Asymptotic Analysis of Portfolio Diversification

Hengxin Cui[*] Ken Seng Tan[†] Fan Yang[‡] Chen Zhou[§]

April 20, 2022

Abstract

In this paper, we investigate the optimal portfolio construction aiming at extracting the most diversification benefit. We employ the diversification ratio based on the Value-at-Risk as the measure of the diversification benefit. With modeling the dependence of risk factors by the multivariate regularly variation model, the most diversified portfolio is obtained by optimizing the asymptotic diversification ratio. Theoretically, we show that the asymptotic solution is a good approximation to the finite-level solution. Our theoretical results are supported by extensive numerical examples. By applying our portfolio optimization strategy to real market data, we show that our strategy provides a fast algorithm for handling a large portfolio, while outperforming other peer strategies in out-of-sample risk analyses.

Keywords: Portfolio optimization, Diversification, Risk management, Multivariate regularly variation

1 Introduction

In order to mitigating risks in portfolios of financial investment, a common tool is the diversification strategy. Large insurance claims and asset returns have been empirically shown to be heavy tailed, i.e. the tail exhibits power-law decay; see, e.g. Loretan and Phillips [31], Gabaix et al. [23], Ibragimov et al. [28] and Hofert and Wüthrich [26]. Moreover, dependence is known to widely exist in financial assets and insurance losses; see for example Embrechts et al. [19] and Acharya et al. [1]. The benefit from a diversification strategy can be reflected in the reduction of dependent tail risks in a diversified portfolio. Value-at-Risk

[*]Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada.

[†]Division of Banking & Finance, Nanyang Business School, Nanyang Technological University, Singapore.

[‡]Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada.

[§]Economics and Research Division, De Nederlandsche Bank, Amsterdam, Netherlands; Erasmus University Rotterdam, Rotterdam, Netherlands.

(VaR) has been widely applied to capture market risk as it quantifies the quantile of a loss for a given time horizon. In this paper, we investigate the optimal portfolio construction aiming at extracting the most diversification benefit for dependent extreme risks based on the VaR measure.

A key difficulty in evaluating the diversification benefit based on the VaR measure is that there is often no explicit formula for calculating the portfolio VaR. Since a portfolio is a linear combination of the underlying risky assets, only if the asset returns follow sum-stable distributions such as the Gaussian distribution or the stable distributions, one can precisely calculate the distribution of the portfolio return, and derive the VaR therefrom. As an alternative, Extreme Value Theory (EVT), in particular, the multivariate regular variation (MRV) model, may provide an explicit approximation to the tail of the distribution of portfolio return; see e.g. Mainik and Rüchendorf [32], Mainik and Embrechts [33] and Zhou [49]. By inverting the approximation formula on the tail of the distribution, one may get an approximation for the VaR measure, when the probability level in VaR is considered to be close to 1. Therefore, the EVT approach opens a new door for investigating the diversification benefit based on the VaR measure.

Nevertheless, the approximation holds only in the limit when the probability level in VaR tending to 1. The approximation nature leaves two difficulties to be handled. Firstly, for heavy-tailed portfolio returns as assumed in the setup of the MRV, when the probability level in VaR tends to 1, the VaR converges to infinity. Consequently, the goal of VaR optimization turns to be minimizing "the VaR in the limit", even if the limit is infinity. It is difficult to provide an economic interpretation for such a mathematical exercise. Secondly, the practical goal for risk managers is to minimize VaR at a given probability level, such as 99% (Basel II) or 99.5% (Solvency II), while "the VaR in the limit" is not of their concern. And, it is not guaranteed that the optimal portfolio based on minimizing " the VaR in the limit" is also close to the practical goal.

The first difficulty can be overcome by proper normalization. For example, one may compare the portfolio VaR to the VaRs of marginal risks. For that purpose, we employ the measure *diversification ratio* (DR), or sometimes with its alternative name: the risk concentration based on VaR; see, for example Degen et al. [11] and Embrechts et al. [18]. The diversification ratio is defined as follows. Let $\boldsymbol{X} := (X_1, \dots X_d)^T$ be a nonnegative random vector indicating the losses of d assets. The value of a portfolio is given by $\boldsymbol{w}^T\boldsymbol{X}$, where the weights satisfy $\boldsymbol{w} = (w_1, w_2, \dots, w_d)^T \in \Sigma^d := \left\{\boldsymbol{x} \in [0,1]^d : x_1 + x_2 + \dots + x_d = 1\right\}$. For this portfolio, the diversification ratio (DR) based on VaR at level $q \in (0,1)$ is defined as

$$\mathrm{DR}_{\boldsymbol{w},q} = \frac{\mathrm{VaR}_q(\boldsymbol{w}^T\boldsymbol{X})}{\sum_{i=1}^d w_i \mathrm{VaR}_q(X_i)}. \tag{1.1}$$

The DR is a measure of diversification benefit in the following sense. Consider the comonotonic case where all assets are completely dependent. Then DR is a constant one regardless how the portfolio is allocated. This is a special case in which any diversification strategy would not reduce the portfolio risk. Consequently, in a general case, $1 - \mathrm{DR}_{\boldsymbol{w},q}$ can be regarded as the diversification benefit. In Cui et al. [7], DR is applied to measure the effect of diversification in catastrophe insurance markets.

The first result in this paper is to show that the DR converges to a finite value for any portfolio as $q \to 1$ under the MRV model. More specifically, by modeling the joint distribution of the random vector $\boldsymbol{X}$ by MRV, we can derive an explicit formula for

$$\mathrm{DR}_{\boldsymbol{w},1} := \lim_{q \uparrow 1} \mathrm{DR}_{\boldsymbol{w},q}$$

with respect to the weight $\boldsymbol{w}$ and the two key elements characterizing the MRV model: the tail index of the marginals and the spectral measure for the tail dependence structure.[1]

This result overcomes the first difficulty regarding the interpretation: one may target minimizing the DR in the limit, which is at a finite level. We show that there exists a unique solution to the optimization problem

$$\boldsymbol{w}^* := \arg\min_{\boldsymbol{w} \in \Sigma^d} \mathrm{DR}_{\boldsymbol{w},1}.$$

A portfolio that minimizes the DR is consequently extracting the most diversification benefit based on the VaR measure. It is also worth noticing that by taking the marginal VaRs in the denominator, the optimal portfolio based on the DR is mainly driven by the dependence structure across the risky assets, while is more robust to changes in marginal risks.

However, the second difficulty raised above remains valid after switching to minimizing $\mathrm{DR}_{\boldsymbol{w},1}$. Is the optimal solution based on minimizing $\mathrm{DR}_{\boldsymbol{w},1}$ close to the practical goal of minimizing $\mathrm{DR}_{\boldsymbol{w},q}$ at a given probability level q? We formalize this question by the following notation.

Practically, with introducing the DR, risk managers aim at solving the following optimization problem:

$$\arg\min_{\boldsymbol{w} \in \Sigma^d} \mathrm{DR}_{\boldsymbol{w},q}. \tag{1.2}$$

Denote the solution to (1.2) by $\boldsymbol{w}_q$. Solving (1.2) directly is computationally intensive. With observations on the joint distribution of the random vector $\boldsymbol{X}$, $\boldsymbol{w}_q$ can be estimated by conducting a numerical grid search. However, such a searching algorithm suffers from the dimensionality curse: the computational burden increases exponentially with respect to the dimension d.

The second main result of this paper is to show how close the solution $\boldsymbol{w}^*$ is from the solution of the original optimization problem $\boldsymbol{w}_q$. First, we show theoretically that

$$\lim_{q \uparrow 1} \boldsymbol{w}_q = \boldsymbol{w}^*. \tag{1.3}$$

The convergence in (1.3) ensures that one may use the solution to the optimization problem in the limit as an approximation to the solution to the original problem with a finite level q

[1] As pointed out by Mainik and Embrechts [33], under the MRV structure, when the tail index is great than 1, $\mathrm{DR}_{\boldsymbol{w},1} < 1$. In other words, the VaR measure possesses subadditivity as $q \to 1$. Hence, diversification is always optimal in this situation and the optimization problem (1.2) is well defined.

close to 1. Further, define the distance between $\boldsymbol{w}_q$ and $\boldsymbol{w}^*$, measured by $\|\boldsymbol{w}_q - \boldsymbol{w}^*\|$ with respect to an arbitrary norm as D_q. In other words, given a finite level of q close to 1, the solution $\boldsymbol{w}_q$ is within an area defined as a D_q radius circle around $\boldsymbol{w}^*$. For a special case of MRV, the Farlie-Gumbel-Morgenstern (FGM) copula, we explicitly determine D_q.

Empirically, with observations on the joint distribution of the random vector $\boldsymbol{X}$, one can estimate the two main components for the MRV: the marginal tail index and the spectral measure. By plugging in the estimates of these two elements, the solution $\boldsymbol{w}^*$ can be estimated using conventional convex optimization method. We show the consistency of the estimator. Notice that the computational burden is much lower than the aforementioned numerical approach for solving $\boldsymbol{w}_q$.

We use a few numerical examples to support our theoretical results and also apply our method to empirical data. We find that portfolio constructed using our approach possess the lowest DR and also suffers low losses in out-of-sample periods, compared to other portfolio optimization strategies.

Our proposed portfolio optimization strategy is comparable to other strategies based on tail risk. Mainik and Rüchendorf [32], proposed to minimize the so-called extreme risk index (ERI),

$$\mathrm{ERI} = \arg\min_{\boldsymbol{w}} \lim_{q\uparrow 1} \frac{\mathrm{VaR}_q(\boldsymbol{w}^T\boldsymbol{X})}{\mathrm{VaR}_q(||\boldsymbol{X}||_1)},$$

which essentially is minimizing the portfolio VaR. This strategy is more sensitive to marginal tail risks and consequently load high on marginals with a low VaR. On the contrary, minimizing DR in (1.1) scales off the effect of marginals and focuses more on the dependence structure.

Another closely related strategy is the so called most diversified portfolio (MDP)

$$\mathrm{MDP} = \arg\min_{\boldsymbol{w}} \frac{\mathrm{var}(\boldsymbol{w}^T\boldsymbol{X})}{\sum_{i=1}^d w_i \mathrm{var}(X_i)},$$

proposed by Choueifaty and Coignard [6]. The MDP method shares the same structure with our approach: it considers the ratio between portfolio risk and the sum of individual risks measured by variances. Since variance is a measure of overall risk rather than focusing on the tail region, the MDP method may fail to capture the extreme risks.

One possible drawback of our portfolio optimization strategy (1.2) is that it only minimizes the risk without taking into account the upper side potential: portfolio returns. Given that the limit of DR is a convex function, it is in fact straightforward to consider the return components simultaneously. For example, consider the "safety-first" criterion proposed by Roy [44], which aims at first constraining the downside risk to a given level and then maximizing the profit. This is equivalent to minimizing risk with a linear constraint on the returns. Comparing this optimization problem with the aforementioned unconstrained convex minimization problem, taking the return into consideration is just to impose an additional linear constraint. It is straightforward to verify that our current results remain valid for the constrained optimization problem. To avoid complicating the discussion, in this paper we opt to focusing on the optimization of DR without considering the return side.

The paper is organized as follows. In Section 2, we provide our main results on the convergence of optimal portfolios. Section 3 discusses the convergence rate of the optimal portfolio. In Section 4, we demonstrate the empirical performance of our strategy based on three numerical examples. Section 5 provides the application of our strategy to real market data. Section 6 concludes the paper and some of the proofs are relegated to Appendix A.

2 Convergence of optimal portfolios

2.1 Preliminaries

2.1.1 The multivariate regular variation model

A nonnegative random vector $\boldsymbol{X}$ is said to be *multivariate regularly varying* (MRV), if there exists a sequence $b_t \to \infty$ and a Radon measure ν on $\mathcal{B}\left([0,\infty]^d \setminus \{\mathbf{0}\}\right)$ such that $\nu\left([0,\infty]^d \setminus \mathbb{R}_+^d\right) = 0$, and

$$\nu_t = t \Pr\left(\frac{\boldsymbol{X}}{b_t} \in \cdot\right) \xrightarrow{\mathrm{v}} \nu(\cdot), \qquad t \to \infty, \tag{2.1}$$

where $\xrightarrow{\mathrm{v}}$ refers to the vague convergence. We additionally assume that the limit measure ν is nondegenerate in the sense that

$$\nu\left(\left\{\boldsymbol{x} \in \mathbb{R}_+^d : x_i > 1\right\}\right) > 0,$$

for all $i = 1, 2, \ldots, d$. For a full account of technical details related to the notion of MRV, the reader is referred to Resnick [43] and Kulik and Soulier [29].

For any arbitrary norm $\|\cdot\|$, let $\mathcal{S}_+^{d-1} = \left\{\boldsymbol{s} \in \mathbb{R}_+^d : \|\boldsymbol{s}\| = 1\right\}$ be the unit sphere. Under the polar transformation, it is equivalently to say that $\boldsymbol{X}$ is MRV if there exists a sequence $b_t \to \infty$, a positive constant c and a probability measure Ψ on $\mathcal{S}_+^{d-1}$ such that for all $x > 0$,

$$\nu_t = t \Pr\left(\|\boldsymbol{X}\| > b_t x, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in \cdot\right) \xrightarrow{\mathrm{v}} c \cdot \rho_\alpha((x,\infty]) \times \Psi, \qquad t \to \infty,$$

where the measure $\rho_\alpha((x,\infty]) = x^{-\alpha}$ and the vague convergence holds on $(0,\infty) \times \mathcal{S}_+^{d-1}$. The measure Ψ is often called the spectral or angular measure. Throughout the paper, we denote that $\boldsymbol{X}$ is MRV with tail index α and spectral measure Ψ by $\boldsymbol{X} \in \mathrm{MRV}_\alpha(\Psi)$. See Section 6.5 of Resnick [43], Section 2.2 of Kulik and Soulier [29] and Soulier [46] for more details on the spectral decomposition of MRV.

Theoretically, it does not matter which norm is chosen for the polar representation. For simplicity, in this paper we consider the ℓ_1-norm $\|\cdot\|_1$. Then $\mathcal{S}_+^{d-1} = \Sigma^d$. Further, by constraining the measures ν_t and ν to the set $A_1 := \left\{\boldsymbol{x} \in \mathbb{R}_+^d : \|\boldsymbol{x}\|_1 > 1\right\}$, the constant c

is normalized to 1. With a proper choice of $b_t = F_R^{\leftarrow}(1-1/t)$ and $R = ||\boldsymbol{X}||_1$, the vague convergence in (2.1) implies the weak convergence on $\mathcal{B}(A_1)$, as

$$\nu_t(\cdot) = \frac{\mathbb{P}(t^{-1}\boldsymbol{X} \in \cdot)}{\mathbb{P}(\|\boldsymbol{X}\|_1 > t)} \xrightarrow{\text{w}} \nu(\cdot)|_{A_1}, \qquad t \to \infty, \tag{2.2}$$

where $\nu|_{A_1}$ is the restriction of ν to the set A_1. Note that ν_t in (2.1) can also be rewritten as a conditional probability $\mathbb{P}(t^{-1}\boldsymbol{X} \in \cdot \mid \|\boldsymbol{X}\|_1 > t)$.

On the one hand, MRV is a semi-parametric model by only assuming a limit relation in the tail region, which allows for a flexible dependence structure across several heavy-tailed random variables. For example, the multivariate student's t-distributions, multivariate α-stable distributions, elliptical distributions with a regularly varying radial component, Archimedean copulas with regularly varying generator and marginals, among others. On the other hand, the nondegenerate MRV model requires all the marginal distributions are of the same level of heavy tailedness. This restriction challenges the application of the model in practice.

2.1.2 Convergence of minimizers

In this subsection, we give a general result on the convergence of minimizers. This is the foundation to prove the main result in this paper. Throughout the paper, for a function $g : S \to \mathbb{R}$, we denote $M(g)$ the set of all the minimizers of g. That is,

$$M(g) = \left\{ x \in S : g(x) = \inf_{y \in Z} g(y) \right\}.$$

A minimizer of g is denoted by $m_g \in M(g)$.

The following result may be known in the literature but we cannot find a proper reference for it. For completeness, we include the proof here.

Lemma 2.1 *Suppose that $\{f_n\}$ is a sequence of lower semi-continuous functions from a compact metric space S to $\overline{\mathbb{R}} = [-\infty, \infty]$, and f_n converges uniformly to a function f. If, in addition, assume that f has a unique minimum point in Z, then*

$$\lim_{n\to\infty} m_{f_n} = \arg\min f. \tag{2.3}$$

Proof. On the compact metric space S, we have that the sequence $\{f_n\}$ is equi-coercive and gamma-converges to f under the conditions of Lemma 2.1. The sequence $\{f_n\}$ is said to be equi-coercive if for any $a \in \mathbb{R}$, there exists a compact set K_a of S such that the subsets $\{f_n \le a\} \subseteq K_a$ for all n. The sequence $\{f_n\}$ is said to gamma-converge to f with respect to the topology of S if $f^+ = f^-$, where

$$f^+(x) = \sup_{U \in N(x)} \limsup_{n\to\infty} \inf_{y \in U} f_n(y)$$

and

$$f^{-}(x) = \sup_{U \in N(x)} \liminf_{n\to\infty} \inf_{y\in U} f_n(y)$$

with $N(x)$ being the set of all open neighborhoods of x in S. Then by Corollary 7.24 in Dal Maso [8], the relation (2.3) holds. ■

2.2 Main results

The first result regards the weak convergence of $\mathrm{DR}_{\boldsymbol{w},q}$ as $q \uparrow 1$, which is a direct consequence of known results in the literature.

