Risk Aggregation with Dependence Uncertainty

Carole Bernard*, Xiao Jiang† and Ruodu Wang‡

November 2013§

Abstract

Risk aggregation with dependence uncertainty refers to the sum of individual risks with known marginal distributions and unspecified dependence structure. We introduce the admissible risk class to study risk aggregation with dependence uncertainty. The admissible risk class has some nice properties such as robustness, convexity, permutation invariance and affine invariance. We then derive a new convex ordering lower bound over this class and give a sufficient condition for this lower bound to be sharp in the case of identical marginal distributions. The results are used to identify extreme scenarios and calculate bounds on Value-at-Risk as well as on convex and coherent risk measures and other quantities of interest in finance and insurance. Numerical illustrations are provided for different settings and commonly-used distributions of risks.

Key-words: dependence structure; aggregate risk; admissible risk; convex risk measures; TVaR; convex order; complete mixability; VaR bounds.

*Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada. (email: c3bernar@uwaterloo.ca).

†Master's Student in the MQF program at the University of Waterloo. (e-mail: x9jiang@uwaterloo.ca).

‡Corresponding author. Department of Statistics and Actuarial Science, University of Waterloo. (email: wang@uwaterloo.ca). Tel: 001 519 888 4567 ext: 31569.

§C. Bernard and R. Wang acknowledge support from the Natural Sciences and Engineering Research Council of Canada. C. Bernard acknowledges support from the Society of Actuaries Centers of Actuarial Excellence Research Grant and from the Humboldt research foundation. The authors would like to thank Keita Owari (University of Tokyo) for discussions on convex risk measures, Bin Wang (Huawei Technologies Co. Ltd.) for a counterexample of convex ordering sharp bounds, Edgars Jakobsons (ETH Zurich) for help on numerical examples, Steven Vanduffel (Vrije Universiteit Brussel) for his helpful suggestions on earlier drafts of this paper and the anonymous referee who helped us to improve the paper. This version contains a few small corrections made in November 2015, March 2016 and November 2019.

1 Introduction

In quantitative risk management, *risk aggregation* refers to the (probabilistic) behavior of an aggregate position $S(\mathbf{X})$ associated with a risk vector $\mathbf{X} = (X_1, \cdots, X_n)$, where $X_1, \cdots, X_n$ are random variables representing individual risks (one-period losses or profits). In this paper, we focus on the most commonly studied aggregate risk position, that is the sum $S = X_1 + \cdots + X_n$, since it has important and self-explanatory financial implications as well as tractable probabilistic properties.

In practice, there exist efficient and accurate statistical techniques to estimate the respective marginal distributions of $X_1, \cdots, X_n$. On the other hand, the joint dependence structure of $\mathbf{X}$ is often much more difficult to capture: there are computational and convergence issues with statistical inference of multi-dimensional data, and the choice of multivariate distributions is quite limited compared to the modelling of marginal distributions. However, an inappropriate dependence assumption can have important risk management consequences. For example, using the Gaussian multivariate copula can result in severely underestimating probability of simultaneous default in a large basket of firms (McNeil et al. (2005)). In this paper, we focus on the case when the marginal distributions of $X_1, \cdots, X_n$ are known and the dependence structure of $\mathbf{X}$ is unspecified. This scenario is referred to as *risk aggregation with dependence uncertainty.* To study the aggregate risk when the information of dependence is unavailable or unreliable, we introduce the concept of *admissible risk* as a possible aggregate risk S with given marginal distributions but unknown dependence structure.

We are particularly interested in the convex order of elements in an admissible risk class. Generally speaking, convex order is consistent with preferences among admissible risks for all risk-avoiding investors. Previous studies on convex order of admissible risks mainly focused on the sharp upper bound for a general number n of individual risks and the sharp lower bound for $n = 2$ (for example, see Denuit et al. (1999), Tankov (2011) and Bernard et al. (2012, 2013a)). In this paper, however, we focus on the sharp lower bound when $n \geqslant 3$, which is known to be an open problem for a long time. We show that the existence of a convex ordering minimal element in an admissible risk class is not guaranteed by providing a counterexample, and give conditions under which it exists. One of the conditions involves checking complete mixability (Wang and Wang (2011)). In the last section we propose a numerical technique to check this property, which suggests that the Gamma and Log-Normal distributions are completely mixable.

As we will show, a convex ordering lower bound can be useful to quantify model risk and in many financial applications. A first application is to quantify model risk in capital requirements. Regulators and companies are usually more concerned about a risk measure $\rho(S)$ (as a measure of risk exposure or as capital requirements needed to hold the position S over a pre-determined period) instead of the exact dependence structure of $\mathbf{X}$ itself. When a given dependence structure is chosen, $\rho(S)$ can be computed exactly. However, when the dependence structure is unspecified, $\rho(S)$ can take a range of possible values, which can then be interpreted as a measure of model uncertainty (Cont (2006)) with the absence of information on dependence. The assessment of aggregate risks S with given marginal distributions and partial information on the dependence structure, has been extensively studied in quantitative risk management. A large part of the literature focuses on properties of a specific risk measure when there is no extra information on the dependence structure, for instance: bounds on the distribution function and the Value-at-Risk (VaR) of S were studied by Embrechts et al. (2003), Embrechts and Puccetti (2006) and Wang et al. (2013), among others; convex ordering bounds on S were studied by for example Denuit et al. (1999), Dhaene et al. (2002) and Wang and Wang (2011). Some numerical methods to approximate bounds on risk measures were recently provided by Puccetti and Rüschendorf (2012), Embrechts et al. (2013), and Puccetti (2013). Another direction in the literature has been to study the case when marginals are fixed, and some extra information on the dependence is available; see Cheung and Vanduffel (2013) for convex ordering bounds with given variance; Bernard et al. (2013b) for VaR bounds with a variance constraint; Kaas et al. (2009) for the worst Value-at-Risk with constraints of positively quadrant dependence and some given measures of dependence for $n = 2$; Embrechts and Puccetti (2006) for bounds on the distribution of S when the copula of $\mathbf{X}$ is bounded by a given copula; Tankov (2011) and Bernard et al. (2012, 2013a) for bounds on S when $n = 2$; see also the Herd index proposed by Dhaene et al. (2012) based on the maximum variance of aggregate risk with estimated marginal variances. We refer to Embrechts and Puccetti (2010) for an overview on risk aggregation with no or partial information on dependence. Note that our work is fundamentally different from the literature on the lower bounds on $\rho(S)$ obtained by conditioning methods (e.g. Rogers and Shi (1995), Kaas et al. (2000), Valdez et al. (2009)).

Another contribution is to show that the convex ordering lower bound gives the explicit expression of the infimum and supremum of VaR (and proves the intuition behind the numerical bounds obtained for example by Embrechts et al. (2013); Puccetti and Rüschendorf (2012)). Convex ordering bounds are also directly related to bounds on convex expectations[1] and on general law-invariant convex risk measures, including coherent risk measures. Convex expectations appear naturally in many practical problems such as basket options (Tankov (2011), d'Aspremont and El Ghaoui (2006), Hobson et al. (2005), Albrecher et al. (2008)), discrete variance options pricing (Keller-Ressel and Clauss (2012)), stop-loss premiums for aggregate risk, variances and expected utilities. Examples are discussed extensively in Section 5. *Coherent risk*

[1] A *convex (concave) expectation* is computed as $\mathbb{E}[f(S)]$ where $f : \mathbb{R} \to \mathbb{R}$ is a convex (concave) function.

measures were introduced in Artzner et al. (1999) and later extended by Delbaen (2002) and Kusuoka (2001), among others. See also McNeil et al. (2005) for an overview. In Section 5, we will discuss how convex ordering bounds lead to bounds on convex and coherent risk measures. An important application for the financial industry is to obtain bounds on the coherent risk measure Tail-Value-at-Risk[2] (TVaR) of a joint portfolio $S = X_1 + \cdots + X_n$, when the dependence between individual assets $X_1, \cdots, X_n$ is unknown. More details and applications are given in Section 5.

The rest of the paper is organized as follows. In Section 2 we introduce the concept of *admissible risk class* and derive its properties. The main results of this paper focus on this class. Section 3 provides a new convex ordering lower bound over the admissible risk class, for both homogeneous and heterogeneous risks. It is shown that under some conditions, this bound is sharp in the homogeneous case. Section 4 gives a connection between the convex ordering lower bound and bounds on the Value-at-Risk. Bounds on convex risk measures and other applications are then given in Section 5. Some numerical illustrations can be found in Section 6. Concluding remarks are given in Section 7.

2 Admissible Risk

Assume that all random variables live in a general atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$. This means that for all $A \subset \Omega$ with $\mathbb{P}(A) > 0$, there exists $B \subsetneq A$ such that $\mathbb{P}(B) > 0$. The atomless assumption is very weak: in our context it is equivalent to that there exists at least one continuously distributed random variable in this space (roughly, $(\Omega, \mathcal{A}, \mathbb{P})$ is not a finite space). In particular, it does not prevent discrete variables to exist. In such a probability space, we can generate sequences of independent random vectors with any distribution. We denote by $L^0(\Omega, \mathcal{A}, \mathbb{P})$ the set of all random variables defined in the atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$. See Proposition 6.9 of Delbaen (2002) for details of atomless probability spaces. More discussions on risk measures defined on such spaces can also be found in this paper.

Throughout the paper, we call *aggregate risk* the sum $S = X_1 + \cdots + X_n$ where X_i are non-negative random variables (individual risks) and n is a positive integer. Here the non-negativity is assumed just for the convenience of our discussion.

As already mentioned before, we consider that for each $i = 1, \cdots, n$ the distribution of X_i is known while the joint distribution of $\mathbf{X} := (X_1, X_2, \cdots, X_n)$ is unknown. In other words, marginal distributions of individual risks X_i are given and their dependence structure (copula) is

[2] The TVaR of S at level $p \in [0, 1)$ is defined as $\mathrm{TVaR}_p(S) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_\alpha(S)\mathrm{d}\alpha$, where VaR is the Value-at-Risk measure.

unspecified. To formulate the problem mathematically, define the *Fréchet class* $\mathfrak{F}_n(F_1, \cdots, F_n)$ as the set of random vectors with given marginal distributions $F_1, \cdots, F_n$,

$$\mathfrak{F}_n(F_1, \cdots, F_n) = \{\mathbf{X} : \ X_i \sim F_i, \ i = 1, \cdots, n\},$$

where $X_i \sim F_i$ means that $X_i \in L^0(\Omega, \mathcal{A}, \mathbb{P})$ has the distribution F_i. The Fréchet class is the most natural setup to describe the case when marginal distributions are known and dependence is unspecified. It was used in the literature when studying copulas and dependence in risk management; we refer to recent review papers of Embrechts and Puccetti (2010) and Dhaene et al. (2002). Note that when more information on the dependence is available, the possible aggregate risks belong to subsets of $\mathfrak{F}_n(F_1, \cdots, F_n)$. However, in this paper we study the aggregate risk when marginal distributions are given and the dependence structure is completely unknown, which is called an *admissible risk*.

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

$$\begin{aligned} \mathfrak{S}_n(F_1, \cdots, F_n) &= \{\text{admissible risk of marginal distributions } F_1, \cdots, F_n\} \\ &= \{X_1 + \cdots + X_n : X_i \sim F_i, \ i = 1, \cdots, n\}. \end{aligned}$$

Remark 2.1. It is clear that $\mathfrak{S}_n(F_1, \cdots, F_n) = \{\mathbf{X}\mathbf{1}_n : \mathbf{X} \in \mathfrak{F}_n(F_1, \cdots, F_n)\}$ where $\mathbf{1}_n$ is the column n-vector with all elements being 1. At a first look, one may think the admissible risk class is a trivial reformulation of the Fréchet class. However, the study of an admissible risk class is completely different from the study of a Fréchet class and is of interest in risk aggregation. For example, the structure of an admissible risk depends highly on the marginal distributions, while the structure of every Fréchet class is clearly marginal-independent and is fully characterized. We believe that this difference is exactly why copula techniques work well for the study of Fréchet classes but not the admissible risk class[3]. On the other hand, considering the aggregate risk is a one-dimensional quantity, the Fréchet class contains redundant n-dimensional information which greatly increases the difficulty of statistical inference.

The definition of admissible risks concerns only the distribution of random variables. Note that if S_1 and S_2 have the same distribution, then $S_1 \in \mathfrak{S}_n(F_1, \cdots, F_n) \Leftrightarrow S_2 \in \mathfrak{S}_n(F_1, \cdots, F_n)$ by the atomless property of the probability space (see Theorem 2.1 (i) below). Hence, the study of $\mathfrak{S}_n(F_1, \cdots, F_n)$ is equivalent to the study of the *admissible distribution class* defined as

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

[3]For example, when $n \geqslant 3$, depending on the marginal distributions, the minimal convex ordering element is obtained with different structures or even does not exist.

