Risk Aggregation

Paul Embrechts and Giovanni Puccetti

1 Motivations and preliminaries

Quantitative Risk Management (QRM) standardly concerns a vector of one-period profit-and-loss random variables $\mathbf{X} = (X_1, \ldots, X_d)'$ defined on some probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. *Risk Aggregation* concerns the study of the aggregate financial position $\Psi(\mathbf{X})$, for some measurable function $\Psi : \mathbb{R}^d \to \mathbb{R}$.

Under the terms of the New Basel Capital Accord (Basel II), internationally active banks are required to set aside capital to offset various types of risks, i.e. market, credit and operational risk; see [4]. Under the new regulations, the vector $\mathbf{X}$ represents the profit-and-loss amounts for particular lines of risk or business, and this over a given period. A risk measure ρ maps the aggregate position $\Psi(\mathbf{X})$ to $\rho(\Psi(\mathbf{X})) \in \mathbb{R}$, to be interpreted as the regulatory capital needed to be able to hold the aggregate position $\Psi(\mathbf{X})$ over this predetermined fixed period. The exact calculation of $\rho(\Psi(\mathbf{X}))$ needs the joint distribution function $F_{\mathbf{X}}$ of $\mathbf{X}$; when such information is not available, special procedures are called for, typically leading to bounds on $\rho(\Psi(\mathbf{X}))$.

Risk Aggregation has often been studied within the framework when only the marginal distributions $F_1, \ldots, F_d$ of the individual risks $X_1, \ldots, X_d$ are available. A multitude of statistical techniques are available for estimating the univariate (marginal) distributions. It is often more difficult to capture statistically the d-variate structure of dependence of the vector $\mathbf{X}$. Recently, especially in the management of operational risk, cases in which further dependence information is available have become relevant.

Paul Embrechts, Senior SFI-Chair
Department of Mathematics, ETH Zurich, Switzerland. e-mail: embrechts@math.ethz.ch

Giovanni Puccetti
Department of Mathematics for Decisions, University of Firenze, Italy.
e-mail: giovanni.puccetti@dmd.unifi.it

In the following, we introduce a general mathematical framework which interpolates between marginal knowledge $(F_1,\dots,F_d)$ and full knowledge of $F_{\mathbf{X}}$. For the purpose of this paper, we disregard the statistical uncertainty related to $F_1,\dots,F_d$ and only concentrate on the probabilistic structure.

1.1 The mathematical framework

We follow the mathematical setup described in [37]. Let $B=\Pi_{i=1}^d B_i$ be the product of d Borel spaces with σ-algebra $\mathscr{B}=\bigotimes_{i=1}^n \mathscr{B}_i$, $\mathscr{B}_i$ being the Borel σ-algebra on B_i. Define $I:=\{1,\dots,d\}$ and let $\xi\subset 2^I$, the power set of I, with $\cup_{J\in\xi} J=I$. For $J\in\xi$, let $F_J\in\mathfrak{F}(B_J)$ be a *consistent* system of probability measures on $B_J=\pi_J(B)=\Pi_{j\in J}B_j$, π_J being the natural projection from B to B_J and $\mathfrak{F}(B_J)$ denoting the set of all probability measures on B_J. Consistency of $F_J, J\in\xi$ means that $J_1,J_2\in\xi, J_1\cap J_2\neq\emptyset$ implies that

$$\pi_{J_1\cap J_2}F_{J_1}=\pi_{J_1\cap J_2}F_{J_2}.$$

Finally, we denote by

$$\mathfrak{F}_\xi=\mathfrak{F}(F_J,J\in\xi)$$

the *Fréchet class* of all probability measures on B having marginals $F_J, J\in\xi$.

Consistency of $F_J,J\in\xi$, is a necessary condition to guarantee that F_ξ is non-empty. When ξ is *regular* (see [41]), then consistency is also sufficient. When the system ξ is non-regular, e.g. $\xi=\{\{1,2\},\{2,3\},\{3,1\}\}$, the Fréchet class $\mathfrak{F}_\xi$ may be empty even with consistent marginals, as illustrated in [37].

In the following, we will consider the case $B_i=\mathbb{R}$, $B=\mathbb{R}^n$. For the sake of notational simplicity, we identify probability measures on these spaces with the corresponding distribution functions. We will study only regular systems of marginals which interpolate between two particular choices of ξ:

- $\xi_d=\{\{1\},\dots,\{d\}\}$, also called the *simple* system of marginals, which defines the Fréchet class
$$\mathfrak{F}_{\xi_d}=\mathfrak{F}(F_1,\dots,F_d).$$
This is the most often used marginal system in Risk Aggregation and the natural setup for the theory of copulas, as discussed in [28].
- $\xi_I=\{I\}$, also called the *trivial* system of marginals, in which
$$\mathfrak{F}_{\xi_I}=\{F_{\mathbf{X}}\},$$
where $F_{\mathbf{X}}$ is the distribution function of the vector $\mathbf{X}$. This case represents the complete dependence information about $\mathbf{X}$.

There are other important cases representing intermediate dependence information between ξ_d and ξ_I. Relevant examples are:

- $\xi_d^M = \{\{2j-1, 2j\}, j = 1, \ldots, d/2\}$ (d even), the *multivariate* system of marginals. This system has the role of the simple marginal system when one studies aggregation of random vectors instead of aggregation of random variables.
- $\xi_d^\star = \{\{1, j\}, j = 2, \ldots, d\}$, the *star-like* system of marginals and
- $\xi_d^= = \{\{j, j+1\}, j = 1, \ldots, d-1\}$, the *serial* system of marginals. These latter two systems are of particular interest when dependence from bivariate datasets is available.

When ξ is a partition of I, i.e. when all sets $J \in \xi$ are pairwise disjoint, we speak about a *non-overlapping* system of marginals, *overlapping* otherwise. According to this definition, ξ_d, ξ_I and ξ_d^M are non-overlapping marginal systems, while $\xi_d^\star$ and $\xi_d^=$ are overlapping. We study Risk Aggregation under incomplete information frameworks in Section 2. Section 3 will focus instead on the problems arising within the complete information system ξ_I.

