XAIA INVESTMENT

Introduction to extreme-value theory and its applications in Finance

Jan-Frederik Mai
XAIA Investment GmbH
Sonnenstraße 19, 80331 München, Germany
jan-frederik.mai@xaia.com

Date: December 11, 2020

Abstract This is a concise survey of the main results and notions of *extreme-value theory*, as well as some remarks as to how these results have found their way into Finance. It is meant to provide a non-detailed, rough overview. For more detailed information and background, the interested readers are referred to the prominent textbook accounts Gumbel (1958); Resnick (1987); Embrechts et al. (1997); Beirlant et al. (2004); De Haan, Ferreira (2006).

1 Motivation via analogy to the central limit theorem

Let X be a random variable and suppose for a minute that we are interested in the "average" behavior of X. To this end, assume we observe independent realizations $X_1, X_2, \ldots$ of X and we consider the arithmetic average $\bar{X}_n := (X_1 + \ldots + X_n)/n$. Under the assumption of a finite first moment of X we known from the law of large numbers that $\bar{X}_n$ converges to a constant μ, the mean of X. This "trivial" limiting behavior is sometimes undesired, because one is sometimes rather interested in finer metrics surrounding the average behavior of X than just a point estimate. The central limit theorem provides the most commonly known mathematical wrapping for this. Under the assumption of existence of the second moment of X, it states that there is a constant $\sigma > 0$, the standard deviation of X, such that $\sqrt{n}\,(\bar{X}_n - \mu)/\sigma$ converges in distribution to a standard normally distributed random variable. In generic terms, the typical practical application of this result is as follows:

(1) **Problem**: We are in need of a model for a random quantity X, which we do not know everything about (we do not know the exact distribution), but we have confidence in point estimates for its mean and standard deviation.

(2) **Focus**: We are primarily interested in the "average" behavior of X, do not care too much about the "unusual" outcomes.

(3) **Solution**: The central limit theorem justifies to use a normal distribution with parameters μ and σ to model X.

A financial example would be obtained if X is the (log-)return of some asset. By construction, this logic focuses on the "average" behavior of X, so it simplifies/ignores the tails in some sense. This is reflected in the assumptions of the central limit theorem, since existence of first and second moment is a condition that prevents very heavy tails. If these assumptions are dropped, one may still obtain a generalized central limit theorem based on limiting distributions which involve additional "tail-heavyness" parameters, the so-called *stable laws*. In order to better relate the upcoming statements of extreme-value theory to the central

limit law, we point out that generalized central limit laws (beyond the normal law, so beyond existence of moments) postulate the existence of constants $a_n \in \mathbb{R}$ and $b_n > 0$ such that $(\bar{X}_n - a_n)/b_n$ converges in distribution. The resulting limiting law is then a stable law, which boils down to a normal distribution in case $a_n = \mu$ and $b_n = \sigma/\sqrt{n}$ is a valid choice (which basically means that second moments exist and tails are light).
On a high level, extreme-value theory provides a logic that is analogous to the central limit theorem, but focusing on the "rare" events rather than on the "average". Precisely, instead of the behavior of the arithmetic average $\bar{X}_n$, its main theorems investigate the behavior of the random variable $\max\{X_1, \ldots, X_n\}$ as $n \to \infty$. Equivalently but formulated slightly differently, it also studies for a given high threshold u the behavior of the distribution of X conditioned to be larger than u, meaning that those realizations of $X_1, X_2, \ldots$ below u are discarded and only the limiting distribution of the remaining "extreme" observations is investigated (so-called *peaks-over-threshold* method). There exist analogous limiting distributions, like the normal distribution in the central limit theorem. In a practical situation that requires a focus on the behavior of X only in the (rare) cases when $X > u$ one may thus with an analogous logic directly consider the respective limiting distributions as models for X.
Both the central limit law and extreme-value theory have multivariate counterparts, focusing on d-dimensional random vectors $\boldsymbol{X} = (X_1, \ldots, X_d)$ rather than on single random variables[1]. In case of the central limit law, the limiting law is a multivariate normal distribution, which is again determined by first and second moments (so only correlations enter the scene in addition to the univariate case). In case of extreme-value theory, in addition to the one-dimensional limiting laws the dependence between the components must necessarily be modeled by certain families of distributions, called *extreme-value copulas*. Unfortunately, the class of extreme-value copulas is not determined by a finite number of parameters (like in the multivariate normal case) but instead is infinite-dimensional. But many low-parametric models have been developed and also many qualitative structural results on this class of dependence models are known. Finally, it is clear that since $\min\{X_1, \ldots, X_n\} = -\max\{-X_1, \ldots, -X_n\}$ it is theoretically sufficient to focus on maxima, but obtains an equivalent theory for minima as well.

