Asymptotic Analysis of Portfolio Diversification

Hengxin Cui[a] Ken Seng Tan[a] Fan Yang[a]* Chen Zhou[b],[c]
[a] Department of Statistics and Actuarial Science, University of Waterloo
Waterloo, ON N2L 3G1, Canada
[b]Economics and Research Division, De Nederlandsche Bank
Amsterdam, The Netherlands
[c]Erasmus University Rotterdam
Rotterdam, The Netherlands

April 20, 2017

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.

1 Introduction

In order to mitigating risks in portfolios of financial investment, a common tool used by risk managers is the diversification strategy. The benefit from a diversification strategy can be reflected in the reduction of tail risks in a diversified portfolio. Guided by regulation rules such as the Basel II and III Accords for banking regulation and the Solvency II Directive for insurance regulation, the Value-at-Risk (VaR) became the main concern of the regulators, and therefore is also adopted by risk managers as the main measure of risks. In this paper, we investigate the optimal portfolio construction aiming at extracting the most diversification benefit 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 the portfolio return; see e.g. Mainik and Rüchendorf [13], Mainik and Embrechts [14] and Zhou [26]. By inverting the approximation formula on the tail of the distribution, one may get an approximation for the VaR measure, when

*Corresponding author: fan.yang@uwaterloo.ca

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, when applying the EVT approach, two difficulties remain to be handled. Both of them are due to the fact that the approximation holds only in the limit when the probability level in VaR tending to 1. Firstly, the EVT approach provides an approximation for "the VaR in the limit" when the probability level in VaR tends to 1. However, 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 portfolio 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. Further, 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 comparing 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. [4] and Embrechts et al. [6]. The diversification ratio is defined as follows. Let $\mathbf{X} := (X_1, \dots X_d)^T$ be a non-negative random vector indicating the losses of d assets. The value of a portfolio is given by $\mathbf{w}^T\mathbf{X}$, where the weights satisfy $\mathbf{w} = (w_1, w_2, \dots, w_d)^T \in \Sigma^d := \left\{\mathbf{x} \in [0,1]^d : x_1 + x_2 + \dots + x_d = 1\right\}$. For this portfolio, the diversification ratio (DR) based on VaR at level $q \in (0,1)$ is defined as

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

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, $1 - \mathrm{DR}_{\mathbf{w},q}$ can be regarded as the diversification benefit.

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 $\mathbf{X}$ by MRV, we can derive an explicit formula for

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

with respect to the weight $\mathbf{w}$ 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

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

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.

[1] As pointed out by Mainik and Embrechts [14], under the MRV structure, when the tail index is great than 1, $\mathrm{DR}_{\mathbf{w},1} < 1$. 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.

However, the second difficulty raised above remains valid after switching to minimizing the DR. Is the optimal solution based on minimizing the DR in the limit close to the practical goal of minimizing the DR at a given probability level? We formalize this question by the following notations.

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

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

Denote the solution to (1.2) by $\mathbf{w}_q$.

We remark that solving (1.2) directly is computationally intensive. With observations on the joint distribution of the random vector $\mathbf{X}$, $\mathbf{w}_q$ can be estimated by conducting a numerical 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 $\mathbf{w}^*$ is from the solution of the original optimization problem $\mathbf{w}_q$. First, we show theoretically that

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

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 close to 1. Further, define the distance between $\mathbf{w}_q$ and $\mathbf{w}^*$, measured by $\|\mathbf{w}_q - \mathbf{w}^*\|$ with respect to an arbitrary norm as D_q. In other words, given a finite level of q close to 1, the solution $\mathbf{w}_q$ is within an area defined as a D_q radius circle around $\mathbf{w}^*$. 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 $\mathbf{X}$, 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 $\mathbf{w}^*$ 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 $\mathbf{w}_q$.

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,

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 [23], 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.

Our proposed portfolio optimization strategy is comparable to other strategies based on tail risk. Mainik and Rüchendorf [13], proposed to minimize the so-called extreme risk index (ERI),

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

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)

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

proposed by Choueifaty and Coignard [2]. 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.

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 two numerical examples. Section 5 provides the application of our strategy to real market data. Proofs are postponed to Section 6.

2 Convergence of optimal portfolios

2.1 Preliminaries

2.1.1 The multivariate regular variation model

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

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

where $\xrightarrow{\mathrm{v}}$ refers to the vague convergence. A possible choice of b_t is $b_t = F_R^{\leftarrow}(1-1/t)$ where $R = ||\mathbf{X}||$ and $\|\cdot\|$ is an arbitrary norm. In this paper, we assume that the limit measure v is nondegenerate in the sense that

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

for all $i = 1, 2, \ldots, d$.

The limiting measure v has the scaling property

$$v(tA) = t^{-\alpha} v(A) \tag{2.2}$$

for all sets $A \in \mathcal{B}\left([0,\infty]^d \setminus \{\mathbf{0}\}\right)$. The scaling property leads to a decomposition of the v measure after a polar coordinate transformation as follows.

For any arbitrary norm $\|\cdot\|$, the polar coordinate transform of a vector $\mathbf{x}$ is

$$T(x) = \left(\|\mathbf{x}\|, \|\mathbf{x}\|^{-1}\mathbf{x}\right). \tag{2.3}$$

Without loss of generality, we can assume that $v(\|x\| \geq 1) = 1$ based on a proper choice of the sequence b_t. Then the scaling property (2.2) leads to the decomposition

$$v = \rho_\alpha \otimes \Psi, \tag{2.4}$$

where $\rho_\alpha(x, \infty) = x^{-\alpha}$ for $x > 0$ and Ψ is a probability measure on the unit sphere $\mathcal{S}_+^{d-1} = \left\{ \mathbf{s} \in \mathbb{R}_+^d : \|\mathbf{s}\| = 1 \right\}$. The measure Ψ in (2.4) is called the spectral or angular measure. Throughout the paper, we denote that $\mathbf{X}$ is MRV with index α and spectral measure Ψ by $\mathbf{X} \in \mathrm{MRV}_\alpha(\Psi)$, which implies the corresponding limit measure v as in (2.4).

Theoretically, it does not matter which norm is chosen in the polar representation (2.3). For simplicity, in this paper we consider the L_1-norm $\|\cdot\|_1$.

Further, by constraining the measures v_t and v to the set $A_1 := \left\{ \mathbf{x} \in \mathbb{R}_+^d : \|\mathbf{x}\|_1 > 1 \right\}$, and taking $b_t = F_R^{\leftarrow}(1 - 1/t)$ with $R = ||\mathbf{X}||_1$, the vague convergence in (2.1) implies the weak convergence on $\mathcal{B}(A_1)$, as

$$v_t(\cdot) = \frac{\Pr\left(t^{-1}\mathbf{X} \in \cdot\right)}{\Pr\left(\|\mathbf{X}\|_1 > t\right)} \xrightarrow{\mathrm{w}} v(\cdot)|_{A_1}, \qquad t \to \infty, \tag{2.5}$$

where $v|_{A_1}$ is the restriction of v to the set A_1. Note that v_t in (2.5) can also be rewritten as a conditional probability $\Pr\left(t^{-1}\mathbf{X} \in \cdot | \, \|\mathbf{X}\|_1 > t\right)$.

