On optimal portfolio diversification with respect to extreme risks

Georg Mainik Ludger Rüschendorf

February 10, 2009

Abstract

Extreme losses of portfolios with heavy-tailed components are studied in the framework of multivariate regular variation. Asymptotic distributions of extreme portfolio losses are characterized by a functional $\gamma_\xi = \gamma_\xi(\Psi, \alpha)$ of the tail index α, the spectral measure Ψ, and the vector ξ of portfolio weights. Existence, uniqueness, and location of the optimal portfolio are analysed and applied to the minimization of risk measures. It is shown that diversification effects are positive for $\alpha > 1$ and negative for $\alpha < 1$. Strong consistency and asymptotic normality are established for a semiparametric estimator of the mapping $\xi \mapsto \gamma_\xi$. Strong consistency is also established for the estimated optimal portfolio.

1 Introduction

Due to empirical evidence which is based on various data it is generally accepted among practitioners that financial assets often exhibit heavy-tailed behaviour and strong kind of dependence. This is in particular true for data sets related to operational risk, insurance risk, currency exchange rates and stock share prices. In some cases empirical data sets even suggest that the losses have infinite means (cf. Moscadelli [30], Nešlehová, Embrechts, and Chavez-Demoulin [31]). As a result of this insight, there has been a lot of scientific activity on modelling and statistical analysis of heavy-tailed distributions with applications to risk management and portfolio optimization.

The characterization of the probabilistic structure of multivariate extremes in the seminal paper by de Haan and Resnick [11] provided a sound theoretical basis for various approaches to the modelling and the estimation of extremal dependence. In particular, modelling concepts based on multivariate extreme value theory and copulas have been developed for an adequate description and analysis of risks and risk portfolios. Comprehensive elaborations on this topic are given in McNeil, Frey, and Embrechts [29] and Malevergne and Sornette [28]. Further relevant developments on the combination of extremal risks and dependence are based on the notions of tail dependence and tail copulas, multivariate excess distributions, and the empirical distribution of excess directions, see among others Falk et al. [21], Schmidt and Stadtmüller [37], Klüppelberg and Resnick [27], Hauksson et al. [23].

The main subject of the present paper is the comparison of portfolio losses in terms of their sensitivity to extremal events. The asset losses are modelled by a multivariate regularly varying random vector X. It turns out that both the sensitivity to extremal events and the probability distribution of extreme portfolio losses can be characterized by a single number that we therefore call *extreme risk index* of the portfolio. The extreme risk index $\gamma_\xi = \gamma_\xi(\alpha, \Psi)$ is a functional of the vector ξ of portfolio weights and the characteristics of the multivariate regular variation of X, given by the tail index α and the spectral measure Ψ. As a consequence, based on a non-parametric estimator of Ψ and an asymptotically normal estimator of α, we propose an estimator of the extreme risk index that is asymptotically normal under natural conditions. Asymptotic normality is established for the estimator of γ_ξ as a function of the portfolio vector ξ and, under weaker conditions, pointwise in ξ. We also establish strong consistency for the estimator of γ_ξ (both pointwise and uniform in ξ) and for the estimated optimal portfolio. These results extend the empirical study on risk aggregation in Hauksson et al. [23].

There has been a lot of recent interest to the problem of portfolio diversification in heavy-tailed models based on copulas and tail dependence [19, 20, 2, 7, 40, 5]. Risk measures related to the Value-at-Risk were found to exhibit contrary behaviour in models with finite and infinite means: while they are typically sub-additive in finite mean models, they exhibit super-additive behaviour in the case of infinite means. Moreover, in the infinite mean case increasing positive dependence between asset losses decreases the losses of the diversified portfolio, which contradicts the "usual" principles of portfolio diversification. This effect was quantified by the limiting relation

$$q_n(\alpha, \beta) := \lim_{u \to \infty} \mathrm{P}\left\{X_1 + \ldots + X_n > u\right\} / \mathrm{P}\left\{X_1 > u\right\} \tag{1}$$

for identically distributed random variables $X_1, \ldots, X_n$ with tail index α and dependence structure given by an Archimedean survival copula with parameter β measuring the degree of positive dependence, cf. Embrechts et al. [20]. In this setting, q_n is found to be increasing in β for $\alpha \geq 1$ and decreasing in β for $\alpha < 1$.

The extreme risk index γ_ξ extends these results to the case of arbitrary, i.e. not necessarily equally weighted portfolios and not necessarily identical marginal distributions. Moreover, γ_ξ is considered for all dependence structures that are possible in the framework of multivariate regular variation. The extreme risk index γ_ξ turns out to be a convex function of the portfolio vector ξ for $\alpha > 1$ and a concave function for $\alpha < 1$.

This implies that for $\alpha > 1$ diversification typically makes the portfolio better, whereas for $\alpha < 1$ diversification makes the portfolio worse. In both cases, higher degree of positive dependence reduces the diversification effects. An important practical consequence of these findings is the fact that in the case of infinite means diversification of the portfolio over different assets is not necessarily desirable.

Further we study the related statistical problems of estimating the extreme risk index and the optimal portfolio. The statistical analysis of the estimators incorporates the theory of empirical measures and empirical processes indexed by functions. This approach is of interest on its own and can be applied to the estimation of further functionals of the spectral measure Ψ and the tail index α.

The paper is structured as follows. Section 2 provides the statement of the problem and some basics on multivariate regular variation that lead to the definition of the extreme risk index. In Section 3 the extreme risk index γ_ξ is introduced and the properties of the function $\xi \mapsto \gamma_\xi$ are studied. The consequences for portfolio diversification are discussed and illustrated. Section 4 is dedicated to the statistical results and the underlying assumptions, whereas Section 5 contains a brief discussion of applications to risk measures. Finally, the conclusions are drawn in Section 6 and the proofs of statistical results are given in Appendix A.

2 Heavy-tailed portfolios

Let $X^{(1)}, \dots, X^{(d)} \in \mathbb{R}_+$ be the losses of some risky assets and let $\xi \in \mathbb{R}_+^d$ represent the weights of the assets in the portfolio, so that the portfolio loss is given by $\xi^\top X$ with $X := (X^{(1)}, \dots, X^{(d)})^\top$. It is obvious that multiplying the portfolio vector ξ by a constant factor $c > 0$ results in multiplication of the portfolio loss by c. Hence the influence of the portfolio composition on the portfolio loss can be studied by considering standardized portfolios. Following the intuition of dividing the whole capital in parts and investing them in different assets, we standardize portfolio vectors by the sums of their components. As a result, the set of portfolio vectors ξ that we need to consider is the unit simplex in $\mathbb{R}_+^d$:

$$\xi \in \Sigma^d := \left\{ x \in \mathbb{R}_+^d : \|x\|_1 = 1 \right\}.$$

The assets $X^{(i)}$ are assumed to be (univariate) *regularly varying* with *tail index* $\alpha > 0$:

$$\forall x > 0 \quad \frac{\mathrm{P}\{X^{(i)} > tx\}}{\mathrm{P}\{X^{(i)} > t\}} \to x^{-\alpha}, \quad t \to \infty. \tag{2}$$

The tail index α characterizes the existence of absolute moments $\mathrm{E}|X^{(i)}|^\beta$: for $\beta < \alpha$ they exist, whereas for $\beta > \alpha$ they explode.

It is well known that heavier tails dominate the influence of the lighter ones on the extremes, making asymptotic analysis of extreme losses trivial if the tail indices are different. Therefore only the case of equal tail indices is considered here. Moreover, we assume X to be *multivariate regularly varying*, i.e. there exists a sequence $a_n \to \infty$ and a Radon measure ν on $\mathcal{B}([0,\infty]^d \setminus \{0\})$ such that $\nu([0,\infty]^d \setminus \mathbb{R}_+^d) = 0$ and, as $n \to \infty$,

$$n\mathrm{P}^{\,a_n^{-1}X} \xrightarrow{\mathrm{v}} \nu \text{ on } \mathcal{B}([0,\infty]^d \setminus \{0\}), \tag{3}$$

where $\xrightarrow{\mathrm{v}}$ denotes the *vague convergence* of Radon measures and $\mathrm{P}^{\,a_n^{-1}X}$ is the probability distribution of $a_n^{-1}X$. For more technical details related to the Borel σ-field $\mathcal{B}([0,\infty]^d \setminus \{0\})$, vague convergence and multivariate regular variation we refer to Resnick [36]. Additionally to (3) we assume that the limit measure ν is non-degenerate in the following sense:

$$\nu\left(\left\{x \in \mathbb{R}_+^d : x^{(i)} > \varepsilon\right\}\right) > 0 \tag{4}$$

for all $\varepsilon > 0$ and $i = 1, \dots, d$. This assumption ensures that all components $X^{(i)}$ are relevant for the extremes of X.

The measure ν exhibits the scaling property

$$\nu(tA) = t^{-\alpha}\nu(A) \tag{5}$$

for all sets $A \in \mathcal{B}\left([0,\infty]^d \setminus \{0\}\right)$ that are bounded away from 0. Furthermore, for any random vector X satisfying (3) the limit measure ν is unique up to a constant factor. The measure ν also characterizes the asymptotic distribution of the componentwise maxima

$$M_n := \left(M^{(1)}, \dots, M^{(d)}\right), \quad M^{(i)} := \max\left\{X_1^{(i)}, \dots, X_n^{(i)}\right\}$$

by the limit relation

$$\mathrm{P}\left\{a_n^{-1}M_n \in [0,x]\right\} \xrightarrow{\mathrm{w}} \exp\left(-\nu\left(\mathbb{R}_+^d \setminus [0,x]\right)\right), \quad x \in \mathbb{R}_+^d \setminus \{0\}.$$

Therefore ν is called *exponent measure*. For more details and other standardizations of the measure ν see Resnick [35].

Another consequence of the scaling property (5) is the product representation of ν in polar coordinates $\tau(x) := (r, s) := (\|x\|, \|x\|^{-1}x)$ with respect to an arbitrary norm $\|\cdot\|$. The induced measure $\nu^\tau := \nu \circ \tau^{-1}$ necessarily satisfies

$$\nu^\tau(\mathrm{d}r \times \mathrm{d}s) = C \cdot \rho_\alpha(\mathrm{d}r) \otimes \Psi(\mathrm{d}s) \tag{6}$$

with some constant $C > 0$, $\rho_\alpha(x, \infty) = x^{-\alpha}$ and a probability measure Ψ on the set $\{s \in \mathbb{R}_+^d : \|s\| = 1\}$. The measure Ψ is called *spectral measure* of ν or X. Since the term "spectral measure" is already used in other areas, Ψ is also referred to as *angular measure*.

As shown in Basrak et al. [3], multivariate regular variation of the loss vector X with tail index α and the non-degeneracy condition (4) imply univariate regular variation of any portfolio loss $\xi^\top X$ with the same tail index α. This property is also inherited by the norm $\|X\|$. A remarkable detail fact that the converse implication is also true in the sense that univariate regular variation of $\xi^\top X$ for all $\xi \in \mathbb{R}^d$ with $\|\xi\| = 1$ implies multivariate regular variation of the random vector X if the common tail index α is positive and X takes values in $\mathbb{R}_+^d$. This sort of Cramér-Wold theorem was established in Basrak et al. [3] and Boman and Lindskog [6].

Although the domain of the spectral measure Ψ depends on the norm $\|\cdot\|$ used for constructing the polar coordinates, the representation (6) is norm-independent in the following sense: if (6) holds for some norm $\|\cdot\|$, then it also holds for any other norm $\|\cdot\|_*$ that is equivalent to $\|\cdot\|$. The tail index α is the same and the spectral measure Ψ_* on the subset $\{s \in \mathbb{R}_+^d : \|s\|_* = 1\}$ of the unit sphere corresponding to $\|\cdot\|_*$ is obtained from Ψ by the following transformation:

$$\Psi_* = \Psi^T, \quad T(s) := \|s\|_*^{-1}s.$$

In the following we consider polar coordinates based on the sum norm $\|\cdot\|_1$ and set the constant C in (6) to 1, which does not lead to any loss of generality. Moreover, due to this standardization, the multivariate regular variation of X can be equivalently written in terms of weak convergence as

$$\mathcal{L}\left\{t^{-1}X \mid \|X\|_1 > t\right\} \xrightarrow{\mathrm{w}} \nu|_{A_1} \text{ on } \mathcal{B}(A_1) \tag{7}$$

for $t \to \infty$, where $\nu|_{A_1}$ is the restriction of ν to the set $A_1 := \{x \in \mathbb{R}_+^d : \|x\|_1 > 1\}$.

Further details on regular variation of functions or random variables and related applications in extreme value theory can be found in the vast literature on these topics. See among others Bingham et al. [4], Resnick [35], Basrak et al. [3], Hult and Lindskog [25], de Haan and Ferreira [10], Resnick [36].