We first give properties of the admissible risk class $\mathfrak{S}_n(F_1, \cdots, F_n)$ and the corresponding admissible distribution class $\mathfrak{D}_n(F_1, \cdots, F_n)$. For simplicity, we denote by $\mathbf{F} = (F_1, \cdots, F_n)$, $\mathbf{G} = (G_1, \cdots, G_n)$, I_A is the indicator function for the set $A \in \mathcal{A}$, and $T_{a,b}$ is an affine operator on univariate distributions such that for $a, b \in \mathbb{R}$,

$$T_{a,b}(\text{distribution of } X) = \text{distribution of } aX + b.$$

We also use $F \otimes G$ to denote the distribution of $X+Y$ where $X \sim F$ and $Y \sim G$ are independent, i.e. $(F \otimes G)(x) = \int_{-\infty}^{x} F(x-y)\mathrm{d}G(y)$, and use $\overset{d}{=}$ and $\overset{d}{\to}$ to denote equality and convergence in law, respectively. We say that a probability space is rich enough if in this probability space, for any random variable X_1 and any copula C, there exist a random vector $\mathbf{X} = (X_1, \ldots, X_n)$ with copula C.

Theorem 2.1 (Properties of the admissible risk class)**.**

(i) (law invariance) Suppose that the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is rich enough. If $S_1 \overset{d}{=} S_2$, then $S_1 \in \mathfrak{S}_n(\mathbf{F}) \Leftrightarrow S_2 \in \mathfrak{S}_n(\mathbf{F})$.

(ii) (convexity) If $S_1 \in \mathfrak{S}_n(\mathbf{F})$, $S_2 \in \mathfrak{S}_n(\mathbf{G})$, then $\mathrm{I}_A S_1 + (1-\mathrm{I}_A)S_2 \in \mathfrak{S}_n(\mathbb{P}(A)\mathbf{F} + (1-\mathbb{P}(A))\mathbf{G})$ for $A \in \mathcal{A}$ independent of S_1 and S_2. In particular,

- *(a) if $S_1, S_2 \in \mathfrak{S}_n(\mathbf{F})$, then $\mathrm{I}_A S_1 + (1-\mathrm{I}_A)S_2 \in \mathfrak{S}_n(\mathbf{F})$ for $A \in \mathcal{A}$ independent of S_1 and S_2;*
- *(b) if $S \in \mathfrak{S}_n(\mathbf{F}) \cap \mathfrak{S}_n(\mathbf{G})$, then $S \in \mathfrak{S}_n(\lambda\mathbf{F} + (1-\lambda)\mathbf{G})$ for $\lambda \in [0,1]$. That is, $\mathfrak{S}_n(\mathbf{F}) \cap \mathfrak{S}_n(\mathbf{G}) \subset \mathfrak{S}_n(\lambda\mathbf{F} + (1-\lambda)\mathbf{G})$ for $\lambda \in [0,1]$.*

(iii) (independent sum) If $S_1 \in \mathfrak{S}_n(\mathbf{F})$ and $S_2 \in \mathfrak{S}_n(\mathbf{G})$ are independent, then $S_1 + S_2 \in \mathfrak{S}_n(F_1 \otimes G_1, \cdots, F_n \otimes G_n)$.

(iv) (dependent sum) If $S_1 \in \mathfrak{S}_n(\mathbf{F})$ and $S_2 \in \mathfrak{S}_m(\mathbf{G})$, then $S_1 + S_2 \in \mathfrak{S}_{n+m}(F_1, \cdots, F_n, G_1, \cdots, G_m)$.

(v) (affine invariance) $S \in \mathfrak{S}_n(\mathbf{F}) \Leftrightarrow aS + b \in \mathfrak{S}_n(T_{a,b_1}F_1, \cdots, T_{a,b_n}F_n)$ for $a, b_i \in \mathbb{R}$, $i = 1, \cdots, n$ and $b = \sum_{i=1}^n b_i$.

(vi) (permutation invariance) Let σ be an n-permutation, then $\mathfrak{S}_n(\mathbf{F}) = \mathfrak{S}_n(\sigma(\mathbf{F}))$.

(vii) (robustness) If $F_i^{(k)} \to F_i$ pointwise when $k \to +\infty$ and for $i = 1, \cdots, n$, then

- *(a) each $S \in \mathfrak{S}_n(\mathbf{F})$ is the weak limit of a sequence $S_k \in \mathfrak{S}_n(F_1^{(k)}, \cdots, F_n^{(k)})$;*
- *(b) each weakly convergent sequence $S_k \in \mathfrak{S}_n(F_1^{(k)}, \cdots, F_n^{(k)})$ has its weak limit $S \in \mathfrak{S}_n(\mathbf{F})$;*

(c) (completeness) If $S_k \in \mathfrak{S}_n(\mathbf{F})$, $k = 1, 2, \cdots$, and $S_k \xrightarrow{d} S$, then $S \in \mathfrak{S}_n(\mathbf{F})$.

It is well-known that the Fréchet class has similar properties. Hence, a straightforward proof is given in Appendix A.

The above theorem can help to identify possible aggregate risks when the marginal distributions are known. To summarize, the admissible risk class $\mathfrak{S}_n(\mathbf{F})$ has good properties: the corresponding distribution class $\mathfrak{D}_n(\mathbf{F})$ is a convex set; the sums of admissible risks are also in some admissible risk classes; the admissible risk class is affine and permutation invariant with respect to marginal distributions; any admissible risk class is complete. Finally, the admissible risk classes are robust with respect to marginal distribution; if the estimation of marginal distribution is almost accurate, the resulting admissible risk class is also almost accurate.

Remark 2.2. If we naturally extend the univariate-distributional operators: addition ($+$), scaler-multiplication ($\cdot$), convolution-type product ($\otimes$), affine operation ($T_{a,b}$) and convergence ($\to$) to the sets of distributions (element-wise operations), then (ii), (iii), (v), (vi) and (vii) can be written in terms of operations on sets of the form $\mathfrak{D}_n(\cdot)$, as follows

(ii) $\lambda\mathfrak{D}_n(\mathbf{F}) + (1-\lambda)\mathfrak{D}_n(\mathbf{G}) \subset \mathfrak{D}_n(\lambda\mathbf{F} + (1-\lambda)\mathbf{G})$ and $\mathfrak{D}_n(\mathbf{F}) \cap \mathfrak{D}_n(\mathbf{G}) \subset \mathfrak{D}_n(\lambda\mathbf{F} + (1-\lambda)\mathbf{G})$;

(iii) $\mathfrak{D}_n(\mathbf{F}) \otimes \mathfrak{D}_n(\mathbf{G}) \subset \mathfrak{D}_n(F_1 \otimes G_1, \cdots, F_n \otimes G_n)$;

(v) $\mathfrak{D}_n(T_{a,b_1}F_1, \cdots, T_{a,b_n}F_n) = T_{a,b}\mathfrak{D}_n(\mathbf{F})$, where $b = \sum_{i=1}^n b_i$;

(vi) $\mathfrak{D}_n(\mathbf{F}) = \mathfrak{D}_n(\sigma(\mathbf{F}))$.

(vii) $\mathfrak{D}_n(F_1^{(k)}, \cdots, F_n^{(k)}) \to \mathfrak{D}_n(\mathbf{F}) \quad$ if $F_i^{(k)} \to F_i$.

Remark 2.3. The general characterization of an admissible risk class is an open problem. Note that the determination of whether S belongs to $\mathfrak{S}_n(F_1, \cdots, F_n)$ is equivalent to the determination of whether $F, F_1, \cdots, F_n$ are jointly mixable as defined in Wang et al. (2013), where F is the distribution function of $-S$. The research on joint mixability is known to be highly challenging and still limited in the existing literature.

As already mentioned in the introduction, the study of the admissible risk class $\mathfrak{S}_n(\mathbf{F})$ is of interest in risk management and has been studied from different aspects. One of the most important issue is to quantify aggregate risk under extreme scenarios. Fortunately, although a full characterization of the admissible risk class is unavailable, extreme scenarios are mathematically tractable. Note that all admissible risks of given marginal distributions $(F_1, \cdots, F_n)$ have the same mean when it exists, it is thus natural to consider variability in the class. In this paper, we measure variability with convex order and focus on extreme scenarios of risks in $\mathfrak{S}_n(\mathbf{F})$ in the sense of convex order.

3 Convex Ordering Bounds on Admissible risks

3.1 Convex order and known results

Convex order is a preference between two random variables valid for all risk-avoiding investors.

Definition 3.1 (Convex order)**.** Let X and Y two random variables with finite mean. X is smaller than Y *in convex order*, denoted by $X \prec_{\text{cx}} Y$, if for every convex functions f,

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

It immediately follows that $X \prec_{\text{cx}} Y$ implies $\mathbb{E}[X] = \mathbb{E}[Y]$. This order is thus well-adapted to our problem as all variables in $\mathfrak{S}_n(\mathbf{F})$ have the same mean. Note that convex order is an order on 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 closely related to stop-loss order, which is involved in insurance premium calculations. More discussions on stochastic orders on aggregate risks can be found in Müller (1997a,b). From now on, our objective is to find convex ordering bounds for the set $\mathfrak{S}_n(F_1, \cdots, F_n)$. Applications are numerous as it will appear clearly in subsequent sections.

We denote by $G^{-1}(t) = \inf\{x : G(x) \geqslant t\}$ for $t \in (0, 1]$ the pseudo-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 ordering bound in $\mathfrak{S}_n(F_1, \cdots, F_n)$ is $F_1^{-1}(U) + \cdots + F_n^{-1}(U)$ where U is a uniform distribution over the interval $(0, 1)$ (that we write $U \sim \mathcal{U}[0, 1]$). The special scenario $\mathbf{X} = (F_1^{-1}(U), \cdots, F_n^{-1}(U))$ is called the comonotonic dependence scenario (Kaas et al. (2009)). We obtain

$$X_1 + X_2 + \cdots + X_n \prec_{\text{cx}} F_1^{-1}(U) + \cdots + F_n^{-1}(U).$$

We refer to Dhaene et al. (2002) for details on comonotonicity.

The rest of the paper focuses on the much more complex issue consisting of determining the lower bound. When there are only two variables, $n = 2$, the minimum is obtained by the counter-monotonic dependence scenario:

$$F_1^{-1}(U) + F_2^{-1}(1 - U) \prec_{\text{cx}} X_1 + X_2$$

where $U \sim \mathcal{U}[0, 1]$. Proofs for this assertion can be found in Meilijson and Nadas (1979), Tchen (1980) and Rüschendorf (1980, 1983). The sharp lower bound for $n \geqslant 3$ is obtained in Wang and Wang (2011) in the special case when marginal distributions are identical with a monotone density function. In general, the lower bound for $n \geqslant 3$ is unknown. Observe that in the counter-monotonic scenario for $n = 2$, when the risk $F_1^{-1}(U)$ is large (U close to 1), the other risk $F_2^{-1}(1-U)$ is small ($1-U$ close to 0). Intuitively speaking, the optimal dependence structure for a convex ordering lower bound should be concentrated as much as possible, and thus a large loss X_1 must be "compensated" by a small loss X_2. This intuition is going to be extended to the case when there are $n \geqslant 3$ risks in Section 3.3 (in the case of homogeneous risks) and in Section 3.4 (in the case of heterogeneous risks).

First a natural question is the existence of a convex ordering (global) minimal element in an admissible risk class. Since convex order is a partial order, the existence of such minimal element is not trivial. For $n = 1$ and $n = 2$, the minimum exists for any marginal distributions. One may think that a convex ordering minimal element always exists in an admissible risk class also for $n \geqslant 3$; for example, the Rearrangement Algorithm (RA), proposed in Puccetti and Rüschendorf (2012) and improved in Embrechts et al. (2013) and Puccetti (2013), can be used to find a convex ordering minimal element without proving its existence. However, it turns out that the existence of a convex ordering minimal element is generally not guaranteed as shown in the following counterexample.

3.2 Existence of the convex ordering minimal element

Example 3.1. Let F_1 be a discrete distribution on $\{0, 3, 8\}$ with equal probability, F_2 be a discrete distribution on $\{0, 6, 16\}$ with equal probability, and F_3 be a discrete distribution on $\{0, 7, 13\}$ with equal probability. In our example, the sample space are divided into three disjoint subsets A_1, A_2, A_3 with equal probability 1/3. Let $\omega_i \in A_i$, $i = 1, 2, 3$. We verify two scenarios:

(a) Consider first the following dependence structure

$$\begin{pmatrix} X_1(\omega_1) & X_2(\omega_1) & X_3(\omega_1) \\ X_1(\omega_2) & X_2(\omega_2) & X_3(\omega_2) \\ X_1(\omega_3) & X_2(\omega_3) & X_3(\omega_3) \end{pmatrix} = \begin{pmatrix} 3 & 16 & 0 \\ 0 & 6 & 13 \\ 8 & 0 & 7 \end{pmatrix}$$

It is easy to verify that the distribution of X_i is F_i, $i = 1, 2, 3$. The distribution of $X_1+X_2+X_3$ is on $\{19, 19, 15\}$ with equal probability. Thus, $\mathbb{E}[(X_1 + X_2 + X_3 - 19)^+] = 0$.

(b) Consider another dependence structure

$$\begin{pmatrix} X_1(\omega_1) & X_2(\omega_1) & X_3(\omega_1) \\ X_1(\omega_2) & X_2(\omega_2) & X_3(\omega_2) \\ X_1(\omega_3) & X_2(\omega_3) & X_3(\omega_3) \end{pmatrix} = \begin{pmatrix} 0 & 16 & 0 \\ 3 & 0 & 13 \\ 8 & 6 & 7 \end{pmatrix}$$

It is easy to verify that the distribution of X_i is F_i, $i = 1, 2, 3$. The distribution of $X_1+X_2+X_3$ is on $\{16, 16, 21\}$ with equal probability. Thus, $\mathbb{E}[(16 - X_1 - X_2 - X_3)^+] = 0$.