2.1.2 Convergence of minimizers

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

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

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

Lemma 2.1 *Suppose that $\{f_n\}$ is a sequence of lower semi-continuous functions from a compact metric space Z 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*

$$\lim_{n \to \infty} m_{f_n} = \arg\min f. \tag{2.6}$$

Proof. On the compact metric space Z, we have that the sequence $\{f_n\}$ is equi-coercive and gamma-converges to f under the conditions of Theorem 2.1. Then by Corollary 7.24 in Dal Maso [3], the relation (2.6) holds. ■

2.2 Main results

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

Proposition 2.1 *Suppose the nonnegative random vector $\mathbf{X}$ has a continuous joint distribution F. Further assume that $\mathbf{X} \in \mathrm{MRV}_\alpha(\Psi)$ with $\alpha > 0$. Then for any $\mathbf{w} \in \Sigma^d$, we have*

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

where

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

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

Proof. Note that

$$\mathrm{DR}_{\mathbf{w},q} = \frac{\mathrm{VaR}_q(\mathbf{w}^T\mathbf{X})/\mathrm{VaR}_q(\|\mathbf{X}\|_1)}{\sum_{i=1}^d w_i \mathrm{VaR}_q(X_i)/\mathrm{VaR}_q(\|\mathbf{X}\|_1)}. \tag{2.7}$$

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

$$\lim_{q\uparrow 1} \frac{\mathrm{VaR}_q(\mathbf{u}^T\mathbf{X})}{\mathrm{VaR}_q(\|\mathbf{X}\|_1)} = \eta_{\mathbf{u}}^{1/\alpha}, \qquad \mathbf{u} \in \Sigma^d, \tag{2.8}$$

which can be found in e.g. Mainik and Rüchendorf [13], Mainik and Embrechts [14] and Zhou [26]. The proposition can be proved by letting $\mathbf{u} = \mathbf{w}$ and $\mathbf{u} = \mathbf{e}_i$ in (2.8). ■

In the following theorem, we develop the uniform convergence of $\mathrm{DR}_{\mathbf{w},q}$, which is essential for proving the convergence of minimizers. It is also an interesting result on its own. The proof is postponed to Section 6.

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

$$\lim_{q\uparrow 1} \sup_{\mathbf{w}\in\Sigma^d} |\mathrm{DR}_{\mathbf{w},q} - \mathrm{DR}_{\mathbf{w},1}| = 0. \tag{2.9}$$

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

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

$$\lim_{q\uparrow 1} \mathbf{w}_q = \mathbf{w}^*, \tag{2.10}$$

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

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

$$\begin{aligned} &\min_{\mathbf{w}} \eta_{\mathbf{w}}^{1/\alpha} \\ &\textbf{s.t.} \quad \textstyle\sum_{i=1}^d w_i \eta_{\mathbf{e}_i}^{1/\alpha} = 1 \text{ with } w_i \geq 0 \text{ for } i = 1, 2, \ldots, d. \end{aligned} \tag{2.11}$$

Since the set of constraints in (2.11) is nonempty, closed and bounded, it is compact. By Theorem 2.4 of Mainik and Embrechts [14], $\eta_{\mathbf{w}}^{1/\alpha}$ is strictly convex when $\alpha > 1$ and $\Psi\left(\{\mathbf{x} : \mathbf{a}^T\mathbf{x} = 0\}\right) = 0$ for any $\mathbf{a} \in \mathbb{R}^d$. Suppose $\mathbf{w}_1$ and $\mathbf{w}_2$ are two different minimal points of the optimization problem. Let $\mathbf{w} = (\mathbf{w}_1 + \mathbf{w}_2)/2$. From the strictly convexity of the object function and compactness of the set of constraints, it follows that $\eta_{\mathbf{w}}^{1/\alpha} < \eta_{\mathbf{w}_1}^{1/\alpha} = \eta_{\mathbf{w}_2}^{1/\alpha}$, which yields a contradiction. Thus, $\mathbf{w}^*$ is unique.

Now we prove (2.10). In the proof of Theorem 6.2, we showed that $\mathrm{VaR}_q(\mathbf{w}^T\mathbf{X})$ is continuous with respect to $\mathbf{w} \in \Sigma^d$ for q large. Then there exists $q^* > 0$ such that $\mathrm{DR}_{\mathbf{w},q}$ is continuous with respect to $\mathbf{w} \in \Sigma^d$ for every $q^* < q < 1$. The desired result follows from Theorem 2.1, the uniqueness of $\mathbf{w}^*$ and Lemma 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 $\Psi\left(\left\{\mathbf{x}:\mathbf{a}^T\mathbf{x}=0\right\}\right)=0$ for any $\mathbf{a}\in\mathbb{R}^d$ means that the spectral measure Ψ does not concentrate on any linear subspace. It ensures the uniqueness of the solution $\mathbf{w}^*$ of the limiting problem $\mathrm{DR}_{\mathbf{w},1}$. But it excludes the special cases such as independent or comonotonic structure of $\mathbf{X}$. If $\mathbf{X}$ has independent structure with regularly varying marginals, then it is not hard to show that

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

By Jensen's inequality, $\mathrm{DR}_{\mathbf{w},1}$ 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 $\mathbf{X}$ is comonotonic, then $\mathrm{DR}_{\mathbf{w},q}=1$ for any $\mathbf{w}$ 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 $\mathbf{X}$ in $\mathbb{R}^d$ is elliptically distributed if it satisfies

$$\mathbf{X}\overset{d}{=}\mu+YB\mathbf{U},\tag{2.12}$$

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

It is well known that if $\mathbf{X}$ is elliptical distributed, then $\mathbf{X}\in\mathrm{MRV}_\alpha(\Psi)$ if and only if $Y\in\mathrm{RV}_{-\alpha}$; for example, see Hult and Lindskog [11]. By Theorem 6.8 of McNeil et al. [17], the subadditivity property of VaR always holds for $0.5\le q<1$. It then follows that $\mathrm{DR}_{\mathbf{w},q}\le 1$, 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 $\mathrm{DR}_{\mathbf{w},q}\le 1$ is ensured by restricting $\alpha>1$. In another word, if $\mathbf{X}$ is elliptical distributed and $Y\in\mathrm{RV}_{-\alpha}$, then Theorem 2.2 holds without any restriction on α.

Actually, elliptical distributions leads to the explicit expressions of $\mathrm{DR}_{\mathbf{w},q}$ and $\mathrm{DR}_{\mathbf{w},1}$. 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 (2.10) without the assumption that Y is regularly varying. A direct calculation yields that

$$\mathrm{VaR}_q\left(\mathbf{w}^T\mathbf{X}\right)=\mathbf{w}^T\mu+\left\|B^T\mathbf{w}\right\|_2F_Z^{\leftarrow}(q),\tag{2.13}$$

where $Z\overset{d}{=}RU_1$. The diversification ratio for elliptical distributions can then be obtained as

$$\mathrm{DR}_{\mathbf{w},q}=\frac{\mathbf{w}^T\mu+\left\|B^T\mathbf{w}\right\|_2F_Z^{\leftarrow}(q)}{\mathbf{w}^T\mu+\sum_{i=1}^{d}w_i\left\|B^T\mathbf{e}_i\right\|_2F_Z^{\leftarrow}(q)}.\tag{2.14}$$

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

$$\lim_{q\uparrow 1}\mathrm{DR}_{\mathbf{w},q}=\frac{\left\|B^T\mathbf{w}\right\|_2}{\sum_{i=1}^{d}w_i\left\|B^T\mathbf{e}_i\right\|_2}:=\mathrm{DR}_{\mathbf{w},1}.\tag{2.15}$$

In the following lemma, we first show that the convergence in (2.15) is indeed uniform, whose proof is postponed to the last section.

