Introduction to Extreme Value Theory. Applications to Risk Analysis & Management

Marie Kratz

Abstract We present an overview of Univariate Extreme Value Theory (EVT) providing standard and new tools to model the tails of distributions. One of the main issues in the statistical literature of extremes concerns the tail index estimation, which governs the probability of extreme occurrences. This estimation relies heavily on the determination of a threshold above which a Generalized Pareto Distribution (GPD) can be fitted. Approaches to this estimation may be classified into two classes, one qualified as 'supervised', using standard Peak Over Threshold (POT) methods, in which the threshold to estimate the tail is chosen graphically according to the problem, the other class collects unsupervised methods, where the threshold is algorithmically determined.
We introduce here a new and practically relevant method belonging to this second class. It is a self-calibrating method for modeling heavy tailed data, which we developed with N. Debbabi and M. Mboup. Effectiveness of the method is addressed on simulated data, followed by applications in nero-science and finance. Results are compared with those obtained by more standard EVT approaches.
Then we turn to the notion of dependence and the various ways to measure it, in particular in the tails. Through examples, we show that dependence is also a crucial topic in risk analysis and management. Underestimating the dependence among extreme risks can lead to serious consequences, as for instance those we experienced during the last financial crisis. We introduce the notion of copula, which splits the dependence structure from the marginal distribution, and show how to use it in practice. Taking into account the dependence between random variables (risks) allows us to extend univariate EVT to mutivariate EVT. We only give the first steps of the latter, to motivate the reader to follow or to participate in the increasing research development on this topic.

AMS 2000 subject classification. 60G70; 62G32; 62P05; 62G20; 91B30

Marie Kratz, ESSEC Business School, CREAR, Paris, France, (e-mail: kratz@essec.edu)

Contents

1 Introduction

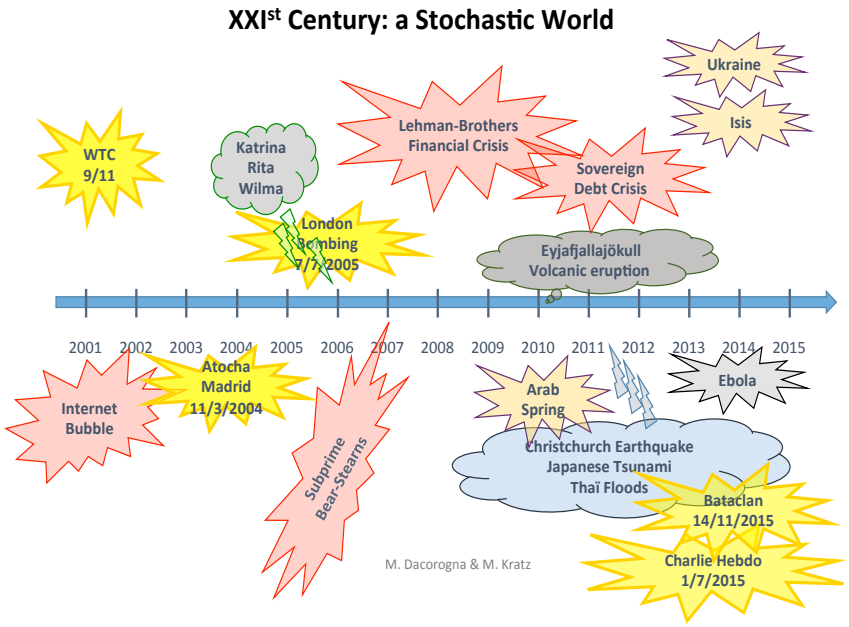


Fig. 1 Some of the extreme events that hit the World between 2001 and 2015

Quantitative risk analysis used to rely, until recently, on classical probabilistic modeling where fluctuations around the average were taken into account. The standard deviation was the usual way to measure risk, like, for instance, in Markowitz portfolio theory[17], or in the Sharpe Ratio[25]. The evaluation of "normal" risks is more comfortable because it can be well modelled and predicted by the Gaussian model and so easily insurable. The series of catastrophes that hit the World at the beginning of this century (see Figure 1, natural (earthquakes, volcano eruption, tsunami, ...) or financial (subprime crisis, sovereign crisis) or political (Arab Spring, ISIS, Ukraine, ...), made it clear that it is crucial nowadays to take also extreme occurrences into account; indeed, although it concerns events that rarely occur (*i.e.* with a very small probability), their magnitude is such that their consequences are dramatic when they hit unprepared societies.

Including extreme risks in probabilistic models is recognized nowadays as a necessary condition for good risk management in any institution, and not restricted anymore to reinsurance companies, who are the providers of covers for natural catastrophes. For instance in finance, minimizing the impact of extreme risks, or even ignoring them because of a small probability of occurrence, has been considered by many professionals and supervisory authorities, as a factor of aggravation of the financial crisis of 2008-09. The American Senate and the Basel Committee

on Banking Supervision confirm this statement in their reports. Therefore, including and evaluating correctly extreme risks has become very topical and crucial for building up the resilience of our societies.

The literature on extremes is very broad; we present here an overview of some standard and new methods in univariate Extreme Value Theory (EVT) and refer the reader to books on the topic [1, 8, 9, 15, 22, 23, 24] and also on EVT with applications in finance or integrated in quantitative risk management [16, 18, 20]. Then we develop the concept of dependence to extend univariate EVT to multivariate EVT. All along applications in various fields including finance, insurance and quantitative risk management, illustrate the various concepts or tools.

1.1 What is Risk?

Risk is a word widely used by many people and not only by professional risk managers. It is therefore useful to spend a bit of time analysing this concept. We start by looking at its definition in common dictionaries. There, we find that it is mainly identified to the notion of danger of loss:

> *The Oxford English Dictionary:* Hazard, a chance of bad consequences, loss or exposure to mischance.
>
> *For financial risks:* "Any event or action that may adversely affect an organization's ability to achieve its objectives and execute its strategies" or, alternatively, "the quantifiable likelihood of loss or less-than-expected returns".
>
> *Webster's College Dictionary (insurance):* "The chance of loss" or "The degree of probability of loss" or "The amount of possible loss to the insuring company" or "A person or thing with reference to the risk involved in providing insurance" or "The type of loss that a policy covers, as fire, storm, etc."

However, strictly speaking, risk is not simply associated to a danger. In its modern acceptance, it can also be seen as an opportunity for a profit. It is the main reason why people would accept to be exposed to risk. In fact, this view started to develop already in the 18th century. For instance, the French philosopher, Etienne de Condillac (1714-1780) defined risk as "The chance of incurring a bad outcome, coupled, with the hope, if we escape it, to achieve a good one".

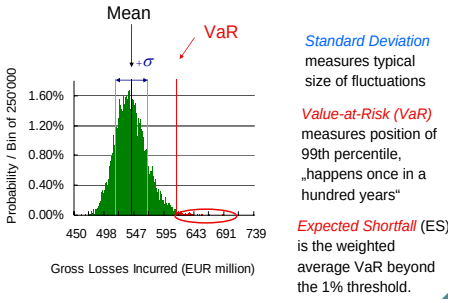
Another concept born from the management of risk is the insurance industry. In the 17th Century, the first insurance for buildings is created after the big London fire (1666). During the 18h century appears the notion that social institutions should protect people against risk. This contributed to the development of life insurance that was not really acceptable by religion at the time. In this context, the definition of 'Insurance' as the transfer of risk from an individual to a group (company) takes its full meaning. The 19th century sees the continuous development of private insurance companies.

Independently of any context, risk relates strongly to the notion of randomness and the uncertainty of future outcomes. The distinction between "uncertainty"' and

"risk" was first introduced by the American economist Frank H. Knight (1885-1972), although they are related concepts. According to Knight, risk can be defined as randomness with knowable probabilities, contrary to uncertainty, which is randomness with unknowable probabilities. We could then define risk as a measurable uncertainty, the 'known-unknowns' according to Donald Rumsfeld's terminology, whereas uncertainty is unmeasurable, the 'unknown-unknowns' (D.Rumsfeld). Of course, research and improved knowledge help to transform some uncertainty into risk with knowable probabilities.

In what follows, we focus on the notion of risk, in particular extreme risk, and its quantification. We choose a possible definition of risk, used within probabilistic framework, namely the variation from the expected outcome over time.

There are numerous ways to measure risk and many risk measures have been developed in the literature. Most modern measures of the risk in a portfolio are statistical quantities describing the conditional or unconditional loss distribution of the portfolio over some predetermined horizon. Here we present only popular ones, used in particular in regulation. We recall their definition and refer to Emmer et al. ([10]) and references therein for their mathematical properties.



Standard Deviation σ: Portfolio theory (Markowitz)

Value-at-Risk (VaR_α) = quantile $q_\alpha(L) = \inf\{q \in \mathbb{R} : \mathbb{P}(L > q) \leq 1-\alpha\}$

Expected Shortfall (ES) (or TVaR)

$$ES_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 q_\beta(L)\,d\beta \underset{F_L \text{ cont}}{=} \mathrm{E}[L \,|\, L \geq q_\alpha(L)]$$

We will focus on the analysis of extreme risks, related to unexpected, abnormal or extreme outcomes.

Many questions arise as: How to model extreme risks? How to study the behavior in the tails of the d.f of the model? How to capture dependency and measure it in risk models? which methods can be used? What about aggregation of risks?

First we present the main concepts of univariate EVT. Then we introduce the issue of dependence among random variables (risks).

1.2 Impact of Extreme Risks

When considering financial assets, because of the existence of a finite variance, a normal approximation is often chosen in practice for the unknown distribution of the yearly log returns, justified by the use of the Central Limit Theorem (CLT), when assuming independent and identically distributed (iid) observations. Such a choice of modeling, in particular using light tail distributions, has shown itself grossly in-

adequate during the last financial crisis when dealing with risk measures because it leads to underestimating the risk.

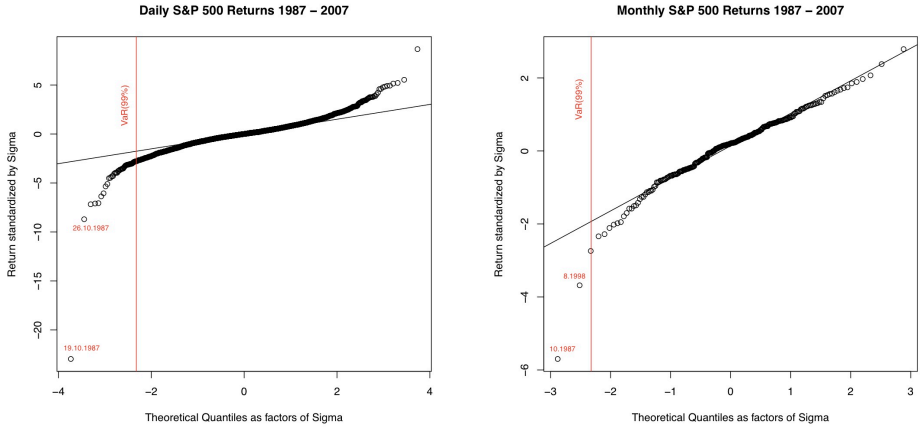


Fig. 2 QQ-plots of the S&P500 daily (left) and monthly (right) log-returns from 1987 to 2007

On Figure 2, the QQ-plot of the S&P 500 daily returns from 1987 to 2007, helps to detect a heavy tail. When aggregating the daily returns into monthly returns, the QQ-plot looks more as a normal one, and the very few observations appearing above the threshold of $\text{VaR}_{99\%}$, among which the financial crises of 1998 and 1987, could almost be considered as outliers, as it is well known that financial returns are symmetrically distributed. Now, look at Figure 3. When adding data from 2008 to

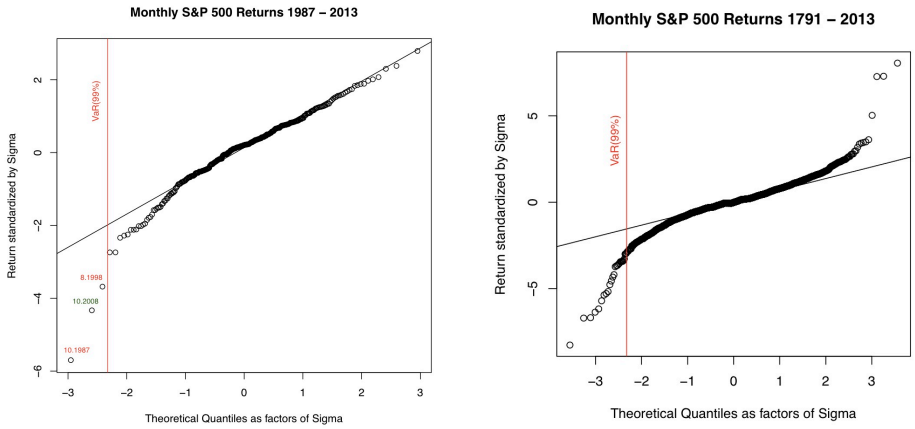


Fig. 3 QQ-plots of the S&P500 monthly log-returns from 1987 (left) or 1791 (right) to 2013

2013, the QQ plot looks pretty the same, *i.e.* normal, except that another "outlier" appears ... with the date of October 2008! Instead of looking again on daily data for the same years, let us consider a larger sample of monthly data from 1791 to 2013 (as compiled by *Global Finance Data*). With a larger sample size, the heavy tail becomes again visible. And now we see that the financial crisis of 2008 does belong to the heavy tail of the distribution and cannot be considered anymore as an outlier. Although it is known, by Feller theorem, that the tail index of the underlying distribution remains constant under aggregation, we clearly see the importance of the sample size to make the tail visible. The figures on the S&P 500 returns illustrate very clearly this issue.

2 Univariate EVT

Let $(X_i)_{i=1,\cdots,n}$ be iid random variables (rv) with parent rv X and continuous cumulative distribution function (cdf) F (that is unknown). The associated order statistics are denoted by $\min_{1\leq i\leq n}(X_i) = X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n-1,n} \leq X_{n,n} = \max_{1\leq i\leq n}(X_i)$.