Note that both $g(x) = (x-19)^+$ and $g(x) = (16-x)^+$ are convex functions. Hence, if there exists a convex ordering minimal element S in the admissible risk class $\mathfrak{S}_3(F_1, F_2, F_3)$, it must satisfy $\mathbb{E}[(S-19)^+] = 0$ and $\mathbb{E}[(16-S)^+] = 0$. However, we can see that when $X_1 = 8$, no matter what values of X_2 and X_3 take, S will be either > 19 or < 16. That means $\mathbb{E}[(S-19)^+] = 0$ and $\mathbb{E}[(16-S)^+] = 0$ cannot be satisfied simultaneously by the same $S \in \mathfrak{S}_3(F_1, F_2, F_3)$.

Remark 3.1. The above example shows that the minimal element w.r.t. convex order may not exist in an admissible risk class.

(i) This observation implies that the minimum of $\mathbb{E}[g(S)]$ over $S \in \mathfrak{S}_n(F_1, \cdots, F_n)$ for different convex functions g or for different $\mathrm{TVaR}_\alpha(S)$ with $\alpha \in (0,1)$ may not be obtained with the same dependence structure. Thus, in particular a global solution to the infimum problem

$$\inf_{S \in \mathfrak{S}_n(F_1, \cdots, F_n)} \mathbb{E}[g(S)]$$

for all convex functions g is not available.

(ii) This observation implies that the Rearrangement Algorithm (RA) of Puccetti and Rüschendorf (2012) may not lead to the minimum value of $\mathbb{E}[g(S)]$ or $\mathrm{TVaR}_\alpha(S)$, since different choices of g or α may lead to different optimal structures.

(iii) In what follows, we provide cases of admissible risk class with a minimal element w.r.t. convex order for homogeneous marginal distributions satisfying some conditions. A remaining question is to find for which marginal distributions $F_1, \cdots, F_n$, the admissible risk class must contain a minimum w.r.t. convex order. The question turns out to be non-trivial to answer.

Under some conditions the minimal element exists and can be characterized. We first start with the case when all risks have the same distribution, $F_1 = \cdots = F_n$ (*homogeneous* risks). This case significantly reduces the complexity of the problem and it is still relevant in practice. For example, it is useful for an insurer who has a portfolio of identically distributed policyholders' individual risks. In another context, it can be used to find bounds on prices of variance options when subsequent stocks' log-returns are identically distributed (see Section 5.4). We then generalize the study to the case when the distributions F_i can be different (case of *heterogeneous* risks).

3.3 Convex ordering lower bounds for homogeneous risks

Let F be a distribution on $\mathbb{R}^+$ with finite mean and n a positive integer. Although we are interested in the case of $n \geqslant 3$, but the theorems in this paper also hold for the cases of $n = 1$

and $n = 2$. We first consider the homogeneous case and give a lower convex ordering bound on $\mathfrak{S}_n(F, \cdots, F)$ in Theorems 3.1 and 3.2. Let us define $H(\cdot)$ and $D(\cdot)$ that are two key quantities in the derivation of this lower bound.

$$\forall x \in \left[0, \frac{1}{n}\right], \quad H(x) = (n-1)F^{-1}((n-1)x) + F^{-1}(1-x), \tag{3.1}$$

$$\forall a \in \left[0, \frac{1}{n}\right), \quad D(a) = \frac{n}{1-na}\int_a^{\frac{1}{n}} H(x)\mathrm{d}x = n\,\frac{\int_{(n-1)a}^{1-a} F^{-1}(y)dy}{1-na}. \tag{3.2}$$

We use the convention that $D\left(\frac{1}{n}\right) = H\left(\frac{1}{n}\right)$ and $H(0) = +\infty$ when the support of F is unbounded. The possible infiniteness of $H(0)$ is for convenience only and will not be problematic in what follows. Note also that $D(a)$ is always finite since $\int_a^{\frac{1}{n}} H(x)\mathrm{d}x \leqslant \int_0^{\frac{1}{n}} H(x) = \mathbb{E}[X_1]$ is finite (as F is a distribution with finite mean). Let us give some intuition about these two quantities. From the last expression of $D(a)$, it is clear that $D(a)$ is directly related to the average sum when its components $(X_1, \cdots, X_n)$ are all in the *middle* of the distribution (also called *body* of the distribution). Precisely,

$$D(a) = \sum_{j=1}^{n} \mathbb{E}\left[X_j \;\middle|\; X_j \in \left[F^{-1}((n-1)a), F^{-1}(1-a)\right]\right] \tag{3.3}$$

because $\mathbb{P}\left(X_j \in [F^{-1}((n-1)a), F^{-1}(1-a)]\right) = 1 - na$ and X_1, $X_2, \cdots, X_n$ all have the same distribution. It is also clear that $H(x)$ and $D(a)$ can be easily calculated for a given distribution F.

Intuitively, the dependence scenario to attain the convex ordering lower bound is constructed such that when one of the X_i is large then all the others are small (all X_i are in the *tails* of the distribution; the pair (X_i, X_j) is counter-monotonic for large X_i and $j \neq i$) and when one of the X_i is of medium size (in the *body* of the distribution) we treat the sum $\sum_i X_i$ as a constant equal to its conditional expectation as in (3.3). Precisely, the lower bound in the coming theorem corresponds exactly to the following dependence structure. The probability space is split into two parts: the *tails* (with probability na for a small value of $a \in [0, 1/n]$) and the *body* (with probability $1 - na$). $H(\cdot)$ gives the values of S in the tails and $D(a)$ is the value of S in the body of the distribution. To this end, for $a \in \left[0, \frac{1}{n}\right]$, we introduce a random variable

$$T_a = H(U/n)\mathrm{I}_{\{U \in [0, na]\}} + D(a)\mathrm{I}_{\{U \in (na, 1]\}}, \tag{3.4}$$

where $U \sim \mathcal{U}[0, 1]$. The atomless assumption of the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ allows us to generate such U, and since we only care about distributions to prove convex order, we do not specify the random variable U. In Theorem 3.1, we prove that T_a is a convex ordering lower bound given that $H(\cdot)$ satisfies a monotonicity property. In the proof of Theorem 3.2, we find the best convex ordering bound and exhibit the worst dependence structure explicitly.

Theorem 3.1 (Convex ordering lower bound for homogeneous risks)**.** *Suppose condition (A) holds:*

(A) *for some* $a \in \left[0, \frac{1}{n}\right]$, $H(x)$ *is non-increasing on the interval* $[0, a]$ *and* $\lim_{x \to a-} H(x) \geqslant D(a)$, *then,*

(i) $T_a \prec_{\mathrm{cx}} S$ *for all* $S \in \mathfrak{S}_n(F, \cdots, F)$;

(ii) $T_u \prec_{\mathrm{cx}} T_v$ *for all* $0 \leqslant u \leqslant v \leqslant \frac{1}{n}$. *Thus, the most accurate lower bound is obtained by the largest* a *such that (A) holds.*

Proof. All proofs are trivial for the case of $n = 1$, so we only provide the case for $n \geqslant 2$.

(i) Let $\mathbf{X} \in \mathfrak{F}_n(F, \cdots, F)$, $S = \mathbf{X}\mathbf{1}_n \in \mathfrak{S}_n(F, \cdots, F)$ and T_a be defined in (3.4). It is straightforward to check

$$\begin{aligned}
\mathbb{E}\left[T_a\right] &= \int_0^{na} H(u/n)\mathrm{d}u + (1 - na)D(a) \\
&= n\int_0^{a} H(u)\mathrm{d}u + n\int_a^{\frac{1}{n}} H(u)\mathrm{d}u \\
&= n\int_0^{\frac{1}{n}} \big((n-1)F^{-1}((n-1)u) + F^{-1}(1-u)\big)\,\mathrm{d}u \\
&= n\int_0^{1} F^{-1}(u)\mathrm{d}u \\
&= \mathbb{E}\left[S\right].
\end{aligned}$$

Let F_S and F_{T_a} be the cdf of S and T_a respectively, and further let $U_1, \cdots, U_n$ be $\mathcal{U}\left[0,1\right]$ random variables such that $F^{-1}(U_i) = X_i$ for $i = 1, \cdots, n$. Such $U_1, \cdots, U_n$ always exist in an atomless probability space. Our goal is to show that

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

Property (3.5) together with $\mathbb{E}[T_a] = \mathbb{E}[S]$ is equivalent to $T_a \prec_{\mathrm{cx}} S$ (for example, see Theorem 2.5 of Bäuerle and Müller (2006)).

To obtain this, denote $A_S(u) = \bigcup_i \{U_i > 1 - u\}$ and let $W(u) = \mathbb{P}(A_S(u))$. Obviously $u \leqslant W(u) \leqslant nu$, W is continuous and non-decreasing (as all the U_i are continuously distributed as $\mathcal{U}(0,1)$, hence for $u > s$, $W(u) - W(s) = \mathbb{P}(A_S(u) \setminus A_S(s)) \leqslant n(u - s)$). For $c \in [0, na]$, let $u^\star = W^{-1}(c)$, it then follows that $c \geqslant u^\star \geqslant c/n$ and $\{U_i \in [1 - c/n, 1]\} \subset \{U_i \in [1 - u^\star, 1]\} \subset A_S(u^\star)$. Note that $\mathbb{P}(A_S(u^\star)) = c$, therefore

$$\mathbb{P}\left(A_S(u^\star) \setminus \{U_i \in [1 - c/n, 1]\}\right) = c - c/n = \mathbb{P}\left(U_i \in [0, (n-1)c/n]\right).$$

Since $X_i = F^{-1}(U_i)$ is non-decreasing in U_i and the above two sets have the same measure, we have

$$\mathbb{E}\left[\mathrm{I}_{\{U_i \in [0,(n-1)c/n]\}} X_i\right] \leqslant \mathbb{E}\left[\mathrm{I}_{A_S(u^\star) \setminus \{U_i \in [1-c/n, 1]\}} X_i\right]. \tag{3.6}$$

It follows that

$$\begin{aligned}
\mathbb{E}\left[\mathrm{I}_{\{U\leqslant c\}}T_a\right] &= \mathbb{E}\left[\mathrm{I}_{\{U\leqslant c\}}H(U/n)\right] \\
&= n\int_0^{c/n}\big((n-1)F^{-1}((n-1)x)+F^{-1}(1-x)\big)\,\mathrm{d}x \\
&= n\int_0^{\frac{(n-1)c}{n}}F^{-1}(t)\mathrm{d}t+n\int_{1-c/n}^1 F^{-1}(t)\mathrm{d}t \\
&= n\mathbb{E}\left[(\mathrm{I}_{\{U_i\in[0,(n-1)c/n]\}}+\mathrm{I}_{\{U_i\in[1-c/n,1]\}})X_i\right] \\
&\leqslant n\mathbb{E}\left[(\mathrm{I}_{A_S(u^\star)\setminus\{U_i\in[1-c/n,1]\}}+\mathrm{I}_{\{U_i\in[1-c/n,1]\}})X_i\right]
\end{aligned}$$

where the inequality follows from (3.6). We then find that $\mathbb{E}\left[\mathrm{I}_{\{U\leqslant c\}}T_a\right]\leqslant n\mathbb{E}\left[\mathrm{I}_{A_S(u^\star)}X_i\right]=\mathbb{E}\left[\mathrm{I}_{A_S(u^\star)}S\right]$. Thus we have

$$\mathbb{E}\left[\mathrm{I}_{\{U\leqslant c\}}T_a\right]\leqslant\mathbb{E}\left[\mathrm{I}_{A_S(u^\star)}S\right]. \tag{3.7}$$

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

$$\mathbb{E}\left[\mathrm{I}_{\{U\leqslant c\}}T_a\right]=\mathbb{E}\left[\mathrm{I}_{\{U\leqslant c\}}H(U/n)\right]=\int_{1-c}^1 F_{T_a}^{-1}(t)\mathrm{d}t. \tag{3.8}$$

Also note that

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

since $\mathbb{P}(A_S(u^\star))=c$. It follows from (3.7), (3.8) and (3.9) that for any $c\in[0,na]$,

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

For $x\in[0,1-na]$, let $G(x)=\int_x^1 F_S^{-1}(t)\mathrm{d}t-\int_x^1 F_{T_a}^{-1}(t)\mathrm{d}t$. Note that $\int_x^1 F_S^{-1}(t)\mathrm{d}t$ is concave, and $F_{T_a}^{-1}(t)=D(a)$ is a constant when $t\in[0,1-na)$, hence $G(x)$ is concave over $[0,1-na)$. Since G is concave, $G(0)=\mathbb{E}[S]-\mathbb{E}[T_a]=0$, and $G(1-na)\geqslant 0$ by (3.10), we have $G(x)\geqslant 0$ over $[0,1-na]$. Thus

$$\int_c^1 F_{T_a}^{-1}(t)\mathrm{d}t\leqslant\int_c^1 F_S^{-1}(t)\mathrm{d}t \tag{3.11}$$

for any $c\in[0,1]$. This implies $T_a\prec_{\mathrm{cx}}S$.