Lemma 2.2 *For elliptically distributed* $\mathbf{X}$ *and* $\mathbf{w} \in \Sigma^d$*, if* $\mu \in l_1$*, the induced norm* $\|B\|_2 = \sup_{x \neq 0} \frac{\|B\mathbf{x}\|_2}{\|\mathbf{x}\|_2} < \infty$ *and random variable* Y *is unbounded, then the convergence in (2.15) is uniform for* $\mathbf{w} \in \Sigma^d$*. Moreover, the mapping* $\mathbf{w} \to \mathrm{DR}_{\mathbf{w},1}$ *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*

$$\lim_{q \uparrow 1} \arg\min_{\mathbf{w} \in \Sigma^d} \frac{\mathrm{VaR}_q(\mathbf{w}^T \mathbf{X})}{\sum_{i=1}^d w_i \mathrm{VaR}_q(X_i)} = \arg\min_{\mathbf{w} \in \Sigma^d} \frac{\left\|B^T \mathbf{w}\right\|_2}{\sum_{i=1}^d w_i \left\|B^T \mathbf{e}_i\right\|_2}. \tag{2.16}$$

Proof. By Lemmas 2.1 and 2.2, we only need to show that the solutions of the minimization problems on both sides of (2.16) exist and are unique. To achieve it, first note that the minimization problem

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

is equivalent to a convex optimization problem

$$\begin{aligned} &\min_{\mathbf{w}} \mathbf{w}^T \mu + \left\|B^T \mathbf{w}\right\|_2 F_Z^{\leftarrow}(q) \\ &\textbf{s.t.} \quad \mathbf{w}^T \mu + \textstyle\sum_{i=1}^d w_i \left\|B^T \mathbf{e}_i\right\|_2 F_Z^{\leftarrow}(q) = 1 \text{ with } w_i \geq 0 \text{ for } i = 1, 2, \ldots, d. \end{aligned} \tag{2.17}$$

Similarly, the minimization problem

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

is equivalent to

$$\begin{aligned} &\min_{\mathbf{w}} \left\|B^T \mathbf{w}\right\|_2 \\ &\textbf{s.t.} \quad \textstyle\sum_{i=1}^d w_i \left\|B^T \mathbf{e}_i\right\|_2 = 1 \text{ with } w_i \geq 0 \text{ for } i = 1, 2, \ldots, d. \end{aligned} \tag{2.18}$$

Denote the constraint sets in (2.17) and (2.18) 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 (2.17) and (2.18) 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 (2.17) and (2.18) exist due to the Weierstrass extreme value theorem.

Next, we show the uniqueness of the solution to (2.18). Due to the convexity, we have for any $\lambda \in (0, 1)$,

$$\left\|B^T (\lambda \mathbf{w}_1 + (1 - \lambda) \mathbf{w}_2)\right\|_2 \leq \lambda \left\|B^T \mathbf{w}_1\right\|_2 + (1 - \lambda) \left\|B^T \mathbf{w}_2\right\|_2. \tag{2.19}$$

The equality in (2.19) holds only when $\mathbf{w}_1 = k\mathbf{w}_2$ for $k \in \mathbb{R}^+$ and $\mathbf{w}_1, \mathbf{w}_2$ nonzero. If both $\mathbf{w}_1$ and $\mathbf{w}_2$ belong to the constraint set C_1 or C_2, then k can only be 1. This means for any $\mathbf{w}_1 \neq \mathbf{w}_2$, the strictly inequality in (2.19) holds. Therefore, the objective function in (2.18) is strictly convex. The uniqueness of the solution then follows from the similarly arguments in the proof of Theorem 2.2. ■

2.4 Estimation of the diversification ratio

When the DR optimization strategy with MRV structure is applied in practice, the estimations of MRV structure and $\mathrm{DR}_{\mathbf{w},1}$ are required. In this section, we propose an estimation procedure and show the consistency of the estimators.

Assume $\mathbf{X} \in \mathrm{MRV}_\alpha(\Psi)$ with $\alpha > 1$. Let $\mathbf{X}_1, \ldots \mathbf{X}_n$ be an i.i.d. sample of $\mathbf{X}$. 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 $\eta_{\mathbf{w}}$ by
$$\widehat{\eta}_{\mathbf{w}} = \int_{\Sigma^d} (\mathbf{w}^T \mathbf{s})^{\widehat{\alpha}} \widehat{\Psi}(d\mathbf{s}).$$
4. Estimate $\mathrm{DR}_{\mathbf{w},1}$ by
$$\widehat{\mathrm{DR}}_{\mathbf{w},1} = \frac{\widehat{\eta}_{\mathbf{w}}^{1/\alpha}}{\sum_{i=1}^d w_i \widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}}.$$

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

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 $\mathbf{X}$ and $\mathbf{X}_i$ with respect to $||\cdot||_1$. That is,

$$(R, S) = \left(||\mathbf{X}||_1, \frac{\mathbf{X}}{||\mathbf{X}||_1} \right). \tag{2.20}$$

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:

$$\widehat{\alpha} = \widehat{\alpha}\left(R_{(1)}, \ldots, R_{(k)}\right).$$

Many existing estimators can be applied here, see for example, Hill [10], Pickands [21], Smith [24], Dekkers et al. [5], 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 $S_{\pi(1)}, \ldots, S_{\pi(k)}$,

$$\widehat{\Psi} = \frac{1}{k} \sum_{j=1}^{k} \delta_{S_{\pi(j)}}, \tag{2.21}$$

where $\delta_{\pi(j)}(\cdot)$ is the Dirac measure.

Lemma 2.3 *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an i.i.d. sample of $\mathbf{X} \in \mathrm{MRV}_\alpha(\Psi)$ with $\alpha > 1$. Assume that the distribution function F_R of R in (2.20) is continuous. If the estimator $\widehat{\alpha}$ is consistent almost surely, and then the estimator $\hat{\mathrm{DR}}_{\mathbf{w},1}$ is consistent uniformly in $\mathbf{w} \in \Sigma^d$, i.e.,*

$$\sup_{\mathbf{w}\in\Sigma^d} \left|\widehat{\mathrm{DR}}_{\mathbf{w},1} - \mathrm{DR}_{\mathbf{w},1}\right| \to 0, \qquad a.s. \tag{2.22}$$

Combining Theorem 2.1 and Lemma 2.3, we obtain the consistency in the optimal portfolio weights in the following theorem.

Theorem 2.4 *Under the conditions of Theorem 2.3 and $\Psi\left(\left\{\mathbf{x} : \mathbf{a}^T\mathbf{x} = 0\right\}\right) = 0$ for any $\mathbf{a} \in \mathbb{R}^d$, the estimator $\widehat{\mathbf{w}}^*$ and the estimated value $\hat{\mathrm{DR}}_{\mathbf{w}^*,1}$ are consistent almost surely, i.e.,*

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

Here we only established consistency. Under some additional conditions, further asymptotic properties for the estimator of $\mathrm{DR}_{\mathbf{w},1}$ can be established in a straightforward way. For example, Theorem 4.5 of Mainik and Rüchendorf [13] shows that, under some additional conditions, for any $\mathbf{w} \in \Sigma^d$, $\sqrt{k}\left(\widehat{\eta}_{\mathbf{w}} - \eta_{\mathbf{w}}\right)$ converges to a distribution $G_{\mathbf{w}}$. Then by the functional delta method (e.g. Theorem 20.8 in Van der Vaart [25]), it is easy to show that $\sqrt{k}\left(\widehat{\mathrm{DR}}_{\mathbf{w},1} - \mathrm{DR}_{\mathbf{w},1}\right)$ converges to a given distribution. However, to establish the convergence in a uniform way is difficult and may be left for future research. Without a uniform asymptotic property on $\widehat{\mathrm{DR}}_{\mathbf{w},1}$ we cannot further investigate the asymptotic property of the optimal portfolio weights.

3 The rate of convergence to the optimal portfolio: an example

In this section, we discuss how $\mathbf{w}^*$ approximates $\mathbf{w}_q$ by determining the convergence rate of (2.10) under some special dependence structure, such as the FGM copula.