2.1 CLT versus EVT

▷ *Mean behavior.* Assuming the existence of the variance σ^2 of X, the Central Limit Theorem (CLT) tells us that the empirical mean $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$, when normalized (since $var(\bar{X}_n) = \frac{1}{n} var(X) \underset{n\to\infty}{\to} 0$), has an asymptotic standard Gaussian distribution (whatever is F):

$$\frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\sqrt{var(\bar{X}_n)}} = \sqrt{\frac{n}{var(X)}}\,(\bar{X}_n - \mathbb{E}(X)) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\sigma^2}} \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0,1)$$

$$\text{i.e. } \lim_{n\to\infty} \mathbb{P}[(\bar{X}_n - b_n)/a_n \leq x] = F_{\mathcal{N}(0,1)}(x) \text{ with } b_n = \mathbb{E}(X), a_n = \sqrt{\frac{var(X)}{n}}.$$

▷ *Extreme behavior.* Instead of looking at the mean behavior, consider now the extreme behavior, with for instance the maximum. Noticing that

$$\mathbb{P}[\max_{1\leq i\leq n} X_i \leq x] = \prod_{i=1}^{n} \mathbb{P}[X_i \leq x] = F^n(x) \underset{n\to\infty}{\to} \begin{cases} 0 \text{ if } F(x) < 1 \\ 1 \text{ if } F(x) = 1\,, \end{cases}$$

could we find, as for the CLT, a linear transformation to avoid such degeneracy, and say that there exist sequences (a_n), (b_n) and a rv Z with cdf H such that $\lim_{n\to\infty} \mathbb{P}[(\max X_i - b_n)/a_n \leq x] = H(x)$? It comes back to look for (a_n) and (b_n), and a non-degenerated cdf H s.t.

$$\mathbb{P}\left[\frac{\max X_i - b_n}{a_n} \leq x\right] = \mathbb{P}[\max_{1\leq i\leq n} X_i \leq a_n x + b_n] = F^n(a_n x + b_n) \underset{n\to\infty}{\simeq} H(x).$$

It can be proved that there is not a unique limit distribution as for the CLT, but three possible asymptotic distributions (whatever is F), namely:

Theorem 1 (The 'three-types theorem'; Fréchet-Fisher-Tippett theorem, 1927-28; Genedenko 1948). *The rescaled sample extreme (max renormalized) has a limiting distribution H that can only be of three types:*

$$\begin{array}{ll} H_{1,a}(x) := \exp\{-x^{-a}\}\mathbb{I}_{(x>0)} & (a>0) : \textbf{Fréchet} \\ H_{2,a}(x) := \mathbb{I}_{(x\geq 0)} + \exp\{-(-x)^{a}\}\mathbb{I}_{(x<0)} & (a>0) : \textbf{Weibull} \\ H_{3,0}(x) := \exp\{-e^{-x}\}, & \forall x \in \mathbb{R} : \textbf{Gumbel} \end{array}$$

(A similar result holds for the minimum).

We can then classify the distributions according to the three possible limiting distributions of the (rescaled) maximum, introducing the notion of *Maximum Domain of Attraction* (MDA):

$$F \in \text{MDA}(H) \quad \Leftrightarrow \quad \exists (a_n) > 0, (b_n) : \forall x \in \mathbb{R}, \lim_{n\to\infty} F^n(a_n x + b_n) = H(x).$$

For instance, for Fréchet, $a_n = F^{-1}(1-1/n)$ and $b_n = 0$. (Note that most of the cdf F we use, usually belong to a MDA.)

To mimick the CLT, the three types of extreme value distribution have been combined into a single three-parameter family (Jenkinson-Von Mises, 1955; Hosking et al., 85) known as **Generalized Extreme Value Distribution** (GEV).

Theorem 2 (The EV theorem). *If $F \in MDA(G)$ then, necessarily, G is of the same type as the GEV cdf H_ξ (i.e. $G(x) = H_\xi(ax+b)$, $a>0$), defined by*

$$H_\xi(x) = \begin{cases} \exp\left[-(1+\xi x)_+^{-\frac{1}{\xi}}\right] & \text{if } \xi \neq 0 \\ \exp(-e^{-x}) & \text{if } \xi = 0 \end{cases}$$

where $y_+ = \max(0, y)$. The parameter ξ, named the tail (or extreme-value) index, determines the nature of the tail distribution: if $\xi > 0$ then H_ξ is Fréchet, if $\xi = 0$ then H_ξ is Gumbel, and if $\xi < 0$ then H_ξ is Weibull.

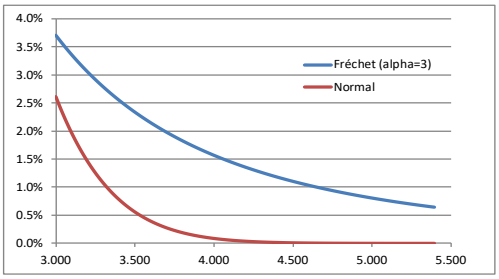
We can write $G(x) = G_{\mu,\sigma,\xi}(x) = \exp\left[-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right]$, for $1+\xi\,\frac{x-\mu}{\sigma} > 0$.

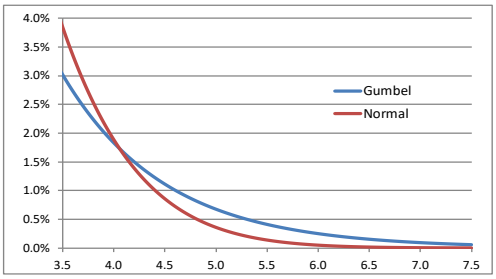
Moments of the GEV: the kth moment exits if $\xi < 1/k$ (in particular if $\mathbb{E}(X) < \infty$ if $\xi < 1$ and $var(X) < \infty$ if $\xi < 1/2$).

Example:

Sample cdf	MDA
Uniform	Weibull
Exponential(1) ($F(x) = 1 - e^{-x}, x > 0$)	Gumbel
Gaussian	Gumbel
Log-normal	Gumbel
Gamma (λ, r)	Gumbel
Cauchy ($F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x$)	Fréchet
Student	Fréchet
Pareto (β) ($F(x) = 1 - x^{-\beta}, x \geq 1, \beta > 0$)	Fréchet

Example of tails of distributions: Fréchet and Gumbel versus Gaussian (normal).





We observe that the tail can vary substantially according to the type of distributions. Here the tail of the Fréchet distributions is moderately heavy ($\alpha = 3$) although it looks much heavier than the Gaussian distribution.

Exercise: Considering the S&P500 daily log returns from 1987 to 2016, compute the chances to find a value smaller than the 2nd minimum, *i.e.* $\mathbb{P}[X < x_{(n-1)}] = \Phi\left(\frac{x_{(n-1)}-\mu}{\sigma}\right)$ (with Φ the standard normal cdf), assuming the data are normally distributed. We obtain the following statistics on S&P500 daily log returns from 1987-2016:

Expected Value (μ)	0.029%	Maximum Value $\left(x_{(1)}\right)$	11.0%
Standard Deviation (σ)	1.192%	Minimum Value $\left(x_{(n)}\right)$	-22.9%
Probability of finding a value smaller than $\left(x_{(n-1)}\right)$ in a Gaussian Model	2.96E-16 (1 over 13.5 billion years)	Second Minimum $\left(x_{(n-1)}\right)$	-9.47%

Characteristic Property of the GEV. A distribution is a GEV if and only if it is *max-stable*, *i.e.* that it satisfies $\max_{1\leq i\leq n} X_i \overset{d}{=} \alpha_n X + \beta_n$, with $\alpha_n > 0$.

For the 3 types of the GEV, we have:

Fréchet: $\max_{1\leq i\leq n} X_i \overset{d}{=} n^{1/\xi} X$; Weibull: $\max_{1\leq i\leq n} X_i \overset{d}{=} n^{-1/\xi} X$; Gumbel: $\max_{1\leq i\leq n} X_i \overset{d}{=} X + \log n$.

2.2 A limit theorem for Extremes: the Pickands theorem

Extracting more information in the tail of the distribution than just that given by the maximum should help for the evaluation of the tail. So considering the kth ($k \geq 1$) largest order statistics, we introduce the notion of 'threshold exceedances' where all data are extreme in the sense that they exceed a high threshold.

Picking up a high threshold $u < x_F^+$ (upper-end point of F), we study all exceedances above u.

Theorem 3 (Pickands Theorem, 1975). *If F does belong to one of the maximum domains of attraction (i.e. the limit distribution of $\max X_i$ is a GEV), then for a sufficiently high threshold u, $\exists\, \beta(u) > 0$ and ξ real number such that the Generalized Pareto Distribution (GPD) $G_{\xi,\beta(u)}$, defined by $\overline{G}_{\xi,\beta(u)}(y) := 1 - G_{\xi,\beta(u)}(y) = \left(1 + \xi \frac{y}{\beta(u)}\right)^{-1/\xi} \mathbb{1}_{(\xi \neq 0)} + e^{-y/\beta(u)} \mathbb{1}_{(\xi=0)}$, is a very good approximation to the excess cdf $F_u(\cdot) := \mathbb{P}[X - u \leq \cdot \,|\, X > u]$*

$$\lim_{u \uparrow x_F^+} \sup_{0 \leq y \leq x_F^+ - u} \left| F_u(y) - G_{\xi,\sigma(u)}(y) \right| = 0,$$

x_F^+ denoting the upper endpoint of F

As for the GEV, we have three cases for the GPD, depending on the sign of the tail index ξ:

- $\xi > 0$: $\overline{G}_{\xi,\beta}(y) \sim c y^{-1/\xi}$, $c > 0$ ("Pareto" tail) : heavy-tail
  (note that $\mathbb{E}(X^k) = \infty$ for $k \geq 1/\xi$).
- $\xi < 0$: $x_G^+ = \beta/|\xi|$ (upper endpoint of G), similar to the Weibull type of the GEV (short-tailed, Pareto type II distribution)
- $\xi = 0$: $\overline{G}_{\xi,\beta}(y) = e^{-y/\beta}$: light-tail (exponential distribution with mean β)

The mean of the GPD is defined for $\xi < 1$ by $\mathbb{E}(X) = \dfrac{\beta}{1-\xi}$.

2.3 Supervised methods in EVT: standard thresholds methods

Univariate Extreme Value Theory (EVT) focuses on the tail distribution evaluation, more precisely on the estimation of the tail index. That is why the first and main question is how to determine the threshold above which observations are considered as extremes. Various methods have been developed to answer this question. We give here their main ideas and refer the reader e.g. to [9] for more details (see also the references therein).

2.3.1 Peak Over Threshold (POT) method

This method developed for the GPD by Davison and Smith ([5]) helps to decide on an appropriate threshold for exceedance-based methods, when looking at the

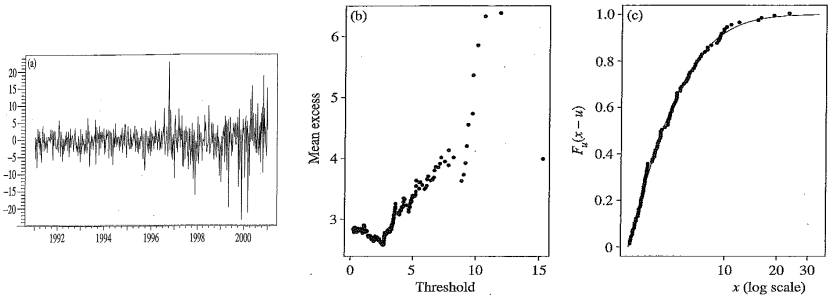
empirical Mean Excess Plot (MEP) . This graphical method can be qualified as supervised.

The mean excess (ME) function defined by $e(u) = \mathbb{E}[X-u|X>u]$ can be computed for any rv X (whenever its expectation exists). For instance, if X is exponentially distributed, then its ME function is a constant. If X is GPD $G_{\xi,\sigma}$, with $\sigma > 0$ and $\xi < 1$, then its ME function is given by $e(u) = \dfrac{\sigma + u\xi}{1-\xi} 1_{(\sigma+u\xi>0)}$.

Hence, via the Pickands theorem, the MEP of X with unknown cdf f should stay reasonably close to a linear function from the threshold u at which the GPD provides a valid approximation to the excess distribution of X: $\mathbb{E}[X-u|\ X > u] \underset{u\to\infty}{\simeq} \sigma(u)/(1-\xi)$. It will be the way to select u, when considering the empirical MEP $\left(v, \dfrac{1}{n_v}\displaystyle\sum_{i=1}^{n_v}(x_{(i)}-v) : v < x_{n,n} \right)$, where the $x_{(i)}$ correspond to the n_v observations that exceed v.

Then, u being chosen, we can use ML or Moments estimators to evaluate the tail index ξ (and the scaling parameter β).

Illustration: Example from Embrechts et al's book ([9])



Data set: time series plot (a) of *AT&T weekly percentage loss data* for the 521 complete weeks in the period 1991-2000

(b) Sample MEP. Selection of the threshold at a loss value of 2.75% (102 exceedances)

(c) Empirical distribution of excesses and fitted GPD, with ML estimators $\hat{\xi} = 0.22$ and $\hat{\beta} = 2.1$ (with Standard Error 0.13 and 0.34, respectively)

2.3.2 Tail index estimators for MDA(Fréchet) distributions

To determine the tail index, other graphical methods than MEP may be used. Various estimators of the tail index have been (and still are) built, starting with the *Hill estimator* (Hill, 1975), a moment estimator (Dekkers et al., 1989), the QQ-estimator (Kratz & Resnick, 1996), ..., the Hill estimator for truncated data (Beirlant et al., 2017), ...

For a sample of size n, the tail index estimators are generally built on the $k = k(n)$ upper order statistics, with $k(n) \to \infty$ such that $k(n)/n \to 0$, as $n \to \infty$.

Choosing k is usually the Achilles heel of all these (graphical) supervised procedures, including the MEP one, as already observed.

Nevertheless it is remarkable to notice that for these methods, no extra information is required on the observations before the threshold (the $n-k$th order statistics).

Let us present two tail index estimators under regular variation framework: the *Hill estimator* ([11], as it is most probably still the most popular, and the *QQ-estimator* ([14]), which is based on a simple and intuitive idea (hence this choice).

Assume $F \in$ MDA(Fréchet) with tail index $\xi > 0$, *i.e.* $\bar{F}$ is regularly varying $RV_{-\alpha}$, with $\xi = \alpha^{-1}$. (Recall that function f belongs to the class RV_ρ of regularly varying functions with index $\rho \in \mathbb{R}$ if $f : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies $\lim_{t\to\infty} f(tx)/f(t) = x^\rho$, for $x > 0$ (see Bingham, Goldie and Teugels (1989).) Consider the threshold $u = X_{n-k,n}$ with $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$.

• The *Hill estimator* $H_{k,n}$ of the tail index $\xi = \alpha^{-1}$ is defined by, and satisfies ([11])

$$H_{k,n} := \frac{1}{k}\sum_{i=0}^{k-1} \log\left(\frac{X_{n-i,n}}{X_{n-k,n}}\right) \xrightarrow[n\to\infty]{P} \xi$$

This estimator is asymptotically normal, with a rate of convergence of $1/\alpha^2$. The Hill estimator can exhibit outrageous bias and graphical aids are often very difficult to interpret accurately. So it is wise to consider alternative methods to supplement information given by the Hill estimator and associated plots. Thus we turn to the QQ-plot.

