MaPhySto
The Danish National Research Foundation:
Network in Mathematical Physics and Stochastics

Research Report
no. 21 October 2004

Thomas Mikosch:
How to model multivariate extremes if one must?

ISSN 1398-2699

How to model multivariate extremes if one must?

THOMAS MIKOSCH[1]

Laboratory of Actuarial Mathematics, University of Copenhagen,
Universitetsparken 5, DK-2100 Copenhagen, Denmark, mikosch@math.ku.dk

Abstract

In this paper we discuss some approaches to modeling extremely large values in multivariate time series. In particular, we discuss the notion of multivariate regular variation as key to modeling multivariate heavy-tailed phenomena. The latter notion has found a variety of applications in queuing theory, stochastic networks, telecommunications, insurance, finance and other areas. We contrast this approach with modeling multivariate extremes by using the multivariate student distribution and copulas.

Key Words and Phrases: Multivariate regular variation, heavy-tailed distribution, extreme value distribution, copula, elliptical distribution

1 Introduction

Over the last few years heavy-tailed phenomena have attracted a lot of attention. Those include turbulences and crashes of the financial and insurance markets, but also strong deviations of weather and climate phenomena from the average behavior. More recently, the Internet and more generally the enormous increase of computer power have led to collections of huge data sets which cannot be handled by classical statistical methods. Among others, teletraffic data (such as on/off times of computers, lengths and transfer times of files, etc.) exhibit not only a non-standard dependence structure which cannot be described by the methods of classical time series analysis, but these data also have clusters of unusually large data. It has been recognized early on in hydrology and meteorology, but also in insurance practice that the distributions of the classical statistical theory (such as the normal and the gamma family) are of restricted use for modeling the data at hand. The description of these data by the median, expectation, variance or by moment related quantities such as the kurtosis and skewness are of rather limited value in this context. For the actuary it is not a priori of interest to know what the expectation and the variance of the data are, be he is mainly concerned with large claims which might arise from scenarios similar to the WTC disaster. Such events are extremely rare and dangerous. It would be silly to use the (truncated) normal or the gamma distributions to capture such an event by a mean-variance analysis. The devil sits in the tail of the distribution. It is the tail of the distribution that costs the insurance and

[1]Thomas Mikosch's research is partially supported by MaPhySto, The Danish Research Foundation: Network in Mathematical Physics and Stochastics, and DYNSTOCH, a research training network under the programme Improving Human Potential financed by The 5th Framework Programme of the European Commission. Financial support by the Danish Research Council (SNF) Grant No 21-01-0546 is gratefully acknowledged by the author.

financial industry billions of dollars (only the WTC disaster has cost the reinsurance industry about 20 billion \$ US by now; see Sigma (2003)). It is the tail of the distributions of the file sizes and transition times of files that causes the unpleasant behavior of our computer networks. Although the Internet is a traffic system which, unlike the German Autobahn, has thousands of extremely fast lanes, this system is not always able to handle the amounts of information to be transferred and, like the German Autobahn, is subject to traffic jam caused by huge files representing movies, pictures, DVDs, CDs which make the difference. They cannot be modeled by the exponential or gamma distributions, very much in contrast to classical queuing and network theory, where the exponential distribution was recognized as adequate for modeling human behavior in telephone or other costumer-service systems. By now there is general agreement that the modern teletraffic systems are well described by distributions with tails much heavier than the exponential distribution; see e.g. Willinger et al. (1995); cf. Mikosch et al. (2002) and Stegeman (2002).

Early on, distributions with power law tails have been used in applications to model extremely large values. The Pareto distribution was introduced in order to describe the distribution of income in a given population (Pareto (1896/97)). Although the world has changed a lot since Pareto suggested this distribution at the end of the 19th century, it still gives a very nice fit to the world income distribution. The Pareto distribution is also a standard distribution for the purposes of reinsurance, where the largest claims of a portfolio are taken care of. The Pareto distribution in its simplest form can be written as

$$\overline{F}(x) = 1 - F(x) = (c/x)^{\alpha}\,, \quad x \geq c\,, \quad \text{some positive } c.$$

As expected this distribution does in general not give a great fit to data in the center of the distribution, but it often captures the large values of the data in a convincing way. Of course, one can shift the distribution to the origin by introducing a location parameter, but the fit in the center would not become much better in this way.

The Pareto distribution appears in a completely different theoretical context, namely as the limit distribution of the excesses of an iid sequence $X_1, \ldots, X_n$ with distribution F above a high threshold. To be more precise, the only limit distribution of the excess distribution of the X_i's is necessarily of the form (up to changes of location and scale)

$$\lim_{u \uparrow x_F, u+x < x_F} P(X_1 - u > x \mid X_1 > u) \quad \to \quad (1 + \xi\, x)_+^{-1/\xi} = \overline{G}_{\xi}(x)\,, \quad x \in \mathbb{R}\,, \tag{1.1}$$

(Pickands (1975), Balkema and de Haan (1974), cf. Embrechts et al. (1997), Section 3.4), where

$$x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$$

is the right endpoint of the distribution F and the shape parameter $\xi \in \mathbb{R}$. For $\xi = 0$ the limit has to be interpreted as the tail of the standard exponential distribution.

