional information available. Section 3 presents the inferential theory for the risk estimators and justifies the use of Robust H-CLUB. Sections 4 and 5 present synthetic and real data analyses to back up the developed theory. Proofs, as well as some additional materials not included here, are presented in supplementary materials (see Appendix A).

2. Robust H-CLUB

This section introduces the Robust H-CLUB procedure. We consider a multivariate stationary time series of asset returns $\mathbf{R}_1, \dots, \mathbf{R}_T$ with $\mathbf{R}_t = (R_{t1}, \dots, R_{td})^T \in \mathbb{R}^d$ for $t = 1, \dots, T$. Let $\Sigma := \text{Cov}(\mathbf{R}_t)$ be the covariance matrix and $\mathbf{D} \in \mathbb{R}^{d \times d}$ be a diagonal matrix with diagonals $\Sigma_{11}^{1/2}, \dots, \Sigma_{dd}^{1/2}$. It is easy to derive $\Sigma = \mathbf{D} \Sigma^0 \mathbf{D}$, where Σ^0 is the correlation matrix of \mathbf{R}_t . For a given portfolio allocation vector $\mathbf{w} \in \mathbb{R}^d$, we aim to construct a confidence interval for $\mathbf{w}^T \Sigma \mathbf{w}$. Throughout this section, our interest is on analyzing heavy-tailed returns, which are common in financial applications.

We exploit the elliptical distribution family to model heavy-tailed data. The elliptical distribution is routinely used in modeling financial data (Owen and Rabinovitch, 1983; Hamada and Valdez, 2004; Frahm and Jaekel, 2007). More specifically, a random vector $\mathbf{Z} \in \mathbb{R}^d$ follows an elliptical distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^d$ and positive definite covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$ if

$$\mathbf{Z} \stackrel{d}{=} \boldsymbol{\mu} + \xi \mathbf{A} \mathbf{U},$$

where $\mathbf{A} \in \mathbb{R}^{d \times d}$ satisfies $\mathbf{A} \mathbf{A}^T = \Sigma$, $\mathbf{U} \in \mathbb{R}^d$ is uniformly distributed on the d -dimensional sphere \mathbb{S}^{d-1} , and ξ is an unspecified nonnegative random variable independent of \mathbf{U} satisfying $\mathbb{E} \xi^2 = d$.

For parameter estimation, we define rank-based Kendall's tau correlation coefficient and quantile-based median absolute deviation estimators. We first introduce the Kendall's tau statistic. Given $\mathbf{R}_1, \dots, \mathbf{R}_T$, the sample and population Kendall's tau matrices $\widehat{\mathbf{T}} = [\widehat{\tau}_{jk}]$ and $\mathbf{T} = [\tau_{jk}]$ are defined as

$$\begin{aligned} \widehat{\tau}_{jk} &:= \frac{2}{T(T-1)} \sum_{t < t'} \text{sign}(R_{tj} - R_{t'j}) \text{sign}(R_{tk} - R_{t'k}), \\ \tau_{jk} &:= \mathbb{E} \text{sign}(R_j - \widetilde{R}_j) \text{sign}(R_k - \widetilde{R}_k), \end{aligned} \tag{2.1}$$

where $\mathbf{R} = (R_1, \dots, R_d)^\top$ and $\tilde{\mathbf{R}} = (\tilde{R}_1, \dots, \tilde{R}_d)^\top$ are two independent copies of \mathbf{R}_1 . Under the elliptical model, the Kendall's tau matrix \mathbf{T} and correlation matrix Σ^0 satisfy (Lindskog et al., 2003):

$$\Sigma_{jk}^0 = \sin\left(\frac{\pi}{2} \tau_{jk}\right). \tag{2.2}$$

Secondly, we define the quantile-based median absolute deviation estimator of the scale parameter. We start with some extra notation. Let $X \in \mathbb{R}$ be a random variable and $\{X_1, \dots, X_T\}$ be T realizations of X . For any $q \in [0, 1]$, we define the population and sample q -quantiles as

$$Q(X; q) := \inf\{x : \mathbb{P}(X \leq x) \geq q\},$$

$$\widehat{Q}(\{X_t\}; q) := X^{(k)}, \quad \text{where } k = \min\left\{t : \frac{t}{T} \geq q\right\}. \tag{2.3}$$

Here $X^{(1)} \leq X^{(2)} \leq \dots \leq X^{(T)}$ is the ordered sequence of X_1, \dots, X_T .

We then define the population and sample median absolute deviations for $\{X_1, \dots, X_T\}$ as the population and sample medians of absolute values of the centered data. The formal definitions are as follows:

$$\sigma_M(X) := Q\left(\left\{\left|X - Q\left(X; \frac{1}{2}\right)\right|\right\}; \frac{1}{2}\right),$$

$$\widehat{\sigma}_M(\{X_t\}_{t=1}^T) := \widehat{Q}\left(\left\{\left|X_t - \widehat{Q}\left(\{X_t\}_{t=1}^T; \frac{1}{2}\right)\right|\right\}_{t=1}^T; \frac{1}{2}\right). \tag{2.4}$$

They are robust alternatives to the population and sample standard deviations. In particular, for an elliptically distributed random vector $\mathbf{R} = (R_1, \dots, R_d)^\top$, Han et al. (2014) prove

$$\frac{\sigma_M(R_1)}{\text{sd}(R_1)} = \frac{\sigma_M(R_2)}{\text{sd}(R_2)} = \dots = \frac{\sigma_M(R_d)}{\text{sd}(R_d)}, \tag{2.5}$$

where for arbitrary random variable X , $\text{sd}(X)$ represents the standard deviation of X .

Under the elliptical model and using the rank- and quantile-based estimators, we propose three robust approaches to construct the confidence interval of $\mathbf{w}^\top \Sigma \mathbf{w}$. Formally speaking, for each proposed robust covariance matrix estimator $\widehat{\Sigma}_{\text{est}}$ and any given $\gamma > 0$, we aim to find a $\widehat{U}_{\text{est}}(\gamma)$ such that

$$\mathbb{P}\left(\mathbf{w}^\top \Sigma \mathbf{w} \in \left[\mathbf{w}^\top \widehat{\Sigma}_{\text{est}} \mathbf{w} - \widehat{U}_{\text{est}}(\gamma), \mathbf{w}^\top \widehat{\Sigma}_{\text{est}} \mathbf{w} + \widehat{U}_{\text{est}}(\gamma)\right]\right) \rightarrow 1 - \gamma,$$

as $T, d \rightarrow \infty$. The proposed approaches correspond to three scenarios where \mathbf{D} has different structures.

Of note, a main strategy throughout the proposed three methods is to separately estimate the marginal standard deviations and bivariate correlation coefficients. This paper focuses on measuring the uncertainty introduced in estimating the correlation coefficients, while assuming that the uncertainty introduced in estimating marginal standard deviations is negligible³.

To measure the uncertainty in estimation of the correlation coefficients, we employ a circular block bootstrap method.

In detail, suppose that we derive a robust marginal standard deviation estimator $\widehat{\mathbf{D}}_{\text{est}}$ of \mathbf{D} . We further derive the correlation matrix estimator $\widehat{\Sigma}_{\text{est}}^0$ of Σ^0 based on a d -dimensional multivariate time series $\mathbf{X}_1, \dots, \mathbf{X}_T$. For any given portfolio allocation vector \mathbf{w} , we propose to estimate $\mathbf{w}^\top \Sigma \mathbf{w}$ by

$$\widehat{\text{Risk}}(\mathbf{w}) := \mathbf{w}^\top \widehat{\Sigma}_{\text{est}} \mathbf{w}, \quad \text{where } \widehat{\Sigma}_{\text{est}} := \widehat{\mathbf{D}}_{\text{est}} \widehat{\Sigma}_{\text{est}}^0 \widehat{\mathbf{D}}_{\text{est}}. \tag{2.6}$$

² Let F and f be the distribution function and density function of X . We will use $Q(X; q)$, $Q(F; q)$, and $Q(f; q)$ interchangeably.

³ This is mainly for the purpose of constructing the bootstrap-based inferential theory.