After the model is specified, an approach to the comparison of extreme portfolio losses is needed. Let us consider events when the portfolio loss $\xi^\top X$ exceeds a high bound $t > 0$. These events can be written as $\{X \in A_{\xi,t}\}$, where

$$A_{\xi,t} := \left\{x \in \mathbb{R}_+^d : \xi^\top x > t\right\}. \tag{8}$$

In order to make the vanishing probabilities $\mathrm{P}\{X \in A_{\xi,t}\}$ comparable for $t \to \infty$ we normalize them by the probability $\mathrm{P}\{X \in A_t\}$, where

$$A_t := \left\{x \in \mathbb{R}_+^d : \|x\|_1 > t\right\}. \tag{9}$$

The sets $A_{\xi,t}$ and A_t can be regarded as different kinds of extremal events: $A_{\xi,t}$ indicates high losses of the portfolio ξ, whereas A_t is a generic extremal event indicating that some components of the vector X produce high losses.

Figure 1: Sets A_1 and $A_{\xi,1}$ in $\mathbb{R}_+^2$ for $\xi = (0.28, 0.72)^\top$

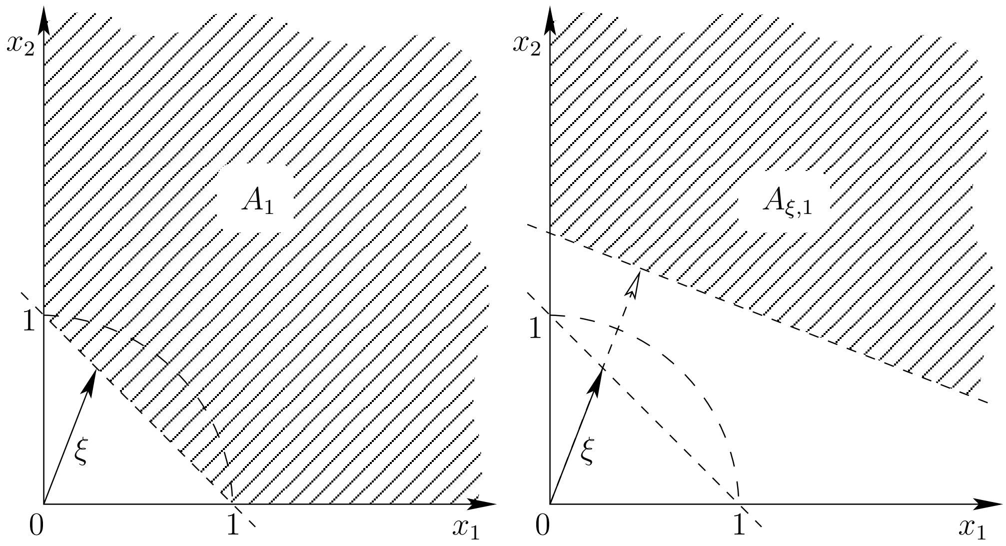


Figure 1 illustrates the sets A_1 and $A_{\xi,1}$ in $\mathbb{R}_+^2$. It should be noted that due to $\|\xi\|_2 \le \|\xi\|_1 = 1$ for all $\xi \in \Sigma^d$ we have

$$\inf_{x \in A_{\xi,1}} \|x\|_2 = \|\xi\|_2^{-1} \ge 1,$$

which means that in the "natural" Euclidean metric on $\mathbb{R}_+^d$ the distance between the set $A_{\xi,1}$ and the origin depends on ξ and is bounded from below by 1. The sets A_t and $A_{\xi,t}$ can be obtained by rescaling:

$$A_t = t \cdot A_1, \quad A_{\xi,t} = t \cdot A_{\xi,1}.$$

It is easy to see that the inequality $\xi^\top x \leq \|x\|_1$ holds for all $\xi \in \Sigma^d$. As a result, we obtain the set inclusion $A_{\xi,t} \subset A_t$ and the representation

$$\frac{\mathrm{P}\{X \in A_{\xi,t}\}}{\mathrm{P}\{X \in A_t\}} = \mathrm{P}\left\{\xi^\top X > t \,|\, \|X\|_1 > t\right\}.$$

Thus the comparison of risks corresponding to portfolios $\xi \in \Sigma^d$ can be reduced to the comparison of conditional probabilities $\mathrm{P}\{X \in A_{\xi,t} | X \in A_t\}$ for large t. According to (7), multivariate regular variation of X yields

$$\begin{aligned} \mathrm{P}\{X \in A_{\xi,t} | X \in A_t\} &= \mathrm{P}\left\{t^{-1}X \in A_{\xi,1} \,|\, \|X\|_1 > t\right\} \\ &\xrightarrow{\mathrm{w}} \nu(A_{\xi,1}). \end{aligned} \tag{10}$$

This means that under the assumption of multivariate regular variation the asymptotic behaviour of portfolio losses can be characterized by the functional

$$\gamma_\xi := \nu(A_{\xi,1}),$$

which quantifies the asymptotic sensitivity of the portfolio ξ to extremal events. Furthermore, for any pair of portfolio vectors $\xi_1, \xi_2 \in \Sigma^d$ relation (10) implies

$$\frac{\mathrm{P}\left\{\xi_1^\top X > t\right\}}{\mathrm{P}\left\{\xi_2^\top X > t\right\}} \to \frac{\gamma_{\xi_1}}{\gamma_{\xi_2}}, \quad t \to \infty,$$

which means that the functional γ_ξ also allows to compare the sensitivity of different portfolios to extremal events.

Moreover, multivariate regular variation of X yields the asymptotic relation of tail probabilities

$$\frac{1 - F_{\xi^\top X}(rt)}{1 - F_{\|X\|_1}(t)} \to \gamma_\xi \cdot r^{-\alpha}, \quad t \to \infty, \tag{11}$$

for all $r > 1$ and (cf. Resnick [35], Proposition 0.8, parts (v) and (vi)) the asymptotic quantile relation

$$\frac{F^{\leftarrow}_{\xi^\top X}(1 - uv)}{F^{\leftarrow}_{\|X\|_1}(1 - v)} \to \gamma_\xi^{1/\alpha} \cdot u^{-1/\alpha}, \quad v \downarrow 0, \tag{12}$$

for all $u \in (0, 1)$. Thus, γ_ξ allows to order both the probabilities of extremal losses and high loss quantiles for all portfolios $\xi \in \Sigma^d$. This means that γ_ξ provides all information that is needed for comparing the influence of the portfolio vector ξ on the severity of extreme losses.

It should also be noted that the scaling relations (11) and (12) allow to estimate probabilities of extremal losses and high loss quantiles and to extrapolate these estimates beyond the observable area. The estimated values can be used in portfolio optimization. An empirical study based on these scaling relations is provided in Hauksson et al. [23].

3 Extreme risk index and portfolio diversification

The results of the previous section justify the following definition.

Definition 3.1. For any portfolio vector $\xi \in \Sigma^d$ the functional

$$\gamma_\xi := \nu\left(A_{\xi,1}\right)$$

is called *extreme risk index* of ξ.

The extreme risk index is a natural way to quantify the influence of the asymptotic dependence structure on extreme portfolio losses in the framework of multivariate regular variation. It complements the available palette of approaches including the coefficient of tail dependence (cf. Joe [26]), the extremal dependence measure (cf. Resnick [34]), the Pickands dependence function (cf. Pickands [33]), and the tail copula (cf. Schmidt and Stadtmüller [37]), which are more focused on applications different from portfolio optimization. Indeed, the coefficient of tail dependence and the extremal dependence measure are very useful for fitting and testing models. However, these functionals are single-number characteristics and therefore they are not able to carry sufficient information about the location of the optimal portfolio. On the other hand, parametrization of dependence structures by dependence functions and tail copulas is based on sets of the form $\mathbb{R}_+^d \setminus [0, x]$ for $x \in \mathbb{R}_+^d$, which are naturally related to simultaneous exceedance of bounds by the components $X^{(i)}$. Thus, the extreme risk index fills the gap for an approach that addresses extremes of portfolio losses directly.

The product structure of the measure ν in polar coordinates yields

$$\begin{aligned}
\gamma_\xi &= \int_{\Sigma^d} \int_{\mathbb{R}_+} 1\left\{\xi^\top \cdot rs > 1\right\} \rho_\alpha(\mathrm{d}r)\Psi(\mathrm{d}s) \\
&= \int_{\Sigma^d} \rho_\alpha \left\{r \in \mathbb{R}_+ : r > 1/\left(\xi^\top s\right)\right\} \Psi(\mathrm{d}s) \\
&= \int_{\Sigma^d} \left(\xi^\top s\right)^\alpha \Psi(\mathrm{d}s). \qquad (13)
\end{aligned}$$

It should be noted that the representation (13) does not depend on the norm $\|\cdot\|$ used for the polar coordinates and the resulting spectral measure $\Psi_{\|\cdot\|}$. However, since the set $A_1 = \{x \in \mathbb{R}_+^d : \|x\| > 1\}$ depends on the norm, setting $\nu(A_{\xi,1}) := 1$ results in rescaling of γ_ξ by a constant factor that depends on the norm and the spectral measure Ψ. A remarkable property of the 1-norm is the fact that the extreme risk index of the equally weighted portfolio does not depend on the spectral measure:

$$\gamma_{d^{-1}(1,\ldots,1)} = \int_{\Sigma^d} \left(d^{-1}\left(s^{(1)} + \ldots + s^{(d)}\right)\right)^{\alpha} \Psi(\mathrm{d}s) = d^{-\alpha}.$$

Now let us consider the problem of finding the portfolio with lowest sensitivity to extremal events. As already shown before, this sort of riskiness is measured by γ_ξ. Therefore we need to minimize the function $\xi \mapsto \gamma_\xi$. The resulting optimization problem is analysed in the following lemma.

Lemma 3.2. *(a) For $\alpha > 1$ the mapping $\xi \mapsto \gamma_\xi$ is convex. The convexity is strict if Ψ does not concentrate the entire mass on a linear subspace of Σ^d.*

(b) For $\alpha = 1$ the mapping $\xi \mapsto \gamma_\xi$ is linear.

(c) For $\alpha \in (0,1)$ the mapping $\xi \mapsto \gamma_\xi$ is concave. The concavity is strict if Ψ does not concentrate the entire mass on a linear subspace of Σ^d.

Proof. Part (a). The convexity of $\xi \mapsto \gamma_\xi$ follows from the convexity of $t \mapsto t^\alpha$ for $t > 0$ and $\alpha \geq 1$. Given $\lambda \in (0,1)$ and $\xi_1, \xi_2 \in \Sigma^d$, we immediately obtain

$$\begin{aligned}
\lambda\gamma_{\xi_1} + (1-\lambda)\gamma_{\xi_2} &= \int \left(\lambda\left(\xi_1^\top s\right)^\alpha + (1-\lambda)\left(\xi_2^\top s\right)^\alpha\right) \Psi(\mathrm{d}s) \\
&\leq \int \left(\lambda\xi_1^\top s + (1-\lambda)\xi_2^\top s\right)^\alpha \Psi(\mathrm{d}s) \\
&= \gamma_{\lambda\xi_1 + (1-\lambda)\xi_2}.
\end{aligned}$$

Strict convexity holds if the upper inequality is strict, i.e. if

$$\int \left(\lambda\left(\xi_1^\top s\right)^\alpha + (1-\lambda)\left(\xi_2^\top s\right)^\alpha\right) \Psi(\mathrm{d}s) < \int \left(\lambda\xi_1^\top s + (1-\lambda)\xi_2^\top s\right)^\alpha \Psi(\mathrm{d}s)$$

for all $\xi_1, \xi_2 \in \Sigma^d$ such that $\xi_1 \neq \xi_2$. Since the mapping $t \mapsto t^\alpha$ is strictly convex for $\alpha > 1$, equality holds only if $\xi_1^\top s = \xi_2^\top s$ almost sure with respect to Ψ. This can also be written as

$$\Psi\left\{s \in \Sigma^d : (\xi_1 - \xi_2)^\top s = 0\right\} = 1,$$

which exactly means that the entire probability mass of Ψ is concentrated on $\Sigma^d \cap (\xi_1 - \xi_2)^\perp$.

Part (b) is trivial since for $\alpha = 1$ the mapping $t \mapsto t^\alpha$ is linear and the mapping $\xi \mapsto \gamma_\xi$ is therefore a composition of linear mappings.

Part (c) is analogous to part (a) due to the strict concavity of $t \mapsto t^\alpha$ for $\alpha \in (0,1)$. $\square$

Consequently, the location of the optimal portfolio

$$\xi^{\text{opt}} := \operatorname*{argmin}_{\xi \in \Sigma^d} \gamma_\xi$$

can be described as follows:

- For $\alpha > 1$ the typical location of ξ^{opt} would be in the interior of Σ^d. The optimal portfolio is unique if there is no mass concentration on linear subspaces under Ψ.

- For $\alpha \leq 1$ the minimum of γ_ξ is achieved in a vertex of Σ^d, i.e. we have

$$\min_{\xi \in \Sigma^d} \gamma_\xi = \min_{i=1,\ldots,d} \gamma_{e_i} \tag{14}$$

  with e_i denoting the ith unit vector.