2 Bounds for functions of risks: the coupling-dual approach

We will focus on those risk measures $\rho(\Psi(\mathbf{X}))$ which are representable as

$$\rho(\Psi(\mathbf{X})) = \mathbb{E}[\psi(\mathbf{X})] = \int \psi \, dF_{\mathbf{X}}, \tag{1}$$

for some measurable function $\psi : \mathbb{R}^d \to \mathbb{R}$. This representation includes some of the most popular risk measures, such as Value-at-Risk. Most importantly, (1) will allow to use the theory of *Mass Transportations* within the context of Risk Aggregation. In this section, we illustrate how to obtain bounds on $\mathbb{E}[\psi(\mathbf{X})]$ under an incomplete information setting, i.e. when the distribution function $F_{\mathbf{X}}$ of the vector $\mathbf{X}$ is not completely specified. Formally, we assume that

$$F_{\mathbf{X}} \in \mathfrak{F}_\xi, \text{ for a fixed } \xi \subset 2^I, \xi \neq I.$$

Since $F_{\mathbf{X}}$ is not uniquely determined, there exist an entire range of values for $\mathbb{E}[\psi(\mathbf{X})]$, which are consistent with the choice of the subgroups $J \in \xi$ of marginals. The infimum and supremum of this range are defined as

$$m_\xi(\psi) := \inf\left\{\int \psi \, dF_{\mathbf{X}} : F_{\mathbf{X}} \in \mathfrak{F}_\xi\right\}, \tag{2a}$$

$$M_\xi(\psi) := \sup\left\{\int \psi \, dF_{\mathbf{X}} : F_{\mathbf{X}} \in \mathfrak{F}_\xi\right\}. \tag{2b}$$

Since the Fréchet class $\mathfrak{F}_\xi$ is convex and the problems (2a) and (2b) are linear on $F_{\mathbf{X}}$, (2a) and (2b) both admit a dual representation. This representation is to be found in the theory of Mass Transportations.

Theorem 2.1. *Let the measurable function ψ be bounded or continuous, then problems* (2a) *and* (2b) *have the following dual counterparts:*

$$m_\xi(\psi) = \sup\left\{ \sum_{J\in\xi} \int f_J \, dF_J : f_J \in L^1(F_J), J \in \xi \text{ with } \sum_{J\in\xi} f_J \circ \pi_J \leq \psi \right\}, \quad (3a)$$

$$M_\xi(\psi) = \inf\left\{ \sum_{J\in\xi} \int g_J \, dF_J : g_J \in L^1(F_J), J \in \xi \text{ with } \sum_{J\in\xi} g_J \circ \pi_J \geq \psi \right\}. \quad (3b)$$

There exist several versions of Theorem 2.1 which are valid under weaker assumptions and more general settings, even non-topological ones. For more details on these versions, a proof of Theorem 2.1 and a complete coverage of the theory of Mass Transportations, we refer to the milestone book [35] and the review paper [38].

According to [23], we call a *coupling* every random vector $\mathbf{X}$ having distribution function $F_{\mathbf{X}} \in \mathfrak{F}_\xi$. Moreover, we call *dual choice* for (3a) any family of functions $\mathbf{f} = \{f_J, J \in \xi\}$ which are admissible for (3a). Analogously, we define a dual choice $\mathbf{g} = \{g_J, J \in \xi\}$ for (3b). By Theorem 2.1, a coupling $\mathbf{X}$ and two dual choices $\mathbf{f}$ and $\mathbf{g}$ satisfy

$$\int \psi \, dF_{\mathbf{X}} \geq m_\xi(\psi) \geq \sum_{J\in\xi} \int f_J dF_J, \quad (4a)$$

$$\int \psi \, dF_{\mathbf{X}} \leq M_\xi(\psi) \leq \sum_{J\in\xi} \int g_J dF_J, \quad (4b)$$

for all $F_{\mathbf{X}} \in \mathfrak{F}_\xi$. A coupling and a dual choice which satisfy (4a) (or (4b)) with two equalities will be called an *optimal coupling* and a *dual solution*, respectively, since they solve problem (2a) (or (2b)).

Equations (4a) and (4b) illustrate the *coupling-dual* approach in Risk Aggregation. Problems (2a) and (2b) are in general very difficult to solve, with some exceptions illustrated in Section 2.2 below. Depending on the system ξ of marginals, the concept of a copula might not be useful and it may be difficult even to identify a single coupling in F_ξ. When the solutions of (2a) and (2b) are unknown, any dual choice satisfying (4a) or (4b) provides a bound on m_ξ or M_ξ.

2.1 Application 1: bounding Value-at-Risk

We now illustrate the usefulness of the dual representations (3a) and (3b), in the case of Value-at-Risk (VaR). VaR is probably the most popular risk measure in finance and insurance, this is no doubt due to its importance within the Basel II capital-adequacy framework; see [4]. The VaR of a profit-and-loss random variable L at the probability (or confidence) level $\alpha \in (0,1)$ is simply the α-quantile of its distribution, defined as

$$\mathrm{VaR}_\alpha(L) = F_L^{-1}(\alpha) = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}, \tag{5}$$

where F_L is the distribution of L. Under the terms of Basel II, banks often measure the risk associated with a portfolio $\mathbf{X} = (X_1, \ldots, X_d)'$ in terms of $\mathrm{VaR}_\alpha(X_1 + \cdots + X_d)$, the VaR of the sum of its marginal components. This is for example the case of operational risk; see [15]. Using our notation, we have $\rho = \mathrm{VaR}$ and $\Psi = +$, the sum operator. Typical values for α are $\alpha = 0.95$ or $\alpha = 0.99$, or even $\alpha = 0.999$ in the case of credit and operational risk. By (5), bounding the VaR of a random variable L from above is equivalent to bounding from below its distribution F_L or, similarly, bounding from above its tail (or survival) function $\overline{F_L} = 1 - F_L$. Roughly speaking, if VaR is used to risk measure L, a *higher tail function* for L means a *more dangerous risk*.