The FGM copula was originally introduced by Morgenstern [20] and investigated by Gumbel [9] and Farlie [7]. 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 [1], Fischer and Klein [8], among others. Here we focus on a high dimensional generalized FGM copula proposed by Cambanis [1], which is defined as

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

The constants $a_{i,j}$, $1 \le i < j \le n$, 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

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

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

We intend to consider the random vector $\mathbf{X}$ 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

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

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

3.1 Tail expansion for the weighted sum

Assume that the random vector $\mathbf{X}$ 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 $\rho \le 0$ 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 $x^{-\alpha}\log x$.

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

Lemma 3.1 *Let $\mathbf{X}$ 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 $\alpha > 1$, $\rho \le 0$ and auxiliary function $A(\cdot)$. Assume that $\mathbf{X}$ follows an n-dimensional generalized FGM copula given by (3.2). Then as $t \to \infty$, we have that*

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

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

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

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

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 m_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*

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

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

Proof. In this proof, all the limits are taken as $q \uparrow 1$. We first derive the second-order expansion of $\mathrm{DR}_{\mathbf{w},q}$. Similar to the proof of Theorem 4.6 in Mao and Yang [16], we have that

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

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 $t = F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)$. 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

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

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 $\mathrm{DR}_{\mathbf{w},q}$.

Immediately from (3.5), the limiting function is

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

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

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

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

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

By Lemma 3.2, we get that

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

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

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

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

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

we obtain that for some constant $C>0$

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

By Lemma 3.2 we get that

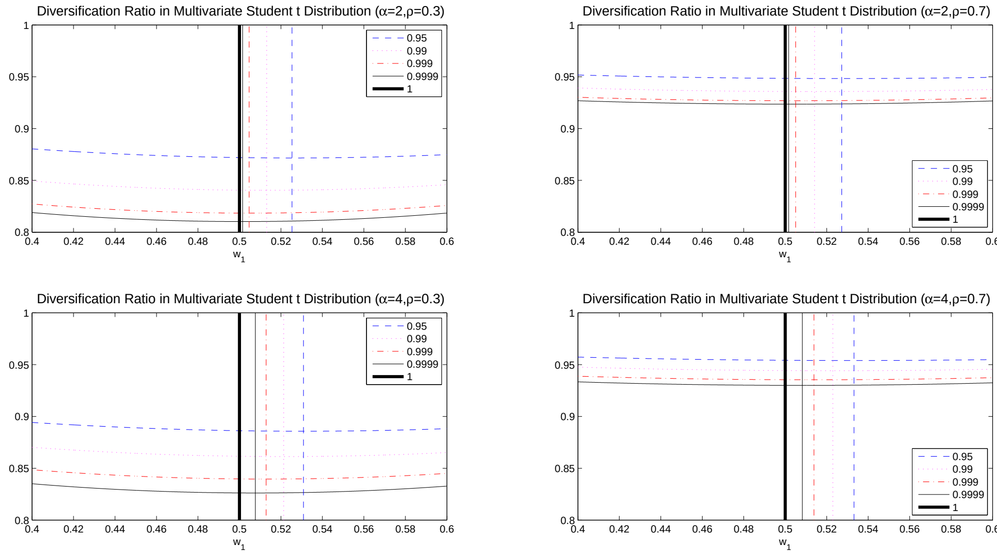
$$(1-q)^{\rho/\alpha}\left\|\mathbf{w}_q - \mathbf{d}^{-1}\right\| = O(1).$$

This completes the proof. ■

4 Numerical examples

In this section, we conduct two numerical examples to examine our theoretical results. The first example is an elliptical distribution–the bivariate Student–t distribution, while the second one is a non-elliptical distribution.

Figure 1: Optimal portfolio from elliptical distribution risk factors



Note: The portfolios are constructed as a linear combination of two risk factors from a bivariate Student-t distribution $t_\alpha(\mu, \Sigma)$ with $\mu = (1,2)^T$ and Σ is $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The $\mathrm{DR}_{\mathbf{w},q}$ of such portfolios for various values of q against the weight w_1 are plotted for different pairs of (α, ρ) with $\alpha = 2, 4$ and $\rho = 0.3, 0.7$ in the four subfigures. The level of q is 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 $\mathrm{DR}_{\mathbf{w},1}$ is indicated by a thick vertical line.

Consider $\mathbf{X}$ follows a bivariate Student-t distribution $t_\alpha(\mu, \Sigma)$, where $\mu = (1,2)^T$ and the covariance matrix Σ is $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Then the marginals both follow Student-t distribution with the degree of freedom α but different shifts 1 and 2.

We construct portfolios as a linear combination of the two risk factors from $\mathbf{X}$ defined above. As discussed in Section 2.3, both $\mathrm{DR}_{\mathbf{w},q}$ and $\mathrm{DR}_{\mathbf{w},1}$ can be explicitly expressed for elliptical distributions as in (2.14) and (2.15), 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 ρ at $\alpha = 2, 4$ and $\rho = 0.3, 0.7$, and plot the results for different pairs of (α, ρ) 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 indicates 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.

First, we observe that $\mathbf{w}_q$ is converging to $\mathbf{w}_1$ as $q \uparrow 1$. This verifies our theoretical result in Theorem 2.2. Second, the absolute difference between $\mathbf{w}_q$ and $\mathbf{w}_1$ 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 (ρ), the heavier the marginal tails reflected in a lower α, the faster the convergence rate. This is in line with our finding in Theorem 3.1: 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 ρ, the slower the convergence rate in the limit relation $\mathbf{w}_q \to \mathbf{w}_1$. Nevertheless, the slow convergence is not of a concern in practice. With a strong dependence at 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.

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 $\mathbf{X} = (X_1, X_2)^T$ is then defined as

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

where $\rho \in (-1, 1)$. Such random vector follows a non-elliptical distribution. In the case that the variance of Y_1 and Y_2 exists, ρ is the correlation coefficient between X_1 and X_2 Under this structure, the diversification ratio $\mathrm{DR}_{\mathbf{w},1}$ can be explicitly calculated. Following Mainik and Embrechts [14], we have that

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

and

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

Hence,

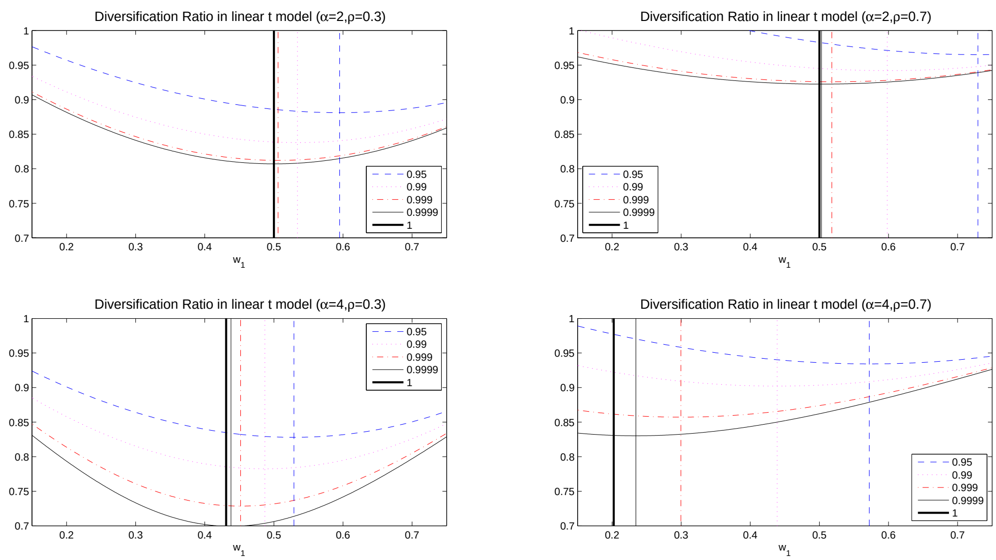
$$\mathrm{DR}_{\mathbf{w},1} = \left(w_1 \left((w_1 + w_2\rho)^\alpha + \left(w_2\sqrt{1-\rho^2}\right)^\alpha \right)^{-\frac{1}{\alpha}} + w_2 \left(\frac{(w_1 + w_2\rho)^\alpha + \left(w_2\sqrt{1-\rho^2}\right)^\alpha}{\rho^\alpha + \sqrt{1-\rho^2}^\alpha} \right)^{-\frac{1}{\alpha}} \right)^{-1}.$$

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

Consider a special case where Y_1 and Y_2 follow a standard Student–t distribution with degree of freedom $\alpha > 1$. By choosing $\alpha = 2, 4$ and $\rho = 0.3, 0.7$, in Figure 2 we plot the calculated diversification ratios $\mathrm{DR}_{\mathbf{w},q}$ 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.