To estimate the asymptotic variance of the estimator $\mathbf{w}^\top \widehat{\Sigma}_{\text{est}} \mathbf{w}$, we adopt a circular block bootstrap procedure introduced in Politis and Romano (1992). First, we extend the sample $\mathbf{X}_1, \dots, \mathbf{X}_T$ periodically by concatenating $\mathbf{X}_{i+T} = \mathbf{X}_i$ for $i \geq 1$. We then randomly select a block of $l = l_T \asymp T^{1-\epsilon_0}$ consecutive observations from the extended sample for some absolute constant $\epsilon_0 < 1$ (e.g., we can pick ϵ_0 to be 1/2). As the financial returns admit only weak dependence, the choice of block size l is not very important. We repeat this process $b = \lfloor T/l \rfloor$ times independently to obtain a sample $\mathbf{X}_1^*, \dots, \mathbf{X}_T^*$, so that for each $k = 0, \dots, b - 1$,

$$\mathbb{P}^*(\mathbf{X}_{kl+1}^* = \mathbf{X}_j, \dots, \mathbf{X}_{(k+1)l}^* = \mathbf{X}_{j+l-1}) = 1/T, \quad \text{for } j = 1, \dots, T,$$

where \mathbb{P}^* is the resampling distribution conditional on $\mathbf{X}_1, \dots, \mathbf{X}_T$. Based on each resampled time series $\mathbf{X}_1^*, \dots, \mathbf{X}_T^*$, we calculate the correlation matrix estimator $\widehat{\Sigma}_{\text{est}}^{0*}$. Let $\widehat{\Sigma}_{\text{est}}^* := \widehat{\mathbf{D}}_{\text{est}} \widehat{\Sigma}_{\text{est}}^{0*} \widehat{\mathbf{D}}_{\text{est}}$ be the estimator of Σ based on the resampled data and $\text{Var}^*(\cdot)$ be the variance operator of the probability mass function \mathbb{P}^* . We estimate the asymptotic variance of $\mathbf{w}^\top \widehat{\Sigma}_{\text{est}} \mathbf{w}$ by

$$\widehat{\sigma}_{\text{est}}^2 := \text{Var}^*(\sqrt{T} \mathbf{w}^\top \widehat{\Sigma}_{\text{est}}^* \mathbf{w}).$$

Remark 2.1. In practice, there are many other nonparametric overlapping and non-overlapping block bootstrap approaches for time series data, including moving block bootstrap, block-of-blocks bootstrap, and stationary bootstrap. We refer to Bühlmann (2002) for an excellent review on this area. We believe, under similar stationary conditions, the block-of-blocks bootstrap and stationary bootstrap approaches shall also work in our context. In particular, for short-memory and stationary time series data, the circular block bootstrap procedure has proven to work very well (Shao and Yu, 1993; Bühlmann, 2002).

Remark 2.2. It is worth comparing the bootstrap-based inference proposed here to the asymptotic methods exploited in H-CLUB (Fan et al., 2015). Specifically, the asymptotic methods aim to calculate the main term of the asymptotic variances of the estimators, and leave the residuals unexplored. However, by discarding the residual term, such an asymptotic approach leads to possible information loss. In contrast, exploiting the block bootstrap procedure to approximate the asymptotic variance intrinsically allows us to exploit all available information.

2.1. Known marginal volatilities

In this section we consider the setting where the marginal standard deviations of \mathbf{R}_t , encoded in \mathbf{D} , are known. While this is an ideal assumption, a practical implementation is to fit a parametric model such as the GARCH(1, 1) model introduced in Bollerslev (1986) to each individual return time series. Such estimates are much more accurate than the nonparametric ones and can be ideally treated as known.

When \mathbf{D} is known, estimating $\mathbf{w}^\top \Sigma \mathbf{w}$ reduces to estimating the correlation matrix Σ^0 . Using (2.2), under the elliptical model, we focus on the covariance matrix estimator $\widehat{\Sigma}$ with $\widehat{\Sigma} := \mathbf{D} \sin(\pi \widehat{\mathbf{T}}/2) \mathbf{D}$. We then estimate $\mathbf{w}^\top \Sigma \mathbf{w}$ via replacing $\widehat{\Sigma}_{\text{est}}$ by $\widehat{\Sigma}$ in (2.6). Let $\widehat{\sigma}^2$ be an estimator of the asymptotic variance σ^2 of $\mathbf{w}^\top \widehat{\Sigma} \mathbf{w}$. We calculate $\widehat{\sigma}^2$ based on the circular block bootstrap method introduced earlier. Let $\Phi(\cdot)$ be the cumulative distribution function of a standard Gaussian random variable. For any given confidence level $1 - \gamma \in (0, 1)$, we define the Robust H-CLUB estimator $\widehat{U}(\gamma)$ as

$$\widehat{U}(\gamma) := \Phi^{-1}(1 - \gamma/2) \sqrt{\widehat{\sigma}^2/T}. \tag{2.7}$$

The corresponding confidence interval for the risk is

$$[\mathbf{w}^\top \widehat{\Sigma} \mathbf{w} - \widehat{U}(\gamma), \mathbf{w}^\top \widehat{\Sigma} \mathbf{w} + \widehat{U}(\gamma)]. \tag{2.8}$$

In Section 3 we will show that, under mild conditions,

$$\widehat{\sigma}^2 = \sigma^2(1 + o_p(1)) \quad \text{and}$$

$$\mathbb{P}\left\{|\mathbf{w}^\top(\widehat{\Sigma} - \Sigma)\mathbf{w}| \leq \widehat{U}_\tau(\gamma)\right\} \rightarrow 1 - \gamma,$$

as T and d go to infinity. Therefore $[\mathbf{w}^\top \widehat{\Sigma} \mathbf{w} - \widehat{U}(\gamma), \mathbf{w}^\top \widehat{\Sigma} \mathbf{w} + \widehat{U}(\gamma)]$ is a valid level $(1 - \gamma)$ 100% interval covering the true $\mathbf{w}^\top \Sigma \mathbf{w}$.

2.2. Additional data

To adapt to current market conditions, we usually pick a short time series such that the asset returns are approximately stationary. However, it is likely that each univariate time series is stationary over a longer time scale than the multivariate time series, and hence we can incorporate extra information into calculation of the marginal standard deviations.

Inspired by this, this section considers a setting where historical information is available. We do not assume the historical data to be multivariately stationary, but only marginally stationary. Formally speaking, let $\mathbf{R}_1, \dots, \mathbf{R}_T$ be the observed stationary multivariate time series, and $\mathbf{H}_1, \dots, \mathbf{H}_{T_h}$ be the available historical data with $\mathbf{H}_t = (H_{t1}, \dots, H_{td})^\top$ and

$$T = O(T_h^{1-\delta}), \quad \text{where } \delta \text{ is an absolute constant.} \tag{2.9}$$

$\mathbf{H}_1, \dots, \mathbf{H}_{T_h}$ could have overlap with $\mathbf{R}_1, \dots, \mathbf{R}_T$. However, \mathbf{H}_t is not necessarily identically distributed to either $\mathbf{H}_{t'}$ or \mathbf{R}_1 for any $t \neq t' \in \{1, \dots, T_h\}$. Instead, we only assume that

$$H_{1j} \stackrel{d}{=} H_{2j} \stackrel{d}{=} \dots \stackrel{d}{=} H_{T_h j} \quad \text{and} \quad \text{Var}(H_{1j}) = \text{Var}(R_{1j}),$$

for $j \in \{1, \dots, d\}$.

We then estimate $\mathbf{w}^\top \Sigma \mathbf{w}$ by separately estimating \mathbf{D} and Σ^0 .

Formally, to estimate \mathbf{D} , we use the historical data $\mathbf{H}_1, \dots, \mathbf{H}_{T_h}$ and derive

$$\widehat{\mathbf{D}}^h = (\widehat{\mathbf{D}}_{11}^h, \dots, \widehat{\mathbf{D}}_{dd}^h), \quad \text{where } \widehat{\mathbf{D}}_{jj}^h := \widehat{\sigma}_{M,j}^h \frac{\widehat{\sigma}_1^h}{\widehat{\sigma}_{M,1}^h}, \tag{2.10}$$

and $\widehat{\sigma}_{M,j}^h = \widehat{\sigma}_M(\{H_{tj}\}_{t=1}^T)$, for $j = 1, \dots, d$, is the median absolute deviation estimator based on $\{H_{tj}\}_{t=1}^T$, and $\widehat{\sigma}_1^h = (\widehat{\text{Var}}(\{H_{t1}\}_{t=1}^T))^{1/2}$ is the Pearson sample standard deviation of $\{H_{t1}\}_{t=1}^T$. To estimate Σ^0 we calculate the Kendall's tau matrix $\widehat{\mathbf{T}}$ based on $\{\mathbf{R}_1, \dots, \mathbf{R}_T\}$.

Remark 2.3. In (2.10), to calculate $\widehat{\mathbf{D}}^h$, we employ the term $\widehat{\sigma}_1^h / \widehat{\sigma}_{M,1}^h$ to approximate the scaling factor between the median absolute deviation and the Pearson's standard deviation. This facilitates theoretical derivations. In practice, we can use, for example, the average version $\sum_{j=1}^d \widehat{\sigma}_j^h / \sum_{j=1}^d \widehat{\sigma}_{M,j}^h$ to estimate the scaling factor. We could also replace the sample standard deviation estimator by the Catoni's M -estimator (Catoni, 2012) or less biased robust approximate quadratic (RA-quadratic) estimator (Fan et al., 2016), which are appealing alternatives in handling heavy-tailed data.