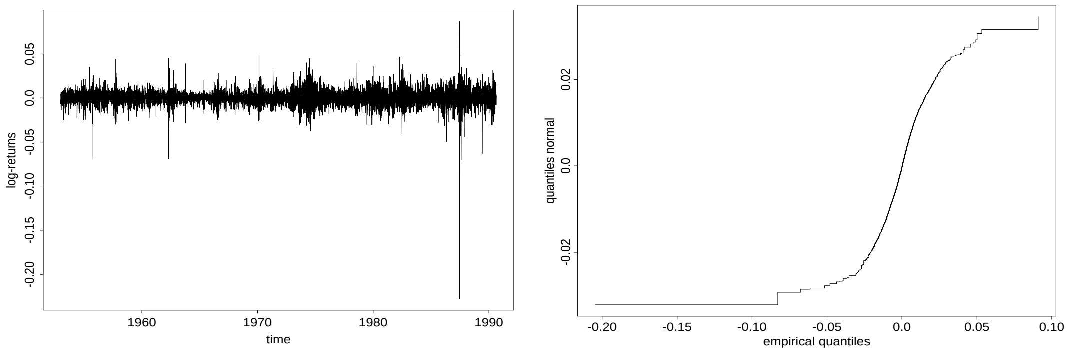

Figure 1.1 *Left: Plot of* 9558 S&P500 *daily log-returns from January* 2, 1953, *to December* 31, 1990. *The year marks indicate the beginning of the calendar year. Right: QQ-plot of the* S&P500 *data against the normal distribution whose mean and variance are estimated from the data. The data come from a distribution which has much heavier left and right tail than the normal distribution.*

The so defined limit distribution G_ξ is called the *generalized Pareto distribution* (GPD). Relation (1.1) holds only for a restricted class of distributions F. Indeed, (1.1) is satisfied if and only if a limit relation of the following type holds for suitable constants $d_n \in \mathbb{R}$, $c_n > 0$, and the partial maxima $M_n = \max(X_1, \ldots, X_n)$ (see Embrechts et al. (1997), Chapter 3):

$$(1.2) \qquad P(c_n^{-1}(M_n - d_n) \le x) \;\to\; \exp\left\{-(1+\xi\,x)_+^{-1/\xi}\right\} = H_\xi(x)\,, \qquad n \to \infty\,, x \in \mathbb{R}\,.$$

For $\xi = 0$ the distribution has to be interpreted as the Gumbel distribution $H_0(x) = \mathrm{e}^{-\mathrm{e}^{-x}}$. The limit distribution is called the *generalized extreme value distribution* and (up to changes of scale and location) it is the only possible non-degenerate limit distribution for centered and normalized maxima of iid sequences. We say that the underlying distribution F belongs to the *maximum domain of attraction of the extreme value distribution* H_ξ ($F \in \mathrm{MDA}(H_\xi)$). The case $\xi > 0$ is particularly interesting for modeling extremes with unlimited values. Then the extreme value distribution H_ξ can be reparametrized and written as the so-called *Fréchet distribution* with $\alpha = \xi^{-1}$:

$$\Phi_\alpha(x) = \mathrm{e}^{-x^{-\alpha}}\,, \quad x > 0\,.$$

Every distribution $F \in \mathrm{MDA}(\Phi_\alpha)$ is completely characterized by the relation

$$(1.3) \qquad \overline{F}(x) = 1 - F(x) = \frac{L(x)}{x^\alpha}\,, \quad x > 0\,,$$

where L is a *slowly varying* function, i.e., L is a positive function on $(0,\infty)$ with property $L(cx)/L(x) \to 1$ as $x \to \infty$ for every $c > 0$. Notice that distributional tails of type (1.3) are a slight generalization of distributions with pure power law tails such as the Pareto distribution. It is a semiparametric description of a large class of distributions; the slowly varying functions L represent a nuisance parameter which is not further specified. This is very much in agreement with real-life data analyzes where it is hard to believe that the data come from a pure Pareto distribution. In particular, L is not specified in any finite interval which leaves the question about the form of the distribution F in its center open. Several distributions with a name have *regularly varying* right tail, i.e., (1.3) holds, e.g. the Pareto, Burr, log-gamma, student, Cauchy, Fréchet and infinite variance stable distributions; see Embrechts et al. (1997), p. 35, for definitions of these distributions.

The stable distributions consist of the only possible non-degenerate limit distributions H for the partial sums $S_n = X_1 + \cdots + X_n$ of an iid sequence (X_i) with distribution F, i.e., there exist constants $c_n > 0$, $d_n \in \mathbb{R}$, such that

$$\lim_{x\to\infty} P(c_n^{-1}(S_n - d_n) \le x) = H(x)\,, \quad x \in \mathbb{R}\,.$$

We say that F belongs to the *domain of attraction* of the stable distribution H ($F \in \mathrm{DA}(H)$). The best known stable distribution is the normal whose domain of attraction contains all F with slowly varying truncated second moment $\int_{|y|\le x} y^2 dF(x)$ (Feller (1971)), i.e., it contains almost all distributions of interest in statistics. The remaining stable distributions are less known; they have infinite variance and so are the members of their domains of attraction. In particular, every infinite variance stable distribution is characterized by a shape parameter $\alpha \in (0,2)$ which appears as the tail parameter of these distributions H_α. Moreover, it also appears in the tails of distributions $F \in \mathrm{DA}(H_\alpha)$:

$$F(-x) \sim p\,\frac{L(x)}{x^\alpha} \quad \text{and} \quad \overline{F}(x) \sim q\,\frac{L(x)}{x^\alpha}\,, \quad x > 0\,, \tag{1.4}$$

where L is slowly varying and $p, q \ge 0$ such that $p + q = 1$. Relation (1.4) is also referred to as *tail balance condition* and F is said to be *regularly varying with index* $\alpha \in (0,2)$.

