ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

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 xxxx

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.

1. Introduction

Let $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T$ be a stationary multivariate time series with $\boldsymbol{R}_t \in \mathbb{R}^d$ 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

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

where Σ denotes the unknown volatility (or covariance) matrix of $\boldsymbol{R}_t$. 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 $\mathbf{w}^{\mathsf{T}}\Sigma\mathbf{w}$ 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 $\mathbf{w}^{\mathsf{T}}\Sigma\mathbf{w}$, Fan et al. (2012) propose to use $\|\mathbf{w}\|_1^2\|\widehat{\Sigma}_{\text{est}} - \Sigma\|_{\max}$[1] as an upper bound of $|\mathbf{w}^{\mathsf{T}}(\widehat{\Sigma}_{\text{est}} - \Sigma)\mathbf{w}|$. 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 $|\mathbf{w}^{\mathsf{T}}(\widehat{\Sigma}_{\text{est}} - \Sigma)\mathbf{w}|$: 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 $|\mathbf{w}^{\mathsf{T}}(\widehat{\Sigma}_{\text{est}} - \Sigma)\mathbf{w}|$ 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 $\mathbf{w}^{\mathsf{T}}\Sigma\mathbf{w}$ when the asset returns $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T$ 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 $\ell_{\max}$ norm $\|\cdot\|_{\max}$ later.

http://dx.doi.org/10.1016/j.jeconom.2016.05.008
0304-4076/

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

Theoretically, when $T, d \to \infty$ and d is possibly much larger than T, we develop an inferential theory of the robust risk estimators. In particular, we show that $\sqrt{T}\mathbf{w}^\mathsf{T}(\widehat{\Sigma}_{\rm est} - \Sigma)\mathbf{w}$ is asymptotically normal with variance σ^2, and the block-bootstrap-based estimator $\widehat{\sigma}^2_{\rm est}$ 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 $\mathbf{v} = (v_1, \ldots, v_d)^\mathsf{T}$ 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 $I, J \in \{1, \ldots, d\}$, 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 $\ell_{\max}$ 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 $f : \mathbb{R} \to \mathbb{R}$ 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 $\mathbf{M} = \mathrm{diag}(\mathbf{M}_1, \ldots, \mathbf{M}_k)$ if $\mathbf{M}$ is block diagonal with diagonal matrices $\mathbf{M}_1, \ldots, \mathbf{M}_k$. For random vectors $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^d$, we write $\boldsymbol{X} \stackrel{\rm d}{=} \boldsymbol{Y}$ if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are identically distributed. Throughout the paper, we use $c, c_1, c_2, \ldots$, and $C, C_1, C_2, \ldots$ 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 $a_n \geq cb_n$ 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 $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T$ with $\boldsymbol{R}_t = (R_{t1}, \ldots, R_{td})^\mathsf{T} \in \mathbb{R}^d$ for $t = 1, \ldots, T$. Let $\Sigma := \mathrm{Cov}(\boldsymbol{R}_t)$ be the covariance matrix and $\mathbf{D} \in \mathbb{R}^{d\times d}$ be a diagonal matrix with diagonals $\Sigma_{11}^{1/2}, \ldots, \Sigma_{dd}^{1/2}$. It is easy to derive $\Sigma = \mathbf{D}\Sigma^0\mathbf{D}$, where Σ^0 is the correlation matrix of $\boldsymbol{R}_t$. For a given portfolio allocation vector $\mathbf{w} \in \mathbb{R}^d$, we aim to construct a confidence interval for $\mathbf{w}^\mathsf{T}\Sigma\mathbf{w}$. 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 $\boldsymbol{Z} \in \mathbb{R}^d$ 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

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

where $\mathbf{A} \in \mathbb{R}^{d\times d}$ satisfies $\mathbf{A}\mathbf{A}^\mathsf{T} = \Sigma$, $\boldsymbol{U} \in \mathbb{R}^d$ is uniformly distributed on the d-dimensional sphere $\mathbb{S}^{d-1}$, and ξ is an unspecified nonnegative random variable independent of $\boldsymbol{U}$ 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 $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T$, the sample and population Kendall's tau matrices $\widehat{\mathbf{T}} = [\widehat{\tau}_{jk}]$ and $\mathbf{T} = [\tau_{jk}]$ are defined as

$$\widehat{\tau}_{jk} := \frac{2}{T(T-1)} \sum_{t<t'} \mathrm{sign}(R_{tj} - R_{t'j})\mathrm{sign}(R_{tk} - R_{t'k}),$$

$$\tau_{jk} := \mathbb{E}\mathrm{sign}(R_j - \widetilde{R}_j)\mathrm{sign}(R_k - \widetilde{R}_k), \tag{2.1}$$