Graphic examples for these facts are given in Figures 2 and 3 with discrete spectral measures $\Psi(w)$ defined by

$$\Psi(w) := \sum_{i=1}^{n(w)} w^{(i)} \delta_{(i-1,\, n(w)-i)/(n(w)-1)}, \tag{15}$$

where w is a vector of weights and $n(w)$ is the size w. There are two major reasons for using discrete spectral measures here. The first one is the easy construction of illustrative examples and the second one is the fact that empirical estimators are discrete. Moreover, graphics based on spectral measures with densities do not exhibit any specific properties.

The results of Lemma 3.2 and the conclusions above have an interesting consequence: if only the losses are accounted, then portfolio diversification does not reduce the danger of extreme losses in the case $\alpha \in (0,1]$. Moreover, for $\alpha < 1$ portfolio diversification typically increases extreme risks. The representation $\gamma_\xi = \int (\xi^\top s)^\alpha \Psi(\mathrm{d}s)$ suggests that these negative effects are stronger in the case of low positive dependence, i.e. when Ψ concentrates the probability mass around the vertices of the unit simplex Σ^d. Analogously, for $\alpha > 1$ low positive dependence makes positive diversification effects stronger. This is illustrated in Figure 4, where γ_ξ is plotted for symmetric 3-point spectral measures Ψ_λ,

$$\Psi_\lambda := \lambda \cdot \delta_{(\frac{1}{2},\frac{1}{2})} + \frac{1}{2}(1-\lambda) \cdot \left(\delta_{(1,0)} + \delta_{(0,1)}\right), \quad \lambda \in [0,1], \tag{16}$$

with parameter λ quantifying the degree of positive dependence.

Figure 2: Left: density of the spectral measure $\Psi(w)$ defined as in (15) for $w = \frac{1}{12}(3, 2, 4, 1, 2)$. Right: resulting extreme risk index $\gamma_\xi(\Psi(w), \alpha)$ and the optimal portfolios (vertical lines) for selected values of α between 2 and 4.

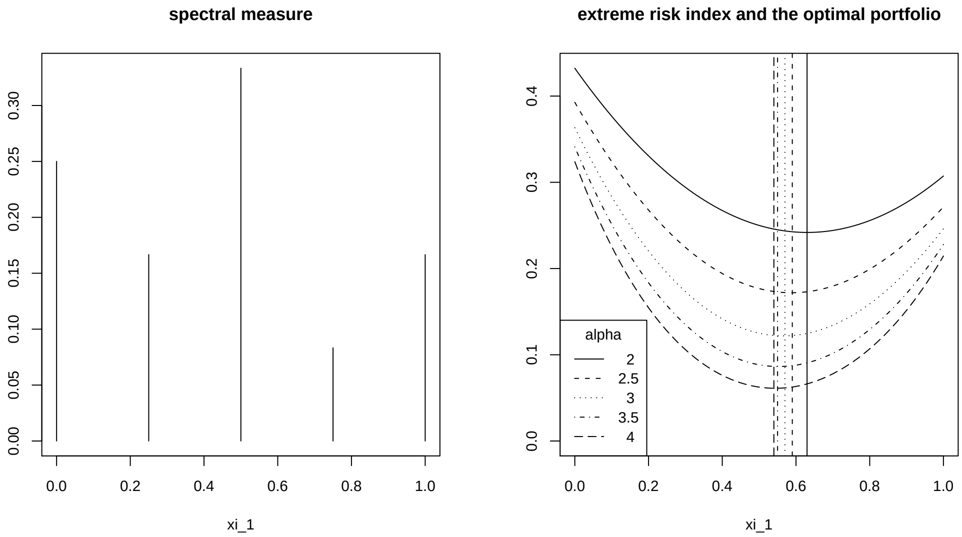
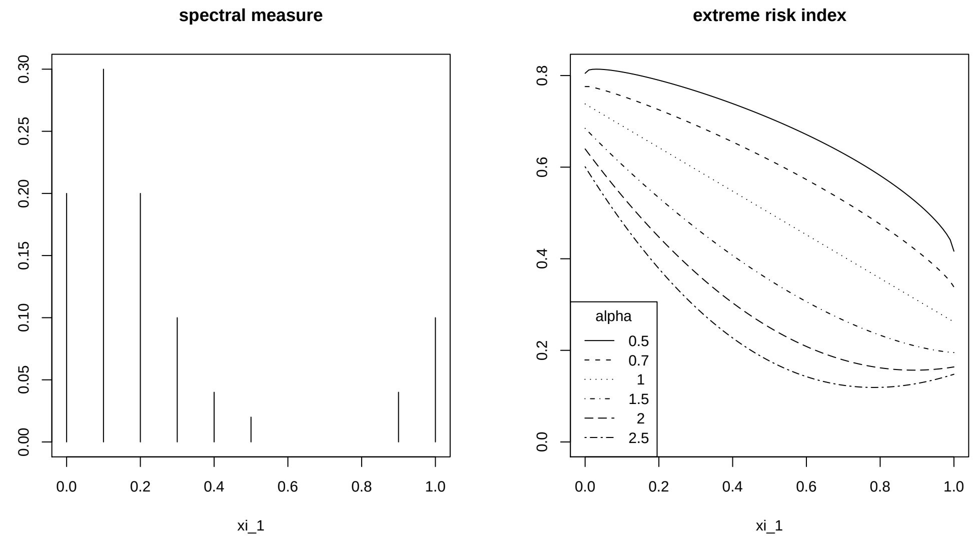


Figure 4 shows that for $d = 2$ and $\alpha > 1$ best diversification effects are achieved if $X^{(1)}$ and $X^{(2)}$ are asymptotically independent, i.e. if $\lambda = 0$, whereas the worst case is $\lambda = 1$, which corresponds to the comonotonic distribution of asset losses. While this behaviour accords with the usual intuition of diversification effects, in the case $\alpha < 1$ the situation is just the opposite. It turns out that for $\alpha < 1$ diversification effects are negative or zero and that the asymptotic independence of asset losses is the worst case for the uniformly diversified portfolio $\xi = (\frac{1}{2}, \frac{1}{2})$, whereas the comonotonic distribution is the best case.

We see that the comonotonic distribution just removes all diversification effects: the positive ones for $\alpha > 1$ and the negative ones for $\alpha < 1$. This implies that in the case of infinite means sensitivity to extremal events can only be optimized by minimizing the number of uncertainty sources and not by diversification.

This remarkable property has been repeatedly observed in settings similar to (1) and vividly discussed in the recent literature [19, 20, 2, 7, 40]. It should be noted that the clear evidence of negative diversification effects for $\alpha < 1$ is restricted to models that account only the asset losses. If the profits are also incorporated, i.e. if $X^{(i)}$ can take positive and negative values as well, then a countermonotonic distribution of $(X^{(1)}, X^{(2)})$ leads to the compensation

Figure 3: Left: density of the spectral measure $\Psi(w)$ defined as in (15) for $w = \frac{1}{50}(10, 15, 10, 5, 2, 1, 0, 0, 0, 2, 5)$. Right: resulting extreme risk index $\gamma_\xi(\Psi(w), \alpha)$ for selected values of α between 0.5 and 2.5.



of losses from one component by the profits from the other one, so that diversification effects may become positive again.

Negative diversification effects in infinite mean models were already noticed in the beginnings of Probability Theory. If, for example, $X_1, \ldots, X_n$ are i.i.d. α-stable random variables, then

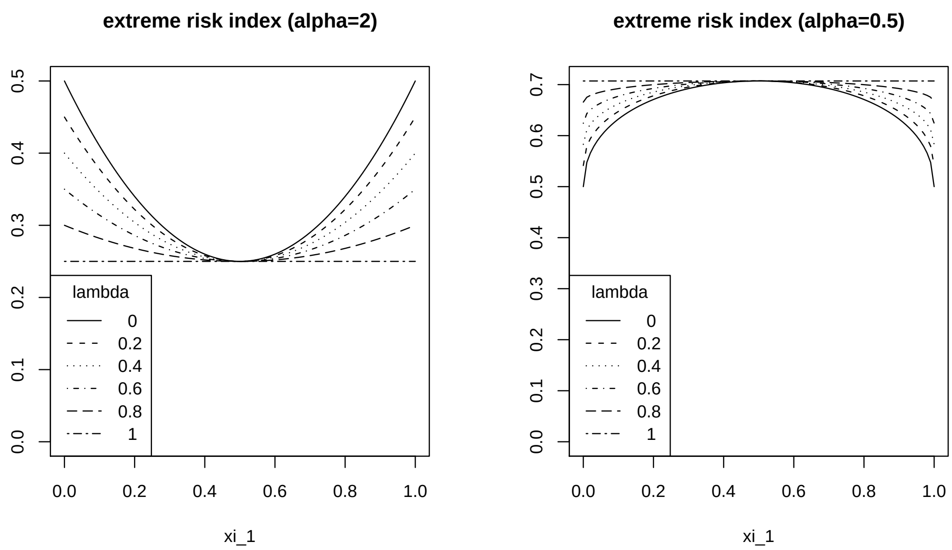
$$n^{-1}(X_1 + \ldots + X_n) \stackrel{\mathrm{d}}{=} n^{(1/\alpha)-1} X_1,$$

which implies negative diversification effects for $\alpha < 1$. A general result on negative diversification effects in similar settings can be obtained from the Marcinkievicz–Zygmund Strong Law of Large Numbers for i.i.d. random variables, cf. Nešlehová et al. [31].

4 Estimation of the extreme risk index and the optimal portfolio

In the following, let X be a multivariate regularly varying random variable and let $X_1, \ldots, X_n$ be an i.i.d. sample of X. Our aim is the estimation of the extreme risk index γ_ξ and the optimal portfolio ξ^{opt}. The representation (13) suggests the following plug-in approach:

Figure 4: Influence of dependence on the extreme risk index for $\alpha > 1$ (left) and $\alpha < 1$ (right) with underlying spectral measures Ψ_λ defined in (16)



1. Estimate the tail index α by an estimator $\hat{\alpha}$.

2. Estimate the spectral measure Ψ by an estimator $\hat{\Psi}$.

3. Estimate γ_ξ by
$$\hat{\gamma}_\xi := \int_{\Sigma^d} \left(\xi^\top s\right)^{\hat{\alpha}} \hat{\Psi}(\mathrm{d}s). \tag{17}$$

4. Obtain an estimate for the optimal portfolio by minimizing $\hat{\gamma}_\xi$:
$$\hat{\xi}^{\mathrm{opt}} := \operatorname*{argmin}_{\xi \in \Sigma^d} \hat{\gamma}_\xi. \tag{18}$$

Since $\hat{\gamma}_\xi$ is obtained by plugging $\hat{\Psi}$ and $\hat{\alpha}$ into the representation (13), the minimization problem for $\hat{\gamma}_\xi$ has the same properties as for γ_ξ and is characterized by Lemma 3.2. For $\hat{\alpha} \leq 1$ the minimization is simplified by (14).

Although the estimation and optimization procedures can be done by approved methods, their result is not trivial. It must be assured that the solutions of the approximating problems yield sensible approximations for both the optimal argument ξ^{opt} and the optimal value $\gamma_{\xi^{\mathrm{opt}}}$.

These results will be obtained from the strong consistency of $\hat{\gamma}_\xi$ uniformly in ξ. Furthermore, we establish asymptotic normality (AN) of $\hat{\gamma}_\xi$ as a function of ξ. The uniform strong consistency and the uniform AN property of $\hat{\gamma}_\xi$ are the central statistical results of the present paper. They are based on well-known consistency and AN results for estimators of α and incorporate the theory of empirical measures and empirical processes indexed by functions.

Before formulating the estimators and stating the central results, some notation is needed. Let (R, S) and (R_i, S_i) denote the polar coordinates of X and X_i with respect to the 1-norm:

$$(R, S) := \left(\|X\|_1, \|X\|_1^{-1} X\right), \quad (R_i, S_i) := \left(\|X_i\|_1, \|X_i\|_1^{-1} X_i\right).$$

In order to avoid technical difficulties we assume that the distribution function of the radial parts is continuous:

$$F_R(t) := \mathrm{P}\{R \leq t\} \in \mathcal{C}(\mathbb{R}).$$

Since this assumption is fulfilled in common applications and models, this restriction is not problematic.

