General Convex Order on Risk Aggregation

Edgars Jakobsons[*], Xiaoying Han[†] and Ruodu Wang[‡]

January 15, 2015

Abstract

Using a general notion of convex order, we derive general lower bounds for risk measures of aggregated positions under dependence uncertainty, and this in arbitrary dimensions and for heterogeneous models. We also prove sharpness of the bounds obtained when each marginal distribution has a decreasing density. The main result answers a long-standing open question and yields an insight in optimal dependence structures. A numerical algorithm provides bounds for quantities of interest in risk management. Furthermore, our numerical results suggest that the bounds obtained in this paper are generally sharp for a broader class of models.

Key-words: risk aggregation; dependence uncertainty; convex order; risk measures; heterogeneous models; joint mixability.

[*]RiskLab, Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland.
Email: edgars.jakobsons@math.ethz.ch

[†]Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849, USA.
Email: xzh0003@auburn.edu

[‡]Corresponding author. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada. Email: wang@uwaterloo.ca

1 Introduction

In quantitative risk management, under the term *risk aggregation* one discusses the statistical behavior of an aggregate position $S(\mathbf{X})$ associated with a risk vector $\mathbf{X} = (X_1, \ldots, X_n)$, where $X_1, \ldots, X_n$ are random variables representing one-period individual risks. The most commonly studied aggregate risk position is the sum $S = X_1 + \cdots + X_n$; it plays an important role in both insurance and finance.

In the quantification of risk aggregation, *model uncertainty* has received much attention recently, especially after the financial crisis of 2008; see discussions in BCBS (2013b). An insufficient understanding of model uncertainty (or manipulation) and its quantitative consequences may lead to wrong conclusions, undermining the efficiency of risk management (for a case study, see US Senate, 2013, Chapter V).

One of the more challenging types of uncertainty in modern risk management relates to *dependence uncertainty*. In practice, marginal distributions are easier to analyze with statistical tools, while multivariate dependence is much more difficult to quantify. We refer to Embrechts et al. (2014) for a comprehensive review on the growing literature on this topic and its impact on the recent framework for banking regulation, as for instance discussed in the Basel documents (BCBS, 2012, 2013a). In insurance regulation, discussions on uncertainty in risk management also take place in Solvency 2 and the Swiss Solvency Test; see for instance Sandström (2010) and SCOR (2008).

To address questions of risk aggregation with dependence uncertainty, Bernard et al. (2014a) introduced the *admissible risk class* as the collection of all possible aggregate risks with fixed, known marginal distributions. A practical illustration of this setup is for instance to be found in the Loss Distribution Approach to Operational Risk; see Embrechts et al. (2013). Though, of course, in this case, there is considerable uncertainty at the level of the estimation of the marginal distributions. If one has additional information on the underlying dependence structure, subsets of the admissible risk class can be used to describe the possible aggregate risks; for the case of information on variance, see Cheung and Vanduffel (2013) and Bernard et al. (2014b).

A main tool in our analysis is *convex ordering*, which in our context is equivalent with *second order stochastic dominance* or *stop-loss ordering*. These concepts from the realm of decision making under uncertainty are consistent with risk-averse risk measurement, especially in the by now classical context of coherent risk measures and its utility-based formulation; see for instance Artzner et al. (1999) and Delbaen (2012). In the context of finding upper and lower bounds for quantities related to risk aggregation, convexity arguments play an important role. In the sequel of this paper *convex bounds* will refer to any bound obtained using arguments based on the concept of convex ordering. It is well-known that the sharp upper convex bound on any admissible risk class is obtained under the comonotonic dependence structure, whereas for the lower convex bound for $n \geqslant 3$ no general solution is known in the literature; see for instance Tchen (1980) and Dhaene et al. (2002). Rüschendorf and Uckelmann (2002) and Section 8.3.1 of Müller and Stoyan (2002) studied special cases of lower convex bounds on risk aggregation for uniform, symmetric and some discrete marginal distributions, and Wang and Wang (2011) studied the case when marginal distributions are identical and have a monotone density. A numerical algorithm (the Rearrangement Algorithm (RA)) for the approximation of the dependence structure leading to a lower convex bound is given in Embrechts et al. (2013). The latter paper contains a general lower convex bound in the homogeneous case, i.e. when all marginal distributions are identical. Furthermore, under some extra conditions, sharpness of this bound is proved; see also Bernard et al. (2014a).

In this paper, we generalize the results of Bernard et al. (2014a) and study a lower convex bound for non-identical marginal distributions. This generalization is particularly important in the practice of financial and insurance risk management, where identical marginal distributions are clearly unrealistic; see Embrechts et al. (2013). It turns out that the problem of finding convex bounds with heterogeneous marginal distributions is considerably more challenging. The new lower convex bound obtained in this paper is based on a new technique of dynamically weighting marginal distributions by finding solutions to related functional equations. The dependence structure that leads to the lower convex bound can be interpreted as a combination of joint mixability, introduced in Wang et al. (2013), and mutual exclusivity, introduced in Dhaene and Denuit (1999) (earlier mathematical results can be found in Dall'Aglio (1972)). We show that this new bound is sharp if each of the marginal distributions has a decreasing density on its support. Numerical results show that the new bounds outperform almost all other results in the literature, and this in great generality.

We remark that although our results work for a broad class of models, including all models with decreasing densities, a universal solution for a lower convex bound is still out of reach at this moment, even for homogeneous models. A full characterization of this lower convex bound for arbitrary distributions would require further research on joint mixability and other negative dependence concepts, a rapidly expanding field of research. A recent review on extremal dependence concepts can be found in Puccetti and Wang (2014).

The rest of the paper is organized as follows. In Section 2 we summarize some preliminaries on admissible risks, complete and joint mixability, and convex order. Section 3 provides a new lower convex bound on risk aggregation for heterogeneous marginal distributions. This bound is shown to be sharp under a monotonicity condition and a condition of joint mixability; in particular, these conditions are satisfied if the marginal distributions have decreasing densities. Insurance and financial applications are then discussed in Section 4. Numerical illustrations and an algorithm are given in Section 5, highlighting the advantages of our results compared to other numerical methods available in the literature. We conclude in Section 6.

2 Preliminaries

2.1 Admissible risk

In this paper, we assume that all random variables are defined on an atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Similar to Bernard et al. (2014a), we call an *aggregate risk* the sum $S = X_1 + \cdots + X_n$ where X_i are non-negative random variables and n is a positive integer. Note that the non-negativity is assumed just for notational convenience, and for our results is equivalent to the assumption that $X_1, \ldots, X_n$ are bounded below (since convex order is invariant under translation $X_i \mapsto a_i + X_i$, $a_i \in \mathbb{R}$).

In this paper we consider the case where for each $i = 1, \ldots, n$ the distribution of X_i is known, while the joint distribution of $\mathbf{X} := (X_1, X_2, \ldots, X_n)$ is unknown. We use the notation $X \sim F$ to indicate that $X \in L^0(\Omega, \mathcal{A}, \mathbb{P})$ has distribution function (df) F.

Definition 2.1 (Admissible risk)**.** An aggregate risk S is called an *admissible risk* of marginal distributions $F_1, \ldots, F_n$ if it can be written as $S = X_1 + \cdots + X_n$ where $X_i \sim F_i$ for $i = 1, \ldots, n$. The *admissible risk class* is defined by the set of admissible risks of given marginal distributions:

$$\mathfrak{S}_n(F_1, \ldots, F_n) = \{X_1 + \cdots + X_n : X_i \sim F_i,\ i = 1, \ldots, n\}.$$

The definition of admissible risks only concerns the distributions of random variables, thus there is a one-to-one relationship between $\mathfrak{S}_n(F_1, \dots, F_n)$ and the *admissible distribution class* defined as

$$\mathfrak{D}_n(F_1, \dots, F_n) = \{\text{distribution of } S : S \in \mathfrak{S}_n(F_1, \dots, F_n)\}.$$

Properties of the admissible risk class $\mathfrak{S}_n(F_1, \dots, F_n)$ were given in Bernard et al. (2014a). A full characterization of $\mathfrak{S}_n(F_1, \dots, F_n)$ is challenging and seems far beyond the reach of current methodology. The admissible risk class identifies what risks are possible when the marginal distributions are known. When all risks have the same distribution, i.e. $F_1 = \cdots = F_n$, we say that the risks are *homogeneous*. When the distributions F_i are allowed to be different, we say that the risks are *heterogeneous*. For simplicity, we denote $\mathbf{F} = (F_1, \dots, F_n)$.

As mentioned in the introduction, the study of the admissible risk class $\mathfrak{S}_n(\mathbf{F})$ is of great interest in risk management and this topic has a long history. One of the most important issues is to quantify aggregate risks under extreme dependence structures. Note that all admissible risks of given marginal distributions $(F_1, \dots, F_n)$ have the same mean if it exists for each F_i. It is thus natural to consider variability in the class. In this paper, we measure variability using convex order and focus on extreme aggregate risks in $\mathfrak{S}_n(\mathbf{F})$ in the sense of convex order.

2.2 Complete and joint mixability

Distributions $F_1, \dots, F_n$ are jointly mixable (JM) (Wang et al., 2013) if there exist $X_i \sim F_i$, $i = 1, \dots, n$ such that $X_1 + \cdots + X_n$ is (almost surely) a constant; such $(X_1, \dots, X_n)$ is called a *joint mix*. Here, we give an equivalent definition using admissible risks.

Definition 2.2 (Joint Mixability and Complete Mixability)**.**

(i) Univariate distributions $F_1, \dots, F_n$ are jointly mixable (JM) if the admissible risk class $\mathfrak{S}_n(F_1, \dots, F_n)$ contains a constant.

(ii) A univariate distribution F is n−completely mixable (n-CM) if the admissible risk class $\mathfrak{S}_n(F, \dots, F)$ contains a constant.

We also that say F is n-CM on an interval I if the conditional distribution of F on I is n-CM, and $F_1, \dots, F_n$ are JM on a hypercube $\prod_{i=1}^n I_i$, if the conditional distributions of $F_1, \dots, F_n$ on intervals $I_1, \dots, I_n$, respectively, are JM.

Some examples and recent theoretical results of CM distributions and JM distributions can be found in Wang and Wang (2011, 2014) and Puccetti et al. (2012, 2013). Complete mixability turns out to be crucial for finding lower convex bounds for homogeneous risks; for a detailed discussion, see Bernard et al. (2014a). In this paper, joint mixability will be used to obtain sharp lower bounds for heterogeneous risks.

2.3 Convex order and existing results

Convex order describes a preference between two random variables, agreed upon by all risk-avoiding investors. Let us recall the definition.

Definition 2.3 (Convex order)**.** Let X and Y be two random variables with finite means. Then X is smaller than Y *in convex order*, denoted by $X \prec_{\text{cx}} Y$, if for all convex functions f,

$$\mathbb{E}[f(X)] \leqslant \mathbb{E}[f(Y)],$$

provided that the two expectations exist.