Banks often have more precise information about the marginal distributions of $\mathbf{X}$, but less about the joint distribution $F_{\mathbf{X}}$. This then immediately translates into the incomplete information setting ξ_d, which defines the Fréchet class $\mathfrak{F}(F_1, \ldots, F_d)$. Within ξ_d, banks are typically interested on a upper bound on $\mathrm{VaR}_\alpha\left(\sum_{i=1}^d X_i\right)$, since this latter amount cannot be calculated exactly. Such a bound can be obtained by solving problem (2b) for a particular choice of the function ψ, in this case $\psi = \psi(s) = 1_{\{x_1 + \cdots + x_d \geq s\}}$, for some $s \in \mathbb{R}$. Thus, we define the function M_{ξ_d} as

$$M_{\xi_d}(s) = \sup\left\{ \int 1_{\{x_1 + \cdots + x_d \geq s\}} dF_{\mathbf{X}}(x_1, \ldots, x_d), F_{\mathbf{X}} \in \mathfrak{F}(F_1, \ldots, F_d) \right\}, s \in \mathbb{R}. \tag{6}$$

Note that the inequality $\geq$ in the definition of the indicator function in (6) is essential in order to guarantee that the supremum is attained; see Remark 3.1(ii) in [13]. With respect to any random vector $(X_1, \ldots, X_d)'$ having distribution $F_{\mathbf{X}} \in \mathfrak{F}(F_1, \ldots, F_d)$, the function M_{ξ_d} obviously satisfies

$$\mathbb{P}[X_1 + \cdots + X_d \geq s] \leq M_{\xi_d}(s) \text{ for all } s \in \mathbb{R}, \tag{7}$$

while, for its inverse $M_{\xi_d}^{-1}$, we have

$$\mathrm{VaR}_\alpha(X_1 + \cdots + X_d) \leq M_{\xi_d}^{-1}(1 - \alpha), \text{ for all } \alpha \in (0,1). \tag{8}$$

According to Theorem 2.1, the dual counterpart of (6) is given by:

$$\begin{aligned} M_{\xi_d}(s) = \inf\Bigg\{ & \sum_{i=1}^d \int f_i \, dF_i : f_i \in L_1(F_i), i \in I \\ & \text{s.t. } \sum_{i=1}^d f_i(x_i) \geq 1_{\{x_1 + \cdots + x_d \geq s\}} \text{ for all } x_i \in \mathbb{R}, i \in I \Bigg\}. \end{aligned} \tag{9}$$

The dual solution for (9) is given in [36] for the sum of two risks ($d = 2$). Independently from this, [25] provided the corresponding optimal coupling. For the sum of more than two risks, (9) seems to be very difficult to solve. The only explicit

results known in the literature are given in [36] for the case of the sum of marginals being all uniformly or binomially distributed.

When the value of $M_\xi(s)$ is unknown, equation (4b) plays a crucial role. In fact, every dual admissible choice in (9) gives un upper bound on $M_\xi(s)$ which, though not sharp, is conservative from a risk management point of view. This is for instance the idea used in [13] to produce bounds on $\mathrm{VaR}_\alpha\left(\sum_{i=1}^d X_i\right)$. The following theorem is a reformulation of Th. 4.2 in the above reference and illustrates the case of a homogeneous risk portfolio, i.e. $F_i = F$ for all $i = 1,\dots,d$.

Theorem 2.2. *Let F be a continuous distribution with non-negative support. If $F_i = F, i = 1,\dots,d$, then, for every $s \geq 0$,*

$$M_{\xi_d}(s) \leq D_{\xi_d}(s) = d \inf_{r\in[0,s/d)} \frac{\int_r^{s-(d-1)r}(1-F(x))dx}{s-dr}. \tag{10}$$

The infimum in (10) can be easily calculated numerically by finding the zero-derivative points of its argument. For $d = 2$, we obtain $M_{\xi_d}(s) = D_{\xi_d}(s)$, the bound given in [36]. The idea of using dual choices to produce bounds on functions of risks was discussed further in [12] (within simple systems with non-homogeneous marginals), [14] (multivariate systems) and [16] (overlapping systems). Bounds produced by a choice of admissible dual functionals are referred to as *dual bounds*. A related study of bounds on VaR can be found in [22].

In Figure 1, we plot the dual bound function D_{ξ_d} for a portfolio of three ($d = 3$) Gamma-distributed risks. In the same figure, we also give the tail function of the random variable $X_1 + X_2 + X_3$ in case of comonotonic ($C_{\mathbf{X}} = M$) and independent ($C_{\mathbf{X}} = \Pi$) marginals; for this notation, see [28], Chapter 5. Note that the two tail functions cross at some threshold $\hat{s}$ and the tail function obtained under comonotonicity lies above the one obtained under independence for all $s > \hat{s}$. We will return on this later in Section 3. Table 1 shows the upper bounds $D^{-1}_{\xi_d}(1-\alpha)$ on the VaR of the Gamma portfolio, as well as exact quantiles in case of independence and comonotonicity. Recall that for comonotonic risks VaR is additive, see also (15) later in the paper. Figure 1 and Table 1 exemplify the fact that, using (7), (8) and (10), we have

$$\mathbb{P}[X_1 + \cdots + X_d \geq s] \leq D_{\xi_d}(s), \text{ i.e. } \mathrm{VaR}_\alpha(X_1 + \cdots + X_d) \leq D^{-1}_{\xi_d}(1-\alpha),$$

for any $(X_1,\dots,X_d)'$ having distribution $F_{\mathbf{X}} \in \mathfrak{F}_{\xi_d}$.

We finally remark that the entire curve $D_{\xi_d}(s)$ is generally obtained within seconds, independently of the number d of variables under study. In general, the computational time of dual bounds strongly depends on the number of non-homogeneous marginals.