Figure 2: Optimal portfolio from non–elliptical distribution risk factors



Note: The portfolios are constructed as a linear combination of two risk factors from a vector $\mathbf{X}$ defined in (4.1) with Y_1 and Y_2 following a standard Student–t distribution with degree of freedom $\alpha > 1$. The $\mathrm{DR}_{\mathbf{w},q}$ of such portfolios for various values of q against the weight w_1 are plotted for different pairs of (α, ρ) with $\alpha = 2, 4$ and $\rho = 0.3, 0.7$ in the four subfigures. The level of q is 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 $\mathrm{DR}_{\mathbf{w},1}$ is indicated by a thick vertical line.

5 Empirical study

In the numerical examples, the limit diversification ratio $\mathrm{DR}_{\mathbf{w},1}$ 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. In Section 2.4, we discuss the estimation methodology for $\mathrm{DR}_{\mathbf{w},1}$. In this section, we apply our estimation method and the optimal portfolio construction procedure to real market data.

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 [10]) 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. [19] 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 [10]). 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 $\text{DR}_{\mathbf{w},q}$ by a numerical search. This is achieved by assigning weights to the 10 stocks on a grid spanning the set σ^10, evaluating $\text{DR}_{\mathbf{w},q}$ at each grid point and taking the weights that corresponds to the minimum diversification ratio. Then we construct the optimal portfolio based on minimizing the estimated $\text{DR}_{\mathbf{w},1}$ using the procedure laid out in Section 2.4.

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 $||\mathbf{w}_q - \mathbf{w}^*||_1/10$. 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. Here, we choose k to be 4% for estimating α and 10% for estimating the spectral measure $\widehat{\Psi}$. Moreover, since we only consider the loss, the estimator for $\eta_{\mathbf{w}}$ is slightly modified to

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

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.

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 $\text{DR}_{\mathbf{w},q}$. This is achieved by assigning weights to the 10 stocks on a grid spanning the set σ^10, evaluating $\text{DR}_{\mathbf{w},q}$ at each grid point and taking the weights that corresponds to the minimum diversification ratio. The second optimal portfolio minimizes the estimated $\text{DR}_{\mathbf{w},1}$ using the procedure laid out in Section 2.4. The numbers reported are the distance calculated by $||\mathbf{w}_q - \mathbf{w}^*||_1/10$ between the two portfolios.

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 less 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 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 $\mathbf{w}_{95\%}$ 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.

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

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 $\mathrm{DR}_{\mathbf{w},1}$	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 $\mathrm{DR}_{\mathbf{w},1}$ based on 10 stocks.

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 strategy laid out in the first empirical study which minimizes the $\mathrm{DR}_{\mathbf{w},95\%}$ 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 [18]), and lastly a simple equal weight strategy.

Figure 3 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. To achieve the tail diversification benefit measured by the diversification ratio, our portfolio optimization strategy gives the best out–of–sample performance.

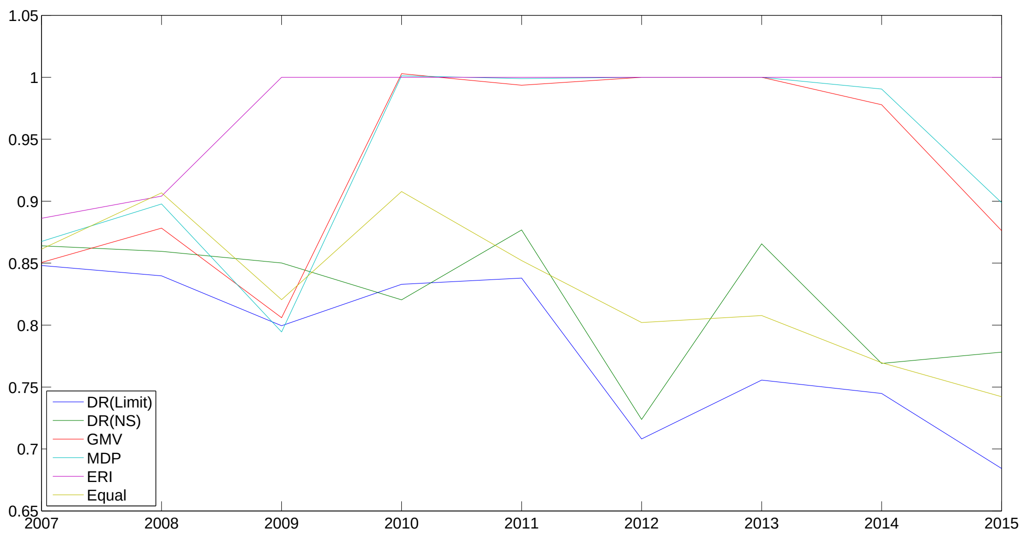
Figure 4 shows the results on the out–of–sample VaR. Our portfolio optimization strategy produces the lowest VaR in 2007 and 2008, but not in the other years. Nevertheless, the VaR of the optimal portfolio from our strategy is never largely above ERI, which minimizes VaR among the six strategies. Furthermore, 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 gives the best out–of–sample performance in terms of risk management.

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 distance between the optimal portfolios obtained from our limit DR optimization 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 fast and better performed algorithm derived from our limit DR optimization strategy.

6 Proofs

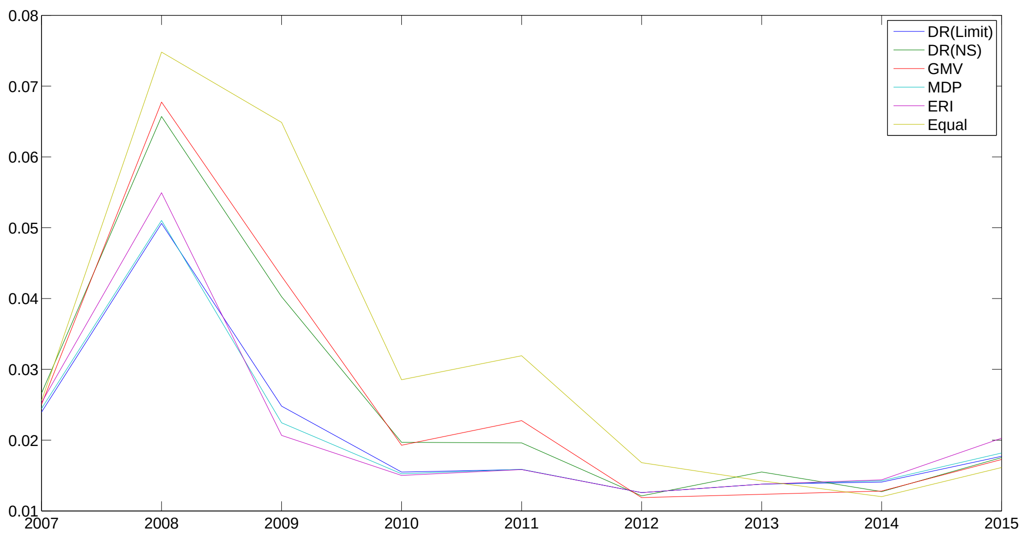
In this section, we first prove Theorem 2.1, which is the key and the most difficult part in the proof of Theorem 2.2, in two steps as Sections 6.1 and 6.2. Then the very last section contains all the proofs of lemmas from previous sections.