In estimating $\mathbf{w}^\top \Sigma \mathbf{w}$, we replace $\widehat{\mathbf{D}}_{\text{est}}$ by $\widehat{\mathbf{D}}^h$, $\widehat{\Sigma}_{\text{est}}^0$ by $\sin(\pi \widehat{\mathbf{T}}/2)$, and $\widehat{\Sigma}_{\text{est}}$ by $\widehat{\Sigma}^h$ in (2.6). For any given $1 - \gamma \in (0, 1)$, we calculate the Robust H-CLUB estimator $\widehat{U}^h(\gamma)$ as

$$\widehat{U}^h(\gamma) = \Phi^{-1}(1 - \gamma/2) \sqrt{\widehat{\sigma}_h^2 / T}, \tag{2.11}$$

where $\widehat{\sigma}_h^2$ is calculated by employing the circular block bootstrap method introduced earlier. The corresponding confidence interval for the risk is

$$[\mathbf{w}^\top \widehat{\Sigma}^h \mathbf{w} - \widehat{U}^h(\gamma), \mathbf{w}^\top \widehat{\Sigma}^h \mathbf{w} + \widehat{U}^h(\gamma)]. \tag{2.12}$$

2.3. Unknown marginal volatilities

This section considers the setting that \mathbf{D} is unknown with no additional data available. In this setting we use a data splitting strategy to separately estimate \mathbf{D} and Σ^0 . More precisely, we estimate \mathbf{D} using the whole dataset:

$$\widehat{\mathbf{D}} = (\widehat{\mathbf{D}}_{11}, \dots, \widehat{\mathbf{D}}_{dd}), \quad \text{with } \widehat{\mathbf{D}}_{jj} := \widehat{\sigma}_{M,j} \frac{\widehat{\sigma}_1}{\widehat{\sigma}_{M,1}}, \tag{2.13}$$

where $\widehat{\sigma}_{M,j} = \widehat{\sigma}_M(\{R_{tj}\}_{t=1}^T)$ for $j = 1, \dots, d$ and $\widehat{\sigma}_1 = (\widehat{\text{Var}}(\{R_{t1}\}_{t=1}^T))^{1/2}$ is the Pearson sample standard deviation of $\{R_{t1}\}_{t=1}^T$. To estimate Σ^0 , we extract a subsequence $\mathbf{R}_{T-T_s+1}, \dots, \mathbf{R}_T$ from the time series $\mathbf{R}_1, \dots, \mathbf{R}_T$, where $T_s \asymp T^{1-\delta}$ with δ a small enough absolute constant. Using this subsequence, we calculate the Kendall's tau matrix $\widehat{\mathbf{T}}^s$. Combining it with $\widehat{\mathbf{D}}$, we obtain a robust covariance matrix estimator

$$\widehat{\Sigma}^s := \widehat{\mathbf{D}} \sin\left(\frac{\pi}{2} \widehat{\mathbf{T}}^s\right) \widehat{\mathbf{D}}.$$

We estimate $\mathbf{w}^\top \Sigma \mathbf{w}$ by replacing $\widehat{\mathbf{D}}_{\text{est}}$, $\widehat{\Sigma}_{\text{est}}^0$, and $\widehat{\Sigma}_{\text{est}}$ by $\widehat{\mathbf{D}}$, $\sin(\frac{\pi}{2} \widehat{\mathbf{T}}^s)$, and $\widehat{\Sigma}^s$ in (2.6). We then obtain a Robust H-CLUB estimator as

$$\widehat{U}^s(\gamma) = \Phi^{-1}(1 - \gamma/2) \sqrt{\widehat{\sigma}_s^2 / T_s}, \tag{2.14}$$

where $\widehat{\sigma}_s^2$ is calculated by employing the circular block bootstrap method. Accordingly, we construct the confidence interval of the risk as

$$[\mathbf{w}^\top \widehat{\Sigma}^s \mathbf{w} - \widehat{U}^s(\gamma), \mathbf{w}^\top \widehat{\Sigma}^s \mathbf{w} + \widehat{U}^s(\gamma)]. \tag{2.15}$$

Remark 2.4. In (2.13) we can employ an average version of the scaling factor, as in Remark 2.3. We also note that the data splitting strategy is mainly proposed for theoretical analysis. In practice, we can set $\delta = 0$ and use the entire data set in calculating $\widehat{\Sigma}^s$ and performing the block bootstrap.

3. Asymptotic theory

In this section we prove that the confidence intervals of $\mathbf{w}^\top \Sigma \mathbf{w}$ corresponding to the three settings discussed in Section 2 have desired coverage probability. In other words, we prove that the Robust H-CLUB estimators proposed in (2.7), (2.11), and (2.14) are asymptotic $(1 - \gamma)$ 100% confidence upper bounds for the risk. It is clear that this problem reduces to calculating the limiting distributions of $\mathbf{w}^\top(\widehat{\Sigma}_{\text{est}} - \Sigma)\mathbf{w}$ for $\widehat{\Sigma}_{\text{est}} = \widehat{\Sigma}$, $\widehat{\Sigma}^h$, and $\widehat{\Sigma}^s$. In the sequel, we adopt the triangular array setting as in Fan and Peng (2004) and Greenshtein and Ritov (2004), and allow the dimension d to increase with the sample size n .

We start with an introduction of the ϕ -mixing coefficient. For a d -dimensional stationary process $\{\mathbf{R}_t\}_{t \in \mathbb{Z}}$, let \mathcal{F}_a^b be the σ -algebra generated by $\mathbf{R}_a, \dots, \mathbf{R}_b$ for $a \leq b$. We define the ϕ -mixing coefficient as follows (Fan and Yao, 2003):

$$\phi(n) := \sup_{B \in \mathcal{F}_{-\infty}^0, A \in \mathcal{F}_n^\infty, \mathbb{P}(B) > 0} |\mathbb{P}(A|B) - \mathbb{P}(A)|.$$

Suppose that $\{\mathbf{R}_1, \dots, \mathbf{R}_T\}$ is a subsequence of the stationary process $\{\mathbf{R}_t\}_{t \in \mathbb{Z}}$. Let F be the distribution function of \mathbf{R}_1 . For $\mathbf{a} := \mathbf{D}\mathbf{w} = (a_1, \dots, a_d)^\top$, let $g : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a kernel function

$$g(\mathbf{R}_t, \mathbf{R}_{t'}) := \frac{\pi}{2} \sum_{j \neq k} a_j a_k \cos\left(\frac{\pi}{2} \tau_{jk}\right) \text{sign}(R_{tj} - R_{t'j}) \text{sign}(R_{tk} - R_{t'k}). \tag{3.1}$$

We further define the following three quantities which will be useful in later sections:

$$g_1(\mathbf{R}_1) := \int g(\mathbf{R}_1, \mathbf{R}_2) dF(\mathbf{R}_2), \tag{3.2}$$

$$\theta := \int g(\mathbf{R}_1, \mathbf{R}_2) dF(\mathbf{R}_1) dF(\mathbf{R}_2) = \mathbf{a}^\top \left\{ \cos\left(\frac{\pi}{2} \mathbf{T}\right) \circ \frac{\pi}{2} \mathbf{T} \right\} \mathbf{a}, \tag{3.3}$$

$$\sigma^2 := 4 \left(\mathbb{E} g_1(\mathbf{R}_1)^2 - \theta^2 + 2 \sum_{h=1}^{\infty} \left\{ \mathbb{E} g_1(\mathbf{R}_1) g_1(\mathbf{R}_{1+h}) \right\} \right). \tag{3.4}$$

In the following, we assume that the elliptical time series model in Section 2 holds. More precisely, we assume

- **(A0).** $\mathbf{R}_1, \dots, \mathbf{R}_T$ are continuous and identically distributed as an elliptical random vector \mathbf{R} with covariance and correlation matrices Σ and Σ^0 .

3.1. Theory for known volatilities

We make the following four assumptions which regulate the portfolio allocation vector \mathbf{w} and the stationary process $\{\mathbf{R}_t\}_{t \in \mathbb{Z}}$:

- (A1)** There exist absolute constants C_1 and C_2 such that $\|\mathbf{w}\|_1 \leq C_1$ and $\|\Sigma\|_{\max} \leq C_2$.
- (A2)** σ is lower bounded by a positive absolute constant.
- (A3)** The process $\{\mathbf{R}_t\}_{t \in \mathbb{Z}}$ is ϕ -mixing with $\phi(n) \leq n^{-1-\epsilon}$ for some $\epsilon > 0$.
- (A4)** $\log d / (T^{1/2}) = o(1)$.

Assumption **(A1)** regulates the portfolio allocation vector \mathbf{w} to prevent extreme positions. It is a common assumption made for stability of the portfolio (Jagannathan and Ma, 2003; Fan et al., 2012, 2015). Assumption **(A2)** guarantees that the portfolio risk cannot be diversified away. This is mild given that the returns are commonly assumed to follow a factor model (Chamberlain, 1983; Fan et al., 2015). Assumption **(A3)** is routinely used in analyzing time series to capture serial dependence strength (Pan and Yao, 2008; Han and Liu, 2013). We also refer the readers to Fan and Yao (2003) and Longla and Peligrad (2012) for time series models satisfying ϕ -mixing conditions. Lastly, Assumption **(A4)** allows d to grow nearly exponentially faster than T and hence is mild.

In the setting of Section 2.1 and Assumptions **(A1)–(A4)**, we derive the limiting distribution of $\mathbf{w}^\top (\widehat{\Sigma} - \Sigma) \mathbf{w}$. The following theorem shows that $\sqrt{T} \mathbf{w}^\top (\widehat{\Sigma} - \Sigma) \mathbf{w} / \sigma$ is asymptotically normal.

Theorem 3.1 (CLT, Known Volatilities). *Assuming that **(A0)–(A4)** hold and in the setting of Section 2.1, we have*

$$\sqrt{T} \mathbf{w}^\top (\widehat{\Sigma} - \Sigma) \mathbf{w} / \sigma \xrightarrow{d} N(0, 1),$$

as both T and d go to infinity.