Proposition 2.1 *Suppose the nonnegative random vector* $\boldsymbol{X} \in \mathrm{MRV}_\alpha(\Psi)$ *with* $\alpha > 0$. *Then for any* $\boldsymbol{w} \in \Sigma^d$, *we have*

$$\lim_{q\uparrow 1} \mathrm{DR}_{\boldsymbol{w},q} = \mathrm{DR}_{\boldsymbol{w},1},$$

where

$$\mathrm{DR}_{\boldsymbol{w},1} = \frac{\eta_{\boldsymbol{w}}^{1/\alpha}}{\sum_{i=1}^d w_i \eta_{\boldsymbol{e}_i}^{1/\alpha}}$$

with $\eta_{\boldsymbol{w}} = \int_{\Sigma^d} (\boldsymbol{w}^T\boldsymbol{s})^\alpha \Psi(d\boldsymbol{s})$ *and* $\boldsymbol{e}_i = (0,...,1,...,0)^T$ *only the ith component being* 1 *for* $i = 1,..,d$.

Proof. Note that

$$\mathrm{DR}_{\boldsymbol{w},q} = \frac{\mathrm{VaR}_q(\boldsymbol{w}^T\boldsymbol{X})/\mathrm{VaR}_q(\|\boldsymbol{X}\|_1)}{\sum_{i=1}^d w_i \mathrm{VaR}_q(X_i)/\mathrm{VaR}_q(\|\boldsymbol{X}\|_1)}. \tag{2.4}$$

For $\boldsymbol{X} \in \mathrm{MRV}_\alpha(\Psi)$ with $\alpha > 0$, it follows that

$$\lim_{q\uparrow 1} \frac{\mathrm{VaR}_q(\boldsymbol{u}^T\boldsymbol{X})}{\mathrm{VaR}_q(\|\boldsymbol{X}\|_1)} = \eta_{\boldsymbol{u}}^{1/\alpha}, \qquad \boldsymbol{u} \in \Sigma^d, \tag{2.5}$$

which can be found in e.g. Mainik and Rüchendorf [32], Mainik and Embrechts [33] and Zhou [49]. The proposition can be proved by letting $\boldsymbol{u} = \boldsymbol{w}$ and $\boldsymbol{u} = \boldsymbol{e}_i$ in (2.5). ■

In the following theorem, we develop the uniform convergence of $\mathrm{DR}_{\boldsymbol{w},q}$, which is essential for proving the convergence of minimizers. It is also an interesting result on its own. The proof is relegated to Appendix A.

Theorem 2.1 *Suppose the nonnegative random vector* $\boldsymbol{X}$ *has a positive joint density function. Further assume that* $\boldsymbol{X} \in \mathrm{MRV}_\alpha(\Psi)$ *with* $\alpha > 0$. *Then*

$$\lim_{q\uparrow 1} \sup_{\boldsymbol{w}\in\Sigma^d} |\mathrm{DR}_{\boldsymbol{w},q} - \mathrm{DR}_{\boldsymbol{w},1}| = 0. \tag{2.6}$$

The main result of this section, in the following theorem, shows that the convergence of a sequence of optimal solutions of $\mathrm{DR}_{\boldsymbol{w},q}$ to the unique minimizer of $\mathrm{DR}_{\boldsymbol{w},1}$.

Theorem 2.2 *Suppose the nonnegative random vector* $\boldsymbol{X}$ *has a positive joint density function. Further assume that* $\boldsymbol{X} \in \mathrm{MRV}_\alpha(\Psi)$ *with* $\alpha > 1$, *and* $\Psi\left(\left\{\boldsymbol{x} : \boldsymbol{a}^T\boldsymbol{x} = 0\right\}\right) = 0$ *for any* $\boldsymbol{a} \in \mathbb{R}^d$. *Then* $\boldsymbol{w}^* = \arg\min \mathrm{DR}_{\boldsymbol{w},1}$ *exists and is unique. Moreover,*

$$\lim_{q\uparrow 1} \boldsymbol{w}_q = \boldsymbol{w}^*, \tag{2.7}$$

where $\boldsymbol{w}_q$ *is a solution of* $\min_{\boldsymbol{w}\in\Sigma^d} \mathrm{DR}_{\boldsymbol{w},q}$.

Proof. The existence $\boldsymbol{w}^*$ is due to the continuity of $\mathrm{DR}_{\boldsymbol{w},1}$ and the compactness of Σ^d. To show the uniqueness, first note that the minimization problem $\min_{\boldsymbol{w}\in\Sigma^d} \mathrm{DR}_{\boldsymbol{w},1}$ is equivalent to

$$\begin{aligned} &\min_{\boldsymbol{w}} \eta_{\boldsymbol{w}}^{1/\alpha} \\ &\boldsymbol{s.t.} \quad \textstyle\sum_{i=1}^d w_i \eta_{\boldsymbol{e}_i}^{1/\alpha} = 1 \text{ with } w_i \geq 0 \text{ for } i = 1, 2, \ldots, d. \end{aligned} \tag{2.8}$$

Since the set of constraints in (2.8) is nonempty, closed and bounded, it is compact. By Theorem 2.4 of Mainik and Embrechts [33], $\eta_{\boldsymbol{w}}^{1/\alpha}$ is strictly convex when $\alpha > 1$ and $\Psi\left(\left\{\boldsymbol{x} : \boldsymbol{a}^T\boldsymbol{x} = 0\right\}\right) = 0$ for any $\boldsymbol{a} \in \mathbb{R}^d$. Suppose $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are two different minimal points of the optimization problem. Let $\boldsymbol{w} = (\boldsymbol{w}_1 + \boldsymbol{w}_2)/2$. From the strictly convexity of the object function and compactness of the set of constraints, it follows that $\eta_{\boldsymbol{w}}^{1/\alpha} < \eta_{\boldsymbol{w}_1}^{1/\alpha} = \eta_{\boldsymbol{w}_2}^{1/\alpha}$, which yields a contradiction. Thus, $\boldsymbol{w}^*$ is unique.

Now we prove (2.7). In the proof of Theorem A.2, we showed that $\mathrm{VaR}_q(\boldsymbol{w}^T\boldsymbol{X})$ is continuous with respect to $\boldsymbol{w} \in \Sigma^d$ for q large. Then there exists $q^* > 0$ such that $\mathrm{DR}_{\boldsymbol{w},q}$ is continuous with respect to $\boldsymbol{w} \in \Sigma^d$ for every $q^* < q < 1$. The desired result follows from Theorem 2.1, the uniqueness of $\boldsymbol{w}^*$ and Lemma 2.1. ■

Remark 2.1 *Assuming a positive joint density function for the random vector* $\boldsymbol{X}$ *is to ensure the distribution is strictly increasing, which is a technical condition needed in the proof.*

A related problem to our setting is the utility maximization problem. By Berge's Maximum Theorem, its maximizers are continuous on the parameters. However, this theorem is not applicable to our problem as it requires that $\mathrm{DR}_{\boldsymbol{w},q}$ *is continuous on* $\boldsymbol{w}$ *and* q *jointly. In fact, under the current conditions (see Theorem 2.2), the continuity of* $\mathrm{DR}_{\boldsymbol{w},q}$ *on* $\boldsymbol{w}$ *and* q *separately does not lead to the continuity of* $\mathrm{DR}_{\boldsymbol{w},q}$ *on* $\boldsymbol{w}$ *and* q *simultaneously. Hence, in the proof of Theorem 2.2, we need to rely on Lemma 2.1 to show the uniform convergence of* $\mathrm{DR}_{\boldsymbol{w},q}$ *to* $\mathrm{DR}_{\boldsymbol{w},1}$ *when* q *is close to* 1 *as in Theorem 2.1.*

2.3 Beyond the main theorem

In our main result, Theorem 2.2, some restrictions are imposed on the index α and spectral measure Ψ to make sure that the optimization problem is well defined. In fact, they are not necessary conditions. In the following through several special cases, we show that the conditions can be relaxed.

The condition $\Psi\left(\left\{\boldsymbol{x}:\boldsymbol{a}^T\boldsymbol{x}=0\right\}\right)=0$ for any $\boldsymbol{a}\in\mathbb{R}^d$ means that the spectral measure Ψ does not concentrate on any linear subspace. It ensures the uniqueness of the solution $\boldsymbol{w}^*$ of the limiting problem $\mathrm{DR}_{\boldsymbol{w},1}$. But it excludes the special cases such as independent or comonotonic structure of $\boldsymbol{X}$. If $\boldsymbol{X}$ has independent structure with regularly varying marginals, then it is not hard to show that

$$\mathrm{DR}_{\boldsymbol{w},1}=\sum_{k=1}^{d} w_k^{\alpha}.$$

By Jensen's inequality, $\mathrm{DR}_{\boldsymbol{w},1}$ is minimized when $w_k=1/d$ for $k=1,2,...,d$, which is unique. Therefore, Theorem 2.2 holds for the independent case. If $\boldsymbol{X}$ is comonotonic, then $\mathrm{DR}_{\boldsymbol{w},q}=1$ for any $\boldsymbol{w}$ or q. There is no optimization problem to consider.

If we restrict ourselves to elliptical distributions, then Theorem 2.2 holds for any $\alpha\in\mathbb{R}$, without any restriction on Ψ, or even without the MRV assumption. In the rest of the section, we focus on this special case.

A random vector $\boldsymbol{X}$ in $\mathbb{R}^d$ is elliptically distributed if it satisfies

$$\boldsymbol{X}\overset{d}{=}\boldsymbol{\mu}+YB\boldsymbol{U}, \tag{2.9}$$

where $\boldsymbol{\mu}\in\mathbb{R}^d$, $B\in\mathbb{R}^{d\times d}$, $\boldsymbol{U}=(U_1,...,U_d)^T$ is uniformly distributed on the Euclidean sphere $\mathbb{S}_2^d$, and Y is a nonnegative random variable that is independent of $\boldsymbol{U}$. The matrix $C:=BB^T$ is called ellipticity matrix of $\boldsymbol{X}$. To avoid degenerate cases, we assume throughout the following that C is positive definite.

It is well known that if $\boldsymbol{X}$ is elliptically distributed, then $\boldsymbol{X}\in\mathrm{MRV}_\alpha(\Psi)$ if and only if $Y\in\mathrm{RV}_{-\alpha}$; for example, see Hult and Lindskog [27]. By Theorem 6.8 of McNeil et al. [36], the subadditivity property of VaR always holds for $0.5\le q<1$. It then follows that $\mathrm{DR}_{\boldsymbol{w},q}\le 1$, which means that diversification is always optimal for $0.5\le q<1$ no matter what distribution Y follows and thus the optimization problem is well defined. In the general MRV case, to have $\mathrm{DR}_{\boldsymbol{w},q}\le 1$ is ensured by restricting $\alpha>1$. In another word, if $\boldsymbol{X}$ is elliptically distributed and $Y\in\mathrm{RV}_{-\alpha}$, then Theorem 2.2 holds without any restriction on α.

Actually, elliptical distributions lead to the explicit expressions of $\mathrm{DR}_{\boldsymbol{w},q}$ and $\mathrm{DR}_{\boldsymbol{w},1}$. This enables us to further relax the assumption of MRV. As long as Y is unbounded, we are able to directly show the convergence of (2.7) without the assumption that Y is regularly varying. A direct calculation yields that

$$\mathrm{VaR}_q\left(\boldsymbol{w}^T\boldsymbol{X}\right)=\boldsymbol{w}^T\boldsymbol{\mu}+\left\|B^T\boldsymbol{w}\right\|_2 F_Z^{\leftarrow}(q), \tag{2.10}$$

where $Z\overset{d}{=}YU_1$. The diversification ratio for elliptical distributions can then be obtained as

$$\mathrm{DR}_{\boldsymbol{w},q}=\frac{\boldsymbol{w}^T\boldsymbol{\mu}+\left\|B^T\boldsymbol{w}\right\|_2 F_Z^{\leftarrow}(q)}{\boldsymbol{w}^T\boldsymbol{\mu}+\sum_{i=1}^{d} w_i\left\|B^T\boldsymbol{e}_i\right\|_2 F_Z^{\leftarrow}(q)}. \tag{2.11}$$

If the random variable Y is unbounded, then by $F_Z^{\leftarrow}(q) \to \infty$ as $q \uparrow 1$, we obtain

$$\lim_{q\uparrow 1} \mathrm{DR}_{\boldsymbol{w},q} = \frac{\left\|B^T \boldsymbol{w}\right\|_2}{\sum_{i=1}^d w_i \left\|B^T \boldsymbol{e}_i\right\|_2} := \mathrm{DR}_{\boldsymbol{w},1}. \tag{2.12}$$

In the following lemma, we first show that the convergence in (2.12) is indeed uniform, whose proof is postponed to the last section.

Lemma 2.2 *For elliptically distributed $\boldsymbol{X}$ and $\boldsymbol{w} \in \Sigma^d$, if $\|\boldsymbol{\mu}\|_1 < \infty$, the induced norm $\|B\|_2 = \sup_{x\neq 0} \frac{\|B\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} < \infty$ and random variable Y is unbounded, then the convergence in (2.12) is uniform for $\boldsymbol{w} \in \Sigma^d$. Moreover, the mapping $\boldsymbol{w} \to \mathrm{DR}_{\boldsymbol{w},1}$ is continuous.*

Now we are ready to show that Theorem 2.2 holds in the most general setting of elliptical distributions by dropping the MRV assumption.

Theorem 2.3 *Under the conditions of Lemma 2.2, we have*

$$\lim_{q\uparrow 1} \arg\min_{\boldsymbol{w}\in\Sigma^d} \frac{\mathrm{VaR}_q(\boldsymbol{w}^T \boldsymbol{X})}{\sum_{i=1}^d w_i \mathrm{VaR}_q(X_i)} = \arg\min_{\boldsymbol{w}\in\Sigma^d} \frac{\left\|B^T \boldsymbol{w}\right\|_2}{\sum_{i=1}^d w_i \left\|B^T \boldsymbol{e}_i\right\|_2}. \tag{2.13}$$

Proof. By Lemmas 2.1 and 2.2, we only need to show that the solutions of the minimization problems on both sides of (2.13) exist and are unique. To achieve it, first note that the minimization problem

$$\min_{\boldsymbol{w}\in\Sigma^d} \frac{\mathrm{VaR}_q(\boldsymbol{w}^T \boldsymbol{X})}{\sum_{i=1}^d w_i \mathrm{VaR}_q(X_i)}$$

is equivalent to a convex optimization problem

$$\begin{aligned} &\min_{\boldsymbol{w}} \boldsymbol{w}^T \boldsymbol{\mu} + \left\|B^T \boldsymbol{w}\right\|_2 F_Z^{\leftarrow}(q) \\ &\boldsymbol{s.t.} \quad \boldsymbol{w}^T \boldsymbol{\mu} + \textstyle\sum_{i=1}^d w_i \left\|B^T \boldsymbol{e}_i\right\|_2 F_Z^{\leftarrow}(q) = 1 \text{ with } w_i \geq 0 \text{ for } i = 1, 2, \ldots, d. \end{aligned} \tag{2.14}$$

Similarly, the minimization problem

$$\min_{\boldsymbol{w}\in\Sigma^d} \frac{\left\|B^T \boldsymbol{w}\right\|_2}{\sum_{i=1}^d w_i \left\|B^T \boldsymbol{e}_i\right\|_2}$$

is equivalent to

$$\begin{aligned} &\min_{\boldsymbol{w}} \left\|B^T \boldsymbol{w}\right\|_2 \\ &\boldsymbol{s.t.} \quad \textstyle\sum_{i=1}^d w_i \left\|B^T \boldsymbol{e}_i\right\|_2 = 1 \text{ with } w_i \geq 0 \text{ for } i = 1, 2, \ldots, d. \end{aligned} \tag{2.15}$$

Denote the constraint sets in (2.14) and (2.15) by C_1 and C_2. It is obvious that C_1 and C_2 are nonempty, closed, convex and bounded. Hence, they are compact by the Heine-Borel theorem. By the triangle inequality and positive homogeneity of $\|\cdot\|_2$, the objective functions in (2.14) and (2.15) are convex over $\mathbb{R}^d$, and they are continuous over the constraint sets

C_1 and C_2; see Rochafellar (2015). By the compactness of the constraint set and continuity of the objective functions, the solutions to (2.14) and (2.15) exist due to the Weierstrass extreme value theorem.

Next, we show the uniqueness of the solution to (2.15). Due to the convexity, we have for any $\lambda \in (0,1)$,

$$\left\|B^T\left(\lambda \boldsymbol{w}_1+(1-\lambda)\,\boldsymbol{w}_2\right)\right\|_2 \leq \lambda\left\|B^T\boldsymbol{w}_1\right\|_2+(1-\lambda)\left\|B^T\boldsymbol{w}_2\right\|_2. \tag{2.16}$$

The equality in (2.16) holds only when $\boldsymbol{w}_1 = k\boldsymbol{w}_2$ for $k \in \mathbb{R}^+$ and $\boldsymbol{w}_1, \boldsymbol{w}_2$ nonzero. If both $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ belong to the constraint set C_1 or C_2, then k can only be 1. This means for any $\boldsymbol{w}_1 \neq \boldsymbol{w}_2$, the strictly inequality in (2.16) holds. Therefore, the objective function in (2.15) is strictly convex. The uniqueness of the solution then follows from the similarly arguments in the proof of Theorem 2.2. ■

3 The Rate of Convergence to the Optimal Portfolio: An Example

In this section, we discuss how $\boldsymbol{w}^*$ approximates $\boldsymbol{w}_q$ by determining the convergence rate of (2.7) under some special dependence structure, such as the FGM copula.