Figure 3: Out–of–sample diversification ratio



Note: Within each five-year window, the optimal portfolio based on the 10 selected stocks in Table 1 is constructed by minimizing $\mathrm{DR}_{\mathbf{w},1}$. 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 steps are repeated for five other strategies, the numerical search strategy for minimizing $\mathrm{DR}_{\mathbf{w},95\%}$ (DR(NS)), global minimum variance (GMV; see, e.g. Merton [18]), the MDP, the ERI, and equal weight strategy (Equal).

Figure 4: Comparison of portfolio risks



Note: Within each five-year window, the optimal portfolio based on the 10 selected stocks in Table 1 is constructed by minimizing $\mathrm{DR}_{\mathbf{w},1}$. 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 steps are repeated for five other strategies, the numerical search strategy for minimizing $\mathrm{DR}_{\mathbf{w},95\%}$ (DR(NS)), global minimum variance (GMV; see, e.g. Merton [18]), the MDP, the ERI, and equal weight strategy (Equal).

6.1 Uniform convergence in Radon measures

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

$$M = \left\{ f_{\mathbf{w}}(\mathbf{x}) = \frac{1}{1 + \mathbf{w}^T\mathbf{x}} : \mathbf{w} \in \Sigma^d,\, \mathbf{x} \in A_1 \right\}. \tag{6.1}$$

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

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

Theorem 6.1 *If $\mathbf{X} \in \mathrm{MRV}_\alpha(\Psi)$ with $\alpha > 0$, then*

$$\lim_{t\to\infty} \sup_{\mathbf{w}\in\Sigma^d} \left| v_t\left(A_{\mathbf{w},1}\right) - v\left(A_{\mathbf{w},1}\right)\right| = 0, \tag{6.2}$$

where ν_t and ν are defined in (2.5).

Proof. Since $A_{\mathbf{w},1} \in \mathcal{B}(A_1)$, by (2.5) we have that $v_t\left(A_{\mathbf{w},1}\right)$ converges weakly to $v\left(A_{\mathbf{w},1}\right)$. To further show the uniform convergence, we apply Theorem 3.4 of Rao [22]. That is we need to verify the following three conditions. (1) The mappings in M defined in (6.1) 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) For each $f_{\mathbf{w}} \in M$, $vf_{\mathbf{w}}^{-1}$ has a continuous distribution. Next, we prove them separately.

(1) By Theorem 1.5 of Lindskog [12], 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 $f_{\mathbf{w}} \in M$ is continuous.

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

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

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

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

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

(3) For $x \in \mathbb{R}_+$, we have

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

which is obviously continuous for any $0 < x < 1/2$. Furthermore, we have

$$\begin{aligned} v(A_1) &= \int_{\Sigma^d}\int_{\mathbb{R}_+} 1_{\{r>1\}} \rho_\alpha(dr)\Psi(d\mathbf{s}) \\ &= \int_{\Sigma^d} \Psi(d\mathbf{s}) = 1. \end{aligned}$$

By far, we have verified the three conditions. By the weak convergence in (2.5) and Theorem 3.4 of Rao [22], we obtain

$$\lim_{t\to\infty}\sup_{\mathbf{w}\in\Sigma^d}|v_t(A_{\mathbf{w},1})-v(A_{\mathbf{w},1})|=0,$$

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

Next corollary is a natural rewriting of relation (6.2). It yields a uniform convergence of the ratio $\Pr\left(\mathbf{w}^T\mathbf{X}>t\right)/\Pr\left(\|\mathbf{X}\|_1>t\right)$ to $\eta_{\mathbf{w}}$. However, only the weak convergence of it is known in the literature.

Corollary 6.1

$$\lim_{t\to\infty}\sup_{\mathbf{w}\in\Sigma^d}\left|\frac{\Pr\left(\mathbf{w}^T\mathbf{X}>t\right)}{\Pr\left(\|\mathbf{X}\|_1>t\right)}-\eta_{\mathbf{w}}\right|=0, \tag{6.3}$$

where

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

Further, the mapping $\mathbf{w}\mapsto\eta_{\mathbf{w}}$ is uniform continuous.

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

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

Moreover $v(A_{\mathbf{w},1})$ is actually

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

The desired result (6.3) then follows. Lastly, since $\eta_{\mathbf{w}}$ is continuous on the compact set Σ^d, it implies the uniform continuity of $\eta_{\mathbf{w}}$. ■

6.2 Uniform convergence in quantiles

In order to show that the convergence in (2.8) is indeed uniform, we first prepare a key lemma. For notational simplicity, we denote

$$l(\mathbf{w},q):=\frac{\mathrm{VaR}_q(\mathbf{w}^T\mathbf{X})}{\mathrm{VaR}_q(\|\mathbf{X}\|_1)}=\frac{F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)}{F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)}, \tag{6.4}$$

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

Lemma 6.1 *Suppose the nonnegative random vector $\mathbf{X}$ has a continuous joint distribution. Further assume that $\mathbf{X}\in\mathrm{MRV}_\alpha(\Psi)$ with $\alpha>0$. Then for any $\varepsilon>0$ there exist $0<\tilde{q}<1$ and $\delta=\delta(\varepsilon)$ such that for all $\mathbf{z},\mathbf{w}\in\Sigma^d$ and $\tilde{q}<p,q<1$satisfying $\|(\mathbf{w},q)-(\mathbf{z},p)\|<\delta$, we have*

$$|l(\mathbf{w},q)-l(\mathbf{z},p)|<\varepsilon. \tag{6.5}$$

Proof. Throughout the proof $\varepsilon > 0$ is arbitrary small but fixed. By the uniform convergence in (6.3), there exists $t_0 > 0$ such that

$$\left| \frac{1 - F_{\mathbf{w}^T\mathbf{X}}(t)}{1 - F_{\|\mathbf{X}\|_1}(t)} - \eta_{\mathbf{w}} \right| < \varepsilon, \tag{6.6}$$

for all $\mathbf{w} \in \Sigma^d$ and $t \geq t_0$. Moreover, by the uniform continuity of $\eta_{\mathbf{w}}$, there exists $\delta_1 > 0$ such that

$$|\eta_{\mathbf{z}} - \eta_{\mathbf{w}}| < \varepsilon \tag{6.7}$$

for all $\mathbf{z}, \mathbf{w} \in \Sigma^d$ satisfying $\|\mathbf{w} - \mathbf{z}\| < \delta_1$. In view of (6.6) and (6.7), we have

$$\begin{aligned} |F_{\mathbf{w}^T\mathbf{X}}(t) - F_{\mathbf{z}^T\mathbf{X}}(t)| &< \left| \frac{F_{\mathbf{w}^T\mathbf{X}}(t) - F_{\mathbf{z}^T\mathbf{X}}(t)}{1 - F_{\|\mathbf{X}\|_1}(t)} \right| \\ &< \left| \frac{1 - F_{\mathbf{z}^T\mathbf{X}}(t)}{1 - F_{\|\mathbf{X}\|_1}(t)} - \eta_{\mathbf{z}} \right| + |\eta_{\mathbf{z}} - \eta_{\mathbf{w}}| + \left| \eta_{\mathbf{w}} - \frac{1 - F_{\mathbf{w}^T\mathbf{X}}(t)}{1 - F_{\|\mathbf{X}\|_1}(t)} \right| \\ &< 3\varepsilon, \end{aligned}$$

for any $t \geq t_0$. Hence $F_{\mathbf{w}^T\mathbf{X}}(t)$ is uniformly continuous in $\mathbf{w}$ when t large enough. Especially, at t_0, by the extreme value theorem, there exists $\mathbf{w}_0$ such that $F_{\mathbf{w}^T\mathbf{X}}(t_0) \geq F_{\mathbf{w}_0^T\mathbf{X}}(t_0)$ for all $\mathbf{w}$. Denote the minimum $F_{\mathbf{w}_0^T\mathbf{X}}(t_0)$ by q_0. Note that q_0 only depends on the choice of t_0. Since $F_{\mathbf{w}^T\mathbf{X}}(t_0) \leq F_{\mathbf{w}^T\mathbf{X}}(t)$ for any $t > t_0$, by letting $q = F_{\mathbf{w}^T\mathbf{X}}(t)$ we have the following equivalence for any $\mathbf{w}$,

$$F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q) \geq t_0 \qquad \Longleftrightarrow \qquad q \geq q_0.$$