It is immediate that $X \prec_{\text{cx}} Y$ implies $\mathbb{E}[X] = \mathbb{E}[Y]$. In the following, we assume that $F_1, \ldots, F_n$ have finite means. Convex order is well-adapted to $\mathfrak{S}_n(\mathbf{F})$ as all variables in $\mathfrak{S}_n(\mathbf{F})$ have the same mean. Note that convex order is an order determined by distributions only, hence we do not really need to specify random variables in our discussion. Convex order on aggregate risks has been extensively studied in actuarial science since it is equivalent to stop-loss order (given that the means of two risks are the same), an important concept in insurance premium calculations. More discussions on stochastic orders on aggregate risks can be found in Müller (1997a,b). We say that T is an *upper (resp. lower) convex bound* of $\mathfrak{S}_n(F_1, \ldots, F_n)$ if $T \succ_{\text{cx}} S$ (resp. $T \prec_{\text{cx}} S$) for all $S \in \mathfrak{S}_n(F_1, \ldots, F_n)$. From now on, our objective is to find convex bounds of the set $\mathfrak{S}_n(F_1, \ldots, F_n)$.

We denote by $G^{-1}(t) = \inf\{x : G(x) \geqslant t\}$ for $t \in (0, 1]$ the generalized inverse function for any monotone function $G : \mathbb{R}^+ \to [0, 1]$, and in addition let $G^{-1}(0) = \inf\{x : G(x) > 0\}$ throughout the paper. A well-known result is that the sharp upper convex bound in $\mathfrak{S}_n(F_1, \ldots, F_n)$ is $F_1^{-1}(U) + \cdots + F_n^{-1}(U)$ where U is uniformly distributed on the interval $(0, 1)$, denoted as $U \sim \mathcal{U}(0, 1)$. The dependence structure of $\mathbf{X} = (F_1^{-1}(U), \ldots, F_n^{-1}(U))$ is called the *comonotonic* scenario. In particular, one has

$$S_n \prec_{\text{cx}} F_1^{-1}(U) + \cdots + F_n^{-1}(U) \text{ for any } S_n \in \mathfrak{S}_n(\mathbf{F}). \tag{2.1}$$

Note that sharpness immediately follows, since for $i = 1, \ldots, n$, $F_i^{-1}(U) \stackrel{\mathrm{d}}{=} X_i$ and hence the above upper bound belongs to $\mathfrak{S}_n(\mathbf{F})$. We refer to Dhaene et al. (2002) for more details on comonotonicity and Deelstra et al. (2011) for a recent review on the applications of comonotonicity in finance and insurance.

The rest of the paper focuses on the much more complex issue of determining the lower convex bound of $\mathfrak{S}_n(\mathbf{F})$. When there are only two variables, $n = 2$, the minimum is obtained by the *counter-monotonic* scenario:

$$F_1^{-1}(U) + F_2^{-1}(1-U) \prec_{\mathrm{cx}} S_2 \text{ for any } S_2 \in \mathfrak{S}_2(F_1, F_2),$$

where $U \sim \mathcal{U}(0,1)$; and this bound is again sharp. The above two convex order results, in the larger class of supermodular functions, date back to W. Hoeffding in the 40s; see Tchen (1980). However, the sharp lower convex bound for $n \geqslant 3$ is missing in the literature due to the fact that counter-monotonicity cannot be generalized to $n \geqslant 3$ without losing its minimality with respect to convex order. In the case when marginal distributions are identical with a monotone density function, the sharp lower bound for general n is obtained in Wang and Wang (2011), together with results on complete mixability. In another special case, when $F_1, \ldots, F_n$ are on $\mathbb{R}^+$ with $\sum_{i=1}^n F_i(0) \geqslant n-1$, the convex minimum is obtained by the the *mutually exclusive* scenario:

$$Y_1 + \cdots + Y_n \prec_{\mathrm{cx}} S_n \text{ for any } S_n \in \mathfrak{S}_n(\mathbf{F}),$$

where $Y_i \sim F_i$ and $\mathbb{P}(Y_i > 0, Y_j > 0) = 0$, $i, j = 1, \ldots, n$, $i \neq j$, i.e. only one random variable can be positive at the same time; see Dhaene and Denuit (1999). However, this assumption means that the distributions $F_1, \ldots, F_n$ have atoms at zero with a very large total mass, and hence it is rather restrictive. Another observation, also restrictive, is that if $F_1, \ldots, F_n$ are JM, then a sharp convex lower bound is based on joint mixability: if $X_1 + \cdots + X_n$ is a constant, where $X_i \sim F_i$, $i = 1, \ldots, n$, then for $S_n \in \mathfrak{S}_n(\mathbf{F})$,

$$X_1 + \cdots + X_n = \mathbb{E}[S_n] \prec_{\mathrm{cx}} S_n \text{ for any } S_n \in \mathfrak{S}_n(\mathbf{F}).$$

However, joint mixability is theoretically difficult to prove, and many distributions are shown to be not JM. Limited results on JM are summarized in Wang et al. (2013) and the more recent Wang and Wang (2014).

One step further, Bernard et al. (2014a) studied more general lower convex bounds over the admissible risk class. Roughly speaking, their idea is to combine complete mixability and mutual exclusivity. We summarize their results below. Let n be a positive integer (although only $n \geqslant 3$ is of interest). Let F be the average of the marginal distributions, i.e.

$$F = \frac{1}{n}\sum_{i=1}^{n} F_i.$$

The following functions $H(x)$, $D(a)$ for $a, x \in [0, \frac{1}{n}]$ and the number c_n are defined in Bernard et al. (2014a):

$$\begin{aligned}
H(x) &= (n-1)F^{-1}((n-1)x) + F^{-1}(1-x), \\
D(a) &= \frac{n}{1-na}\int_{(n-1)a}^{1-a} F^{-1}(y)dy, \\
c_n &= \min\left\{c \in \left[0, \tfrac{1}{n}\right] : H(c) \leqslant D(c)\right\}, \\
T_a &= H(U/n)\mathrm{I}_{\{U\in[0,na]\}} + D(a)\mathrm{I}_{\{U\in(na,1]\}}.
\end{aligned}$$

Roughly speaking, H represents the sum in a nearly mutual exclusive scenario, where one large risk is coupled with $n-1$ small risks. D represents the sum in a scenario of joint mix, where the sum is exactly equal to its mean. The structure of T_a can be interpreted as a combination of H and D, and c_n is a threshold distinguishing the two scenarios.

For some $a \in [0, \frac{1}{n}]$, Bernard et al. (2014a) used the following assumptions (A), (A') and (B):

(A) $H(x)$ is non-increasing on $[0, a]$ and $\lim_{x\to a-} H(x) \geqslant D(a)$.

(A') $H(x)$ is non-increasing on the interval $[0, c_n]$.

(B) The distribution F is n-CM on the interval $I = [F^{-1}((n-1)c_n), F^{-1}(1-c_n)]$.

The assumption (A) is used to obtain a convex lower bound for $\mathfrak{S}_n(F_1, \ldots, F_n)$, and the assumptions (A') and (B) are used to obtain sharpness for the homogeneous model $\mathfrak{S}_n(F, \ldots, F)$. Note that (A) always holds trivially for $a = 0$, and (A') is stronger than (A). The main results in Bernard et al. (2014a) are summarized in the following theorem.

Theorem 2.1 (Bernard et al. (2014a))**.**

(i) Suppose (A) holds for some $a \in [0, \frac{1}{n}]$*, then* $T_a \prec_{\mathrm{cx}} S$ *for all* $S \in \mathfrak{S}_n(F_1, \ldots, F_n)$*.*

(ii) In the homogeneous case $F_1 = \cdots = F_n = F$*,* $T_{c_n} \in \mathfrak{S}_n(F_1, \ldots, F_n)$ *if (A') and (B) holds.*

Theoretically, one only obtains the sharpness of the above bounds for homogeneous risks. As pointed out and illustrated numerically in Bernard et al. (2014a), the sharpness of the bound obtained in Theorem 2.1 generally fails to hold for heterogeneous risks, in particular when the marginal distributions are significantly different. A sharp convex lower bound for heterogeneous risks seems out of reach by their methodology.

In this paper, we give a new lower convex bound for heterogeneous risks which is sharp under a monotonicity condition and a JM condition. Our result is based on a new technique involving a dynamical weighting of the marginal distributions.

3 Convex Bounds on Risk Aggregation

Throughout we suppose $F_1, \ldots, F_n$ are continuous distributions on $\mathbb{R}_+$ with finite means, and $n \geqslant 3$ is a positive integer. Without loss of generality, we can assume all distributions have left end-point at 0. Since convex order is invariant under shifting by constants, this is equivalent to assuming all of their supports are bounded from below. We denote by $\bar{F}_i$ the survival function of F_i, i.e. $\bar{F}_i = 1 - F_i$. In all our discussions, the terms *decreasing* and *increasing* are understood in the strict sense.

Our idea to construct an optimal structure is inspired by the arguments of Bernard et al. (2014a). In order to have a convexly small element, intuitively one wants the sum $S = X_1 + \cdots + X_n$ to be concentrated around its mean, e.g. a smaller variance is preferred by taking a quadratic f in Definition 2.3. Typically, for financial risks $\mathbb{P}(X_i > \mathbb{E}[S]) > 0$ for some i (large losses are possible). When $X_i > \mathbb{E}[S]$, S must be greater than its mean, so we would like all the other X_i to be as small as possible, so that the sum S is minimized in convex order. This motivates constructing a dependence structure where, when one of the X_i is large, the others are set to be small ("nearly mutually exclusive"); and when one of X_i is of medium size, all the others are also of medium size and the sum is concentrated around a constant ("nearly jointly mixable"). This idea is very similar to the construction in Bernard et al. (2014a) which originates from that in Wang and Wang (2011) and indeed forms the basis for the Rearrangement Algorithm in Embrechts et al. (2013). However, although the idea of complete mixability in Bernard et al. (2014a) can be naturally generalized to joint mixability for heterogeneous risks, the construction of the "nearly mutually exclusive scenario" for heterogeneous risks is unclear and cannot be easily generalized. It turns out that to construct such a dependence structure one needs to dynamically assign weights to each margin, while keeping the sum small in convex order. Below we provide a rigorous mathematical formulation for the above idea. At first this weighting may seem rather non-intuitive; further in the paper its importance will hopefully become clear.

3.1 Main results

First, we introduce the following functional equations (E1)-(E2): for $y, y_1, \ldots, y_n \colon (0,1) \to \mathbb{R}_+$, such that for $x \in (0,1)$,

(E1) $\sum_{i=1}^n \bar{F}_i(y_i(x)) = x$,

(E2) $F_i(y_i(x) - y(x)) + \bar{F}_i(y_i(x)) = x \quad$ for each $i = 1, \ldots, n$.

Equations (E1)-(E2) will be key to the rest of the paper. In the following, we shall continue our discussion assuming that (E1)-(E2) have at least one solution. A condition which guarantees the existence and uniqueness of such a solution is:

(F) Each F_i, $i = 1, \ldots, n$, has a decreasing density on its support $[0, \infty)$.

Condition (F) includes, for instance, Pareto and Exponential distributions. The main results in this paper do not require (F), but do require (E1)-(E2) to have a solution; the reader is suggested to keep (F) in mind for a primary example. Our numerical results show that (E1)-(E2) have a solution for a wider class of distributions relevant in practice.

Lemma 3.1. *The system of functional equations (E1)-(E2) has a unique solution if (F) holds.*

We put the rather tedious proofs of Lemma 3.1 and Lemma 3.2 below at the end of this section.