Further we denote by $k = k(n)$ the number of the observations in the sample $X_1, \ldots, X_n$ that the estimates of tail related parameters are based on. These are the observations with highest absolute values, i.e. the ones associated with the k upper order statistics $R_{n:1}, \ldots, R_{n:k}$ of the radial parts $R_1, \ldots, R_n$. The growth of k is linked to n by the following assumption:

$$k(n) \to \infty, \quad \frac{k(n)}{n} \to 0.$$

Let $i(n, 1), \ldots, i(n, k)$ denote the indices corresponding to the k observations with greatest values of R_i, ordered as they appear in the sample. Then we have

$$1 \leq i(n, 1) < \ldots < i(n, k) \leq n$$

and there exists a permutation π of the tuple $(1, \ldots, k)$ such that

$$\left(R_{i(n,1)}, \ldots, R_{i(n,k)}\right) = \left(R_{n:\pi(1)}, \ldots, R_{n:\pi(k)}\right). \tag{19}$$

The subsample $X_{i(n,1)}, \ldots, X_{i(n,k)}$ contains all information that is needed for estimating γ_ξ. By (13), γ_ξ can be written as

$$\Psi f := \int_{\Sigma^d} f(s)\Psi(\mathrm{d}s), \tag{20}$$

where

$$f(s) := f_{\xi,\alpha}(s) := \left(\xi^\top s\right)^\alpha.$$

The function f is estimated by

$$\hat{f} := f_{\xi,\hat{\alpha}}$$

with an estimator $\hat{\alpha}$ obtained from the upper order statistics of radial parts,

$$\hat{\alpha} = \hat{\alpha}\left(R_{n:1}, \ldots, R_{n:k}\right), \tag{21}$$

which can be based on various approaches (cf. Hill [24], Pickands [32], Smith [38], Dekkers et al. [13]). The spectral measure Ψ is estimated by the empirical measure of the angular parts $S_{i(n,1)}, \ldots, S_{i(n,k)}$:

$$\hat{\Psi} := \mathbb{P}_n := \frac{1}{k} \sum_{j=1}^{k} \delta_{S_{i(n,j)}}. \tag{22}$$

There is vast literature on the estimation of the exponent measure ν and the spectral measure Ψ, incorporating methods based on convergence of point processes (cf. de Haan and Resnick [9]) and empirical processes (cf. Einmahl et al. [16, 17, 18], de Haan and Sinha [12], Schmidt and Stadtmüller [37]). However, there is no reference that would cover the asymptotic behaviour of $\hat{\gamma}_\xi$. Indeed, estimation of the exponent measure ν on sets related to portfolio losses is studied only in [12] for estimators that are essentially different from $\hat{\gamma}_\xi$. Moreover, estimation of Ψ with respect to function classes containing functions $f_{\xi,\alpha}$ and uniform consistency, which is needed in portfolio optimization have not been considered so far.

The following theorem states strong consistency of $\hat{\gamma}_\xi$ uniformly in $\xi \in \Sigma^d$ and, under weaker conditions, pointwise in ξ.

Theorem 4.1. *Let $X_1, \ldots, X_n$ be i.i.d. multivariate regularly varying random variables with tail index $\alpha \in (0, \infty)$ and spectral measure Ψ and assume that the distribution function F_R of the radial parts is continuous.*

(a) If the estimator $\hat{\alpha}$ is consistent almost surely,

$$\hat{\alpha} \to \alpha \text{ P-}a.s., \tag{23}$$

and

$$\sup_{\xi \in \Sigma^d} \left| \mathrm{E}\hat{\Psi} f_{\xi,\alpha} - \Psi f_{\xi,\alpha} \right| \to 0, \tag{24}$$

then the estimator $\hat{\gamma}_\xi$ is consistent uniformly in $\xi \in \Sigma^d$ almost surely:

$$\sup_{\xi \in \Sigma^d} |\hat{\gamma}_\xi - \gamma_\xi| \to 0 \quad \text{P-}a.s.$$

(b) If only (23) is satisfied, then the almost sure consistency of $\hat{\gamma}_\xi$ holds pointwise:

$$\forall \xi \in \Sigma^d \quad |\hat{\gamma}_\xi - \gamma_\xi| \to 0 \quad \text{P-}a.s.$$

Remark 4.2. Since the functions $f_{\xi,\alpha}$ are bounded by 1 and for any fixed $\alpha \geq 1$ the function class $\{f_{\xi,\alpha} : \xi \in \Sigma^d\}$ is uniformly Lipschitz, condition (24) is satisfied for any $\alpha \geq 1$. See Remark A.5 for more details.

It is well known that uniform convergence of functions implies convergence of their minima to the minimum of the limit function in the case when the limit function has a unique minimum. Hence, as a consequence of Theorem 4.1, we obtain the following result.

Corollary 4.3. *Suppose that the conditions of Theorem 4.1(a) are satisfied and the optimal portfolio ξ^{opt} is unique. Then the estimator $\hat{\xi}^{\text{opt}}$ and the estimated optimal value $\hat{\gamma}_{\hat{\xi}^{\text{opt}}}$ are consistent almost surely:*

$$\hat{\xi}^{\text{opt}} \to \xi^{\text{opt}} \text{ P-}a.s., \quad \hat{\gamma}_{\hat{\xi}^{\text{opt}}} \to \gamma_{\xi^{\text{opt}}} \text{ P-}a.s.$$

It was already noted that condition (24) is satisfied for $\alpha \geq 1$, which makes the applications easier in this case. However, since for $\alpha \leq 1$ the optimization problem can be reduced to the minimization of $\hat{\gamma}_\xi$ in the vertices of Σ^d, condition (24) is crucial only for applications where there is no clear evidence for $\alpha > 1$ or $\alpha \leq 1$.

The rest of this section is dedicated to the asymptotic normality (AN) results for the estimator $\hat{\gamma}_\xi$, which are are based on the following assumption.

Condition 4.4. *At least one of the following assumptions is fulfilled:*

(a) The tail index α is positive and the spectral measure Ψ has no mass on the boundary of Σ^d:

$$\alpha \in (0, \infty), \quad \Psi\left(\partial\Sigma^d\right) = 0.$$

(b) The tail index α is not smaller than 1 bounded from above:

$$\alpha \in [1, \alpha^*], \quad \alpha^* < \infty.$$

The next theorem states the AN result in a process version with index $\xi \in \Sigma^d$ and, under weaker conditions, pointwise in ξ. For a definition of the Brownian bridge on a function class we refer to the formulation of the Donsker property (39) in Appendix A.

Theorem 4.5. *Let $X_1, \ldots, X_n$ be i.i.d. multivariate regularly varying random variables with tail index α and spectral measure Ψsatisfying Condition 4.4. Further assume that the distribution function F_R of the radial parts is continuous.*

(a) Suppose that the estimator $\hat{\alpha}$ is asymptotically normal,

$$\sqrt{k}\,(\hat{\alpha}-\alpha) \xrightarrow{\mathrm{w}} Y \sim \mathcal{N}\left(\mu_\alpha, \sigma_\alpha^2\right), \tag{25}$$

and that there exists a mapping $b \in l^\infty(\Sigma^d)$ such that

$$\sqrt{k}(\mathrm{E}\hat{\Psi} f_{\xi,\alpha} - \Psi f_{\xi,\alpha}) \to b(\xi) \quad \text{in } l^\infty\left(\Sigma^d\right). \tag{26}$$

Then

$$\sqrt{k}\,(\hat{\gamma}_\xi - \gamma_\xi) \xrightarrow{\mathrm{w}} \mathbb{G}_\Psi f_{\xi,\alpha} + b(\xi) + c_{\xi,\alpha} Y \quad \text{in } l^\infty(\Sigma^d), \tag{27}$$

where $\mathbb{G}_\Psi$ is a Brownian bridge on the function class $\left\{f_{\xi,\alpha} : \xi \in \Sigma^d\right\}$ "with time" Ψ, $b(\xi)$ is the asymptotic bias term from (26), Y is a Gaussian random variable which is independent from $\mathbb{G}_\Psi$ and distributed according to (25), and $c_{\xi,\alpha}$ is given by

$$c_{\xi,\alpha} = \int_{\Sigma^d} \left(\xi^\top s\right)^\alpha \log\left(\xi^\top s\right) \Psi(\mathrm{d}s).$$

(b) Suppose that (25) is satisfied and that

$$\sqrt{k}(\mathrm{E}\hat{\Psi} f_{\xi_i,\alpha} - \Psi f_{\xi_i,\alpha}) \to b(\xi_i) \in \mathbb{R} \tag{28}$$

holds for $\xi_1, \ldots, \xi_p \in \Sigma^d$. Then

$$\sqrt{k}\left(\hat{\gamma}_{\xi_1} - \gamma_{\xi_1}, \ldots, \hat{\gamma}_{\xi_p} - \gamma_{\xi_p}\right) \xrightarrow{\mathrm{w}} \mathcal{N}\left(\mu(\alpha, \xi_1, \ldots, \xi_p), \sigma(\alpha, \xi_1, \ldots, \xi_p)\right) \tag{29}$$

for all $\alpha \geq 0$. The expectations $\mu^{(i)}(\alpha, \xi_1, \ldots, \xi_p)$ are given by

$$\mu^{(i)} = b(\xi_i) + c_{\alpha,\xi_i}\mu_\alpha, \quad i = 1, \ldots, p,$$

and the covariances $\sigma_{i,j}(\alpha, \xi_1, \ldots, \xi_p)$ are equal to

$$\Psi\left[\left(f_{\xi_i,\alpha} - \Psi f_{\xi_i,\alpha}\right)\left(f_{\xi_j,\alpha} - \Psi f_{\xi_j,\alpha}\right)\right] + c_{\xi_i,\alpha} c_{\xi_j,\alpha} \sigma_\alpha,$$

where μ_α is the mean and σ_α^2 is the variance of the random variable Y in (25).

It is well known that the estimators of α mentioned above the statement of Theorem 4.1 are asymptotically normal under appropriate conditions specifying convergence rate of the distribution $\mathcal{L}(t^{-1}R | R > t)$ for $t \to \infty$. A comprehensive elaboration on this topic is given in de Haan and Ferreira [10]. For original results see (among others) Davis and Resnick [8], Drees [14], Dekkers et al. [13], Smith [38] and Drees et al. [15].

Condition (26) can be understood as a second order condition related to the weak convergence of the angular parts $S_{i(n,1)}, \ldots, S_{i(n,k)}$. Since multivariate regular variation leaves convergence rates completely unspecified, similar conditions are necessary for establishing asymptotic normality in regularly varying models.

5 Applications to risk minimization

This section is dedicated to the application of the extreme risk index γ_ξ and its estimates to risk minimization. We show that the portfolio ξ^{opt} obtained by minimization of γ_ξ is asymptotically optimal with respect to some well-known and rather natural *risk measures*.

The ordering of large quantiles of the portfolio loss $\xi^\top X$ by γ_ξ (cf. (12)) has immediate consequences on risk measures such as the *Value-at-Risk* (VaR) and the *Expected Shortfall* (ES). Recall the definition of VaR and ES for an asset loss Y with distribution function F_Y. The Value-at-Risk at the confidence level $1-\lambda$ for (typically small) $\lambda \in (0,1)$ is defined by

$$\mathrm{VaR}_{1-\lambda}(Y) := F_Y^{\leftarrow}(1-\lambda)$$

and the Expected Shortfall at the confidence level $1-\lambda$ is defined by

$$\mathrm{ES}_{1-\lambda}(Y) := \mathrm{E}\left[Y | Y > \mathrm{VaR}_{1-\lambda}\right],$$

if the expectation exists. If F_Y is continuous, then the Expected Shortfall can be represented as

$$\mathrm{ES}_{1-\lambda}(Y) = \frac{1}{\lambda}\int_{1-\lambda}^{1} F_Y^{\leftarrow}(u)\mathrm{d}u. \tag{30}$$

Now let us consider $\mathrm{VaR}_{1-\lambda}(\xi^\top X)$ and $\mathrm{ES}_{1-\lambda}(\xi^\top X)$ for $\lambda \downarrow 0$ when the loss vector X with non-negative components is multivariate regularly varying. Due to (12) we obtain that ξ^{opt} is the optimal portfolio if we want to minimize $\mathrm{VaR}_{1-\lambda}$ with respect to extreme risks, i.e. for $\lambda \downarrow 0$. Moreover, for $\alpha > 1$ we have

$$\lim_{\lambda\downarrow 0} \frac{\mathrm{ES}_{1-\lambda}(\xi^\top X)}{\mathrm{VaR}_{1-\lambda}(\xi^\top X)} = \frac{\alpha}{\alpha-1}. \tag{31}$$

This asymptotic relation is a consequence of the Karamata theorem (cf. Proposition 1.5.10 in Bingham et al. [4]). Consequently, ξ^{opt} also minimizes $\mathrm{ES}_{1-\lambda}(\xi^\top X)$ for $\lambda \downarrow 0$.

The asymptotic result (31) can be generalized to the class of *spectral risk measures*. Spectral risk measures were introduced in Acerbi [1] as weighted averages of loss quantiles,

$$M_\phi(Y) := \int_0^1 F_Y^{-1}(p)\phi(p)\mathrm{d}p,$$

where the weight function $\phi : [0,1] \to \mathbb{R}$ is an *admissible risk spectrum*, i.e. it is non-negative, non-decreasing, and satisfies $\int_0^1 \phi(p)\mathrm{d}p = 1$.