Regular variation also occurs in a surprising way in solutions to stochastic recurrence equations. We consider here the simplest one-dimensional case. Assume the *stochastic recurrence equation*

$$Y_t = A_t Y_{t-1} + B_t\,, \quad t \in \mathbb{Z}\,, \tag{1.5}$$

has a strictly stationary causal solution, where $((A_t, B_t))$ constitute an iid sequence of non-negative random variables. Causality refers to the fact that Y_t is a function only of (A_s, B_s), $s \le t$. A sufficient condition for the existence of such a solution is given by $E\log^+ A_0 < \infty$, $E\log^+ B_0 < \infty$ and $E\log A_0 < 0$. Equations of type (1.5) occur in the context of financial time series models. For example, the celebrated (2003 Nobel prize winning) ARCH and GARCH models of Engle (1982) and Bollerslev (1986) can be embedded in a stochastic recurrence equation. We

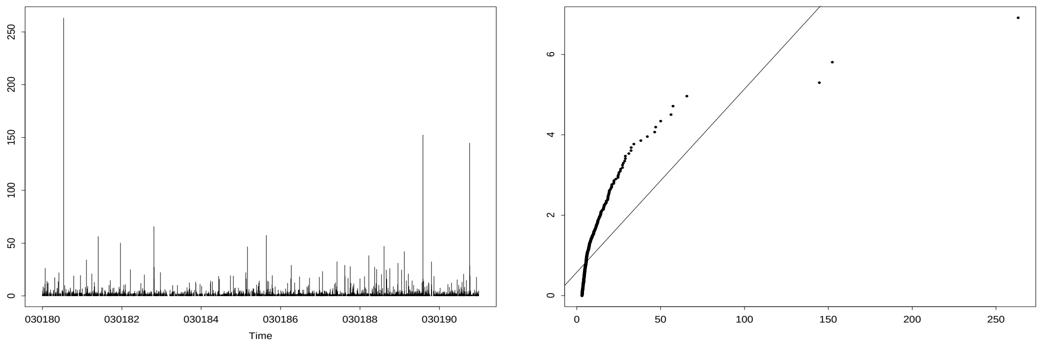

Figure 1.2 2493 *Danish fire insurance claims in Danish Kroner from the period* 1980–1992*. The data* (*left*) *and a QQ-plot of the data against standard exponential quantiles* (*right*)*. The data have tail much heavier than the exponential distribution.*

illustrate this with the GARCH(1,1) model (*generalized autoregressive conditionally heteroscedastic model of order* (1,1)) which is given by the equation

$$X_t = \sigma_t\, Z_t\,,\quad \sigma_t^2 = \alpha_0 + \alpha_1\, X_{t-1}^2 + \beta_1 \sigma_{t-1}^2\,,\quad t \in \mathbb{Z}\,.$$

Here (Z_t) is an iid sequence with $EZ_1 = 0$ and $\text{var}(Z_1) = 1$ and $\alpha_0 > 0$, α_1, β_1 are non-negative parameters. Obviously, $Y_t = \sigma_t^2$ satisfies the stochastic recurrence equation (1.5) with $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$ and $B_t = \alpha_0$. An important result by Kesten (1973) (see also Goldie (1991)) says that under general conditions on the distribution of A_0 the equation

$$EA_0^\kappa = 1 \tag{1.6}$$

has a unique positive solution κ_1 and then for some $c > 0$,

$$P(Y_0 > x) \sim c\, x^{-\kappa_1}\,,\quad x \to \infty\,. \tag{1.7}$$

In particular, the mentioned ARCH and GARCH models have marginal distribution with regularly varying tail of type (1.7). For a GARCH(1,1) model, (1.6) turns into

$$E(\alpha_1 Z_0^2 + \beta_1)^\kappa = 1\,,$$

which has a solution κ_1, e.g. when Z_0 is normally distributed. Hence $Y_0 = \sigma_0^2$ satisfies (1.6) and a standard argument on regular variation implies that $P(X_0 > x) \sim c' x^{-\kappa_1/2}$. This is a rather surprising result which says that light-tailed input (noise) can cause heavy-tailed output in a non-linear time series. Such a result is impossible for linear processes (such as ARMA processes) driven by iid noise; see e.g. Embrechts et al. (1997), Appendix A3, or Mikosch and Samorodnitsky (2000). We refer to Basrak et al. (2000a,b) as general references on GARCH and regular variation, to Embrechts et al. (1997), Section 8.4, for an introduction to stochastic recurrence equations and the tails of their solutions. See also Mikosch (2003) for a survey paper on financial time series models, their extremes and regular variation.