(ii) For $0\leqslant u\leqslant v\leqslant\frac{1}{n}$, it can be easily checked that the distribution of T_u is a fusion of the distribution of T_v, and thus $T_u\prec_{\mathrm{cx}}T_v$ (see Theorem 2.8 of Bäuerle and Müller (2006) for the definition of a fusion and a proof of this assertion).

□

Remark 3.2. We formulate the two following observations:

1. Assumption (A) in Theorem 3.1 is often verified in practice. Notice indeed that $H(0) = +\infty$ when the support of F is unbounded. For such distributions F, it is reasonable to assume that $H(x)$ is non-decreasing over some interval $[0, a]$ as it will be illustrated with examples later.

2. For a bounded distribution F, if $H(0) < D(0)$, we have $a = 0$ and $T_0 = \mathbb{E}[S]$.

3. It is straightforward to check that if (A) holds, then the distribution function of T_a is

$$F_{T_a}(t) = \mathbb{P}(T_a \leqslant t) = (1 - nH^{-1}(t))\mathrm{I}_{\{t \geqslant D_a\}}. \tag{3.12}$$

Theorem 3.1 (ii) shows that the most accurate convex ordering lower bound over $\mathfrak{S}_n(F, \cdots, F)$ is obtained by T_a with the largest possible a. The next theorem characterizes the sharpness of this bound, which is closely connected with the concept of Complete Mixability (CM) introduced by Wang and Wang (2011).

Definition 3.2 (Complete Mixability)**.** A distribution function F on $\mathbb{R}$ is *n-completely mixable* (n-CM) if there exist n random variables $X_1, \ldots, X_n$ identically distributed as F such that

$$X_1 + \cdots + X_n = n\mu \tag{3.13}$$

for some $\mu \in \mathbb{R}$ called a *center* of F. A distribution function F on $\mathbb{R}$ is called *n-CM on an interval I* (finite or infinite) if the conditional distribution of F on I is n-CM.

As F has finite mean, if F is n-CM, then its center is unique and equal to the mean. Note that F is n-CM is equivalent to $n\mathbb{E}[X] \in \mathfrak{S}_n(F, \cdots, F)$, where $X \sim F$. Some straightforward examples and properties of completely mixable distributions are given in Wang and Wang (2011) and Puccetti et al. (2012). By Theorem 3.1, one needs to find the largest possible a to get the most accurate lower bound. This motivates us to define c_n by

$$c_n = \inf\left\{c \in \left[0, \frac{1}{n}\right] : H(c) \leqslant D(c)\right\}. \tag{3.14}$$

Note that c_n is the largest possible a satisfying $\lim_{x \to a-} H(x) \geqslant D(a)$. When F is a continuous distribution, $H(c_n) = D(c_n)$. On the other hand, c_n is exactly the smallest possible a such that F on $I = \left[F^{-1}((n-1)a), F^{-1}(1-a)\right]$ satisfies the mean condition necessary for the CM property. See, for example, (7) in Proposition 2.1 of Wang and Wang (2011) for more details on this condition.

Theorem 3.2 (Sharp convex ordering lower bound for homogeneous risks)**.** *Suppose*
(A') *$H(x)$ is non-increasing on the interval $[0, c_n]$, where c_n is given by* (3.14)
then $T_{c_n} \prec_{\mathrm{cx}} S$ for all $S \in \mathfrak{S}_n(F, \cdots, F)$. Moreover, $T_{c_n} \in \mathfrak{S}_n(F, \cdots, F)$ that is T_{c_n} is sharp if

(B) holds:

(B) *F is n-CM on the interval* $I = \left[F^{-1}((n-1)c_n), F^{-1}(1-c_n)\right]$.

Proof. $T_{c_n} \prec_{\mathrm{cx}} S$ follows from Theorem 3.1 by noting that $\lim_{x\to c_n-} H(x) \geqslant D(c_n)$ from the definition of c_n in (3.14). Let us prove the second half of the theorem. When condition (B) holds, that is F is n-CM on I, there exist random variables $Y_1, \cdots, Y_n$ from the conditional distribution F on I such that $Y_1 + \cdots + Y_n$ is a constant. Thus, as Y has finite mean (because F has finite mean), $Y_1 + \cdots + Y_n = n\mathbb{E}(Y_1) = D(c_n)$ by (3.2) and (3.3). Now we construct $S \in \mathfrak{S}_n(F_1, \cdots, F_n)$ which has the same distribution as T_{c_n}, by imposing a special dependence structure. For each i, when $X_i \in I$ (the *body* part), we let $X_i = Y_i$ and when $X_i \notin I$ (the *tail* part), we let (X_i, X_j) be counter-monotonic for each $j \neq i$. That is,

$$X_i = \mathrm{I}_{\{U > nc_n\}} Y_i + \mathrm{I}_{\{U \leqslant nc_n\}} F^{-1}(V_i), \tag{3.15}$$

where $U \sim \mathcal{U}[0,1]$, $(V_1, \cdots, V_n)$ is independent of U and uniformly distributed on the line segments

$$\mathcal{O} = \bigcup_{k=1}^{n} \{(v_1, \cdots, v_n) : v_j = (n-1)(1-v_k),\ v_k \in [1-c_n, 1],\ j = 1, \cdots, n,\ j \neq k\}. \tag{3.16}$$

We can check that V_i is uniformly distributed on $[0, (n-1)c_n] \cup [1-c_n, 1]$, and thus the distribution of $F^{-1}(V_i)$ is the conditional distribution of F on $\mathbb{R}^+ \setminus I$. Moreover by construction, Y_i has the conditional distribution of F on I. It follows that $X_i \sim F$. Then

$$\begin{aligned} S &= \sum_{i=1}^{n} \left(\mathrm{I}_{\{U > nc_n\}} Y_i + \mathrm{I}_{\{U \leqslant nc_n\}} F^{-1}(V_i)\right) \\ &= \mathrm{I}_{\{U > nc_n\}} D(c_n) + \mathrm{I}_{\{U \leqslant nc_n\}} \sum_{i=1}^{n} F^{-1}(V_i). \end{aligned}$$

Note that

$$\sum_{i=1}^{n} F^{-1}(V_i) = \sum_{i=1}^{n} \mathrm{I}_{\{V_i \geqslant 1-c_n\}} (F^{-1}((n-1)(1-V_i)) + F^{-1}(V_i)) = \sum_{i=1}^{n} \mathrm{I}_{\{V_i \geqslant 1-c_n\}} H(1-V_i),$$

and for $t > 0$,

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{n} F^{-1}(V_i) \leqslant t\right) &= \mathbb{P}\left(\sum_{i=1}^{n} \mathrm{I}_{\{V_i \geqslant 1-c_n\}} H(1-V_i) \leqslant t\right) \\ &= \mathbb{E}\left(\sum_{i=1}^{n} \mathrm{I}_{\{V_i \geqslant 1-c_n\}} \mathbb{P}(H(1-V_i) \leqslant t | V_i \geqslant 1-c_n)\right) \\ &= \mathbb{P}(H(1-V_1) \leqslant t | V_1 \geqslant 1-c_n) \\ &= \mathbb{P}\left(H(V) \leqslant t\right) \end{aligned}$$

for some $V \sim \mathcal{U}[0, c_n]$, independent of U. Note that the second equality holds because $\{V_i \geqslant 1-c_n\}$ are mutually exclusive. Therefore, $S \overset{d}{=} \mathrm{I}_{\{U > nc_n\}} D(c_n) + \mathrm{I}_{\{U \leqslant nc_n\}} H(V) \overset{d}{=} T_{c_n}$, and thus $T_{c_n} \in \mathfrak{S}_n(F, \cdots, F)$. $\square$

Theorem 3.3 (Necessity and sufficiency of condition (B)). *Suppose $H(x)$ is strictly decreasing on $[0, c_n]$, then*

(i) $T_{c_n} \in \mathfrak{S}_n(F, \cdots, F)$ *if and only if (B) holds.*

(ii) $T_a \notin \mathfrak{S}_n(F, \cdots, F)$ *for all* $a < c_n$*.*

Proof. (i) The "$\Leftarrow$" part follows directly from Theorem 3.2. Let us show the "$\Rightarrow$" part. We begin by showing this assertion in the discrete case. Let F be any continuous distribution on $\mathbb{R}^+$, with F^{-1} strictly increasing. Let G be the distribution of $F^{-1}(V)$ where V is a discrete uniform distribution on $\{0, \frac{1}{K}, \cdots, \frac{K-1}{K}\}$ for some large number $K > n$ and let $\hat{T}_{c_n}$ be defined as T_{c_n} with F replaced by G:

$$\hat{T}_{c_n} = \hat{H}(U/n)\mathrm{I}_{\{U \in [0, nc_n]\}} + \hat{D}\mathrm{I}_{\{U \in (nc_n, 1]\}}, \tag{3.17}$$

where $\hat{H}(x) = (n-1)G^{-1}((n-1)x) + G^{-1}(1-x)$, $U \sim \mathcal{U}[0,1]$,

$$c_n = \inf\left\{c \in \left\{0, \frac{1}{K}, \cdots, \left\lfloor\frac{K}{n}\right\rfloor\frac{1}{K}\right\} : \hat{H}(c) \leqslant \frac{n}{1-nc}\int_c^{\frac{1}{n}} \hat{H}(x)\mathrm{d}x\right\},$$

and $\hat{D} := \frac{n}{1-nc_n}\int_{c_n}^{\frac{1}{n}} \hat{H}(x)\mathrm{d}x$ is a constant.

Note that $G^{-1}(t) = F^{-1}(t)$ for $t = 0, \frac{1}{K}, \cdots, \frac{K-1}{K}$, and $G^{-1}(x) = F^{-1}(\frac{\lfloor xK \rfloor}{K})$ for $x \in [0,1)$. Thus, $H(t) = \hat{H}(t)$ for $t = 0, \frac{1}{K}, \cdots, \frac{K-1}{K}$, and the interval $I = \left[G^{-1}((n-1)c_n), G^{-1}(1-c_n)\right]$. Note that since G is discrete, this function $\hat{H}$ is not non-increasing, but this would not hurt our proof since we are not using the results in convex order. To simulate the strict decreasing property, we assume

$$\min_{i \leqslant x < i+1} \hat{H}\left(\frac{x}{K}\right) > \max_{i+1 \leqslant x < i+2} \hat{H}\left(\frac{x}{K}\right) \quad \text{for } i = 0, \cdots, Kc_n - 2. \tag{3.18}$$

Suppose $\hat{T}_{c_n} = \mathbf{X}\mathbf{1}_n \in \mathfrak{S}_n(G, \cdots, G)$ for $\mathbf{X} \in \mathfrak{F}_n(G, \cdots, G)$. Let us show that this implies G is n-CM on I. Note that by definition of $\hat{T}_{c_n}$ and (3.18),

$$\mathbb{P}\left[\hat{T}_{c_n} - G^{-1}\left(1 - \frac{1}{K}\right) \in \left[(n-1)G^{-1}(0), (n-1)G^{-1}\left((n-1)\frac{1}{K}\right)\right)\right] = \frac{n}{K},$$

and

$$\mathbb{P}\left[\hat{T}_{c_n} > (n-1)G^{-1}\left((n-2)\frac{1}{K}\right) + G^{-1}\left(1 - \frac{1}{K}\right)\right] = 0.$$

This implies that when one of X_i takes the value $G^{-1}\left(1 - \frac{1}{K}\right)$, all the others must take values in $\left[G^{-1}(0), G^{-1}\left((n-1)\frac{1}{K}\right)\right)$, by observing that $G^{-1}(x)$ is strictly increasing. Using this argument again, we obtain that when one of X_i takes the value $G^{-1}\left(1 - \frac{2}{K}\right)$, all the others must take values in $\left[G^{-1}\left((n-1)\frac{1}{K}\right), G^{-1}\left((n-1)\frac{2}{K}\right)\right)$. Eventually, we have that for all $1 \leqslant j < Kc_n$, when X_i takes the value $G^{-1}\left(1-\frac{j}{K}\right)$, all the others must take values in $\left[G^{-1}\left((n-1)\frac{j-1}{K}\right), G^{-1}\left((n-1)\frac{j}{K}\right)\right)$. The remaining part is

$$\mathbb{P}\left[\hat{T}_{c_n} = \hat{D}\right] = 1 - nc_n.$$

Let $A = \{\hat{T}_{c_n} = \hat{D}\}$. The conditional distribution of X_i on A is exactly the conditional distribution G on I, since $\{X_i \notin I\}$ has been contained in the set A^c. Since $\hat{T}_{c_n}$ is a constant on A, we have G is n-CM on I. The above proof shows that for a discrete distribution G, if G^{-1} is strictly increasing and $\hat{H}$ satisfies (3.18), then T_{c_n} is admissible implies that the conditional distribution is n-CM on I. To prove the case of F being continuous, we can simply replace $\frac{1}{K}$ by an infinitesimal dt, and the condition (3.18) becomes that H is strictly decreasing. Note that H being strictly increasing is sufficient for F^{-1} to be strictly increasing on $[1-nc_n, 1]$, which is sufficient for our proof.