Since the random vector $\mathbf{X}$ has a continuous joint distribution, one can conclude that $F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)$ is continuos in q for all $\mathbf{w}$. That is, for some $\mathbf{w}$ and any $q \geq q_0$ (because we focus on the tail risk), there exists $\delta(\mathbf{w}, q) > 0$ such that

$$\left| F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(p) - F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q) \right| < \varepsilon, \tag{6.8}$$

for any p satisfying $|p - q| \leq \delta(\mathbf{w}, q)$. Similarly, $F^{\leftarrow}_{\|\mathbf{X}\|_1}(q)$ is also continuous in q. Then $l(\mathbf{w}, q)$ is continuous in q. Because of the weak convergence in (2.8), the limit case $l(\mathbf{w}, 1)$ is well defined to be $\eta_{\mathbf{w}}^{1/\alpha}$. Also note that there exists q' such that $F^{\leftarrow}_{\|\mathbf{X}\|_1}(q) > 1$ for all $q \geq q'$. Therefore,

$$l(\mathbf{w}, q) \leq F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q), \qquad q \geq q'. \tag{6.9}$$

Letting $\tilde{q} = q' \vee q_0$, we are ready to show (6.5), the uniform continuity, holds on $\mathbb{S} :=\Sigma^d \otimes [\tilde{q}, 1]$. Due to the relation (6.9), the compactness of $\mathbb{S}$ and the finiteness of $l(\mathbf{w}, q)$, it suffices to show that $F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)$ is continuous in both $\mathbf{w}$ and q on $\mathbb{S}$. By the uniform continuity of $F_{\mathbf{w}^T\mathbf{X}}(t)$ in $\mathbf{w}$, for $\delta(\mathbf{w}, q)/2$ with $\delta(\mathbf{w}, q)$ determined such that (6.8) holds, there exists δ_2 such that

$$F_{\mathbf{w}^T\mathbf{X}}(t) - \delta(\mathbf{w}, q)/2 \leq F_{\mathbf{z}^T\mathbf{X}}(t) \leq F_{\mathbf{w}^T\mathbf{X}}(t) + \delta(\mathbf{w}, q)/2, \tag{6.10}$$

for $\|\mathbf{w} - \mathbf{z}\| < \delta_2$ and $t \geq t_0$. Then, for $|p - q| \leq \delta(\mathbf{w}, q)/2$, we obtain the lower bound of $F^{\leftarrow}_{\mathbf{z}^T\mathbf{X}}(p)$ as

$$\begin{aligned} F^{\leftarrow}_{\mathbf{z}^T\mathbf{X}}(p) &= \inf \{x \geq t_0 : F_{\mathbf{z}^T\mathbf{X}}(x) \geq p\} \\ &\geq \inf \{x \geq t_0 : F_{\mathbf{w}^T\mathbf{X}}(x) + \delta(\mathbf{w}, q)/2 \geq q - \delta(\mathbf{w}, q)/2\} \\ &= F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q - \delta(\mathbf{w}, q)), \end{aligned} \tag{6.11}$$

where the second step is due to (6.10). Similarly, for the upper bound of $F^{\leftarrow}_{\mathbf{z}^T\mathbf{X}}(p)$, we have

$$\begin{aligned} F^{\leftarrow}_{\mathbf{z}^T\mathbf{X}}(p) &= \inf\left\{x \geq t_0 : F_{\mathbf{z}^T\mathbf{X}}\left(x\right) \geq q\right\} \\ &\leq \inf\left\{x \geq t_0 : F_{\mathbf{w}^T\mathbf{X}}\left(x\right) - \delta(\mathbf{w}, q)/2 \geq q + \delta(\mathbf{w}, q)/2\right\} \\ &= F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q + \delta(\mathbf{w}, q)). \end{aligned} \tag{6.12}$$

Therefore, combining (6.11), (6.12) and (6.8) yields that $F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)$ is continuous in both $\mathbf{w}$ and q because

$$\begin{aligned} \left|F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q) - F^{\leftarrow}_{\mathbf{z}^T\mathbf{X}}\left(p\right)\right| &\leq \left|F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q + \delta(\mathbf{w}, q)) - F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)\right| \vee \left|F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q - \delta(\mathbf{w}, q))) - F^{\leftarrow}_{\mathbf{w}^T\mathbf{X}}(q)\right| \\ &< \varepsilon, \end{aligned}$$

for $\|\mathbf{w} - \mathbf{z}\| < \delta_2$ and $|p - q| \leq \delta(\mathbf{w}, q)/2$. Hence, $l(\mathbf{w}, q)$ is uniform continuous on $\mathbb{S}$. ∎

Now we are ready to show that the convergence in (2.8) is uniform.

Theorem 6.2 *Suppose the nonnegative random vector* $\mathbf{X}$ *has a continuous joint distribution* F*. Further assume that* $\mathbf{X} \in \mathrm{MRV}_\alpha(\Psi)$ *with* $\alpha > 0$*. Then*

$$\lim_{q\uparrow 1} \sup_{\mathbf{w}\in\Sigma^d} \left| \frac{\mathrm{VaR}_q(\mathbf{w}^T\mathbf{X})}{\mathrm{VaR}_q(\|\mathbf{X}\|_1)} - \eta_{\mathbf{w}}^{1/\alpha} \right| = 0. \tag{6.13}$$

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

$$\left|l(\mathbf{w}, q) - \eta_{\mathbf{w}}^{1/\alpha}\right| \leq \left|l(\mathbf{w}, q) - l\left(\mathbf{v}, q\right)\right| + \left|l(\mathbf{v}, q) - \eta_{\mathbf{v}}^{1/\alpha}\right| + \left|\eta_{\mathbf{v}}^{1/\alpha} - \eta_{\mathbf{w}}^{1/\alpha}\right|, \tag{6.14}$$

where $l(\mathbf{w}, q)$ is defined as in (6.4). By properly choosing $\mathbf{v}$, if the three terms can be shown to be arbitrarily small for any $\mathbf{w} \in \Sigma^d$ as q close to 1, then (6.13) is proved. In the following we show how $\mathbf{v}$ can be determined.