2 Summary of extreme-value theory

We distinguish the univariate case (Subsection 2.1) and the multivariate case (Subsection 2.2).

2.1 The univariate case

Let F be the distribution function of a random variable X, i.e. $F(x) = \mathbb{P}(X \leq x)$ for $x \in \mathbb{R}$. We denote by u_F the right end point of the support of F, which might possibly be infinity. There are two important theorems in univariate extreme-value theory, the *extremal types theorem* and the *Pickands-Balkema-de Haan theorem*, which are briefly recalled. To this end, the following definitions are of importance.

[1] In fact, people even study the same concepts on more general spaces than $\mathbb{R}^d$, like function spaces for instance.

Definition 2.1 (Possible extreme-value distributions)
Let $\xi \in \mathbb{R}$ be a shape parameter.

(a) The distribution function

$$H_\xi(x) = \begin{cases} \exp\Big\{ -\big(1+\xi\,x\big)^{-\frac{1}{\xi}} \Big\} & \text{, if } \xi \neq 0 \\ \exp\Big\{ -e^{-x} \Big\} & \text{, if } \xi = 0 \end{cases},$$

which has support

$$\begin{cases} x \in \mathbb{R} & \text{, if } \xi = 0 \text{ (\textit{Gumbel type})} \\ x \in [-1/\xi, \infty) & \text{, if } \xi > 0 \text{ (\textit{Fréchet type})} \\ x \in (-\infty, -1/\xi] & \text{, if } \xi < 0 \text{ (\textit{Weibull type})} \end{cases}$$

is called *(standard) generalized extreme-value distribution*.

(b) The distribution function

$$G_\xi(x) = \begin{cases} 1 - \big(1+\xi\,x\big)^{-\frac{1}{\xi}} & \text{, if } \xi \neq 0 \\ 1 - e^{-x} & \text{, if } \xi = 0 \end{cases},$$

which has support

$$\begin{cases} x \in [0, \infty) & \text{, if } \xi = 0 \text{ (\textit{Gumbel type})} \\ x \in [0, \infty) & \text{, if } \xi > 0 \text{ (\textit{Fréchet type})} \\ x \in [0, -1/\xi] & \text{, if } \xi < 0 \text{ (\textit{Weibull type})} \end{cases}$$

is called *(standard) generalized Pareto distribution*.

Figure 1 illustrates these distribution functions, making explicit that the tail-heavyness increases in ξ.
Intuitively, the first important theorem states that if $X_1, X_2, \ldots$ are independent and identically distributed random variables with distribution function F, and if there exist normalizing constants $a_n \in \mathbb{R}$, $b_n > 0$ such that $(\max\{X_1, \ldots, X_n\} - a_n)/b_n$ converges in distribution as $n \to \infty$, the limiting distribution is necessarily a generalized extreme-value distribution. It is called the extremal types theorem and can be formulated as follows.

Theorem 2.2 (Fisher, Tippett (1928) and Gnedenko (1943))
If there exist constants $a_n \in \mathbb{R}$ and $b_n > 0$ such that $H(x) := \lim_{n\to\infty} F(b_n\,x + a_n)^n$ exists (pointwise), then necessarily $H(x) = H_\xi\big((x-\mu)/\sigma\big)$ for some $\xi \in \mathbb{R}$ and location parameters $\mu \in \mathbb{R}$, $\sigma > 0$. Such F is said to be in the *domain of attraction of* H_ξ.

The assumption of the theorem on the existence of a_n, b_n needs not hold in general, but holds very often. Concretely, the following can be said regarding a classification of the three different types:

- **Gumbel type** $\xi = 0$: This occurs typically if X is light-tailed, as is the case for many "common" probability distributions like the normal, the exponential, or the Gamma distribution. An analytical criterion on F to be in the domain of attraction of H_0 is known, but very complicated to formulate.