It is interesting to study how dual bounds vary within different marginal systems having the same univariate marginals. To this aim, we now consider d risks $X_1,\dots,X_d$ which we assume to be Pareto distributed with tail parameter θ, i.e.

$$F_i(x) = \mathbb{P}[X_i \leq x] = 1 - (1+x)^{-\theta},\ x \geq 0,\ i = 1,\dots d. \tag{11}$$

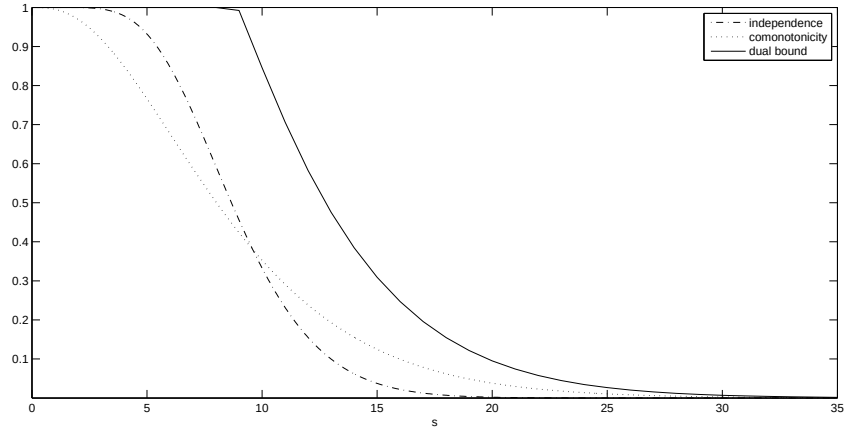

Fig. 1 Plot of the tail function $\mathbb{P}[X_1+X_2+X_3 \geq s]$ for a $\Gamma(3,1)$-portfolio under independence and comonotonic scenarios. We also plot the upper dual bound function $D_{\xi_d}(s)$.

α	$C_{\mathbf{X}} = \Pi$	$C_{\mathbf{X}} = M$	dual bound
0.90	13.00	15.97	19.80
0.95	14.44	18.89	22.57
0.99	17.41	25.22	28.67
0.999	21.16	33.69	36.97

Table 1 $\mathrm{VaR}_\alpha(X_1+X_2+X_3)$ for a $\Gamma(3,1)$-portfolio under independence and comonotonicity, for some levels α of interest. We also give the corresponding upper dual bounds $D^{-1}_{\xi_d}(1-\alpha)$.

Together with the non-overlapping marginal system ξ_d studied above, we consider the overlapping star-like system ξ^*_d. Under ξ^*_d, we assume that each of the $d-1$ subvectors $(X_1,X_i), i=2,\dots,d$, is coupled by a Frank copula C^F_δ with parameter $\delta = 1$. Within the system ξ^*_d, bounds on $\mathrm{VaR}_\alpha(\sum_{i=1}^d X_i)$ are obtained by integration of particular dual bounds in ξ_d. For more details on this technique, we refer to [16].

In Table 2, we give upper VaR limits $D^{-1}_{\xi^\star_d}(\alpha)$ for Frank-Pareto portfolios of increasing dimensions. As quantile levels, we take $\alpha = 0.99$ and $\alpha = 0.999$. For comparison, the comonotonic quantiles are also given. Considering the absolute values reported in Table 2, the overlapping bounds are smaller than the corresponding bounds obtained in a non-overlapping setting. The reason is clear: switching from a non-overlapping simple system to a overlapping star-like marginal system means reducing the Fréchet class of attainable risks, i.e. having more information about the dependence structure of the portfolio $\mathbf{X}$. Formally, we have $\mathfrak{F}_{\xi^\star_d} \subset \mathfrak{F}_{\xi_d}$. Under the extra information represented by $\xi^\star_d$, less capital is needed to offset the underlying portfolio risk.

Detailed studies of the quality of the dual bounds have been presented in [13] for $D^{-1}_{\xi_d}$, and in [16] for $D^{-1}_{\xi^*_d}$.

	$\alpha = \mathbf{0.99}$		$\alpha = \mathbf{0.999}$	
d	*overlapping*	*non-overlapping*	*overlapping*	*non-overlapping*
3	29.98	46.70	95.17	156.98
4	51.82	70.75	167.24	248.98
5	78.46	98.44	253.83	348.55
6	108.99	129.36	352.62	458.76
7	143.03	178.20	463.35	578.66
8	180.12	218.27	584.19	707.54
9	220.14	261.00	712.03	844.81
10	262.83	306.27	850.30	990.00

Table 2 Upper bounds on Value-at-Risk for the sum of d Pareto(2)-distributed risks within the *overlapping* star-like $\xi_d^\star$ and the *non-overlapping* marginal system ξ_d. Under the star-like system, the bivariate marginals are coupled by a Frank copula with parameter $\delta = 1$.

Open problems

The search for $M_{\xi_d}(s)$, i.e. for the largest VaR over $\mathfrak{F}(F_1,\dots,F_d)$, is open when $d > 2$. The proof of the optimality of the dual functionals for the case $d = 2$, given in [36], is based on Strassen's theorem (see Th. 11 in [40]). Unfortunately, Strassen's theorem does not have an obvious extension to the product of more than two marginal spaces; see [39] and references therein.

The search for $m_{\xi_d}(s)$, i.e. for the smallest VaR over $\mathfrak{F}(F_1,\dots,F_d)$, is again open when $d > 2$. For general dimensions d, several authors have obtained an elementary lower bound for $m_{\xi_d}(s)$; see for instance [7]. In models of actuarial interest, in [15] it is shown that the last mentioned lower bound does not depend on d. Therefore, a better bound on $m_{\xi_d}(s)$ is needed.

Finally, VaR dual bounds of the type (10) are needed for more general aggregating functionals Ψ.

2.2 Application 2: supermodular functions

In the simple marginal setting $\xi = \xi_d$, there are some functionals ψ for which the solutions of problems (2a) and (2b) are known. They form the class $\mathscr{S}_d$ of supermodular functions.

Definition 2.1. A measurable function $\psi : \mathbb{R}^d \to \mathbb{R}$ is said to be *supermodular* if

$$\psi(\mathbf{u}\wedge\mathbf{v}) + \psi(\mathbf{u}\vee\mathbf{v}) \geq \psi(\mathbf{u}) + \psi(\mathbf{v}), \text{ for all } \mathbf{u},\mathbf{v} \in \mathbb{R}^d,$$

where $\mathbf{u}\wedge\mathbf{v}$ is the componentwise minimum of $\mathbf{u}$ and $\mathbf{v}$, and $\mathbf{u}\vee\mathbf{v}$ is the componentwise maximum of $\mathbf{u}$ and $\mathbf{v}$.