(ii) By (3.2), we know $D(a)$ is a strictly decreasing function of a. Suppose $a < c_n$ and let $c = \frac{1}{2}a + \frac{1}{2}c_n$, then $c < \frac{1}{n}$ and $D(a) > D(c)$. It is straightforward to check that

$$\mathbb{E}[(T_a - D(a))^+] = \mathbb{E}[T_a] - D(a) = \mathbb{E}[T_c] - D(a) < \mathbb{E}[(T_c - D(a))^+]$$

since $\mathbb{P}(T_c < D(a)) \geqslant \mathbb{P}(T_c = D(c)) \geqslant 1 - nc > 0$. This shows $T_c \not\prec_{\text{cx}} T_a$ by the definition of convex order. Since $c < c_n$, we have $H(c) \geqslant D(c)$, and by Theorem 3.1 $T_c \prec_{\text{cx}} S$ for any $S \in \mathfrak{S}_n(F, \cdots, F)$. Thus we conclude that $T_a \notin \mathfrak{S}_n(F, \cdots, F)$ for $a < c_n$.

□

To summarize, using our result together with the comonotonic upper bound, sharp convex ordering bounds for $S \in \mathfrak{S}_n(F, \cdots, F)$ are given by

$$T_{c_n} \prec_{\text{cx}} S \prec_{\text{cx}} nF^{-1}(U)$$

where $U \sim \mathcal{U}[0,1]$. $nF^{-1}(U)$ is always admissible, and when (A') and (B) hold, T_{c_n} is also admissible. Thus the upper and lower bounds are both sharp.

Recall that in the proof of Theorem 3.2, an optimal structure for $\mathbf{X}$ to attain $X_1 + \cdots + X_n \overset{d}{=} T_{c_n}$ can be described as follows. The probability space are divided into two parts:

- For each j, if X_j is large, then each X_i, $i \neq j$ is small and (X_i, X_j) is counter-monotonic. This part has probability nc_n.

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

From the proof of Theorem 3.3, we can see that when $H(x)$ is strictly decreasing on $[0, c_n]$, the optimal structure can only be the one described above. It may be interesting to point out that the well-known counter-monotonic scenario, which only exists for $n = 2$, is a special case of the above optimal structure (as we have seen, on the set $\{X_1 > F^{-1}(c_n)\}$, X_1 and X_2 are counter-monotonic; on the set $\{X_1 \leqslant F^{-1}(c_n)\}$, S is a constant, which implies X_1 and X_2 are also counter-monotonic). Also the copula Q_n^F defined in Wang and Wang (2011) and Wang et al. (2013) for F with a decreasing density is also a special case of the above optimal structure.

One may wonder how the conditions in Theorems 3.1 and 3.2 are satisfied by commonly used distributions of risks. The following proposition gives some criteria to verify properties (A), (A') and (B) theoretically.

Proposition 3.4 (Criteria for properties (A), (A') and (B)).

(i) If $H(x)$ is convex, then it is non-increasing on $[0, c_n]$ and (A') holds.

(ii) If F admits a non-increasing conditional density on $I = \left[F^{-1}((n-1)c_n), F^{-1}(1-c_n)\right]$, then (B) holds.

(iii) If F admits a non-increasing density, then (A') and (B) hold.

(iv) If F^{-1} is C^2 over $[0,1)$ and $\lim_{x\to 1+} \left[F^{-1}\right]'(x) = +\infty$, then $H(x)$ is decreasing in a neighborhood of 0 (and therefore (A) holds for small value of a).

Proof. (i) Since $\lim_{x\to c_n+} H(x)$ is not less than the average of $H(x)$ on $[c_n, \frac{1}{n}]$, $H(x)$ is non-increasing in a left neighborhood of c_n. This means $H(x)$ is non-increasing on $[0, c_n]$ by the convexity of $H(\cdot)$.

(ii) This follows from Corollary 2.9 in Wang and Wang (2011).

(iii) With F being concave, F^{-1} is convex and $H(x)$ is convex. The conclusion follows from (i) and (ii). Theorem 3.2 is thus a generalization of Theorem 3.4 in Wang and Wang (2011).

(iv) We have that

$$H'(x) \quad = \quad (n-1)^2 \left[F^{-1}\right]'((n-1)x) - \left[F^{-1}\right]'(1-x)$$

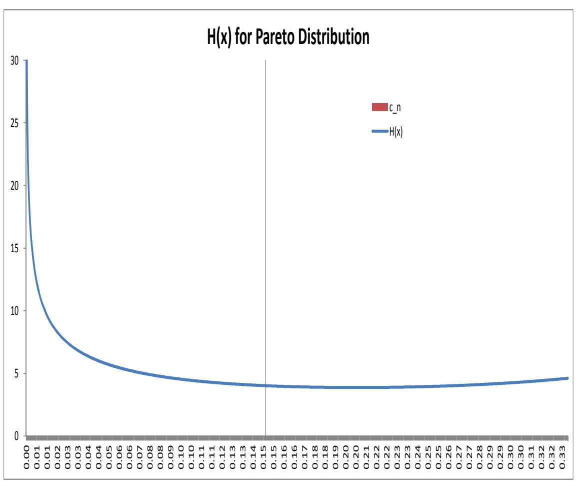
For $x \to 0$, $\left[F^{-1}\right]'(1-x)$ tends to $+\infty$, and $(n-1)^2 \left[F^{-1}\right]'((n-1)x)$ is bounded (since F^{-1} is C^2 over $[0,1)$). Therefore $H'(x) \leqslant 0$ in a neighborhood of 0.

□

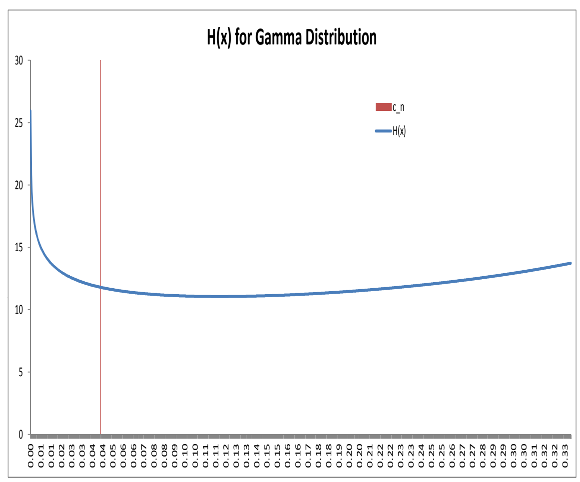
Example 3.2. For a uniform distribution F, although $H(x)$ is in fact increasing, we have $c_n = 0$ and (A') and (B) are satisfied, as the density function of F is a non-increasing function. In that case, $T_0 = \mathbb{E}[S]$ is the sharp convex ordering lower bound over $\mathfrak{S}_n(F, \cdots, F)$. This is because $\mathcal{U}[0,1]$ is n-CM.

Example 3.3. Pareto and Gamma distributions (with shape parameter $\alpha \leqslant 1$) satisfy (A') and (B) since they have non-increasing densities.

Numerical results suggest that most commonly used distributions, such as Pareto, Log-Normal and Gamma distributions satisfy (A') and (B). In practice, it is simple to check condition (A') for a given F either theoretically or numerically. For example, in the case of the Pareto, Gamma or Log-Normal distributions, we represent the function $H(x)$ in Figure 1 and observe immediately that $H(x)$ first decreases, passes the point c_n and then increases (moreover, $H(x)$ is likely to be convex). The vertical lines in Figure 1 display c_n and it is clear that $H(x)$ is decreasing on $[0, c_n]$ in the four panels. In general, condition (B) in Theorem 3.2 is theoretically difficult to prove, except for several known classes given in Wang and Wang (2011) and Puccetti et al. (2012). We propose a numerical technique to check this condition in Section 6.



Panel A: $H(x)$ for Pareto(1,3), $n = 3$.

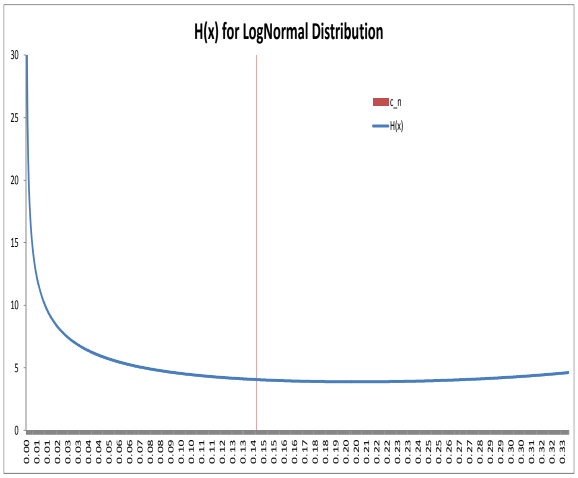


Panel B: $H(x)$ for Gamma(2,0.5), $n = 3$.

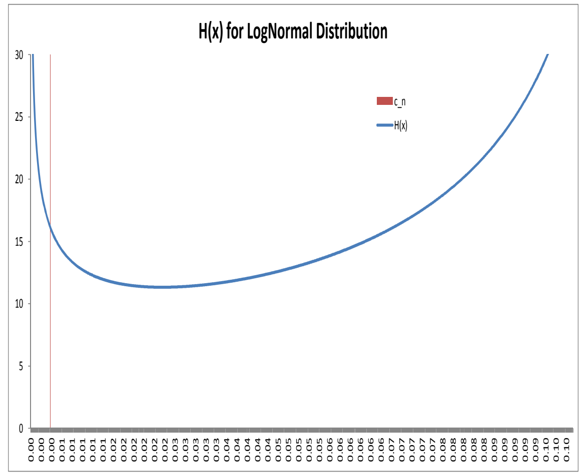
3.4 Convex ordering lower bounds for heterogeneous risks

In this section we give a lower bound for heterogeneous risks based on Theorem 3.1. Consider n distributions $F_1, F_2, \cdots, F_n$ on $\mathbb{R}^+$ with finite mean. We here look for a convex ordering lower bound over the heterogeneous set $\mathfrak{S}_n(F_1, F_2 \cdots, F_n)$ of admissible risks. This result is then illustrated by numerical examples in Section 6.2. Let

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



Panel C: $H(x)$ for Log-Normal(0,1), $n = 3$.



Panel D: $H(x)$ for Log-Normal(0,1), $n = 10$.

Figure 1: $H(x)$ as a function of x for the Pareto, Gamma and LogNormal distributions.

and $H(\cdot)$ and $D(\cdot)$ be defined by (3.1) and (3.2) similarly as in the previous section for homogeneous risks. Next, define $T_a = H(U/n)\mathrm{I}_{\{U\in[0,na]\}} + D(a)\mathrm{I}_{\{U\in(na,1]\}}$ as in (3.4), and we use the same condition (A) as in Theorem 3.1 for $H(\cdot)$ and $D(\cdot)$.

We have the following theorem as a generalization of Theorem 3.1.

Theorem 3.5 (Convex ordering lower bound for heterogeneous risks)**.**

(i) $\mathfrak{S}_n(F_1, \cdots, F_n) \subset \mathfrak{S}_n(F, \cdots, F)$.

(ii) Suppose (A) holds, then $T_a \prec_{\text{cx}} S$ *for all* $S \in \mathfrak{S}_n(F_1, \cdots, F_n)$.

Proof. (i) Let σ_k, $k = 1, 2, \cdots, n!$ be all different n-permutations. By Theorem 2.1 (i)(b) and (iv), we have

$$\mathfrak{S}_n(F_1, \cdots, F_n) = \bigcap_{k=1}^{n!} \mathfrak{S}_n(\sigma_k(F_1, \cdots, F_n)) \subset \mathfrak{S}_n\left(\sum_{k=1}^{n!} \lambda_k \sigma(F_1, \cdots, F_n)\right),$$

where $\lambda_k \geqslant 0$, $k = 1, 2, \cdots, n!$ and $\sum_{k=1}^{n!} \lambda_k = 1$. Take $\lambda_k = \frac{1}{n!}$ for all k then we get $\mathfrak{S}_n(F_1, \cdots, F_n) \subset \mathfrak{S}_n(F, \cdots, F)$.

(ii) By Theorem 3.1 and (i), $T_a \prec_{\text{cx}} S$ for all $S \in \mathfrak{S}_n(F, \cdots, F)$, and hence $T_a \prec_{\text{cx}} S$ for all $S \in \mathfrak{S}_n(F_1, \cdots, F_n)$.

□

Remark 3.3. Theorem 3.5 (i) states that if a risk is admissible of marginal distributions $F_1, \cdots, F_n$, then it is admissible of marginal distributions $F, \cdots, F$. Unlike the bound in Theorem 3.2, the sharpness of the bound in Theorem 3.5 (ii) is difficult to characterize. In general, the set $\mathfrak{S}_n(F_1, \cdots, F_n)$ may be a proper subset of $\mathfrak{S}_n(F, \cdots, F)$, and hence T_{c_n} may belong to $\mathfrak{S}_n(F, \cdots, F)$ but may not belong to $\mathfrak{S}_n(F_1, \cdots, F_n)$. Numerical evidence of this strict inclusion is given in Section 6, where we also observe that the bound tends to be more precise when the marginal distributions F_1, F_2,...,F_n are similar.