The FGM copula was originally introduced by Morgenstern [39] and investigated by Gumbel [24] and Farlie [21]. The FGM copula is defined as

$$C(u,v)=uv(1+\theta(1-u)(1-v)),\quad (u,v)\in[0,1]^2, \tag{3.1}$$

where $\theta \in [-1,1]$ is a dependence parameter. This model has been generalized in various ways, for example, from two dimensions to higher dimensions or with more general form of $(1-u)(1-v)$ in (3.1); see Cambanis [5], Fischer and Klein [22], among others. Here we focus on a high dimensional generalized FGM copula proposed by Cambanis [5], which is defined as

$$C(u_1,\ldots,u_n)=\prod_{k=1}^{n}u_k\left(1+\sum_{1\leq i<j\leq n}a_{ij}(1-u_i)(1-u_j)\right),\quad (u_1,\ldots,u_n)\in[0,1]^n. \tag{3.2}$$

The constants $a_{i,j}$, $1 \leq i < j \leq n$, are so chosen that $C(u_1,\ldots,u_n)$ is a proper copula. A necessary and sufficient condition on $a_{i,j}$'s is that they satisfy a set of 2^n inequalities

$$1+\sum_{1\leq i<j\leq n}\epsilon_i\epsilon_j a_{ij}\geq 0 \quad \text{for all } (\epsilon_1,\ldots,\epsilon_n)\in\{-1,1\}^n.$$

A FGM copula defined as in (3.2) is asymptotically independent.

We intend to consider the random vector $\boldsymbol{X}$ following FGM copula with identical regularly varying marginals. For that purpose we need a second-order convergence in Proposition

2.1. This further requires the second-order expansion of tail probabilities of the weighted sum

$$\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t) = \Pr\left(\boldsymbol{w}^T\boldsymbol{X} > t\right),$$

where $F_{\boldsymbol{w}^T\boldsymbol{X}} = 1 - \overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}$ is the distribution function of $\boldsymbol{w}^T\boldsymbol{X}$. In the next subsection, we present this result.

3.1 Tail expansion for the weighted sum

Assume that the random vector $\boldsymbol{X}$ has a common marginal distribution function $G = 1-\overline{G}$. Further, assume $\overline{G}$ to be *second-order regularly varying* (2RV), denoted by $\overline{G} \in 2\mathrm{RV}_{-\alpha,\rho}$. That is, there exist some $\rho \leq 0$ and a measurable function $A(\cdot)$, which does not change sign eventually and converges to 0, such that, for all $x > 0$,

$$\lim_{t\to\infty} \frac{\overline{G}(tx)/\overline{G}(t) - x^{-\alpha}}{A(t)} = x^{-\alpha}\frac{x^{\rho}-1}{\rho} =: H_{-\alpha,\rho}(x). \tag{3.3}$$

When $\rho = 0$, $H_{-\alpha,\rho}(x)$ is understood as $x^{-\alpha}\log x$.

To better understand the condition of 2RV, a simple example of $\overline{G} \in 2\mathrm{RV}_{-\alpha,\rho}$ is $\overline{G}(x) = Ax^{-\alpha}(1 + Bx^{\rho})$, where $A > 0$ and $B \in \mathbb{R}$. The smaller ρ means $\overline{G}(x)$ behaves more like a power function $Ax^{-\alpha}$ and hence the faster convergence rate in the asymptotic theory. This also explains the faster convergence rate of the optimal portfolio in Theorem 3.1 for smaller ρ.

For simplicity, here we only consider the case $\alpha > 1$ which implies that $\boldsymbol{X}$ has a finite mean. The results for $0 < \alpha \leq 1$ can be obtained in a similar way. The proof of the next lemma is collected in the appendix.

Lemma 3.1 *Let $\boldsymbol{X}$ be a nonnegative random vector with identically distributed marginal with common distribution function G satisfying that $\overline{G} \in 2\mathrm{RV}_{-\alpha,\rho}$ with $\alpha > 1$, $\rho \leq 0$ and auxiliary function $A(\cdot)$. Assume that $\boldsymbol{X}$ follows an n-dimensional generalized FGM copula given by (3.2). Then as $t \to \infty$, we have that*

$$\begin{aligned}
&\frac{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t)}{\overline{G}(t)} - \sum_{k=1}^{d} w_k^{\alpha} \\
&= \begin{cases} \alpha t^{-1}\mu_G^{*}(1+o(1)), & \rho < -1, \\ (1+Q_{\boldsymbol{a}})\sum_{k=1}^{d} H_{-\alpha,\rho}\left(w_k^{-1}\right)A(t)(1+o(1)), & \rho \geq -1, \end{cases}
\end{aligned} \tag{3.4}$$

where $H_{-\alpha,\rho}(\cdot)$ is given in (3.3), $Q_{\boldsymbol{a}} = \sum_{1\leq i<j\leq n} a_{ij}$, $\mu_G = \int_0^\infty x dF(x)$, $\mu_{G^2} = \int_0^\infty x dF^2(x)$,

and

$$
\begin{aligned}
\mu_G^* = &(1+Q_{\boldsymbol{a}})\,\mu_G \sum_{k\neq l} w_k^\alpha w_l \\
&+ \sum_{i<j} a_{i,j} \left(\sum_{k,l=i,j} \left(\sum_{l\neq k} \mu_{G^2} w_k^\alpha w_l - \mu_G w_k \sum_{m\neq i,j} w_m^\alpha - 2\mu_G w_k^\alpha w_l - \mu_G w_k^\alpha w_l \right)\right) \\
&- \sum_{i<j} a_{i,j} \sum_{k\neq i,j} \sum_{l\neq k,i,j} \mu_G w_k^\alpha w_l .
\end{aligned}
$$

Further, the convergence in (3.4) is uniform for all $\boldsymbol{w} \in \Sigma^d$.

3.2 Convergence rate

We first show a general lemma regarding the convergence rate of minimizers under the setup of Lemma 2.1. Define the distance between f_n and f as $D_n = ||f_n - f||_\infty$, where $||\cdot||_\infty$ is the supremum norm. The distance between m_{f_n} and $\arg\min f$ is defined as $||m_{f_n} - \arg\min f||_\Box$ for a norm $||\cdot||_\Box$ on the space Z. Since Z is a metric space, all the norms on Z are equivalent in the sense that there exist constants c_1 and c_2 such that

$$c_1||x||_\Box \leq ||x||_\Diamond \leq c_2||x||_\Box, \qquad x \in Z,$$

for any two norms $||\cdot||_\Box$ and $||\cdot||_\Diamond$ on Z. In case no confusion arises, the norm index ∞ or $\Box$ is dropped in the rest of the paper.

Lemma 3.2 *Under the assumptions of Lemma 2.1, we have for n large*

$$||m_{f_n} - \arg\min f|| < C\sqrt{D_n},$$

where $D_n = ||f_n - f||_\infty$ *and* C *is a constant.*

Lemma 2.1 shows that m_{f_n}, the minimizer of function f_n, can be approximated by the minimizer of the limiting function f, which is usually much easier to calculate. The result in Lemma 3.2 further explores how good the approximation is. In practice, if we can determine D_n, which is related to the second-order expansion of f_n, then the error of the approximation can be determined.

Now we are ready to determine the convergence rate of the optimal portfolio under the FGM copula.

Theorem 3.1 *Under the conditions of Lemma 3.1, we have that as* $q \uparrow 1$,

$$(1-q)^{(-1\vee\rho)/\alpha} \left\| \boldsymbol{w}_q - \boldsymbol{d}^{-1} \right\| = O(1),$$

where $\boldsymbol{w}_q$ *is a solution of* $\min_{\boldsymbol{w}\in\Sigma^d} \mathrm{DR}_{\boldsymbol{w},q}$, *and* $\boldsymbol{d}^{-1} = (1/d, ..., 1/d)^T$.

Proof. In this proof, all the limits are taken as $q \uparrow 1$. We first derive the second-order expansion of $\mathrm{DR}_{\boldsymbol{w},q}$. Similar to the proof of Theorem 4.6 in Mao and Yang [35], we have that

$$U\left(\frac{1}{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}(q))}\right) = G^{\leftarrow}(q) + o(A(G^{\leftarrow}(q))),$$

where $U(\cdot)$ is the tail quantile function of G defined as $U(\cdot) = (1/\overline{G})^{\leftarrow}(\cdot) = G^{\leftarrow}(1 - 1/\cdot)$. For simplicity, denote $t = F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}(q)$. It is easy to see that $t \to \infty$ as $q \uparrow 1$. Then noting that $U(1/\overline{G}(t)) = t + o(A(t))$ and by the uniform convergence of (3.3), it follows that

$$\begin{aligned}
\mathrm{DR}_{\boldsymbol{w},q} &= \frac{F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}(q)}{G^{\leftarrow}(q)} = \frac{U(1/\overline{G}(t))}{U(1/\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t))} + o(A(t)) \\
&= \left(\frac{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t)}{\overline{G}(t)}\right)^{1/\alpha} + H_{1/\alpha,\rho/\alpha}\left(\frac{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t)}{\overline{G}(t)}\right)\alpha^{-2}A(U(1/\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t)))(1+o(1)) \\
&= \begin{cases} \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha}\left(1 + \mu_G^*\left(\sum_{k=1}^d w_k^\alpha\right)^{-1/\alpha-1}(G^{\leftarrow}(q))^{-1}(1+o(1))\right), & \rho < -1, \\ \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha}(1 + \tau_\alpha A(G^{\leftarrow}(q))(1+o(1))), & \rho > -1. \end{cases}
\end{aligned} \tag{3.5}$$

where

$$\tau_\alpha = \frac{(1+Q_a)\sum_{k=1}^d H_{-\alpha,\rho}\left(w_k^{-1}\right)}{\alpha\sum_{k=1}^d w_k^\alpha} + \frac{\left(\sum_{k=1}^d w_k^\alpha\right)^{\rho/\alpha}}{\rho\alpha}.$$

This gives the second-order expansion of $\mathrm{DR}_{\boldsymbol{w},q}$.

Immediately from (3.5), the limiting function is

$$\lim_{q\uparrow 1}\mathrm{DR}_{\boldsymbol{w},q} = \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha} = \mathrm{DR}_{\boldsymbol{w},1}.$$

By Jensen's inequality, $\mathrm{DR}_{\boldsymbol{w},1}$ is uniquely minimized at $\boldsymbol{d}^{-1} = (1/d, ..., 1/d)^T$. If $\rho < -1$, then

$$\mathrm{DR}_{\boldsymbol{w},q} - \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha} = \mu_G^*\left(\sum_{k=1}^d w_k^\alpha\right)^{-1}(G^{\leftarrow}(q))^{-1}(1+o(1)).$$

By Lemma 3.1, the above convergence is uniform. Hence, we have that for some constant $C > 0$

$$\left|\mathrm{DR}_{\boldsymbol{w},q} - \left(\sum_{k=1}^d w_k^\alpha\right)^{1/\alpha}\right| < C\,(G^{\leftarrow}(q))^{-1}.$$

By Lemma 3.2, we get that

$$(1-q)^{-1/\alpha}\left\|\boldsymbol{w}_q - \boldsymbol{d}^{-1}\right\| = O(1).$$

Similarly, if $\rho > -1$, then

$$\mathrm{DR}_{\boldsymbol{w},q} - \left(\sum_{k=1}^{d} w_k^{\alpha}\right)^{1/\alpha} = \left(\sum_{k=1}^{d} w_k^{\alpha}\right)^{1/\alpha} \tau_\alpha A(G^{\leftarrow}(q))(1+o(1)).$$

Since for any $\boldsymbol{w} \in \Sigma^d$

$$\tau_\alpha \leq \frac{(1+Q_a)\,\rho d^{(\alpha-1)^2/\alpha} + d^{\rho(1-\alpha)/\alpha}}{\rho\alpha},$$

we obtain that for some constant $C > 0$

$$\left| \mathrm{DR}_{\boldsymbol{w},q} - \left(\sum_{k=1}^{d} w_k^{\alpha}\right)^{1/\alpha} \right| < CA(G^{\leftarrow}(q)).$$

By Lemma 3.2 we get that

$$(1-q)^{\rho/\alpha} \left\| \boldsymbol{w}_q - \boldsymbol{d}^{-1} \right\| = O(1).$$

This completes the proof. ∎

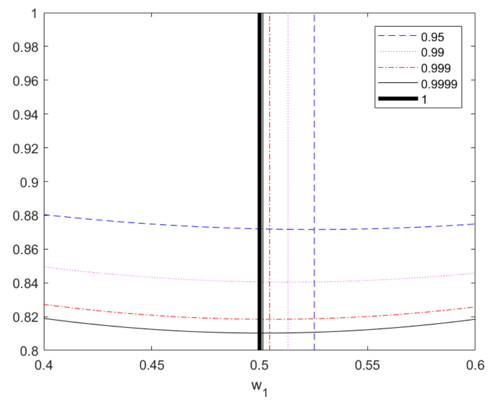
4 Numerical examples

In this section, we conduct three numerical examples to examine our theoretical results. The first two examples are elliptical distributions involving 2- and 3-dimensional Student's t-distributions, while the third one is a non-elliptical distribution.

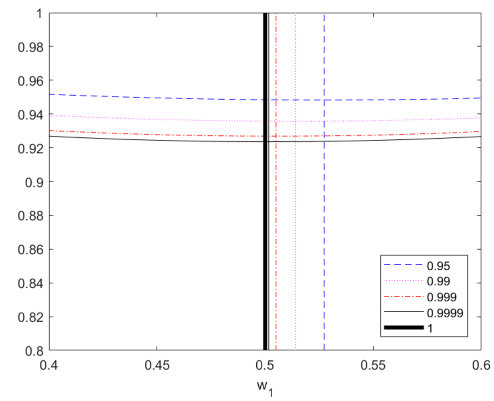
Consider $\boldsymbol{X}$ follows a bivariate Student's t-distribution[2] $t_\alpha(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (1,2)^T$ and the scale matrix Σ is $\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$. Then both marginals follow Student's t-distribution with the degree of freedom α but with different shifts 1 and 2.

We construct portfolios as a linear combination of the two risk factors from $\boldsymbol{X}$ defined above. As discussed in Section 2.3, both $\mathrm{DR}_{\boldsymbol{w},q}$ and $\mathrm{DR}_{\boldsymbol{w},1}$ can be explicitly expressed for elliptical distributions as in (2.11) and (2.12), which are used in this example. In Figure 1, we plot the diversification ratio of such portfolios for various values of q against the weight w_1. For the parameters, we choose α and r at $\alpha = 2, 4$ and $r = 0.3, 0.7$, and plot the results for different pairs of (α, r) in the four subfigures in Figure 1. The level of q is set to 0.95, 0.99, 0.999 and 0.9999. For each q level, we indicate the optimal portfolio weight on w_1 by a vertical line, which is given at the lowest point of the convex diversification ratio curve. Notice that due to the different shifts, the optimal portfolio at a finite q level tends to load higher on the first dimension with a lower mean. However, as $q \to 1$, the difference in the mean plays no role in the limit of the diversification ratio. Therefore, due to symmetry, the optimal portfolio for $q = 1$ load equal weights on the two dimensions. We indicate this optimal solution for the limit diversification ratio by a thick vertical line located at 0.5.

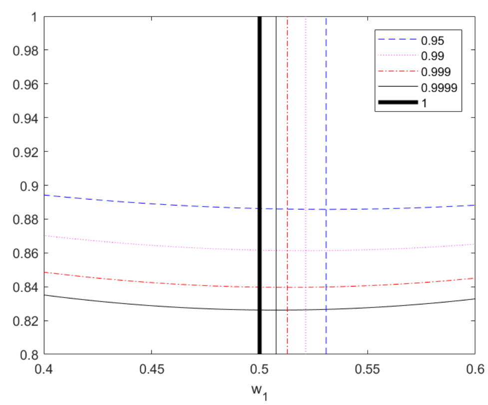
[2] The left tail of a Student's t-distribution does not play a role in our analysis. It can be understood as a truncated Student's t-distribution with a mass point at 0.



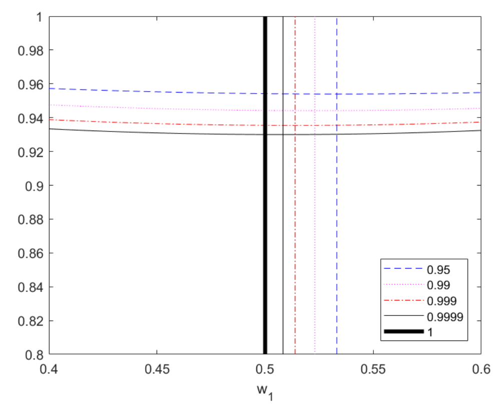
(a) $\alpha = 2,\ r = 0.3$



(b) $\alpha = 2,\ r = 0.7$



(c) $\alpha = 4,\ r = 0.3$



(d) $\alpha = 4,\ r = 0.7$

Figure 1: Optimal portfolio based on 2-d elliptical distribution risk factors. The portfolios are constructed as a linear combination of two risk factors from a bivariate Student's t-distribution $t_\alpha(\boldsymbol{\mu}, \Sigma)$ with $\mu = (1,2)^T$ and Σ is $\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$. The $\mathrm{DR}_{\boldsymbol{w},q}$ of such portfolios for various values of q against the weight w_1 are plotted for different pairs of (α, r) with $\alpha = 2, 4$ and $r = 0.3, 0.7$ in the four subfigures. The levels of q are set to 0.95, 0.99, 0.999 and 0.9999. For each q level, the optimal portfolio weight on w_1 is indicated by a vertical line of different style. The optimal solution for $\mathrm{DR}_{\boldsymbol{w},1}$ is indicated by a thick vertical line.