As a consequence of (30), $\mathrm{ES}_{1-\lambda}$ is a spectral risk measure. Thus (31) can be viewed as a limit relation for the rescaled and properly normalized risk spectrum. Analogously, for any admissible risk spectrum $\phi_1 : [0,1] \to \mathbb{R}$ the transformations $\tau_\lambda : u \mapsto 1 - \lambda^{-1}(1-u)$ for $\lambda \in (0,1)$ and $u \in [1-\lambda, 1]$ induce a family of rescaled admissible risk spectra ϕ_λ defined by

$$\begin{aligned} \phi_\lambda(u) &= \tau'_\lambda(u)\phi_1(\tau_\lambda(u)) \cdot 1_{[0,1]}(\tau_\lambda(u)) \\ &= \lambda^{-1}\phi_1\left(1 - \lambda^{-1}(1-u)\right) \cdot 1_{[1-\lambda,1]}(u). \end{aligned} \tag{32}$$

This notion leads to the following generalization of (31).

Lemma 5.1. *Let Y be a continuously distributed random variable on $\mathbb{R}_+$ and suppose that Y is regularly varying with tail index $\alpha > 1$. Further let ϕ_λ be admissible risk spectra defined in (32) with ϕ_1 satisfying*

$$\forall t \in (1,\infty) \quad \phi_1(1-1/t) \le K \cdot t^{-1/\alpha+1-\varepsilon} \tag{33}$$

for some $K > 0$ and $\varepsilon > 0$. Then

$$\lim_{\lambda\downarrow 0} \frac{M_{\phi_\lambda}(Y)}{\mathrm{VaR}_{1-\lambda}(Y)} = \int_1^\infty t^{1/\alpha-2}\phi_1(1-1/t)\,\mathrm{d}t. \tag{34}$$

Proof. We have

$$M_{\phi_\lambda}(Y) = \int_{1-\lambda}^1 F_Y^{\leftarrow}(u)\phi_\lambda(u)\mathrm{d}u.$$

Applying (32) and substituting $u = \tau_\lambda^{-1}(1-1/t) = 1-\lambda/t$, we obtain

$$M_{\phi_\lambda}(Y) = \int_1^\infty F_Y^{\leftarrow}(1-\lambda/t)\phi_1(1-1/t)t^{-2}\mathrm{d}t$$

and, denoting $g_\lambda(t) := F_R^{\leftarrow}(1-\lambda/t)/F_R^{\leftarrow}(1-\lambda)$,

$$\frac{M_{\phi_\lambda}(Y)}{\mathrm{VaR}_{1-\lambda}(Y)} = \int_1^\infty g_\lambda(t)t^{-2}\phi_1(1-1/t)\,\mathrm{d}t. \tag{35}$$

Since regular variation of Y with tail index α implies regular variation of the function $t \mapsto F^{\leftarrow}(1-1/t)$ with index $1/\alpha$, we have $g_\lambda(t) \to t^{1/\alpha}$ for $\lambda \downarrow 0$ pointwise in $t \in [1,\infty)$ and the integrand in (35) converges pointwise to the integrand in (34). Moreover, (33) implies

$$g_\lambda(t)t^{-2}\phi_1(1-1/t) \le K \cdot t^{-1/\alpha-\varepsilon/2}g_\lambda(t) \cdot t^{-1-\varepsilon/2}$$

and the uniform convergence theorem for regularly varying functions (cf. Theorem 1.5.2 in Bingham et al. [4]) yields $t^{-1/\alpha-\varepsilon/2}g_\lambda(t) \to t^{-\varepsilon/2}$ for $\lambda \downarrow 0$ uniformly in $t \in [1,\infty)$. Finally, since $t^{-\varepsilon/2}$ is bounded for $t \ge 1$ and $t^{-1-\varepsilon/2}$ is integrable between 1 and ∞, there exists an integrable bound for $g_\lambda(t)t^{-2}\phi_1(1-1/t)$ on $[1,\infty)$ and the dominated convergence theorem completes the proof. □

As a consequence, we obtain that minimization of $M_{\phi_\lambda}(\xi^\top X)$ for $\lambda \downarrow 0$ can be reduced to the minimization of γ_ξ.

6 Conclusions

The problem of characterizing the distribution tails of portfolio losses is solved in the case when the asset losses are non-negative and multivariate regularly varying. Both the sensitivity of portfolios to extremal events and the severity of extreme losses are characterized by the extreme risk index $\gamma_\xi = \gamma_\xi(\Psi, \alpha)$. The problem of portfolio optimization with respect to extreme losses is solved and available results on negative diversification effects for $\alpha < 1$ are generalized to multivariate regularly varying models with non-negative components. The extreme risk γ_ξ is applied to the asymptotic minimization of the Value-at-Risk, Expected Shortfall, and general spectral risk measures. A semi-parametric approach to the estimation of the mapping $\xi \mapsto \gamma_\xi$ is proposed and the resulting estimators are proved to be strong consistent and asymptotically normal under natural conditions. As a result, strong consistency of the estimated optimal portfolio is obtained.

A Proofs

As already mentioned above, the estimator $\hat{\gamma}_\xi$ can be written as

$$\hat{\gamma}_\xi = \mathbb{P}_n \hat{f} := \int_{\Sigma^d} \hat{f}(s)\mathbb{P}_n(\mathrm{d}s),$$

where $\hat{f} = f_{\xi,\hat{\alpha}}$ and $\mathbb{P}_n$ is the empirical measure of the subsample $S_{i(n,1)}, \ldots, S_{i(n,k)}$. Therefore it is natural to study $\hat{\gamma}_\xi$ in the framework of empirical measures indexed by functions. The strong consistency and the asymptotic normality of $\hat{\gamma}_\xi$ can be viewed as special versions of the *Glivenko–Cantelli* and the *Donsker* theorems (cf. van der Vaart and Wellner [39] and references therein).

Let $\mathbb{P}_{k,\Psi}$ denote the empirical measure corresponding to k i.i.d. random variables with probability distribution Ψ:

$$\mathbb{P}_{k,\Psi} := \frac{1}{k}\sum_{i=1}^{k} \delta_{Y_i}, \quad Y_1, \ldots, Y_k \text{ i.i.d.} \sim \Psi. \tag{36}$$

A function class $\mathcal{F}$ is called *Glivenko–Cantelli* if the Glivenko–Cantelli theorem holds for $\mathbb{P}_{k,\Psi}$ uniformly in $f \in \mathcal{F}$:

$$\mathbb{P}_{k,\Psi} \to \Psi \text{ P-a.s. in } l^\infty(\mathcal{F}), \quad k \to \infty. \tag{37}$$

Let $\mathbb{G}_{k,\Psi}$ denote the empirical process corresponding to $\mathbb{P}_{k,\Psi}$:

$$\mathbb{G}_{k,\Psi} := \sqrt{k}\left(\mathbb{P}_{k,\Psi} - \Psi\right). \tag{38}$$

A function class $\mathcal{F}$ is called *Donsker* if the Donsker theorem holds for $\mathbb{G}_{k,\Psi}$ uniformly in $f \in \mathcal{F}$,

$$\mathbb{G}_{k,\Psi} \xrightarrow{\mathrm{w}} \mathbb{G}_{\Psi} \text{ in } l^{\infty}(\mathcal{F}), \quad k \to \infty, \tag{39}$$

where $\mathbb{G}_{\Psi}$ is the Brownian bridge "with time" Ψ, i.e.

$$(\mathbb{G}_{\Psi} f_1, \ldots \mathbb{G}_{\Psi} f_m) \sim \mathcal{N}(0, C)$$

and $C = (C_{i,j})$ is given by

$$C_{i,j} := \Psi\left[\big(f_i - \Psi f_i\big)\big(f_j - \Psi f_j\big)\right] = \Psi f_i f_j - \Psi f_i \Psi f_j.$$

There are two major problems that do not allow us to apply the standard Glivenko–Cantelli and Donsker theorems to the empirical measure of the subsample $S_{i(n,1)}, \ldots, S_{i(n,k)}$ and the resulting empirical process: the lack of independence between $S_{i(n,1)}, \ldots, S_{i(n,k)}$ and the fact that the underlying probability measure varies with n. Therefore a special version is needed which is suitable for $\mathcal{L}(S_{i(n,1)}, \ldots, S_{i(n,k)})$. The following lemma gives insight into the structure of this probability distribution and provides a basis for the following results.

Lemma A.1. *Suppose that the distribution function F_R of the radial part $R = \|X\|_1$ is continuous and consider the $(k+1)$-st upper order statistic of $R_1, \ldots, R_n$ transformed by F_R:*

$$U_n := F_R\left(R_{n:k+1}\right). \tag{40}$$

Then, for any $u \in (0,1)$,

$$\mathcal{L}\left(\left(S_{i(n,1)}, \ldots, S_{i(n,k)}\right) | U_n = u\right) = \otimes_{i=1}^{k} \Psi_u,$$

where

$$\Psi_u := \mathcal{L}\left(S | F_R(R) > u\right).$$

Proof. The continuity of F_R implies that the ordered sample indices $i(n,1), \ldots, i(n,k)$ and the permutation π satisfying (19) are unique almost surely. Moreover, the random variables

$$Y_i := F_R(R_i), \quad i = 1, \ldots, n,$$

are independent and uniformly distributed on $(0,1)$, whereas by construction of Y_i we have $R_i = F_R^{\leftarrow}(Y_i)$ almost surely and

$$\left(Y_{i(n,1)}, \ldots, Y_{i(n,k)}\right) = \left(Y_{n:\pi(1)}, \ldots, Y_{n:\pi(k)}\right).$$

It is well known (cf. Hajós and Rényi [22]) that the conditional probability distribution of $Y_{n:1}, \ldots, Y_{n:k}$ given $Y_{n:k+1} = u$ is equal to the probability distribution of the order statistic of k i.i.d. random variables that are uniformly distributed on $(u,1)$. Due to the fact that the permutation π is uniformly distributed on the permutation group $\mathcal{S}^k$ we obtain that the subsample $Y_{i(n,1)}, \ldots, Y_{i(n,k)}$ is conditionally i.i.d.:

$$\mathcal{L}\left(Y_{i(n,1)}, \ldots, Y_{i(n,k)} \,|\, U_n = u\right) = \otimes_{i=1}^{k} \mathrm{unif}(u,1).$$

Since for a random variable $Z \sim \mathrm{unif}(u,1)$ the probability distribution of $F_R^{\leftarrow}(Z)$ is given by $\mathcal{L}(R|F_R(R) > u)$, we obtain

$$\mathcal{L}\left(R_{i(n,1)}, \ldots, R_{i(n,k)} |\, U_n = u\right) = \otimes_{i=1}^{k} \mathcal{L}(R|F_R(R) > u).$$

Finally, this conditional i.i.d. property is inherited by the subsample $X_{i(n,1)}, \ldots, X_{i(n,k)}$ and its angular parts $S_{i(n,1)}, \ldots, S_{i(n,k)}$ in the following sense:

$$\mathcal{L}\left(\left(S_{i(n,1)}, \ldots, S_{i(n,k)}\right) |\, U_n = u\right) = \otimes_{i=1}^{k} \mathcal{L}(S|F_R(R) > u).$$

□

An immediate consequence of Lemma A.1 is the following representation of $\mathcal{L}(S_{i(n,1)}, \ldots, S_{i(n,k)})$ as a mixture of product measures.

Corollary A.2. *If F_R is continuous, then*

$$\mathrm{P}\left\{\left(S_{i(n,1)}, \ldots, S_{i(n,k)}\right) \in A\right\} = \int_0^1 \Psi_u^k(A)\, \mathrm{d}\mathrm{P}^{U_n}(u) \tag{41}$$

for $A \in \mathcal{B}\left(\left(\Sigma^d\right)^k\right)$, where P^{U_n} is the probability distribution of U_n and

$$\Psi_u^k := \otimes_{i=1}^{k} \Psi_u, \quad u \in [0,1].$$

Since $F_R^{\leftarrow}(u) \to \infty$ for $u \uparrow 1$, the behaviour of Ψ_u for $u \uparrow 1$ is related to the regular variation of X. One easily obtains the following result.

Lemma A.3. *Suppose that the random variable X is multivariate regularly varying. Then*

$$\Psi_u \xrightarrow{\mathrm{w}} \Psi, \quad u \uparrow 1.$$

Proof. The measure Ψ_u is obtained from the measure

$$\mu_u := \mathcal{L}\left(t(u)^{-1}X|R > t(u)\right), \quad t(u) := F_R^{\leftarrow}(u),$$

by the transformation $\tau : x \mapsto \|x\|_1^{-1} x$:

$$\Psi_u = \mu_u^\tau.$$

The representation (7) of multivariate regular variation implies

$$\mu_u \xrightarrow{\mathrm{w}} \nu|_{A_1},$$

where $\nu|_{A_1}$ is the restriction of ν to the set $A_1 = \{x \in \mathbb{R}_+^d : \|x\|_1 > 1\}$. Hence, the Continuous Mapping Theorem yields

$$\mu_u^\tau \xrightarrow{\mathrm{w}} (\nu|_{A_1})^\tau = \Psi.$$

$\square$

Now let us recapitulate the results we have obtained so far. We have the representation $\hat{\gamma}_\xi = \mathbb{P}_n f_{\xi,\hat{\alpha}}$, where $\mathbb{P}_n$ is the empirical measure of $S_{i(n,1)}, \ldots, S_{i(n,k)}$, indexed by elements of the function class

$$\mathcal{F} := \left\{ f_{\xi,\alpha} : \alpha \in (0,\infty), \xi \in \Sigma^d \right\}. \tag{42}$$

Due to Corollary A.2 we know that the empirical measure $\mathbb{P}_n$ is a mixture of empirical measures constructed from i.i.d. observations:

$$\mathcal{L}(\mathbb{P}_n) = \int_0^1 \mathcal{L}(\mathbb{P}_{k,\Psi_u}) \, \mathrm{d}\mathrm{P}^{U_n}(u).$$

Moreover, we have $\Psi_u \xrightarrow{\mathrm{w}} \Psi_1 := \Psi$ as $u \uparrow 1$ and it is well known that $U_n \uparrow 1$ P-almost surely. Thus the consistency and the asymptotic normality of the estimator $\mathbb{P}_n f_{\xi,\hat{\alpha}}$ are related to the uniformity of the Glivenko–Cantelli and Donsker properties (37) and (39) of the class $\mathcal{F}$ in the underlying probability measure $\Psi^* \in \{\Psi_u : u \in (0,1]\}$.