4 Equivalence of Convex Ordering and Value-at-Risk Bounds

The convex ordering lower bound in an admissible risk class can directly be used to derive the lower bound on any convex functional and examples include TVaR and will be studied in Section 5. It turns out that convex ordering bounds are also closely related to VaR bounds. The key result of this section is the equivalence between the convex ordering lower bound and the bounds on the Value-at-Risk (VaR) and is given in Corollary 4.7. The intuition of Corollary 4.7 below has already been observed recently, such as in the numerical algorithm for bounds on VaR in Embrechts et al. (2013). However, to our knowledge, there is so far no rigorous proof of this conclusion. We note that some lemmas used in this section are obtained in literature in other forms within different contexts. Here for the sake of completeness we provide all the proofs.

Throughout, we denote by $L^0(\Omega, \mathcal{A}, \mathbb{P})$ the set of random variables in the atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Traditionally, the Value-at-Risk of level p is defined as

$$\mathrm{VaR}_p(X) = \inf\{x : \mathbb{P}(X \leqslant x) \geqslant p\}, \qquad p \in (0, 1), \tag{4.1}$$

and the upper Value-at-Risk of level p is defined as

$$\mathrm{VaR}^*_p(X) = \inf\{x : \mathbb{P}(X \leqslant x) > p\}, \qquad p \in (0, 1).$$

Both definitions are needed at this moment for a mathematical discussion. However Corollary 4.7 clarifies the relationship.

We first show the existence of extreme elements in an admissible risk class with respect to VaR_p and VaR^*_p.

Lemma 4.1. $\mathrm{VaR}^*_p : L^0(\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ *is an upper semi-continuous function and* $\mathrm{VaR}_p : L^0(\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ *is a lower semi-continuous function.*

Proof. Suppose random variables $S_k \to S_0$ weakly. Denote the distribution function of S_k by F_k for $k = 0, 1, \cdots$. By definition $\mathrm{VaR}^*_p(S_k) = \inf\{x : F_k(x) > p\}$ and $\mathrm{VaR}_p(S_k) = \inf\{x : F_k(x) \geqslant p\}$ for all $k = 0, 1, \cdots$. Since $G(F) := \inf\{x : F(x) > p\}$ is an upper semi-continuous function of F, $H(F) := \inf\{x : F(x) \geqslant p\}$ is a lower semi-continuous function of F, and $F_k \to F_0$ weakly, we have $\limsup_{k\to\infty} \mathrm{VaR}^*_p(S_k) \leqslant \mathrm{VaR}^*_p(S_0)$ and $\liminf_{k\to\infty} \mathrm{VaR}_p(S_k) \geqslant \mathrm{VaR}_p(S_0)$. □

In the following, we denote

$$\overline{\mathrm{VaR}_p^*} = \sup_{S \in \mathfrak{S}_n(F_1,\cdots,F_n)} \mathrm{VaR}_p^*(S)$$

and

$$\overline{\mathrm{VaR}_p} = \sup_{S \in \mathfrak{S}_n(F_1,\cdots,F_n)} \mathrm{VaR}_p(S).$$

Lemma 4.2. *There exists* $T \in \mathfrak{S}_n(F_1,\cdots,F_n)$ *such that* $\mathrm{VaR}_p^*(T) = \sup_{S \in \mathfrak{S}_n(F_1,\cdots,F_n)} \mathrm{VaR}_p^*(S)$, *and there exists* $T \in \mathfrak{S}_n(F_1,\cdots,F_n)$ *such that* $\mathrm{VaR}_p(T) = \inf_{S \in \mathfrak{S}_n(F_1,\cdots,F_n)} \mathrm{VaR}_p(S)$.

Proof. By definition, there exists a sequence of $T_k \in \mathfrak{S}_n(F_1,\cdots,F_n)$, $k = 1,2,\cdots$, such that $\mathrm{VaR}_p^*(T_k) \to \overline{\mathrm{VaR}_p^*}$. First we note that any admissible risk class is a complete set (see Section 2). Thus, there is a subsequence of $\{T_k\}$ which converges to an admissible risk $T \in \mathfrak{S}_n(F_1,\cdots,F_n)$. We only need to show that $\mathrm{VaR}_p^*(T) \geqslant \limsup_{k\to\infty} \mathrm{VaR}_p^*(T_k) = \overline{\mathrm{VaR}_p^*}$. This is directly implied by the upper semi-continuity of VaR_p^*. The case for the infimum of VaR is similar. □

In the following we suppose $F_1,\cdots,F_n$ are continuous cdf. We denote $F_{i,p}$ for $p \in (0,1)$ as the conditional distribution of F_i on $[F_i^{-1}(p),\infty)$ (upper tail), and F_i^p for $p \in (0,1)$ as the conditional distribution of F_i on $(-\infty, F_i^{-1}(p))$ (lower tail). We denote the (essential) supremum and infimum of the range of a random variable X by $\sup X$ and $\inf X$, respectively. The next lemma is intuitive and formalizes the fact that the supremum for VaR is obtained by only studying the distributions beyond the VaR. This is used for example in the algorithm presented in Embrechts et al. (2013).

Lemma 4.3. *Suppose* $F_1,\cdots,F_n$ *are continuous. For* $p \in (0,1)$,

$$\sup_{S \in \mathfrak{S}_n(F_1,\cdots,F_n)} \mathrm{VaR}_p^*(S) = \sup\{\inf S : S \in \mathfrak{S}_n(F_{1,p},\cdots,F_{n,p})\}.$$

Proof. By Lemma 2.1 in Wang et al. (2013), there exists a vector $\mathbf{X} = (X_1,\cdots,X_n) \in \mathfrak{F}_n(F_1,\cdots,F_n)$ such that $T = X_1 + \cdots + X_n$, $\{T \geqslant \mathrm{VaR}_p^*(T)\} = \{X_i \geqslant F_i^{-1}(a)\}$ for each $i = 1,\cdots,n$ and $\mathrm{VaR}_p^*(T) = \sup_{S \in \mathfrak{S}_n(F_1,\cdots,F_n)} \mathrm{VaR}_p^*(S)$. Note that there exists $S_p \in \mathfrak{S}_n(F_{1,p},\cdots,F_{n,p})$ such that $\inf S_p = \sup\{\inf S : S \in \mathfrak{S}_n(F_{1,p},\cdots,F_{n,p})\}$, by the same argument as in the proof of Lemma 4.2.

Define $Z = T\mathrm{I}_{\{T<\mathrm{VaR}_p^*(T)\}} + S_p\mathrm{I}_{\{T\geqslant \mathrm{VaR}_p^*(T)\}}$. It is easy to see that $Z \in \mathfrak{S}_n(F_1,\cdots,F_n)$ and $\mathrm{VaR}_p^*(Z) = \inf S_p$. On the other hand, let T' have the conditional distribution of T on the set $\{T \geqslant \mathrm{VaR}_p^*(T)\}$. It follows that $T' \in \mathfrak{S}_n(F_{1,p},\cdots,F_{n,p})$. Hence $\mathrm{VaR}_p^*(T) = \inf T' \leqslant \inf S_p$ by the definition of S_p. In conclusion, $\mathrm{VaR}_p^*(T) \leqslant \inf S_p = \mathrm{VaR}_p^*(Z) \leqslant \mathrm{VaR}_p^*(T)$ since $\mathrm{VaR}_p^*(T) = \overline{\mathrm{VaR}_p^*}$, and hence $\overline{\mathrm{VaR}_p^*} = \inf S_p$. □

Lemma 4.4. *If* F_i^{-1}, $i = 1,\cdots,n$ *are continuous, then* $\overline{\mathrm{VaR}_p^*}$ *is a continuous function of* $p \in (0,1)$.

Proof. Take $(Y_1, \cdots, Y_n) \in \mathfrak{F}_n(F_{1,p}, \cdots, F_{n,p})$ such that $\inf(Y_1 + \cdots + Y_n) = \sup\{\inf S : S \in \mathfrak{S}_n(F_{1,p}, \cdots, F_{n,p})\}$. This is always possible by Lemma 4.2. Write $T_1 = Y_1 + \cdots + Y_n$ and let $U_i = F_{i,p}(Y_i)$, $i = 1, \cdots, n$. For $\varepsilon \in (0, p)$, we take $T_2 = \sum_{i=1}^n F_{i,p-\varepsilon}^{-1}(U_i) \in \mathfrak{S}_n(F_{1,p-\varepsilon}, \cdots, F_{n,p-\varepsilon})$. Since $F_{1,p-\varepsilon}^{-1}, \cdots, F_{n,p-\varepsilon}^{-1}$ are continuous and bounded from below, we can conclude that $\inf T_1 - \inf T_2 \to 0$ as $\varepsilon \searrow 0$. Thus, $\overline{\mathrm{VaR}_{p-\varepsilon}^*} \geqslant \inf(Z_1 + \cdots + Z_n) \to \inf(Y_1 + \cdots + Y_n) = \overline{\mathrm{VaR}_p^*}$ as $\varepsilon \searrow 0$. $\square$

The next lemma shows that the supremum (infimum) of VaR_p^* and VaR_p are the same in an admissible risk class (although they may not be both attained).

Lemma 4.5. *Suppose F_i has positive density on its support, $i = 1, \cdots, n$. For $p \in (0, 1)$,*

$$\sup_{S \in \mathfrak{S}_n(F_1, \cdots, F_n)} \mathrm{VaR}_p(S) = \sup_{S \in \mathfrak{S}_n(F_1, \cdots, F_n)} \mathrm{VaR}_p^*(S), \quad \inf_{S \in \mathfrak{S}_n(F_1, \cdots, F_n)} \mathrm{VaR}_p(S) = \inf_{S \in \mathfrak{S}_n(F_1, \cdots, F_n)} \mathrm{VaR}_p^*(S).$$

Proof. By definition, $\overline{\mathrm{VaR}_{p-\varepsilon}^*} \leqslant \overline{\mathrm{VaR}_p} \leqslant \overline{\mathrm{VaR}_p^*}$. Since F_i has positive density and F_i^{-1} is continuous, VaR_p^* is continuous and we have $\overline{\mathrm{VaR}_p} = \overline{\mathrm{VaR}_p^*}$. The proof for the infimum is the same. $\square$

Lemma 4.5 confirms that we do not need to distinguish between VaR_p and VaR_p^* when we study the supremum and infimum over an admissible risk class. In summary, the following theorem is now straightforward and it is obtained by combining the results of Lemmas 4.3 and 4.5.

Theorem 4.6. *Suppose F_i has positive density on its support, $i = 1, \cdots, n$, then*

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

and

$$\inf_{S \in \mathfrak{S}_n(F_1, \cdots, F_n)} \mathrm{VaR}_p(S) = \inf\{\sup S : S \in \mathfrak{S}_n(F_1^p, \cdots, F_n^p)\}.$$

The corollary below connects the convex ordering bounds and bounds on the Value-at-Risk.

Corollary 4.7. *Suppose F_i has positive density on its support, $i = 1, \cdots, n$.*

(a) Suppose S_p is a convex ordering minimal element in $\mathfrak{S}_n(F_{1,p}, \cdots, F_{n,p})$ for $p \in (0, 1)$

$$\sup_{S \in \mathfrak{S}_n(F_1, \cdots, F_n)} \mathrm{VaR}_p(S) = \inf S_p.$$

(b) Suppose S_p is a convex ordering minimal element in $\mathfrak{S}_n(F_1^p, \cdots, F_n^p)$ for $p \in (0, 1)$

$$\inf_{S \in \mathfrak{S}_n(F_1, \cdots, F_n)} \mathrm{VaR}_p(S) = \sup S_p.$$

Remark 4.1. The above corollary shows that the technique used to find convex ordering minimal elements in an admissible risk class in this paper can also be applied to find the bounds for the Value-at-Risk. Note that as pointed out in Section 3, the convex ordering minimal element in an admissible risk class may not exist.

We apply the results obtained in Section 3 to the bounds on VaR for homogeneous risks. Throughout this section and the next section, n is a positive integer (although only $n \geqslant 3$ is of interest). For any distribution F, we use the conditions (A), (A') and (B) introduced in Section 3:

(A) For some $a \in \left[0, \frac{1}{n}\right]$, $H(x)$ 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 $I = \left[F^{-1}((n-1)c_n), F^{-1}(1-c_n)\right]$.

Here, for consistency, $H(x)$ and $D(a)$ are defined as in Section 3.3 for given F (when the marginals are inhomogeneous, let $F = \frac{1}{n}\sum_{i=1}^n F_i$), and c_n is defined by (3.14). (A) is used for both homogeneous and heterogeneous risks, while (A') and (B) are used only for homogeneous risks. In the following Corollary, when we say (A), (A') or (B) holds for F_p (or F^p), we mean that the distribution F in (A), (A') or (B) should be replaced by F_p (or F^p) wherever applicable.

Corollary 4.8. *Suppose $F_1, \cdots, F_n$ are continuous distributions. Denote by $F_p = \frac{1}{n}\sum_{i=1}^n F_{i,p}$ and $F^p = \frac{1}{n}\sum_{i=1}^n F_i^p$ for $p \in (0, 1)$.*

(a) If (A) holds for the distribution F_p and some $a > 0$, then

$$\sup_{S\in\mathfrak{S}_n(F_1,\cdots,F_n)} \mathrm{VaR}_p(S) \leqslant \frac{n\int_{(n-1)a}^{1-a} F_p^{-1}(y)\mathrm{d}y}{1-na}.$$

Moreover, in the homogeneous case $F_1 = \cdots = F_n = F$, the above bound is sharp for $a = c_n$ if (A') and (B) hold for F_p, and F has positive density on its support.