• The *QQ-estimator* (Kratz & Resnick), $Q_{k,n}$, of the tail index ξ:

The QQ-method is based on the following simple observation: if we suspect that the n-sample X comes from the continuous cdf F, then the plot of

$$\left\{\left(\frac{i}{n+1}, F(X_{i,n})\right), 1 \le i \le n\right\}$$

should be roughly linear, hence also the QQ-plot of $\{(F^{\leftarrow}(\frac{i}{n+1}), X_{i,n}), 1 \le i \le n\}$ (considering the theoretical quantile $F^{\leftarrow}(\frac{i}{n+1})$ and the corresponding quantile $X_{i,n}$ of the empirical distribution function).

If $F = F_{\mu,\sigma}(x) = F_{0,1}(\frac{x-\mu}{\sigma})$, since $F^{\leftarrow}_{\mu,\sigma}(y) = \sigma F^{\leftarrow}_{0,1}(y) + \mu$, the plot of

$$\{(G^{\leftarrow}_{0,1}(\frac{i}{n+1}), X_{i,n}), 1 \le i \le n\}$$

should be approximately a line of slope σ and intercept μ.

Take the example of a n-sample Pareto(α) distributed ($\bar{F}(x) = x^{-\alpha}$); then, for $y > 0$, $F_{0,\alpha}(y) := \mathbb{P}[\log X_1 > y] = e^{-\alpha y}$ and the plot of

$$\left\{(F^{\leftarrow}_{0,1}(\frac{i}{n+1}), \log X_{i,n}), 1 \le i \le n\right\} = \left\{\left(-\log\left(1-\frac{i}{n+1}\right), \log X_{i,n}\right), 1 \le i \le n\right\}$$

should be approximately a line with intercept 0 and slope α^{-1}.

Now, just use the least squares estimator for the slope (SL), namely

$$SL(\{(x_i,y_i),1\le i\le n\})=\frac{\sum_{i=1}^n x_iy_i-\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2-\bar{x}^2}$$

to conclude that, for the Pareto example, an estimator of $\alpha^{-1}(=\xi)$ is

$$\widehat{\alpha^{-1}}=\frac{\sum_{i=1}^n -\log(\frac{i}{n+1})\{n\log X_{n-i+1,n}-\sum_{j=1}^n \log X_{n-j+1,n}\}}{n\sum_{i=1}^n(-\log(\frac{i}{n+1}))^2-(\sum_{i=1}^n -\log(\frac{i}{n+1}))^2},$$

which we call the QQ-estimator.

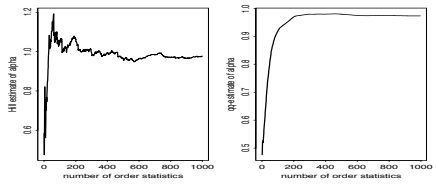
This method can be extended from Pareto to the general case $\bar{F}\sim RV_{-\alpha}$; we can define the QQ-estimator $Q_{k,n}$ of the tail index $\xi=\alpha^{-1}$, based on the upper k order statistics, by

$$\begin{aligned}Q_{k,n}&=SL(\{(-\log(1-\frac{i}{k+1}),\log X_{n-k+i,n}),1\le i\le k\})\\&=\frac{\sum_{i=1}^k -\log\left(\frac{i}{k+1}\right)\left\{k\log\left(X_{n-i+1,n}\right)-\sum_{j=1}^k\log\left(X_{n-j+1,n}\right)\right\}}{k\sum_{i=1}^k\left(-\log\left(\frac{i}{k+1}\right)\right)^2-\left(\sum_{i=1}^k-\log\left(\frac{i}{k+1}\right)\right)^2}\end{aligned}$$

and we can prove ([14]) that the QQ-estimator is weakly consistent ($Q_{k,n}\xrightarrow[n\to\infty]{P}\xi$) and asymptotically normal with a rate of convergence of $1/(2\alpha^2)$ (which is larger than for the Hill, but the Hill estimator exhibits considerable bias in certain circumstances). Whenever the threshold u is determined (corresponding to a kth order statistics), we can estimate the parameters, in particular the tail index.

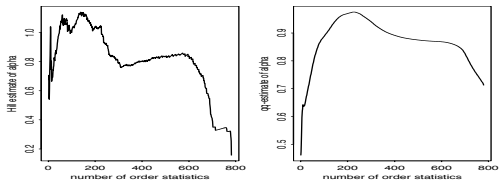
Illustration: Comparison of the Hill plot and the QQ–plot of estimates of α.

- On Pareto(1) simulated data (sample size $n=1000$)



The QQ–plot shows $\widehat{\alpha^{-1}}\simeq 0.98$. It seems a bit less volatile than the Hill plot.

- On real data:



The Hill plot is somewhat inconclusive, whereas the QQ–plot indicates a value of about 0.97

The QQ-method in practice:

1. Make a QQ–plot of all the data (empirical vs theoretical quantile)
2. Choose k based on visual observation of the portion of the graph that looks linear
3. Compute the slope of the line through the chosen upper k order statistics and the corresponding exponential quantiles.

Alternatively, *for Hill and QQ methods:*

1. Plot $\{(k, \widehat{\alpha^{-1}}(k)), 1 \leq k \leq n\}$
2. Look for a stable region of the graph as representing the true value of α^{-1}.

Using those graphical (supervised) methods to determine u or, equivalently k, is an art as well as a science and the estimate of α is usually rather sensitive to the choice of k (but this is the price to pay for having a method to fit the tail without using any information before u).

Let us turn to another method, which answers this concern and provides an automatic (algorithmic) determination of the threshold u (but requiring, in this case, the data information before u).

2.4 A self-calibrated method for heavy-tailed data

It is based on a paper developed with N. Debbabi and M. Mboup ([6]).

We assume continuous (smooth transitions) and, with no loss of generality, right heavy tailed data (a similar treatment being possible on the left tail) belonging to the MDA(Fréchet).

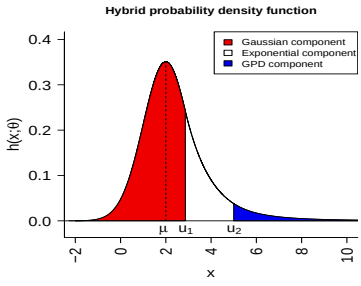
Whereas one of the motivations for this new method is to be able to determine the threshold above which we fit the GPD in an unsupervised way, it will also provide a good fit for the entire distribution.

We introduce a hybrid model to fit the whole distribution underlying heavy tailed data. The idea is to consider both the mean and tail behaviors, and to use limit theorems for each one (as suggested and developed analytically in [12]), in order to make the model as general as possible. Therefore, we introduce a Gaussian distribution for the mean behavior, justified by the Central Limit Theorem (CLT), and a GPD for the tail (justified by the Pickands theorem). Then we bridge the gap between mean and asymptotic behaviors by inserting an exponential distribution used as a leverage to give full meaning of tail threshold to the junction point between the GPD and its exponential neighbour.

We assume that the hybrid model distribution (which belongs to the Fréchet MDA) has a density that is C^1. It is the only assumption that is needed (no assumption on the dependence of the data). This model, denoted by G-E-GPD (Gaussian-Exponential-Generalized Pareto Distribution), is characterized by its pdf h expressed as:

$$h(x;\theta)=\begin{cases}\gamma_1\, f(x;\mu,\sigma), & \text{if } x\le u_1,\\ \gamma_2\, e(x;\lambda), & \text{if } u_1\le x\le u_2,\\ \gamma_3\, g(x-u_2;\xi,\beta), & \text{if } x\ge u_2,\end{cases}$$

where f is the Gaussian pdf (μ,σ^2), e is the exponential pdf with intensity λ, g is the GPD pdf with tail index ξ and scaling parameter β, and γ_1, γ_2, γ_3 are the weights (evaluated from the assumption).



Combining the facts that we are in the MDA(Fréchet) and that h is a C^1 pdf gives rise to six equations relating all model parameters:

$$\begin{cases}\beta=\xi\,u_2;\quad \lambda=\frac{1+\xi}{\beta};\ u_1=\mu+\lambda\sigma^2;\\ \gamma_1=\gamma_2\,\dfrac{e(u_1;\lambda)}{f(u_1;\mu,\sigma)};\quad \gamma_2=\left[\xi\,e^{-\lambda u_2}+\left(1+\lambda\,\dfrac{F(u_1;\mu,\sigma)}{f(u_1;\mu,\sigma)}\right)e^{-\lambda u_1}\right]^{-1};\quad \gamma_3=\beta\,\gamma_2\,e(u_2;\lambda).\end{cases}$$

Consequently, the vector of the free parameters is reduced to $\theta=[\mu,\sigma,u_2,\xi]$.

Remark. The main component in this hybrid model is the GPD one (for heavy tail), the mean behavior having to be adapted to the context. For instance, for *insurance claims*, we have replaced the Gaussian component with a *Lognormal* one (lognormal-E-GPD hybrid model).

Pseudo-code of the algorithm for the G-E-GPD parameters estimation

Here we describe the iterative algorithm, which self-calibrates the G-E-GPD model, in particular the tail threshold above which a Fréchet distribution fits the extremes. We study its convergence, proving analytically the existence of a stationary point, then numerically that the stationary point is attractive and unique.

1: Initialization of $\widetilde{p}^{(0)}=[\widetilde{\mu}^{(0)},\widetilde{\sigma}^{(0)},\widetilde{u}_2^{(0)}]$, α, $\varepsilon>0$, and k_{max}, then initialization of $\widetilde{\xi}^{(0)}$ (recall that $\theta=[\mu,\sigma,u_2,\xi]$):

$$\widetilde{\xi}^{(0)}\leftarrow \underset{\xi>0}{argmin}\left\|H(y;\theta\mid\widetilde{p}^{(0)})-H_n(y)\right\|_2^2,$$

where H_n is the empirical cdf of X (and distance computed on $y=(y_j)_{1\le j\le m}$) and $y=(y_j)_{1\le j\le m}$ is a generated sequence of synthetic increasing data of size m (that may be different from n), with a logarithmic step, in order to increase the number of points above the tail threshold u_2: $y_j=\min_{1\le i\le n}(x_i)+(\max_{1\le i\le n}(x_i)-\min_{1\le i\le n}(x_i))\log_{10}\left(1+\dfrac{9(j-1)}{m-1}\right)$.

2: Iterative process:

- $k \leftarrow 1$

  Step 1 - Estimation of $\widetilde{p}^{(k)}$: $\widetilde{p}^{(k)} \leftarrow \underset{\substack{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_+^* \\ u_2\in\mathbb{R}_+}}{argmin} \left\| H(y;\theta \mid \widetilde{\xi}^{(k-1)}) - H_n(y) \right\|_2^2$

  Step 2 - Estimation of $\widetilde{\xi}^{(k)}$: $\widetilde{\xi}^{(k)} \leftarrow \underset{\xi>0}{argmin} \left\| H(y;\theta \mid \widetilde{p}^{(k)}) - H_n(y) \right\|_2^2$.

- $k \leftarrow k+1$
  until $\Big(d(H(y;\theta^{(k)}), H_n(y)) < \varepsilon \text{ and } d(H(y_{q_\alpha};\theta^{(k)}), H_n(y_{q_\alpha})) < \varepsilon \Big)$ or $\big(k = k_{max}\big)$.

  where ε is a positive real that is small enough, y_{q_α} represents the observations above a fixed high quantile q_α of arbitrary order $\alpha \geq 80\%$ associated with H and $d(a,b)$ denotes the distance between a and b, chosen in this study as the Mean Squared Error (MSE); it can be interpreted as the Cramér-von-Mises test of goodness of fit.

3: Return $\theta^{(k)} = \big[\widetilde{\mu}^{(k)}, \widetilde{\sigma}^{(k)}, \widetilde{u}_2^{(k)}, \widetilde{\xi}^{(k)}\big]$.

Performance of the method (algorithm) tested via MC simulations

To study the performance of the algorithm to self-calibrate the G-E-GPD model, we build on MC simulations. To do so, we proceed in 4 steps:

1. Consider
   $\{X^q = (X_p^q)_{1\leq p\leq n}\}_{1\leq q\leq N}$: training sets of length n
   $\{Y^q = (Y_p^q)_{1\leq p\leq l}\}_{1\leq q\leq N}$: test sets of length l
   with a G-E-GPD parent distribution with a fixed parameters vector θ.
2. On each training set X^q, $1 \leq q \leq N$, evaluate $\widetilde{\theta}^q = [\widetilde{\mu}^q, \widetilde{\sigma}^q, \widetilde{u_2}^q, \widetilde{\xi}^q]$ using our algorithm
3. Compute the empirical mean $\tilde{a}$ and variance $\tilde{S}_a$ of estimates of each parameter a over the N training sets. To evaluate the performance of the estimator $\tilde{a}$, we use two criteria:

   (i) MSE expressed for any a as: $\text{MSE}_a = \frac{1}{N}\sum_{q=1}^N (\widetilde{a}^q - a)^2$; a small value of MSE highlights the reliability of parameters estimation using the algorithm.

   (ii)Test on the mean (with unknown variance): $\left| \begin{array}{l} H0 : \widetilde{a} = a \\ H1 : \widetilde{a} \neq a \end{array} \right.$
   (use for instance the normal test for a large sample)

4. Compare the hybrid pdf h (with the fixed θ) with the corresponding estimated one $\widetilde{h}$, using $\widetilde{\theta}^q$ on each test set Y^q. To do so, compute the average of the log-likelihood function $\mathscr{D}$, over N simulations, between $h(Y^q;\widetilde{\theta}^q)$ and $h(Y^q;\theta)$: $\mathscr{D} = \frac{1}{Nl}\sum_{q=1}^N\sum_{p=1}^l \log\Big(h(Y_p^q;\theta)/\widetilde{h}(Y_p^q;\widetilde{\theta}^q)\Big)$. The smallest the value of $\mathscr{D}$ is, the most trustworthy is the algorithm.

Several MC simulations have been performed varying θ and n, to test the robustness of the algorithm (see [6], §4 and Appendix B).

Application in Neuroscience - Neural data

We consider the data corresponding to twenty seconds, equivalent to $n = 3.10^5$ observations, of real extracellular recording of neurons activities. The information to be extracted from these data (spikes or action potentials) lies on the extreme behaviors (left and right) of the data.