While the uniform Donsker property of $\mathcal{F}$ provides that convergence of empirical processes constructed from i.i.d. observations is uniform in the underlying probability measure, the uniform *pre-Gaussian* property allows to extend the convergence of empirical processes to the case when the underlying probability distribution converges for $k \to \infty$. Let $\mathcal{P}$ be a class of probability measures on Σ^d. The function class $\mathcal{F}$ is pre-Gaussian uniformly in $\Psi \in \mathcal{P}$ if the following two conditions are satisfied:

$$\sup_{\Psi \in \mathcal{P}} \mathrm{E} \|\mathbb{G}_\Psi\|_{l^\infty(\mathcal{F})} < \infty$$

and

$$\lim_{\delta\downarrow 0} \sup_{\Psi\in\mathcal{P}} \mathrm{E} \sup_{\rho_\Psi(f,g)<\delta} |\mathbb{G}_{k,\Psi}(f) - \mathbb{G}_{k,\Psi}(g)| = 0,$$

where $\rho_\Psi(f)$ is the seminorm $\|f - \Psi f\|_{\Psi,2}$.

The following lemma states that the class $\mathcal{F}$ is *universally* Glivenko–Cantelli, Donsker and pre-Gaussian, i.e. that these properties are uniform over the class of all probability measures on $(\Sigma^d, \mathcal{B}(\Sigma^d))$.

Lemma A.4. *The function class $\mathcal{F}$ is universally Glivenko–Cantelli, Donsker and pre-Gaussian.*

Proof. Let us first note that $\mathcal{F}$ is measurable (i.e. all $f \in \mathcal{F}$ are measurable) and that $\mathcal{F}$ is uniformly bounded by 1:

$$\forall s \in \Sigma^d, \forall f \in \mathcal{F}: \quad f(s) \leq 1.$$

Therefore we can take

$$F(s) := 1_{\Sigma^d}(s) \tag{43}$$

as an *envelope function* for $\mathcal{F}$.

According to van der Vaart and Wellner [39], $\mathcal{F}$ is universally Glivenko–Cantelli if it satisfies the *entropy condition*

$$\forall \varepsilon > 0 \quad \sup_{Q\in\mathcal{Q}_n} \log N\left(\varepsilon\|F\|_{Q,1}, \mathcal{F}, L^1(Q)\right) = o(n), \tag{44}$$

where $\mathcal{Q}_n$ denotes the class of all discrete probability measures on Σ^d with atoms of size integer multiples of $1/n$. Moreover, $\mathcal{F}$ is universally Donsker and pre-Gaussian if it satisfies the *uniform entropy condition*

$$\int_0^\infty \sup_{Q\in\mathcal{Q}} \sqrt{\log N\left(\varepsilon \left\|F\right\|_{Q,2}, \mathcal{F}, L^2(Q)\right)} \mathrm{d}\varepsilon < \infty, \tag{45}$$

where $\mathcal{Q}$ denotes the class of all discrete probability measures on Σ^d. The *covering number* $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ is defined as the minimal number of balls $\{g : \|g - f\| < \varepsilon\}$ of radius ε needed to cover the class $\mathcal{F}$. The *entropy* is the logarithm of the covering number. For more details we refer to van der Vaart and Wellner [39]. The covering numbers are defined in Section 2.1.1, whereas conditions (44) and (45) are related to Theorems 2.8.1 and 2.8.3 respectively.

The verification of (44) and (45) can be based on the properties of *Vapnik–Červonenkis (VC)* classes of sets and functions. For definitions of these objects and detailed results we refer to van der Vaart and Wellner [39], Section 2.6.

Let us first consider the structure of the class $\mathcal{F}$. Any $f_{\xi,\alpha} \in \mathcal{F}$ is obtained by composition of a linear and a monotone function:

$$f_{\xi,\alpha} = g_\alpha \circ h_\xi,$$

where

$$g_\alpha : [0,1] \to [0,1], \ t \mapsto t^\alpha$$

and

$$h_\xi : \Sigma^d \to [0,1], \ s \mapsto \xi^\top s.$$

It is easy to see that the function class

$$\mathcal{H} := \left\{ h_\xi : \xi \in \Sigma^d \right\}$$

is a subset of a finite-dimensional vector space of functions. Hence it is a *VC-Major* class. Since the functions g_α are monotone, the class $\mathcal{F}$ is also VC-Major. Furthermore, the uniform boundedness of $\mathcal{F}$ implies that it is a *VC-hull* class. This property implies that $\mathcal{F}$ satisfies (45) and (44). For more details we refer to Section 2.6 of van der Vaart and Wellner [39], especially Theorem 2.6.9 and Lemmas 2.6.13, 2.6.15, and 2.6.20. □

The next step is the application of Lemma A.4 to the estimator $\hat{\gamma}_\xi$. We start with the strong consistency.

Proof of Theorem 4.1. Part (a). Consider the decomposition

$$\hat{\gamma}_\xi - \gamma_\xi = \left(\hat{\Psi} f_{\xi,\hat{\alpha}} - \mathrm{E}\hat{\Psi} f_{\xi,\hat{\alpha}} \right) + \left(\mathrm{E}\hat{\Psi} f_{\xi,\hat{\alpha}} - \Psi f_{\xi,\hat{\alpha}} \right) + \left(\Psi f_{\xi,\hat{\alpha}} - \Psi f_{\xi,\alpha} \right). \quad (46)$$

First we show that

$$\Psi f_{\xi,\hat{\alpha}} - \Psi f_{\xi,\alpha} \to 0 \text{ P-a.s.}$$

uniformly in ξ. Since

$$|\Psi f_{\xi,\hat{\alpha}} - \Psi f_{\xi,\alpha}| \leq \|f_{\xi,\hat{\alpha}} - f_{\xi,\alpha}\|_\infty$$

for all $\xi \in \Sigma^d$, it suffices to show that

$$\sup_{\xi \in \Sigma^d} \|f_{\xi,\hat{\alpha}} - f_{\xi,\alpha}\|_\infty \to 0 \text{ P-a.s.} \quad (47)$$

Consider the partial derivative of $f_{\xi,\alpha}$ in α:

$$\frac{\partial}{\partial \alpha} \left(\xi^\top s \right)^\alpha = \left(\xi^\top s \right)^\alpha \log \left(\xi^\top s \right).$$

Since $\xi^\top s$ ranges in $[0,1]$, we obtain

$$\left|\frac{\partial}{\partial\alpha} f_{\xi,\alpha}(s)\right| \leq \sup_{t\in[0,1]} |t^\alpha \log t| = |t_0^\alpha \log(t_0)|\,,$$

where $t_0 = \exp(-1/\alpha)$. Due to the strong consistency of $\hat\alpha = \hat\alpha(n)$ we have $\hat\alpha > \alpha/2$ P-a.s. for n exceeding a sufficiently large bound n_0 and therefore

$$\forall \xi \in \Sigma^d \quad \|f_{\xi,\hat\alpha} - f_{\xi,\alpha}\|_\infty \leq 2\,(e\cdot\alpha)^{-1}\,|\hat\alpha - \alpha| \quad \text{P-a.s.} \tag{48}$$

for $n \geq n_0$. Hence (47) follows from the strong consistency of $\hat\alpha$.

Now consider the second term on the right side of (46). We have

$$\mathrm{E}\hat\Psi f_{\xi,\hat\alpha} - \Psi f_{\xi,\hat\alpha} = \left(\mathrm{E}\hat\Psi - \Psi\right) f_{\xi,\alpha} + \left(\mathrm{E}\hat\Psi - \Psi\right)[f_{\xi,\hat\alpha} - f_{\xi,\alpha}]$$

Due to

$$\left|\left(\mathrm{E}\hat\Psi - \Psi\right)[f_{\xi,\hat\alpha} - f_{\xi,\alpha}]\right| \leq 2\,\|f_{\xi,\hat\alpha} - f_{\xi,\alpha}\|_\infty$$

and (47) we only need to show that $(\mathrm{E}\hat\Psi - \Psi)f_{\xi,\alpha} \to 0$ uniformly in ξ, which is provided by assumption (24).

Finally, let us consider the first term on the right side of (46). Since $\hat\Psi = \mathbb{P}_n$, the mixture representation (41) yields

$$\begin{aligned}
\mathcal{L}\left(\hat\Psi f_{\xi,\hat\alpha} - \mathrm{E}\hat\Psi f_{\xi,\hat\alpha}\right) &= \mathcal{L}(\mathbb{P}_n f_{\xi,\hat\alpha} - \mathrm{E}\mathbb{P}_n f_{\xi,\hat\alpha}) \\
&= \int_{[0,1]} \mathcal{L}\left(\mathbb{P}_n f_{\xi,\hat\alpha} - \mathrm{E}\mathbb{P}_n f_{\xi,\hat\alpha} \,|\, U_n = u\right) \mathrm{dP}^{U_n}(u) \\
&= \int_{[0,1]} \mathcal{L}\left((\mathbb{P}_{k,\Psi_u} - \Psi_u)\, f_{\xi,\hat\alpha}\right) \mathrm{dP}^{U_n}(u).
\end{aligned} \tag{49}$$

Due to the universal Glivenko–Cantelli Property of $\mathcal{F}$ (cf. Lemma A.4) we have

$$(\mathbb{P}_{k,\Psi_u} - \Psi_u)\, f \to 0 \text{ P-a.s.}$$

uniformly in Ψ_u and $f \in \mathcal{F}$. Applied to the representation (49), this yields

$$\hat\Psi f_{\xi,\hat\alpha} - \mathrm{E}\hat\Psi f_{\xi,\hat\alpha} \to 0 \text{ P-a.s.}$$

uniformly in ξ. Hence all terms in (46) vanish uniformly in ξ almost surely and the proof of part (a) is finished.

Part (b) follows along the lines of the proof of part (a). If the assumption (24) is dropped, we only need to verify

$$\mathrm{E}\hat\Psi f_{\xi,\alpha} - \Psi f_{\xi,\alpha} \to 0 \tag{50}$$

pointwise in ξ. Recall that the mixture representation (41) yields

$$\mathcal{L}(\mathbb{P}_n) = \int_{[0,1]} \mathcal{L}(\mathbb{P}_{k,\Psi_u})\,\mathrm{dP}^{U_n}(u)$$

and that we have $U_n \uparrow 1$ P-a.s. and $\Psi_u \xrightarrow{\mathrm{w}} \Psi$ for $u \uparrow 1$. As a result, we obtain the weak convergence

$$\mathbb{P}_n \xrightarrow{\mathrm{w}} \Psi.$$

Since $\hat{\Psi} = \mathbb{P}_n$ and all functions $f_{\xi,\alpha}$ are continuous, we obtain (50) pointwise in ξ. □

Remark A.5. It was noted in Remark 4.2 that the assumption (24) is satisfied for all $\alpha \geq 1$. This is due to the fact that weak convergence of separable Borel measures on a metric space $\mathcal{Z}$ is metrizable by the *bounded Lipschitz metric*

$$d_{BL_1}(L_1, L_2) := \sup_{h \in BL_1} \left| \int h\,\mathrm{d}L_1 - \int h\,\mathrm{d}L_2 \right|, \tag{51}$$

where BL_1 is the set of all functions $h \in l^\infty(\mathcal{Z})$ that are uniformly bounded by 1 and Lipschitz with factor 1:

$$\sup_{z \in \mathcal{Z}} |h(z)| \leq 1,$$
$$|h(z_1) - h(z_2)| \leq \|z_1 - z_2\|_{\mathcal{Z}}$$

(cf. van der Vaart and Wellner [39], Chapter 1.12). It is easy to verify that the function class $\{f_{\xi,\alpha} : \xi \in \Sigma^d\}$ is uniformly Lipschitz for any $\alpha \geq 1$. Hence weak convergence $\hat{\Psi} \xrightarrow{\mathrm{w}} \Psi$ implies that (24) is satisfied for any $\alpha \geq 1$.