By Lemma 6.1 and the uniform continuity of $\eta_{\mathbf{w}}$, for any $\varepsilon > 0$, there exist $\delta > 0$ and $0 < \tilde{q} < 1$ such that for any $\mathbf{w}, \mathbf{z} \in \Sigma^d$ satisfying $\|\mathbf{w} - \mathbf{z}\| < \delta$ and all $q \geq \tilde{q}$, we have

$$|l(\mathbf{w}, q) - l(\mathbf{z}, q)| < \varepsilon. \tag{6.15}$$

and

$$\left|\eta_{\mathbf{w}}^{1/\alpha} - \eta_{\mathbf{z}}^{1/\alpha}\right| < \varepsilon. \tag{6.16}$$

That is, δ is so chosen that both (6.15) and (6.16) hold. Now we are ready to determine $\mathbf{v}$ in (6.14) by constructing open coverings. Let $B_{\mathbf{w},\delta}$ denote the open ball of $\mathbf{w}$; that is $B_{\mathbf{w},\delta} = \{\mathbf{z} \in \Sigma^d : \|\mathbf{w} - \mathbf{z}\| < \delta\}$. Then the collection of all the sets $B_{\mathbf{w},\delta}$ for each $\mathbf{w}$ is an open cover of Σ^d. By the compactness, there exists a finite subcover denoted by $B_{\mathbf{w}_1,\delta}, \ldots, B_{\mathbf{w}_m,\delta}$. For each selected $\mathbf{w}_i$, by the limit relation in (2.8), there exists $0 < q_i < 1$ such that

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

for all $q_i \leq q < 1$. Let $q^* = \max\{\tilde{q}, q_1, \ldots, q_m\}$. For any $\mathbf{w} \in \Sigma^d$, one can find i such that $\mathbf{w} \in B_{\mathbf{w}_i,\delta}$, which means $\|\mathbf{w} - \mathbf{w}_i\| < \delta$. This $\mathbf{w}_i$ is the proper choice of $\mathbf{v}$ in (6.14) since each term on the right-hand side of (6.14) is smaller than ε for all $q^* \leq q < 1$. This completes the proof. ∎

Now we are ready to prove Theorem 2.1.

Proof of Theorem 2.1. Since the convergence $\lim_{q\uparrow 1} \mathrm{VaR}_q(X_i)/\mathrm{VaR}_q(\|\mathbf{X}\|_1) = \eta_{e_i}^{1/\alpha}$ is independent of $\mathbf{w}$, applying Theorem 6.2 to the rewriting in (2.7) we obtain the desired result. ∎

6.3 Proofs of lemmas

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

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

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

For every $\mathbf{w} \in \Sigma^d$, since $\mu \in l_1$, there exists $N_1 > 0$ such that $\mathbf{w}^T\mu < \|\mu\|_1 < N_1$. Since $\|B\|_2 < \infty$, there exists $N_2, N_3 > 0$ such that

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

and

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

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 $|\mathrm{DR}_{\mathbf{w},q} - \mathrm{DR}_{\mathbf{w},1}| < \varepsilon$ for every $q > q_0$ and for every $\mathbf{w}$ in Σ^d follows.

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

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

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

Proof of Lemma 2.3. First note that

$$\begin{aligned}
&\sup_{\mathbf{w}\in\Sigma^d}\left|\hat{\mathrm{DR}}_{\mathbf{w},1}-\mathrm{DR}_{\mathbf{w},1}\right| \\
&=\sup_{\mathbf{w}\in\Sigma^d}\left|\frac{\widehat{\eta}_{\mathbf{w}}^{1/\alpha}\sum_{i=1}^d w_i\eta_{\mathbf{e}_i}^{1/\alpha}-\eta_{\mathbf{w}}^{1/\alpha}\sum_{i=1}^d w_i\widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}}{\left(\sum_{i=1}^d w_i\widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}\right)\left(\sum_{i=1}^d w_i\eta_{\mathbf{e}_i}^{1/\alpha}\right)}\right| \\
&\le \sup_{\mathbf{w}\in\Sigma^d}\left|\frac{\left(\widehat{\eta}_{\mathbf{w}}^{1/\alpha}-\eta_{\mathbf{w}}^{1/\alpha}\right)\sum_{i=1}^d w_i\eta_{\mathbf{e}_i}^{1/\alpha}}{\left(\sum_{i=1}^d w_i\widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}\right)\left(\sum_{i=1}^d w_i\eta_{\mathbf{e}_i}^{1/\alpha}\right)}\right|+\sup_{\mathbf{w}\in\Sigma^d}\left|\frac{\eta_{\mathbf{w}}^{1/\alpha}\left(\sum_{i=1}^d w_i\eta_{\mathbf{e}_i}^{1/\alpha}-\sum_{i=1}^d w_i\widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}\right)}{\left(\sum_{i=1}^d w_i\widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}\right)\left(\sum_{i=1}^d w_i\eta_{\mathbf{e}_i}^{1/\alpha}\right)}\right|. \qquad (6.17)
\end{aligned}$$

Thus, to show that (6.17) converges to 0 almost surely, the key is the strong consistency of $\widehat{\eta}_{\mathbf{w}}$ uniformly in $\mathbf{w}$. This is ensured by Theorem 4.1(a) of Mainik and Rüchendorf [13] if

$$\lim_{q\uparrow 1}\sup_{\mathbf{w}\in\Sigma^d}|\Psi_q f_{\mathbf{w},\alpha}-\Psi f_{\mathbf{w},\alpha}|=0, \qquad (6.18)$$

where Ψ_q is the conditional angular distribution of $S|F_R(R)>q$ for $q\in(0,1)$ and $f_{\mathbf{w},\alpha}(\mathbf{s})=(\mathbf{w}^T\mathbf{s})^\alpha$. Now we show that (6.18) holds under the current conditions. Note that for any $\mathbf{s}\in\Sigma^d$, we have

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

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

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

where in the second step we used the polynomial expansion formula. This means that the function class $\{f_{\mathbf{w},\alpha}:\mathbf{w}\in\Sigma^d\}$ is uniformly Lipschitz for any $\alpha>1$. Then by Remark A.5 of Mainik and Rüchendorf [13], the uniform convergence in (6.18) holds. Hence, $\widehat{\eta}_{\mathbf{w}}$ converges to $\eta_{\mathbf{w}}$ uniformly in $\mathbf{w}\in\Sigma^d$ almost surely. Further, by the continuity of the mapping $\widehat{\eta}_{\mathbf{w}}\longmapsto\widehat{\eta}_{\mathbf{w}}^{1/\alpha}$, we have

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

and

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

Further notice that $\sum_{i=1}^d w_i\eta_{\mathbf{e}_i}^{1/\alpha}$ and $\sum_{i=1}^d w_i\widehat{\eta}_{\mathbf{e}_i}^{1/\alpha}$ 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 (6.17) 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 $S_t=\{(x_1,\dots,x_d)\in\mathbb{R}_+^d:\sum_{i=1}^d w_ix_i\ge t\}$. We can split $\overline{F}_{\mathbf{w}^T\mathbf{X}}(t)$ as

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

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 $w_1X_1^* + \cdots + w_dX_d^*$, 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 [15] 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 [15] leads to

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

Similarly,

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

and

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

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

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

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 [15]. 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

$$|f(m_{f_n}) - f(m_f)| < 2D_n. \tag{6.19}$$

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

$$f(\mathbf{x}) = f(m_f) + \frac{1}{2}(\mathbf{x} - m_f)^T \nabla^2 f(m_f)(\mathbf{x} - m_f) + o\left(||\mathbf{x} - m_f||_2^2\right), \tag{6.20}$$

where we used the multi-index notation and $\nabla^2 f(m_f)$ is the Hessian matrix of f at m_f. Since $D_n \to 0$ as $n \to \infty$ by Theorem 2.1, m_{f_n} is in a small neighborhood of m_f for large n. It then follows from the expansion in (6.20) that

$$|f(m_{f_n}) - f(m_f)| > \frac{c}{2}||m_{f_n} - m_f||_2^2, \tag{6.21}$$

where c is the smallest eigenvalue of $\nabla^2 f(m_f)$. Combining (6.19) and (6.21) leads to the desired result. ■

References

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

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

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

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

[5] 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.

[6] 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.

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

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

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

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

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

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

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

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

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

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

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

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

[19] 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.

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

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

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

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

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

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

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