When $d = 2$, a function $c : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is supermodular if and only if

$$\psi(x_1,y_1)+\psi(x_2,y_2)\geq\psi(x_1,y_2)+\psi(x_2,y_1),\ \text{for all } x_2\geq x_1,\ y_2\geq y_1. \qquad (12)$$

Recall that, for any set of univariate distributions $F_1,\dots,F_d$, there exists a *comonotonic coupling* $\mathbf{X}^M$, i.e. a random vector having marginals $F_1,\dots,F_d$ and copula M.

Theorem 2.3. *For given univariate distributions $F_1,\dots,F_d$, denote by $\mathbf{X}^M$ a comonotonic coupling having these marginals. Let $\psi:\mathbb{R}^d\to\mathbb{R}$ be right-continuous. Then*

$$\mathbb{E}\left[\psi\left(\mathbf{X}^M\right)\right]=\sup\left\{\int\psi\,dF_{\mathbf{X}}:F_{\mathbf{X}}\in\mathfrak{F}(F_1,\dots,F_d)\right\},\ \textit{for all } F_1,\dots,F_d, \qquad (13)$$

if and only if $\psi\in\mathscr{S}_d$.

Proof. The *if* part follows from [35, Remark 3.1.3], but many authors have derived the same result under different regularity conditions: see for instance [24] and [5]. For the *only if* part, see [34]. □

The most popular supermodular function is the product $\times(\mathbf{x})=\Pi_{i=1}^n x_i$. When $\psi=\times$, Theorem 2.3 gives the well-known result that a multivariate comonotonic distribution maximizes correlation between its marginals.

Note that Theorem 2.3 applies to a large class of interesting functionals, including $\psi(\mathbf{x})=\sum_{i=1}^d h_i(x_i)$, where the h_i's are non-decreasing (see [29]) and $\psi(\mathbf{x})=h(\sum_{i=1}^d x_i)$ for h non-decreasing and convex; see [26, pp. 150–155]. In insurance, $\sum_{i=1}^d h_i(x_i)$ and $h(\sum_{i=1}^d x_i)$ can be interpreted, respectively, as the risk positions for a reinsurance treaty with individual retention functions h_i, and a reinsurance treaty with a global retention function h.

We remark that the functional $\psi=1_{\{\sum_{i=1}^d x_i\geq s\}}$, which defines the worst-VaR problem (6), is not supermodular and hence does not satisfy the assumption of Theorem 2.3. Hence, it may happen that a comonotonic coupling does *not* maximize the VaR of the sum of d risks, as we will study in details in Section 3 below.

Open Problems

For $d=2$, the infimum in (13) is attained by the *countermonotonic* distribution $W(F_1,F_2)$. Since $W(F_1,\dots,F_d)$ is not a proper distribution when $d>2$, the search for the infimum of $\mathbb{E}[\psi(\mathbf{X})]$ among the Fréchet class $\mathfrak{F}(F_1,\dots,F_d)$ remains open for a variety of functionals ψ. Especially for $\psi=\times$, Roger Nelsen (private communication) remarked that the solution of this last mentioned problem would have important consequences in the theory of dependence measures.

3 The calculation of the distribution of the sum of risks

In the trivial system of marginals $\xi=\xi_I$, we have that $\mathfrak{F}_{\xi_I}=\{F_{\mathbf{X}}\}$. This setting represents complete probabilistic information about the portfolio $\mathbf{X}$ of risks held. In

fact, from a theoretical point of view, the knowledge of $F_{\mathbf{X}}$ completely determines the distribution of the random variable $\Psi(\mathbf{X})$. In practice, we will see that things are more complicated.

The system ξ_I is particularly important in stress-testing, i.e. when one has different models for $F_{\mathbf{X}}$ and wants to stress-test the distribution of $\Psi(\mathbf{X})$. Especially in the context of the current (credit) crisis, financial institutions often have information on the marginal distributions of the underlying risks but want to stress-test the interdependence between these risks, for instance assuming different copula scenarios.

In the following, we will study the case of the sum of risks, i.e. $\Psi = +$. Thus, we will focus on the computation of the distribution of $\Psi(\mathbf{X}) = \sum_{i=1}^{d} X_i$, i.e.

$$\mathbb{P}[X_1 + \cdots + X_d \leq s] = \int_{\mathfrak{I}(s)} dF_{\mathbf{X}}(x_1, \ldots, x_d), s \in \mathbb{R} \tag{14}$$

where $\mathfrak{I}(s) = \{\mathbf{x} \in \mathbb{R}^d : \sum_{i=1}^{d} x_i \leq s\}$.

The computation of (14) is a rather onerous task. In the literature, there exist several methods to calculate (14) when the marginals X_i are independent. In some rare cases, it is possible to write the integral in (14) in closed form. For general marginals, one can for instance rely on the Fast Fourier Transforms; see [8] and the references therein for a discussion within a risk management context.

Much less is known when the X_i's are dependent. Indeed, when $\mathbf{X}$ has a general copula $C_{\mathbf{X}}$, one often has to rely on integration tools like Monte Carlo and Quasi-Monte Carlo methods. When $F_{\mathbf{X}}$ has a density function $f_{\mathbf{X}}$, these methods approximate (14) by the average of $f_{\mathbf{X}}$ evaluated at M points $x_1, \ldots, x_M$ filling up $\mathfrak{I}(s)$ in a convenient way, i.e.

$$\int_{\mathfrak{I}(s)} dF_{\mathbf{X}}(x_1, \ldots, x_d) \simeq \frac{1}{M} \sum_{i=1}^{M} f_{\mathbf{X}}(x_i).$$

If the x_i's are chosen to be (pseudo) randomly distributed, this is the *Monte Carlo* (MC) method. If the x_i's are chosen as elements of a low-discrepancy sequence, this is the *Quasi-Monte Carlo* (QMC) method. A *low-discrepancy* sequence is a totally deterministic sequence of vectors that generates representative samples from a uniform distribution on given subsets. Compared to Monte Carlo methods, the advantage of using quasi-random sequences is that points cannot cluster coincidentally on some region of the set. Using Central Limit Theorem arguments, it is possible to show that traditional MC has a convergence rate of $O(M^{-1/2})$, and this independently of the number of dimensions d. QMC can be much faster than MC with errors approaching $O(M^{-1})$ for a smooth underlying density. For details on the theory of *rare event simulation* within MC methods, we refer the reader to the monographs [3], [18] and [27]. For an introduction to QMC methods, see for instance [32]. A comprehensive overview of both methods is given in [42]. Note that all the techniques mentioned above warrant considerable expertise and, more importantly, need to be tailored to the specific problem under study. In particular, the

implementation very much depends on the functional form of $f_{\mathbf{X}}$ (either direct, or through the marginals and a copula).