The central part in the proof of Theorem 4.5 is the weak convergence of the empirical process related to the subsample $S_{i(n,1)}, \ldots, S_{i(n,k)}$:

$$\mathbb{G}_n := \sqrt{k}\,(\mathbb{P}_n - \mathrm{P}_n)\,, \tag{52}$$

where the probability measure P_n is defined as the expectation of $\mathbb{P}_n$:

$$\mathrm{P}_n f := \mathrm{E}\mathbb{P}_n f. \tag{53}$$

With the universal Donsker property and pre-Gaussianity at hand, convergence of $\mathbb{G}_n$ is obtained from the convergence of the conditioning random variable U_n.

Lemma A.6. *Suppose that Condition 4.4 is satisfied. Then the empirical process $\mathbb{G}_n$ converges to a Brownian Bridge "with time" Ψ:*

$$\mathbb{G}_n \xrightarrow{\mathrm{w}} \mathbb{G}_\Psi \quad \textit{in } l^\infty(\mathcal{F}). \tag{54}$$

Proof. Since the empirical process $\mathbb{G}_n$ is constructed from the subsample $S_{i(n,1)}, \ldots, S_{i(n,k)}$, the mixture representation (41) of $\mathcal{L}(S_{i(n,1)}, \ldots, S_{i(n,k)})$ implies

$$\mathcal{L}\left(\mathbb{G}_n\right) = \int_{[0,1]} \mathcal{L}\left(\mathbb{G}_{k,\Psi_u}\right) \mathrm{dP}^{U_n}(u). \tag{55}$$

Moreover, we already know that $U_n \uparrow 1$ P-a.s. and $\Psi_u \xrightarrow{\mathrm{w}} \Psi$ for $u \uparrow 1$.

Let us consider a sequence u_k in $(0,1)$ such that $u_k \uparrow 1$ for $k \to \infty$ and the empirical processes $\mathbb{G}_{k,\Psi_k}$ with the underlying measure $\Psi_k := \Psi_{u_k}$. As shown in Lemma A.4, the class $\mathcal{F}$ is universally Donsker and pre-Gaussian. According to Lemma 2.8.7 in van der Vaart and Wellner [39], the convergence

$$\mathbb{G}_{k,\Psi_k} \xrightarrow{\mathrm{w}} \mathbb{G}_{\Psi} \tag{56}$$

holds if the class $\mathcal{F}$ and the sequence Ψ_k satisfy

$$\forall \varepsilon > 0 \quad \limsup_{k\to\infty} \Psi_k \left[F^2 \cdot 1\{F \geq \varepsilon\sqrt{k}\}\right] = 0 \tag{57}$$

and

$$\sup_{f,g\in\mathcal{F}} \left|\rho_{\Psi_k}(f-g) - \rho_{\Psi}(f-g)\right| \to 0, \tag{58}$$

where $\rho_\Psi(f)$ denotes the seminorm $\|f - \Psi f\|_{\Psi,2}$. Since the envelope function F of $\mathcal{F}$ is bounded, condition (57) is trivial and we only need to verify (58). For $h := (f-g)$ we have

$$\begin{aligned}
\left|\rho_{\Psi_k}(h) - \rho_\Psi(h)\right| &= \left|\Psi_k h^2 - \left(\Psi_k h\right)^2 - \left(\Psi h^2 - (\Psi h)^2\right)\right| \\
&= \left|\left(\Psi_k h^2 - \Psi h^2\right) - \left(\Psi_k h - \Psi h\right) \cdot \left(\Psi_k h + \Psi h\right)\right| \\
&\leq \left|\Psi_k h^2 - \Psi h^2\right| + \left|\Psi_k h - \Psi h\right| \cdot O(1).
\end{aligned}$$

Thus, due to $(f-g)^2 \leq 2f^2 + 2g^2$ and $|f-g| \leq |f| + |g|$, it suffices to show that for $k \to \infty$

$$\sup_{f\in\mathcal{F}} \left|\Psi_k f^2 - \Psi f^2\right| \to 0 \quad \text{and} \quad \sup_{f\in\mathcal{F}} \left|\Psi_k f - \Psi f\right| \to 0.$$

Since $f \in \mathcal{F}$ implies $f^2 \in \mathcal{F}$, we only need to verify

$$\sup_{f\in\mathcal{F}} \left|\Psi_k f - \Psi f\right| \to 0, \quad k \to \infty. \tag{59}$$

Consider sets

$$B_\delta := \left\{s \in \Sigma^d : s^{(i)} > \delta \text{ for all } i = 1, \ldots, d\right\}, \quad \delta > 0.$$

If Condition 4.4(a) is satisfied, i.e. we have $\Psi(\partial\Sigma^d) = 0$, then for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\Psi(\Sigma^d \setminus B_\delta) < \varepsilon/4.$$

Since the number of atoms of Ψ is countable, δ can always be chosen such that

$$\Psi\left(\partial B_\delta\right) = 0.$$

Hence, $\Psi_k \xrightarrow{\mathrm{w}} \Psi$ implies $\Psi_k(\Sigma^d \setminus B_\delta) \to \Psi(\Sigma^d \setminus B_\delta)$ and therefore

$$\begin{aligned}
|\Psi_k f - \Psi f| &\leq |\Psi_k\left[f \cdot 1_{B_\delta}\right] - \Psi\left[f \cdot 1_{B_\delta}\right]| + |\Psi_k\left[f \cdot (1 - 1_{B_\delta})\right] - \Psi\left[f \cdot (1 - 1_{B_\delta})\right]| \\
&\leq |\Psi_k\left[f \cdot 1_{B_\delta}\right] - \left[\Psi f \cdot 1_{B_\delta}\right]| + \Psi_k\left(\Sigma^d \setminus B_\delta\right) + \Psi\left(\Sigma^d \setminus B_\delta\right) \\
&\leq |\Psi_k\left[f \cdot 1_{B_\delta}\right] - \Psi\left[f \cdot 1_{B_\delta}\right]| + \frac{3}{4}\varepsilon
\end{aligned}$$

for sufficiently large k. Hence we only need to verify

$$\sup_{f\in\mathcal{F}} |\Psi_k\left[f \cdot 1_{B_\delta}\right] - \Psi\left[f \cdot 1_{B_\delta}\right]| \to 0, \quad k \to \infty.$$

Due to the metrization of weak convergence by the bounded Lipschitz metric (cf. Remark A.5) it suffices to show that the function class $\mathcal{F}$ is uniformly Lipschitz on B_δ, i.e. that there exists $K > 0$ such that

$$\forall f \in \mathcal{F}, \forall s_1, s_2 \in B_\delta : \quad |f(s_1) - f(s_2)| \leq K\left|s_1 - s_2\right|. \tag{60}$$

Since all $f \in \mathcal{F}$ are differentiable on B_δ it suffices to show that the partial derivatives of $f \in \mathcal{F}$ are uniformly bounded on B_δ. We have

$$\frac{\partial}{\partial s^{(i)}} f_{\xi,\alpha}(s) = \alpha\left(\xi^\top s\right)^{\alpha-1} \cdot \xi^{(i)}.$$

Due to $\xi^{(i)} \leq 1$ and $s \in B_\delta$ we obtain

$$\sup_{s\in B_\delta} \left|\frac{\partial}{\partial s^{(i)}} f_{\xi,\alpha}(s)\right| \leq \sup_{t\in(\delta,1-\delta)} \alpha \cdot t^{\alpha-1}.$$

The term on the right side is uniformly bounded for $\alpha \in (0, \infty)$ due to

$$\sup_{\alpha\geq 1} \sup_{t\in(\delta,1-\delta)} \alpha \cdot t^{\alpha-1} = \sup_{\alpha\geq 1} \alpha(1-\delta)^{\alpha-1} < \infty$$

and

$$\sup_{\alpha\in(0,1)} \sup_{t\in(\delta,1-\delta)} \alpha \cdot t^{\alpha-1} = \sup_{\alpha\in(0,1)} \alpha \cdot \delta^{\alpha-1} < \infty.$$

As a consequence, we obtain (60), which implies (59) and (58). Hence we obtain the convergence (56).

If Condition 4.4(b) is satisfied, i.e. we have $\alpha \in [1, \alpha^*]$, then the class of index functions f can be reduced to

$$\mathcal{F}_{\alpha^*} := \left\{ f_{\xi,\alpha} : \xi \in \Sigma^d, \alpha \in [1, \alpha^*] \right\}.$$

Thus (58) is simplified to

$$\sup_{f,g \in \mathcal{F}_{\alpha^*}} \left| \rho_{\Psi_k}(f-g) - \rho_\Psi(f-g) \right| \to 0,$$

which can be obtained from the uniform Lipschitz property of $f \in \mathcal{F}_{2\alpha^*}$. Since $f \in \mathcal{F}$ are differentiable on Σ^d, the uniform Lipschitz property follows from

$$\sup_{f \in \mathcal{F}_{2\alpha^*}} \sup_{s \in \Sigma^d} \left| \frac{\partial}{\partial s^{(i)}} f(s) \right| = 2\alpha^* < \infty.$$

Hence (58) is verified and we obtain (56).

Now let us finish the proof by combination of (56) with the mixture representation (55). It was already mentioned above that weak convergence is metrized by the bounded Lipschitz metric d_{BL_1} (cf. Remark A.5). Hence it suffices to show that

$$d_{BL_1}(\mathbb{G}_n, \mathbb{G}_\Psi) \to 0.$$

Let $h \in BL_1$. Then the mixture representation (55) yields

$$\begin{aligned}
&\left| \int h \,\mathrm{d}\mathcal{L}(\mathbb{G}_n) - \int h \,\mathrm{d}\mathcal{L}(\mathbb{G}_\Psi) \right| \\
&\quad = \left| \int_{[0,1]} \int h \,\mathrm{d}\mathcal{L}(\mathbb{G}_{k,\Psi_u}) \,\mathrm{dP}^{U_n}(u) - \int h \,\mathrm{d}\mathcal{L}(\mathbb{G}_\Psi) \right| \\
&\quad = \left| \int_{[0,1]} \left(\int h \,\mathrm{d}\mathcal{L}(\mathbb{G}_{k,\Psi_u}) - \int h \,\mathrm{d}\mathcal{L}(\mathbb{G}_\Psi) \right) \mathrm{dP}^{U_n}(u) \right| \\
&\quad \le \int_{[0,u_0)} \left| \int h \,\mathrm{d}\mathcal{L}(\mathbb{G}_{k,\Psi_u}) - \int h \,\mathrm{d}\mathcal{L}(\mathbb{G}_\Psi) \right| \mathrm{dP}^{U_n}(u) \\
&\qquad + \int_{[u_0,1]} \left| \int h \,\mathrm{d}\mathcal{L}(\mathbb{G}_{k,\Psi_u}) - \int h \,\mathrm{d}\mathcal{L}(\mathbb{G}_\Psi) \right| \mathrm{dP}^{U_n}(u) \\
&\quad \le 2\mathrm{P}\left\{ U_n < u_0 \right\} + \sup_{u \ge u_0} d_{BL_1}(\mathbb{G}_{\Psi_u}, \mathbb{G}_\Psi).
\end{aligned}$$

Given a fixed $\varepsilon > 0$, there exists $u_0 \in (0,1)$ and $n_0 \in \mathbb{N}$ such that

$$d_{BL_1}(\mathbb{G}_{k,\Psi_u}, \mathbb{G}_\Psi) < \varepsilon/2$$

for $u \geq u_0$ and $n \geq n_0$. Since $U_n \to 1$ P-a.s., the index n_0 can be enlarged (if necessary) so that

$$\mathrm{P}\left\{U_n < u_0\right\} < \frac{\varepsilon}{4}$$

for $n \geq n_0$. Now, for $n \geq n_0$, we obtain

$$d_{BL_1}\left(\mathbb{G}_n, \mathbb{G}_\Psi\right) = \sup_{h \in BL_1} \left| \int h \, \mathrm{d}\mathcal{L}\left(\mathbb{G}_n\right) - \int h \, \mathrm{d}\mathcal{L}\left(\mathbb{G}_\Psi\right) \right| < \varepsilon,$$

which implies $d_{BL_1}(\mathbb{G}_n, \mathbb{G}_\Psi) \to 0$. □

The preceding lemma allows us to prove the AN property of $\hat{\gamma}_\xi$.