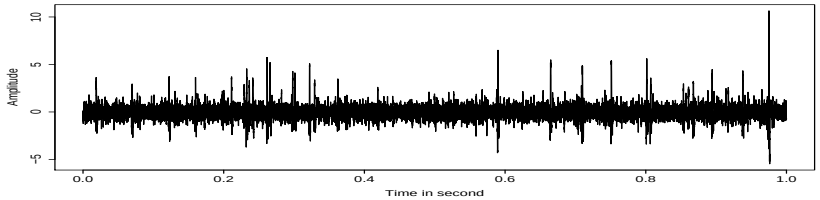


Fig. 4 One second of neural data, extracellularly recorded.

Since the neural data can be considered as symmetric, it is sufficient to evaluate the right side of the distribution with respect to its mode.

In Table 1, we present the results obtained with the self-calibrating method, the MEP, Hill and QQ methods. Since the three graphical approaches fit only the tail distribution, the comparison of the methods will focus on the goodness-of-fit of the GPD component. As observed in this table, the MSE between the estimated cdf and the empirical one, using only data above the selected threshold, is small enough for the four methods ensuring a reliable modeling of extremes. The GPD threshold and the estimated tail index are of the same order of magnitude for all methods; it confirms that our algorithm works in the right direction.

Table 1 Comparison between the self-calibrating method and the three graphical methods: MEP, Hill and QQ ones. N_{u_2} represents the number of observations above u_2. The distance gives the MSE between the empirical (tail or full respectively) distribution and the estimated one from a given model (GPD or hybrid G-E-GPD respectively). The neural data sample size is $n = 3 \times 10^5$.

Model	tail index (ξ)	threshold (u_2)	N_{u_2}	distance (tail distr.)	distance (full distr.)
GPD	MEP (PWM): 0.3326	$1.0855 = q_{93.64\%}$	19260	$3.26\ 10^{-6}$	
GPD	Hill-estimator: 0.599	$1.0855 = q_{93.64\%}$	19260	$2.07\ 10^{-6}$	
GPD	QQ-estimator: 0.5104	$1.0671 = q_{93.47\%}$	19871	$1.26\ 10^{-5}$	
G-E-GPD	Self-calibrating method: 0.5398	$1.0301 = q_{92.9\%}$	21272	$7.79\ 10^{-6}$	$9.31\ 10^{-5}$

We can also notice the good performance of these methods through Figure 5, where we plot the empirical quantile function and the estimated ones using the self-calibrating method and the various graphical ones. However, the advantage of our method is that it is unsupervised, *i.e.* it does not need the intervention of the user to select the threshold manually. Moreover it provides a good fit between the hybrid cdf estimated on the entire data sample (the right side for this data set) and the empirical cdf, with a MSE of order 10^{-5}.

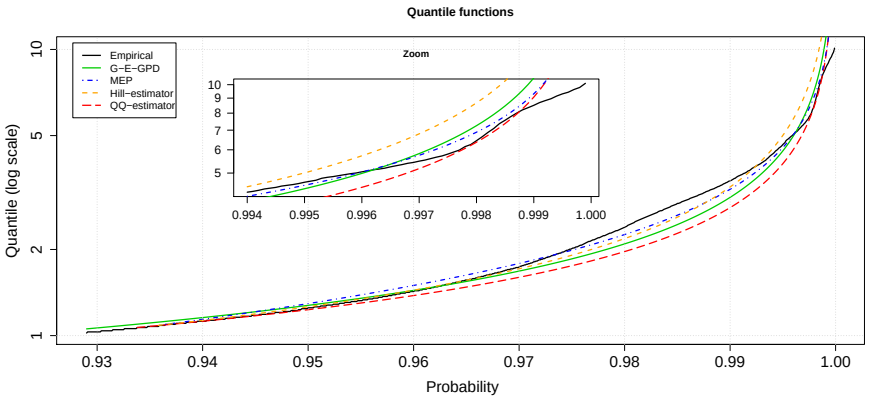


Fig. 5 Neural data: Comparison between the empirical quantile function and the estimated ones via the self-calibrating method and the graphical methods.

Application in Finance - S&P 500 data

Consider the S&P500 log-returns from January 2, 1987 to February 29, 2016, corresponding to $n = 7348$ observations, available in the tseries package of the R programming language. It is well known that log-returns of financial stock indices exhibit left and right heavy tails, with a slight different tail index from one to the other. It is important in such context to evaluate the nature of tail(s) in order to compute the capital needed by a financial institution to cover their risk, often expressed as a Value-at-Risk (*i.e.* a quantile) of high order.

The S&P500 log-returns being essentially symmetric around zero (representing the data mode), we kept the Gaussian component to model the mean behavior when applying the self-calibrating method. We modelled the negative log returns and the positive ones, respectively, then the full data set. When focusing on tails, we also compare our results with those obtained with MEP, Hill, and QQ methods. We present them in Tables 2 and 3. We observe that all methods offer a good fit of the tail distribution. However, the advantage of the self-calibrating method is that it does not need the intervention of the user to select the threshold manually, which is a considerable advantage in practice.

Now, to underline the good performance of the self-calibrating method even in the case when data are autocorrelated with a long memory, we apply it on the S&P500 absolute log-returns. Indeed, it is well known that the absolute value of financial returns are autocorrelated (see Figure 6), but also that their extremes are not (for a thorough discussion of this point and empirical evidences, see Hauksson et al. (2001)). In time of crisis, as e.g. in 2008-09, we observe an increase of the dependence between various financial indices, in particular in the extremes. This is to be distinguished from a dependence of the extremes within a univariate financial index, which is not observed (Hauksson et al. (2001)).

Table 2 Lower tail modeling of the S&P500 log-returns. Comparison between the self-calibrating method and the three graphical methods: MEP, Hill and QQ ones, applied on the right side of the S&P500 opposite log-returns ($-X$). N_{u_2} represents the number of observations above the tail threshold u_2. The distance gives the MSE between the empirical tail (from u_2), or positive side (for $x \geq 0$) respectively, distribution and the estimated one from a given model (GPD, or hybrid G-E-GPD respectively).

Model	tail index (ξ)	threshold (u_2)	N_{u_2}	distance (tail distr.)	distance (positive distr.)
GPD	MEP: 0.3640	$0.0270 = q_{98.36\%}$	120	$1.19\ 10^{-7}$	
GPD	Hill-estimator: 0.3601	$0.0301 = q_{98.84\%}$	86	$6.43\ 10^{-8}$	
GPD	QQ-estimator: 0.3813	$0.0313 = q_{99.00\%}$	74	$3.54\ 10^{-8}$	
G-E-GPD	Self-calibrating method: 0.3545	$0.0289 = q_{98.63\%}$	100	$2.64\ 10^{-7}$	$3.11\ 10^{-6}$

Table 3 Upper tail modeling of the S&P500 log-returns. Comparison between the self-calibrating method and the graphical methods applied on the right side of the S&P500 log-returns.

Model	tail index (ξ)	threshold (u_2)	N_{u_2}	distance (tail distr.)	distance (positive distr.)
GPD	MEP: 0.2715	$0.0209 = q_{96.89\%}$	229	$4.91\ 10^{-7}$	
GPD	Hill-estimator: 0.3225	$0.0288 = q_{98.84\%}$	86	$4.42\ 10^{-7}$	
GPD	QQ-estimator: 0.2859	$0.0321 = q_{99.03\%}$	71	$5.06\ 10^{-8}$	
G-E-GPD	Self-calibrating method: 0.3360	$0.0266 = q_{98.51\%}$	109	$3.89\ 10^{-7}$	$2.49\ 10^{-6}$

A comparison of the results obtained with our self-calibrating method and the graphical EVT ones is depicted in Table 4. In Figure 7, we also give a comparison of the estimated quantile function using the G-E-GPD method and the graphical (MEP,

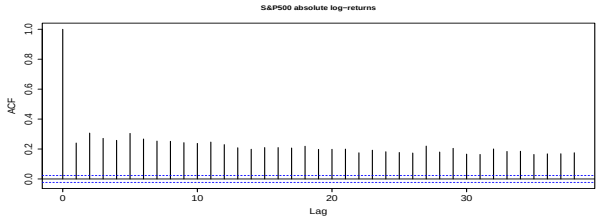


Fig. 6 AutoCorrelation Function (ACF) of the S&P500 absolute log-returns.

Hill and QQ) ones. Through Table 4 and Figure 7, we can highlight once again the good performance of the self-calibrating method to estimate the tail distribution as well as the entire distribution of autocorrelated data. Note that the estimate of the tail index is of the same order as those of the upper and lower tail indices evaluated in the previous section.

Table 4 Comparison between the self-calibrating method and the three graphical methods: MEP, Hill and QQ ones. The S&P500 absolute log-returns data sample size is $n = 7348$.

Model	tail index (ξ)	threshold (u_2)	N_{u_2}	distance (tail distr.)	distance (full distr.)
GPD	MEP: 0.3025	$0.0282 = q_{97.21\%}$	206	$1.78\ 10^{-7}$	
GPD	Hill-estimator: 0.3094	$0.0382 = q_{98.85\%}$	85	$4.49\ 10^{-8}$	
GPD	QQ-estimator: 0.3288	$0.0323 = q_{98.14\%}$	137	$6.01\ 10^{-8}$	
G-E-GPD	Self-calibrating method: 0.3331	$0.0290 = q_{97.49\%}$	184	$2.00\ 10^{-7}$	$1.05\ 10^{-5}$

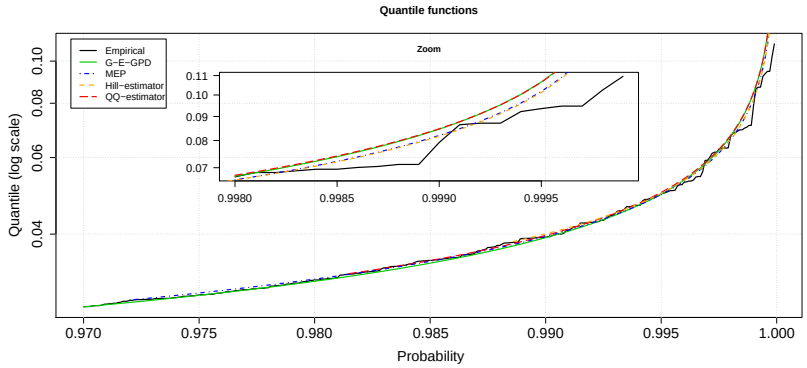


Fig. 7 S&P 500 absolute log-returns data: Comparison between the empirical quantile function and the estimated ones via the self-calibrating method and the graphical methods.

3 Dependence

3.1 Motivation

3.1.1 Impact of the dependence on the Diversification Benefit

The diversification performance is at the heart of the strategy of a company. It can be measured, for a portfolio of N risks, via the diversification benefit $D_{n,\alpha}$ at a threshold α ($0 < \alpha < 1$) defined by $D_{n,\alpha} = 1 - \dfrac{\rho_\alpha(\sum_{i=1}^n L_i)}{\sum_{i=1}^n \rho_\alpha(L_i)}$, where ρ denotes a risk measure. This indicator, not universal as it depends on the number of the risks undertaken and on the chosen risk measure ρ, helps to determine the optimal portfolio of the company since the diversification reduces the risk and thus enhances the performance. This is key to both insurances and financial institutions.

Before developing an example in the insurance context (it would be the same for investment banks) to point out the impact of the dependence on the diversification benefit, let us recall some standard notions in insurance.

Insurance framework.

In insurance, the risk is priced based on the knowledge of the loss probability distribution. The occurrence of a loss L being random, we define it as a random variable (rv) on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ (note that, in insurance context, we often use *risk* and *loss* for one another).

The role of capital for an insurance company is to ensure that the company can pay its liability even in the worst cases, up to some threshold.

It means to define the capital to put behind the risk. That is why we introduce a risk measure (say ρ), defined on the loss distribution, in order to estimate the capital needed to ensure payment of the claim up to a certain confidence level.

Now let us define some useful quantities, as:

Risk-adjusted-capital. The risk can be defined as the deviation from the expectation, hence the notion of Risk-Adjusted-Capital (RAC), say K, which is a function of the risk measure ρ associated to the risk L, defined by $K = \rho(L) - \mathbb{E}[L]$.

Risk Loading. An insurance is a company in which shareholders can invest. They expect a return on investment. So the insurance firm has to make sure that the investors receive their dividends. It corresponds to the cost of capital, η, that the insurance company must charge on its premium. Consider a portfolio of N similar policies. The risk loading per policy, say R, is defined as $R = \eta \dfrac{K_N}{N} = \eta \left(\dfrac{\rho(L^{(N)})}{N} - \mathbb{E}[L_1] \right)$, where K_N is the capital assigned to the entire portfolio, $L^{(N)} = \sum_{i=1}^N L_i$ is the total loss of the portfolio, and $L_1 = L$ is loss incurred by 1 policy (of the portfolio).

Technical risk premium. For one policy case, incurring a loss L, the technical premium, P, that needs to be paid can be defined by $\mathbb{P} = \mathbb{E}(L) + \eta K + e$, where η is the return expected by shareholders before tax, K is the RAC (*i.e.* the capital

assigned to this risk), ηK corresponds to the Risk loading (per policy), and e are the expenses incurred by the insurer to handle this case.
Assuming that the expenses are a small portion of the expected loss, *i.e.* $e = a\mathbb{E}[L]$ with $0 < a \ll 1$, then the premium can be written as $P = (1+a)\mathbb{E}[L] + R$.
Generalizing to a portfolio of N similar (iid) policies, the total loss is $L^{(N)} = \sum_{i=1}^{N} L_i$, hence the premium for one policy in the portfolio becomes:

$$P = \frac{(1+a)\mathbb{E}[L^{(N)}] + \eta K_N}{N} = (1+a)\,\mathbb{E}[L] + \eta\,\frac{K_N}{N},$$

where $\eta \frac{K_N}{N}$ is the risk loading per policy.

Let us then develop our toy model (see [2, 3]) to show the dependence impact on the diversification benefit.

Suppose an insurance company has underwritten N policies of a given risk. To price these policies, the company must know the underlying probability distribution of this risk. Assume that each policy is exposed n times to this risk, thus in a portfolio of N policies, the risk may occur $n \times N$ times.

Let us introduce a sequence $(X_i, i = 1, \ldots, Nn)$ of rv's X_i to model the occurrence of the risk, with a given severity l (for simplicity, take it deterministic). Hence the total loss amount, say L, associated to this portfolio is given by $L = l\sum_{i=1}^{Nn} X_i := l\,S_{Nn}$.

We are going to consider three models for the occurrence of the risk, depending on the dependence structure.