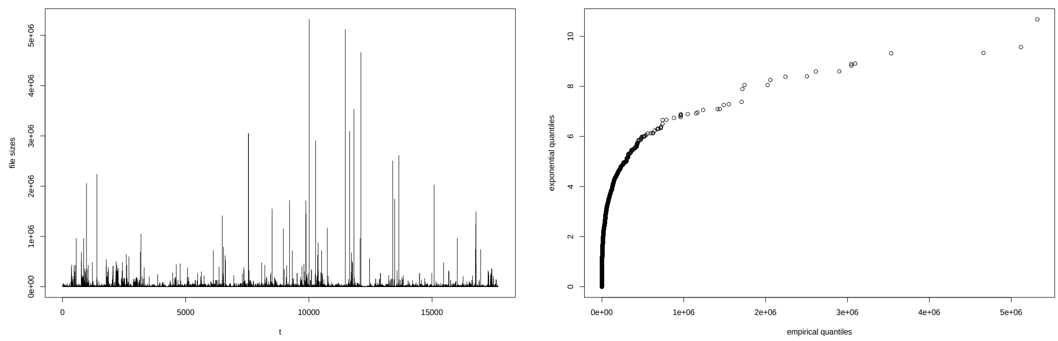

Figure 1.3 *A time series of measured file sizes handled by a webserver. The data* (*left*) *and a QQ-plot of the data against standard exponential quantiles* (*right*)*. The data have tail much heavier than the exponential distribution.*

The conclusion of this introduction should be that one-dimensional distributions with regularly varying tails are very natural for modeling extremal events when large values are involved. Since there exists some theoretical background why these distributions occur in different contexts, it is only consequent to fit them to real-life data. On the other hand, distributions such as the gamma, the exponential or the normal distributions are less appropriate for fitting extremes when large values occur.

It is the aim of the next section to discuss multivariate regular variation as a suitable tool for modeling multivariate extremes. We continue in Section 3 by discussing some of the alternative approaches such as the multivariate student distribution and copulas for modeling multivariate data with very large values.

2 Multivariate regularly varying distributions

Over the last few years there has been some search for multivariate distributions which might be appropriate for modeling very large values such as present in financial or insurance portfolios. The multivariate Gaussian distribution is not an appropriate tool in this context since there is strong evidence that heavy-tailed marginal distributions are present. Nevertheless it has become standard in risk management to apply the multivariate Gaussian distribution, e.g. for calculating the Value at Risk (VaR). A major reason for the use of the normal distribution is its "simplicity": linear combinations of the components of a normally distributed vector $\mathbf{X}$ are normal and the distribution of $\mathbf{X}$ is completely determined by its mean and covariance structure.

It is the aim of this section to introduce multivariate distributions which have, in a sense, power law tails in all directions. We will also indicate that these distributions appear in a natural way, e.g. as domain of attraction conditions for weakly converging partial sums and componentwise maxima of iid vectors.

To start with, we rewrite the defining property of a one-dimensional regularly varying distribution F (see (1.4)) as follows: for every $t > 0$,

$$\lim_{x\to\infty} \frac{P(x^{-1}X_1 \in (t,\infty])}{P(|X_1| > x)} = q\, t^{-\alpha} = \mu(t,\infty]\,, \tag{2.8}$$

$$\lim_{x\to\infty} \frac{P(x^{-1}X_1 \in [-\infty,t))}{P(|X_1| > x)} = p\, t^{-\alpha} = \mu[-\infty,-t]\,. \tag{2.9}$$

These relations are immediate from the properties of regularly varying functions; see e.g. Bingham et al. (1987). Notice that the right hand expressions can be interpreted as the μ-measure of the sets $(t,\infty]$ and $[-\infty,-t)$, where μ is defined on the Borel sets of $\overline{\mathbb{R}}\backslash\{0\}$:

$$d\mu(x) = \alpha\,[p\,|x|^{-\alpha-1}\, I_{[-\infty,0)}(x) + q\, x^{-\alpha-1} I_{(0,\infty]}(x)]\, dx\,.$$

Relations (2.8) and (2.9) can be understood as convergence of measures:

$$\mu_x(\cdot) = \frac{P(x^{-1}X_1 \in \cdot)}{P(|X_1| > x)} \stackrel{v}{\to} \mu(\cdot)\,, \quad x \to \infty\,, \tag{2.10}$$

where $\stackrel{v}{\to}$ refers to *vague convergence* on the Borel σ-field of $\overline{\mathbb{R}}\backslash\{0\}$. This simply means in our context that $\mu_x(A) \to \mu(A)$ for every Borel set $A \subset \overline{\mathbb{R}}\backslash\{0\}$ which is bounded away from zero and satisfies $\mu(\partial A) = 0$. We refer to Kallenberg (1983) or Resnick (1987) for the definition and properties of vaguely converging measures.

Relation (2.10) allows one to extend the notion of regular variation to Euclidean space. Indeed, we say that the vector $\mathbf{X}$ with values in $\mathbb{R}^d$ and its distribution are *regularly varying with limiting measure* μ if the relation

$$\mu_x(\cdot) = \frac{P(x^{-1}\mathbf{X} \in \cdot)}{P(|\mathbf{X}| > x)} \stackrel{v}{\to} \mu(\cdot)\,, \quad x \to \infty\,, \tag{2.11}$$

holds for a non-null measure μ on the Borel σ-field of $\overline{\mathbb{R}}^d \setminus \{\mathbf{0}\}$. Again, this relation means nothing but $\mu_x(A) \to \mu(A)$ for any set $A \subset \overline{\mathbb{R}}^d \setminus \{\mathbf{0}\}$ which is bounded away from zero and satisfies $\mu(\partial A) = 0$. Regular variation of $\mathbf{X}$ implies regular variation of $|\mathbf{X}|$ with a positive index α and therefore

$$\mu_x(t\,A) = \frac{P(x^{-1}\mathbf{X} \in tA)}{P(|\mathbf{X}| > tx)}\,\frac{P(|\mathbf{X}| > tx)}{P(|\mathbf{X}| > x)} \to \mu(A)\, t^{-\alpha}\,.$$

This means that μ satisfies the *homogeneity property* $\mu(tA) = t^{-\alpha}\mu(A)$, and we therefore also say that $\mathbf{X}$ is *regularly varying with index* α.