Proof of Theorem 4.5. Part (a). We need to show that the asymptotic normality (25) of $\hat{\alpha}$ and the second order condition (26) yield weak convergence of $\sqrt{k}(\hat{\gamma}_\xi - \gamma_\xi) = \sqrt{k}(\mathbb{P}_n \hat{f} - \Psi f)$ to the Gaussian process in (27). Consider the decomposition

$$\begin{aligned}
&\sqrt{k}\left(\mathbb{P}_n \hat{f} - \Psi f\right) \\
&\quad = \sqrt{k}\left(\mathbb{P}_n \hat{f} - \mathrm{P}_n \hat{f}\right) + \sqrt{k}\left(\mathrm{P}_n \hat{f} - \Psi \hat{f}\right) + \sqrt{k}\left(\Psi \hat{f} - \Psi f\right) \\
&\quad = \mathbb{G}_n f_{\xi,\hat{\alpha}} + \sqrt{k}\left(\mathrm{P}_n - \Psi\right) f_{\xi,\hat{\alpha}} + \Psi \sqrt{k}\left(f_{\xi,\hat{\alpha}} - f_{\xi,\alpha}\right).
\end{aligned} \tag{61}$$

Recall the arguments that justified inequality (48). Since asymptotic normality of $\hat{\alpha}$ implies $\hat{\alpha} \xrightarrow{\mathrm{P}} \alpha$, for any $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\mathrm{P}\left\{\hat{\alpha} < \alpha/2\right\} < \varepsilon \tag{62}$$

for $n \geq n_0$. This yields

$$\mathrm{P}\left\{\|f_{\xi,\hat{\alpha}} - f_{\xi,\alpha}\| > 2(e \cdot \alpha)^{-1} \left|\hat{\alpha} - \alpha\right|\right\} < \varepsilon \tag{63}$$

for $n \geq n_0$ and therefore $\|f_{\xi,\hat{\alpha}} - f_{\xi,\alpha}\|_\infty \xrightarrow{\mathrm{P}} 0$. Hence Lemma A.6 implies

$$\left(\mathbb{G}_n f_{\xi,\hat{\alpha}}\right)_{\xi \in \Sigma^d} \xrightarrow{\mathrm{w}} \left(\mathbb{G}_\Psi f_{\xi,\alpha}\right)_{\xi \in \Sigma^d} \quad \text{in } l^\infty\left(\Sigma^d\right)$$

for all $\alpha > 0$.

Now consider the second term in (61). Due to the asymptotic normality of $\hat{\alpha}$, inequality (63) implies

$$\|f_{\xi,\hat{\alpha}} - f_{\xi,\alpha}\|_\infty = O_{\mathrm{P}}(1/\sqrt{k})$$

and therefore

$$\sqrt{k}(\mathrm{P}_n - \Psi) f_{\xi,\hat{\alpha}} = \sqrt{k}(\mathrm{P}_n - \Psi)\left[f_{\xi,\alpha} + O_{\mathrm{P}}(1/\sqrt{k})\right].$$

Hence assumption (26) yields

$$\sqrt{k}(\mathrm{P}_n - \Psi) f_{\xi,\hat{\alpha}} \xrightarrow{\mathrm{P}} b(\xi)$$

uniformly in ξ.

The asymptotic distribution of $\Psi\sqrt{k}\left(\hat{f} - f\right)$ is obtained from the asymptotic normality (25) of $\hat{\alpha}$. Recall that $f_{\xi,\alpha} = (\xi^\top s)^\alpha$ and $\xi^\top s$ ranges in $[0, 1]$. Taylor expansion yields

$$\left(t^{\hat{\alpha}} - t^\alpha\right) = (\hat{\alpha} - \alpha)\, t^\alpha \log t + \frac{1}{2}\,(\hat{\alpha} - \alpha)^2\, t^{\alpha^*} \log^2 t$$

with some α^* between α and $\hat{\alpha}$. Recall that asymptotic normality of $\hat{\alpha}$ implies (62). Since the mappings $t \mapsto t^\alpha \log t$ and $t \mapsto t^\alpha \log^2 t$ are bounded on $[0, 1]$ uniformly in $\alpha > \delta$ for any fixed $\delta > 0$, we obtain

$$\begin{aligned}
\sqrt{k}\left(\hat{f}(s) - f(s)\right) &= \sqrt{k}\left(\left(\xi^\top s\right)^{\hat{\alpha}} - \left(\xi^\top s\right)^{\alpha}\right) \\
&= \sqrt{k}\,(\hat{\alpha} - \alpha)\left(\xi^\top s\right)^\alpha \log\left(\xi^\top s\right) + \sqrt{k}\, O_{\mathrm{P}}\,(\hat{\alpha} - \alpha)^2 \\
&\xrightarrow{\mathrm{w}} Z(s) := Y \cdot \left(\xi^\top s\right)^\alpha \log\left(\xi^\top s\right)
\end{aligned}$$

as an l^∞-function of $\xi \in \Sigma^d$.

Due to the continuity of the mapping $g \mapsto \Psi g$ for $g \in l^\infty(\Sigma^d)$ we can apply the Continuous Mapping Theorem and obtain

$$\Psi\sqrt{k}\,(f_{\xi,\hat{\alpha}} - f_{\xi,\alpha}) \xrightarrow{\mathrm{w}} \Psi Z = Y \cdot c_{\xi,\alpha} \quad \text{in } l^\infty(\Sigma^d),$$

where

$$c_{\xi,\alpha} := \Psi\left(\left(\xi^\top s\right)^\alpha \log\left(\xi^\top s\right)\right).$$

Furthermore, ΨZ and $\mathbb{G}_\Psi f$ are independent. This follows from the asymptotic independence of the radial parts $R_{n:1}, \dots, R_{n:k}$ and the angular parts $S_{i(n,1)}, \dots, S_{i(n,k)}$ of the extreme subsample $X_{i(n,1)}, \dots, X_{i(n,k)}$. Hence we obtain (27).

The result of part (b) is just the finite-dimensional convergence of marginal distributions in part (a). Since replacing the assumption (26) by (28) affects only the middle term in (61), resulting in the replacement of the uniform convergence to $b(\xi)$ by pointwise convergence, the pointwise asymptotic normality (29) follows immediately along the lines of the proof for the part (a). □

References

[1] Carlo Acerbi. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518, July 2002.

[2] Philippe Barbe, Anne-Laure Fougères, and Christian Genest. On the tail behavior of sums of dependent risks. *Astin Bull.*, 36(2):361–373, 2006. ISSN 0515-0361.

[3] Bojan Basrak, Richard A. Davis, and Thomas Mikosch. A characterization of multivariate regular variation. *Ann. Appl. Probab.*, 12(3): 908–920, 2002. ISSN 1050-5164.

[4] Nicholas H. Bingham, Charles M. Goldie, and Józef L. Teugels. *Regular Variation*, volume 27 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, Cambridge, 1987. ISBN 0-521-30787-2.

[5] Klaus Böcker and Claudia Klüppelberg. Modelling and measuring multivariate operational risk with Lévy copulas. *J. Operational Risk*, 3(2): 3–27, 2008.

[6] Jan Boman and Filip Lindskog. Support theorems for the Radon transform and Cramér-Wold theorems. *J. Theoret. Probab.*, 2008. ISSN 0894-9840 (Print) 1572-9230 (Online). URL `http://www.springerlink.com/content/105605/?Content+Status=Accepted`.

[7] Jón Daníelsson, Bjørn N. Jorgensen, Mandira Sarma, and Casper G. de Vries. Sub-additivity re-examined: the case for Value-at-Risk. *EURANDOM Reports*, 2005. ISSN 1389-2355. URL `http://www.eurandom.nl/reports/2005/006MSreport.pdf`.

[8] Richard Davis and Sidney Resnick. Tail estimates motivated by extreme value theory. *Ann. Statist.*, 12(4):1467–1487, 1984. ISSN 0090-5364.

[9] L. de Haan and Sidney I. Resnick. Estimating the limit distribution of multivariate extremes. *Comm. Statist. Stochastic Models*, 9(2):275–309, 1993. ISSN 0882-0287.

[10] Laurens de Haan and Ana Ferreira. *Extreme Value Theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006. ISBN 978-0-387-23946-0; 0-387-23946-4.

[11] Laurens de Haan and Sidney I. Resnick. Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 40 (4):317–337, 1977.

[12] Laurens de Haan and Ashoke Kumar Sinha. Estimating the probability of a rare event. *Ann. Statist.*, 27(2):732–759, 1999. ISSN 0090-5364.

[13] Arnold L. M. Dekkers, John H. J. Einmahl, and Laurens de Haan. A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, 17(4):1833–1855, 1989. ISSN 0090-5364.

[14] Holger Drees. Refined Pickands estimators of the extreme value index. *Ann. Statist.*, 23(6):2059–2080, 1995. ISSN 0090-5364.

[15] Holger Drees, Ana Ferreira, and Laurens de Haan. On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.*, 14(3): 1179–1201, 2004. ISSN 1050-5164.

[16] John H. J. Einmahl, Laurens de Haan, and Xin Huang. Estimating a multidimensional extreme-value distribution. *J. Multivariate Anal.*, 47 (1):35–47, 1993. ISSN 0047-259X.

[17] John H. J. Einmahl, Laurens de Haan, and Ashoke Kumar Sinha. Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.*, 70(2):143–171, 1997. ISSN 0304-4149.

[18] John H. J. Einmahl, Laurens de Haan, and Vladimir I. Piterbarg. Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.*, 29(5):1401–1423, 2001. ISSN 0090-5364.

[19] Paul Embrechts, Dominik Lambrigger, and Mario V. Wüthrich. Multivariate extremes and the aggregation of dependent risks: Examples and counter-examples. *Extremes*, 2008. ISSN 1386-1999 (Print) 1572-915X (Online). URL `http://www.springerlink.com/content/102890/?Content+Status=Accepted`.

[20] Paul Embrechts, Johanna Nešlehová, and Mario V. Wüthrich. Additivity properties for Value-at-Risk under Archimedean dependence and heavy-tailedness. *Insurance: Mathematics and Economics*, in press, 2008. ISSN 0167-6687. URL `http://www.sciencedirect.com/science/article/B6V8N-4T5JHS0-1/2/839309e3ccb58c06c9221e7115c8623e`.

[21] Michael Falk, Jürg Hüsler, and Rolf-Dieter Reiss. *Laws of Small numbers: Extremes and Rare Events*, volume 23 of *DMV Seminar*. Birkhäuser, Basel, 1994. ISBN 3-7643-5071-7.

[22] G. Hajós and A. Rényi. Elementary proofs of some basic facts concerning order statistics. *Acta Math. Acad. Sci. Hungar.*, 5:1–6, 1954. ISSN 0001-5954.

[23] Hoskuldur A. Hauksson, Michel M. Dacorogna, Thomas Domenig, Ulrich A. Müller, and Gennady Samorodnitsky. Multivariate Extremes, Aggregation and Risk Estimation. *SSRN eLibrary*, 2000. URL `http://ssrn.com/paper=254392`.

[24] Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3(5):1163–1174, 1975. ISSN 0090-5364.

[25] Henrik Hult and Filip Lindskog. Regular variation for measures on metric spaces. *Publ. Inst. Math. (Beograd) (N.S.)*, 80(94):121–140, 2006. ISSN 0350-1302.

[26] Harry Joe. *Multivariate models and dependence concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1997. ISBN 0-412-07331-5.

[27] Claudia Klüppelberg and Sidney I. Resnick. The Pareto copula, aggregation of risks, and the emperor's socks. *J. Appl. Probab.*, 45(1):67–84, 2008. ISSN 0021-9002.

[28] Yannick Malevergne and Didier Sornette. *Extreme Financial Risks*. Springer-Verlag, Berlin, 2006. ISBN 978-3-540-27264-9; 3-540-27264-X.

[29] Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management*. Princeton Series in Finance. Princeton University Press, Princeton, NJ, 2005. ISBN 0-691-12255-5.

[30] Marco Moscadelli. The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee. Temi di discussione (Economic working papers) 517, Bank of Italy, Economic Research Department, July 2004. URL `http://ideas.repec.org/p/bdi/wptemi/td_517_04.html`.

[31] Johanna Nešlehová, Paul Embrechts, and Valerie Chavez-Demoulin. Infinite-mean models and the LDA for operational risk. *J. Operational Risk*, 1(1):3–25, 2006.

[32] James Pickands, III. Statistical inference using extreme order statistics. *Ann. Statist.*, 3:119–131, 1975. ISSN 0090-5364.

[33] James Pickands, III. Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, volume 49, pages 859–878, 894–902, 1981. With a discussion.

[34] Sidney Resnick. The extremal dependence measure and asymptotic independence. *Stoch. Models*, 20(2):205–227, 2004. ISSN 1532-6349.

[35] Sidney I. Resnick. *Extreme Values, Regular Variation, and Point Processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York, 1987. ISBN 0-387-96481-9.

[36] Sidney I. Resnick. *Heavy-Tail Phenomena*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007. ISBN 978-0-387-24272-9; 0-387-24272-4.

[37] Rafael Schmidt and Ulrich Stadtmüller. Non-parametric estimation of tail dependence. *Scand. J. Statist.*, 33(2):307–335, 2006. ISSN 0303-6898.

[38] Richard L. Smith. Estimating tails of probability distributions. *Ann. Statist.*, 15(3):1174–1207, 1987. ISSN 0090-5364.

[39] Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, New York, 1996. ISBN 0-387-94640-3. Corrected 2nd printing 2000.

[40] Mario V. Wüthrich. Asymptotic Value-at-Risk estimates for sums of dependent random variables. *Astin Bull.*, 33(1):75–92, 2003. ISSN 0515-0361.