(a) *A first simple model, under the iid assumption*

Assume the X_i's are iid (independent, identically distributed) with parent rv denoted by X, Bernoulli distributed $\mathscr{B}(p)$, *i.e.* the loss $L_1 = lX$ occurs with some probability p:

$$X = \begin{cases} 1 \text{ with probability} & p \\ 0 \text{ with probability} & 1-p \end{cases}$$

Hence the total loss amount $L = l\,S_{Nn}$ follows a binomial distribution $\mathscr{B}(Nn, p)$. We can then deduce the risk loading for an increasing number N of policies in the portfolio: $R = \eta\,\left(\frac{\rho(L)}{N} - lnp\right)$, in order to determinate the risk premium the insurance will ask to a customer if he buys this insurance policy. The relative risk loading per policy is then $\frac{R}{\mathbb{E}[L^{(1)}]} = \eta\,\left(\frac{\rho(L)}{lnp} - 1\right)$.

Numerical application. We choose for instance the number of times one policy is exposed to the risk as $n = 6$ and the unit loss l is fixed to $l = 10$ Euros.

Computing the loss distribution, we obtain the results presented in Table 5. We observe that the probability that the company will turn out paying more than the expectation $\mathbb{E}(L) = lnp = 10$, is of more than 26%. It makes then clear why the technical premium cannot be reduced to $\mathbb{E}(L)$.

Table 5 Distribution of the loss $L = lS_{1n}$ for one policy ($N = 1$) with $n = 6$ and $p = 1/6$

number of losses k	Policy Loss $l\,X(\omega)$	Probability Mass $\mathbb{P}[S_{1n} = k]$	Cdf $\mathbb{P}[S_{1n} \leq k]$
0	0	33.490%	33.490%
1	10	40.188%	73.678%
2	20	20.094%	93.771%
3	30	5.358%	99.130%
4	40	0.804%	99.934%
5	50	0.064%	99.998%
6	60	0.002%	100.000%

Now we compute the risk loading per policy as a function of the number N of policies in the portfolio for both risk measures VaR and TVaR, and when taking $p = 1/6$ (fair game), $1/4$ and $1/2$, respectively. We assume that the cost (of capital) is $\eta = 15\%$, and that the risk measure is computed at threshold $\alpha = 99\%$. Results are given in Table 6.

Table 6 The Risk loading per policy as a function of the number N of policies in the portfolio (with $n = 6$)

Risk measure ρ	**Number N of Policies**	**Risk Loading R per policy with probability** $p = 1/6$	$p = 1/4$	$p = 1/2$
VaR				
	1	3.000	3.750	4.500
	5	1.500	1.650	1.800
	10	1.050	1.200	1.350
	50	0.450	0.540	0.600
	100	0.330	0.375	0.420
	1'000	0.102	0.117	0.135
	10'000	0.032	0.037	0.043
TVaR				
	1	3.226	3.945	4.500
	5	1.644	1.817	1.963
	10	1.164	1.330	1.482
	50	0.510	0.707	0.675
	100	0.372	0.425	0.476
	1'000	0.116	0.134	0.154
	10'000	0.037	0.042	0.049
$\mathbb{E}[\mathbf{L}]/\mathbf{N}$		*10.00*	*15.00*	*30.00*

We observe that, in the case of independent risks, the risk loading R is a decreasing function of the number of policies N, even with a biased dice. With 10'000 policies, R is divided by 100, whatever is the choice of the risk measure ρ, with

slightly higher values for TVaR than VaR.

(b) *Introducing a structure of dependence to reveal a systematic risk*

We introduce two types of structure of dependence between the risks, in order to explore the occurrence of a systematic risk and, as a consequence, the limits to diversification.

We still consider the sequence $(X_i, i = 1, \ldots, Nn)$ to model the occurrence of the risk, with a given severity l, for N policies, but do not assume anymore that the X_i's are independent (but identically distributed, for sake of simplicity). We assume that the occurrence of the risks X_i's depends on another phenomenon, represented by a random variable (rv), say U. Depending on the intensity of the phenomenon, *i.e.* the values taken by U, a risk X_i has more or less chances to occur.

Suppose that the dependence between the risks is totally captured by U that is identified to the occurrence of a state of systematic risk. Consider, w.l.o.g., that U can take two possible values denoted by 1 and 0: $U \overset{d}{\sim} \mathscr{B}(\tilde{p})$, $0 < \tilde{p} << 1$, where $\tilde{p}$ is chosen very small since we want to explore rare events. We present two examples of models (i.e. two types of dependence).

(i) *A dependent model, but conditionally independent*

The occurrence of the risks $(X_i)_i$ is modeled by a Bernoulli rv whose parameter is chosen depending on U and such that the conditional rv's $X_i \mid U$ are independent. Since U takes 2 possible values, the same holds for the parameter of the Bernoulli distribution of the conditionally independent rv's $X_i \mid U$, namely

$$X_i \mid (U = 1) \overset{d}{\sim} \mathscr{B}(q) \quad \text{and} \quad X_i \mid (U = 0) \overset{d}{\sim} \mathscr{B}(p)$$

where we choose $q >> p$, so that whenever U occurs (*i.e.* $U = 1$ (crisis state)), it has a big impact in the sense that there is a higher chance of loss. We include this effect in order to have a systematic risk (non-diversifiable) in our portfolio. Hence the mass probability distribution f_S of the total amount of losses S_{Nn} appears as a mixture of two mass probability distributions $f_{\tilde{S}_q}$ and $f_{\tilde{S}_p}$ of conditional independent rv's $\tilde{S}_q := S_{Nn} \mid (U = 1) \overset{d}{\sim} \mathscr{B}(Nn, q)$ and $\tilde{S}_p := S_{Nn} \mid (U = 0) \overset{d}{\sim} \mathscr{B}(Nn, p)$, respectively:

$$f_S = \tilde{p}\, f_{\tilde{S}_q} + (1 - \tilde{p})\, f_{\tilde{S}_p}.$$

Note that $\tilde{p} = 0$ gives back the normal state.

Numerical application. As for example (a), we take $n = 6$ and $p = 1/n$. Moreover we choose the loss probability during the crisis to be $q = 1/2$, and explore different probabilities $\tilde{p}$ of occurrence of a crisis. In Table 7, the results illustrate well the effect of the non-diversifiable risk. When the probability of occurrence of a crisis is high, the diversification does not play a significant role anymore already with 100 contracts in the portfolio. For $\tilde{p} \geq 1\%$, the risk loading barely changes when there is a large number of policies (starting at $N = 1000$) in the portfolio, for both VaR and TVaR. The non-diversifiable term dominates the risk. For lower probability $\tilde{p}$ of occurrence of a crisis, the choice of the risk measure

Table 7 The Risk loading per policy as a function of the probability of occurrence of a systematic risk in the portfolio using VaR and TVaR measures with $\alpha = 99\%$. The probability of giving a loss in a state of systematic risk is chosen to be $q = 50\%$.

Risk measure ρ	**Number** N **of Policies**	**Risk Loading** R **in a normal state**	**with occurrence of a crisis state**			
		$\tilde{p} = 0$	$\tilde{p} = 0.1\%$	$\tilde{p} = 1.0\%$	$\tilde{p} = 5.0\%$	$\tilde{p} = 10.0\%$
VaR						
	1	3.000	2.997	4.469	4.346	5.693
	5	1.500	1.497	2.070	3.450	3.900
	10	1.050	1.047	1.770	3.300	3.450
	50	0.450	0.477	1.410	3.060	3.030
	100	0.330	0.327	1.605	3.000	2.940
	1'000	0.102	0.101	2.549	2.900	2.775
	10'000	0.032	0.029	2.837	2.866	2.724
TVaR						
	1	3.226	3.232	4.711	4.755	5.899
	5	1.644	1.707	2.956	3.823	4.146
	10	1.164	1.266	2.973	3.578	3.665
	50	0.510	0.760	2.970	3.196	3.141
	100	0.372	0.596	2.970	3.098	3.020
	1'000	0.116	0.396	2.970	2.931	2.802
	10'000	0.037	0.323	2.970	2.876	2.732
E[L]/N		*10.00*	*10.02*	*10.20*	*11.00*	*12.00*

matters. For instance, when choosing $\tilde{p} = 0.1\%$, the risk loading, compared to the normal state, is multiplied by 10 in the case of TVaR, for $N = 10'000$ policies, and hardly moves in the case of VaR! This effect remains, but to a lower extend, when diminishing the number of policies. It is clear that the VaR measure does not capture well the crisis state, while TVaR is sensitive to the change of state, even with such a small probability and a high number of policies.

(ii)*A more realistic model setting to introduce a systematic risk*

We adapt further the previous setting to a more realistic description of a crisis. At each of the n exposures to the risk, in a state of systematic risk, the entire portfolio will be touched by the same increased probability of loss, whereas, in a normal state, the entire portfolio will be subject to the same equilibrium probability of loss.

For this modeling, it is more convenient to rewrite the sequence $(X_i,\ i = 1,\ldots,Nn)$ with a vectorial notation, namely $(\mathbf{X}_j,\ j = 1,\ldots,n)$ where the vector $\mathbf{X}_j$ is defined by $\mathbf{X}_j = (X_{1j},\ldots,X_{Nj})^T$. Hence the total loss amount S_{Nn} can be rewritten as

$$S_{Nn} = \sum_{j=1}^{n} \tilde{S}^{(j)} \quad \text{where} \quad \tilde{S}^{(j)} \text{ is the sum of the components of } \mathbf{X}_j \,:\, \tilde{S}^{(j)} = \sum_{i=1}^{N} X_{ij}.$$

We keep the same notation for the Bernoulli rv U determining the state and for its parameter $\tilde{p}$. But now, instead of defining a normal ($U=0$) or a crisis ($U=1$) state on each element of $(X_i,\, i=1,\ldots,Nn)$, we do it on each vector $\mathbf{X}_j$, $1\le j\le n$. It comes back to define a sequence of iid rv's $(U_j,\, j=1,\ldots,n)$ with parent rv U. We deduce that $\tilde{S}^{(j)}$ follows a Binomial distribution whose probability depends on U_j:

$$\tilde{S}^{(j)} \mid (U_j=1) \overset{d}{\sim} \mathscr{B}(N,q) \quad \text{and} \quad \tilde{S}^{(j)} \mid (U_j=0) \overset{d}{\sim} \mathscr{B}(N,p),$$

and these conditional rv's are independent.
Let us introduce the event A_l defined, for $l=0,\ldots,n$, as

$$\begin{aligned} A_l &:= \{l \,\text{vectors}\, \mathbf{X}_j \,\text{are exposed to a crisis state and}\, n-l \,\text{to a normal state}\} \\ &= \Big(\sum_{j=1}^{n} U_j = l\Big) \end{aligned}$$

whose probability is given by

$$\mathbb{P}(A_l) = \mathbb{P}\Big(\sum_{j=1}^{n} U_j = l\Big) = \binom{n}{l}\, \tilde{p}^{\,l}\, (1-\tilde{p})^{n-l}.$$

We can then write, with, by conditional independence,

$$\tilde{S}_q^{(l)} = \sum_{j=1}^{l} \Big(\tilde{S}^{(j)} \mid U_j=1\Big) \overset{d}{\sim} \mathscr{B}(Nl,q)$$

and

$$\tilde{S}_p^{(n-l)} = \sum_{j=1}^{n-l} \Big(\tilde{S}^{(j)} \mid U_j=0\Big) \overset{d}{\sim} \mathscr{B}(N(n-l),p),$$

that

$$\mathbb{P}(S_{Nn}=k) = \sum_{l=0}^{n} \mathbb{P}(S_{Nn}=k|A_l)\mathbb{P}(A_l) = \sum_{l=0}^{n} \binom{n}{l} \tilde{p}^{\,l}\, (1-\tilde{p})^{n-l}\, \mathbb{P}\big[\tilde{S}_q^{(l)} + \tilde{S}_p^{(n-l)} = k\big].$$

Numerical example revisited:
In this case, we cannot directly use an explicit expression for the distributions, so we go through Monte-Carlo simulations.
At each of the n exposures to the risk, first choose between a normal or a crisis state. Since, we take here $n=6$, the chances of choosing a crisis state when $\tilde{p}=0.1\%$ is very small. To get enough of the crisis states, we need to do enough simulations, and then average over all the simulations. The results shown in Table 8 are obtained with 10 million simulations (we ran it also with 1 and 20 million simulations to check the convergence).

Table 8 The Risk loading per policy as a function of the probability of occurrence of a systematic risk in the portfolio using VaR and TVaR measures with $\alpha = 99\%$. The probability of giving a loss in a state of systematic risk is chosen to be $q = 50\%$.

Risk measure ρ	**Number** N **of Policies**	**Risk Loading** R **in a normal state**	**with occurrence of a crisis state**			
		$\tilde{p} = 0$	$\tilde{p} = 0.1\%$	$\tilde{p} = 1.0\%$	$\tilde{p} = 5.0\%$	$\tilde{p} = 10.0\%$
VaR						
	1	3.000	2.997	2.969	4.350	4.200
	5	1.500	1.497	1.470	1.650	1.800
	10	1.050	1.047	1.170	1.350	1.500
	50	0.450	0.477	0.690	0.990	1.200
	100	0.330	0.357	0.615	0.945	1.170
	1'000	0.102	0.112	0.517	0.882	1.186
	10'000	0.032	0.033	0.485	0.860	1.196
	100'000	0.010	0.008	0.475	0.853	1.199
TVaR						
	1	3.226	3.232	4.485	4.515	4.448
	5	1.644	1.792	1.870	2.056	2.226
	10	1.164	1.252	1.342	1.604	1.804
	50	0.510	0.588	0.824	1.183	1.408
	100	0.375	0.473	0.740	1.118	1.358
	1'000	0.116	0.348	0.605	1.013	1.295
	10'000	0.037	0.313	0.563	0.981	1.276
	100'000	0.012	0.301	0.550	0.970	1.269
$\mathbb{E}[L]/N$		*10.00*	*10.02*	*10.20*	*11.00*	*12.00*

The diversification due to the total number of policies is more effective for this model than for the previous one, but we still experience a part which is not diversifiable. We also computed the case with 100'000 policies (since via Monte Carlo simulations). As expected, the risk loading in the normal state continues to decrease. In this state, it decreases by $\sqrt{10}$. However, except for $\tilde{p} = 0.1\%$ in the VaR case, the decrease becomes very slow when we allow for a crisis state to occur. The behavior of this model is more complex than the previous one, but more realistic, and we reach also the non-diversifiable part of the risk. For a high probability of occurrence of a crisis (1 every 10 years), the limit with VaR is reached already at 100 policies, while, with TVaR, it continues to slowly decrease. Concerning the choice of risk measure, we see a similar behavior as in the previous case for the case $N = 10'000$ and $\tilde{p} = 0.1\%$: VaR is unable to catch the possible occurrence of a crisis state, which shows its limitation as a risk measure. Although we know that there is a part of the risk that is non-diversifiable, VaR does not catch it really when $N = 10'000$ or $100'000$ while TVaR does not decrease significantly between $10'000$ and $100'000$ reflecting the fact that the risk cannot be completely diversified away.