In the following, we assume $(y, y_1, \ldots, y_n)$ is a solution to (E1)-(E2), and define three functions which play a key role in this paper. For $X_i \sim F_i$, $i = 1, \ldots, n$,

$$h(x) = \sum_{i=1}^n y_i(x) - (n-1)y(x) \text{ for } x \in (0,1),$$

$$d(a) = \frac{1}{1-a} \sum_{i=1}^n \mathbb{E}\left[X_i \mathrm{I}_{\{y_i(a) - y(a) \leqslant X_i \leqslant y_i(a)\}}\right] \text{ for } a \in (0,1), \quad d(0) = \sum_{i=1}^n \mathbb{E}[X_i],$$

$$s_n = \inf\{s \in (0,1) : h(s) \leqslant d(s)\}.$$

The lemma below lists some useful properties of the solution $(y, y_1, \ldots, y_n)$.

Lemma 3.2. *Suppose $(y, y_1, \ldots, y_n)$ is a solution to (E1)-(E2), and (F) holds. Then on the interval $(0, s_n)$ for each $i = 1, \ldots, n$,*

(i) $0 < y < y_i$,

(ii) y *and* y_i *are decreasing,*

(iii) $y_i - y$ *is increasing,*

(iv) h *is decreasing.*

Remark 3.1. A numerical procedure (provided in Section 5.2) can be applied to find a solution to (E1)-(E2), when $F_1, \ldots, F_n$ have densities. Moreover, we shall also see that the monotonicity condition (F) is not a necessary one and it is possible to obtain solutions to (E1)-(E2) in more general cases. For instance, in the homogeneous model $F_1 = \cdots = F_n =: F$, one can easily check that $y_1(x) = \cdots = y_n(x) = F^{-1}(x/n)$, and $y(x) = F^{-1}(x/n) - F^{-1}((n-1)x/n)$ for $x \in (0, 1)$ give a solution to (E1)-(E2) which satisfies (i)-(iii) in Lemma 3.2 on (0,1). This does not require the assumption (F). Property (iv) is not guaranteed in general, but it is satisfied by a large class of distributions.

In all the following discussions and results, we throughout assume

(C) $F_1, \ldots, F_n$ are continuous distributions on $\mathbb{R}_+$ and (E1)-(E2) have a solution $(y, y_1, \ldots, y_n)$ which satisfies properties (i)-(iv) in Lemma 3.2.

From Lemmas 3.1 and 3.2 we have seen that (F) is sufficient for (C). Condition (C) can be easily verified numerically for given marginal distributions $F_1, \ldots, F_n$. Indeed, (C) is not restrictive; numerical illustrations suggest that it is satisfied by almost all distributions used in quantitative risk management. See Section 5 for a discussion. However, it is theoretically difficult to show that (C) is satisfied by general choices of distributions; even in the homogeneous case only numerical verification is available, as discussed in Bernard et al. (2014a). Note that we do not assume that the distributions $F_1, \ldots, F_n$ have unbounded support, nor do we assume the uniqueness of $(y, y_1, \ldots, y_n)$; we will simply need one solution to (E1)-(E2). Although the uniqueness of $(y, y_1, \ldots, y_n)$ is not guaranteed, we will see that h is unique on $(0, s_n)$ under some extra conditions to be formulated later.

Note that if (C) holds, then $\bar{F}_i(y_i(\cdot))$ is continuous, almost everywhere differentiable on $(0,1)$, and $0 < \mathrm{d}\bar{F}_i(y_i(x))/\mathrm{d}x < 1$ for each $i = 1, \ldots, n$; this can be seen from (E1) and the fact that y_i is decreasing. We will use this fact frequently in the subsequent proofs. We first provide some properties of the function $h(x)$.

Lemma 3.3. *For $a \in (0,1)$, we have*

$$\int_0^a h(u)\,\mathrm{d}u = \sum_{i=1}^n \mathbb{E}[X_i(\mathrm{I}_{\{X_i > y_i(a)\}} + \mathrm{I}_{\{X_i < y_i(a) - y(a)\}})]. \tag{3.1}$$

Proof. We have that

$$\begin{aligned}\mathbb{E}[X_i(\mathrm{I}_{\{X_i > y_i(a)\}} + \mathrm{I}_{\{X_i < y_i(a) - y(a)\}})] &= \int_{y_i(a)}^{\infty} x\mathrm{d}F_i(x) + \int_0^{y_i(a)-y(a)} x\mathrm{d}F_i(x) \\ &= \int_a^0 y_i(t)\mathrm{d}F_i(y_i(t)) + \int_0^a (y_i(t) - y(t))\,\mathrm{d}F_i(y_i(t) - y(t)).\end{aligned} \tag{3.2}$$

By (E2), it follows that

$$\begin{aligned}\int_0^a (y_i(t) - y(t))\,\mathrm{d}F_i(y_i(t) - y(t)) &= \int_0^a (y_i(t) - y(t))\,\mathrm{d}(t - 1 + F_i(y_i(t))) \\ &= \int_0^a (y_i(t) - y(t))\,\mathrm{d}t + \int_0^a y_i(t)\mathrm{d}F_i(y_i(t)) - \int_0^a y(t)\mathrm{d}F_i(y_i(t)).\end{aligned} \tag{3.3}$$

From (3.2)-(3.3) and (E1), we have that

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}[X_i(\mathrm{I}_{\{X_i > y_i(a)\}} + \mathrm{I}_{\{X_i < y_i(a) - y(a)\}})] &= \sum_{i=1}^n \int_0^a (y_i(t) - y(t))\,\mathrm{d}t - \sum_{i=1}^n \int_0^a y(t)\mathrm{d}F_i(y_i(t)) \\ &= \int_0^a \left(\sum_{i=1}^n y_i(t) - ny(t)\right)\mathrm{d}t + \int_0^a y(t)\mathrm{d}t \\ &= \int_0^a h(t)\mathrm{d}t.\end{aligned}$$

□

The function $h(x)$ plays a key role in the construction of a sharp lower convex bound in $\mathfrak{S}_n(F_1, \ldots, F_n)$. In order to see this, we first define a candidate for the lower bound. For $a \in [0, s_n]$, let

$$R_a = h(U)\mathrm{I}_{\{U \in (0,a)\}} + d(a)\mathrm{I}_{\{U \in [a,1)\}}, \tag{3.4}$$

where $U \sim \mathcal{U}(0,1)$. Since convex order depends only on distributions, we are only interested in the distribution of R_a and do not specify the random variable U. Note that R_a is a generalization of the random variable T_a defined in Section 2. When $U < a$, R_a is the random variable $h(U)$; when $U \geqslant a$, R_a is a constant $d(a)$. The relationship between R_a and $\mathfrak{S}_n(F_1, \ldots, F_n)$ will be discussed later. Intuitively, $1 - a$ is the mass of the atom of R_a at $d(a)$, so the smaller a is, the smaller R_a is in convex order, since it has more mass at a constant.

In the rest of the paper, we will use the following condition (D). It is parallel to (B) in Section 2.

(D) $F_1, \ldots, F_n$ are JM on the hypercube $\prod_{i=1}^n [y_i(s_n) - y(s_n), y_i(s_n)] =: \prod_{i=1}^n I_i$.

The study of joint mixability is a separate research field in probability theory; see Wang et al. (2013) and Wang and Wang (2014). The assumption (F) is sufficient for (D), as was recently shown in Wang and Wang (2014, Theorem 3.2). A numerical procedure to test for joint mixability is provided in Puccetti and Wang (2015).

We give some properties of the random variable R_a in the following lemma.

Lemma 3.4. *Suppose (C) holds, then*

(a) $\mathbb{E}[R_a] = \mathbb{E}[S]$ *for any* $S \in \mathfrak{S}_n(F_1, \ldots, F_n)$,

(b) $R_u \prec_{\mathrm{cx}} R_v$ *for* $0 \leqslant u < v \leqslant 1$,

(c) $R_{s_n} \in \mathfrak{S}_n(F_1, \ldots, F_n)$ *if (D) holds.*

Proof. (a) This follows from the definition of $d(a)$ and (3.1).

(b) We have $R_u \prec_{\mathrm{cx}} R_v$ since R_u is a *fusion* of R_v (see Theorem 2.8 of Bäuerle and Müller (2006) and Theorem 3.1 of Bernard et al. (2014a)).

(c) If (D) holds, there exist random variables $Y_1, \ldots, Y_n$ s.t. Y_i has the conditional distribution of F_i on I_i and $Y_1 + \cdots + Y_n$ is a constant. Moreover,

$$Y_1 + \cdots + Y_n = \sum_{i=1}^n \mathbb{E}[X_i \, \mathrm{I}_{\{X_i \in [y_i(s_n) - y(s_n), y_i(s_n)]\}}]/(1 - s_n) = d(s_n),$$

by the definition of d. We now construct $S \in \mathfrak{S}_n(F_1, \ldots, F_n)$ with the same distribution as R_{s_n} by imposing a particular dependence structure. Let $U \sim \mathcal{U}(0, 1)$ be independent of $Y_1, \ldots, Y_n$. We briefly explain the main idea behind the construction before moving forward to the rigorous setting:

- On the set $\{U \geqslant s_n\}$, of probability $1 - s_n$, we let $X_i = Y_i \in I_i$ for each i. We call this the *body* part of the dependence structure.
- On the set $\{U < s_n\}$, of probability s_n, which we call the *tail* part of the dependence structure, we let exactly one X_i be in the right tail region, i.e. $X_i > y_i(s_n)$, and all the others be in the left tail region, i.e. $X_j < y_j(s_n) - y(s_n)$, $j \neq i$, being counter-monotonic to X_i.

To construct the random variables rigorously, let K be a discrete random variable such that $\mathbb{P}(K = i|U) = \mathrm{d}\bar{F}_i(y_i(u))/\mathrm{d}u|_{u=U}$, $i = 1, \dots, n$. K is properly defined due to the diffentiability of $F_i(y_i(u))$ and (E1). We construct for $i = 1, \dots, n$:

$$X_i = \mathrm{I}_{\{U<s_n\}}\left(y_i(U) - y(U)\mathrm{I}_{\{K\neq i\}}\right) + \mathrm{I}_{\{U\geqslant s_n\}}Y_i,$$

and check that $X_i \sim F_i$. For $q < y_i(s_n) - y(s_n)$:

$$\begin{aligned}\mathbb{P}(X_i \leqslant q) &= \mathbb{P}\big(F_i(y_i(U) - y(U)) \leqslant F_i(q),\ K \neq i\big)\\ &= \int_0^{s_n} \mathrm{I}_{\{u-\bar{F}_i(y_i(u))\leqslant F_i(q)\}}\left(1 - \frac{\mathrm{d}\bar{F}_i(y_i(u))}{\mathrm{d}u}\right)\mathrm{d}u\\ &= \int_0^{F_i(y_i(s_n)-y(s_n))} \mathrm{I}_{\{w\leqslant F_i(q)\}}\mathrm{d}w = F_i(q),\end{aligned}$$

using (E2) twice and substitution $w = u - \bar{F}_i(y_i(u))$. Similarly, for $q > y_i(s_n)$:

$$\begin{aligned}P(X_i > q) &= P\big(F_i(y_i(U)) > F_i(q),\ K = i\big)\\ &= \int_0^{s_n} \mathrm{I}_{\{F_i(y_i(u))>F_i(q)\}}\frac{\mathrm{d}\bar{F}_i(y_i(u))}{\mathrm{d}u}\mathrm{d}u\\ &= \int_0^{\bar{F}_i(y_i(s_n))} \mathrm{I}_{\{w<1-F_i(q)\}}\mathrm{d}w = 1 - F_i(q),\end{aligned}$$

using substitution $w = \bar{F}_i(y_i(u))$. Thus X_i has the required distribution on $\mathbb{R}^+ \setminus I_i$, and by construction also on the interval I_i.