Now define the sets

$$A(t,S) = \{\mathbf{x} : |\mathbf{x}| > t\,, \widetilde{\mathbf{x}} \in S\}\,,$$

where $t > 0$ and $S \subset \mathbb{S}^{d-1}$, the unit sphere of $\mathbb{R}^d$ with respect to a given norm, and

$$\widetilde{\mathbf{x}} = \mathbf{x}/|\mathbf{x}|\,.$$

The defining property of regular variation also implies that

$$
\begin{aligned}
(2.12) \qquad \mu_x(A(t,S)) &= \frac{P(|\mathbf{X}| > tx\,, \widetilde{\mathbf{X}} \in S)}{P(|\mathbf{X}| > x)} \\
&\to \ \mu(A(t,S)) = t^{-\alpha}\mu(A(1,S))\,,
\end{aligned}
$$

where we assume that $\mu(\partial A(1,S)) = 0$. Since sets of the form $A(t,S)$ generate vague convergence in $\overline{\mathbb{R}}^d\backslash\{\mathbf{0}\}$, (2.11) and (2.12) both define regular variation of $\mathbf{X}$. The totality of the values $\mu(A(1,S))$, for any Borel set $S \subset \mathbb{S}^{d-1}$, defines a probability measure $P(\mathbf{\Theta} \in \cdot)$ on the Borel σ-field of $\mathbb{S}^{d-1}$, the so-called *spectral measure* of μ. The spectral measure $P(\mathbf{\Theta} \in \cdot)$ and the index $\alpha > 0$ completely determine the measure μ.

The notion of multivariate regular variation is a natural extension of one-dimensional regular variation. Indeed, regular variation of iid $\mathbf{X}_i$'s with index $\alpha \in (0,2)$ is equivalent to the property that centered and normalized partial sums $\mathbf{X}_1 + \cdots + \mathbf{X}_n$ converge in distribution to an α-stable random vector; see Rvačeva (1962). Moreover, the properly normalized and centered componentwise partial maxima of the $\mathbf{X}_i$'s, i.e.,

$$
\Big(\max_{i=1,\ldots,n} X_i^{(1)}, \ldots, \max_{i=1,\ldots,n} X_i^{(d)}\Big)
$$

converge in distribution to a multivariate extreme value distribution whose marginals are of the type Fréchet Φ_α for some $\alpha > 0$ if and only if the vector of the componentwise positive parts of $\mathbf{X}_1$ is regularly varying with index α; see de Haan and Resnick (1977), Resnick (1987). Moreover, Kesten (1973) proves that, under general conditions, a unique strictly stationary solution to the d-dimensional stochastic recurrence equation

$$
\mathbf{Y}_t = \mathbf{A}_t\,\mathbf{Y}_{t-1} + \mathbf{B}_t\,, \quad t \in \mathbb{Z}\,,
$$

exists and satisfies

$$
(2.13) \qquad P((\mathbf{x}, \mathbf{Y}_0) > x) \sim c(\mathbf{x})\,x^{-\kappa_1}\,, \quad x \to \infty\,,
$$

for some positive $c(\mathbf{x}), \kappa_1$, provided $\mathbf{x} \neq \mathbf{0}$ has non-negative components. Here $((\mathbf{A}_t, \mathbf{B}_t))$ is an iid sequence, where $\mathbf{A}_t$ are $d \times d$ matrices and $\mathbf{B}_t$ are d-dimensional vectors, both with non-negative entries. Kesten's result implies in particular, that the one-dimensional marginal distribution of a GARCH process is regularly varying; see the discussion in Section 1. Unfortunately, the definition of multivariate regular variation of a vector $\mathbf{Y}_0$ in the sense of (2.11) or (2.12) and the definition via linear combinations of the components of $\mathbf{Y}_0$ in the sense of (2.13) are in general not known to be equivalent; see Basrak et al. (2000a), Hult (2003).

Well known multivariate regularly varying distributions are the multivariate student, Cauchy and F-distributions as well as the extreme value distributions with Fréchet marginals and the multivariate α-stable distributions with $\alpha \in (0,2)$. In the context of extreme value theory for multivariate data the spherical representation of the limiting measure μ (see (2.12)) has a nice interpretation. Indeed, we see that

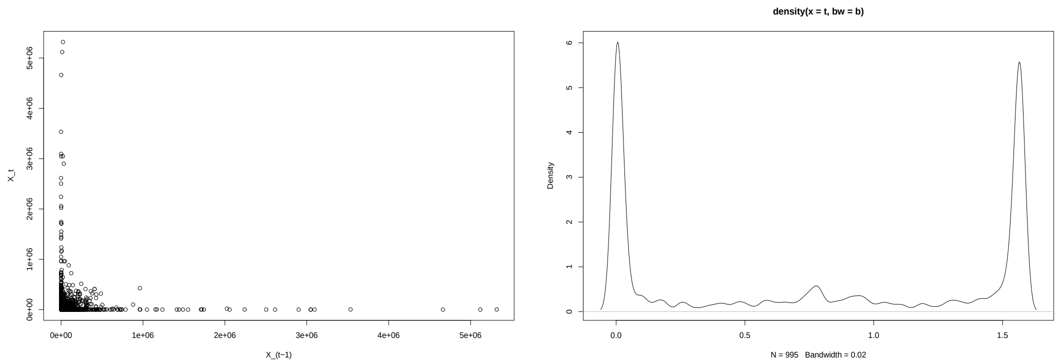

Figure 2.1 *Left: Scatterplot* $\mathbf{X}_t = (X_{t-1}, X_t)$ *of the teletraffic data from Figure 1.3. Right: Estimated spectral density on* $[0, \pi/2]$*. The density is estimated from the values* $\mathbf{X}_t$ *with* $|\mathbf{X}_t| > 80000$*. The density has two clear peaks at the angles* 0 *and* $\pi/2$*, indicating that the components of* $\mathbf{X}_t$ *behave like iid components far away from 0.*