The re-tailoring of the rule to be iterated, from example to example, is common also to other numerical techniques for the estimation of (14), such as *quadrature methods*; see [6] and [33] for a review. However, in the computation of multidimensional integrals as in (14), numerical quadrature rules are typically less efficient than MC and QMC since they must be applied iteratively.

A simple and competitive tool for the computation of the distribution function of a sum of random variables is the AEP algorithm introduced in [1]. If one knows the distribution $F_{\mathbf{X}}$ of $\mathbf{X}$, it is very easy to compute the $F_{\mathbf{X}}$-measure of hypercubes in $\mathbb{R}^d$. Thus, the authors of [1] propose a decomposition of $\mathfrak{I}(s)$ via a infinite union of (possibly overlapping) hypercubes and hence compute (14) in terms of the algebraic sum of the probability masses contained in them.

In the MC and QMC methods described above, the final estimates contain a source of randomness. Instead, the AEP algorithm is completely determinist because it is solely based on the geometrical properties of $\mathfrak{I}(s)$. Moreover, the accuracy of MC and QMC methods is generally lost for problems in which the density $f_{\mathbf{X}}$ is not smooth or cannot be given in closed form, and comes at the price of an adaptation of the sampling algorithm to the specific example under study. The AEP algorithm however can handle in a uniform way any joint distribution $F_{\mathbf{X}}$ and does not require existence or smoothness of a density $f_{\mathbf{X}}$. As illustrated in [1], AEP performs better than QMC in dimensions $d = 2,3$ and slightly worse for dimensions $d = 4,5$. In these latter dimensions, however, programming a QMC sequence is much more demanding than using AEP. At the time being, AEP cannot be applied for $d > 5$ due to computational complexity (memory).

We set $d = 3$ and we use AEP to provide estimates for the tail and the quantile (VaR) function of the sum $S_3 = X_1 + X_2 + X_3$. For pedagogical reasons, we assume the marginals F_i of the portfolio to be Pareto distributed with tail parameter $\theta_i > 0$. We consider the two dependence scenarios obtained by coupling the Pareto marginals either by the independent copula $C_{\mathbf{X}} = \Pi$ or via the comonotonic copula $C_{\mathbf{X}} = M$. In the following, we use the fact that VaR is additive under comonotonicity; see Prop. 3.1 in [9]. This means that, for a comonotonic vector (X_1^M, X_2^M, X_3^M), we have

$$\mathrm{VaR}_\alpha(X_1^M + X_2^M + X_3^M) = \mathrm{VaR}_\alpha(X_1) + \mathrm{VaR}_\alpha(X_2) + \mathrm{VaR}_\alpha(X_3). \tag{15}$$

Denote by F_Π the distribution of S_3 obtained under independence between the X_i's and by F_M the distribution of S_3 obtained under comonotonicity between the X_i's. $\overline{F_\Pi}$ and $\overline{F_M}$, respectively, are the corresponding tail functions. We study two different cases: when the X_i's have finite or infinite first moment.

The finite-mean case. In Figure 2 (left), we plot $\overline{F_\Pi}$ and $\overline{F_M}$ when the Pareto tail parameter θ for the marginal distributions is set to 2 (the X_i's have finite first moment). We note that the two curves $\overline{F_\Pi}$ and $\overline{F_M}$ cross once at some high threshold $s = \hat{s}$. For $s < \hat{s}$, we have that $\overline{F_M}(s) < \overline{F_\Pi}(s)$. Recalling (5), this means that

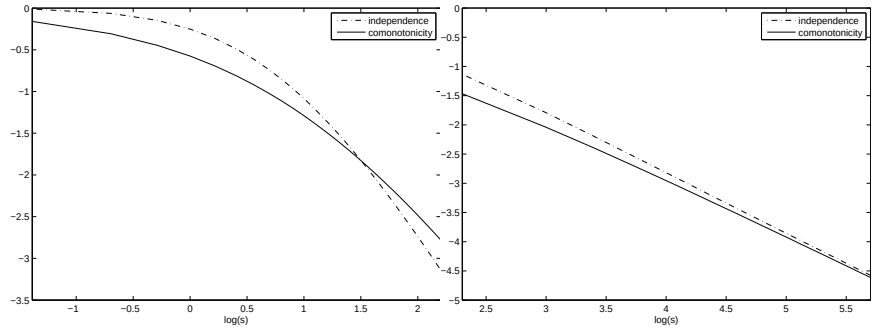

Fig. 2 Log/log plots of the tail function of $X_1+X_2+X_3$, under independence and comonotonicity. The X_i's are distributed as a Pareto(2) (left) and as a Pareto(1) (right).

$$F_M^{-1}(\alpha) < F_{\Pi}^{-1}(\alpha), \text{ for all } \alpha < \hat{\alpha} = F_M(\hat{s}) = F_{\Pi}(\hat{s}), \tag{16}$$

i.e. for the lower levels $\alpha < \hat{\alpha}$, $\mathrm{VaR}_\alpha(S_3)$ is larger under independence between the X_i's. For $\alpha > \hat{\alpha}$, inequality (16) is obviously reversed and we have that

$$F_{\Pi}^{-1}(\alpha) \leq F_M^{-1}(\alpha), \text{ for all } \alpha \geq \hat{\alpha}, \tag{17}$$

i.e. for the higher levels $\alpha > \hat{\alpha}$, $\mathrm{VaR}_\alpha(S_3)$ is larger under comonotonicity between the marginals. Recalling (15), and for the independent vector $(X_1^{\Pi}, X_2^{\Pi}, X_3^{\Pi})$, inequality (17) can be written as

$$\mathrm{VaR}_\alpha(X_1^{\Pi}+X_2^{\Pi}+X_3^{\Pi}) \leq \mathrm{VaR}_\alpha(X_1)+\mathrm{VaR}_\alpha(X_2)+\mathrm{VaR}_\alpha(X_3), \text{ for all } \alpha \geq \hat{\alpha}, \tag{18}$$

i.e. VaR is subadditive in the tail of F_{Π}. When $\theta > 1$, [19] illustrates that this tail behavior can be extended to more general dependence and marginal scenarios.