Discussion - comparison of the methods.

Table 9 Summary of the analytical results (expectation and variance per policy) for the 3 cases of biased games ($L = l\,S_{Nn}$)

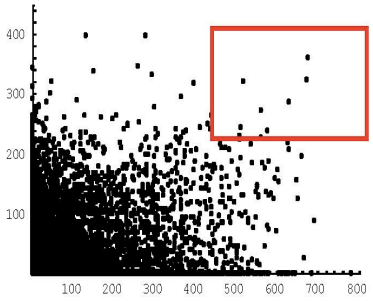
Case	Expectation $\frac{1}{N}\mathbb{E}(L)$	Variance $\frac{1}{N^2}var(L)$
(a)	$ln\ q$	$\frac{\mathbf{l^2 n}}{\mathbf{N}}\ q(1-q)$
(b)-(i)	$ln\Big(\tilde{p}\ q\ +\ (1-\tilde{p})\ p\Big)$	$\frac{\mathbf{l^2 n}}{\mathbf{N}}\Big(q(1-q)\tilde{p}+p(1-p)(1-\tilde{p})\Big)+\mathbf{l^2 n^2}(q-p)^2\tilde{p}(1-\tilde{p})$
(b)-(ii)	$ln\Big(\tilde{p}\ q\ +\ (1-\tilde{p})\ p\Big)$	$\frac{\mathbf{l^2 n}}{\mathbf{N}}\Big(q(1-q)\tilde{p}+p(1-p)(1-\tilde{p})\Big)+\mathbf{l^2 n}\ (q-p)^2\tilde{p}(1-\tilde{p})$

In Table 9, we see in the first case (a) that the variance decreases with increasing N, while both other cases (b) (i and ii) contain a term in the variance that does not depend on N. Those two cases are those containing a systematic risk component that cannot be diversified. Note that the variance $var_2(L)$ of L in the case (b)-(i) contains a non-diversified part that corresponds to n times the non-diversified part of $var_3(L)$ in the case (b)-(ii).

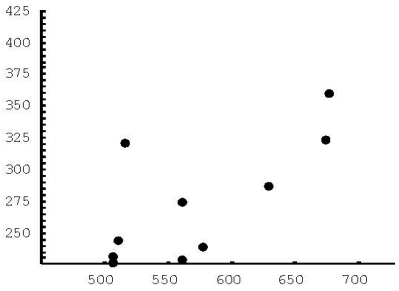
To conclude, we have seen the effect of diversification on the pricing of insurance risk through a simple modeling that allows for a straightforward analytical evaluation of the impact of the non-diversified part. In real life, risk takers have to pay special attention to the effects that can weaken the diversification benefits, hence affect greatly the risk loading of the risk premium (as seen here). Various examples can illustrate this situation, as, for instance, for motor insurance, the appearance of a hail storm may hit a big number of cars at the same time and thus cannot be diversified among the various policies, or for life insurance, pandemic or mortality trend would affect the entire portfolio and cannot be diversified away, or the financial crisis suddenly increases the dependence between risks (systemic risk). There is a saying among traders: "Diversification works the best when you need it the least". Understanding the dependence between risks is crucial for solid risk management. For portfolio management, we need to include both the single risk model and the dependence model.

3.1.2 Type of dependence

Consider a portfolio of political risks, the two largest ones being those of China and Hong-Kong, with 22.5% (linear) correlation. A customer asks a reinsurer for a cover of those extreme risks, providing him their marginal distributions and the following simulations results:



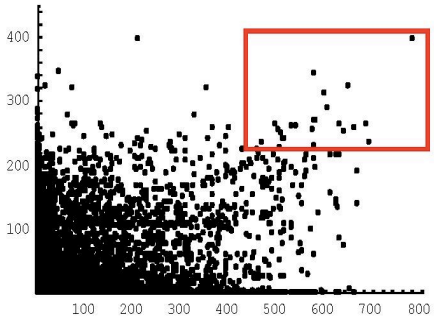
X-axis: China; Y-axis: Hong-Kong
($r = 22.5\%$)



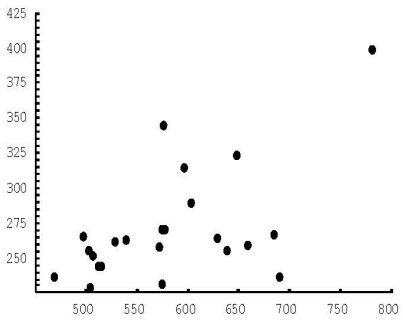
10 extreme events

Applying the reinsurance structure (rectangle) to the customer's simulations, we find 10 relevant events in it. However, in this model, the conditional probability for Hong-Kong to default on the risk, given that China defaults with probability of 1/200 years (ie 0.5%), would give a probability less than 5% , which is totally unrealistic, given the political situation of dependence of Hong-Kong on China!

Hence the reinsurer decides to study this portfolio, using the same margins, but suggesting a dependence structure via a Clayton copula, calibrating it to have the same linear correlation of 22.5%. He obtains a much more realistic conditional probability of default of 60%, which gives 21 relevant events in the reinsurance structure. Applying simply a non-linear dependence structure increases by a factor 2 the number of events and by a factor 3 the average loss for the reinsurer. Of course, the price of such a cover would be much higher than what the customer expected given his model.



X-axis: China; Y-axis: Hong-Kong
($r = 22.5\%$)



21 extreme events

This example shows that the type of dependence considered for the modeling matters a lot when considering the risk!

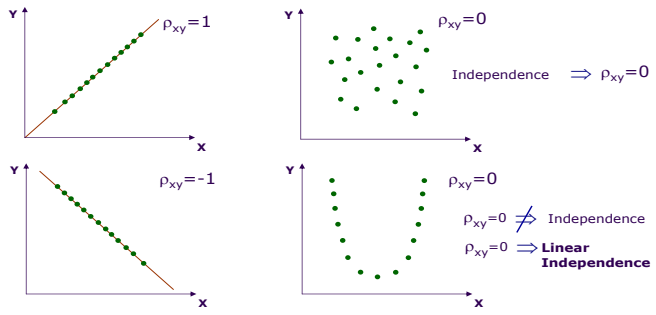
3.2 Notion of dependence

How to analyze a phenomenon in view of understanding it better, then modeling it? Modeling is a simplification but must not be a reduction! It is the fundamental basis of a scientific approach. 'Everything should be made as simple as possible, but not simpler' (Saying attributed to Albert Einstein).

We proceed from simplest tools to more elaborated ones, when needed. In terms of dependence, in a multivariate context, it means to look at the rv's from independence to linear dependence to non-linear dependence. Studying the dependence between risks is essential for understanding their real impacts and consequences. There exists many ways of describing dependence or association between rv's, e.g. linear correlation coefficient, rank correlations (Kendall's tau, Spearman's rho), ...

Let us present a brief historical overview. Dependence has always been a topic in probability and statistics when looking at what is called a multivariate framework. Notions like linear correlation or copula, for instance, were introduced to treat this problem.

Fig. 8 Linear versus Stochastic Independence



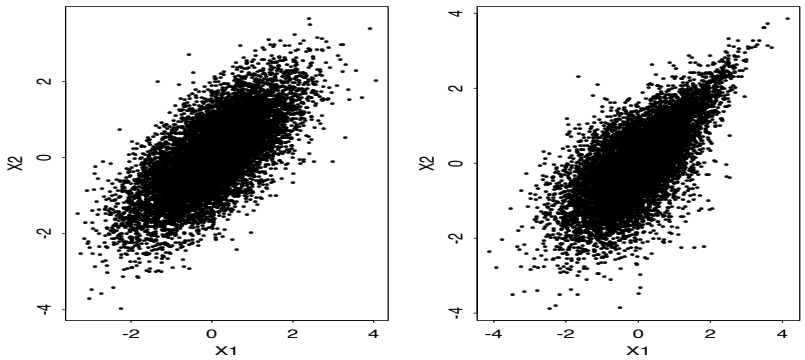
- In 1895: Karl Pearson[21] formalized mathematically the notion of linear correlation (first introduced by Galton in the context of biometric studies). If independence implies linear independence, the converse is false (except in the elliptical case), as illustrated on Figure 8.

- In 1959: Abe Sklar ([26]) introduced (in the context of probability theory to solve a theoretical problem posed by Fréchet) the more general concept of dependence structure, called also *copula*, separating this structure from the margins.

For instance, consider two random vectors having the same standard normal margins and a linear correlation of 70%, but a different dependence structure, a Gaussian copula and a Gumbel one, respectively. We clearly see in Figure 9 how different they are.

It emphasizes the fact that knowing the marginal distributions and linear correlation is not enough for determining the joint distribution, except for elliptical distributions (as e.g. the Gaussian ones).

Fig. 9 Scatterplots of (X_1, X_2) with normal margins, linear correlation $\rho = 70\%$, and, respectively, a Gaussian copula (left plot) and a Gumbel copula (right plot)



- In 1984: Paul Deheuvels[7] introduced the notion of extreme-value copula.

- From the 70's, diverse types of dependence have been studied in mathematical statistics and probability.

- From the 21st century, those dependence tools have been introduced in the industry: copulas turn out to become an important tool for applications and the evaluation of risks in insurance and reinsurance (and later in finance: non-linear tools cannot/should not be ignored anymore, especially after the second most severe financial crisis starting in 2008)

- After the 2008 financial crisis, Extreme Value Theory (EVT) finally enters the financial world (academics and professionals). The fact that risks are more interdependent in extreme situations led to the development of the notion of systemic risks, risks that would affect the entire system as well as the notion of systematic risks, where components are present in all other risks.

The world has changed a lot, from the end of the 19th-early 20th century, where using the concept of linear correlation (Pearson) was of great help, to nowadays, where world is getting more complex, and more and more interconnected (see [4] for a discussion of this point). In the next few years, research in statistics and probability will have to make significant progress in this area if we want to master the risk at an aggregate level. We have seen that societal demand goes in this direction, looking for protection at a global level.

3.3 Copulas

Definition 1. A *copula* is a multivariate distribution function $C : [0,1]^d \to [0,1]$ with standard uniform margins (or a distribution with such a df), i.e. $C(1,\cdots,1,u_i,1,\cdots,1) = u_i$, $\forall i \in \{1,...,d\}$, $u_i \in [0,1]$.

Sklar showed in [26] how a unique copula C fully describes the dependence of X proving the following theorem.

Theorem 4 (Sklar's theorem, 1959).
Let F be a joint cdf with margins $(F_i, i = 1, \cdots, d)$. There exists a copula C such that

$$F(x_1, \cdots, x_d) = C(F_1(x_1), \cdots, F_d(x_d)), \quad \forall x_i \in \mathbb{R}, \quad i = 1, \cdots, d.$$

If the margins are continuous then C is unique.
Conversely, if C is a copula and $(F_i,\ 1 \leq i \leq d)$ are univariate d.f., then F defined above is a multivariate df with margins $F_1, \cdots, F_d$.

Proof as an exercise.
As a consequence, we can give another definition of a copula.

Definition 2. The copula of $(X_1, \cdots, X_d)$ (or F) is the cdf C of $(F_1(X_1), \cdots, F_d(X_d))$.
We sometimes refer to C as the *dependence structure of F*.

Here is a useful way to express Sklar's theorem in dimension 2:

Theorem 5 (Sklar - dim 2). *Let F be a joint cdf with margins (F_1, F_2). The copula C associated to F can be written as*

$$\begin{aligned} C(u_1, u_2) &= C(F_1(x_1), F_2(x_2)) \\ C(u_1, u_2) &= F(x_1, x_2) \\ C(u_1, u_2) &= F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \end{aligned}$$

If the margins are continuous then C is unique.

Copulas satisfy a property of invariance, very useful in practice, and which is not satisfied by the linear correlation.

Property 1 (Property of Invariance). C is invariant under strictly increasing transformations of the marginals. If $T_1, \cdots, T_d$ are strictly increasing, then $(T_1(X_1), \cdots, T_d(X_d))$ has the same copula as $(X_1, \cdots, X_d)$.

As for probability distributions, we can define the notion of density function, when existing.

Definition 3. The *density function c of a copula C* is defined by

$$c(u_1, \cdots, u_d) = \frac{\partial^d C(u_1, \cdots, u_d)}{\partial u_1 \cdots \partial u_d}.$$

The density function of a bivariate distribution can be written in terms of the density function c of the associated copula and in terms of the density functions f_1 and f_2 of the margins:

$$f(x_1, x_2) = c\Big(F_1(x_1), F_2(x_2)\Big) f_1(x_1) f_2(x_2).$$

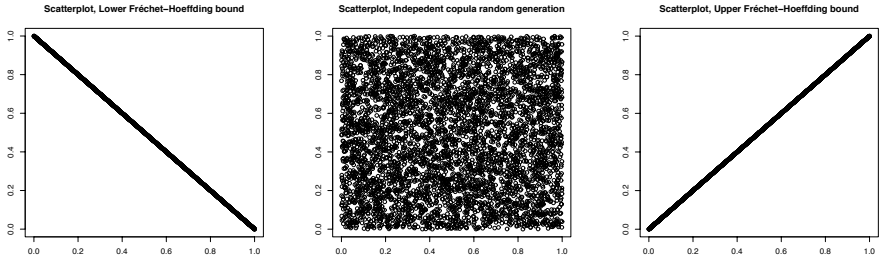
Using this definition, we can prove that: *the product copula characterize the independence between two r.v.* More generally,

$X_1,\dots,X_d$ are mutually independent $\Longleftrightarrow$ their copula C satisfies $C(u_1,\dots,u_d) = \prod_{i=1}^{d} u_i$.

As the linear correlation, which is bounded between -1 and 1, a copula also admits bounds, named *Fréchet-Hoeffding bounds*:

$$\max\Big(\sum_{i=1}^{d} u_i + 1 - d\ ;\ 0\Big) \ \le\ C(u)\ \le\ \min_{1\le i\le d} u_i = \mathbb{P}[U \le u_1,\cdots,U \le u_d]$$

where $u = (u_1,\dots,u_d)$ and U is uniformly distributed on [0,1].