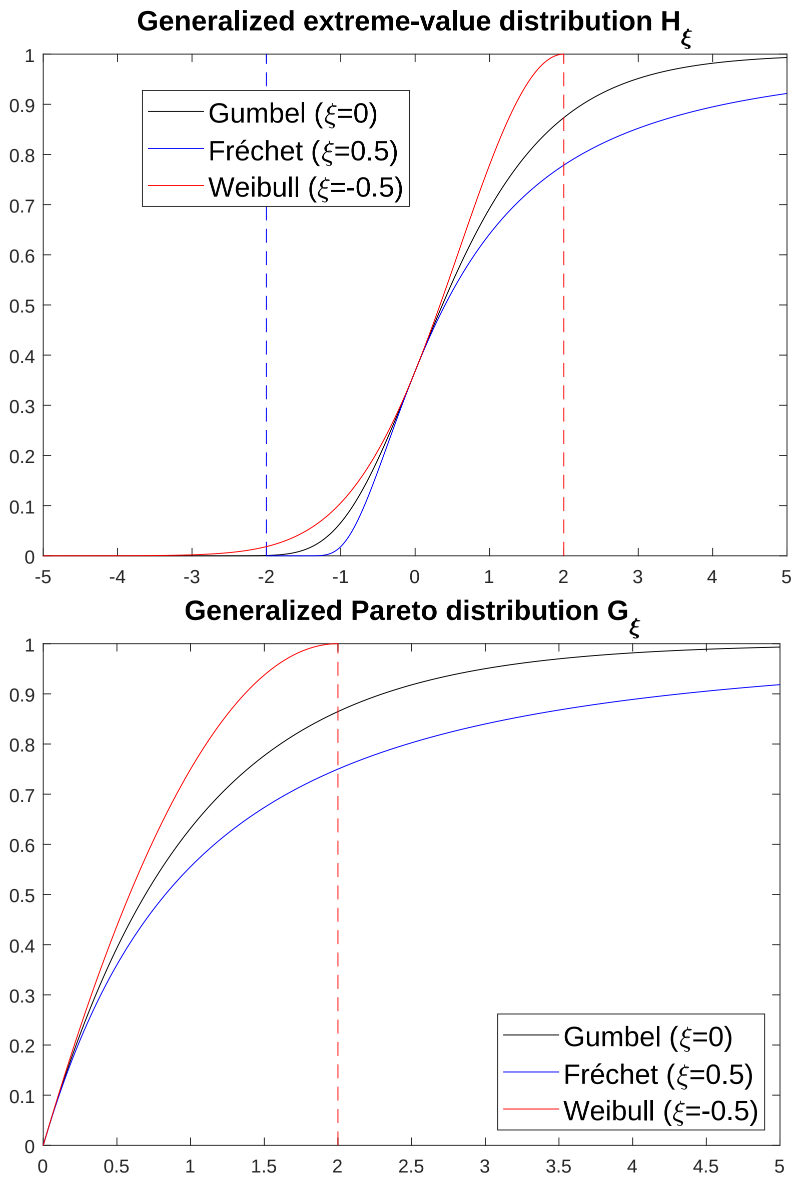

Fig. 1: Visualization of the distribution functions H_ξ and G_ξ. The dotted vertical lines indicate the end points of the supports of the respective distributions (if any).

- **Fréchet type $\xi > 0$:** This occurs typically if X is heavy-tailed. Concretely, F is in the domain of attraction of H_ξ if and only if $u_F = \infty$ and $1 - F$ is *regularly varying with index* $-1/\xi$, meaning that $1 - F(x) = L(x)\, x^{-1/\xi}$ for a function L with the defining property that $\lim_{x\to\infty} L(c\,x)/L(x) = 1$ for arbitrary $c > 0$. Such function L is called *slowly varying* and, intuitively, the tails of L are "light", so that the parameter ξ really determines the "heavyness" of the tail of F. Nevertheless, the introduction of L leaves more freedom to the modeler of some F than directly using the limit H_ξ.

- **Weibull type $\xi < 0$:** This occurs typically if X is bounded, like for a uniform distribution on a closed interval. Concretely, F is in the domain of attraction of H_ξ if and only if $u_F < \infty$ and the function $x \mapsto 1 - F(u_F - 1/x)$ is regularly varying with index ξ.

For the second important theorem, we are interested only in the tail of F, that is in the excess over a (high) threshold $u < u_F$ defined for $x \in [0, u_F - u]$ by

$$F_u(x) := \frac{F(x+u) - F(u)}{1 - F(u)} = \mathbb{P}(X - u \leq x \,|\, X > u).$$

Under the same assumptions as in the extremal types theorem, it states that F_u for large u behaves like a generalized Pareto distribution with shape parameter ξ.

Theorem 2.3 (Balkema, de Haan (1974) and Pickands (1975)) If F is in the domain of attraction of H_ξ, then there exists a (measurable) function $\sigma(u)$ such that

$$\lim_{u \to u_F} \sup_x \Big| F_u(x) - G_\xi\Big(\frac{x}{\sigma(u)}\Big) \Big| = 0.$$

Theorem 2.3 justifies to use the generalized Pareto distribution to model the tail of a random variable, just like the central limit law justifies to model the average behavior of a random variable via the normal distribution. As already mentioned, the case $\xi = 0$ corresponds to light tails, $\xi > 0$ corresponds to heavy tails, and $\xi < 0$ even to bounded tails.