The infinite-mean case. Figure 2 (right) shows the same plot as Figure 2 (left), but now the Pareto tail parameter $\theta = 1$ (the X_i's have infinite first moment). We note that $\overline{F_M}(s) < \overline{F_{\Pi}}(s)$ for all $s \in \mathbb{R}$. Therefore, all the quantiles of S_3 under independence are larger than the corresponding quantiles under comonotonicity and inequality (18) is reversed:

$$\mathrm{VaR}_\alpha(X_1^{\Pi}+X_2^{\Pi}+X_3^{\Pi}) > \mathrm{VaR}_\alpha(X_1)+\mathrm{VaR}_\alpha(X_2)+\mathrm{VaR}_\alpha(X_3), \text{ for all } \alpha \in (0,1). \tag{19}$$

This shows that, in general, VaR *may fail to be subadditive*. Typical frameworks in which VaR shows a superadditive behavior are: marginals with infinite mean or skew distributions (as in this case) and/or marginals coupled by a non-elliptical copula; see [28]. An early interesting read on this is [11]. In [10], a mathematical summary of the issue is given within extreme value theory using the concept of multivariate regular variation.

Possible superadditivity is an important conceptual deficiency of Value-at-Risk. In fact, VaR has been heavily criticized by many authors for not being a *coherent*

measure of risk; see the seminal paper [2]. Many other authors have discussed desirable properties which a general risk measure ρ has to satisfy. Textbook treatments are [28] and [17].

	$X_i \backsim$ Pareto(2)			$X_i \backsim$ Pareto(1.3)			$X_i \backsim$ Pareto(1)		
α	Π	Cl	M	Π	Cl	M	Π	Cl	M
0.80	3.92	**4.21**	3.71	8.90	**9.36**	7.35	16.69	**17.21**	12.00
0.90	5.87	6.45	**6.49**	15.36	**16.54**	14.63	33.20	**35.05**	27.00
0.99	18.37	19.62	**27.00**	84.08	87.34	**100.65**	308.21	**315.25**	297.00
0.999	55.92	57.37	**91.87**	477.44	481.80	**606.28**	3012.97	**3025.00**	2997.00

Table 3 $\mathrm{VaR}_\alpha(X_1+X_2+X_3)$ under different dependence scenarios for three different Pareto portfolios. For a fixed level α and Pareto parameter θ, the largest VaR value is bold-faced.

Finally, in Table 3, we show the quantiles for S_3 for different levels of probabilities, under several marginal and dependence scenarios. Along with independence ($C_{\mathbf{X}} = \Pi$) and comonotonicity ($C_{\mathbf{X}} = M$), we study the case in which the copula of $\mathbf{X}$ is of Clayton type ($C_{\mathbf{X}} = Cl$). There are various points to remark about:

- The behavior of the tail function $\overline{F_{Cl}}$ of S_3 under the Clayton scenario is similar to the behavior of $\overline{F_{\Pi}}$ studied above. When $\theta > 1$, $\overline{F_{Cl}}$ and $\overline{F_M}$ cross once. This can be seen from the fact that, for $\theta = 2$ and $\theta = 1.3$, the comonotonic quantiles are smaller than the Clayton ones when the quantile level α is small, while they are larger when α is large. In this case, VaR under the Clayton scenario shows subadditivity in the tail.
  For $\theta = 1$, we have that $\overline{F_M}(s) \leq \overline{F_{Cl}}(s)$ for all $s \in \mathbb{R}$, hence VaR under the Clayton model is superadditive at all levels α. We also note that the intersection point between the Clayton and the comonotonic curve goes to infinity as the tail parameter θ approaches 1 from above. When $\theta = 1$, the two curves do not cross.
- Since the marginal distributions of the X_i's are fixed, the first moment of the sum S_3 does not depend on the copula $C_{\mathbf{X}}$. When $\theta > 1$, two different distributions for S_3 have the same *finite* mean and therefore *cannot* be *stochastically ordered*; see Sect. 1.2 in [30] for the definition of stochastic order and its properties. As a consequence, two different distributions for S_3 must cross. The case illustrated in Figure 2, in which the intersection point is unique, is typical for two random variables which are *stop-loss ordered*; see Th. 1.5.17 and Def. 1.5.1 in [30] (in this last reference the authors use the equivalent terminology *increasing-convex order* to indicate the stop-loss order).
  When $\theta = 1$, we have that $\mathbb{E}[S_3] = +\infty$ and it is possible that $\overline{F_M} < \overline{F_{\Pi}}$, i.e. the distribution of S_3 under independence is stochastically larger than the distribution of S_3 under comonotonicity, as illustrated in Figure 2 (right). For general distributions, both the change of behavior with respect to stochastic dominance and superadditivity of VaR in the tail seem to be strictly related to the existence of first moments. For some further discussions on this phenomenon, see [21], [31] and [20].

- For Pareto marginals of the form (11), the quantile function of S_3 can be given in closed form under the independence and comonotonic assumptions. Things are different when one assumes a Clayton-type dependence. In this latter case, the computation of the distribution and the VaRs of S_3 requires one of the integration techniques described above in this section. In particular, the quantiles in Table 3 have been obtained via AEP.

Open problems

In insurance and finance, there is a increasing need of software being able to compute the distribution of $\Psi(\mathbf{X})$ when the distribution of $\mathbf{X}$ is known. The authors of [1] are working on a extension of AEP to general increasing functionals Ψ. Moreover, efficiency of AEP for dimensions $d > 5$ needs to be improved. Finally, AEP and its competitors open the way to the computational study of large and non-homogeneous risk portfolios.

Acknowledgements The first author would like to thank the Swiss Finance Institute for financial support. The second author thanks both RiskLab and the FIM at the Department of Mathematics of the ETH Zurich for financial support and kind hospitality.