$\mathbf{X}$ is regularly varying with index α and spectral measure $P(\mathbf{\Theta} \in \cdot)$ if and only if for sets $S \subset \mathbb{S}^{d-1}$ with $P(\mathbf{\Theta} \in \partial S) = 0$ and $t > 0$,

$$\lim_{x\to\infty} \frac{P(|\mathbf{X}| > tx)}{P(|\mathbf{X}| > x)} = t^{-\alpha} \quad \text{and} \quad \lim_{x\to\infty} P(x^{-1}\mathbf{X} \in S \mid |\mathbf{X}| > x) = P(\mathbf{\Theta} \in S)\,.$$

This means that the radial and the spherical parts of a regularly varying vector $\mathbf{X}$ become "independent" for values $|\mathbf{X}|$ far away from the origin. For an iid sample of multivariate regularly varying $\mathbf{X}_i$'s the spectral measure tells us about the likelihood of the directions of those $\mathbf{X}_i$'s which are farthest away from zero.

The spectral measure can be estimated from data; we refer to de Haan and Resnick (1993), de Haan and de Ronde (1998) and Einmahl et al. (2001). For two-dimensional vectors $\mathbf{X}_t$ we can write $\mathbf{\Theta} = (\cos(\Phi), \sin(\Phi))$ and estimate the distribution of Φ on $[-\pi, \pi]$. Assuming a density of Φ, one can estimate this *spectral density*; see e.g. Figures 2.1 and 2.2 for some attempts. Figure 2.1 shows the typical shape of a spectral measure when the components of $\mathbf{X}_t$ are independent. Then the spectral measure is concentrated at the intersection of the unit sphere with the axes, i.e., the spectral measure is discrete. This is in contrast to the case of extremal dependence when the spectral measure is concentrated not only at the intersection with the axes. Figure 2.2 gives a typical shape of a spectral density where the components of the vectors $\mathbf{X}_t$ exhibit extremal dependence. The two valleys at the angles $-\pi/4$ and $3\pi/4$ show that X_{t-1} and X_t are not extreme together when they have different signs. However, the peak at $-3\pi/4$ shows that X_{t-1} and X_t are quite likely to be extreme and negative. There is a peak at zero indicating that X_{t-1} can be large and positive, whereas X_t is "less extreme" the next day.

The aim of this section was to explain that the notion of regularly varying distribution is very natural in the context of extreme value theory. As the notion of one-dimensional regular variation, it arises as domain of attraction condition for

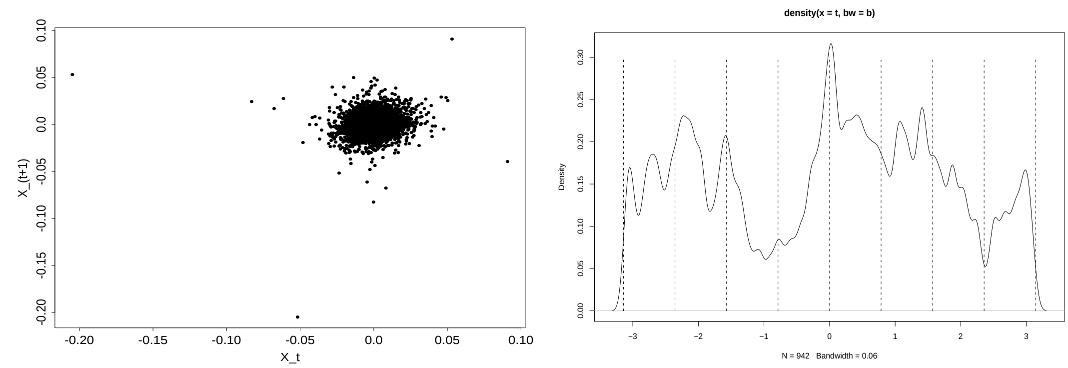

Figure 2.2 *Left: Scatterplot* (X_{t-1}, X_t) *of the S&P500 data from Figure 1.1. Right: Estimated spectral density on* $[-\pi, \pi]$*. The vertical lines indicate multiples of* $\pi/4$*.*

partial maxima and partial sums of iid vectors $\mathbf{X}_t$. There exist, however, enormous statistical problems if one wants to fit regularly varying distributions. Among others, one needs large sample sizes (thousands of data, say) in order to come up with reasonable statistical answers. Successful applications of multivariate extreme value theory have been conducted in dimensions 2 and 3; see e.g. the survey paper by de Haan and de Ronde (1998) for hydrological applications. The limitations of the method are due to the fact that one has to estimate multivariate measures.

In the one-dimensional case, the generalized extreme value distribution (see (1.2)) is the only limit distribution for normalized and centered partial maxima of iid data. This is equivalent to the weak convergence of the excesses to the generalized Pareto distribution. Modern extreme value statistics mostly focuses on fitting the generalized Pareto distribution (GPD) via the excesses (so-called POT – peaks over threshold method; see Embrechts et al. (1997), Chapter 6). Although desirable, in the multivariate case a general result of Balkema-deHaan-Pickands type (see (1.1)) is not available so far, but see Tajvidi (1996) and Balkema and Embrechts (2004) for some approaches.