2.2 The multivariate case

Now let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a random vector. We denote its (multivariate) distribution function by F, i.e. $F(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x})$ with "$\leq$" understood componentwise for vectors, and the distribution function of component X_i is denoted by F_i. The two main statements Theorems 2.2 and 2.3 as well as the concept of regular variation have multivariate counterparts.
First, regarding generalizations of Theorem 2.2, we consider independent copies $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots$ of $\boldsymbol{X}$. If there exist constants $a_{n,i} \in \mathbb{R}$, $b_{n,i} > 0$ such that the random vector of re-scaled componentwise maxima

$$\left(\frac{\max\{X_1^{(1)}, \ldots, X_1^{(n)}\} - a_{n,1}}{b_{n,1}}, \ldots, \frac{\max\{X_d^{(1)}, \ldots, X_d^{(n)}\} - a_{n,d}}{b_{n,d}} \right)$$

converges in distribution to some random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$, then necessarily

- each component Y_i has a generalized extreme-value distribution H_{ξ_i} (by the univariate theory),
- the distribution function C of the random vector of ranks of $\boldsymbol{Y}$, i.e. $\big(H_{\xi_1}(Y_1), \ldots, H_{\xi_d}(Y_d)\big)$, is an *extreme-value copula*, meaning that it satisfies

$$C(\boldsymbol{u}^t) = C(\boldsymbol{u})^t, \quad t > 0, \quad \boldsymbol{u} \in [0,1].$$

There exist many extreme-value copulas, i.e. possible dependence structures between the components of $\boldsymbol{Y}$. The following structural result provides a feeling for how many there are.

Theorem 2.4 (De Haan, Resnick (1977), Pickands (1981)) An extreme-value copula C can be written as

$$C(\boldsymbol{u}) = \exp\Big\{ - \|{-\log(\boldsymbol{u})}\|_{\boldsymbol{Q}} \Big\}, \quad \boldsymbol{u} \in [0,1],$$

where $\|.\|_{\boldsymbol{Q}}$ denotes a norm on $\mathbb{R}^d$, defined by

$$\|\boldsymbol{x}\|_{\boldsymbol{Q}} := d\,\mathbb{E}\big[\max\{|x_1|\,Q_1,\ldots,|x_d|\,Q_d\}\big],$$

with $\boldsymbol{Q} = (Q_1,\ldots,Q_d)$ a random vector with non-negative components satisfying $Q_1 + \ldots + Q_d = 1$ and $\mathbb{E}[Q_i] = 1/d$ for all components i. There is a one-to-one relationship between such $\boldsymbol{Q}$ and extreme-value copulas C.

"Typical" norms applied in other branches of mathematics arise as special cases of the family of norms $\|.\|_{\boldsymbol{Q}}$ in Theorem 2.4. For instance, by far the most popular extreme-value copulas in applications rely on the ℓ_p-norm $\|\boldsymbol{x}\|_{\boldsymbol{Q}} := (|x_1|^p + \ldots + |x_d|^p)^{1/p}$, respectively generalizations thereof. The standard ℓ_p-norm is indeed obtained for a specific choice of $\boldsymbol{Q}$ and is called *Gumbel copula*, but hierarchical versions thereof are also popular in order to account for asymmetries. Furthermore, there exists quite some literature that is dedicated to designing specific families of extreme-value copulas, investigating the relationships with the involved norms, the simulation of random vectors $\boldsymbol{Y}$ with given extreme-value copula, and to the estimation of an extreme-value copula based on observed data.[2]
The following list contains some qualitative properties that every extreme-value copula has, and which can be quantified in various ways (mentioned only in non-mathematical, intuitive terms here):

- The components of $\boldsymbol{Y}$ exhibit non-negative association. For instance, a correlation coefficient between Y_i and Y_j cannot be negative.
- There is a low probability that all components of $\boldsymbol{Y}$ are jointly small. Intuitively, this is because the limiting model $\boldsymbol{Y}$ focuses on componentwise maxima, thus provides no model flexibility away from the maxima. Like the multivariate normal distribution provides no flexibility to model heavy tails.
- There is a high probability that all components of $\boldsymbol{Y}$ are jointly large. Different specifications of C essentially control how high this probability is, and specifies the finer metrics of such joint occurrences between subgroups of the components.