References

1. Arbenz, P., Embrechts, P., Puccetti, G.: The AEP algorithm for the fast computation of the distribution of the sum of dependent random variables (2009). Preprint, ETH Zurich
2. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Math. Finance **9**(3), 203–228 (1999)
3. Asmussen, S., Glynn, P.W.: Stochastic Simulation: Algorithms and Analysis, vol. 57. Springer, New York (2007)
4. Basel Committee on Banking Supervision: International Convergence of Capital Measurement and Capital Standards. Bank for International Settlements, Basel (2006)
5. Cambanis, S., Simons, G., Stout, W.: Inequalities for $Ek(X,Y)$ when the marginals are fixed. Z. Wahrsch. Verw. Gebiete **36**(4), 285–294 (1976)
6. Davis, P.J., Rabinowitz, P.: Methods of Numerical Integration. Academic Press, Orlando, FL (1984). Second edition
7. Denuit, M., Genest, C., Marceau, É.: Stochastic bounds on sums of dependent risks. Insurance Math. Econom. **25**(1), 85–104 (1999)
8. Embrechts, P., Frei, M.: Panjer recursion versus FFT for compound distributions. Math. Meth. Oper. Res. **69**, 497–508 (2009)
9. Embrechts, P., Höing, A., Juri, A.: Using copulae to bound the Value-at-Risk for functions of dependent risks. Finance Stoch. **7**(2), 145–167 (2003)
10. Embrechts, P., Lambrigger, D.D., Wütrich, M.V.: Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. Extremes **12**, 107–127 (2009)
11. Embrechts, P., McNeil, A.J., Straumann, D.: Correlation and dependence in risk management: properties and pitfalls. In: M. Dempster (ed.) Risk Management: Value at Risk and Beyond, pp. 176–223. Cambridge Univ. Press, Cambridge (2002)
12. Embrechts, P., Puccetti, G.: Aggregating risk capital, with an application to operational risk. Geneva Risk Insur. Rev. **31**(2), 71–90 (2006)

13. Embrechts, P., Puccetti, G.: Bounds for functions of dependent risks. Finance Stoch. **10**(3), 341–352 (2006)
14. Embrechts, P., Puccetti, G.: Bounds for functions of multivariate risks. J. Mult. Analysis **97**(2), 526–547 (2006)
15. Embrechts, P., Puccetti, G.: Aggregating operational risk across matrix structured loss data. J. Operational Risk **3**(2), 29–44 (2008)
16. Embrechts, P., Puccetti, G.: Bounds for the sum of dependent risks having overlapping marginals (2009). Forthcoming in J. Multivariate Anal.
17. Föllmer, H., Schied, A.: Stochastic Finance. Walter de Gruyter, Berlin (2004). Second edition
18. Glasserman, P.: Monte Carlo Methods in Financial Engineering. Springer-Verlag, New York (2004)
19. Ibragimov, R.: Portfolio diversification and value at risk under thick-tailedness. Quantitative Finance **9**(5), 565–580 (2009)
20. Ibragimov, R., Walden, J.: Portfolio diversification under local and moderate deviations from power laws. Insurance Math. Econom. **42**(2), 594–599 (2008)
21. Jaworski, P.: Value at risk in the presence of the power laws. Acta Phys. Polon. B **36**(8), 2575–2587 (2005)
22. Jaworski, P.: Bounds for value at risk—the approach based on copulas with homogeneous tails. Mathware Soft Comput. **15**(1), 113–124 (2008)
23. Lindvall, T.: Lectures on the Coupling Method. Wiley, New York (1992)
24. Lorentz, G.G.: An inequality for rearrangements. Amer. Math. Monthly **60**, 176–179 (1953)
25. Makarov, G.D.: Estimates for the distribution function of the sum of two random variables with given marginal distributions. Theory Probab. Appl. **26**, 803–806 (1981)
26. Marshall, A.W., Olkin, I.: Inequalities: Theory of Majorization and its Applications. Academic Press, New York (1979)
27. McLeish, D.L.: Monte Carlo Simulation and Finance. John Wiley & Sons, Hoboken, NJ (2005)
28. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techiniques, Tools. Princeton University Press, Princeton, NJ (2005)
29. Müller, A.: Stop-loss order for portfolios of dependent risks. Insurance Math. Econom. **21**(3), 219–223 (1997)
30. Müller, A., Stoyan, D.: Comparison Methods for Stochastic Models and Risks. Wiley, Chichester (2002)
31. Nešlehová, J., Embrechts, P., Chavez-Demoulin, V.: Infinite-mean models and the LDA for operational risk. J. Operational Risk **1**(1), 3–25 (2006)
32. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods, *CBMS-NSF Regional Conference Series in Applied Mathematics*, vol. 63. SIAM, Philadelphia (1992)
33. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes: the Art of Scientific Computing. Cambridge University Press, Cambridge (2007). Third Edition
34. Puccetti, G., Scarsini, M.: Multivariate comonotonicity (2009). Forthcoming in J. Multivariate Anal.
35. Rachev, S.T., Rüschendorf, L.: Mass Transportation Problems. Vol. I-II. Springer-Verlag, New York (1998)
36. Rüschendorf, L.: Random variables with maximum sums. Adv. in Appl. Probab. **14**(3), 623–632 (1982)
37. Rüschendorf, L.: Bounds for distributions with multivariate marginals. In: Stochastic Orders and Decision under Risk, vol. 19, pp. 285–310. Inst. Math. Statist., Hayward, CA (1991)
38. Rüschendorf, L.: Fréchet-bounds and their applications. In: Advances in Probability Distributions with Given Marginals, vol. 67, pp. 151–187. Kluwer Acad. Publ., Dordrecht (1991)
39. Shortt, R.M.: Strassen's marginal problem in two or more dimensions. Z. Wahrsch. Verw. Gebiete **64**(3), 313–325 (1983)
40. Strassen, V.: The existence of probability measures with given marginals. Ann. Math. Statist **36**, 423–439 (1965)
41. Vorob'ev, N.N.: Consistent families of measures and their extensions. Theory of Probability and its Applications **7**(2), 147–163 (1962)
42. Weinzierl, S.: Introduction to Monte Carlo Methods (2000). Eprint arXiv:hep-ph/0006269