We also check that $S = X_1 + \dots + X_n$ has the required distribution:

$$S = \mathrm{I}_{\{U<s_n\}}\left(\sum_{i=1}^n y_i(U) - \sum_{i=1}^n y(U)\mathrm{I}_{\{K\neq i\}}\right) + \mathrm{I}_{\{U\geqslant s_n\}}d(s_n) = \mathrm{I}_{\{U<s_n\}}h(U) + \mathrm{I}_{\{U\geqslant s_n\}}d(s_n) \stackrel{\mathrm{d}}{=} R_{s_n}.$$

This completes the proof. □

The following theorem is the main result of this paper. We will show that R_a is a lower convex bound on the set $\mathfrak{S}_n(F_1, \dots, F_n)$ and this bound is sharp for $a = s_n$ under the JM condition (D).

Theorem 3.1 (Convex ordering lower bound and sharpness)**.** *Suppose (C) holds, then*

(i) $R_a \prec_{\mathrm{cx}} S$ *for all* $S \in \mathfrak{S}_n(F_1, \ldots, F_n)$ *and all* $a \in [0, s_n]$, *and*

(ii) R_{s_n} *is the sharp lower convex bound in* $\mathfrak{S}_n(F_1, \ldots, F_n)$ *if and only if (D) holds.*

Proof. Recall that we only consider $n \geqslant 3$. The idea of our proof in part (i) is similar to the proof of Theorem 3.1 in Bernard et al. (2014a).

(i) Let $S = X_1 + \cdots + X_n$ with $X_i \sim F_i$ be any random variable in $\mathfrak{S}_n(F_1, \ldots, F_n)$ and R_a be defined in (3.4). By Lemma 3.4(a), we have $\mathbb{E}[R_a] = \mathbb{E}[S]$. Let F_S and F_{R_a} be the df of S and R_a respectively. Our goal is to show that

$$\int_c^1 F_{R_a}^{-1}(t)\,\mathrm{d}t \leqslant \int_c^1 F_S^{-1}(t)\,\mathrm{d}t, \quad \forall c \in (0,1). \tag{3.5}$$

It is well-known that property (3.5) together with $\mathbb{E}[R_a] = \mathbb{E}[S]$ is equivalent to $R_a \prec_{\mathrm{cx}} S$ (see for instance Bäuerle and Müller, 2006, Theorem 2.5).

To prove (3.5), define $A_S(u) = \bigcup_i \{X_i > y_i(u)\}$ and let $W(u) = \mathbb{P}(A_S(u))$. By (E1) and since $\mathbb{P}$ is subadditive, $W(u) \leqslant u$ holds; moreover, $0 \leqslant W(u+\epsilon) - W(u) \leqslant \epsilon$, so W is continuous and non-decreasing. For $c \in (0, a]$, let $u^* = W^{-1}(c)$ (the generalized inverse), so $W(u^*) = c$ and thus $u^* \geqslant c$. Hence $\{X_i > y_i(c)\} \subset \{X_i > y_i(u^*)\} \subset A_S(u^*)$. Therefore

$$\mathbb{P}\left(A_S(u^*) \setminus \{X_i > y_i(c)\}\right) = c - \bar{F}_i(y_i(c)) \overset{\text{(E2)}}{=} F_i(y_i(c) - y(c)) = \mathbb{P}\left(X_i < y_i(c) - y(c)\right).$$

Since the above two sets have the same measure, we have

$$\mathbb{E}[\mathrm{I}_{\{X_i < y_i(c) - y(c)\}} X_i] \leqslant \mathbb{E}[\mathrm{I}_{A_S(u^*) \setminus \{X_i > y_i(c)\}} X_i]. \tag{3.6}$$

It follows by Lemma 3.3 that, for $c \in (0, a]$,

$$\begin{aligned}
\mathbb{E}[\mathrm{I}_{\{U<c\}} R_a] &= \mathbb{E}[\mathrm{I}_{\{U<c\}} h(U)] \\
&= \sum_{i=1}^n \mathbb{E}[(\mathrm{I}_{\{X_i < y_i(c) - y(c)\}} + \mathrm{I}_{\{X_i > y_i(c)\}}) X_i] \\
&\leqslant \sum_{i=1}^n \mathbb{E}[(\mathrm{I}_{A_S(u^*) \setminus \{X_i > y_i(c)\}} + \mathrm{I}_{\{X_i > y_i(c)\}}) X_i] \\
&= \mathbb{E}[\mathrm{I}_{A_S(u^*)} S],
\end{aligned}$$

where the inequality follows from (3.6). Thus we have

$$\mathbb{E}[\mathrm{I}_{\{U<c\}} R_a] \leqslant \mathbb{E}[\mathrm{I}_{A_S(u^*)} S]. \tag{3.7}$$

Note that $h(x)$ is non-increasing on $(0, a)$ and $\lim_{x\to a-} h(x) \geqslant d(a)$. Thus for $c \in (0, a]$,

$$\mathbb{E}[\mathrm{I}_{\{U<c\}} R_a] = \mathbb{E}[\mathrm{I}_{\{U<c\}} h(U)] = \int_{1-c}^{1} F_{R_a}^{-1}(t)\,\mathrm{d}t. \tag{3.8}$$

Also note that, since $\mathbb{P}(A_S(u^*)) = c$,

$$\mathbb{E}[\mathrm{I}_{A_S(u^*)} S] \leqslant \int_{1-c}^{1} F_S^{-1}(t)\,\mathrm{d}t. \tag{3.9}$$

It follows from (3.7), (3.8) and (3.9) that for any $c \in (0, a]$,

$$\int_{1-c}^{1} F_{R_a}^{-1}(t)\,\mathrm{d}t \leqslant \int_{1-c}^{1} F_S^{-1}(t)\,\mathrm{d}t. \tag{3.10}$$

For $x \in [0, 1-a]$, let $G(x) = \int_x^1 F_S^{-1}(t)\,\mathrm{d}t - \int_x^1 F_{R_a}^{-1}(t)\,\mathrm{d}t$. Note that $F_S^{-1}(t)$ is non-decreasing, and $F_{R_a}^{-1}(t) = d(a)$ is constant on $t \in [0, 1-a]$, hence $G(x)$ is concave on $[0, 1-a]$. Hence, with $G(0) = \mathbb{E}[S] - \mathbb{E}[R_a] = 0$ and $G(1-a) \geqslant 0$ by (3.10), we have that $G(x) \geqslant 0$ on $[0, 1-a]$. Thus

$$\int_{c}^{1} F_{R_a}^{-1}(t)\,\mathrm{d}t \leqslant \int_{c}^{1} F_S^{-1}(t)\,\mathrm{d}t \tag{3.11}$$

for any $c \in (0, 1)$, and hence $R_a \prec_{\mathrm{cx}} S$.

(ii) $\Leftarrow$: This is a direct result of (i) and Lemma 3.4(c).

$\Rightarrow$: Suppose $R_{s_n} \overset{\mathrm{d}}{=} S \in \mathfrak{S}_n(F_1, \ldots, F_n)$, so $S = X_1 + \cdots + X_n$ for some $X_i \sim F_i$, $i = 1, \ldots, n$. Thus (3.6) is an equality for each i and $c \in (0, s_n]$. This implies that $A_S(u^*) \setminus \{X_i > y_i(c)\} = \{X_i < y_i(c) - y(c)\}$, hence $A_S(u^*) = \{X_i < y_i(c) - y(c)\} \cup \{X_i > y_i(c)\}$ for each i and $c \in (0, s_n]$, up to a difference of a $\mathbb{P}$-null set. As a consequence, S has the same construction as in the proof of Lemma 3.4 (c) on the set $A := \{X_i < y_i(s_n) - y(s_n)\} \cup \{X_i > y_i(s_n)\}$. Therefore, $S\mathrm{I}_A \overset{\mathrm{d}}{=} h(U)\mathrm{I}_A$ for some $U \sim \mathcal{U}(0, s_n)$ independent of $X_1, \ldots, X_n$. Since $S \overset{\mathrm{d}}{=} R_{s_n}$, we have that $S\mathrm{I}_{A^c} = d(s_n)\mathrm{I}_{A^c}$, that is,

$$\sum_{i=1}^{n} X_i \mathrm{I}_{A^c} \overset{\text{a.s.}}{=} d(s_n)\mathrm{I}_{A^c},$$

where $A^c = \cap_{i=1}^{n} \{y_i(s_n) - y(s_n) \leqslant X_i \leqslant y_i(s_n)\}$, and thus (D) holds. □

Note that when (C) and (D) hold, s_n and $d(s_n)$ is uniquely determined. Moreover, since the smallest element in $\mathfrak{S}_n(F_1, \ldots, F_n)$ with respect to convex order is unique in law, h is unique on $(0, s_n)$. Even if (D) is not satisfied, Theorem 3.1 (i) always provides a lower convex bound, which in case of $s_n = 0$, reduces to a constant.

Together with the classical result on comonotonicity (2.1), this yields the convex bounds for $S \in \mathfrak{S}_n(F_1, \ldots, F_n)$ under (C):

$$R_{s_n} \prec_{\mathrm{cx}} S \prec_{\mathrm{cx}} F_1^{-1}(U) + \cdots + F_n^{-1}(U),$$

where $U \sim \mathcal{U}(0, 1)$. When (D) holds, the upper and lower bounds are both sharp. Note that (F) implies (C) and (D), hence in the case of decreasing densities the problem is fully solved.

Remark 3.2. If $F_1 = \cdots = F_n = F$, one solution to (E1)-(E2) is given by $y_i(x) = F_i^{-1}(x/n)$ (see Remark 3.1). In that case, we have that $h_i(x) = H(x/n)$, $d(a) = D(a/n)$, $s_n = nc_n$, $R_a = T_{na}$ with H, D, c_n and T defined in Section 2. Thus, Theorem 3.1 implies Theorem 2.1 which is the main result in Bernard et al. (2014a).

Remark 3.3. In this section we only discussed the case for $n \geqslant 3$. As explained in Section 2.3, the case for $n = 2$ is well-known. Note, however, that our result is also valid for $n = 2$, and it reduces to the counter-monotonic scenario when conditions (C) and (D) hold. This can be seen by the construction using counter-monotonicity in the proof of Lemma 3.4 (c), and the fact that joint mixability implies counter-monotonicity in the case of $n = 2$.

Similar to the homogeneous risks in Bernard et al. (2014a), the optimal dependence structure for heterogeneous risks can be described as follows. The probability space is divided into two subsets:

- For each i, if X_i is large, then each X_j, $j \neq i$ is small and (X_i, X_j) are counter-monotonic. This part has probability s_n. Since only one of X_i can be large in any outcome, this part represents *mutual exclusivity*.

- For each i, if X_i is of medium size, then each X_j, $j \neq i$ is also of medium size, and the sum $X_1 + \cdots + X_n$ is a constant. This part has probability $1 - s_n$ and represents *joint mixability*.