Second, in analogy to the univariate Fréchet case, the notion of multivariate regular variation is a model for random vectors with heavy tails. On first glimpse it may appear reasonable to replace the single parameter ξ in the univariate case by a vector $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_d)$ of heavyness indices for all components, like this is done in the extreme-value copula approach above. However, in some financial applications one is specifically interested in studying the behavior of a weighted sum of risks and it is well

[2] In my personal perception, huge parts of this research are inspired more by the interesting mathematics involved than by the need for these methods in concrete applications. On the one hand, this is probably due to the fact that there exist many cross-relationships with different mathematical concepts like infinite divisibility and random measures, as well as harmonic analysis and norms. On the other hand, this might be due to the fact that this is an active field of research far from being completely understood, so researchers still investigate structural properties.

known that the heaviest tail dominates the weighted sum. Thus, one typically focuses on equally heavy tails for all individual risks in order to make a finer comparison or allocation investigation meaningful at all. Concretely, a random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ is said to be *(multivariate) regularly varying with index* $-1/\xi > 0$ if

$$\mathbb{P}\Big(\|\boldsymbol{Y}\| > x\, u,\ \frac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|} \in \mathrm{d}\boldsymbol{q} \,\Big|\, \|\boldsymbol{Y}\| > u \Big) \longrightarrow x^{-\frac{1}{\xi}}\, \rho(\mathrm{d}\boldsymbol{q}), \qquad (1)$$

as $u \to \infty$, where this limit is meant to be in some meaningful sense of measure convergence, and the limiting measure ρ is defined on Borel sets of the unit $\|.\|$-sphere for some given norm $\|.\|$ on $\mathbb{R}^d$ (whose choice plays no particular role). The measure ρ is called *spectral measure*. In words (1) means that conditioned on the fact that the norm of $\boldsymbol{Y}$ is larger than some (large) u, the measure ρ determines the angle of $\boldsymbol{Y}$ while the excess overshoot of the norm $\|\boldsymbol{Y}\|$ across the threshold u has regularly varying decay with index $-1/\xi$. In analogy to the univariate Fréchet case, such regularly varying random vectors are well-justified and very flexible models for random vectors whose components exhibit heavy tails. The shape parameter ξ models the tail-heavyness of the components, and the measure ρ intuitively plays the analogous role as does the extreme-value copula C of Theorem 2.4 in the aforementioned first multivariate approach. In fact, the probability measure of the vector $\boldsymbol{Q}$ associated with C in Theorem 2.4 is a measure on the ℓ_1-simplex, i.e. on the unit sphere with respect to the ℓ_1-norm, hence quite a similar object as ρ. In general, if one defines $\boldsymbol{Y}$ according to the first approach via standardized Fréchet marginals (i.e. $\xi_i = \xi > 0$ for all $i = 1, \ldots, d$) and an arbitrary extreme-value copula, then one obtains a random vector that is regularly varying with index $1/\xi$.[3] However, there are also other regularly varying random vectors, since the concept or multivariate regular variation is in analogy to the univariate case more general than just “working with the limit”. Thus, despite that this concept focuses on the Fréchet case $\xi > 0$ it is somehow more general on the dependence level than the extreme-value copula approach. A nice, elementary introduction to multivariate regular variation, including a list of references, can be found in Mikosch (2005).

3 How are these theoretical results applied in Finance?

The following are typical applications of such models in Finance:

- **Estimation of the Value-at-Risk associated with some asset:** Based on a time series of log returns $X_1, X_2, \ldots$ associated with some asset, one seeks to estimate an α-quantile for low α, say $\alpha = 1\%$ or 5%. Intuitively, this corresponds to estimating a (low) return level z such that the percentage $1-\alpha$ of all historically observed returns are above z. Converting such a value z into a cash amount, this number is usually called the *Value-at-Risk* and is used as a regulatory risk measure. Technically, there are several well-developed estimation techniques for this task, essentially making use of Theorems 2.2 and 2.3. For instance, according to Theorem 2.2 it is justified to assume that $-\max\{X_1, \ldots, X_k\}$ approximately has

[3] We recall this logic for the mathematically interested reader in the Appendix.

a distribution function of the type $x \mapsto H_\xi\big((x-a)/b\big)$ with unknown $a, \xi \in \mathbb{R}$ and $b > 0$. Notice that a, b result from the fact that the normalizing constants in Theorem 2.2 are unknown. The most common technique is the *block maxima method*. It consists of partitioning the sample $-X_1, \ldots, -X_n$ of observations into $k < n$ blocks, in each of the k blocks computing the minimal observation, and thus forming a new set of k observations that are interpreted as realizations of a random variable with distribution function $x \mapsto H_\xi\big((x-a)/b\big)$. Then standard parameter estimation techniques (method of moments or likelihood methods) are used to estimate the three parameters, and a quantile is computed from the resulting estimates. An alternative method is the *peaks-over-threshold method*, relying on Theorem 2.3. It basically relies on discarding observations smaller than a given threshold and fitting a generalized Pareto distribution to the remaining observations. Optimal choices of the threshold and further technical issues are not further discussed in this brief survey, see Brodin, Klüppelberg (2014) for more details and references.

Intuitively, the difficulty in this kind of application is that one wishes to estimate a quantity that is potentially so far out in the tail that one has only very few (or even none) observations. Somehow, one thus has to "extrapolate" in a reasonable way, how bad it can become based on the observations. Extreme-value theory provides a justifiable way to accomplish that, and may thus be considered the best one can do in this regard.