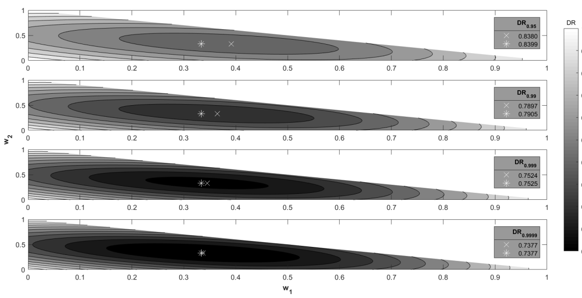
First, we observe that $\boldsymbol{w}_q$ converges to $\boldsymbol{w}_1$ as $q \uparrow 1$. This verifies our theoretical result as established in Theorem 2.2. Second, the absolute difference between $\boldsymbol{w}_q$ and $\boldsymbol{w}_1$ remains at a low level across all subfigures. For example, when focusing on approximating the optimal portfolio based on diversification ratio at $q = 0.99$ level, if one takes the optimal weight for the limit diversification ratio 0.5 as an approximation, then she makes an error for loading 2% less on the first dimension. Third, given the level of dependence (r), the heavier the marginal tails reflected in a lower α, the faster the convergence rate. This is in line with Theorem 3.1's finding: α plays a role in the speed of convergence, the higher the α, the slower the speed of convergence. Lastly, when fixing the level of heavy-tailedness (α), the more dependence reflected in a higher r, the slower the convergence rate in the limit relation $\boldsymbol{w}_q \to \boldsymbol{w}_1$. Nevertheless, the slow convergence is not of a concern in practice. With a strong dependence in the first place, the room for diversification benefit is limited. As a result, the diversification ratio is in general at a high level and is less sensitive to the variation of the weights. Therefore, with a strong dependence, although the solution in the limit $(0.5, 0.5)^T$ might not be close to the optimal solution at a finite q, investing in the portfolio $(0.5, 0.5)^T$ would not result in a large increase in diversification ratio at a finite q level, compared to the actual optimal portfolio.

We now proceed to a 3-dimensional example with the portfolio constructed as a linear combination of the three risk factors from $\boldsymbol{X}$ following a 3-dimensional Student's t-distribution $t_\alpha(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (1, 2, 3)^T$ and the scale matrix Σ is $\begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix}$. We choose the same parameters as above, that is $\alpha = 2, 4$ and $r = 0.3, 0.7$. Figure 2 shows the contour plots of DR of such portfolios for various values of q and weights w_1 and w_2. More specifically, in the uppermost graph of Figure 2a, the circles represent the contours of $\mathrm{DR}_{\boldsymbol{w},q}$ having the same value as $\mathrm{DR}_{\boldsymbol{w}_{0.95},0.95}$. The location of $\boldsymbol{w}_{0.95}$ is marked by the "cross" sign and $\boldsymbol{w}_1 = (1/3, 1/3, 1/3)^T$ is marked by the "star" sign in the graph. The values of $\mathrm{DR}_{\boldsymbol{w}_{0.95},0.95}$ and $\mathrm{DR}_{\boldsymbol{w}_1,0.95}$ are reported in the legend. In the remaining graphs of Figure 2a, the contours are plotted at $\mathrm{DR}_{\boldsymbol{w}_{0.99},0.99}$, $\mathrm{DR}_{\boldsymbol{w}_{0.999},0.999}$ and $\mathrm{DR}_{\boldsymbol{w}_{0.9999},0.9999}$, respectively. The shaded bar on the right of Figure 2a shows the values of DR represented by different shades. Similar conclusions can be obtained for the 3-d example. Comparing with the 2-d example, the convergence speed of the 3-d case is not significantly slower.

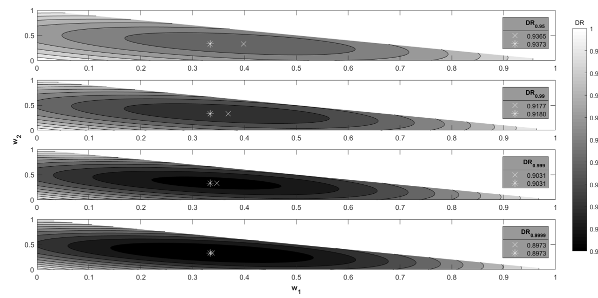
Next, we study a different numerical example based on a non-elliptical distribution. We construct the example using linear combinations of heavy-tailed random variables. Let Y_1 and Y_2 be two i.i.d. random variables with regularly varying tails. A random vector $\boldsymbol{X} = (X_1, X_2)^T$ is then defined as

$$\boldsymbol{X} = A\boldsymbol{Y}, \quad A := \begin{pmatrix} 1 & 0 \\ r & \sqrt{1 - r^2} \end{pmatrix}, \tag{4.1}$$

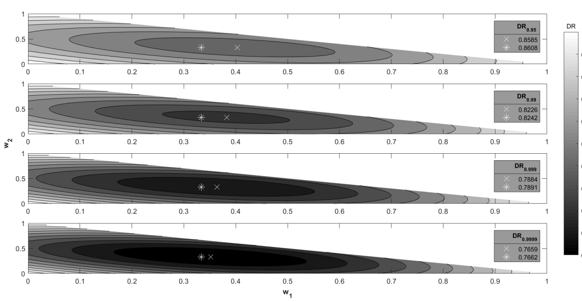
where $r \in (-1, 1)$. Such random vector follows a non-elliptical distribution. In the case that the variance of Y_1 and Y_2 exists, r is the correlation coefficient between X_1 and X_2 Under this structure, the diversification ratio $\mathrm{DR}_{\boldsymbol{w},1}$ can be explicitly calculated. Following



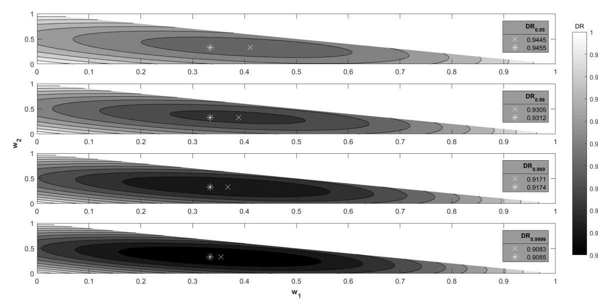
(a) $\alpha = 2,\ r = 0.3$



(b) $\alpha = 2,\ r = 0.7$



(c) $\alpha = 4,\ r = 0.3$



(d) $\alpha = 4,\ r = 0.7$

Figure 2: Optimal portfolio based on 3-d elliptical distribution risk factors. The portfolios are constructed as a linear combination of three risk factors from a 3-dimensional Student's t-distribution $t_\alpha(\boldsymbol{\mu}, \Sigma)$ with $\mu = (1, 2, 3)^T$ and Σ is $\begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix}$. Contours of DR at $\mathrm{DR}_{\boldsymbol{w}_q, q}$ of $q = 0.95$, 0.99, 0.999 and 0.9999 are plotted for different pairs of (α, r) with $\alpha = 2, 4$ and $r = 0.3, 0.7$ in the four subfigures. The location of w_q is marked by the "cross" sign and w_1 is marked by the "star" sign. The values of $\mathrm{DR}_{\boldsymbol{w}_q, q}$ and $\mathrm{DR}_{\boldsymbol{w}_1, q}$ are reported in the legend. The shaded bar on the right of each subfigure shows the values of DR represented by different shades.

Mainik and Embrechts [33], we have that

$$\frac{\eta_{\boldsymbol{w}}}{\eta_{\boldsymbol{e}_1}} = (w_1 + w_2 r)^\alpha + \left(w_2\sqrt{1-r^2}\right)^\alpha,$$

and

$$\frac{\eta_{\boldsymbol{w}}}{\eta_{\boldsymbol{e}_2}} = \frac{(w_1 + w_2 r)^\alpha + \left(w_2\sqrt{1-r^2}\right)^\alpha}{r^\alpha + \sqrt{1-r^2}^\alpha}.$$

Hence,

$$\mathrm{DR}_{\boldsymbol{w},1} = \left(w_1 \left((w_1 + w_2 r)^\alpha + \left(w_2\sqrt{1-r^2}\right)^\alpha \right)^{-\frac{1}{\alpha}} + w_2 \left(\frac{(w_1 + w_2 r)^\alpha + \left(w_2\sqrt{1-r^2}\right)^\alpha}{r^\alpha + \sqrt{1-r^2}^\alpha} \right)^{-\frac{1}{\alpha}} \right)^{-1}.$$

We use this formula to determine $\mathrm{DR}_{\boldsymbol{w},1}$. Since the expression for $\mathrm{DR}_{\boldsymbol{w},q}$ is less explicit, its calculation is based on simulations.

Consider a special case where Y_1 and Y_2 follow a standard Student's t-distribution with degree of freedom $\alpha > 1$. By choosing $\alpha = 2, 4$ and $r = 0.3, 0.7$, in Figure 3 we plot the calculated diversification ratios $\mathrm{DR}_{\boldsymbol{w},q}$ against the loading on X_1, w_1 for various values of q: 0.95, 0.99, 0.999 and 0.9999. The optimal weight for each q level is again marked by a corresponding vertical line, with thick vertical line indicating the optimal weight for the limit case $q = 1$.

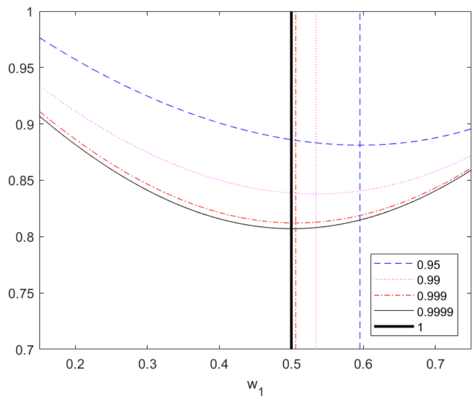
All four observations in the elliptical case remain qualitatively valid for the non-elliptical case. Quantitatively, the distance between the optimal solutions for finite q and the limit case can be far apart. For example, in the worst case scenario when the lower tail index meets the stronger dependence (right bottom subfigure), the distance between the optimal weight for $q = 0.99$ and that for $q = 1$ is around 0.25. In this case, the optimal portfolio in the limit is not a good approximation for that based on a finite q. To summarize, we recommend using the optimal portfolio based on the limit diversification ratio particularly for the case with low cross-sectional dependence and heavy marginal tails.

5 Empirical study

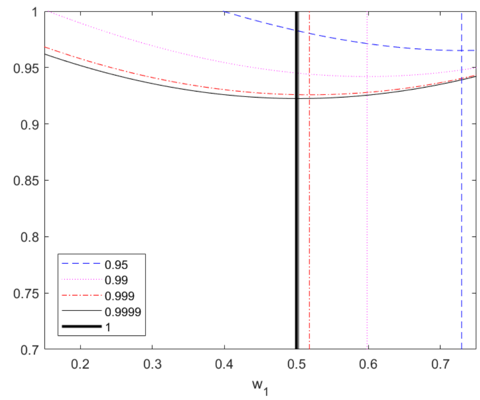
In the numerical examples, the limit diversification ratio $\mathrm{DR}_{\boldsymbol{w},1}$ can be calculated explicitly. With real data application, we need to estimate this function using historical data, and then consider the optimal portfolio based on the estimated diversification ratio. Next, we first discuss the estimation methodology for $\mathrm{DR}_{\boldsymbol{w},1}$. Then we apply our estimation method and the optimal portfolio construction procedure to real market data.

5.1 Estimation of the diversification ratio

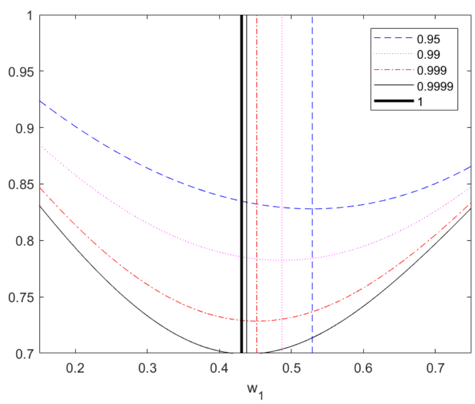
When the DR optimization strategy with MRV structure is applied in practice, the estimations of MRV structure and $\mathrm{DR}_{\boldsymbol{w},1}$ are required. In this subsection, we propose an



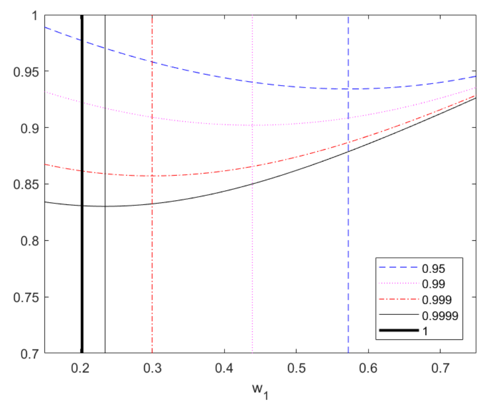
(a) $\alpha = 2,\ r = 0.3$



(b) $\alpha = 2,\ r = 0.7$



(c) $\alpha = 4,\ r = 0.3$



(d) $\alpha = 4,\ r = 0.7$

Figure 3: Optimal portfolio from non-elliptical distribution risk factors. The portfolios are constructed as a linear combination of two risk factors from a vector $\boldsymbol{X}$ defined in (4.1) with Y_1 and Y_2 following a standard Student's t-distribution with degree of freedom $\alpha > 1$. The $\mathrm{DR}_{\boldsymbol{w},q}$ of such portfolios for various values of q against the weight w_1 are plotted for different pairs of (α, r) with $\alpha = 2, 4$ and $r = 0.3, 0.7$ in the four subfigures. The levels of q are set to 0.95, 0.99, 0.999 and 0.9999. For each q level, the optimal portfolio weight on w_1 is indicated by a vertical line of different style. The optimal solution for $\mathrm{DR}_{\boldsymbol{w},1}$ is indicated by a thick vertical line.

estimation procedure and show the consistency of the estimators. Our estimation method is also consistent with that of Mainik and Rüchendorf [32].

Assume $\boldsymbol{X} \in \mathrm{MRV}_\alpha(\Psi)$ with $\alpha > 1$. Let $\boldsymbol{X}_1, \ldots \boldsymbol{X}_n$ be an i.i.d. sample of $\boldsymbol{X}$. By Theorem 2.1, we propose the following estimation procedure.

1. Estimate the tail index α by an estimator $\widehat{\alpha}$.
2. Estimate the spectral measure Ψ by an estimator $\widehat{\Psi}$.
3. Estimate $\eta_{\boldsymbol{w}}$ by
$$\widehat{\eta}_{\boldsymbol{w}} = \int_{\Sigma^d} (\boldsymbol{w}^T \boldsymbol{s})^{\widehat{\alpha}} \widehat{\Psi}(d\boldsymbol{s}).$$
4. Estimate $\mathrm{DR}_{\boldsymbol{w},1}$ by
$$\widehat{\mathrm{DR}}_{\boldsymbol{w},1} = \frac{\widehat{\eta}_{\boldsymbol{w}}^{1/\alpha}}{\sum_{i=1}^d w_i \widehat{\eta}_{\boldsymbol{e}_i}^{1/\alpha}}.$$

With the estimated diversification ratio, we can obtain an optimal portfolio by minimizing $\widehat{\mathrm{DR}}_{\boldsymbol{w},1}$. Denote the optimal portfolio weights following this procedure as $\hat{\boldsymbol{w}}^*$.

More specifically, in the first two steps, we use standard estimators for α and Ψ as follows. Let (R, S) and (R_i, S_i) denote the polar coordinates of $\boldsymbol{X}$ and $\boldsymbol{X}_i$ with respect to $||\cdot||_1$. That is,
$$(R, S) = \left(||\boldsymbol{X}||_1, \frac{\boldsymbol{X}}{||\boldsymbol{X}||_1}\right). \tag{5.1}$$
Assume in this section that the distribution function of R is continuous. Choose an intermediate sequence k such that
$$k\,(n) \to \infty, \qquad \frac{k\,(n)}{n} \to 0.$$
We use the observations corresponding to the top k order statistics of $R_1, \ldots, R_n$ for estimating α and Ψ. Denote the k upper order statistics of $R_1, \ldots, R_n$ by $R_{(1)} \geq \ldots \geq R_{(k)}$. The tail index α is estimated by some usual estimator as a function of these order statistics:
$$\widehat{\alpha} = \widehat{\alpha}\left(R_{(1)}, \ldots, R_{(k)}\right).$$
When $\alpha > 0$, many estimators can be applied here such as Hill estimator (Hill [25]), Pickands estimator (Pickands [41]), the maximum likelihood estimator (e.g. Smith [45]), and the moment estimator (Dekkers et al. [12]). Improvements of the aforementioned standard estimators have been proposed to be better applied in practice, for example bias-reduced Hill estimator (e.g. Peng [40] and Caeiro et al. [4]), estimation for tail index with covariates (e.g. Wang and Tsai [48] and Daouia et al. [9]), estimation of tail index for non-iid samples (e.g. Drees [13] and Einmahl et al. [15]), among others. They all possess consistency and asymptotic normality.

Next, let $\pi(1), \ldots, \pi(k)$ denote the indices corresponding to $R_{(1)}, \ldots, R_{(k)}$ in the original sequence $R_1, \ldots, R_n$. These indices are used to identify each "angle" $S_{\pi(j)}$ corresponding to $R_{(j)}$. The spectral measure Ψ is estimated by the empirical measure of the angular parts $S_{\pi(1)}, \ldots, S_{\pi(k)}$,

$$\widehat{\Psi} = \frac{1}{k} \sum_{j=1}^{k} \delta_{S_{\pi(j)}}, \tag{5.2}$$

where $\delta_{\pi(j)}(\cdot)$ is the Dirac measure. See Chapter 9 of Resnick [43] for more details. Other estimation methods of the spectral measure, especially for the bivariate case, can be found in e.g. Einmahl et al. [16], and Chapter 9 of Beirlant et al. [3], Chapter 7 of de Haan and Ferreira [10], Einmahl and Segers [17], and Eastoe et al. [14].