- The optimal dependence structure is a joint mix if $s_n = 0$, and it is a "nearly mutually exclusive structure" if $s_n = 1$.

Different from the optimal structure in Bernard et al. (2014a), the optimal structure in our paper involves functions $y_1, \ldots, y_n$. Essentially, the function $\bar{F}_i(y_i(x))$ represents the weight assigned to each individual risk X_i due to inhomogeneity. In the homogeneous case, it is x/n, which means that each individual risk is assigned an equal weight.

3.2 Discussion on assumptions

In this section, we briefly discuss some issues related to the assumptions imposed for the main results. With some more complicated technical details, some of the assumptions can be relaxed slightly.

3.2.1 Continuity of the marginal distributions

The continuity assumed in this section is for ease of notation in (E1)-(E2). In the case when $F_1, \dots, F_n$ are possibly discontinuous, (E1)-(E2) needs to be replaced by the following two equations. For $t_1, \dots, t_n : (0,1) \to (0,1)$, and $y : (0,1) \to \mathbb{R}_+$,

(G1) $\sum_{i=1}^n t_i(x) = x$, and

(G2) $F_i^{-1}(1 - t_i(x)) - F_i^{-1}(x - t_i(x)) = y(x)$ for each $i = 1, \dots, n$.

That is, $t_i(x) = \bar{F}_i(y_i(x))$ in the continuous case. Then a solution to (G1)-(G2) may exist which satisfies the properties in Lemma 3.2 (substituting $y_i(x) = F_i^{-1}(1 - t_i(x))$). In this case, Theorem 3.1 is still valid.

For instance, one can add some probability mass to the marginal distributions at zero. A particular example concerns mutual exclusivity. Suppose that $F_1, \dots, F_n$ are compatible with mutual exclusivity, i.e. $\sum_{i=1}^n F_i(0) \geqslant n - 1$, and each F_i is continuous on $[0, \infty)$. Let $y(x) = \sup\{t : \sum_{i=1}^n \bar{F}_i(t) > x\}$ and $t_i(x) = \bar{F}_i(y(x))$. We can easily check that $y, t_1, \dots, t_n$ is a solution to (G1)-(G2) which satisfies (i)-(iv) in Lemma 3.2 with strict inequalities replaced by non-strict inequalities. Theorem 3.1 holds for this case, and the optimal structure is given by mutual exclusivity. This is consistent with the construction explained after Theorem 3.1, which contains a "nearly mutually exclusive" component.

3.2.2 Monotonicity of h

One key property to show that $R_{s_n} \prec_{\mathrm{cx}} S$ for all $S \in \mathfrak{S}_n(F_1, \dots, F_n)$ is that h has to be non-increasing on $(0, s_n)$. Similar to the homogeneous models in Bernard et al. (2014a) (see conditions (A) and (A') in Section 2), if we assume that all properties in Lemma 3.2 hold on $(0, a)$ for some $a < s_n$ (in particular, h is decreasing on $(0, a)$), we still have the results in Theorem 3.1 (i): $R_a \prec_{\mathrm{cx}} S$ for all $S \in \mathfrak{S}_n(F_1, \dots, F_n)$. This can be used to obtain convex bounds when (C) does not hold, i.e. in case of less regularity of $F_1, \dots, F_n$.

However, one only obtains a sharp bound if $a = s_n$, since $R_a \notin \mathfrak{S}_n(F_1, \ldots, F_n)$ for $a < s_n$. This is implied by a necessary condition for joint mixability (Wang and Wang, 2014, Theorem 2.1). We skip a detailed discussion on cases when (C) does not hold.

3.3 Proofs of Lemma 3.1 and 3.2

In this section we give proofs of Lemmas 3.1 and 3.2.

Proof of Lemma 3.1. Fix any $x \in (0, 1)$ and inductively define the following functions for $z \in (0, x]$:

$$\varphi_i(z) = F_i^{-1}(1 - z_i) - F_i^{-1}(x - z_i), \quad i = 1, \ldots, n, \tag{3.12}$$

where $z_1(z) = z$, and for $i = 2, \ldots, n$, $z_i(z)$ denotes the unique solution to

$$\varphi_{i-1}(z - z_i) = F_i^{-1}(1 - z_i) - F_i^{-1}(x - z_i). \tag{3.13}$$

Below we will show that $z_i(z)$, $i = 1, \ldots, n$ are well-defined. The interpretation is that the function $\varphi_k(z)$ returns the value $y(x)$ for the system where (E1) is replaced by $\sum_{i=1}^k \bar{F}_i(y_i(x)) = z$, and (E2) is relaxed to hold only for $i = 1, \ldots, k$. Similarly, $z_k(z)$ returns the probability mass in the right tail of margin k in this modified system.

Since each F_i is strictly concave, F_i^{-1} is strictly convex, so $F_i^{-1}(1 - \cdot) - F_i^{-1}(x - \cdot)$ is a continuous and decreasing function, with right limit ∞ at $0+$. Clearly, $\varphi_1(\cdot)$ is such a function. For induction, assume φ_{i-1} also is and let

$$\delta_i(z, z_i) = \varphi_{i-1}(z - z_i) - F_i^{-1}(1 - z_i) - F_i^{-1}(x - z_i).$$

Then $\delta_i(z, z_i)$ is continuous in each argument, decreasing in z and increasing in z_i. Moreover, $\delta_i(z, 0+) = -\infty$ and $\delta_i(z, z-) = \infty$, so for each $z \in (0, x]$ there is a unique solution $z_i(z) \in (0, z)$ of (3.13) i.e. of $\delta_i(z, z_i) = 0$, and $z_i(z)$ is continuous and increasing. Hence, by (3.12) also φ_i is a continuous and decreasing function, with right limit ∞ at $0+$, completing the induction. Hence $z_i(z)$ and $\varphi_i(z)$, $i = 1, \ldots, n$ are well-defined functions on $z \in (0, x]$.

Finally, let $z_n^* = z_n(x)$ and $z_i^* = z_i(x - \sum_{j=i+1}^n z_j^*)$ for $i = n-1, n-2, \ldots, 1$. Setting $y = \varphi_n(x)$ and $y_i = F_i^{-1}(1 - z_i^*)$, $i = 1, \ldots, n$ yields the values of functions $y, y_1, \ldots, y_n$ at the point x. $\square$

In the following we will use Lemma 5.6 of Wang and Wang (2014), which we state below.

Lemma 3.5 (Wang and Wang (2014))**.** *If a distribution F has a decreasing density f on the support* $[0, L]$*,* $L \in \mathbb{R}^+$*, then writing* $A = Lf(0)$ *and* $B = Lf(L)$*, the mean* μ *of* F *satisfies*

$$\mu \geqslant L\frac{AB + 1 - 2B}{2(A - B)}.$$

The above lemma is obtained using a piecewise linear upper bound on F.

Proof of Lemma 3.2. (i) Suppose $y_i(x) \leqslant 0$ for some $x \in (0, 1)$. Then $\bar{F}_i(y_i(x)) = 1 > x$ which violates (E1). Suppose $y_i(x) - y(x) \leqslant 0$ for some $x \in (0, 1)$. Then $F_i(y_i(x) - y(x)) = 0$, so $\bar{F}_i(y_i(x)) = x$ by (E2), and $\bar{F}_j(y_j(x)) = 0$ for $j \neq i$ by (E1). This contradicts the assumption that densities are decreasing (hence positive) on $[0, \infty)$.

(ii) Define a function $f : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ by

$$f(y_1, \dots, y_n, y) = \left(F_1(y_1 - y) + 1 - F_1(y_1), \dots, F_n(y_n - y) + 1 - F_n(y_n), n - \sum_{i=1}^{n} F_i(y_i)\right)^\top.$$

Since $f(y_1(x), \dots, y_n(x), y(x)) = (x, \dots, x)^\top$ for $x \in (0, 1)$, the total derivative wrt. x is

$$\frac{\partial f(y_1, \dots, y_n, y)}{\partial (y_1, \dots, y_n, y)^\top} \cdot \frac{\mathrm{d}(y_1, \dots, y_n, y)^\top}{\mathrm{d}x} = (1, \dots, 1)^\top.$$

Writing $a_i = f_i(y_i - y)$ and $b_i = f_i(y_i)$, the Jacobian matrix $\frac{\partial f(x, y_1, \dots, y_n, y)}{\partial (y_1, \dots, y_n, y)^\top}$ takes the form

$$J(y_1, \dots, y_n, y) = \begin{pmatrix} a_1 - b_1 & 0 & \dots & 0 & -a_1 \\ 0 & a_2 - b_2 & 0 & \dots & -a_2 \\ \vdots & 0 & a_3 - b_3 & \dots & \vdots \\ 0 & \vdots & \vdots & \ddots & -a_n \\ -b_1 & -b_2 & \dots & -b_n & 0 \end{pmatrix}.$$

Let $c = \sum_{i=1}^{n} \frac{a_i b_i}{a_i - b_i}$ and δ_{ij} be the Kronecker delta: $\delta_{ij} = \mathrm{I}_{\{i=j\}}$. The inverse J^{-1} is then given by

$$J^{-1}_{ij} = -\frac{1}{c}\frac{a_i}{a_i - b_i}\frac{b_j}{a_j - b_j} + \frac{\delta_{ij}}{a_i - b_i}, \qquad J^{-1}_{i,n+1} = -\frac{1}{c}\frac{a_i}{a_i - b_i},$$

$$J^{-1}_{n+1,j} = -\frac{1}{c}\frac{b_j}{a_j - b_j}, \qquad J^{-1}_{n+1,n+1} = -\frac{1}{c},$$

for $i, j = 1, \dots, n$. Hence from $\mathrm{d}(y_1, \dots, y_n, y)^\top/\mathrm{d}x = J^{-1} \cdot (1, \dots, 1)^\top$, we obtain

$$y' = -\frac{1}{c}\left(\sum_{j=1}^{n} \frac{b_j}{a_j - b_j} + 1\right) < 0 \quad \text{and} \quad y_i' = \frac{1 + a_i y'}{a_i - b_i}.$$

To show $y_i' < 0$, or equivalently $a_i y' < -1$, it remains to prove the inequality

$$a_i\left(\sum_{j=1}^{n}\frac{b_j}{a_j-b_j}+1\right) > c, \quad i=1,\dots,n.$$

We apply Lemma 3.5 to the conditional distributions of F_i on intervals $[y_i-y, y_i]$, with supports shifted to the origin, i.e. we use $\hat{F}_i(z) = F_i(z+(y_i-y))/(1-x)$ for $z \in [0,y]$. For each i, denote by μ_i the mean of $\hat{F}_i$ and note that the three quantities y, $a_i y/(1-x)$ and $b_i y/(1-x)$ correspond to L, A and B in Lemma 3.5 applied to $\hat{F}_i$. Lemma 3.5 yields

$$\mu_j \geqslant y\frac{a_j b_j y/(1-x)+(1-x)/y-2b_j}{2(a_j-b_j)}, \quad j=1,\dots,n. \tag{3.14}$$

Define

$$c_0 = \sum_{j=1}^{n}\frac{1}{a_j-b_j} \quad \text{and} \quad c_1 = \sum_{j=1}^{n}\frac{b_j}{a_j-b_j},$$

then (3.14) sums up to

$$\sum_{j=1}^{n}\mu_j \geqslant y\left(\frac{cy/(1-x)+c_0(1-x)/y}{2}-c_1\right).$$

Since $x \in (0, s_n)$, we have that $h(x) = \sum_{i=1}^{n} y_i - (n-1)y > d(x) = \sum_{i=1}^{n}(\mu_i + y_i - y)$ so $\sum_{i=1}^{n}\mu_i < y$. Hence, using the AM-GM (Arithmetic Mean-Geometric Mean) inequality we obtain $1 > \sum_{i=1}^{n}\mu_i/y \geqslant \sqrt{cc_0} - c_1$. Rearranging $1 > \sqrt{cc_0} - c_1$, we have

$$\frac{(c_1+1)^2}{c_0} > c. \tag{3.15}$$

Finally, by $a_i > 1/y > b_j$, $\forall j \neq i$, it follows that

$$c_1 + 1 = \sum_{j\neq i}\frac{b_j}{a_j-b_j} + \frac{b_i}{a_i-b_i} + 1 < \sum_{j\neq i}\frac{a_i}{a_j-b_j} + \frac{a_i}{a_i-b_i} = a_i c_0.$$

As a consequence, we obtain $a_i(c_1+1) > c$ from (3.15), and thus $y_i' < 0$, $i=1,\dots,n$.