- **Qualitative investigations on the aggregation of risk measures:** Since the financial crisis 2008, traditional risk measurement approaches are criticized sharply. One of the main points of criticism is that common methods of aggregating risk measures for single positions to a portfolio risk measurement ignore heavy tails. From a mathematical point of view, this relies precisely on the fact that such approaches rely on the multivariate normal distribution (correlations), and thus somehow by construction ignore heavy tails. The methods of extreme-value theory provide a well-justified mathematical wrapping to demonstrate (theoretically) how false such traditional aggregation methods can be, and the (high-ranked) academic literature is packed with such examples. For instance, it is well-known that the tail of a portfolio of risks is heavy already if at least one single constituent is heavy-tailed. Speaking figuratively, having to mix a drink from several unlabeled bottles, knowing that exactly one of them contains deadly poison, it is certainly not the best idea to mix in a little bit of every bottle! This simple observation stands in glaring contrast to the diversification benefit paradigm that is implicit in the multivariate normal distribution approaches, such as Markowitz for instance. Such investigations stay mainly academic, pointing the finger at bad practice. However, since the multivariate extreme-value theory is rather complex, concrete practical improvements, relying on such tools, are not very popular to the best of my knowledge. There still seems to be a gap between academia and practice. The reason is that finding a reasonable trade-off between practicability and realism is rather difficult. To provide some examples nevertheless, Stărică (1999); Hauksson et al. (2001) use it to empirically analyze tails of high-frequency foreign exchange data, and Mainik, Rüschendorf (2010) have applied the setting of regularly varying random vectors $\boldsymbol{Y}$ in (1) to minimize portfolio risk.

- **Heavy-tailed models in general:** In several financial applications that traditionally rely on the multivariate normal distribution people have changed to alternative distributions with heavier tails, sometimes taken from multivariate extreme-value theory. This means that the methods of extreme-value theory are used for reasons that deviate from its original narrow focus of "estimating tail event probabilities". For instance, certain low-parametric extreme-value copulas have been applied as substitute for the Gaussian copula (of a multivariate normal distribution) when pricing CDOs or modeling multivariate log-returns. Such approaches introduce heavy tails into the distributions under consideration, which is their main motivation. However, it is important to be aware that such applications only use some desirable aspects of extreme-value models, but in situations which are not the classical extreme-value applications. For instance, when modeling a vector of log-returns of several assets with an extreme-value copula one typically does not interpret this vector as the realization of componentwise maxima of independent realizations of some other vector. Instead, one only defines "some" model with the distinctive property that certain cataclysmic events have non-negligible probabilities. This can be a totally justifiable reason for choosing this model, but this justification then needs to be done case-by-case.

Appendix For the mathematically interested reader we explain that a random vector $\boldsymbol{Y}$ whose marginals are of the form H_ξ for some positive ξ and whose copula is an extreme-value copula is multivariate regularly varying, and explain how the spectral measure ρ is related to the probability measure of $\boldsymbol{Q}$ in Theorem 2.4. The presented logic originates from the seminal reference De Haan, Resnick (1977). Fix $\xi > 0$ and let $\boldsymbol{Q}$ be a random vector such as in Theorem 2.4, and denote by $\|.\|_{\boldsymbol{Q}}$ the associated norm, and by C the associated extreme-value copula. Let $\boldsymbol{Y}$ be a random vector with distribution function defined by

$$F(\boldsymbol{y}) := \mathbb{P}(\boldsymbol{Y} \leq \boldsymbol{y}) = C\big(e^{-y_1^{-1/\xi}}, \ldots, e^{-y_d^{-1/\xi}}\big) = e^{-\|\boldsymbol{y}^{-1/\xi}\|_{\boldsymbol{Q}}}.$$

Notice that $\mathbb{P}(Y_i \leq y_i) = H_\xi\big((y_i - 1)/\xi\big)$, so each component Y_i is modeled by a standardized Fréchet distribution with heavyness parameter ξ. As a first technical step, we notice that there exists a random vector $\boldsymbol{R}$ taking values in the standard unit simplex, unique in law, such that

$$\mathbb{E}[\max_k\{Q_k\, x_k\}] = \mu_\xi\, \mathbb{E}[\max_k\{R_k^{1/\xi}\, x_k\}], \quad \mu_\xi := \mathbb{E}[\|\boldsymbol{Q}\|_\xi],$$

holds for arbitrary $\boldsymbol{x} \geq 0$. Notice that $\boldsymbol{R}^{1/\xi}$ takes values in the standard ℓ_ξ-simplex. The uniqueness of such $\boldsymbol{R}$ may for instance be deduced from (Ressel, 2012, Theorem 4(ii)), and clearly

$$\mathbb{P}(\boldsymbol{R} \in A) = \mathbb{E}\Big[1_{\big\{\frac{\boldsymbol{Q}^\xi}{\|\boldsymbol{Q}\|_\xi^\xi} \in A\big\}} \frac{\|\boldsymbol{Q}\|_\xi}{\mu_\xi}\Big].$$