(b) We always have

$$\inf_{S\in\mathfrak{S}_n(F_1,\cdots,F_n)} \mathrm{VaR}_p(S) \geqslant (n-1)(F^p)^{-1}(0) + (F^p)^{-1}(1).$$

Moreover, in the homogeneous case $F_1 = \cdots = F_n = F$, the above bound is sharp if (A') and (B) hold for F^p.

Remark 4.2. Corollary 4.8 (b) holds trivially true without any assumptions for homogeneous risks. It is somewhat surprising that with (A') and (B), this trivial bound is sharp. The main result of explicit VaR bounds for tail-monotone densities in Wang et al. (2013) is directly implied by Corollary 4.8 and the complete mixability of monotone densities. In this paper, our proof is much simpler than the proof in Wang et al. (2013).

Remark 4.3. The results in this section benefited from earlier communications and discussions with Steven Vanduffel. Corollary 4.7 was also obtained independently in Bernard et al. (2013b) for VaR^*.

5 Bounds on Convex Risk Measures and Convex Expectations

Apart from the bounds on the Value-at-Risk, our results on convex order also apply naturally to bounds on convex risk measures and in particular on coherent risk measures as well as on convex expectations.

5.1 Convex and coherent risk measures

A risk measure is a mapping from random variables to real numbers, which can be used as capital requirement to regulate risk assumed by market participants. For a detailed introduction on risk measures and more specifically on coherent risk measures, we refer to Artzner et al. (1999). Consider a *risk measure* as $\rho : L^0(\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R} \cup \{\infty\}$. Most discussions focus on risk measures on $L^p(\Omega, \mathcal{A}, \mathbb{P})$ for $p \in [1, \infty]$. Delbaen (2009) studied the case of non-integrable random variables, and proved that there exist no finite convex risk measures defined on $L^p(\Omega, \mathcal{A}, \mathbb{P})$ for $p \in [0, 1)$. Since convex order is defined for L^1 random variables, we restrict our discussion on $\rho : L^1(\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$. Let $X, X_1, X_2, \cdots \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. Recall the following properties of a risk measure $\rho(\cdot)$

(1) Monotonicity: if $X_1 \leqslant X_2$ then $\rho(X_1) \leqslant \rho(X_2)$.

(2) Translation invariance: $\rho(X + m) = \rho(X) + m$ for $m \in \mathbb{R}$.

(3) Subadditivity: $\rho(X_1 + X_2) \leqslant \rho(X_1) + \rho(X_2)$.

(4) Positive homogeneity: $\rho(\lambda X) = \lambda \rho(X)$ for $\lambda > 0$.

(5) Convexity: $\rho(\lambda X_1 + (1 - \lambda)X_2) \leqslant \lambda \rho(X_1) + (1 - \lambda)\rho(X_2)$ for $\lambda \in [0, 1]$.

(6) Law invariance: if $X_1 \stackrel{d}{=} X_2$, then $\rho(X_1) = \rho(X_2)$.

(7) L^1-Fatou property: if $X_n \to X$ in L^1, then $\rho(X) \leqslant \liminf \rho(X_n)$.

The importance of properties (1-5) in risk management is rather obvious[4] and well explained in Artzner et al. (1999). (6) is most often assumed and is used when connecting convex order

[4]Note that these properties are formulated to apply to non-negative risks. Some authors assume $\rho(-1) = 1$ and apply risk measures on losses variables that are negative.

and risk measures (for example, see Bäuerle and Müller (2006)). Recall that convex order is based on distributional properties only and thus (6) is needed for applying our results to convex order. (7) is a continuity property on the risk measure $\rho(\cdot)$ with respect to convergence in L^1 (it is typically assumed when studying risk measures on an atomless probability space). Some justifications for the continuity property (7) can be found in Delbaen (2002) and Bäuerle and Müller (2006).

A risk measure is *coherent* if it satisfies properties (1-4). It immediately follows that a coherent risk measure satisfies also (5). Recall that a coherent risk measure has the typical dual representation

$$\rho(X) = \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[X]$$

where $\mathcal{Q}$ is some family of probability measures on Ω. This was introduced in Artzner et al. (1999) in a finite state probability space and discussed in Delbaen (2002) in a more general probability space.

A risk measure on $L^\infty(\Omega, \mathcal{A}, \mathbb{P})$ is called a *convex risk measure*, defined in Föllmer and Schied (2002), if it satisfies properties (1,2,5) (relaxing subadditivity and positive homogeneity). A dual representation is also given in the same paper. The concept was later studied in Svindland (2008) and Kaina and Rüschendorf (2009), for more general probability spaces. A recent review of convex and coherent measures can be found in Föllmer and Schied (2010).

5.2 Bounds on convex risk measures of admissible risk

In practice, information about dependence is limited. Bounds on a convex (or coherent) risk measure $\rho(S)$ over the admissible risk class $\mathfrak{S}_n(F_1, \cdots, F_n)$ are thus of much importance in risk management. The consistency of convex order and convex risk measures is given in Theorem 4.3 of Bäuerle and Müller (2006). Since it is well-known that the convex ordering upper bound of $\mathfrak{S}_n(F_1, \cdots, F_n)$ is given by the comonotonic scenario of $\mathbf{X}$, a sharp upper bound on $\rho(S)$ over $S \in \mathfrak{S}_n(F_1, \cdots, F_n)$ is $\rho(nF^{-1}(U))$ where $U \sim \mathcal{U}[0,1]$ and it is well-discussed in the literature (for a review, see Dhaene et al. (2006)). On the other hand, the lower bound on $\rho(S)$ over $S \in \mathfrak{S}_n(F_1, \cdots, F_n)$ is unknown in the literature except for $n = 2$. Using the results in Section 3, we are able to give a lower bound on $\rho(S)$, as follows.

Corollary 5.1 (Bounds on convex risk measures of admissible risk)**.** *For every risk measure ρ satisfying (5-7), i.e. law-invariant, convex, L^1-Fatou, if (A) holds, then*

$$\inf_{S \in \mathfrak{S}_n(F_1, \cdots, F_n)} \rho(S) \geqslant \rho(T_a), \tag{5.1}$$

where T_a is defined by (3.4)*. Moreover, in the homogeneous case $F_1 = \cdots = F_n = F$, if (A') and (B) hold, then the above bound is sharp for $a = c_n$.*

Proof. The inequality (5.1) is a corollary of Theorem 3.5 in this paper and Theorem 4.3 of Bäuerle and Müller (2006). The sharpness in the homogeneous case is implied by Theorem 3.2. $\square$

Remark 5.1. Note that we assume finite means for $F, F_1, \cdots, F_n$, thus only the behavior of ρ on $L^1(\Omega, \mathcal{A}, \mathbb{P})$ matters. In Corollary 5.1, we do not require ρ to satisfy (1,2), and thus ρ is not necessarily a convex risk measure as defined in Föllmer and Schied (2002) and does not necessarily have a financial interpretation. A law-invariant coherent risk measure with the Fatou property is thus only a special case in this corollary.

5.3 Bounds on TVaR of admissible risk

The Tail Value-at-Risk (TVaR; it has other names such as CTE, AVaR, CVaR and ESF in different contexts) is defined as

$$\mathrm{TVaR}_p(X) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_\alpha(X)\mathrm{d}\alpha, \qquad p \in [0,1).$$

As it satisfies (1-7), it is a coherent risk measure. Furthermore, every risk measure on $L^1(\Omega, \mathcal{A}, \mathbb{P})$ satisfying (1-7) has a representation of

$$\rho(X) = \sup_{\mu \in P_0} \int_0^1 \mathrm{TVaR}_p(X)\mu(\mathrm{d}p) \tag{5.2}$$

where P_0 is a compact, convex set of probability measures on $[0,1]$ (for this result, see Bäuerle and Müller (2006); Kusuoka (2001)). Due to the increasing importance of TVaR in risk management (see e.g. Panjer (2006) and recent regulations of Basel Committee on Banking Supervision (2010, 2012)) and the representation (5.2) of law-invariant coherent risk measures, bounds for $\mathrm{TVaR}_p(S)$ are of practical interest.

Theorem 5.2 (Bounds on TVaR of admissible risk)**.**

(a) For $p \in [0,1]$, *if (A) holds, then*

$$\inf_{S \in \mathfrak{S}_n(F_1,\cdots,F_n)} \mathrm{TVaR}_p(S) \geqslant \begin{cases} \frac{1}{1-p}[\mathbb{E}[S] - pD(a)] & p \leqslant 1-na; \\ \frac{n}{1-p}\int_0^{(1-p)/n} H(x)\mathrm{d}x & p > 1-na. \end{cases} \tag{5.3}$$

(b) In the homogeneous case $F_1 = \cdots = F_n = F$, *the bound* (5.3) *is sharp for* $a = c_n$ *if (A') and (B) hold.*

(c) In the homogeneous case $F_1 = \cdots = F_n = F$, *if (A) holds for* $a \geqslant \frac{1-p}{n}$, *then*

$$\inf_{S \in \mathfrak{S}_n(F,\cdots,F)} \mathrm{TVaR}_p(S) = \frac{n}{1-p}\int_0^{(1-p)/n} H(x)\mathrm{d}x \tag{5.4}$$

if

$$\inf_{S\in\mathfrak{S}_n(F_J,\cdots,F_J)} \mathbb{P}\left(S > H\left(\frac{1-p}{n}\right)\right) = 0, \tag{5.5}$$

where F_J is the conditional distribution of F on $J = \left[F^{-1}\left(\frac{(n-1)(1-p)}{n}\right), F^{-1}\left(1-\frac{1-p}{n}\right)\right]$.

Proof. (a) By Corollary 5.1 we have $\mathrm{TVaR}_p(S) \geqslant \mathrm{TVaR}_p(T_a)$. We just need to verify that

$$\mathrm{TVaR}_p(T_a) = \begin{cases} \frac{1}{1-p}[\mathbb{E}[S] - pD(a)] & p \leqslant 1-na; \\ \frac{n}{1-p}\int_0^{(1-p)/n} H(x)\mathrm{d}x & p > 1-na, \end{cases}$$

which can be directly calculated from the distribution of T_a in (3.12).

(b) It is follows from (a) and Corollary 5.1.

(c) We only need to show that there exists a random variable $S \in \mathfrak{S}_n(F,\cdots,F)$ such that $\mathrm{TVaR}_p(S) \leqslant \frac{n}{1-p}\int_0^{(1-p)/n} H(x)\mathrm{d}x$. Let $\mathbf{Y} = (Y_1,\cdots,Y_n) \in \mathfrak{F}_n(F_J,\cdots,F_J)$ and $T = \mathbf{Y}\mathbf{1}_n$ such that $\mathbb{P}\left(T > H\left(\frac{1-p}{n}\right)\right) = 0$. Such T always exists since $\inf_{T\in\mathfrak{S}_n(F_J,\cdots,F_J)} \mathbb{P}(T > t)$ is attainable by $T \in \mathfrak{S}_n(F_J,\cdots,F_J)$ for any $t \in \mathbb{R}$; see the introduction of Rüschendorf (1983). Let $\mathbf{Z} = (Z_1,\cdots,Z_n) \in \mathfrak{F}_n(F_{J^c},\cdots,F_{J^c})$ where F_{J^c} is the conditional distribution of F on $\mathbb{R}^+ \setminus J$ and $W = \mathbf{Z}\mathbf{1}_n$ such that $\mathbb{P}\left(W \geqslant H\left(\frac{1-p}{n}\right)\right) = 1$. Such W always exists and can be constructed by (3.16). Let $S = (1-\mathrm{I}_A)T + \mathrm{I}_A W$ for $i = 1,\cdots,n$, where $A \in \mathcal{A}$ is independent of T and W and $\mathbb{P}(A) = p$. It is easy to check that $F = (1-p)F_J + pF_{J^c}$, and hence $S \in \mathfrak{S}_n(F,\cdots,F)$ by Theorem 2.1 (i). Since $\mathbb{P}\left(S > H\left(\frac{1-p}{n}\right)\right) = p\mathbb{P}\left(W > H\left(\frac{1-p}{n}\right)\right) \leqslant p$, we have $\mathrm{VaR}_p(S) \leqslant H\left(\frac{1-p}{n}\right)$ and

$$\mathrm{TVaR}_p(S) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_\alpha(S)\mathrm{d}\alpha \leqslant \frac{1}{1-p}\mathbb{E}[W] = \frac{n}{1-p}\int_0^{(1-p)/n} H(x)\mathrm{d}x.$$

□

Remark 5.2. In practice, typical values of p for TVaR are close to 1, such as 0.95, 0.99 and 0.995. From Theorem 5.2 we can tell that the lower bound on $\mathrm{TVaR}_p(S)$ for $p \geqslant 1-na$ is $\frac{n}{1-p}\int_0^{(1-p)/n} H(x)\mathrm{d}x$ and it does not depend on a any more. In the homogeneous case, for p close to 1, the sharpness of the lower bound (5.4) does not require the conditions (A') and (B). Instead, it is guaranteed by a much weaker condition (5.5) (note that the condition (5.5) is weaker than (A') and (B) since (B) requires $\sup_{S\in\mathfrak{S}_n(F_I,\cdots,F_I)} \mathbb{P}(S = D(c_n)) = 1$) and the monotonicity of $H(x)$ in a neighborhood of 0. The bounds (5.3) and (5.4) are also very easy to compute.