Proof. We provide a sketch of the proof here, and defer the details to the supplementary materials (see Appendix A). We split the proof to three parts. First, denoting $\mathbf{a} = \mathbf{D} \mathbf{w}$, using Taylor expansion entry-wisely on $\sin(\pi \widehat{\mathbf{T}}/2)$ at $\sin(\pi \mathbf{T}/2)$, we decompose $\mathbf{w}^\top (\widehat{\Sigma} - \Sigma) \mathbf{w}$ to two parts:

$$\begin{aligned} \mathbf{w}^\top (\widehat{\Sigma} - \Sigma) \mathbf{w} &= \mathbf{a}^\top \left\{ \sin\left(\frac{\pi}{2} \widehat{\mathbf{T}}\right) - \sin\left(\frac{\pi}{2} \mathbf{T}\right) \right\} \mathbf{a} \\ &= \underbrace{\mathbf{a}^\top \left\{ \cos\left(\frac{\pi}{2} \mathbf{T}\right) \circ \frac{\pi}{2} (\widehat{\mathbf{T}} - \mathbf{T}) \right\} \mathbf{a}}_{A_1} \\ &\quad + \underbrace{\mathbf{a}^\top \left\{ -\frac{1}{2} [\sin(\theta_{jk})] \circ \left(\frac{\pi}{2}\right)^2 (\widehat{\mathbf{T}} - \mathbf{T}) \circ (\widehat{\mathbf{T}} - \mathbf{T}) \right\} \mathbf{a}}_{A_2}, \end{aligned}$$

where for each $j, k \in \{1, \dots, d\}$, θ_{jk} lies between τ_{jk} and $\widehat{\tau}_{jk}$. Secondly, we show A_2 is asymptotically ignorable. Lastly, we prove the main term, A_1 , is a U-statistic. Using the results in Yoshihara (1976), we then show A_1 's asymptotical normality under mixing conditions. Combining the above three assertions, we complete the proof. \square

The following theorem verifies that $\widehat{\sigma}^2$ calculated using the circular block bootstrap approach is a consistent estimator of σ^2 . This result, combined with Theorem 3.1 and Slutsky's theorem, confirms that $\sqrt{T} \mathbf{w}^\top (\widehat{\Sigma} - \Sigma) \mathbf{w} / \widehat{\sigma}$ converges weakly to the standard Gaussian. Accordingly, the confidence interval in (2.8) gives a reliable coverage probability.

Theorem 3.2 (Bootstrap, Known Volatilities). *Under Assumptions **(A0)–(A4)**, we have*

$$\widehat{\sigma}^2 = \sigma^2 (1 + o_p(1)),$$

and accordingly, for any given $\gamma \in (0, 1)$, as $T, d \rightarrow \infty$, we have

$$\mathbb{P}\left(\mathbf{w}^\top \Sigma \mathbf{w} \in \left[\mathbf{w}^\top \widehat{\Sigma} \mathbf{w} - \widehat{U}(\gamma), \mathbf{w}^\top \widehat{\Sigma} \mathbf{w} + \widehat{U}(\gamma)\right]\right) \rightarrow 1 - \gamma.$$

Proof. The proof follows the same strategy as in the proof of Theorem 3.1, by separately studying the main term (a U-statistic) and the remainder term. The remainder term proves to be asymptotically ignorable, and the main term is handled via a bootstrap consistency proof for U-statistics. The details are put in the supplementary materials (see Appendix A). \square

The above two theorems only assume that the marginal second moments exist. Therefore, the Robust H-CLUB estimator naturally handles heavy-tailed data.

3.2. Theory with additional data

In this section we study the setting in Section 2.2. When \mathbf{D} is unknown, we require additional assumptions. First, the following three assumptions require that d does not grow too fast compared to n and the given time series $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ (either $\{\mathbf{R}_t\}_{t \in \mathbb{Z}}$ or $\{\mathbf{H}_t\}_{t \in \mathbb{Z}}$) is ϕ -mixing with an exponentially decaying serial dependence:

- **(A5).** $\max\{\sqrt{\log d / T^\delta}, \log d / (T^{1/2})\} = o(1)$.
- **(A6).** The process $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ is ϕ -mixing with $\phi(n) \leq C_1 \exp(-C_2 n^r)$ for some absolute constants $C_1, C_2, r > 0$.
- **(A7).** Letting $a = \max(1, 1/r)$, we require that $\log d = o(T^{1/(2a+3)})$.

Recall that δ is defined in (2.9) for characterizing the length of historical data. Secondly, we require that the returns' $(4 + \epsilon_1)$ th moments exist for some absolute constant $\epsilon_1 > 0$, and the density functions are bounded away from zero around the median:

- **(A8).** For any $j \in \{1, \dots, d\}$, $\mathbb{E}|X_{1j}|^{4+\epsilon_1} \leq C_0 < \infty$ for some constant $\epsilon_1, C_0 > 0$.
- **(A9).** Let f_j and \bar{f}_j be the density functions of X_j and $|X_j - Q(X_j; 1/2)|$. For any $j \in \{1, \dots, d\}$, we require $\inf_{|x-Q(f; 1/2)| < \kappa} f(x) \geq \eta$ for some positive absolute constants κ and η , and any $f \in \{f_j, \bar{f}_j\}$.

Under **(A0)–(A2)** and **(A5)–(A9)**, the next theorem shows that $\sqrt{T} \mathbf{w}^\top (\widehat{\Sigma}^h - \Sigma) \mathbf{w}$ is asymptotically normal.

Theorem 3.3 (CLT, Unknown Volatilities with Additional Data). *Assume that Assumptions **(A0)–(A2)** hold. In addition, assume that Assumptions **(A5)–(A7)** hold for both $\{\mathbf{R}_t\}_{t \in \mathbb{Z}}$ and the additional data $\{\mathbf{H}_t\}_{t \in \mathbb{Z}}$, and Assumptions **(A8)–(A9)** hold for $\{\mathbf{H}_t\}_{t \in \mathbb{Z}}$. Then in the setting of Section 2.2, we have*

$$\sqrt{T} \mathbf{w}^\top (\widehat{\Sigma}^h - \Sigma) \mathbf{w} / \sigma \xrightarrow{d} N(0, 1),$$

as both T and d go to infinity.

The next theorem shows that $\hat{\sigma}_h^2$ is a consistent estimator of σ^2 and accordingly the confidence interval in (2.12) is valid.

Theorem 3.4 (Bootstrap, Unknown Volatilities with Additional Data). Under the assumptions of Theorem 3.3, we have

$$\hat{\sigma}_h^2 = \sigma^2\{1 + o_p(1)\},$$

and accordingly, for any given $\gamma \in (0, 1)$, as $T, d \rightarrow \infty$, we have

$$\mathbb{P}\left(\mathbf{w}^\top \Sigma \mathbf{w} \in [\mathbf{w}^\top \hat{\Sigma}^h \mathbf{w} - \hat{U}^h(\gamma), \mathbf{w}^\top \hat{\Sigma}^h \mathbf{w} + \hat{U}^h(\gamma)]\right) \rightarrow 1 - \gamma.$$

Of note, the main strategy for proving Theorems 3.3 and 3.4 is to decompose $\mathbf{w}^\top (\hat{\Sigma}^h - \Sigma) \mathbf{w}$ to two parts. The first part corresponds to $\mathbf{w}^\top (\hat{\Sigma} - \Sigma)$ in Theorem 3.1 and could be similarly studied. The second part involves the standard deviation estimator $\hat{\mathbf{D}}^h$, proving to be asymptotically ignorable. The details are deferred to the supplementary materials (see Appendix A).

3.3. Theory with unknown marginal volatilities

Lastly we study the setting in Section 2.3. Under this setting, we use a data splitting strategy and conduct inference only on a subsequence of length $T^{1-\delta}$. The next theorem justifies the use of such an approach.

Theorem 3.5 (CLT, Unknown Marginal Volatilities). Assume that Assumptions (A0)–(A2) hold and Assumptions (A5)–(A9) hold for $\{\mathbf{R}_t\}_{t \in \mathbb{Z}}$. Then, under the setting of Section 2.3, we have

$$\sqrt{T_s} \mathbf{w}^\top (\hat{\Sigma}^s - \Sigma) \mathbf{w} / \sigma \xrightarrow{d} N(0, 1).$$

Furthermore, the bootstrap-based estimator $\hat{\sigma}_s^2$ proves to be a consistent estimator of σ^2 .

Theorem 3.6 (Bootstrap, Unknown Marginal Volatilities). Under the assumptions of Theorem 3.5, we have

$$\hat{\sigma}_s^2 = \sigma^2\{1 + o_p(1)\},$$

and accordingly, for any given $\gamma \in (0, 1)$, as $T, d \rightarrow \infty$, we have

$$\mathbb{P}\left(\mathbf{w}^\top \Sigma \mathbf{w} \in [\mathbf{w}^\top \hat{\Sigma}^s \mathbf{w} - \hat{U}^s(\gamma), \mathbf{w}^\top \hat{\Sigma}^s \mathbf{w} + \hat{U}^s(\gamma)]\right) \rightarrow 1 - \gamma.$$

Remark 3.7. Compared to the method in Fan et al. (2015), the Robust H-CLUB estimator gains substantial robustness since it only assumes that the $(4 + \epsilon_1)$ th moments exist for the marginal returns. In comparison, Fan et al. (2015) require a strong exponentially decaying rate in the tails. (See Assumption 4.1 therein, although note that for their sample covariance based approach only bounded 8th moments are actually required.) Such assumptions are often too restrictive and rarely satisfied in real applications. In addition, our approach is more robust to noise contamination. The Robust H-CLUB estimator attains the power to handle heavy-tailed data at the cost of a small T^δ efficiency. This is due to the data splitting strategy, which is an artifact of the proof. In practice, we find that the method introduced in Section 2.3 performs well.