Further, we observe that

$$\begin{aligned} \left\|\boldsymbol{y}^{-1/\xi}\right\|_{\boldsymbol{Q}} &= d\,\mathbb{E}[\max_k\{Q_k\,y_k^{-1/\xi}\}] = d\,\mu_\xi\,\mathbb{E}[\max_k\{R_k^{1/\xi}\,y_k^{-1/\xi}\}] \\ &= d\,\mu_\xi\,\mathbb{E}[\min_k\{y_k/R_k\}^{-1/\xi}] = d\,\mathbb{E}\Big[\int_{\min_k\{y_k/q_k\}}^{\infty} \xi\,r^{-\frac{1}{\xi}-1}\,\mathrm{d}r\Big] \\ &= d\,\mu_\xi \int\int_{\min_k\{y_k/q_k\}}^{\infty} \frac{1}{\xi}\,r^{-\frac{1}{\xi}-1}\,\mathrm{d}r\,\mathbb{P}(\boldsymbol{R} \in \mathrm{d}\boldsymbol{q}). \end{aligned} \tag{2}$$

Now we define a (Radon) measure ν_ξ on $[\boldsymbol{0}, \infty] \setminus \{\boldsymbol{0}\}$ as follows:

$$\nu_\xi\big(B(r, A)\big) := d\,\mu_\xi\,r^{-1/\xi}\,\mathbb{P}(\boldsymbol{R} \in A),$$

where $r > 0$ and A is a subset of the standard unit simplex, and

$$B(r, A) := \{\boldsymbol{z}\,:\,\|\boldsymbol{z}\|_1 > r,\,\boldsymbol{z}/\,\|\boldsymbol{z}\|_1 \in A\}.$$

By construction, we have

$$\begin{aligned} \nu_\xi\big(B(s, A)\big) &= d\,\mu_\xi\,s^{-1/\xi}\,\mathbb{P}(\boldsymbol{R} \in A) \\ &= d\,\mu_\xi \int_s^\infty \frac{1}{\xi}\,r^{-\frac{1}{\xi}-1}\,\mathrm{d}r\,\mathbb{P}(\boldsymbol{R} \in A), \end{aligned}$$

or, formulated differently, under the polar coordinate transformation $T\,:\,\boldsymbol{x} \mapsto (\|\boldsymbol{x}\|_1\,, \boldsymbol{x}/\,\|\boldsymbol{x}\|_1)$ we have

$$\nu_\xi \circ T^{-1}(\mathrm{d}r, \mathrm{d}\boldsymbol{q}) = d\,\mu_\xi\,\frac{1}{\xi}\,r^{-\frac{1}{\xi}-1}\,\mathrm{d}r\,\mathbb{P}(\boldsymbol{R} \in \mathrm{d}\boldsymbol{q}).$$

Summarizing, since

$$T\big([\boldsymbol{0}, \boldsymbol{y}]^c\big) = \big\{(r, \boldsymbol{q})\,:\,r > \min_k\{y_k/q_k\}\big\},$$

we observe from (2) that

$$\left\|\boldsymbol{y}^{-1/\xi}\right\|_{\boldsymbol{Q}} = \nu_\xi \circ T^{-1}\Big(T\big([\boldsymbol{0}, \boldsymbol{y}]^c\big)\Big) = \nu_\xi\big([\boldsymbol{0}, \boldsymbol{y}]^c\big).$$

We observe that $F(n^\xi\,\boldsymbol{y})^n = F(\boldsymbol{y})$ holds for arbitrary $n \in \mathbb{N}$. This property is called *max-stability* and is precisely what characterizes limiting extreme-value distributions. Using the fact that $\lim_{t\searrow 0}(1 - \exp\{-t\,x\})/t = x$ for arbitrary $x \geq 0$ in $(*)$ and $n^\xi\,[\boldsymbol{0}, \boldsymbol{y}]^c = [\boldsymbol{0}, n^\xi\,\boldsymbol{y}]^c$ in the second equality below

$$\begin{aligned} \lim_{n\to\infty} n\,\mathbb{P}\big(\boldsymbol{Y} \in n^\xi\,[\boldsymbol{0}, \boldsymbol{y}]^c\big) &= \lim_{n\to\infty} n\,\big(1 - F(n^\xi\,\boldsymbol{y})\big) \\ &\overset{(*)}{=} -\log\big(F(\boldsymbol{y})\big) = \nu_\xi\big([\boldsymbol{0}, \boldsymbol{y}]^c\big). \end{aligned} \tag{3}$$