The upper Fréchet-Hoeffding bound $C_u(u_1,\dots,u_d) := \min_{1\le i\le d} u_i$ is a copula for any d. It describes the perfect dependence, named also *comotonicity*:
$X_i \overset{a.s.}{=} T_i(X_1)$, with T_i strictly increasing function, $i = 2,\dots,d \Longleftrightarrow C_u$ satisfies $C_u(u_1,\dots,u_d) := \min_{1\le i\le d} u_i$.

The lower Fréchet-Hoeffding bound $C_l(u) := \max\Big(\sum_{i=1}^{d} u_i + 1 - d\ ;\ 0\Big)$ is a copula for $d = 2$, but not for all $d > 2$. For $d = 2$, C_l describes the perfect negative dependence, named also *countercomotonicity*: $X_2 \overset{a.s.}{=} T(X_1)$, with T strictly decreasing function $\Longleftrightarrow C_l(u_1,u_2) := \max(u_1+u_2-1,0)$.

Examples of copulas. Since any type of dependence structure can exist, the same can be said about copulas, that is why many new copulas are introduced by researchers. Here let us define three standard classes of copulas, already in use among practitioners, among which the Extreme Value (EV) copulas (which can overlap the two other classes).

- Elliptical or normal mixture copulas, as for instance:
  - The Gaussian copula (often used in financial modeling); in dimension 2, with parameter $\alpha \in (-1,1)$, it is defined via Sklar's theorem by

$$C(u,v) = \frac{1}{2\pi\sqrt{1-\alpha^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} exp\left\{-\frac{x^2 - 2\alpha xy + y^2}{2(1-\alpha^2)}\right\} dxdy$$

    where Φ denotes the standard normal distribution.

– The Student-t copula is the distribution of $(T(X_i), i = 1, \cdots, d)$ where T is the t-cdf and $(X_i, i = 1, \ldots, d)$ has a joint t-distribution. For $d = 2$, it can be expressed (via Sklar's theorem) as:

$$C(u,v) = \frac{1}{2\pi\sqrt{1-\alpha^2}} \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \left(1 + \frac{x^2 - 2\alpha xy + y^2}{2(1-\alpha^2)}\right)^{-(\nu+2)/2} dxdy$$

for $\alpha \in (-1,1)$ and degrees of freedom $\nu \geq 2$.

- Archimedean copulas.
Definition. An *Archimedean copula* C is defined by

$$C(u_1, \ldots, u_d) = \psi^{-1}\big(\psi(u_1) + \cdots + \psi(u_d)\big)$$

where $\psi :]0,1] \to [0,\infty)$ is continuous, strictly decreasing, convex, and satisfies $\psi(1) = 0$ and $\lim_{t\to 0} \psi(t) = +\infty$; set $\psi^{-1}(t) = 0$ if $\psi(0) \leq t \leq +\infty$. We call ψ the *strict generator of* C.
Archimedean copulas are *exchangeable*, *i.e.* invariant under permutation.

Examples:

– Gumbel copula, defined in dimension 2 by:

$$C_\beta^{Gu}(u,v) = \exp\big\{-\big((-\log u)^\beta + (-\log v)^\beta\big)^{1/\beta}\big\}, \quad \text{with } \beta \geq 1.$$

When $\beta = 1$: $C_1^{Gu}(u,v) = uv$ pointing out the independence of the variables.
When $\beta \to \infty$, the variables tend to be comonotonic: $C_\infty^{Gu}(u,v) = \min(u,v)$.

– Clayton copula , defined in dimension 2 by:

$$C_\beta^{Cl}(u,v) = \big(u^{-\beta} + v^{-\beta} - 1\big)^{-1/\beta}, \quad \text{with } \beta > 0.$$

We have $\lim_{\beta \searrow 0} C_\beta^{Cl}(u,v) = uv$, independence case, and $\lim_{\beta\to\infty} C_\infty^{Cl}(u,v) = \min(u,v)$, comonotonic case.

- Extreme Value (EV) Copulas
A copula C is said to be an Extreme Value (EV) copula if it satisfies the *max-stability* characteristic property:

$$\forall \gamma > 0, \quad C^\gamma(u_1, \ldots, u_d) = C(u_1^\gamma, \ldots, u_d^\gamma).$$

In dimension 2, an alternative definition is the following:

$$C(u,v) = \exp\left\{(\log u + \log v) A\left(\frac{\log u}{\log u + \log v}\right)\right\}$$

where A, called the dependence (or Pickands) function, is convex on $[0,1]$ and satisfies $A(0) = A(1) = 1$ and $\max(1-\omega, \omega) \leq A(\omega) \leq 1$, $\forall \omega \in [0,1]$.
The function A can be defined from the EV copula C by setting

$$A(w) = -\ln C\left(e^{-w}, e^{-(1-w)}\right), \quad w \in [0,1].$$

Bounds have also been provided for A. If the upper bound is reached, *i.e.* if $A(w) = 1$, $\forall w$, then C is the independence copula. If the lower bound is reached, *i.e.* if $A(w) = \max(w, 1-w)$, then it is a comonotonicity copula.

Examples of EV copulas: independence copula, comonotonicity copula, Gumbel copula (it is a parametric EV copula; the Gumbel copula model is sometimes known as the logistic model), Galambos copula (as defined below).

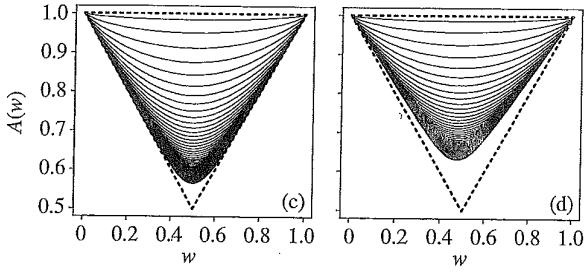


Fig. 10 Plot of dependence function for (c) the symmetric Galambos ($\alpha = \beta = 1$), spanning the whole range from independence to comonotonicity, and (d) the asymmetric Galambos copula with $\alpha = 0.9$ and $\beta = 0.8$; the limit as $\theta \to 0$ is the independence model, whereas as $\theta \to \infty$, it is no longer the comonotonicity model. Dashes lines show boundaries of the triangle in which the dependence function must reside; solid lines show dependence functions for a range of θ values running from 0.2 to 5 in steps of size 0.1

Let us consider the dependence function A, introduced by Galambos and defined, for $0 \leq w \leq 1$, with $0 \leq \alpha, \beta \leq 1$ and $\theta > 0$, by

$$A(w) = 1 - \left((\alpha w)^{-\theta} + \left(\beta(1-w)\right)^{-\theta}\right)^{-1/\theta}.$$

We can check that A is a convex function having the right bounds for the definition of an EV copula, so that we can create an EV copula from A. We obtain the bivariate EV copula, named Galambos copula (this copula model is sometimes known as the negative logistic model),

$$C^{Gal}_{\theta,\alpha,\beta}(u,v) = uv\,exp\left\{\left((-\alpha \ln u)^{-\theta} + (-\beta \ln v)^{-\theta}\right)^{-1/\theta}\right\}.$$

It is represented in Figure 10 ([18], p. 313).

3.4 Notion of Rank Correlation

Let us introduce two rank correlations, the Spearman's rho ρ_S and the Kendall's tau ρ_τ, which can also be expressed in terms of copulas (see e.g. [18], §5.2).

Let C denote the copula of (X_1, X_2), and ρ the Pearson (linear) correlation of X_1 and X_2.

- The *Spearman's rho* ρ_S is defined by

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)) = \rho(\text{copula})$$

and also by

$$\rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 \big(C(u_1, u_2) - u_1 u_2\big) du_1 du_2.$$

- The *Kendall's tau* ρ_τ is defined by

$$\rho_\tau(X_1, X_2) = 2\mathbb{P}\big[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0\big] - 1$$

with $(\tilde{X}_1, \tilde{X}_2)$ an independent copy of (X_1, X_2), and also by

$$\rho_\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1.$$

Case of elliptical models: Suppose $X = (X_1, X_2)$ has any elliptical distribution (e.g. X has a Student distribution $t_2(\nu, \mu, \Gamma)$). Then $\rho_\tau(X_1, X_2) = \dfrac{2}{\pi} \arcsin\big(\rho(X_1, X_2)\big)$. Note that if X_i has infinite variance, then $\rho(X_1, X_2)$ can be interpreted as $\frac{\Gamma_{1,2}}{\sqrt{\Gamma_{1,1}\Gamma_{2,2}}}$.

Properties of rank correlations.

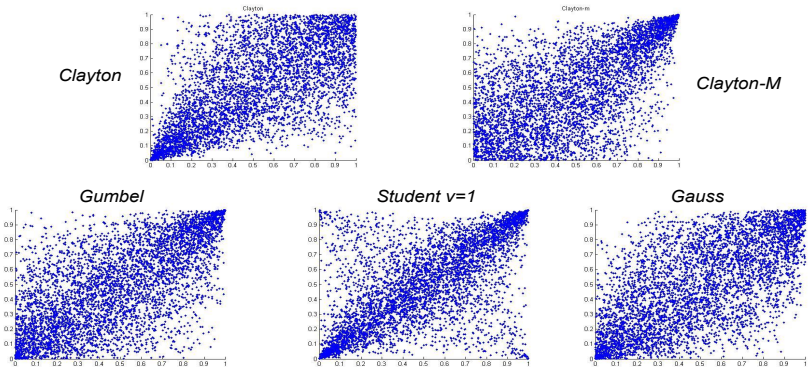
We can enunciate the following properties for the Spearman's rho ρ_S. The same holds true for Kendall's tau ρ_τ. But those properties are not shared by the linear correlation.

1. ρ_S depends only on the copula of (X_1, X_2);
2. ρ_S is invariant under strictly increasing transformations of the rv's;
3. $\rho_S(X_1, X_2) = 1 \Leftrightarrow C(X_1, X_2)$ is comonotonic;
4. $\rho_S(X_1, X_2) = -1 \Leftrightarrow C(X_1, X_2)$ is countermonotonic.

3.5 Ranked Scatterplots

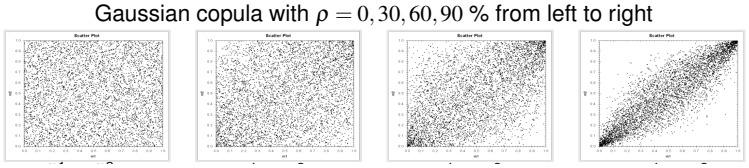
Let us draw ranked scatterplots with different copulas.

First we consider archimedean copulas, namely Clayton, Clayton-mirror (i.e. when we flip it) and Gumbel, and elliptical ones, namely Student copula with $\nu = 1$ and Gaussian one. For all of them, we choose the Kendall's tau $\rho_\tau = 50\%$.

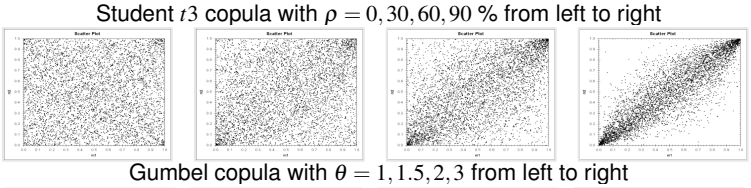


Then we consider the same margins and play with the parameters of the copulas to see their impact.

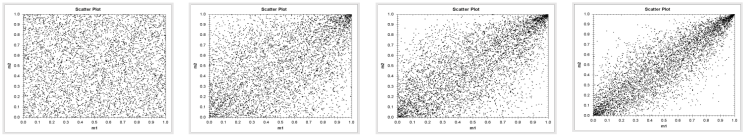
Gaussian copula with $\rho = 0, 30, 60, 90$ % from left to right



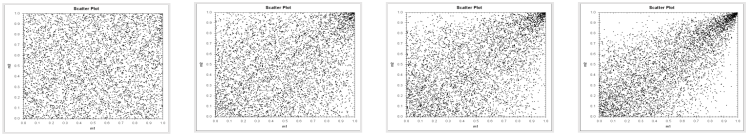
Student $t3$ copula with $\rho = 0, 30, 60, 90$ % from left to right



Gumbel copula with $\theta = 1, 1.5, 2, 3$ from left to right



Survival (Mirror) Clayton copula with $\theta = 0.1, 0.5, 1, 2$ from left to right



3.6 Other type of dependence: Tail or Extremal dependence

The objective is to measure the dependence in joint tail of bivariate distribution. Let C denote the copula of the random vector (X_1, X_2).

- *Coefficient of upper tail dependence.* When the limit exists, it is defined as

$$\lambda_u(X_1, X_2) = \lim_{\alpha \to 1} \mathbb{P}\big[X_2 > VaR_\alpha(X_2) \,|\, X_1 > VaR_\alpha(X_1)\big]$$

and, as function of the copula C,

$$\lambda_u(X_1,X_2) = \lim_{\alpha\to 1} \frac{1-2\alpha+C(\alpha,\alpha)}{1-\alpha}.$$

- *Coefficient of lower tail dependence.* When the limit exists, it is defined as

$$\lambda_l(X_1,X_2) = \lim_{\alpha\to 0} \mathbb{P}\big[X_2 \leq VaR_\alpha(X_2) \mid X_1 \leq VaR_\alpha(X_1)\big]$$

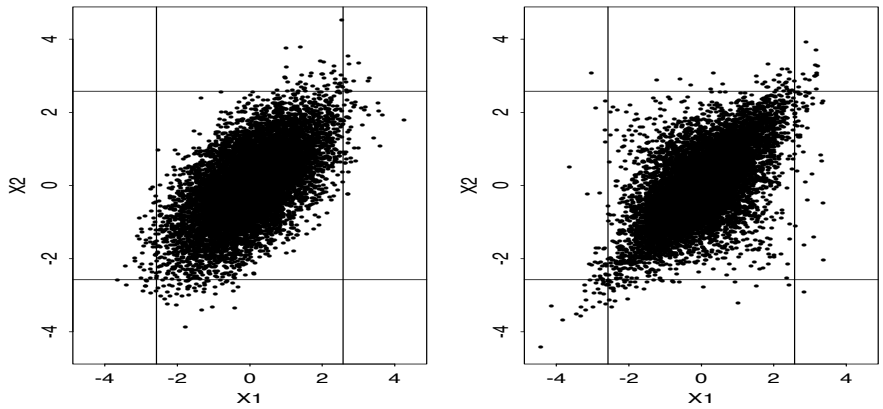
and, as function of C,

$$\lambda_l(X_1,X_2) = \lim_{\alpha\to 0} \frac{C(\alpha,\alpha)}{\alpha}.$$

Properties and terminology

1. $\lambda_u \in [0,1]$ and $\lambda_l \in [0,1]$;
2. For elliptical copulas, $\lambda_u = \lambda_l := \lambda$. Note that this is true for all copulas with radial symmetry, *i.e.* such that $(U_1,U_2) =^d (1-U_1,1-U_2)$;
3. If $\lambda_u \in (0,1]$, then there exits an upper tail dependence and if $\lambda_l \in (0,1]$, there exits a lower tail dependence;
4. $\lambda_u = 0$ means that there is asymptotic independence in the upper tail and $\lambda_l = 0$ means that there is asymptotic independence in lower tail.