(iii) Since $(y_i - y)' = (1+b_i y')/(a_i-b_i)$, it suffices to show that $b_i y' > -1$, i.e. $b_i(c_1+1) < c$. This can be seen from

$$b_i(c_1+1) = b_i\left(\sum_{j\neq i}\frac{b_j}{a_j-b_j}+\frac{b_i}{a_i-b_i}+1\right) = \sum_{j\neq i}\frac{b_i b_j}{a_j-b_j}+\frac{a_i b_i}{a_i-b_i} < \sum_{j=1}^{n}\frac{a_j b_j}{a_j-b_j} = c.$$

(iv) A straightforward calculation yields

$$h' = \sum_{i=1}^{n} y_i' - (n-1)y' = c_0 + y'\left(\sum_{i=1}^{n}\frac{a_i}{a_i-b_i} - n + 1\right) = c_0 + y'(c_1+1) = c_0 - \frac{(c_1+1)^2}{c},$$

so $h' < 0$ by inequality (3.15). □

4 Applications to Quantitative Risk Management

4.1 Bounds on convex and coherent risk measures

Convex and coherent risk measures (Artzner et al. (1999), Föllmer and Schied (2002)) are powerful mathematical tools used to calculate capital requirement for a financial institution. The consistency between convex order and convex risk measures can be found in Bäuerle and Müller (2006). As a consequence of Theorem 3.1, we find the lower bound for convex and coherent risk measures over the admissible risk class. For general definitions of coherent and convex risk measures, we refer to Föllmer and Schied (2011, Chapter 4). Recall the definition of R_{s_n} in (3.4) on page 12.

Corollary 4.1 (Bounds on convex risk measures)**.** *For any law-invariant and convex risk measure ρ, if (C) holds, then*

$$\inf_{S\in\mathfrak{S}_n(F_1,\dots,F_n)} \rho(S) \geqslant \rho(R_{s_n}), \tag{4.1}$$

and the above inequality is an equality if (D) holds.

One of the most commonly used coherent risk measures is the Expected Shortfall (ES), also known as Tail Value-at-Risk (TVaR) or Conditional VaR (CVaR) in the actuarial literature. The ES of S at level p is defined as

$$\mathrm{ES}_p(S) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_\alpha(S)\,\mathrm{d}\alpha, \quad p \in [0,1),$$

where VaR_p is another popular risk measure, the Value-at-Risk (VaR) at level p:

$$\mathrm{VaR}_p(X) = \inf\{x : \mathbb{P}(X \leqslant x) \geqslant p\}, \quad p \in (0,1). \tag{4.2}$$

Here we give a lower bound on ES over the admissible risk class. For the notation, see page 10.

Corollary 4.2 (Bounds on ES)**.** *For $p \in [0,1)$, if (C) holds, then*

$$\inf_{S\in\mathfrak{S}_n(F_1,\dots,F_n)} \mathrm{ES}_p(S) \geqslant \begin{cases} \frac{1}{1-p}(\mathbb{E}[S] - pd(s_n)), & p \leqslant 1 - s_n, \\ \frac{1}{1-p}\int_0^{1-p} h(x)\,\mathrm{d}x, & p > 1 - s_n, \end{cases} \tag{4.3}$$

and the above inequality is an equality if (D) holds.

4.2 Bounds on expectations of convex functions

A *convex (concave) expectation* of a random variable X is defined as $\mathbb{E}[f(X)]$ where $f : \mathbb{R} \to \mathbb{R}$ is a convex (concave) function. By the definition of convex order, we have a straightforward corollary about the lower bound on a convex expectation (or upper bound on a concave expectation) over the admissible risk class $\mathfrak{S}_n(F_1, \dots, F_n)$,

$$\mathbb{E}[f(S)] = \mathbb{E}[f(X_1 + X_2 + \cdots + X_n)]. \tag{4.4}$$

Recall that when f is convex, the upper bound can be computed explicitly with the comonotonic dependence structure.

Corollary 4.3 (Bounds on convex expectations)**.** *For a convex function f, if (C) holds, then*

$$\inf_{S \in \mathfrak{S}_n(F_1,\dots,F_n)} \mathbb{E}[f(S)] \geqslant \int_0^{s_n} f(h(x))\,\mathrm{d}x + (1 - s_n) f(d(s_n)), \tag{4.5}$$

and the above inequality is an equality if (D) holds.

Note that when $s_n = 0$, (4.5) degenerates to Jensen's inequality. However, Lemma 3.4 (b) implies $R_0 \prec_{\mathrm{cx}} R_{s_n}$, so (4.5) always gives a better bound than Jensen's inequality when $s_n > 0$.

Another inequality about the lower bound on convex expectation is given in Cheung and Lo (2013), summarized as follows:

Proposition 4.1 (Cheung and Lo (2013))**.** *Let $X_1, \dots, X_n$ be non-negative random variables, $S = X_1 + \cdots + X_n$, and f be a convex function such that $\mathbb{E}[f(S)]$ exists.*

(i) We have

$$\mathbb{E}[f(S)] \geqslant L_1 := \sum_{i=1}^{n} \mathbb{E}[f(X_i)] - (n-1) f(0). \tag{4.6}$$

(ii) If f is strictly convex, then equality holds in (4.6) if and only if $X_1, \dots, X_n$ are mutually exclusive random variables.

Inequality (4.6) provides an easily calculated lower bound on $\mathbb{E}[f(S)]$, $S \in \mathfrak{S}_n(F_1, \dots, F_n)$. The bound (4.6) is sharp only if $F_1, \dots, F_n$ are compatible with mutual exclusivity. When mutual exclusivity is not compatible, it is not clear which bound in (4.5)-(4.6) dominates the other. For instance, for bounded distributions with $h(0) \leqslant d(0)$, (4.5) becomes Jensen's inequality which is not strictly comparable to (4.6). In the numerical illustration in Section 5, we will compare the two bounds.

Remark 4.1. Cheung and Lo (2013) provide two proofs of (4.6) based on the Breeden-Litzenberger formula and mutual exclusivity. We remark that it can also be obtained from Karamata's inequality (Karamata, 1932) that

$$f(S) + (n-1)f(0) \geqslant \sum_{i=1}^{n} f(X_i),$$

which is a stronger statement than (4.6).

4.3 Bounds on Value-at-Risk

The search for bounds on Value-at-Risk under dependence uncertainty is a topic of considerable interest in theory as well as practice; see Embrechts et al. (2014) for a review. Due the fact that VaR does not respect convex order, optimization problems with respect to VaR have always been challenging. A connection between the bounds on VaR and the lower convex bound over an admissible risk class has been given in Bernard et al. (2014a,b).

We suppose that $F_1, \ldots, F_n$ are continuous dfs with positive density. For each $i = 1, \ldots, n$, we let $F_{i,p}$ for $p \in (0,1)$ be the conditional distribution of F_i on $[F_i^{-1}(p), \infty)$, and let F_i^p for $p \in (0,1)$ be the conditional distribution of F_i on $[0, F_i^{-1}(p))$. Theorem 4.6 of Bernard et al. (2014a) gives that

$$\sup_{S \in \mathfrak{S}_n(F_1,\ldots,F_n)} \mathrm{VaR}_p(S) = \sup\{\text{ess-inf}(S) : S \in \mathfrak{S}_n(F_{1,p}, \ldots, F_{n,p})\},$$

and

$$\inf_{S \in \mathfrak{S}_n(F_1,\ldots,F_n)} \mathrm{VaR}_p(S) = \inf\{\text{ess-sup}(S) : S \in \mathfrak{S}_n(F_1^p, \ldots, F_n^p)\},$$

where ess-inf(X) and ess-sup(X) are the essential infimum and the essential supremum of the support of a random variable X. With this connection, the proofs of the following two corollaries are trivial, hence we omit them here.

In the following corollary, we use $(\mathrm{C})_p$ (resp. $(\mathrm{D})_p$) if $F_{1,p}, \ldots, F_{n,p}$ satisfy (C) (resp. (D)). The function h_p and the quantity $s_{n,p}$ are the corresponding h and s_n defined for the distributions $F_{1,p}, \ldots, F_{n,p}$.

Corollary 4.4 (Upper bound on VaR)**.** *For $p \in (0,1)$, if $(C)_p$ holds, then*

$$\sup_{S \in \mathfrak{S}_n(F_1,\ldots,F_n)} \mathrm{VaR}_p(S) \leqslant h_p(s_{n,p}), \tag{4.7}$$

and the above inequality is an equality if $(D)_p$ holds.

Note that (F) implies (C) and (D). For $(C)_p$ and $(D)_p$ to hold, it suffices to require that each $F_1, \dots, F_n$ has a decreasing density beyond its p-quantile. As p is typically close to 1, this assumption is satisfied by all practical examples. Because of its practical relevance, we state it as a separate corollary; the homogeneous version of this result is given in Wang et al. (2013).

Corollary 4.5 (Upper bound on VaR for tail-decreasing densities). *For $p \in (0, 1)$, suppose that each $F_1, \dots, F_n$ has a decreasing density beyond its p-quantile, then*

$$\sup_{S \in \mathfrak{S}_n(F_1,\dots,F_n)} \mathrm{VaR}_p(S) = h_p(s_{n,p}). \tag{4.8}$$

Similar result holds for the lower bound on VaR. In the following corollary, we use $(C)^p$ (resp. $(D)^p$) if $F_1^p, \dots, F_n^p$ satisfy (C) (resp. (D)). Let μ_i^p be the mean of F_i^p, $i = 1, \dots, n$.

Corollary 4.6 (Lower bound on VaR). *For $p \in (0, 1)$, if $(C)^p$ holds, then*

$$\inf_{S \in \mathfrak{S}_n(F_1,\dots,F_n)} \mathrm{VaR}_p(S) \geqslant \max\left\{\max_{i=1,\dots,n}\left\{F_i^{-1}(p) + \sum_{j\neq i} F_j^{-1}(0)\right\}, \sum_{i=1}^{n} \mu_i^p\right\}, \tag{4.9}$$

and the above inequality is an equality if $(D)^p$ holds.