Lemma 5.1 *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be an i.i.d. sample of $\boldsymbol{X} \in \mathrm{MRV}_\alpha(\Psi)$ with $\alpha > 1$. Assume that the distribution function F_R of R in (5.1) is continuous. If the estimator $\widehat{\alpha}$ is consistent almost surely, then the estimator $\widehat{\mathrm{DR}}_{\boldsymbol{w},1}$ is consistent uniformly in $\boldsymbol{w} \in \Sigma^d$, i.e.,*

$$\sup_{\boldsymbol{w} \in \Sigma^d} \left| \widehat{\mathrm{DR}}_{\boldsymbol{w},1} - \mathrm{DR}_{\boldsymbol{w},1} \right| \to 0, \qquad a.s. \tag{5.3}$$

Combining Theorem 2.1 and Lemma 5.1, we obtain the consistency in the optimal portfolio weights in the following theorem.

Theorem 5.1 *Under the conditions of Theorem 5.1 and $\Psi\left(\left\{\boldsymbol{x} : \boldsymbol{a}^T\boldsymbol{x} = 0\right\}\right) = 0$ for any $\boldsymbol{a} \in \mathbb{R}^d$, the estimator $\widehat{\boldsymbol{w}}^*$ and the estimated value $\hat{\mathrm{DR}}_{\boldsymbol{w}^*,1}$ are consistent almost surely, i.e.,*

$$\widehat{\boldsymbol{w}}^* \to \boldsymbol{w}^*, \qquad a.s.; \qquad \widehat{\mathrm{DR}}_{\boldsymbol{w}^*,1} \to \mathrm{DR}_{\boldsymbol{w}^*,1}, \qquad a.s.$$

Here we only established consistency. Under some additional conditions, further asymptotic properties for the estimator of $\mathrm{DR}_{\boldsymbol{w},1}$ can be established in a straightforward way. For example, Theorem 4.5 of Mainik and Rüchendorf [32] shows that, under some additional conditions, for any $\boldsymbol{w} \in \Sigma^d$, $\sqrt{k}\left(\widehat{\eta}_{\boldsymbol{w}} - \eta_{\boldsymbol{w}}\right)$ converges to a multivariate Gaussian distribution $G_{\boldsymbol{w}}$. Then by the functional delta method (e.g. Theorem 20.8 in Van der Vaart [47]), it is easy to show that $\sqrt{k}\left(\widehat{\mathrm{DR}}_{\boldsymbol{w},1} - \mathrm{DR}_{\boldsymbol{w},1}\right)$ converges to a Gaussian distribution as well. However, to establish the convergence in an uniform way is difficult and may be left for future research. Without a uniform asymptotic property on $\widehat{\mathrm{DR}}_{\boldsymbol{w},1}$ we cannot further investigate the asymptotic property of the optimal portfolio weights.

5.2 Real data analysis

The dataset consists of underlying stocks in the S&P 500 index that have a full trading history throughout the period from January 2, 2002 to December 31, 2015. This results in 425 stocks. We construct the continuously compounded loss returns of these stocks. That is, if the price of asset i at time t is denoted by $P_i(t)$, then the log loss at time t for asset i, denoted by $X_i(t)$ is given by

$$X_i(t) = -\log\left(\frac{P_i(t)}{P_i(t-1)}\right).$$

We conduct three empirical studies. Firstly, we demonstrate the difference between the optimal portfolio constructed based on minimizing a diversification ratio at a finite q level and that based on minimizing the limit diversification ratio. Secondly, we show that our proposed methodology has the advantage of bearing less computational burden. Lastly, we evaluate the out-of-sample performance between our portfolio optimization procedure and those existing in the literature.

The first empirical study is set up as follows. To avoid dimensional curse in the numerical search strategy (see below), we select 10 stocks from the dataset that share a similar level of tail index. Notice that having the same marginal tail index is a necessary condition for MRV. We estimate the tail indices of the 425 stocks using the Hill estimator (Hill [25]) as

$$\widehat{\alpha} = \frac{k}{\sum_{n=1}^{k}\log\left(R_{(n)}/R_{(k+1)}\right)}.$$

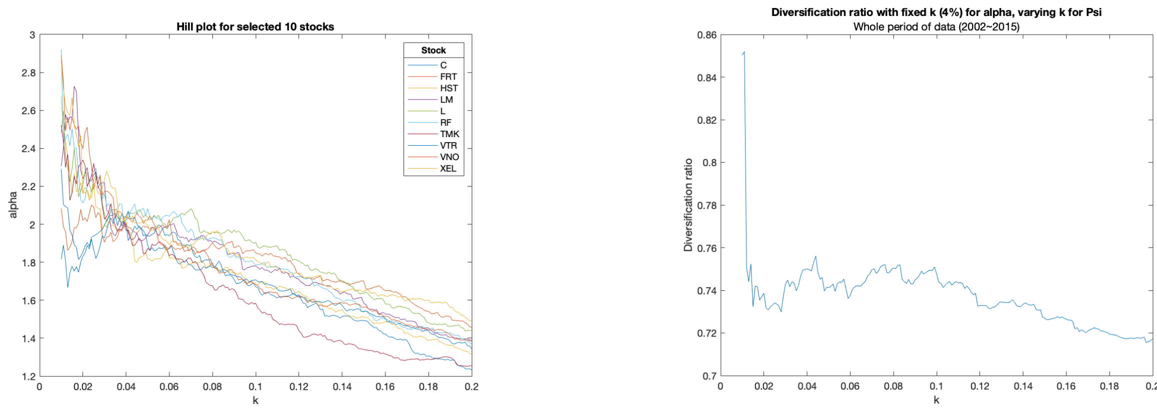
We select 10 stocks with the lowest estimates that are not significantly different from each other. Here, to test whether the 10 stocks have significantly different tail indices, we employ the test constructed in Moore et al. [38] for testing tail index equivalence. In other words, we select 10 stocks with the lowest estimates while not being rejected by this test. The reason for selecting stocks with lower α follows from the numerical example: the approximation works better when α is lower. The selected stocks are given in Table 1, where the estimate of α and its standard deviation (std) for each stock are provided. From Table 1, we observe that the point estimates of the tail index range from 1.989 to 2.040.

Table 1: Tail index estimates for the 10 selected stocks

Stock	C	FRT	HST	LM	L	RF	TMK	VTR	VNO	XEL
$\hat{\alpha}$	1.989	2.000	2.002	2.007	2.012	2.014	2.019	2.036	2.036	2.040
std	0.168	0.169	0.169	0.170	0.170	0.170	0.171	0.172	0.172	0.172

Note: The table shows the tail index estimates for 10 selected stocks within the S&P 500 index based on their daily returns in the period from January 2, 2002 to December 31, 2015. The tail indices are estimated using the Hill estimator (Hill [25]). The second row reports the standard deviations of the estimates.

Our empirical analysis is based on daily data in each five-year window, namely, 2002–2006, 2003–2007, etc. Within each window, for a given q level, we first construct the optimal portfolio that minimizes $\mathrm{DR}_{\boldsymbol{w},q}$ by a numerical search. This is achieved by assigning weights to the 10 stocks on a grid spanning the set Σ^{10}, evaluating $\mathrm{DR}_{\boldsymbol{w},q}$ at each grid point and



(a) Hill plots for all 10 stocks (b) DR against $\widehat{\Psi}$ estimated at various k

Figure 4: The impact of k on estimating the tail index α and the diversifcation ratio Ψ

taking the weights that corresponds to the minimum diversification ratio. Then we construct the optimal portfolio based on minimizing the estimated $\mathrm{DR}_{\boldsymbol{w},1}$ using the procedure laid out in Section 5.1.

The numerical search strategy gives a numerical optimal while our portfolio optimization strategy gives an approximation to that. To evaluate the difference between the two optimal portfolios, we use $||\boldsymbol{w}_q - \boldsymbol{w}^*||_1/10$. This distance indicates the average error made on the weight for one stock. We conduct this analysis for nine different windows and four different levels of q: 0.95, 0.975, 0.99 and 0.999.

In the estimation procedure, we need to select the intermediate sequence k. It should be chosen by balancing the bias and variance of the estimation. In Figure 4a the tail index estimated by the Hill estimator for each of the selected 10 stocks are plotted against various k. When $k = 4\%$, the estimations of the tail indices are the closest to each other. Hence, we choose k to be 4% for estimating α. In Figure 4b, by having α being estimated at 4%, the values of $\mathrm{DR}_{\boldsymbol{w},1}$ is plotted against various k at which the spectral measure Ψ is estimated. We choose k to be 10% for estimating the spectral measure $\widehat{\Psi}$. Moreover, since we only consider the loss, the estimator for $\eta_{\boldsymbol{w}}$ is slightly modified to

$$\widehat{\eta}_{\boldsymbol{w}} = \frac{1}{k}\sum \left(\boldsymbol{w}^T S_{\pi(j)}\right)^{\widehat{\alpha}}.$$

Table 2 shows the results on the error made using our optimization procedure. We observe that the distance is decreasing as q increases. This is in line with our theoretical result.

Next, we turn to analyzing the computation time for obtaining the optimal portfolio. For this analysis, we use only data in the most recent six windows and only consider $q = 0.95$. To show that the computational burden for the numerical search strategy largely depends on the number of stocks, we also perform the numerical search when using fewer stocks, namely the first 3, 5, and 8 stocks in Table 1. In contrast, we perform our portfolio optimization strategy always based on 10 stocks. The computation time of all the experiments run in

Table 2: Average error made on the weight for each stock

q	02-06	03-07	04-08	05-09	06-10	07-11	08-12	09-13	10-14
95%	0.1348	0.1091	0.125	0.0673	0.0868	0.0967	0.1447	0.1426	0.0941
97.50%	0.0838	0.0978	0.0967	0.0638	0.0795	0.0663	0.0985	0.0668	0.0802
99%	0.0837	0.0861	0.0858	0.0573	0.0636	0.0476	0.0834	0.0642	0.0731
99.9%	0.0442	0.0582	0.0688	0.0444	0.0397	0.0435	0.0435	0.0538	0.044

Note: Within in each five-year window, for a given q level, two portfolios are constructed. The numerical search strategy provides the first optimal portfolio that minimizes $\mathrm{DR}_{\boldsymbol{w},q}$. This is achieved by assigning weights to the 10 stocks on a grid spanning the set Σ^{10}, evaluating $\mathrm{DR}_{\boldsymbol{w},q}$ at each grid point and taking the weights that corresponds to the minimum diversification ratio. The second optimal portfolio minimizes the estimated $\mathrm{DR}_{\boldsymbol{w},1}$ using the procedure laid out in Section 5.1. The numbers reported are the distance calculated by $||\boldsymbol{w}_q - \boldsymbol{w}^*||_1/10$ between the two portfolios.

Matlab 2013a on a Thinkpad T430 (dual core, 2.6GHz CPU, 4GB of memory) computer is reported in Table 3. We observe that as the number of stocks increasing, the computation time for $\boldsymbol{w}_{95\%}$ increases significantly. On the contrary, our portfolio optimization strategy for 10 stocks takes even less time than that using the numerical search for 3 stocks.

Table 3: Computation time

Strategy		05-09	06-10	07-11	08-12	09-13	10-14
Numerical search	3 Stocks	0.350s	0.310s	0.261s	0.249s	0.231s	0.235s
Numerical search	5 Stocks	0.483s	0.402s	0.417s	0.391s	0.570s	0.612s
Numerical search	8 Stocks	1.226s	1.265s	1.594s	0.861s	1.463s	1.397s
Numerical search	10 Stocks	2.418s	2.799s	3.673s	2.022s	2.016s	2.383s
Minimizing $\mathrm{DR}_{\boldsymbol{w},1}$	10 Stocks	0.218s	0.189s	0.164s	0.175s	0.304s	0.166s

Note: Within each five-year window, the numerical search strategy is performed for minimizing the DR with $q = 0.95$ based on 3, 5, 8 and 10 stocks. The computation time are reported in the first four rows. The last row reports the computation time when performing the portfolio optimization strategy minimizing $\mathrm{DR}_{\boldsymbol{w},1}$ based on 10 stocks.

Finally, we perform an out-of-sample analysis comparing our portfolio optimization strategy with those in the literature. Within each five-year window, we perform our strategy to construct the optimal portfolio based on the 10 selected stocks in Table 1. Then we hold this portfolio for one year, and calculate the diversification ratio at 95% and the 95% VaR using the one-year out-of-sample data. We focus on $q = 95\%$ here because one-year loss data (roughly 250 daily observations) do not permit an accurate estimation of tail risk measures with a higher probability level. With a similar setup, we also apply the numerical search

Figure 5: Out-of-sample diversification ratio. Within each five-year window, the optimal portfolio based on the 10 selected stocks in Table 1 is constructed by minimizing $\mathrm{DR}_{\boldsymbol{w},1}$. These weights are held for one year. The diversification ratio at 95% is reported using the one-year out-of-sample data and named as DR(Limit) in the figure. The same procedures are repeated for five other strategies, the numerical search strategy for minimizing $\mathrm{DR}_{\boldsymbol{w},95\%}$ (DR(NS)), global minimum variance (GMV; see, e.g. Merton [37]), the MDP, the ERI, and equal weight strategy (Equal).

strategy laid out in the first empirical study which minimizes the $\mathrm{DR}_{\boldsymbol{w},95\%}$ within each five-year window, and evaluates the out-of-sample performance of this strategy. In addition, we apply four other strategies as competitors for out-of-sample performance, namely, the ERI, the MDP, global minimum variance (see, e.g. Merton [37]), and lastly a simple equal weight strategy.

Figure 5 shows the results on the out-of-sample diversification ratios. Our strategy produces consistently the lowest diversification ratio only except in 2009, where our strategy yields a diversification ratio slightly above that derived from the MDP, and in 2010 slightly higher than that derived from the numerical research strategy. This shows that our strategy is stable to achieve the tail diversification benefit in the out-of-sample experiments.

Figure 6 shows the results on the out-of-sample VaR. Our portfolio optimization strategy produces the lowest VaR in 2007 and 2008, but not in the other years. Nevertheless, the VaR of the optimal portfolio from our strategy is never largely above ERI, which minimizes VaR among the six strategies. Furthermore, it matters the most to get an optimal portfolio with the lowest risk in the period ahead of the crisis. Therefore, we conclude that our strategy also achieves good out-of-sample performance in terms of low portfolio risk, especially during the crisis period.

From all three empirical studies, we conclude that the computation burden of our portfolio optimization strategy is much lower than the numerical search. Although there is a moderate discrepancy between the optimal portfolios obtained from our limit DR optimiza-

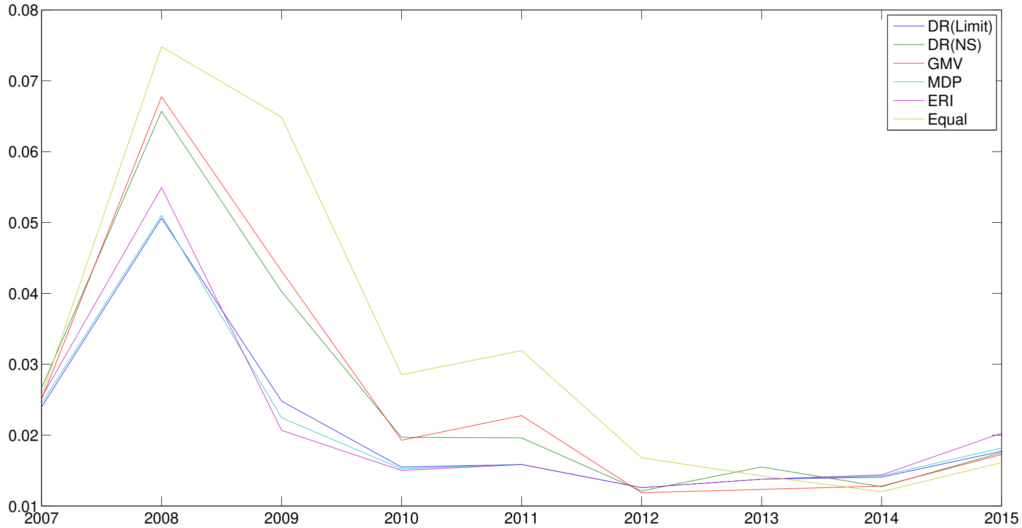


Figure 6: Comparison of portfolio risks. Within each five-year window, the optimal portfolio based on the 10 selected stocks in Table 1 is constructed by minimizing $\mathrm{DR}_{\boldsymbol{w},1}$. These weights are held for one year. The 95% VaR is reported using the one-year out-of-sample data and named as DR(Limit) in the figure. The same procedures are repeated for five other strategies, the numerical search strategy for minimizing $\mathrm{DR}_{\boldsymbol{w},95\%}$ (DR(NS)), global minimum variance (GMV; see, e.g. Merton [37]), the MDP, the ERI, and equal weight strategy (Equal).

tion strategy and the numerical search strategy, it turns out in the out-of-sample analysis that our strategy outperforms. It is therefore worth bearing the errors on the weights while using the faster and better performed algorithm derived from our limit DR optimization strategy.