The data splitting strategy allows the portfolio allocation vector to be random. More specifically, suppose that $\hat{\mathbf{w}}$ is calculated based on the data $\mathbf{R}_1, \dots, \mathbf{R}_T$. The next theorem shows that $\sqrt{T_s} \hat{\mathbf{w}}^\top (\hat{\Sigma}^s - \Sigma) \hat{\mathbf{w}}$ is asymptotically normal under assumptions outlined below.

Corollary 3.1. Under the assumptions in Theorem 3.5, let $\hat{\mathbf{w}} = (\hat{w}_1, \dots, \hat{w}_d)^\top$ be an estimator of $\mathbf{w} = (w_1, \dots, w_d)^\top$ satisfying

$$\mathbb{P}(|\hat{w}_j/w_j - 1| > t) \leq 2 \exp(-CTt^2) \tag{3.5}$$

for some absolute constant C , any $j \in \{1, \dots, d\}$, and any $t > 0$. We then have, as $T, d \rightarrow \infty$,

$$\sqrt{T_s} \hat{\mathbf{w}}^\top (\hat{\Sigma}^s - \Sigma) \hat{\mathbf{w}} / \sigma \xrightarrow{d} N(0, 1).$$

Proof. Compared to the proofs of the above theorems, for proving Corollary 3.1, the only extra effort is on bounding the gap between $\hat{\mathbf{w}}^\top (\hat{\Sigma}^s - \Sigma) \hat{\mathbf{w}}$ and $\mathbf{w}^\top (\hat{\Sigma}^s - \Sigma) \mathbf{w}$. In the supplementary materials, we show this gap is asymptotically ignorable (see Appendix A), and the rest follows from Theorem 3.5.

In this case, we can also employ a similar circular block bootstrap procedure to estimate the asymptotic variance of $\sqrt{T_s} \hat{\mathbf{w}}^\top (\hat{\Sigma}^s - \Sigma) \hat{\mathbf{w}}$.

4. Simulations on synthetic data

In this section we examine the finite-sample performance of the Robust H-CLUB estimators on synthetically generated data with heavy tails and noise contamination. We calculate several statistics of the estimators, following those used in Fan et al. (2015), to show the quality of the estimators. Our analysis shows that Robust H-CLUB performs well in all of the cases considered when compared to the full-confidence bound $\xi_T = \|\mathbf{w}\|_1^2 \|\hat{\Sigma}_{\text{est}} - \Sigma\|_{\text{max}}$. We observe that the 95% confidence intervals given by our proposed method are much tighter than the absolute bound given by ξ_T . We also demonstrate that Robust H-CLUB outperforms H-CLUB in the presence of heavy-tailed data. In particular, we show that H-CLUB does not achieve coverage proportions of 95% in the heavy-tailed setting, while the performance of Robust H-CLUB is consistently reliable. Lastly, we show that Robust H-CLUB also performs competitively when applied to the Gaussian data.

4.1. Calibration and parameter selection

To calibrate the parameters governing data generation in our model, we use the daily returns of the S&P 500's top 100 stocks ranked by market capitalization (as of June 29th, 2012), and the 3-month Treasury bill rates, sourced from the COMPUSTAT database (www.compustat.com) and the CSRP database (www.crsp.com), respectively. We consider the excess returns $\{\tilde{\mathbf{y}}_t\}$ over the period from July 1, 2008 to June 29, 2012. We extract the following features:

1. $\{d_i^\dagger\}_{i=1}^{100}$ with d_i^\dagger equal to the sample standard deviation of the i th stock.
2. $\Sigma^{0\dagger} = \{\Sigma_{ij}^{0\dagger}\}_{i,j=1}^{100}$, the sample correlation matrix of the observations $\tilde{\mathbf{y}}_t$.

From these, we extract the mean and variance of $\{d_i^\dagger\}_{i=1}^{100}$, denoted respectively by μ_{d^\dagger} and $\sigma_{d^\dagger}^2$. We also compute the average and standard deviation of all pairwise correlations, denoted respectively by $\mu_{\Sigma^{0\dagger}}$ and $\sigma_{\Sigma^{0\dagger}}^2$. These parameters are used to generate correlation matrices and marginal variances later on.

We also have several tuning parameters to select. We choose $T_h = \lceil T^{1/(1-\delta_h)} \rceil$ with $\delta_h = 0.1$ as the parameter determining the quantity of historical data available to the estimator $\hat{\Sigma}^h$, $l = \lceil T^{1-\epsilon_0} \rceil$ with $\epsilon_0 = 0.5$ as the parameter controlling the block size in the block bootstrap, $N_{\text{bootstrap}} = 50$ as the number of bootstrapped datasets generated, and $T_s = \lceil T^{1-\delta} \rceil$ with $\delta = 0.01$ as the parameter controlling the data-splitting used in the estimator $\hat{\Sigma}^s$. For calculating $\hat{\Sigma}^h$ and $\hat{\Sigma}^s$, we employ the ‘‘average version’’ procedure in Remark 2.3.

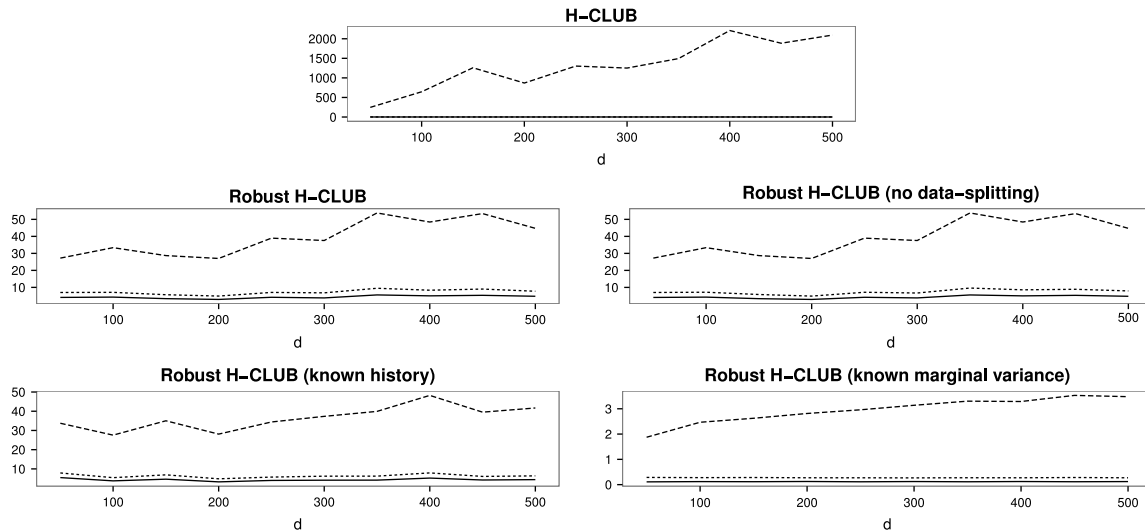


Fig. 1. Averages of $\Delta = |\mathbf{w}^T(\widehat{\Sigma}_{\text{est}} - \Sigma)\mathbf{w}|$ (solid curve), $\widehat{U}(0.05) = 2\sqrt{\text{Var}(\mathbf{w}^T\widehat{\Sigma}_{\text{est}}\mathbf{w})}$ (dashed curve), and $\xi_T = \|\mathbf{w}\|_1^2 \|\widehat{\Sigma}_{\text{est}} - \Sigma\|_{\text{max}}$ (large-dash curve) for $c = 1.0$. Horizontal axis shows dimension of problem, i.e., portfolio size. Vertical axis shows the calculated averaged values.

4.2. Simulation

For each given gross exposure constraint $c := \|\mathbf{w}\|_1$, we set $T = 300$ and allow d to range from 50 to 500 in multiples of 50. For each value of d we conduct 200 iterations of the same procedure: Generate a model, synthesize data from that model, and then calculate estimates based on the synthesized data. In detail, the data are first generated from an elliptical vector autoregressive model (Fan et al., 2011b; Qiu et al., 2015b) of order one and marginal multivariate t -distribution with 5 degrees of freedom. Secondly, 1% of them are further corrupted by a multiplier of independent $\text{Unif}(1, 15)$ noise. These two steps are designed to challenge non-robust method H-CLUB and to highlight the advantages of our new method. The detailed simulation scheme is deferred to the supplementary materials (see Appendix A).

We plot the averages of $\Delta = |\mathbf{w}^T(\widehat{\Sigma}_{\text{est}} - \Sigma)\mathbf{w}|$, $\xi_T = \|\mathbf{w}\|_1^2 \|\widehat{\Sigma}_{\text{est}} - \Sigma\|_{\text{max}}$, and $\widehat{U}(0.05) = 2\sqrt{\text{Var}(\mathbf{w}^T\widehat{\Sigma}_{\text{est}}\mathbf{w})}$ against d for every estimator considered and for $c = 1$, $c = 1.6$, and $c = 2$ to examine the effects of the dimensionality and gross exposure on risk assessment.

Next, for $d = 200$ and $d = 500$, we calculate the following quantities over the 100,000 portfolios (500 portfolios over 200 synthetic datasets; see Fan et al. (2015) for details how to generate these portfolios): the coverage proportion, defined as the fraction of the samples in which the 95% confidence interval contains the true risk $R(\mathbf{w}) = (\mathbf{w}^T \Sigma \mathbf{w})^{1/2}$; the ratio of bounds defined as $\text{RE}_1 := \xi_T / (2\sqrt{\widehat{\sigma}^2/T})$; and the relative error defined as $\text{RE}_2 := \sqrt{\widehat{\sigma}^2/T} / (2\mathbf{w}^T \Sigma \mathbf{w})$. See Fan et al. (2015). Again, we compute these for $c = 1, 1.6$, and 2 . The measure RE_1 compares the upper bound with the half width of the 95% confidence interval, whereas RE_2 is the half width of the 95% confidence interval for the portfolio risk $(\mathbf{w}^T \Sigma \mathbf{w})^{1/2}$ divided by the portfolio risk itself. The former depicts how inefficient the confidence upper bound is and the latter measures how informative the constructed confidence interval is.