The inequality (4.9) is straightforward, whereas the sharpness under $(C)^p$ and $(D)^p$ is not.

5 Numerical Illustration

5.1 Results and comparison

In this section we explain how to calculate the bounds obtained in the previous sections, and present the results from a numerical case study considering 10 different sets of marginal distributions. We compare the results from the method based on Corollaries 4.2 and 4.3 with their alternatives: the Rearrangement Algorithm (RA) (Puccetti and Rüschendorf (2012), Embrechts et al. (2013)), the bound given in Bernard et al. (2014a) based on an approximation with homogeneous risks, and the bound (4.6) given in Cheung and Lo (2013). Note that the sharpness of the latter two bounds is either not justified or fails to hold for most of the heterogeneous models of relevance. The RA values can be used as a good approximation to the exact values of the bounds considered.

In Table 1 the bounds for the sets of margins considered in Bernard et al. (2014a) are listed, and in Table 2 five further cases are presented. Three quantities are calculated: the variance

$\mathbb{E}[f(S)] = \mathbb{E}[(S-K)^2]$ where $K = \sum_{i=1}^n \mathbb{E}[X_i]$; the European call option price (or excess of loss) $\mathbb{E}[f(S)] = \mathbb{E}[(S-K)_+]$, and ES of S at level 95%. It was mentioned in Cheung and Lo (2013) that their bound on ES involves a complicated optimization procedure for heterogeneous models which is unavailable in their paper, hence for ES we only compare Corollary 4.2 with the RA and the approximation in Bernard et al. (2014a).

We observe in Tables 1 and 2 that the variance bounds from Cheung and Lo (2013) are quite poor (in case a negative bound was obtained, it was replaced by the trivial bound $\mathrm{Var}(S) \geqslant 0$). The bounds on option price are closer, but still far from sharp. In contrast, the bounds from Corollaries 4.2 and 4.3 and the RA are in close agreement, which suggests that the JM condition (D) holds for the considered cases, and the bounds we obtain are sharp. In Table 2 more dissimilar distributions (e.g. from different families) are considered, for which the homogeneous approximation bounds in Bernard et al. (2014a) are significantly worse than the ones obtained by the method presented in this paper. The greatest deviation can be observed in the case Pareto-LogN-Gamma (the last column of Table 2). For this case, in the left panel of Figure 1 the plots for the functions $h(x)$, $d(x)$ (used for sharp bounds in Theorem 3.1) are given, along with $H(x/n)$ and $D(x/n)$ in Bernard et al. (2014a), which are used for the homogeneous approximation bounds. Notice that the distributions of the constructed lower convex bounds using the two methods differ significantly, and this leads to the discrepancies in Table 2. In the right panel the weights $\bar{F}_i(y_i(x))$ from the solution to the functional equations (E1)-(E2) are plotted. For comparison, the implied values from the homogeneous approximation are shown in gray.

In the following section we provide details of the implementation.

5.2 Solving the functional equations

In order to solve the functional equations (E1)-(E2), we first obtain a system of ordinary differential equations (ODEs) and an initial condition (as in the Implicit Function Theorem). Differentiating (E1)-(E2) wrt. x yields a system of ODEs

(E1') $-\sum_{i=1}^n f_i(y_i(x))y_i'(x) = 1,$

(E2') $\big(f_i(y_i(x)-y(x)) - f_i(y_i(x))\big)y_i'(x) - f_i(y_i(x)-y(x))y'(x) = 1 \quad$ for each $i = 1, \dots, n,$

where f_i is the density corresponding to F_i. For given $y_i(x)$ and $y(x)$, this is a system of $n+1$ linear equations of the form $A \cdot (y_1', \dots, y_n', y')^\top = b$ (easy to solve, see the proof of Lemma 3.2).

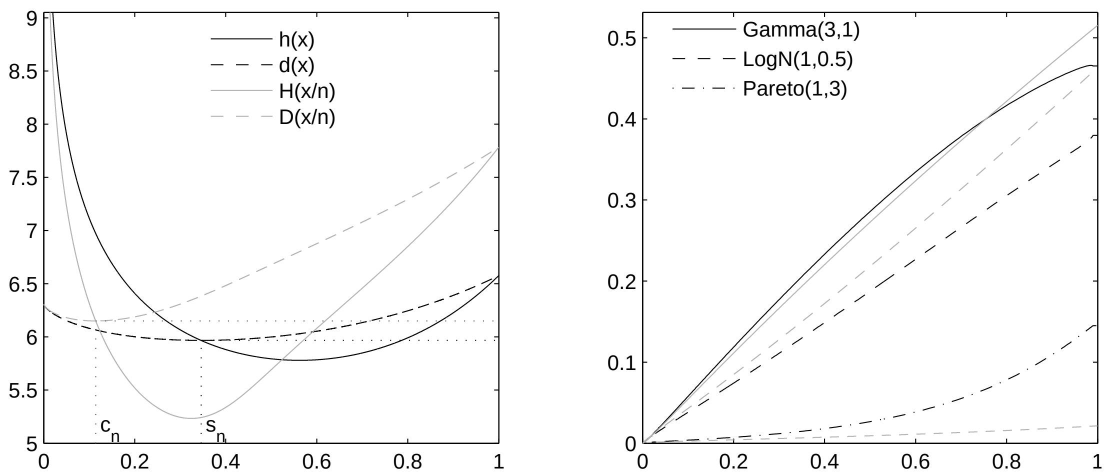

Figure 1: Left panel: plots of functions $h(x)$, $d(x)$ (used for sharp bounds) and functions $H(x/n)$, $D(x/n)$ from Bernard et al. (2014a). Right panel: $\bar{F}_i(y_i(x))$ from the solution to the functional equations (E1)-(E2), in gray - implied values from the homogeneous approximation.

Thus, if an initial condition is available, we can solve the ODEs (E1')-(E2') using an Euler-type scheme.

We find an initial condition $y_i(\epsilon), y(\epsilon)$ at $x = \epsilon > 0$ (small), using an approximate method:

1. Solve $\sum_{i=1}^{n} \bar{F}_i(y^{(M)}) = \epsilon$ for $y^{(M)} \in \mathbb{R}$,

2. Let $y_i^{(m)} = F_i^{-1}(\epsilon - \bar{F}_i(y^{(M)}))$ (so that $F_i(y_i^{(m)}) + \bar{F}_i(y^{(M)}) = \epsilon$),

3. Let $y(\epsilon) = y^{(M)} - \min_{1 \leqslant i \leqslant n} y_i^{(m)}$ and $y_i(\epsilon) = y_i^{(m)} + y(\epsilon)$.

Note that we effectively start by satisfying (E1) with identical $y_i \equiv y^{(M)}$ and then shifting the y_i to the right, in order to satisfy (E2). Since the density in the right tail is typically smaller than in the left tail, this shift does not significantly reduce the remaining mass in the right tail. Moreover, for margins with the largest $\bar{F}_i(y^{(M)})$, the y_i are likely to be shifted the least. This method provides very accurate initial conditions in the considered cases (error in (E1) less than $\epsilon \cdot 10^{-4}$ for $\epsilon = \Delta x/2 = 10^{-6}/2$).

Using the initial condition at $x = \epsilon$ obtained in this manner, we solve the ODEs for $x \in [\epsilon, 1]$. This yields the solution $y_i(x), y(x)$ on a non-uniform grid of x values (we use the function `ode113` in Matlab, which is a variable step-size algorithm). Finally, we use linear interpolation to compute $y_i(x)$ and $y(x)$ for the desired grid-points $x \in \{(k - \frac{1}{2})\Delta x : k = 1, \dots, 1/\Delta x\}$. We chose the step size $\Delta x = 10^{-6}$, as is typical in the literature. In all considered cases the interpolated results satisfy the equations (E1) and (E2) with an absolute error of order 10^{-8} or less (for $x \in (0, s_n)$).

In Figure 2 the solutions (quantiles) $y_i(x) - y(x)$ and $y_i(x)$ for x up to $s_n = 0.61$ are plotted over the densities of three different Log-Normal margins (the second case in Table 2). We notice that more probability in the right tail of LogN(1, 1) is combined with the left tails of the other two margins, i.e. $\bar{F}_2(y_2(x))$ is the largest term on the left-hand side of (E1). The remaining support intervals (shown in white) are of length $y(s_n)$, and $h(s_n) = d(s_n)$ holds.

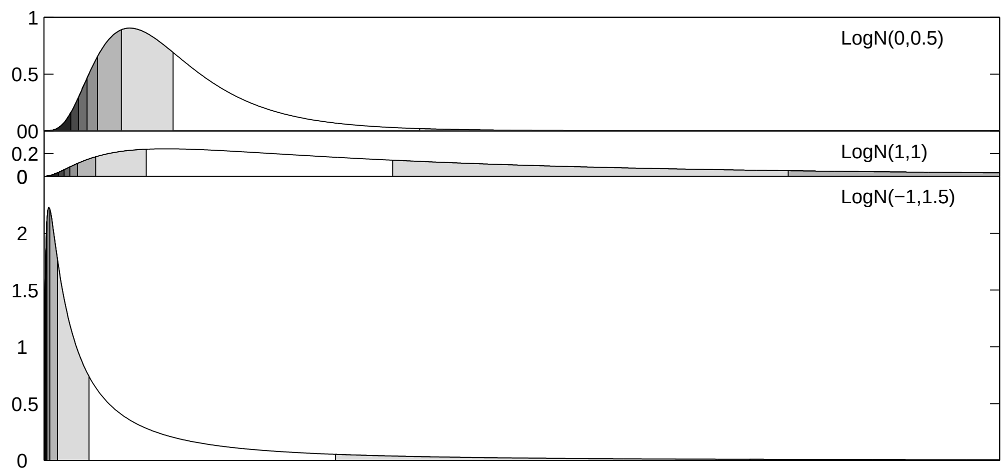

Figure 2: Quantiles $y_i(x) - y(x)$ and $y_i(x)$ for $x = 1\%, 2.5\%, 5\%, 10\%, 25\%, 61\%$.

Finally, to check that the discretization is fine enough, we compared the numerical means of the constructed convexly minimal elements R_{s_n} with the theoretical means. In all cases the relative error wrt. the theoretical means was of the order 10^{-6}.

5.3 Computation times

The computations were performed on a Lenovo X1 laptop with Intel Core i7 2GHz × 4 processor and 8GB RAM. Computation times are summarized in Table 3. The range of computation times between the ten considered cases are as follows. RA: 28 – 81s, homogeneous approximation bounds from Bernard et al. (2014a): 373 – 667s, sharp bounds from Corollaries 4.3 and 4.2: 5 – 10s. The considerably longer computation times for the homogeneous approximation bounds are due to the fact that we need to compute the inverse of $F = \frac{1}{n} \sum F_i$, which is computationally expensive, and efficient algorithms are available only for special cases, see e.g. Castellacci (2012). To test the scalability of the algorithms, we also computed the bounds for higher values of n for margins ($n/2$ different Pareto and $n/2$ different LogN). For $n = 20$ the computation time for RA was 1 hour (note that the time is random due to a random initial rearrangement), for homogeneous approximation 13 minutes and for the sharp bounds 13s. Furthermore, it was possible to compute the sharp bounds for up to $n = 100$ different margins in less than 1 minute. At this high number of margins, s_n is very small and the optimal sum R_{s_n} is almost identical to its mean R_0.