6 Conclusion

This paper aims at constructing optimal portfolios by extracting the most diversification benefit, measured by the DR measure based on the VaR. Practically, risk manager is interested in an optimal portfolio weights $\boldsymbol{w}_q = \underset{\boldsymbol{w}\in\Sigma^d}{\arg\min}\ \mathrm{DR}_{\boldsymbol{w},q}$. Recognizing it was computationally intensive to solve this problem directly, we proposed to approximate the optimal portfolio by seeking $\boldsymbol{w}^* = \underset{\boldsymbol{w}\in\Sigma^d}{\arg\min}\ \mathrm{DR}_{\boldsymbol{w},1}$. When the underlying loss vector $\boldsymbol{X}$ followed MRV, we theoretically shown that $\lim_{q\uparrow 1} \boldsymbol{w}_q = \boldsymbol{w}^*$, in which the convergence ensures that one may use $\boldsymbol{w}^*$ as an approximation to $\boldsymbol{w}_q$ with a finite level q close to 1. Moreover, for a special case of MRV, the FGM copula, we explicitly determined the distance between $\boldsymbol{w}_q$ and $\boldsymbol{w}^*$. Numerically, through 2-d and 3-d Student's t-distributions, we examined the theoretical results that $\boldsymbol{w}_q$ indeed converges to $\boldsymbol{w}^*$ as q getting close to 1. Using observed stock data, we empirically examined the out-of-sample performance of our optimal portfolio and compared it with other portfolio optimization strategies such as optimizing ERI, MDP, global minimum variance and equal weight strategies. Our DR strategy has much lower computation burden than the numerical search and it outperforms other strategies during the crisis period by producing the lowest loss. Possible future work includes adding the profit component to the DR strategy, which only takes the downside risk into consideration in the current form.

A Appendix

In this section, we first prove Theorem 2.1, which is the key and the most difficult part in the proof of Theorem 2.2, in two steps as Sections A.1 and A.2. Then the very last section contains all the proofs of lemmas from previous sections.

A.1 Uniform convergence in Radon measures

Define a family of mappings from $A_1 = \left\{ \boldsymbol{x} \in \mathbb{R}_+^d : \|\boldsymbol{x}\|_1 > 1 \right\}$ to $\mathbb{R}_+$ as

$$M = \left\{ f_{\boldsymbol{w}}(\boldsymbol{x}) = \frac{1}{1 + \boldsymbol{w}^T \boldsymbol{x}} : \boldsymbol{w} \in \Sigma^d,\, \boldsymbol{x} \in A_1 \right\}. \tag{A.1}$$

Note that the construction of the mappings in M is not unique. Let $A_{\boldsymbol{w},1}$ denote the events where the portfolio loss $\boldsymbol{w}^T \boldsymbol{X}$ exceeds 1, namely for $\boldsymbol{w} \in \Sigma^d$,

$$A_{\boldsymbol{w},1} = \left\{ \boldsymbol{x} \in \mathbb{R}_+^d : \boldsymbol{w}^T \boldsymbol{x} > 1, \right\}.$$

Theorem A.1 *If $\boldsymbol{X} \in \mathrm{MRV}_\alpha(\Psi)$ with $\alpha > 0$, then*

$$\lim_{t \to \infty} \sup_{\boldsymbol{w} \in \Sigma^d} \left| \nu_t \left(A_{\boldsymbol{w},1} \right) - \nu \left(A_{\boldsymbol{w},1} \right) \right| = 0, \tag{A.2}$$

where ν_t and ν are defined in (2.2).

Proof. Since $A_{\boldsymbol{w},1} \in \mathcal{B}\left(A_1\right)$, by (2.2) we have that $\nu_t \left(A_{\boldsymbol{w},1} \right)$ converges weakly to $\nu \left(A_{\boldsymbol{w},1} \right)$. To further show the uniform convergence, we apply Theorem 3.4 of Rao [42]. That is we need to verify the following three conditions. (1) The mappings in M defined in (A.1) are continuous mappings from a separable metric space to $\mathbb{R}_+$. (2) The family M is relative compact; that is every sequence in M on a compact subset of A_1 has a subsequence that converges uniformly. (3) $v f_{\boldsymbol{w}}^{-1}$ has a continuous marginal distribution for each $f_{\boldsymbol{w}} \in M$, where $v f_{\boldsymbol{w}}^{-1}$ is a measure on $\mathcal{B}(\mathbb{R}_+)$ such that $v f_{\boldsymbol{w}}^{-1}(E) = v \left(f_{\boldsymbol{w}}^{-1}(E) \right)$ for any $E \in \mathcal{B}(\mathbb{R}_+)$. Next, we prove them separately.

(1) By Theorem 1.5 of Lindskog [30], there exists a metric $\overline{\rho}$ such that $(A_1, \overline{\rho})$ is a locally compact, complete and separable metric space. It is easy to see that each $f_{\boldsymbol{w}} \in M$ is continuous.

(2) Note that for $\boldsymbol{x}, \boldsymbol{y} \in A_1$, we have $\boldsymbol{w}^T\boldsymbol{x}, \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{y} > \mathbf{0}$. Then, by Cauchy-Schwarz inequality,

$$|f_{\boldsymbol{w}}(\boldsymbol{x}) - f_{\boldsymbol{w}}(\boldsymbol{y})| = \left| \frac{\boldsymbol{w}^T(\boldsymbol{x}-\boldsymbol{y})}{(1+\boldsymbol{w}^T\boldsymbol{x})(1+\boldsymbol{w}^T\boldsymbol{y})} \right| \leq \sqrt{d}\,\|\boldsymbol{x}-\boldsymbol{y}\|_2 .$$

For arbitrary $\varepsilon > 0$, we can choose $\delta < \varepsilon/\sqrt{d}$, which is independent of f, $\boldsymbol{x}$ and $\boldsymbol{y}$, such that when $\|\boldsymbol{x}-\boldsymbol{y}\|_2 < \delta$, we have $|f_{\boldsymbol{w}}(\boldsymbol{x}) - f_{\boldsymbol{w}}(\boldsymbol{y})| < \varepsilon$. This shows that M is equicontinuous at each $\boldsymbol{x} \in A_1$. Moreover, M is uniformly bounded as for each $\boldsymbol{x} \in A_1$,

$$\sup_{f_{\boldsymbol{w}} \in M} \{f_{\boldsymbol{w}}(\boldsymbol{x})\} = \sup_{\boldsymbol{w} \in \Sigma^d} \left\{ \frac{1}{1+\boldsymbol{w}^T\boldsymbol{x}} \right\} < \frac{1}{2}.$$

Therefore, from the Arzelà-Ascoli theorem, we know M is relatively compact.

(3) From (2), $f_{\boldsymbol{w}} < \frac{1}{2}$ for any $f_{\boldsymbol{w}} \in M$. Then for any $0 < y < 1/2$, we have

$$\begin{aligned} v f_{\boldsymbol{w}}^{-1}((0,y)) &= \int_{\Sigma^d}\int_{\mathbb{R}_+} 1_{\left\{r\boldsymbol{w}^T\boldsymbol{s} > \frac{1}{y}-1\right\}} \rho_\alpha(dr)\Psi(d\boldsymbol{s}) \\ &= \left(\frac{1}{y}-1\right)^{-\alpha} \int_{\Sigma^d} \left(\boldsymbol{w}^T\boldsymbol{s}\right)^\alpha \Psi(d\boldsymbol{s}), \end{aligned}$$

which is obviously continuous for any $0 < y < 1/2$. Furthermore, by definition we have $\nu(A_1) = 1$.

So far, we have verified the three conditions. By the weak convergence in (2.2) and Theorem 3.4 of Rao [42], we obtain

$$\lim_{t\to\infty} \sup_{\boldsymbol{w}\in\Sigma^d} |\nu_t(A_{\boldsymbol{w},1}) - \nu(A_{\boldsymbol{w},1})| = 0,$$

where the supremum is taken over all sets $A_{\boldsymbol{w},1}$ of the form $A_{\boldsymbol{w},1} = \left\{\boldsymbol{x} \in \mathbb{R}_+^d : f_{\boldsymbol{w}}(x) < \frac{1}{2}\right\} = \left\{\boldsymbol{x} \in \mathbb{R}_+^d : \boldsymbol{w}^T\boldsymbol{x} > 1\right\}$ with $\boldsymbol{w} \in \Sigma^d$. ■

Next corollary is a natural rewriting of relation (A.2). It yields a uniform convergence of the ratio $\Pr\left(\boldsymbol{w}^T\boldsymbol{X} > t\right)/\Pr\left(\|\boldsymbol{X}\|_1 > t\right)$ to $\eta_{\boldsymbol{w}}$. However, only the weak convergence of it is known in the literature.

Corollary A.1

$$\lim_{t\to\infty} \sup_{\boldsymbol{w}\in\Sigma^d} \left| \frac{\Pr\left(\boldsymbol{w}^T\boldsymbol{X} > t\right)}{\Pr\left(\|\boldsymbol{X}\|_1 > t\right)} - \eta_{\boldsymbol{w}} \right| = 0, \tag{A.3}$$

where

$$\eta_{\boldsymbol{w}} = \int_{\Sigma^d} (\boldsymbol{w}^T\boldsymbol{s})^\alpha \Psi(d\boldsymbol{s}).$$

Further, the mapping $\boldsymbol{w} \mapsto \eta_{\boldsymbol{w}} : \Sigma^d \to (0,1)$ *is uniform continuous.*

Proof. First note that $A_{\boldsymbol{w},t} = tA_{\boldsymbol{w},1}$. Since $A_{\boldsymbol{w},1} \subset \mathcal{B}(A_1)$ for $\boldsymbol{w} \in \Sigma^d$, we have that

$$\nu_t(A_{\boldsymbol{w},1}) = \frac{\Pr\left(\frac{\boldsymbol{X}}{t} \in A_{\boldsymbol{w},1}\right)}{\Pr(\|\boldsymbol{X}\|_1 > t)} = \frac{\Pr(\boldsymbol{X} \in A_{\boldsymbol{w},t})}{\Pr(\|\boldsymbol{X}\|_1 > t)}.$$

Moreover $\nu(A_{\boldsymbol{w},1})$ is actually

$$\nu(A_{\boldsymbol{w},1}) = \int_{\Sigma^d} (\boldsymbol{w}^T\boldsymbol{s})^\alpha \Psi(d\boldsymbol{s}) = \eta_{\boldsymbol{w}}.$$

The desired result (A.3) then follows. Lastly, since $\eta_{\boldsymbol{w}}$ is continuous on the compact set Σ^d, it implies the uniform continuity of $\eta_{\boldsymbol{w}}$ on Σ^d. ■

A.2 Uniform convergence in quantiles

In order to show that the convergence in (2.5) is indeed uniform, we first prepare a key lemma. For notational simplicity, we denote

$$l(\boldsymbol{w}, q) := \frac{\mathrm{VaR}_q(\boldsymbol{w}^T\boldsymbol{X})}{\mathrm{VaR}_q(\|\boldsymbol{X}\|_1)} = \frac{F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}(q)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q)}, \tag{A.4}$$

where $F_{\boldsymbol{w}^T\boldsymbol{X}}$ is the distribution function of $\boldsymbol{w}^T\boldsymbol{X}$ and $F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}(q) = \mathrm{VaR}_q(\boldsymbol{w}^T\boldsymbol{X})$.

Lemma A.1 *Suppose the nonnegative random vector $\boldsymbol{X}$ is continuously distributed with a positive joint density function. Further assume that $\boldsymbol{X} \in \mathrm{MRV}_\alpha(\Psi)$ with $\alpha > 0$. Given $\boldsymbol{w} \in \Sigma^d$, for any $\varepsilon > 0$ there exist $0 < \tilde{q} < 1$ and δ such that for all $\tilde{q} < q < 1$ and $\boldsymbol{z} \in \Sigma^d$ satisfying $\|\boldsymbol{w} - \boldsymbol{z}\| < \delta$, we have*

$$|l(\boldsymbol{w}, q) - l(\boldsymbol{z}, q)| < \varepsilon. \tag{A.5}$$

Proof. We start by showing that for any $\varepsilon_1 > 0$, there exist $t_0(\varepsilon_1)$ and $\delta(\varepsilon_1)$ such that for all $t > t_0$ and all $\boldsymbol{w}, \boldsymbol{z} \in \Sigma^d$ with $\|\boldsymbol{w} - \boldsymbol{z}\| < \delta$, we have

$$\left|\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t) - \overline{F}_{\boldsymbol{z}^T\boldsymbol{X}}(t)\right| < \varepsilon_1 \overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t). \tag{A.6}$$

Note that $\eta_{\boldsymbol{w}} > 0$ for every $\boldsymbol{w} \in \Sigma^d$. Since Σ^d is compact, there exists $\underline{\eta} > 0$ such that $\eta_{\boldsymbol{w}} > \underline{\eta} > 0$. Further, $\eta_{\boldsymbol{w}}$ is uniform continuous on Σ^d by Corollary A.1. That is, for any $\varepsilon_1 > 0$, there exists $\delta(\varepsilon_1)$ such that for all $\boldsymbol{w}, \boldsymbol{z} \in \Sigma^d$ with $\|\boldsymbol{w} - \boldsymbol{z}\| < \delta$, we have

$$|\eta_{\boldsymbol{w}} - \eta_{\boldsymbol{z}}| < \frac{\underline{\eta}}{6}\varepsilon_1. \tag{A.7}$$

Again, by Corollary A.1, there exists $t_0(\varepsilon_1)$ such that for all $t > t_0$ and all $\boldsymbol{w} \in \Sigma^d$

$$\left|\frac{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t)}{\overline{F}_{\|\boldsymbol{X}\|_1}(t)} - \eta_{\boldsymbol{w}}\right| < \frac{\underline{\eta}}{6}\varepsilon_1 \wedge \frac{\underline{\eta}}{2}, \tag{A.8}$$

which implies that

$$\frac{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t)}{\overline{F}_{\|\boldsymbol{X}\|_1}(t)} > \eta_{\boldsymbol{w}} - \frac{\eta}{2} > \frac{\eta}{2}. \tag{A.9}$$

Then, combing (A.7), (A.8) and (A.9), for all $t > t_0$ and all $\boldsymbol{w}, \boldsymbol{z} \in \Sigma^d$ with $\|\boldsymbol{w} - \boldsymbol{z}\| < \delta$,

$$\begin{aligned}
\left|\frac{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t) - \overline{F}_{\boldsymbol{z}^T\boldsymbol{X}}(t)}{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t)}\right| &= \left|\frac{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t) - \overline{F}_{\boldsymbol{z}^T\boldsymbol{X}}(t)}{\overline{F}_{\|\boldsymbol{X}\|_1}(t)}\right| \cdot \frac{\overline{F}_{\|\boldsymbol{X}\|_1}(t)}{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t)} \\
&\leq \left(\left|\frac{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t)}{\overline{F}_{\|\boldsymbol{X}\|_1}(t)} - \eta_{\boldsymbol{w}}\right| + |\eta_{\boldsymbol{w}} - \eta_{\boldsymbol{z}}| + \left|\eta_{\boldsymbol{z}} - \frac{\overline{F}_{\boldsymbol{z}^T\boldsymbol{X}}(t)}{\overline{F}_{\|\boldsymbol{X}\|_1}(t)}\right|\right) \cdot \frac{\overline{F}_{\|\boldsymbol{X}\|_1}(t)}{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t)} \\
&< \left(\frac{\eta}{6}\varepsilon_1 + \frac{\eta}{6}\varepsilon_1 + \frac{\eta}{6}\varepsilon_1\right)\frac{2}{\eta} = \varepsilon_1,
\end{aligned}$$

which yields (A.6).

Next, for the chosen $t_0(\varepsilon_1)$, denote $q_0 = \sup_{\boldsymbol{z}\in\Sigma^d} F_{\boldsymbol{z}^T\boldsymbol{X}}(t_0(\varepsilon_1))$. Then for any $q_0 < q < 1$ and all $\boldsymbol{z} \in \Sigma^d$,

$$F^{\leftarrow}_{\boldsymbol{z}^T\boldsymbol{X}}(q) \geq F^{\leftarrow}_{\boldsymbol{z}^T\boldsymbol{X}}(q_0) \geq t_0. \tag{A.10}$$

By (A.6) and (A.10), it leads to that for all $q > q_0$ and $\|\boldsymbol{w} - \boldsymbol{z}\| < \delta$,

$$\left|\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(F^{\leftarrow}_{\boldsymbol{z}^T\boldsymbol{X}}(q)) - (1-q)\right| < \varepsilon_1(1-q).$$

By the monotonicity of $F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}(q)$, we obtain

$$F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}\left(q(1+\varepsilon_1) - \varepsilon_1\right) < F^{\leftarrow}_{\boldsymbol{z}^T\boldsymbol{X}}(q) < F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}\left(q(1-\varepsilon_1) + \varepsilon_1\right). \tag{A.11}$$

Finally we handle $|l(\boldsymbol{w}, q) - l(\boldsymbol{z}, q)|$ in (A.5). We only discuss the upper bound of $l(\boldsymbol{w}, q) - l(\boldsymbol{z}, q)$ in this step as the lower bound can be derived in a similar way. By (A.11),

$$\begin{aligned}
&l(\boldsymbol{w}, q) - l(\boldsymbol{z}, q) \\
\leq& \frac{F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}(q)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q)} - \frac{F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}\left(q(1+\varepsilon_1) - \varepsilon_1\right)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q)} \\
=& \left(\frac{F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}(q)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q)} - \frac{F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}\left(q(1+\varepsilon_1) - \varepsilon_1\right)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q(1+\varepsilon_1) - \varepsilon_1)}\right) + \frac{F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}\left(q(1+\varepsilon_1) - \varepsilon_1\right)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q(1+\varepsilon_1) - \varepsilon_1)}\left(1 - \frac{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}\left(q(1+\varepsilon_1) - \varepsilon_1\right)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q)}\right) \\
:=& I_1 + I_2,
\end{aligned}$$

where

$$I_1 = l(\boldsymbol{w}, q) - l(\boldsymbol{w}, q(1+\varepsilon_1) - \varepsilon_1),$$

and

$$I_2 = l(\boldsymbol{w}, q(1+\varepsilon_1) - \varepsilon_1)\left(1 - \frac{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}\left(q(1+\varepsilon_1) - \varepsilon_1\right)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q)}\right).$$

We show that $I_1 < \varepsilon/2$ and $I_2 < \varepsilon/2$.