Lastly, we repeat the previous calculations of coverage proportions, RE_1 and RE_2 in a setting where the data are generated from a Gaussian distribution without any noise contamination. This allows us to examine the degree of efficiency loss for robustness when the data are normal. In this setting, we also calculate the ratio $\widehat{U}(0.05)/\Delta$ as a measure of how tight the H-CLUB estimate is relative to the theoretical minimum bound.

4.3. Results

In Figs. 1 and 2, we plot the average risk estimation errors along with the estimated error bounds with gross exposure $c = 1, 1.6$, and 2 , using estimators $\widehat{\Sigma}_{\text{est}} = \widehat{\Sigma}, \widehat{\Sigma}^h, \widehat{\Sigma}^s$, and $\widehat{\Sigma}_{T_s=T}^s$. Note that $c = 1.6$ results in an average 130% long positions and 30% short positions, which is commonly used in practice. For comparison, we also plot these values for the sample covariance matrix estimator \mathbf{S} with the H-CLUB procedure of Fan et al. (2015), which we do not expect to be robust to heavy tailed data.

From these plots, we see that:

- The dashed curve lies above the solid curve throughout, an indication of the validity of the 95% bound given by $\widehat{U}(0.05)$. It is interesting to note that this still holds for the H-CLUB estimator, but this is in the average sense. As we will see in Table 1, however, the H-CLUB estimator fails to attain 95% coverage.
- The crude bound ξ_T is much larger than either the true error Δ or the 95% confidence bound $\widehat{U}(0.05)$. This discrepancy increases with d , but also with c as we can see by comparing Fig. 1 with Fig. 2. This is quantified in Table 2.
- For large d the crude bound on the H-CLUB estimator is 50–60 times larger than on any of the robust estimators. This suggests inaccurate estimation of the sample covariance in the presence of heavy tails and contamination.

Table 1 illustrates the coverage of each estimator, defined as the proportion of samples in which the 95% confidence interval captures the true variance $\mathbf{w}^T \Sigma \mathbf{w}$. It can be seen that all the Robust H-CLUB estimators have coverage proportions of approximately 95%. However, the H-CLUB estimator has substantially lower coverage. It is not sufficiently robust to give a valid bound under the current setting (the data were generated with heavy tails and corrupted by a scale mixture).

We make further comparisons between the Robust H-CLUB estimators we have proposed. Table 2 illustrates averages and standard deviations of the ratio $\text{RE}_1 = \xi_T/\widehat{U}(0.05)$: the ratio between the full confidence bound and the Robust H-CLUB estimate. These serve to quantify some of our observations made on Figs. 1 and 2 – in particular, that the ratio $\xi_T/\widehat{U}(0.05)$ increases strongly with c and weakly with d .

We observe that:

- The value of RE_1 is considerably greater than 1, reflecting the fact that the confidence interval given by Robust H-CLUB is

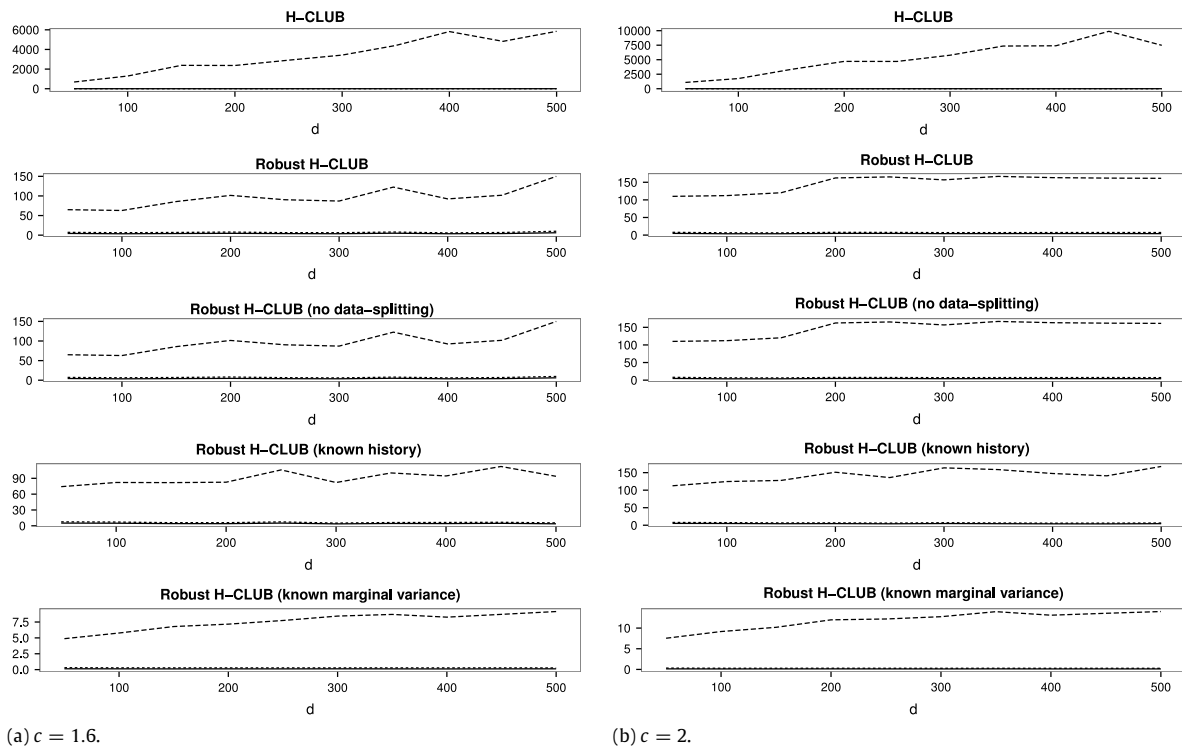


Fig. 2. Averages of $\Delta = \|\mathbf{w}^\top(\widehat{\Sigma}_{\text{est}} - \Sigma)\mathbf{w}\|$ (solid curve), $\widehat{U}(0.05) = 2\sqrt{\widehat{\text{Var}}(\mathbf{w}^\top \widehat{\Sigma}_{\text{est}} \mathbf{w})}$ (dashed curve) and $\xi_T = \|\mathbf{w}\|_1^2 \|\widehat{\Sigma}_{\text{est}} - \Sigma\|_{\max}$ (large-dash curve) for $c = 1.6$ and $c = 2$. Horizontal axis shows dimension of problem, i.e., portfolio size. Vertical axis shows the calculated averaged values.

Table 1
Empirical coverage proportion for 95% confidence intervals in settings of data drawn from t_5 distribution with 1% noise contamination, taken over 200 samples with $T = 300$.

	$d = 200$			$d = 500$		
	$c = 1.0$	$c = 1.6$	$c = 2.0$	$c = 1.0$	$c = 1.6$	$c = 2.0$
Coverage $\widehat{\mathbf{S}}$	77.96%	75.17%	70.22%	80.25%	82.01%	78.05%
Coverage $\widehat{\Sigma}^{s**}$	94.53%	92.49%	94.97%	95.51%	96.71%	95.29%
Coverage $\widehat{\Sigma}_{T_s=T}^{s**}$	94.98%	93.16%	95.35%	94.32%	97.29%	96.32%
Coverage $\widehat{\Sigma}^{h**}$	93.78%	93.13%	93.02%	92.95%	95.33%	94.88%
Coverage $\widehat{\Sigma}^{**}$	93.53%	94.42%	92.19%	91.97%	92.59%	93.11%

* H-CLUB.
** Robust H-CLUB.

Table 2
Averages and standard deviations (in parentheses) of $RE_1 := \xi_T / (2\sqrt{\widehat{\sigma}^2/T})$ over 200 samples.

	$d = 200$			$d = 500$		
	$c = 1.0$	$c = 1.6$	$c = 2.0$	$c = 1.0$	$c = 1.6$	$c = 2.0$
RE_1 $\widehat{\Sigma}^{s**}$	6.01 (2.21)	15.88 (7.08)	23.34 (8.34)	6.94 (2.28)	18.08 (6.73)	28.21 (9.57)
RE_1 $\widehat{\Sigma}_{T_s=T}^{s**}$	5.88 (2.08)	15.65 (7.15)	23.25 (8.45)	7.33 (2.64)	17.69 (6.38)	27.73 (9.43)
RE_1 $\widehat{\Sigma}^{h**}$	6.43 (2.32)	16.49 (6.04)	24.24 (9.89)	7.46 (2.90)	18.70 (6.16)	29.33 (10.32)
RE_1 $\widehat{\Sigma}^{**}$	10.13 (2.98)	28.10 (8.86)	42.18 (12.87)	13.29 (3.81)	35.16 (9.40)	52.20 (13.81)

** Robust H-CLUB.

much tighter than that given by the crude bound. In most cases the value of RE_1 reflects a difference of scale of an order of

magnitude between the intervals given by the Robust H-CLUB estimate and by ξ_T .