Table 1: RA results vs theoretical bounds

	Pareto$(1, \alpha_i)$, $n = 3$ ($\alpha_1, \alpha_2, \alpha_3$)		$X_i \sim$ Log-Normal$(\frac{i}{10}, 1)$ n		
	(3, 4, 5)	(3.5, 4, 4.5)	3	5	10
	Variance				
RA	0.6057	0.3352	9.0990	10.3695	15.4077
Bernard et al. (2014a)	0.6063	0.3360	9.1075	10.3803	15.3885
Cheung and Lo (2013)	0.3114	0.0537	0	0	0
Corollary 4.3	0.6063	0.3359	9.1085	10.3938	15.5071
	Option price with strike $K = \sum_{i=1}^{n} \mathbb{E}[X_i]$				
RA	0.1727	0.1413	0.7701	0.6419	0.4668
Bernard et al. (2014a)	0.1726	0.1413	0.7700	0.6405	0.4604
Cheung and Lo (2013)	0.1653	0.1340	0.6726	0.5450	0.4047
Corollary 4.3	0.1727	0.1413	0.7701	0.6420	0.4670
	ES at level 0.95				
RA	6.4259	5.8758	16.0777	21.7849	39.0253
Bernard et al. (2014a)	6.4257	5.8760	16.0778	21.7765	38.8893
Corollary 4.2	6.4261	5.8760	16.0782	21.7860	39.0288

Table 2: RA results vs theoretical bounds

	$X_i \sim \text{Pareto}(1, \alpha_i)$ $n = 10$ $\alpha = 3, \ldots, 12$	$X_1 \sim \text{LogN}(0, \frac{1}{2})$ $X_2 \sim \text{LogN}(1, 1)$ $X_3 \sim \text{LogN}(-1, \frac{3}{2})$	$X_1 \sim \text{LogN}(0, 1)$ $X_2 \sim \text{LogN}(1, 1)$ $X_3 \sim \text{Pareto}(1, 3)$	$X_1 \sim \text{Gamma}(2, \frac{1}{2})$ $X_2 \sim \text{Gamma}(3, \frac{1}{3})$ $X_3 \sim \text{Gamma}(4, \frac{1}{4})$	$X_1 \sim \text{Pareto}(1, 3)$ $X_2 \sim \text{LogN}(1, \frac{1}{2})$ $X_3 \sim \text{Gamma}(3,1)$
	Variance				
RA	0.3680	33.9531	27.8480	0.0258	0.7915
Bernard et al. (2014a)	0.3688	32.5465	27.3492	0.0199	0.4876
Cheung and Lo (2013)	0	22.6945	18.9757	0	0
Corollary 4.3	0.3686	33.9848	27.8620	0.0258	0.7916
	Option price with strike $K = \sum_{i=1}^{n} \mathbb{E}[X_i]$				
RA	0.0729	1.5156	1.3957	0.0305	0.2474
Bernard et al. (2014a)	0.0728	1.4305	1.3613	0.0235	0.1283
Cheung and Lo (2013)	0.0710	1.3745	1.3206	0.0132	0.1230
Corollary 4.3	0.0729	1.5156	1.3957	0.0305	0.2474
	ES at level 0.95				
RA	13.4667	25.4528	24.9180	3.5553	10.5449
Bernard et al. (2014a)	13.4650	25.1146	24.8130	3.4551	9.8726
Corollary 4.2	13.4669	25.4534	24.9187	3.5554	10.5449

Table 3: Computation times, in seconds

Case	RA	Bernard et al. (2014a)	Corollaries 4.2 and 4.3
Table 1, case 1	28	438	5
Table 1, case 2	29	434	5
Table 1, case 3	30	375	5
Table 1, case 4	38	399	6
Table 1, case 5	81	456	8
Table 2, case 1	79	667	8
Table 2, case 2	32	373	6
Table 2, case 3	28	393	6
Table 2, case 4	39	572	10
Table 2, case 5	38	459	7

6 Conclusion

We give a general lower bound on the aggregate risk with respect to convex order for heterogeneous marginal distributions. The bound is shown to be sharp when the marginal distributions all have decreasing densities. The new result partially answers an open question that has existed in the theory of dependence modeling for a long time. Although the proposed lower convex bound is generally implicit and involves solving a non-trivial functional equation, it helps to understand the safest dependence structure with respect to convex order. As opposed to comonotonicity, which is often treated as the most dangerous dependence structure, the safest dependence structure can be interpreted as a combination of joint mixability and mutual exclusivity. This is indeed not surprising if one realizes that joint mixability and mutual exclusivity both give the safest dependence structure when they are compatible with the marginal distributions. Our results directly lead to bounds on quantities including convex and coherent risk measures, the expectation of convex functions and the Value-at-Risk of the aggregate risk. A numerical procedure is provided to identify the distribution representing the lower convex bound and to compute corresponding quantities of interest.

We remark that there are still quite a few open questions on this new method, concerning conditions (C) and (D). Numerical evidence seems to suggest that the main theorem holds for much more general classes of distributions than those with decreasing densities. A theoretical proof of such statements is still beyond our knowledge, and is expected to be very challenging. The new results obtained in this paper are closely associated with the development of theory of joint mixability, on which many open questions are left unanswered at this moment.

Acknowledgement

The authors thank an anonymous referee and Paul Embrechts for valuable comments and suggestions. E. Jakobsons thanks RiskLab Switzerland and the Swiss Finance Institute for financial support. X. Han thanks the Department of Statistics and Actuarial Science, University of Waterloo for financial support during her visit in Waterloo, Summer 2014. R. Wang acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC No. 435844), and the Forschungsinstitut für Mathematik (FIM) at ETH Zurich during his visits in 2013 and 2014.

References

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.

BCBS (2012). *Consultative Document May 2012. Fundamental review of the trading book*. Basel Committee on Banking Supervision. Basel: Bank for International Settlements.

BCBS (2013a). *Consultative Document October 2013. Fundamental review of the trading book: A revised market risk framework*. Basel Committee on Banking Supervision. Basel: Bank for International Settlements.

BCBS (2013b). *Consultative Document July 2013. The regulatory framework: Balancing risk sensitivity, simplicity and comparability*. Basel Committee on Banking Supervision. Basel: Bank for International Settlements.

Bäuerle, N. and Müller, A. (2006). Stochastic orders and risk measures: Consistency and bounds. *Insurance Mathematics and Economics*, **38**(1), 132–148.

BERNARD, C., JIANG, X. AND WANG, R. (2014a). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108.

BERNARD, C., RÜSCHENDORF, L. AND VANDUFFEL, S. (2014b). Value-at-Risk bounds with variance constraints. *Preprint*, University of Freiburg.

CASTELLACCI, G. (2012). A formula for the quantiles of mixtures of distributions with disjoint supports. *Available at SSRN-2055022*.

CHEUNG, K. C., AND LO, A. (2013). General lower bounds on convex functionals of aggregate sums. *Insurance: Mathematics and Economics*, **53**(3), 884–896.

CHEUNG, K. C. AND VANDUFFEL, S. (2013). Bounds for sums of random variables when the marginal distributions and the variance of the sum are given. *Scandinavian Actuarial Journal*, **2013**(2), 103–118.

DALL'AGLIO, G. (1972). Fréchet classes and compatibility of distribution functions. *Symposia Mathematica*, **9**, 131–150.

DEELSTRA, G., DHAENE J. AND VANMAELE, M. (2011). An overview of comonotonicity and its applications in finance and insurance, in: *Advanced Mathematical Methods for Finance,* Di Nunno, G. and Øksendal B. (Eds.), Springer, 2011.

DELBAEN, F. (2012). *Monetary Utility Functions*. Osaka University Press.

DHAENE, J. AND DENUIT, M. (1999). The safest dependence structure among risks. *Insurance: Mathematics and Economics*, **25**(1), 11-21.

DHAENE, J., DENUIT, M., GOOVAERTS, M.J., KAAS, R. AND VYNCHE, D. (2002). The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, **31**(1), 3-33.

EMBRECHTS, P., PUCCETTI, G. AND RÜSCHENDORF, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750–2764.

EMBRECHTS, P., PUCCETTI, G., RÜSCHENDORF, L., WANG, R. AND BELERAJ, A. (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25–48.

FÖLLMER, H. AND SCHIED, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, **6**, 429–447.

Föllmer, H. and Schied, A. (2011). *Stochastic Finance: An Introduction in Discrete Time.* Third ed., Walter de Gruyter, Berlin/New York.

KARAMATA, J. (1932) Sur une inégalité relative aux fonctions convexes, *Publ. Math. Univ. Belgrade* **1**, 145–148.

MÜLLER, A. (1997). Stop-loss order for portfolios of dependent risks. *Insurance: Mathematics and Economics*, **21**, 219–223.

MÜLLER, A. (1997). Stochastic orderings generated by integrals: A unified study. *Advances in Applied Probability*, **29**, 414–428.

MÜLLER, A. AND STOYAN, D. (2002). *Comparison Methods for Statistical Models and Risks*. Wiley, England.

PUCCETTI, G. AND RÜSCHENDORF L. (2012). Computation of sharp bounds on the distribution of a function of dependent risks. *Journal of Computational and Applied Mathematics*, **236** (7), 1833-1840.

PUCCETTI, G., WANG, B. AND WANG, R. (2012). Advances in complete mixability. *Journal of Applied Probability,* **49**(2), 430–440.

PUCCETTI, G., WANG, B. AND WANG, R. (2013). Complete mixability and asymptotic equivalence of worst-possible VaR and ES estimates. *Insurance: Mathematics and Economics*, **53**(3), 821–828.

PUCCETTI, G. AND WANG, R. (2014). General extremal dependence concepts. *Preprint*, University of Firenze.

PUCCETTI, G. AND WANG, R. (2015). Detecting complete and joint mixability. *Journal of Computational and Applied Mathematics*, **280**, 174–187.

RÜSCHENDORF, L. AND UCKELMANN, L. (2002). Variance minimization and random variables with constant sum, in: *Distributions with Given Marginals*. Cuadras, et al. (Eds.), Kluwer, 211–222.

SANDSTRÖM, A. (2010). *Handbook of Solvency for Actuaries and Risk Managers: Theory and Practice*. Taylor & Francis US, Florida.

SCOR (2008). *From Principle-based Risk Management to Solvency Requirements*. Second ed., SCOR, Zurich. SCOR Switzerland AG.

US SENATE (2013). *JPMorgan Chase Whale Trades: A Case History of Derivatives Risks and Abuses.* Committee on Homeland Security and Governmental Affairs.

SHAKED, M., SHANTHIKUMAR, J.G. (2007). *Stochastic Orders*. Springer Series in Statistics.

TCHEN, A. (1980). Inequalities for distributions with given marginals. *The Annals of Probability*, **8**(4): 814–827.

WANG, B. AND WANG, R. (2011). The complete mixability and convex minimization problems for monotone marginal distributions. *Journal of Multivariate Analysis*, **102**, 1344–1360.

WANG, B. AND WANG, R. (2014). Joint mixability. *Preprint*, University of Waterloo.

WANG, R., PENG, L. AND YANG, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, **17**(2), 395–417.