For I_1, note the random vector $\boldsymbol{X}$ is continuously distributed with a positive joint density function. By using the change of variables, the density functions for random variables $\|\boldsymbol{X}\|_1$ and $\boldsymbol{w}^T\boldsymbol{X}$ can be shown to be positive as well, which implies that $F_{\|\boldsymbol{X}\|_1}(t)$ and $F_{\boldsymbol{w}^T\boldsymbol{X}}(t)$ are strictly increasing in t. By Proposition 1 (7) in Embrechts and Hofert [20], we have that $F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q)$ and $F^{\leftarrow}_{\boldsymbol{w}^T\boldsymbol{X}}(q)$ are both continuous in q for any fixed $\boldsymbol{w}$. Moreover, from (2.5), $l(\boldsymbol{w},1)$ can be continuously defined as $\eta_{\boldsymbol{w}}^{1/\alpha}$. Thus, given $\boldsymbol{w}$, $l(\boldsymbol{w},q)$ is uniformly continuous in q when $q\in[1/2,1]$. That is, there exists $\lambda_1(\boldsymbol{w},\varepsilon)$ such that when $1/2\le p,q\le 1$ and $|p-q|<\lambda_1$, we have

$$|l(\boldsymbol{w},q)-l(\boldsymbol{w},p)|<\frac{\varepsilon}{2}.$$

Then, for $|q-(q(1+\varepsilon_1)-\varepsilon_1)|<\lambda_1$ or $q>1-\lambda_1/\varepsilon_1$, we obtain that $I_1<\varepsilon/2$.

For I_2, we first show that $l(\boldsymbol{w},q(1+\varepsilon_1)-\varepsilon_1)$ is bounded. Since $\lim_{q\to 1} l(\boldsymbol{w},q)=l(\boldsymbol{w},1)$, there exists $\lambda_2(\boldsymbol{w})$ such that when $1-(q(1+\varepsilon_1)-\varepsilon_1)<\lambda_2$ or $1>q>1-\lambda_2/(1+\varepsilon_1)$, we have

$$|l(\boldsymbol{w},q(1+\varepsilon_1)-\varepsilon_1)-l(\boldsymbol{w},1)|<1.$$

Denote $M_0=\sup_{\boldsymbol{w}\in\Sigma^d} l(\boldsymbol{w},1)$. We obtain

$$l(\boldsymbol{w},q(1+\varepsilon_1)-\varepsilon_1)<M_0+1,\qquad \text{for } q>1-\lambda_2/(1+\varepsilon_1). \tag{A.12}$$

Finally, we consider $1-\frac{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q(1+\varepsilon_1)-\varepsilon_1)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q)}$ in term I_2. It is known that if $\boldsymbol{X}\in\mathrm{MRV}_\alpha(\Psi)$ then $\|\boldsymbol{X}\|_1\in\mathrm{RV}_{-\alpha}$; e.g. see Basrak et al. [2]. Thus,

$$\lim_{q\to 1}\frac{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q(1+\varepsilon_1)-\varepsilon_1)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q)}=(1+\varepsilon_1)^{1/\alpha}.$$

By Proposition B.1.10 of de Haan and Ferreira [10], there exists $q_3(\varepsilon)<1$ such that for all $q>q_3(\varepsilon)$ we have

$$\left|\frac{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q(1+\varepsilon_1)-\varepsilon_1)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q)}-(1+\varepsilon_1)^{1/\alpha}\right|<\frac{1}{M_0+1}\frac{\varepsilon}{4}.$$

Moreover, when ε_1 is so chosen that

$$\left|1-(1+\varepsilon_1)^{1/\alpha}\right|<\frac{1}{M_0+1}\frac{\varepsilon}{4}, \tag{A.13}$$

it leads to that

$$\left|\frac{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q(1+\varepsilon_1)-\varepsilon_1)}{F^{\leftarrow}_{\|\boldsymbol{X}\|_1}(q)}-1\right|<\frac{\varepsilon}{2\,(M_0+1)},\qquad \text{for } q>q_3(\varepsilon). \tag{A.14}$$

Combining (A.12) and (A.14), $I_2<\varepsilon/2$ for $q>1-\lambda_2/(1+\varepsilon_1)\vee q_3(\varepsilon)$.

To sum up, given $\boldsymbol{w}$, for arbitrary $\varepsilon > 0$, and for any ε_1 so chosen that (A.13) holds, there exist δ, $q_0, \lambda_1, \lambda_2$, and q_3 such that for all $\boldsymbol{z} \in \Sigma^d$ with $\|\boldsymbol{w} - \boldsymbol{z}\| < \delta$ and for all q satisfying that

$$1 > q > q_0 \vee \left(1 - \frac{\lambda_1}{\varepsilon_1}\right) \vee \left(1 - \frac{\lambda_2}{1+\varepsilon_1}\right) \vee q_3,$$

we have $l(\boldsymbol{w}, q) - l(\boldsymbol{z}, q) < \varepsilon$. The other side of the inequality can be derived similarly. ■

Now we are ready to show that the convergence in (2.5) is uniform.

Theorem A.2 *Suppose the nonnegative random vector $\boldsymbol{X}$ is continuously distributed with a positive joint density function. Further assume that $\boldsymbol{X} \in \mathrm{MRV}_{-\alpha}(\Psi)$ with $\alpha > 0$. Then*

$$\lim_{q \uparrow 1} \sup_{\boldsymbol{w} \in \Sigma^d} \left| \frac{\mathrm{VaR}_q(\boldsymbol{w}^T \boldsymbol{X})}{\mathrm{VaR}_q(\|\boldsymbol{X}\|_1)} - \eta_{\boldsymbol{w}}^{1/\alpha} \right| = 0. \tag{A.15}$$

Proof. Consider the decomposition for some $\boldsymbol{v} \in \Sigma^d$

$$\left|l(\boldsymbol{w}, q) - \eta_{\boldsymbol{w}}^{1/\alpha}\right| \le |l(\boldsymbol{w}, q) - l(\boldsymbol{v}, q)| + \left|l(\boldsymbol{v}, q) - \eta_{\boldsymbol{v}}^{1/\alpha}\right| + \left|\eta_{\boldsymbol{v}}^{1/\alpha} - \eta_{\boldsymbol{w}}^{1/\alpha}\right|, \tag{A.16}$$

where $l(\boldsymbol{w}, q)$ is defined as in (A.4). By properly choosing $\boldsymbol{v}$, if the three terms can be shown to be arbitrarily small for any $\boldsymbol{w} \in \Sigma^d$ as q close to 1, then (A.15) is proved. In the following we show how $\boldsymbol{v}$ can be determined.

By Lemma A.1 and the uniform continuity of $\eta_{\boldsymbol{w}}$ on Σ^d, for any $\varepsilon > 0$, there exist $\delta(\boldsymbol{w}) > 0$ and $0 < \tilde{q}(\boldsymbol{w}) < 1$ such that for any $\boldsymbol{w}, \boldsymbol{z} \in \Sigma^d$ satisfying $\|\boldsymbol{w} - \boldsymbol{z}\| < \delta(\boldsymbol{w})$ and all $q \ge \tilde{q}(\boldsymbol{w})$, we have

$$|l(\boldsymbol{w}, q) - l(\boldsymbol{z}, q)| < \varepsilon. \tag{A.17}$$

and

$$\left|\eta_{\boldsymbol{w}}^{1/\alpha} - \eta_{\boldsymbol{z}}^{1/\alpha}\right| < \varepsilon. \tag{A.18}$$

That is, $\delta(\boldsymbol{w})$ is so chosen that both (A.17) and (A.18) hold. Now we are ready to determine $\boldsymbol{v}$ in (A.16) by constructing open coverings. Let $B_{\boldsymbol{w},\delta(\boldsymbol{w})}$ denote the open ball of $\boldsymbol{w}$; that is $B_{\boldsymbol{w},\delta(\boldsymbol{w})} = \{\boldsymbol{z} \in \Sigma^d : \|\boldsymbol{w} - \boldsymbol{z}\| < \delta(\boldsymbol{w})\}$. Then the collection of all the sets $B_{\boldsymbol{w},\delta(\boldsymbol{w})}$ for each $\boldsymbol{w}$ is an open cover of Σ^d. By the compactness, there exists a finite subcover denoted by $B_{\boldsymbol{w}_1,\delta(\boldsymbol{w}_1)}, \ldots, B_{\boldsymbol{w}_m,\delta(\boldsymbol{w}_m)}$. For each selected $\boldsymbol{w}_i$, by the limit relation in (2.5), there exists $0 < q_i < 1$ such that

$$\left|l(\boldsymbol{w}_i, q) - \eta_{\boldsymbol{w}_i}^{1/\alpha}\right| < \varepsilon,$$

for all $q_i \le q < 1$. Let $q^* = \max\{\tilde{q}(\boldsymbol{w}_1), \ldots, \tilde{q}(\boldsymbol{w}_m), q_1, \ldots, q_m\}$. For any $\boldsymbol{w} \in \Sigma^d$, one can find i such that $\boldsymbol{w} \in B_{\boldsymbol{w}_i,\delta(\boldsymbol{w}_i)}$, which means $\|\boldsymbol{w} - \boldsymbol{w}_i\| < \delta(\boldsymbol{w}_i)$. This $\boldsymbol{w}_i$ is the proper choice of $\boldsymbol{v}$ in (A.16) since each term on the right-hand side of (A.16) is smaller than ε for all $q^* \le q < 1$. This completes the proof. ■

Now we are ready to prove Theorem 2.1.

Proof of Theorem 2.1. Since the convergence $\lim_{q\uparrow 1} \mathrm{VaR}_q(X_i)/\mathrm{VaR}_q(\|\boldsymbol{X}\|_1) = \eta_{e_i}^{1/\alpha}$ is independent of $\boldsymbol{w}$, applying Theorem A.2 to the rewriting in (2.4) we obtain the desired result. ■

A.3 Proofs of lemmas

Lastly, we present the proofs of lemmas from previous sections.

Proof of Lemma 2.2. To prove $\mathrm{DR}_{\boldsymbol{w},q} \overset{\text{unif}}{\longrightarrow} \mathrm{DR}_{\boldsymbol{w},1}$, we need to show for any given $\varepsilon > 0$, there exists a number q_0 such that $|\mathrm{DR}_{\boldsymbol{w},q} - \mathrm{DR}_{\boldsymbol{w},1}| < \varepsilon$ for every $q > q_0$ and for every $\boldsymbol{w}$ in Σ^d. Note the rewriting

$$|\mathrm{DR}_{\boldsymbol{w},q} - \mathrm{DR}_{\boldsymbol{w},1}| = \left| \frac{\boldsymbol{w}^T \boldsymbol{\mu} \left(\sum_{i=1}^d w_i \left\| B^T \boldsymbol{e}_i \right\|_2 - \left\| B^T \boldsymbol{w} \right\|_2 \right)}{\left(\boldsymbol{w}^T \boldsymbol{\mu} + \sum_{i=1}^d w_i \left\| B^T \boldsymbol{e}_i \right\|_2 F_Z^{\leftarrow}(q) \right) \sum_{i=1}^d w_i \left\| B^T \boldsymbol{e}_i \right\|_2} \right|.$$

For every $\boldsymbol{w} \in \Sigma^d$, since $\|\boldsymbol{\mu}\|_1 < \infty$, there exists $N_1 > 0$ such that $\boldsymbol{w}^T \boldsymbol{\mu} < \|\mu\|_1 < N_1$. Since $\|B\|_2 < \infty$, there exists $N_2, N_3 > 0$ such that

$$0 < \sum_{i=1}^d w_i \left\| e_i^T B \right\|_2 < \sum_{i=1}^d \left\| B^T \boldsymbol{e}_i \right\|_2 < d \|B\|_2 < N_2,$$

and

$$\left\| B^T \boldsymbol{w} \right\|_2 < \|B\|_2 < N_3.$$

Since Y is unbounded, there exists $0 < q_0 < 1$ such that

$$F_Z^{\leftarrow}(q) > \frac{N_1 (N_2 + N_3)}{N_2^2 \varepsilon} - \frac{N_1}{N_2},$$

for every $q > q_0$. Combining the above analysis, the desired result $|\mathrm{DR}_{\boldsymbol{w},q} - \mathrm{DR}_{\boldsymbol{w},1}| < \varepsilon$ for every $q > q_0$ and for every $\boldsymbol{w}$ in Σ^d follows.

Next, we show that $\mathrm{DR}_{\boldsymbol{w},1}$ is continuous. For $\boldsymbol{w}, \boldsymbol{v} \in \Sigma^d$, we have that

$$\begin{aligned}
&|\mathrm{DR}_{\boldsymbol{w},1} - \mathrm{DR}_{\boldsymbol{v},1}| \\
\leq &\left| \frac{\left\| B^T (\boldsymbol{w} - \boldsymbol{v}) \right\|_2 \sum_{i=1}^d v_i \left\| B^T \boldsymbol{e}_i \right\|_2 + \left\| B^T \boldsymbol{v} \right\|_2 \left(\sum_{i=1}^d v_i \left\| B^T \boldsymbol{e}_i \right\|_2 - \sum_{i=1}^d w_i \left\| B^T \boldsymbol{e}_i \right\|_2 \right)}{\left(\sum_{i=1}^d w_i \|B^T \boldsymbol{e}_i\|_2 \right) \left(\sum_{i=1}^d v_i \|B^T \boldsymbol{e}_i\|_2 \right)} \right| \\
\leq &\frac{\|B\|_2 \|\boldsymbol{w} - \boldsymbol{v}\|_1 \sum_{i=1}^d v_i \left\| B^T \boldsymbol{e}_i \right\|_2}{\left(\sum_{i=1}^d w_i \|B^T \boldsymbol{e}_i\|_2 \right) \left(\sum_{i=1}^d v_i \|B^T \boldsymbol{e}_i\|_2 \right)} + \frac{\left\| B^T \boldsymbol{v} \right\|_2 \|\boldsymbol{w} - \boldsymbol{v}\|_1 \max_{1 \leq i \leq d} \left\| B^T \boldsymbol{e}_i \right\|_2}{\left(\sum_{i=1}^d w_i \|B^T \boldsymbol{e}_i\|_2 \right) \left(\sum_{i=1}^d v_i \|B^T \boldsymbol{e}_i\|_2 \right)} \\
= &\|\boldsymbol{w} - \boldsymbol{v}\|_1 \frac{\|B\|_2 \sum_{i=1}^d v_i \left\| B^T \boldsymbol{e}_i \right\|_2 + \left\| B^T \boldsymbol{v} \right\|_2 \max_{1 \leq i \leq d} \left\| B^T \boldsymbol{e}_i \right\|_2}{\left(\sum_{i=1}^d w_i \|B^T \boldsymbol{e}_i\|_2 \right) \left(\sum_{i=1}^d v_i \|B^T \boldsymbol{e}_i\|_2 \right)}.
\end{aligned}$$

Since $\|B\|_2 < \infty$ and BB^T is positive definite, the fraction in the last step is bounded. Therefore for fixed $\boldsymbol{w}$, when $\|\boldsymbol{w} - \boldsymbol{v}\|_1$ is small enough, we have $|\mathrm{DR}_{\boldsymbol{w},1} - \mathrm{DR}_{\boldsymbol{v},1}| < \varepsilon$. This proves the mapping $\boldsymbol{w} \to \mathrm{DR}_{\boldsymbol{w},1}$ is continuous. ∎

Proof of Lemma 5.1. First note that

$$
\begin{aligned}
&\sup_{\boldsymbol{w}\in\Sigma^d}\left|\hat{\mathrm{DR}}_{\boldsymbol{w},1}-\mathrm{DR}_{\boldsymbol{w},1}\right| \\
&=\sup_{\boldsymbol{w}\in\Sigma^d}\left|\frac{\widehat{\eta}_{\boldsymbol{w}}^{1/\alpha}\sum_{i=1}^d w_i\eta_{\boldsymbol{e}_i}^{1/\alpha}-\eta_{\boldsymbol{w}}^{1/\alpha}\sum_{i=1}^d w_i\widehat{\eta}_{\boldsymbol{e}_i}^{1/\alpha}}{\left(\sum_{i=1}^d w_i\widehat{\eta}_{\boldsymbol{e}_i}^{1/\alpha}\right)\left(\sum_{i=1}^d w_i\eta_{\boldsymbol{e}_i}^{1/\alpha}\right)}\right| \\
&\le\sup_{\boldsymbol{w}\in\Sigma^d}\left|\frac{\left(\widehat{\eta}_{\boldsymbol{w}}^{1/\alpha}-\eta_{\boldsymbol{w}}^{1/\alpha}\right)\sum_{i=1}^d w_i\eta_{\boldsymbol{e}_i}^{1/\alpha}}{\left(\sum_{i=1}^d w_i\widehat{\eta}_{\boldsymbol{e}_i}^{1/\alpha}\right)\left(\sum_{i=1}^d w_i\eta_{\boldsymbol{e}_i}^{1/\alpha}\right)}\right|+\sup_{\boldsymbol{w}\in\Sigma^d}\left|\frac{\eta_{\boldsymbol{w}}^{1/\alpha}\left(\sum_{i=1}^d w_i\eta_{\boldsymbol{e}_i}^{1/\alpha}-\sum_{i=1}^d w_i\widehat{\eta}_{\boldsymbol{e}_i}^{1/\alpha}\right)}{\left(\sum_{i=1}^d w_i\widehat{\eta}_{\boldsymbol{e}_i}^{1/\alpha}\right)\left(\sum_{i=1}^d w_i\eta_{\boldsymbol{e}_i}^{1/\alpha}\right)}\right|.
\end{aligned}
\tag{A.19}
$$