We mention in passing that the notion of multivariate regular variation also allows for extensions to function spaces. De Haan and Lin (2002) applied regular variation in the space $\mathbb{D}[0, 1]$ of càdlàg functions on $[0, 1]$ in order to describe the extremal behavior of continuous stochastic processes. Hult and Lindskog (2004) extended these results to càdlàg jump functions with the aim of describing the extremal behavior of, among others, Lévy processes with regularly varying Lévy measure.

3 A discussion of some alternative approaches to multivariate extremes

Various other classes of multivariate distributions have been proposed in the context of risk management with the aim of finding a realistic "heavy-tailed" distribution. Those include the *multivariate student distribution*; see e.g. Glasserman (2004), Section 9.3. As mentioned above, these distributions are regularly varying. Their spectral measure is completely determined by their covariance structure; any student distribution has representation in law

$$\mathbf{X} \stackrel{d}{=} \frac{1}{\sqrt{\chi_d^2/d}}\, A\, \widetilde{\mathbf{Z}}\,,$$

where $AA' = \Sigma$ is a covariance matrix Σ (for $d > 2$ the covariance matrix of $\mathbf{X}$ exists and is given by $d\Sigma/(d-2)$), χ_d^2 is χ^2-distributed with d degrees of freedom and $\mathbf{Z}$ consists of iid standard normal random variables. Moreover, $\mathbf{Z}$ and χ_d^2 are independent. This may be attractive as regards the statistical properties of such distributions, but it is questionable whether the extremal dependence structure of financial data is determined by covariances. The student distribution has a rather limited flexibility as regards modeling the directions of the extremes.

A second approach has been suggested by using *copulas*. For simplicity, assume in the sequel that the vector $\mathbf{X} = (X, Y) \stackrel{d}{=} \mathbf{X}_t$ is two-dimensional and its components have continuous distributions F_X and F_Y, respectively. Define for any distribution function G its *quantile function* by

$$G^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : G(x) \geq t\}\,, \quad t \in (0,1)\,.$$

Then

$$P(F_X^{\leftarrow}(X) \leq x\,, F_Y^{\leftarrow}(Y) \leq y) = C(x,y)\,,$$

is a distribution on $(0,1)^2$, referred to as the copula of (X,Y). Of course, $F_X^{\leftarrow}(X)$ and $F_Y^{\leftarrow}(Y)$ are uniformly distributed on $(0,1)$ and the dependence between X and Y sits in the copula function C. Copulas have been used in extreme value statistics for several decades; see e.g. the survey paper by de Haan and de Ronde (1998) or the monograph by Galambos (1987). The purpose of copulas in extreme value statistics is to transform the marginals of the vectors $\mathbf{X}_t$ to distributions with comparable size; otherwise the extremal behavior of an iid sequence $(\mathbf{X}_t)$ would be determined only by the extremes of one dominating component. (Another standard method is to transform the marginals of $\mathbf{X}_t$ to standard Fréchet (Φ_1) marginals.) Thus the transformation of the data to equal marginals makes statistical sense if we cannot be sure that the marginals have comparable tails.

It is however wishful thinking if one believes that copulas help one to simplify the statistical analysis of multivariate extremes. The main problem in multivariate extreme value statistics for data with very large values is the estimation of the spectral measure (which gives one the likelihood of the directions of extremes) and the index of regular variation (which gives one the likelihood of the distance from the origin where extremes occur) or, equivalently, the estimation of the measure μ. It is an illusion to believe that one can estimate these quantities "in a simpler way" by introducing copulas.

We list here various problems which arise by using copulas.

1. *What is a reasonable choice of a copula?* If one accepts that the extreme value theory outlined above makes some sense, one should search for copulas which correspond to multivariate extreme value distributions or multivariate regularly varying distribution. Since copulas stand for any dependence structure between X and Y it is not a priori clear which copulas obey this property. The choice of some ad hoc copula such as the popular arithmetic copula is completely arbitrary.

2. *Should one use extreme value copulas?* So-called *extreme value copulas* have been suggested as possible candidates for copulas for modeling extremes, such as the popular Gumbel copula. These copulas are obtained by transforming the marginals of some very specific parametric multivariate extreme value distribution to the unit cube, imposing some very specific parametric dependence structure on the extremes, very much in the spirit of assuming a multivariate student distribution as discussed above. The fit of an extreme value copula needs to be justified by verifying that the data come from an extreme value generating mechanism. For example, if we consider the annual maximal heights of sea waves at different sites along the Dutch coast, it can be reasonable to fit a multivariate extreme value distribution to these multivariate data. Then the data are extremes themselves. In other cases it is questionable to apply extreme value copulas.

3. *How do we transform the marginals to the unit cube?* Since we do not know $F_X^{\leftarrow}$ or $F_Y^{\leftarrow}$ we would have to take surrogates. The empirical quantile functions are possible candidates. However, the empirical distribution function has bounded support. This means we would not be able to capture extremal behavior outside the range of the sample. Any other approach, for example by fitting GPDs to the marginals and inverting them to the unit cube can go wrong as well, as long as we have not got any theoretical justification for the approach. Moreover, if we are interested in a practical statistical problem we also have to back-transform the marginals from the unit cube to the original problem. There we make another error. In some cases a theoretical justification for such an approach has been given in the context of 2- or 3-dimensional extreme value statistics; see again the survey paper by de Haan and de Ronde (1998). In the latter paper it is also mentioned that the same kind of problems arises if one wants to transform the marginals to other distributions such as the standard Fréchet distribution.