Table 3

Averages and standard deviations (in parentheses) of $RE_2 = \sqrt{\hat{\sigma}^2/T}/2\mathbf{w}^T \Sigma \mathbf{w}$ over 200 samples.

	$d = 200$			$d = 500$		
	$c = 1.0$	$c = 1.6$	$c = 2.0$	$c = 1.0$	$c = 1.6$	$c = 2.0$
RE_2	55.34%	50.28%	52.17%	49.72%	47.84%	56.71%
$\widehat{\Sigma}^{s**}$	(71.99%)	(66.54%)	(60.54%)	(55.49%)	(65.51%)	(68.31%)
RE_2	56.89%	50.89%	53.43%	47.89%	49.22%	58.27%
$\widehat{\Sigma}_{T_s=T}^s$	(78.69%)	(65.95%)	(64.50%)	(53.49%)	(69.21%)	(72.86%)
RE_2	41.21%	48.26%	52.05%	52.14%	44.32%	42.28%
$\widehat{\Sigma}^{h**}$	(47.40%)	(57.16%)	(55.07%)	(59.70%)	(50.83%)	(42.38%)
RE_2	2.07%	2.03%	2.00%	2.04%	2.00%	2.02%
$\widehat{\Sigma}^{**}$	(0.33%)	(0.31%)	(0.33%)	(0.37%)	(0.32%)	(0.31%)

** Robust H-CLUB.

Table 4

Coverage proportion and means of RE_1, RE_2 and $\widehat{U}(0.05)/\Delta$ for 200 samples when returns are drawn from Gaussian distributions without noise contamination, using $d = 500$.

n	Coverage			RE_1			RE_2			$\widehat{U}(0.05)/\Delta$		
	c	1.0	1.6	2.0	1.0	1.6	2.0	1.0	1.6	2.0	1.0	1.6
S^*	95.83%	94.59%	92.32%	8.22	21.60	33.39	3.98%	3.97%	3.95%	5.20	6.32	4.87
$\widehat{\Sigma}^{s**}$	95.28%	93.63%	95.07%	9.12	23.94	36.10	7.00%	6.81%	7.08%	6.84	5.70	6.42
$\widehat{\Sigma}_{T_s=T}^s$	96.90%	94.09%	94.78%	8.95	23.96	36.51	7.09%	6.82%	7.13%	6.50	6.15	6.37
$\widehat{\Sigma}^{h**}$	93.79%	97.24%	94.41%	9.62	24.97	38.13	4.79%	4.77%	4.81%	5.57	5.50	6.06
$\widehat{\Sigma}^{**}$	95.53%	93.01%	91.84%	12.43	33.46	50.01	1.96%	1.90%	1.91%	5.02	5.80	4.88

* H-CLUB.

** Robust H-CLUB.

- The ratio RE_1 increases with our ability to accurately estimate the marginal standard deviations. Note that $RE_1(\widehat{\Sigma}) > RE_1(\widehat{\Sigma}^h) \approx RE_1(\widehat{\Sigma}_{T_s=T}^s) \approx RE_1(\widehat{\Sigma}^s)$, which suggests that strong knowledge of the marginal standard deviations substantially improves performance, while having additional historical data or using $T_s = T$ has little if any effect.
- The value of RE_1 increases strongly with c and weakly with d . This suggests that the accuracy benefits of using the Robust H-CLUB estimates over the crude bound are particularly substantial for larger portfolios and those with higher gross exposure.

Table 3 summarizes the relative error (RE_2), which shows how informative our confidence intervals for the true portfolio risks are. Similarly to Table 2, we show the mean and standard deviation of RE_2 calculated over 200 simulations with 500 randomly generated portfolios per simulation (i.e., 100,000 portfolios total).

Here we see a similar pattern to before. Values are substantially better when the marginal standard deviations are known, but between the other estimation procedures there is little difference. This statement comes from the observation that $RE_2(\widehat{\Sigma}) \ll RE_2(\widehat{\Sigma}^h) \approx RE_2(\widehat{\Sigma}_{T_s=T}^s) \approx RE_2(\widehat{\Sigma}^s)$. The phenomenon is as expected since more information must be known for calculating $\widehat{\Sigma}$. We also observe that here the value of RE_2 does not appear to vary much with either c or d .

For our last set of results on synthetic data, we show in Table 4 that the Robust H-CLUB estimators are still competitive with the H-CLUB estimator when the data are drawn from a Gaussian distribution without noise contamination. In this table we present coverage proportions, means of RE_1 and RE_2 , as well as the mean of the ratio between the 95% bound and the value it is upper bounding, with this ratio given by $\widehat{U}(0.05)/\Delta$. These are calculated over 200 randomly generated models.

5. An empirical study

In this section we examine the behavior of the Robust H-CLUB estimators when applied to real-world data. We use the daily

excess returns of 100 industrial portfolios formed on size and book-to-market ratio, as available on the website of Kenneth French. We use the subset of data spanning from July 1, 2008 to June 29, 2012. For each 21 day period (nominal month), we use the preceding 63 days' data to estimate the covariance matrix via the Robust H-CLUB estimator with data-splitting ($\widehat{\Sigma}^s$), the Robust H-CLUB estimator with no data splitting ($\widehat{\Sigma}_{T_s=T}^s$), and the Robust H-CLUB estimator with known history ($\widehat{\Sigma}^h$). For the matrix of additional observations used in the latter estimator, we use the preceding 4 months (84 days) of returns data. Note that all other parameters selected here are as in the previous section. Finally, we also estimate the covariance via the sample covariance based H-CLUB estimator (S) for comparison.

We track the performance of the estimators on three portfolios: one portfolio with equal weighting ($\widehat{\mathbf{w}} = (1/100, \dots, 1/100)$), and two portfolios of minimum variance with gross exposure $c = 1$ and $c = 1.6$, as given by

$$\widehat{\mathbf{w}} = \arg \min_{\mathbf{w}^T \mathbf{1} = 1, \|\mathbf{w}\|_1 = c} \mathbf{w}^T \widehat{\Sigma}_{est} \mathbf{w}.$$

Note that on occasion the estimated covariance matrix is not positive definite, leading to problems in solving for the portfolio of minimum variance. In these cases, we coerce the estimated covariance matrix to be positive definite using Higham's algorithm (Higham, 2002) before calculating the minimum variance portfolio.

The portfolios of minimum variance are calculated at the start of each nominal month. The actual risk during the holding month for each $\widehat{\mathbf{w}}$ as defined above is then

$$R(\widehat{\mathbf{w}}) = (\widehat{\mathbf{w}}^T \Sigma \widehat{\mathbf{w}})^{1/2} \quad \text{and} \quad \Sigma = \frac{1}{21} \sum_{t=1}^{21} \mathbf{y}_t \mathbf{y}_t^T,$$

where $\{\mathbf{y}_t\}_{t=1}^T$ are the centralized daily returns over the holding month. This is calculated for each month in the four year period of study.

For each estimator and portfolio strategy, we consider five quantities. These quantities are summarized via their means

Table 5
Annualized true and estimated risk errors calculated on the 100 Fama–French portfolios.

Strategy	Average of $\Delta (\times 10^{-4})$	Average of $\hat{U}(0.05) (\times 10^{-4})$	Average of true risk	True risk error	Estimated risk error upper bound
S(H-CLUB)					
Equal weighted	2.28	1.58	27.36%	8.03%	5.55%
Min. variance ($c = 1$)	0.96	0.71	18.17%	5.19%	3.87%
Min. variance ($c = 1.6$)	0.45	0.34	13.26%	3.46%	2.60%
$\hat{\Sigma}^s$(Robust H-CLUB)					
Equal weighted	2.21	2.48	27.36%	8.04%	9.05%
Min. variance ($c = 1$)	0.84	1.31	18.38%	4.81%	7.73%
Min. variance ($c = 1.6$)	0.49	0.63	13.73%	4.12%	5.94%
$\hat{\Sigma}_{T=1}^s$(Robust H-CLUB – no data-splitting)					
Equal weighted	2.21	2.48	27.36%	8.03%	9.15%
Min. variance ($c = 1$)	0.84	1.31	18.37%	4.79%	7.70%
Min. variance ($c = 1.6$)	0.49	0.67	13.68%	4.08%	6.24%
$\hat{\Sigma}^h$(Robust H-CLUB – known history)					
Equal weighted	2.43	2.49	27.36%	8.76%	8.95%
Min. variance ($c = 1$)	0.97	1.04	18.57%	5.22%	6.07%
Min. variance ($c = 1.6$)	0.55	0.54	14.01%	4.16%	4.22%

Note: $\Delta = |\mathbf{w}^T(\hat{\Sigma}_{est} - \Sigma)\mathbf{w}|$, $\hat{U}(0.05) = 2 \times (\text{Var}(\mathbf{w}^T \hat{\Sigma}_{est} \mathbf{w}))^{1/2}$. True Risk is $\sqrt{252} \times R(\mathbf{w})$. True Risk Error is $\sqrt{252} \times |(\mathbf{w}^T \hat{\Sigma}_{est} \mathbf{w})^{1/2} - (\mathbf{w}^T \Sigma \mathbf{w})^{1/2}|$, and Estimated risk error upper bound is $\sqrt{252} \times \hat{U}(0.05) / \sqrt{4\mathbf{w}^T \hat{\Sigma}_{est} \mathbf{w}}$. The factor of $\sqrt{252}$ is present to convert the risks to annualized values.

(calculated over the whole study period) in Table 5. We compare the first two columns of Table 5 and provide several observations.