Thus, to show that (A.19) converges to 0 almost surely, the key is the strong consistency of $\widehat{\eta}_{\boldsymbol{w}}$ uniformly in $\boldsymbol{w}$. This is ensured by Theorem 4.1(a) of Mainik and Rüchendorf [32] if

$$\lim_{q\uparrow 1}\sup_{\boldsymbol{w}\in\Sigma^d}|\Psi_q f_{\boldsymbol{w},\alpha}-\Psi f_{\boldsymbol{w},\alpha}|=0, \tag{A.20}$$

where Ψ_q is the conditional angular distribution of $S|F_R(R)>q$ for $q\in(0,1)$ and $f_{\boldsymbol{w},\alpha}(\boldsymbol{s})=(\boldsymbol{w}^T\boldsymbol{s})^\alpha$. Now we show that (A.20) holds under the current conditions. Note that for any $\boldsymbol{s}\in\Sigma^d$, we have

$$0<(\boldsymbol{w}^T\boldsymbol{s})^\alpha\le\boldsymbol{w}^T\boldsymbol{s}\le\boldsymbol{w}^T\mathbf{1}=1.$$

For any $\boldsymbol{s}_1,\boldsymbol{s}_2\in\Sigma^d$, it follows that

$$
\begin{aligned}
|f_{\boldsymbol{w},\alpha}(s_1)-f_{\boldsymbol{w},\alpha}(s_2)| &= |(\boldsymbol{w}^T\boldsymbol{s}_1)^\alpha-(\boldsymbol{w}^T\boldsymbol{s}_2)^\alpha| \\
&\le |(\boldsymbol{w}^T\boldsymbol{s}_1)-(\boldsymbol{w}^T\boldsymbol{s}_2)|d \\
&\le d|\boldsymbol{s}_1-\boldsymbol{s}_2|,
\end{aligned}
$$

where in the second step we used the polynomial expansion formula. This means that the function class $\left\{f_{\boldsymbol{w},\alpha}:\boldsymbol{w}\in\Sigma^d\right\}$ is uniformly Lipschitz for any $\alpha>1$. Then by Remark A.5 of Mainik and Rüchendorf [32], the uniform convergence in (A.20) holds. Hence, $\widehat{\eta}_{\boldsymbol{w}}$ converges to $\eta_{\boldsymbol{w}}$ uniformly in $\boldsymbol{w}\in\Sigma^d$ almost surely. Further, by the continuity of the mapping $\widehat{\eta}_{\boldsymbol{w}}\longmapsto\widehat{\eta}_{\boldsymbol{w}}^{1/\alpha}$, we have

$$\sup_{\boldsymbol{w}\in\Sigma^d}\left|\widehat{\eta}_{\boldsymbol{w}}^{1/\alpha}-\eta_{\boldsymbol{w}}^{1/\alpha}\right|\to 0,\qquad a.s.,$$

and

$$\sup_{\boldsymbol{w}\in\Sigma^d}\left|\sum_{i=1}^d w_i\eta_{\boldsymbol{e}_i}^{1/\alpha}-\sum_{i=1}^d w_i\widehat{\eta}_{\boldsymbol{e}_i}^{1/\alpha}\right|=\sup_{\boldsymbol{w}\in\Sigma^d}\left|\sum_{i=1}^d w_i\left(\eta_{\boldsymbol{e}_i}^{1/\alpha}-\widehat{\eta}_{\boldsymbol{e}_i}^{1/\alpha}\right)\right|\to 0,\qquad a.s.$$

Further notice that $\sum_{i=1}^d w_i\eta_{\boldsymbol{e}_i}^{1/\alpha}$ and $\sum_{i=1}^d w_i\widehat{\eta}_{\boldsymbol{e}_i}^{1/\alpha}$ are uniformly bounded away from 0 because both the empirical measure $\widehat{\Psi}$ and the limit measure Ψ are non-degenerated. Combining all these, we obtain that (A.19) converges to 0 almost surely, which yields the desired result. ■

Proof of Lemma 3.1. In this proof the limit is taken as $t \to \infty$. For $t > 0$, denote the region $S_t = \{(x_1, \ldots, x_d) \in \mathbb{R}_+^d : \sum_{i=1}^d w_i x_i \geq t\}$. We can split $\overline{F}_{\boldsymbol{w}^T \boldsymbol{X}}(t)$ as

$$\begin{aligned}\overline{F}_{\boldsymbol{w}^T \boldsymbol{X}}(t) &= \int_{S_t} \mathrm{d}\left(\prod_{k=1}^d G_k(x_k)\right) + \sum_{i<j} a_{i,j} \int_{S_t} \mathrm{d}\left((1 - G_i(x_i))(1 - G_j(x_j)) \prod_{k=1}^d G_k(x_k)\right) \\ &= I(t) + \sum_{i<j} a_{i,j} J_{i,j}(t),\end{aligned}$$

where $G_k(x) = G(x/w_k)$ for $k = 1, ..., d$. The term $I(t)$ can be understood as the survival distribution function of $w_1 X_1^* + \cdots + w_d X_d^*$, where $X_1^*, \ldots, X_d^*$ are i.i.d. with common distribution function G. For $I(t)$, it follows from Theorems 4.7 of Mao and Ng [34] that,

$$\frac{I(t)}{\overline{G}(t)} = \sum_{k=1}^d w_k^\alpha + \sum_{k=1}^d H_{-\alpha,\rho}\left(w_k^{-1}\right) A(t)(1 + o(1)) + \alpha t^{-1} \mu_G \sum_{k \neq l} w_k^\alpha w_l (1 + o(1)).$$

For $J_{i,j}(t)$'s, note that it suffices to study $J_{1,2}(t)$ by symmetry. Then we have

$$\begin{aligned}J_{1,2}(t) =& I(t) - \int_{S_t} \mathrm{d}\left(G_1^2(x_1) \prod_{k=2}^d G_k(x_k)\right) - \int_{S_t} \mathrm{d}\left(G_2^2(x_2) \prod_{k \neq 2} G_k(x_k)\right) \\ &+ \int_{S_t} \mathrm{d}\left(G_1^2(x_1) G_2^2(x_2) \prod_{k=3}^n G_k(x_k)\right) \\ =& I(t) - J_{1,2}^{(1)}(t) - J_{1,2}^{(2)}(t) + J_{1,2}^{(3)}(t).\end{aligned}$$

Note that $\overline{G}_k(x) = \overline{G}(x/w_k) \sim w_k^\alpha \overline{G}(t)$ and $\overline{G_1^2}(t)/\overline{G}_1(t) \to 2$. Since $\alpha \geq 1$, by regarding $G_1^2(\cdot)$ as a distribution function, Proposition 3.7 of Mao and Ng [34] leads to

$$\begin{aligned}&J_{1,2}^{(1)}(t) \\ &= (2w_1^\alpha + w_2^\alpha + \cdots + w_d^\alpha)\,\overline{G}(t) + o\left(\overline{G}(t) A(t)\right) \\ &+ \alpha t^{-1} \left(2 w_1^\alpha \sum_{k=2}^d w_k \mu_G + w_1 \mu_{G^2} \sum_{k=2}^d w_k^\alpha + \sum_{k,l \geq 2, k \neq l} w_k^\alpha w_l \mu_G\right) \overline{G}(t)(1 + o(1)).\end{aligned}$$

Similarly,

$$\begin{aligned}&J_{1,2}^{(2)}(t) \\ &= (w_1^\alpha + 2w_2^\alpha + \cdots + w_d^\alpha)\,\overline{G}(t) + o\left(\overline{G}(t) A(t)\right). \\ &+ \alpha t^{-1} \left(2 w_2^\alpha \sum_{k \neq 2} w_k \mu_G + w_2 \mu_{G^2} \sum_{k \neq 2} w_k^\alpha + \sum_{k,l \neq 2, k \neq l} w_k^\alpha w_l \mu_G\right) \overline{G}(t)(1 + o(1)).\end{aligned}$$

and

$$
\begin{aligned}
J_{1,2}^{(3)}(t) &= (2w_1^\alpha + 2w_2^\alpha + \cdots + w_d^\alpha)\,\overline{G}(t) + o\left(\overline{G}(t)A(t)\right) \\
&\quad + \alpha t^{-1}\left(2\sum_{l=1}^{2}\sum_{k\neq l} w_l^\alpha w_k \mu_{G^2} + 2\sum_{l=1}^{2}\sum_{k=3}^{d} w_l^\alpha w_k \mu_G\right)\overline{G}(t)(1+o(1)) \\
&\quad + \alpha t^{-1}\left(\sum_{l=1}^{2}\sum_{k=3}^{d} w_k^\alpha w_l \mu_{G^2} + \sum_{k,l\geq 3, k\neq l} w_k^\alpha w_l \mu_G\right)\overline{G}(t)(1+o(1)).
\end{aligned}
$$

Combining all the asymptotics for $I(t)$, $J_1(t)$, $J_2(t)$ and $J_3(t)$ yields that

$$
\begin{aligned}
\frac{\overline{F}_{\boldsymbol{w}^T\boldsymbol{X}}(t)}{\overline{G}(t)} - \sum_{k=1}^{d} w_k^\alpha &= (1+Q_{\boldsymbol{a}})\frac{I(t)}{\overline{G}(t)} + \frac{\sum_{i<j} a_{i,j}\left(-J_{ij}^{(1)}(t) - J_{ij}^{(2)}(t) + J_{ij}^{(3)}(t)\right)}{\overline{G}(t)} - \sum_{k=1}^{d} w_k^\alpha \\
&= \begin{cases} \alpha t^{-1}\mu_G^*(1+o(1)), & \rho < -1, \\ (1+Q_{\boldsymbol{a}})\sum_{k=1}^{d} H_{-\alpha,\rho}\left(w_k^{-1}\right)A(t)(1+o(1)), & \rho \geq -1. \end{cases}
\end{aligned}
$$

This completes the proof of (3.4).

The uniform convergence of (3.4) follows immediately from checking that for the limit relations in Proposition 3.7 and Theorems 4.7 of Mao and Ng [34]. The details are omitted here but are available upon request. ■

Proof of Lemma 3.2. In this proof we denote $\arg\min f$ by m_f for notational simplicity. By the definition of D_n, for any n, $|f_n(m_f) - f(m_f)| < D_n$. It follows that

$$
f_n(m_{f_n}) \leq f_n(m_f) < f(m_f) + D_n.
$$

Again, by $|f_n(m_{f_n}) - f(m_{f_n})| < D_n$ we have

$$
f(m_{f_n}) < f_n(m_{f_n}) + D_n < f(m_f) + 2D_n.
$$

Deriving the similar inequalities for the other side yields that

$$
|f(m_{f_n}) - f(m_f)| < 2D_n. \tag{A.21}
$$

By the Taylor's theorem, for any $\boldsymbol{x}$ in a small neighborhood of m_f we obtain that

$$
f(\boldsymbol{x}) = f(m_f) + \frac{1}{2}(\boldsymbol{x} - m_f)^T \boldsymbol{\nabla}^2 f(m_f)(\boldsymbol{x} - m_f) + o\left(||\boldsymbol{x} - m_f||_2^2\right), \tag{A.22}
$$

where we used the multi-index notation and $\boldsymbol{\nabla}^2 f(m_f)$ is the Hessian matrix of f at m_f. Since $D_n \to 0$ as $n \to \infty$ by Lemma 2.1, m_{f_n} is in a small neighborhood of m_f for large n. It then follows from the expansion in (A.22) that

$$
|f(m_{f_n}) - f(m_f)| > \frac{c}{2}||m_{f_n} - m_f||_2^2, \tag{A.23}
$$

where c is the smallest eigenvalue of $\boldsymbol{\nabla}^2 f(m_f)$. Combining (A.21) and (A.23) leads to the desired result. ■

References

[1] Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *The Review of Financial Studies*, **30**(1):2–47.

[2] Basrak, B., Davis, R. A., and Mikosch, T. (2002). A characterization of multivariate regular variation. *The Annals of Applied Probability*, **12**(3), 908–920.

[3] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2004). Statistics of extremes: theory and applications (Vol. 558). John Wiley & Sons.

[4] Caeiro, F., Gomes, M. I., and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat*, **3**(2), 113-136.

[5] Cambanis, S. (1977). Some properties and generalizations of multivariate Eyraud-Gumbel-Morgenstern distributions. *Journal of Multivariate Analysis*, **7**(4), 551–559.

[6] Choueifaty, Y. and Coignard, Y. (2008). Toward maximum diversification. *The Journal of Portfolio Management*, **35**(1), 40–51.

[7] Cui, H., Tan, K., and Yang, F. (2021). Diversification in catastrophe insurance markets. *ASTIN Bulletin*, 1-26. doi:10.1017/asb.2021.18.

[8] Dal Maso, G. (2012). An introduction to Γ-convergence (Vol. **6**). Springer Science & Business Media.

[9] Daouia, A., Gardes, L., & Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, **19**(5B), 2557-2589.

[10] De Haan, L. and Ferreira, A. (2007). Extreme value theory: an introduction. Springer Science & Business Media.

[11] Degen, M., Lambrigger, D. D., and Segers, J. (2010). Risk concentration and diversification: second-order properties. *Insurance: Mathematics and Economics*, **46**(3), 541–546.

[12] Dekkers, A. L., Einmahl, J. H., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, 1833–1855.

[13] Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *The Annals of Applied Probability*, **10**(4), 1274-1301.

[14] Eastoe, E. F., Heffernan, J. E., and Tawn, J. A. (2014). Nonparametric estimation of the spectral measure, and associated dependence measures, for multivariate extreme values using a limiting conditional representation. *Extremes*, **17**(1), 25-43.

[15] Einmahl, J. H., de Haan, L., & Zhou, C. (2016). Statistics of heteroscedastic extremes. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 78(1), 31-51.

[16] Einmahl, J. H., Piterbarg, V. I., and de Haan, L. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *The Annals of Statistics*, **29**(5), 1401-1423.

[17] Einmahl, J. H., and Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics*, 2953-2989.

[18] Embrechts, P., Lambrigger, D. D., and Wüthrich, M. V. (2009). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes*, **12**(2), 107–127.

[19] Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 1:176–223.

[20] Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research*, **77**(3), 423–432.

[21] Farlie, D.J.G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, **47**, 307–323.

[22] Fischer, M. and Klein, I. (2007). Constructing generalized FGM copulas by means of certain univariate distributions. *Metrika*, **65**(2), 243–260.

[23] Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H.E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature* 423, 267–270.

[24] Gumbel, E.J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association.* **55**, 698–707.

[25] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, **3**(5), 1163–1174.

[26] Hofert, M. and Wuthrich, M. V. (2011). Statistical review of nuclear power accidents. *Asia-Pacific Journal of Risk and Insurance*, 7(1).

[27] Hult, H., and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied probability*, **34**(3), 587–608.

[28] Ibragimov, R., Jaffee, D., and Walden, J. (2009). Nondiversification traps in catastrophe insurance markets. *Review of Financial Studies*, **22**(3):959–993.

[29] Kulik, R., and Soulier, P. (2020). Heavy-tailed time series. Berlin: Springer.

[30] Lindskog, F. (2004). Multivariate extremes and regular variation for stochastic processes (Doctoral dissertation, Diss., Mathematische Wissenschaften, Eidgenossische Technische Hochschule ETH Zurich, Nr. 15319, 2004).

[31] Loretan, M., and Phillips, P. C. (1994). Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *Journal of Empirical Finance*, **1**(2), 211–248.

[32] Mainik, G., and Rüchendorf, L. (2010). On optimal portfolio diversification with respect to extreme risks. *Finance and Stochastics*, **14**(4), 593–623.

[33] Mainik, G., and Embrechts, P. (2013). Diversification in heavy-tailed portfolios: properties and pitfalls. *Annals of Actuarial Science*, **7**(01), 26–45.

[34] Mao, T., and Ng, K. W. (2015). Second-order properties of tail probabilities of sums and randomly weighted sums. *Extremes*, **18**(3), 403–435.

[35] Mao, T., and Yang, F. (2015). Risk concentration based on Expectiles for extreme risks under FGM copula. *Insurance: Mathematics and Economics*, **64**, 429–439.

[36] McNeil, A. J., Frey, R., and Embrechts, P. (2015). Quantitative Risk Management: Concepts, Techniques and Tools: Concepts, Techniques and Tools. Princeton university press.

[37] Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of financial and quantitative analysis*, 7(04), 1851-1872.

[38] Moore, K., Sun, P., De Vries, C. G., and Zhou, C. (2013). The cross-section of tail risks in stock returns. Available at SSRN 2240131.

[39] Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. Mitteilungsblatt für Mathematische Statistik, **8**(1), 234-235.

[40] Peng, L. (1998). Asymptotically unbiased estimators for the extreme-value index. *Statistics & Probability Letters*, **38**(2), 107-115.

[41] Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 119–131.

[42] Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 659–680.

[43] Resnick, S. I. (2007). Heavy-tail phenomena: probabilistic and statistical modeling. Springer Science & Business Media.

[44] Roy, A. D. (1952). Safety first and the holding of assets. *Econometrica: Journal of the Econometric Society*, 431–449.

[45] Smith, R. L. (1987). Estimating tails of probability distributions. *The Annals of Statistics*, 1174–1207.

[46] Soulier, P. (2022). The tail process and tail measure of continuous time regularly varying stochastic processes. *Extremes*, **25**(1), 107-173.

[47] Van der Vaart, A. W. (2000). Asymptotic statistics (Vol. **3**). Cambridge university press.

[48] Wang, H., and Tsai, C. L. (2009). Tail index regression. *Journal of the American Statistical Association*, **104**(487), 1233-1240.

[49] Zhou, C. (2010). Dependence structure of risk factors and diversification effects. *Insurance: Mathematics and Economics*, **46**(3), 531–540.