We obtain from this for (almost) arbitrary Borel sets A the limit $\lim_{n\to\infty} n\,\mathbb{P}(\boldsymbol{Y} \in n^\xi\,A) = \nu_\xi(A)$. In particular, making use of $n^\xi\,B(s, A) = B(n^\xi\,s, A)$ we observe

$$\begin{aligned} \lim_{n\to\infty} n\,\mathbb{P}\Big(\|\boldsymbol{Y}\|_1 > n^\xi\,s,\,\frac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|_1} \in A\Big) &= \lim_{n\to\infty} n\,\mathbb{P}\big(\boldsymbol{Y} \in B(n^\xi\,s, A)\big) \\ &= \nu_\xi\big(B(s, A)\big) = d\,\mu_\xi\,s^{-1/\xi}\,\mathbb{P}(\boldsymbol{R} \in A). \end{aligned}$$

Setting $s = 1$ and $A = S_d$ equal to the standard unit simplex, we obtain with limit rules

$$\lim_{n\to\infty} \frac{\mathbb{P}\Big(\|\boldsymbol{Y}\|_1 > n^{\xi}\, s,\ \frac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|_1} \in S_d \Big)}{\mathbb{P}\Big(\|\boldsymbol{Y}\|_1 > n^{\xi},\ \frac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|_1} \in S_d \Big)} = s^{-1/\xi}.$$

And now combining these two limits, we actually obtain (1) with threshold substitution $u = n^{\xi}$ and spectral measure given by $\rho(\mathrm{d}\boldsymbol{q}) = d\,\mu_{\xi}\,\mathbb{P}(\boldsymbol{R} \in \mathrm{d}\boldsymbol{q})$.
The precisely identical logic with an arbitrary random vector $\tilde{\boldsymbol{Y}}$ that lies in the domain of attraction of $\boldsymbol{Y}$ (which equals an extreme-value limit) also leads to a multivariate regularly varying model with identical spectral measure. This generalization relies on the possibility to generalize the limiting relationship (3).

References

A. BALKEMA, L. DE HAAN, Residual life time at great age, Annals of Probability 2 (1974) pp. 792–804.

J. BEIRLANT, Y. GOEGEBEUR, J. SEGERS, J. TEUGELS, Statistics of extremes: theory and applications, Wiley, Chichester (2004).

L. DE HAAN, A. FERREIRA, Extreme value theory - an introduction, Springer, New York (2006).

L. DE HAAN, S.I. RESNICK, Limit theory for multivariate sample extremes, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 40 (1977) pp. 317–337.

P. EMBRECHTS, C. KLÜPPELBERG, T. MIKOSCH, Modelling extremal events for insurance and finance, Springer, Berlin (1997).

R.A. FISHER, L.H.C. TIPPETT, Limiting forms of the frequency distribution of the largest and smallest member of a sample, Proceedings of the Cambridge Philosophical Society 24 :2 (1928) pp. 180–190.

B.V. GNEDENKO, Sur la distribution limite du terme maximum d'une serie aleatoire, Annals of Mathematics 44 :3 (1943) pp. 423–453.

E. GUMBEL, Statistics for extremes, Columbia University Press, New York (1958).

H. HAUKSSON, M. DACOROGNA, T. DOMENIG, U. MÜLLER, G. SAMORODNITSKY, Multivariate extremes, aggregation and risk estimation, Quantitative Finance 1:1 (2001) pp. 79–95.

E. BRODIN, C. KLÜPPELBERG, Extreme Value Theory in Finance, in Encyclopedia in Quantitative Risk Analysis and Assessment, John Wiley & Sons (2014) pp. 1–11.

G. MAINIK, L. RÜSCHENDORF, On optimal portfolio diversification with respect to extreme risks, Finance and Stochastics 14 (2010) pp. 593–623.

T. MIKOSCH, How to model multivariate extremes if one must?, Statistica Neerlandica 59:3 (2005) pp. 324–338.

J. PICKANDS, Statistical inference using extreme order statistics, Annals of Statistics 3 (1975) pp. 119–131.

J. PICKANDS, Multivariate extreme value distributions, Proceedings of the 43rd Session ISI, Buenos Aires (1981) pp. 859–878.

S.I. RESNICK, Extreme values, regular variation and point processes, Springer, New York (1987).

P. RESSEL, Homogeneous distributions - and a spectral representation of classical mean values and stable tail dependence functions, Journal of Multivariate Analysis 117 (2012) pp. 246–256.

C. STĂRICĂ, Multivariate extremes for models with constant conditional correlations, Journal of Empirical Finance 6 (1999) pp. 515–553.