- The values of Δ are comparable among the four estimators considered. This suggests that all estimators are similar in their estimations of the covariance matrix Σ , and that differences between them lie in their ability to accurately conduct inference on $\mathbf{w}^T \hat{\Sigma}_{est} \mathbf{w}$ (i.e., construct a valid confidence interval for the risk).
- The H-CLUB estimator does not give a valid upper bound, as $\hat{U}(0.05)$ is less than Δ throughout.
- For the Robust H-CLUB estimators, $\hat{U}(0.05)$ is greater than Δ in all but one of the cases. This is broadly consistent with the expectation that the value of $\hat{U}(0.05)$ for the robust estimators is a 95% upper bound of the estimation error for the portfolio variance. We note that for the single discrepancy ($\hat{\Sigma}^h$, on the minimum variance portfolio with $\|\mathbf{w}\|_1 = 1.6$), the value of $\hat{U}(0.05)$ still only falls below Δ by a small margin.

Lastly, the estimated risk error upper bound should provide a high-confidence level upper bound for the true risk error $|(\mathbf{w}^T \Sigma \mathbf{w})^{1/2} - (\mathbf{w}^T \hat{\Sigma}_{est} \mathbf{w})^{1/2}|$ (we can see this simply by applying the delta method to the results of, e.g., Theorem 3.6). The last two columns of Table 5 show that the Robust H-CLUB estimators hold true to this, with the estimated risk error upper bound uniformly bounding the true risk error in all cases. However, the H-CLUB estimator does not yield a good upper bound, with the estimated risk error upper bound uniformly falling below the true risk error. This is again evidence for the strength of the proposed robust estimators in the presence of heavy-tailed or noisy data.

Acknowledgments

We thank Huitong Qiu for discussions. Jianqing Fan’s research was supported by the National Science Foundation DMS-1406266 and DMS-1206464. Fang Han’s research was supported by National Institute of Biomedical Imaging and Bioengineering (NIBIB) R01-EB012547. Han Liu’s research was supported by the NSF CAREER Award DMS-1454377, NSF IIS-1546482, NSF IIS-1408910, NSF IIS-1332109, NIH R01-MH102339, NIH R01-GM083084, and NIH R01-HG06841. Byron Vickers was supported by NIH R02-GM072611-10.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.jeconom.2016.05.008>.

References

Agarwal, A., Negahban, S., Wainwright, M.J., 2012. Noisy matrix decomposition via convex relaxation: optimal rates in high dimensions. *Ann. Statist.* 40 (2), 1171–1197.

Bai, J., Liao, Y., 2015. Efficient estimation of approximate factor models via regularized maximum likelihood. *J. Econometrics* 191, 1–18.

Barndorff-Nielsen, O.E., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64 (2), 253–280.

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

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

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31 (3), 307–327.

Bühlmann, P., 2002. Bootstraps for time series. *Statist. Sci.* 17 (1), 52–72.

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

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

Catoni, O., 2012. Challenging the empirical mean and empirical variance: a deviation study. *Ann. Inst. Henri Poincaré, Probab. Stat.* 48 (4), 1148–1185.

Chamberlain, G., 1983. A characterization of the distributions that imply mean-variance utility functions. *J. Econom. Theory* 29 (1), 185–201.

Chang, C., Tsay, R.S., 2010. Estimation of covariance matrix via the sparse Cholesky factor with lasso. *J. Statist. Plann. Inference* 140 (12), 3858–3873.

Chen, X., Xu, M., Wu, W., 2013. Covariance and precision matrix estimation for high-dimensional time series. *Ann. Statist.* 41 (6), 2994–3021.

Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quant. Financ.* 1 (2), 223–236.

Fan, J., Fan, Y., Lv, J., 2008. High dimensional covariance matrix estimation using a factor model. *J. Econometrics* 147 (1), 186–197.

Fan, J., Li, Q., Wang, Y., 2016. Estimation of high-dimensional mean regression in absence of symmetry and light-tail assumptions. *J. Royal Statistical Society: Series B* in press.

Fan, J., Liao, Y., Mincheva, M., 2011. High dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.* 39 (6), 3320–3356.

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

Fan, J., Liao, Y., Shi, X., 2015. Risks of large portfolios. *J. Econometrics* 186 (2), 367–387.

Fan, J., Lv, J., Qi, L., 2011b. Sparse high dimensional models in economics. *Ann. Rev. Econ.* 3, 291–317.

Fan, J., Peng, H., 2004. Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* 32 (3), 928–961.

- Fan, J., Yao, Q., 2003. *Nonlinear Time series: Nonparametric and Parametric Methods*. Springer.
- Fan, J., Zhang, J., Yu, K., 2012. Vast portfolio selection with gross-exposure constraints. *J. Amer. Statist. Assoc.* 107 (498), 592–606.
- Frahm, G., Jaekel, U., 2007. Tyler's M-estimator, Random Matrix Theory, and Generalized Elliptical Distributions with Applications to Finance. Technical Report, Helmut Schmidt University.
- Fryzlewicz, P., 2013. High-dimensional volatility matrix estimation via wavelets and thresholding. *Biometrika* 100 (4), 921–938.
- Gómez, K., Gallón, S., 2011. Comparison among high dimensional covariance matrix estimation methods. *Revista Colombiana Estadíst.* 34 (3), 567–588.
- Greenshtein, E., Ritov, Y., 2004. Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli* 10 (6), 971–988.
- Hamada, M., Valdez, E., 2004. CAPM and Option Pricing with Elliptical Distributions. School of Finance and Economics, University of Technology, Sydney.
- Han, F., Liu, H., 2013. Principal component analysis on non-gaussian dependent data. In: *Proceedings of the 30th International Conference on Machine Learning*. pp. 240–248.
- Han, F., Liu, H., 2014. Scale-invariant sparse PCA on high-dimensional meta-elliptical data. *J. Amer. Statist. Assoc.* 109 (505), 275–287.
- Han, F., Liu, H., 2016. Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. *Bernoulli* (in press).
- Han, F., Lu, J., Liu, H., 2014. Robust Scatter Matrix Estimation for High Dimensional Distributions with Heavy Tails. Technical Report, Princeton University.
- Han, F., Lu, H., Liu, H., 2015. A direct estimation of high dimensional stationary vector autoregressions. *J. Mach. Learn. Res.* 16, 3115–3150.
- Higham, N.J., 2002. Computing the nearest correlation matrix—a problem from finance. *IMA J. Numer. Anal.* 22 (3), 329–343.
- Hsu, D., Kakade, S.M., Zhang, T., 2011. Robust matrix decomposition with sparse corruptions. *IEEE Trans. Inform. Theory* 57 (11), 7221–7234.
- Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: why imposing the wrong constraints helps. *J. Finance* 58 (4), 1651–1684.
- Lai, T.L., Xing, H., Chen, Z., 2011. Mean-variance portfolio optimization when means and covariances are unknown. *Ann. Appl. Stat.* 5 (2A), 798–823.
- Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empir. Financ.* 10 (5), 603–621.
- Lindskog, F., McNeil, A., Schmock, U., 2003. Kendall's tau for elliptical distributions. *Credit Risk: Measurement, Evaluation and Management* 149–156.
- Loh, P.-L., Wainwright, M.J., 2012. High-dimensional regression with noisy and missing data: provable guarantees with nonconvexity. *Ann. Statist.* 40 (3), 1637–1664.
- Longla, M., Peligrad, M., 2012. Some aspects of modeling dependence in copula-based markov chains. *J. Multivariate Anal.* 111, 234–240.
- Mitra, R., Zhang, C.-H., 2014. Multivariate analysis of nonparametric estimates of large correlation matrices. ArXiv preprint arXiv:1403.6195.
- Owen, J., Rabinovitch, R., 1983. On the class of elliptical distributions and their applications to the theory of portfolio choice. *J. Finance* 38 (3), 745–752.
- Pan, J., Yao, Q., 2008. Modelling multiple time series via common factors. *Biometrika* 95 (2), 365–379.
- Pesaran, M.H., Zaffaroni, P., 2008. Optimal Asset Allocation with Factor Models for Large Portfolios. Technical Report, CESifo Working Paper.
- Politis, D.N., Romano, J.P., 1992. A circular block-resampling procedure for stationary data. In: *Exploring the Limits of Bootstrap*. John Wiley, New York, pp. 263–270.
- Qiu, H., Han, F., Liu, H., Caffo, B., 2015a. Robust portfolio optimization. In: *Proceedings of the 28th Annual Conference on Neural Information Processing Systems*.
- Qiu, H., Xu, S., Han, F., Liu, H., Caffo, B., 2015b. Robust estimation of transition matrices in high dimensional heavy-tailed vector autoregressive processes. In: *Proceedings of The 32nd International Conference on Machine Learning*. pp. 1843–1851.
- Shao, Q.-M., Yu, H., 1993. Bootstrapping the sample means for stationary mixing sequences. *Stochastic Process. Appl.* 48 (1), 175–190.
- Wegkamp, M., Zhao, Y., 2016. Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas. *Bernoulli* 22 (2), 1184–1226.
- Xiao, H., Wu, W.B., 2012. Covariance matrix estimation for stationary time series. *Ann. Statist.* 40 (1), 466–493.
- Yoshihara, K.-I., 1976. Limiting behavior of U-statistics for stationary, absolutely regular processes. *Probab. Theory Related Fields* 35 (3), 237–252.
- Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales. *J. Amer. Statist. Assoc.* 100 (472), 1394–1411.