Fig. 11 Gaussian (left) and Student $t3$ (right) copulas with same margins and parameter $\rho = 70\%$. Quantiles lines are given for 0.5% and 99.5%.



Examples.

1. We can prove that a Gaussian copula with parameter ρ is asymptotically independent (*i.e.* $\lambda = 0$) whenever $|\rho| < 1$;

2. A t-copula with parameter ρ is tail dependent whenever $\rho > -1$, whatever is the number of degrees of freedom ν. Its coefficient of (lower and upper) tail dependence is given by: $\lambda = 2\bar{t}_{\nu+1}\left(\sqrt{1+\nu}\sqrt{\frac{1-\rho}{1+\rho}}\right)$;
3. The Gumbel copula with parameter β is upper tail dependent for $\beta > 1$, and this upper tail dependence is measured by $\lambda_u = 2 - 2^{1/\beta}$;
4. The Clayton copula with parameter β is lower tail dependent for $\beta > 0$, and $\lambda_l = 2^{-1/\beta}$.

The properties of symmetric tail dependence, as well as of asymptotic tail dependence for the Gaussian copula and upper tail dependence for the Student copula, are well illustrated in Figure 11.

Numerical Example showing the impact of the choice of copula.

We already provided in §3.1.2 an example with political risks where we observed how much the choice of the dependence structure would impact the results. In that example we considered a Gaussian dependence versus a Clayton one.

Let us give another example where we compare in Table 10 the joint tail probability at finite levels of two copulas which are both elliptical. This is an example developed by McNeil et al. in [18].

Table 10 Left table: Joint Tail Probabilities $\mathbb{P}[X_1 > VaR_\alpha(X_1), X_2 > VaR_\alpha(X_2)]$ for $\alpha = 95, 99, 99.5, 99.9\%$, respectively. Right table: Joint Tail Probabilities $\mathbb{P}[X_i > VaR_{99\%}(X_i), i = 1,\ldots,d]$ for $d = 2,3,4,5$ respectively, when taking equal correlations. For both tables: The copula C of the random vector is either Gaussian (denoted by N) or Student t with 3 possible degrees of freedom $\nu = 8,4,3$ (the smaller is ν, the heavier is the tail) and parameter $\rho = 50\%$ or 70%. Note that for the Student cases, only the factor by which Gaussian (N) joint tail probability must be multiplied, is given.

ρ	C	Quantile			
		95%	99%	99.5%	99.9%
0.5	N	1.21×10^{-2}	1.29×10^{-3}	4.96×10^{-4}	5.42×10^{-5}
0.5	t8	1.20	1.65	1.94	3.01
0.5	t4	1.39	2.22	2.79	4.86
0.5	t3	1.50	2.55	3.26	5.83
0.7	N	1.95×10^{-2}	2.67×10^{-3}	1.14×10^{-3}	1.60×10^{-4}
0.7	t8	1.11	1.33	1.46	1.86
0.7	t4	1.21	1.60	1.82	2.52
0.7	t3	1.27	1.74	2.01	2.83

ρ	C	Dimension d			
		2	3	4	5
0.5	N	1.29×10^{-3}	3.66×10^{-4}	1.49×10^{-4}	7.48×10^{-5}
0.5	t8	1.65	2.36	3.09	3.82
0.5	t4	2.22	3.82	5.66	7.68
0.5	t3	2.55	4.72	7.35	10.34
0.7	N	2.67×10^{-3}	1.28×10^{-3}	7.77×10^{-4}	5.35×10^{-4}
0.7	t8	1.33	1.58	1.78	1.95
0.7	t4	1.60	2.10	2.53	2.91
0.7	t3	1.74	2.39	2.97	3.45

Let us illustrate those results, giving the financial interpretation suggested in [18].

Consider daily returns on 5 financial instruments and suppose that we believe that all correlations between returns are equal to 50%. However, we are unsure about the best multivariate model for these data. On one hand, if returns follow a multivariate Gaussian distribution then the probability that on any day all returns fall below their 1% quantiles is 7.48×10^{-5}. In the long run such an event will happen once every 13369 trading days on average, that is roughly once every 51.4 years (assuming 260 trading days in a year). On the other hand, if returns follow a multivariate t distribution with 4 degrees of freedom then such an event will happen 7.68 times more often, that is roughly once every 6.7 years, which would induce a very different

behavior in terms of risk management! During the subprime crisis, this was the problem of the too high rating given to the CDOs (Collateralized Debt Obligation) by the rating agencies, who only considered linear correlation for the dependence between the risks.

4 Multivariate EVT

Let us end those notes by giving a brief idea about the basis on which EVT has been extended in the multivariate setting. It is a research domain which has aroused an increasing interest this past decade, in particular due to its practical use.

4.1 MEV distribution

Some notation. Let $X_1,\cdots,X_i,\cdots,X_n$ be iid random vectors in $\mathbb{R}^d$, each X_i ($i=1,\cdots,n$) having its components denoted by $X_{ij}, j=1,\cdots,d$; they could be interpreted as losses of d different types. Let F be the joint df of any random vector X_i and $F_1,\cdots,F_d$ be its marginal cdf's. Let $M_{nj}=\max\limits_{1\leq i\leq n} X_{ij}$, for $j=1,\cdots,d$; it is the maximum of the jth component, and M_n be the d-random vector the vector of componentwise block maxima, i.e. with components M_{ij}, $i=1,\cdots,n$.

The main question that might be asked, when going from univariate EVT to multivariate one, is which underlying multivariate cdf's F are attracted to which MEV distributions H?

Definition 4. If there exist vectors of normalizing constants (of dimension d) $c_n>0$ and d_n such that $(M_n-d_n)/c_n$ converges in distribution to a random vector with joint (non-degenerated) cdf H, *i.e.*

$$\mathbb{P}\left[\frac{M_n-d_n}{c_n}\leq x\right]=F^n(c_nx+d_n)\quad\underset{n\to\infty}{\longrightarrow}\quad H(x),\quad x\in\mathbb{R}^d,$$

we say that F is in the Maximum Domain of Attraction of H, written $F\in MDA(H)$, and we refer to H as a *MEV* (Multivariate Extreme Value) distribution.

If H has non-degenerate margins, then

- these margins are univariate EV distributions of one of the three types, by application of univariate EVT;
- via Sklar's theorem, H has a copula, which is unique if the margins are continuous.

Theorem 6. *If $F\in MDA(H)$ for some F and H with GEV margins, then the unique copula C of H satisfies the scaling property:*

$$C^\gamma(u)=C(u^\gamma),\quad \forall u\in\mathbb{R}^d,\ \forall\gamma>0,$$

which means that C is an extreme value (EV) copula (as defined previously); it can then be the copula of a MEV distribution.

4.2 Copula Domain of Attraction

We can enunciate the following asymptotic theorem.

Theorem 7 (Galambos, 1987).
Let $F_i, i = 1, \cdots, d$, be some continuous marginals cdf's and C some copula. Let define $F(x) = C\big(F_1(x_1), \cdots, F_d(x_d)\big)$ and let $H(x) = C_0\big(H_1(x_1), \cdots, H_d(x_d)\big)$ be a MEV distribution with EV copula C_0. Then we have

$$F \in MDA(H) \text{ if and only if } \begin{cases} F_i \in MDA(H_i) \text{ for } i = 1, \cdots, d, \\ \text{and} \\ \lim\limits_{t\to\infty} C^t\left(u^{1/t}\right) = C_0(u), \quad u \in [0,1]^d. \end{cases}$$

Notice that:
- the marginal distributions of F determine the margins of the MEV limit but are irrelevant to the determination of its dependence structure;
- the copula C_0 of the limiting MEV distribution is determined solely by the copula C of the underlying distribution.

Definition 5. If $\lim\limits_{t\to\infty} C^t\left(u^{1/t}\right) = C_0(u)$, $u \in [0,1]^d$, for some C and some EV copula C_0, then we say that C belongs to the copula domain of attraction of C_0: $C \in CDA(C_0)$.

Upper tail dependence and CDA.

Proposition 1. *Let C be a bivariate copula with upper tail-dependence coefficient λ_u. Assume that $C \in MDA(C_0)$ for some EV copula C_0 with Pickands (dependence) function A. Then λ_u is also the upper tail-dependence coefficient of C_0 and is related to its dependence function by $\lambda_u = 2(1 - A(1/2))$.*

Proof. First, let us prove that C and C_0 have the same λ_u. To do so, we just need to check that $\lim\limits_{\alpha\to 1} \dfrac{1-C(\alpha,\alpha)}{1-\alpha} = \lim\limits_{\alpha\to 1} \dfrac{1-C_0(\alpha,\alpha)}{1-\alpha}$. We have, using the definition of $C \in CDA(C_0)$,

$$\begin{aligned} \lim_{\alpha\to 1} \frac{1-C_0(\alpha,\alpha)}{1-\alpha} &= \lim_{\alpha\to 1} \frac{\log C_0(\alpha,\alpha)}{1-\alpha} = \lim_{\alpha\to 1}\lim_{t\to\infty} \frac{\log\left(t[1-C(\alpha^{1/t},\alpha^{1/t})]\right)}{1-\alpha} \\ &= \lim_{\alpha\to 1}\lim_{s\to 0^+} \frac{1-C(\alpha^s,\alpha^s)}{-s\log(\alpha)} = \lim_{\alpha\to 1}\lim_{s\to 0^+} \frac{1-C(\alpha^s,\alpha^s)}{-\log(\alpha^s)} = \lim_{\beta\to 1^-} \frac{1-C(\beta,\beta)}{1-\beta}. \end{aligned}$$

hence the result. The converse is straightforward. □

Consequence: $\lambda_u = 0 \Rightarrow A(1/2) = 1 \Rightarrow A \equiv 1$ (since A convex function) $\Leftrightarrow C_0$ is the independence copula.

5 Conclusion

In these notes, we have explored the EVT both in the univariate as well as in the multivariate case by looking at dependence between rv's. We have seen that there are

mature methods for determining accurately the shape of the tail of the distribution. There are also methods to backtest statistically if the model captures it correctly. This can be done when choosing Expected Shortfall as a risk measure (see e.g. [13] and references therein). We pointed out through examples the importance for good risk management of accounting for extreme risks, but also of correctly modeling the non-linear dependence when present in the data or in the process to be studied. We hope to have shown that there is no excuse anymore to ignore EVT in quantitative risk management.

Acknowledgements I would like to thank the organizers of this workshop, Professors Konstantin Borovkov, Kais Hamza and Alexander Novikov, for the invitation to lecture on this topic.

References

1. BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J., TEUGELS, J.: Statistics of Extremes: Theory and Applications. Wiley (2004).
2. BUSSE, M., DACOROGNA, M. AND KRATZ, M.: Does risk diversification always work ? The answer through simple modelling. SCOR Paper **24** (2013)
3. BUSSE, M., DACOROGNA, M. AND KRATZ, M.: The impact of systemic risk on the diversification benefits of a risk portfolio. Risks **2**, 260-276 (2014)
4. DACOROGNA M. AND KRATZ M.: Living in a stochastic world and managing complex risks. Available on SSRN at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2668468` (2015)
5. DAVISON, A. AND SMITH, R.: Models for exceedances over high thresholds. J. Royal Stat. Soc. Series B **52** (3), 393-442 (1990)
6. DEBBABI, N., KRATZ, M. AND MBOUP, M.: A self-calibrating method for heavy tailed data modelling. Application in neuroscience and finance. Submitted Preprint (arXiv1612.03974v2) (2017)
7. DEHEUVELS, P.: Probabilistic aspects of multivariate extremes. In: J. Tiago de Oliveira (Ed.) Statistical extremes and applications (1984)
8. DE HAAN L. AND FERREIRA A.: Extreme Value Theory: An Introduction. Springer (2007)
9. EMBRECHTS P., KLÜPPELBERG C. AND MIKOSCH T.: Modelling Extremal Events for Insurance and Finance. Springer (1997)
10. EMMER S., KRATZ M. AND TASCHE D.: What is the best risk measure in practice? A comparison of standard measures. J. Risk **18:2**, 31-60 (2015)
11. HILL B.: A simple approach to inference about the tail of a distribution. Ann. Statist. **3**, 1163-1174 (1975)
12. KRATZ, M.: Normex, a new method for evaluating the distribution of aggregated heavy tailed risks. Application to risk measures. Extremes **17:4**, Special issue on *Extremes and Finance* (Guest Ed. P. Embrechts), 661-691 (2014)
13. KRATZ M., LOK Y.H. AND MCNEIL A.: Multinomial VaR Backtests: A simple implicit approach to backtesting expected shortfall. J. Banking and Finance **88**, 393-407 (2018)
14. KRATZ M. AND RESNICK S.: The QQ-estimator and heavy tails. Stoch. Models **12**, 699-724 (1996)
15. LEADBETTER R., LINDGREN G. AND ROOTZÉN H.: Extremes and Related Properties of Random Sequences and Processes. Springer (1983)
16. F. LONGIN (ED.): Extreme events in finance. A Handbook of Extreme Value Theory and its Applications. Wiley (2016)
17. MARKOWITZ H.: Portfolio Selection. J. Finance, **7**(1), 77-91 (1952)

18. MCNEIL A., FREY R. AND EMBRECHTS P.: Quantitative Risk Management. Princeton Series in Finance (2005; 2016, 2nd Ed.)
19. NELSEN R. An introduction to copulas. Springer (1999)
20. NOVAK S. Extreme Value Methods with Applications to Finance. Chapman & Hall, CRC Press (2011)
21. PEARSON K.: Notes on Regression and Inheritance in the Case of Two Parents. Royal Society Proceedings **58**, 240-242 (1895)
22. REISS R-D. AND THOMAS M.: Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields. Birkhäuser Verlag (1997; 2007, 2nd Ed.)
23. RESNICK S.: Extreme Values, Regular Variation, and Point Processes. Springer (1987; 2008, 2nd Ed.)
24. RESNICK S.: Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. Springer (2006)
25. SHARPE H. F.: The Sharpe Ratio. The Journal of Portfolio Management, **21**(1). 49-58 (1994)
26. SKLAR, A.: Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris (ISUP) **8**, 229-231 (1959)