4. *Do copulas overcome the curse of dimensionality?* One can fit any parametric copula to any high-dimensional data set. As explained above, in general, one will fit a distribution which has nothing in common with the extremal

structure of the data. But given the copula is in agreement with the extremal dependence structure of the data, copula fitting faces the same problems as multivariate extreme value statistics which so far can give honest answers for 2- or 3-dimensional problems.

The discussion about which multivariate non-Gaussian distributions are reasonable and mathematically tractable models for extremes is not finished. The aim of this paper was to recall that there exists a probabilistic theory for multivariate extremes that can serve as the basis for honest statistical techniques. Whether one gains by other ad hoc methods is an open question.

References

[1] Balkema, A.A. and Haan, L. de (1974), Residual lifetime at great age, *Annals of Probability* **2**, 792–804.

[2] Balkema, G. and Embrechts, P. (2004), Multivariate excess distributions, ETH Zurich, Department of Mathematic, Technical Report.

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

[4] Basrak B., R.A. Davis and T. Mikosch (2002b), Regular variation of GARCH processes, *Stochastic Processes and their Applications* **99**, 95–116.

[5] Bingham N.H., C.M. Goldie and J.L. Teugels (1987), *Regular variation*, Cambridge University Press.

[6] Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307–327.

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

[8] Embrechts P., C. Klüppelberg and T. Mikosch (1997), *Modelling extremal events for insurance and finance,* Springer, Berlin.

[9] Engle, R.F. (1982), Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation, *Econometrica* **50**, 987–1007.

[10] Feller, W. (1971), *An introduction to probabilit theory and its applications,* vol. II, second edition, Wiley, New York.

[11] Galambos, J. (1987), *Asymptotic theory of extreme order statistics*, 2nd edition, Krieger, Malabar (Florida).

[12] Glasserman, P. (2004), *Monte-Carlo methods in financial enginieering,* Springer, New York.

[13] Goldie, C.M. (1991), Implicit renewal theory and tails of solutions of random equations, *Annals of Applied Probability* **1**, 126–166.

[14] de Haan L. and T. Lin (2002), On convergence toward an extreme value limit in $C[0,1]$, *Annals of Probability* **29**, 467–483.

[15] de Haan L. and S.I. Resnick (1977), Limit theory for multivariate sample extremes, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **40** 317–337.

[16] de Haan L. and S.I. Resnick (1993), Estimating the limiting distribution of multivariate extremes, *Communications in Statistics: Stochastic Models* **9**, 275–309.

[17] de Haan L. and J. de Ronde (1998), Sea and wind: multivariate extremes at work, *Extremes* **1**, 7–45.

[18] Hult, H. (2003), *Topics on fractional Brownian motion and regular variation for stochastic processes,* PhD Thesis, KTH Stockhom, Department of Mathematics.

[19] Hult, H. and F. Lindskog (2004), Extremal behavior for regularly varying stochastic processes, *Stochastic Processes and their Applications* to appear.

[20] Kallenberg, O. (1983), *Random measures,* 3rd edition, Akademieverlag, Berlin.

[21] Kesten, H. (1973), Random difference equations and renewal theory for products of random matrices, *Acta Mathematica* **131**, 207–248.

[22] Mikosch, T. and G. Samorodnitsky (2000), The supremum of a negative drift random walk with dependent heavy-tailed steps, *Annals of Applied Probability* **10**, 1025–1064.

[23] Mikosch T., A. Stegeman, S. Resnick and H. Rootzén (2003), Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Annals of Applied Probability* **12**, 23–68.

[24] Mikosch, T. (2003), Modeling dependence and tails of financial time series, in: B. Finkenstädt and H. Rootén (eds.), *Extreme values in finance, telecommunications, and the environment,* Chapman & Hall, Boca Raton, 185–286

[25] Pareto, V. (1896/97), *Cours d'économie politique professé à l'université de Lausanne.*

[26] Pickands, J. III (1975), Statistical inference using extreme order statistics, *Annals of Statistics* **3**, 119–131.

[27] Resnick, S.I. (1987), *Extreme values, regular variation, and point processes,* Springer, New York.

[28] Rvačeva, E.L. (1962), On domains of attraction of multi-dimensional distributions, *Selected Translations in Mathematical Statistics and Probability Theory*, American Mathematical Society, Providence, R.I. **2**, 183–205.

[29] Sigma (2003), Tables on the major losses 1970-2002, *Sigma publications No 2*, Swiss Re, Zürich, p. 34. The publication is available under www.swissre.com.

[30] Stegeman, A. (2002), *The nature of the beast,* PhD Thesis, University of Groningen, Department of Mathematics and Information Technology.

[31] Tajvidi, N. (1996), *Characterization and some statistical aspects of univariate and multivariate generalised Pareto distributions*, PhD Thesis, Chalmers University Gothenburg, Department of Mathematics.

[32] Willinger W., M.S. Taqqu, M. Leland and D. Wilson (1995), Self-similarity through high variability: statistical analysis of ethernet lan traffic at the source level, *Computer Communications Review* **25**, 100–113. Proceedings of the ACM/SIGCOMM'95, Cambridge, MA.