As an illustration of the results on TVaR, Figure 2 displays a numerical comparison of TVaR for Log-Normal(0,1) risks under three dependence scenarios: comonotonic risks, independent risks and the convex lower bound.

Remark 5.3. We can always use discrete distributions to approximate the marginal distributions $F_1, \cdots, F_n$. When a discrete approximation is used, the optimization over all possible dependence structures becomes a finite-state problem, and hence it can be solved numerically. For example, Puccetti (2013) used the Rearrangement Algorithm (RA) to calculate the bounds on TVaR over the admissible risk class. There are three notable facts about the merits of our theoretical results compared to the RA approximation. First, our result gives an explicit form and a sharpness condition, while the RA only gives a numerical approximation. Second, although being easy to implement, there is yet no proof that the RA approximation converges to the sharp lower bound on the TVaR as the number of discretization steps m goes to infinity. Third, the RA becomes slow when the dimension n or the number of discretization steps m is large. Our method only needs to numerically find c_n and the complexity does not depend on n. We provide some numerical examples in Section 6.

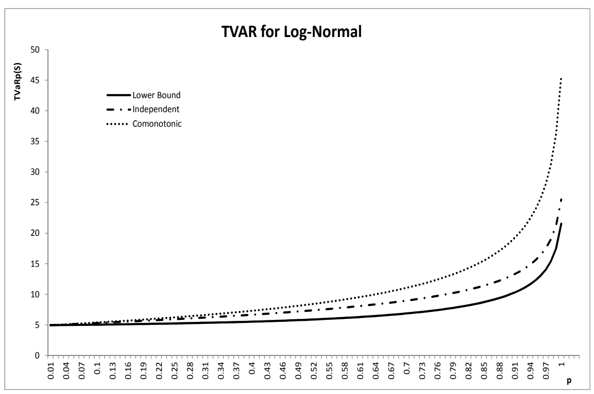


Figure 2: $\mathrm{TVaR}_p(S)$, $p \in [0, 0.995)$ for Log-Normal(0,1) risks, $n = 3$.

5.4 Convex expectation and applications in finance and insurance

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. If f is convex and bounded, then $\mathbb{E}[f(X)]$ satisfies (5-7) and thus is a risk measure as described in Corollary 5.1. Theoretically, $\mathbb{E}[f(X)]$ can be infinity. By 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

$\mathfrak{S}_n(F_1, \cdots, F_n)$,

$$\mathbb{E}\left[f(S)\right] = \mathbb{E}\left[f(X_1 + X_2 + \cdots + X_n)\right], \tag{5.6}$$

regardless of $\mathbb{E}[f(S)]$ being finite or infinite. Recall that when f is convex, the upper bound can be computed explicitly with the comonotonic dependence scenario.

Corollary 5.3 (Bounds on convex expectations of admissible risk)**.** *For a convex function f, if (A) holds, then*

$$\inf_{S \in \mathfrak{S}_n(F_1,\cdots,F_n)} \mathbb{E}\left[f(S)\right] \geqslant n \int_0^a f(H(x))\mathrm{d}x + (1 - na) f(D(a)). \tag{5.7}$$

Specifically, in the homogeneous case

$$\inf_{S \in \mathfrak{S}_n(F,\cdots,F)} \mathbb{E}\left[f(S)\right] \geqslant n \int_0^a f(H(x))\mathrm{d}x + (1 - na) f(D(a)), \tag{5.8}$$

and moreover, the equality in (5.8) *holds for $a = c_n$ if (A') and (B) hold.*

Remark 5.4. Corollary 5.3 can be seen as a generalization of Jensen's inequality as (5.7) is simply Jensen's inequality when $a = 0$. It can also be seen as a generalization of Theorem 3.5 of Wang and Wang (2011), where monotone densities were assumed.

Although finite convex expectations can be viewed mathematically as a special case of law-invariant risk measures, the application and financial interpretation of convex expectations are different from those of risk measures. Some quantities of interest that can be viewed as a convex or concave expectation of the aggregate risk S include the variance of a joint portfolio, the price of a European basket option, the expected utility of a joint portfolio, the stop-loss premium of an aggregate loss; the price of a European option on the realized variance of an asset price process, the expected n-period return, and some convex risk measures such as the entropic risk measure $\rho(X) = \frac{1}{\theta} \log \mathbb{E}[e^{\theta X}]$. Bounds on convex or concave expectations help to analyze risks under best or worst case scenarios when the information on dependence is unreliable.

The last section gives some further illustration and proposes a method to check property (B) numerically.

6 Numerical Illustrations

Considering the conditions (A), (A') and (B) are sometimes difficult to check, we give some numerical illustrations in this section. As mentioned in Remark 5.3, a natural idea is to construct a discretization of the marginal distributions $F_1, \cdots, F_n$, then the optimization over all possible dependence structures becomes a finite-state problem and is always solvable. For each discretization, we find the optimal discrete structure with respect to minimal convex order, and compare some quantities such as variance and TVaR with our theoretical results.

The Rearrangement Algorithm (RA) introduced in Puccetti and Rüschendorf (2012) and also used in Embrechts et al. (2013) and Puccetti (2013) is a quick algorithm to provide discrete numerical approximations for the optimal structure with respect to minimal convex order. In the following, the RA is used to approximate the lower bound on $\mathbb{E}[f(S)]$ for some convex functions f and for $\mathrm{TVaR}_p(S)$ when $p = 0.95$. We compare the RA approximation with the lower bound suggested by Theorem 5.2 and Corollary 5.3. The numerical results suggest that the lower bound for homogeneous risks is very likely to be sharp.

6.1 Homogeneous case

In this section, we compare the RA approximation with the lower bound suggested by Corollary 5.3 and Theorem 5.2 for different settings of homogeneous risks. We take the number of discretization steps in the RA as $m = 10^6$.

Numerical results are given in Table 1 in 6 different settings. For each setting, we also give the TVaR under the assumption of independence and comonotonicity to show the impact of various dependence assumptions. Note that the Gamma and Lognormal distributions above do not have a decreasing density and therefore theoretically we do not know whether they satisfy (B), while for the Pareto distributions we know the bounds given in Theorem 5.2 and Corollary 5.3 are sharp.

From Table 1, we conclude that the bounds obtained for homogeneous risks in Theorems 3.2 and 5.2 are very likely to be sharp for all above distributions.

6.2 Heterogeneous Case

For the heterogeneous case, in Table 2 we give numerical examples of two portfolios of different Pareto risks, three portfolio of different Lognormal risks, and a mixed portfolio of Pareto, Lognornal and Gamma risks. We take the number of discretization steps in the RA as $m = 10^6$.

From Table 2, it appears that the bounds given in Theorem 3.5 are quite accurate in general. Note also that the theoretical bounds tend to be more precise when the distributions are similar. This confirms the intuition provided when deriving the lower bound for heterogeneous risks in Section 3.4.

Table 1: RA results vs theoretical bounds, homogeneous case

We consider 6 different settings: Pareto(1,3), Pareto(1,4) with $n = 4$, Gamma(2,0.5), Gamma(3,1) with $n = 3$, Log-Normal(0,1) with $n = 3$ and $n = 10$. Three quantities are calculated: the variance: $f(S) = (S - n\mu)^2$ where $\mu = \mathbb{E}[X_1]$, the European call option prices $f(S) = (S - K)^+$ when $K = n\mu$, and the TVaR of S at level 95%.

	Pareto(θ, α); n		Gamma(α, β); n		LogN(μ, σ^2); n	
	(1,3); 4	(1,4); 4	(2,0.5); 3	(3,1); 3	(0,1); 3	(0,1); 10
	Variance					
RA	1.3545	0.2615	0.7466	0.0986	5.9521	3.3022
Corollary 5.3	1.3545	0.2615	0.7466	0.0986	5.9521	3.3022
	Option price with strike $K = n\mu$					
RA	0.2321	0.1113	0.1866	0.0510	0.6232	0.1978
Corollary 5.3	0.2321	0.1113	0.1866	0.0510	0.6232	0.1978
	TVaR at level 0.95					
RA	9.4804	7.0015	15.1154	10.0061	13.0523	20.3763
Theorem5.2	9.4803	7.0015	15.1154	10.0061	13.0524	20.3762
Independent	11.0973	8.1136	24.2280	16.2629	16.4025	35.7371
Comonotonic	16.2845	11.2782	35.5077	22.8052	25.6710	85.5700

Table 2: RA results vs theoretical bounds, heterogeneous case

Three quantities are calculated: the variance: $f(S) = (S - K)^2$ where $K = \sum_{i=1}^{n} \mathbb{E}[X_i]$, the European call option price $f(S) = (S - K)^+$, and the TVaR of S at level 95%.

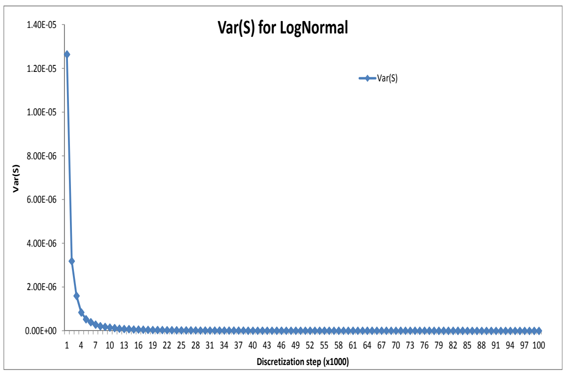
	$X_i \sim$ Pareto$(1, \alpha_i)$, $i = 1, 2, 3$		$X_i \sim$LogN$(i/10, 1)$, $i = 1, \cdots, n$			X_1 ~Pareto(1,3)
	$(\alpha_1, \alpha_2, \alpha_3)$		n			X_2 ~LogN(1,0.5)
	(3, 4, 5)	(3.5, 4, 4.5)	3	5	10	X_3 ~Gamma(3,1)
	Variance					
RA	0.6028	0.3343	9.0882	10.3479	15.3348	0.7889
Corollary 5.3	0.5888	0.3314	9.0677	10.3132	15.1851	0.5041
	Option price with strike $K = \sum_{i=1}^{n} \mathbb{E}[X_i]$					
RA	0.1726	0.1413	0.7701	0.6419	0.4665	0.2474
Corollary 5.3	0.1725	0.1412	0.7699	0.6404	0.4604	0.1389
	TVaR at level 0.95					
RA	6.4255	5.8755	16.0766	21.7828	39.0200	10.5445
Theorem5.2	6.4235	5.8748	16.0749	21.7729	38.8892	9.9818
Independent	7.0484	6.6550	20.1592	31.0085	66.2389	14.1519
Comonotonic	9.1636	8.6144	31.4534	58.3221	154.4790	19.4410

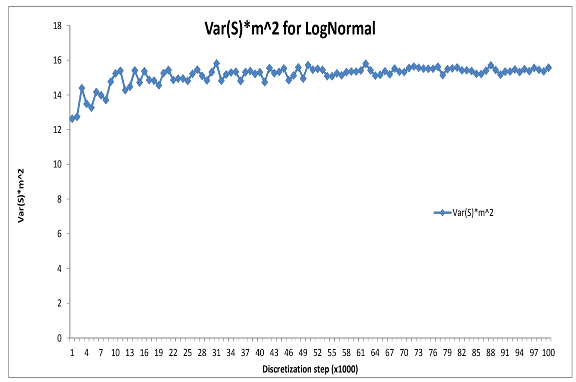
6.3 Checking condition (B)

Recall that Condition (B) in Theorem 3.2 corresponds to checking that F is n-CM on the interval $I = [F^{-1}((n-1)c_n), F^{-1}(1-c_n)]$. This is equivalent to

$$\underline{\mathrm{Var}}(S) := \inf_{S \in \mathfrak{S}_n(F_I, \cdots, F_I)} \mathrm{Var}(S) = 0 \tag{6.1}$$

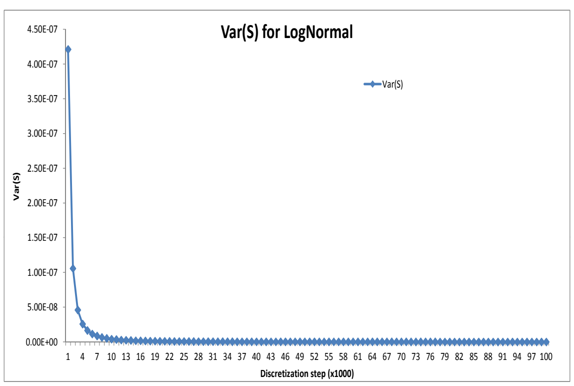
where F_I is the conditional distribution of F on I. Since the RA gives a discrete approximation of the optimal dependence structure, (6.1) holds if the RA approximation of $\underline{\mathrm{Var}}(S)$, denoted by $\underline{\mathrm{Var}}(S)_m$ goes to zero when the number of discretization steps m goes to infinity (however, in the opposite direction, (6.1) does not imply that $\underline{\mathrm{Var}}(S)_m \to 0$ since the convergence of the RA approximation is not proved). To illustrate this convergence of the rearrangement algorithm, we represent in Figures 3 and 4 the variance of the sum of n risks for different distributions as a function of the discretization step m.

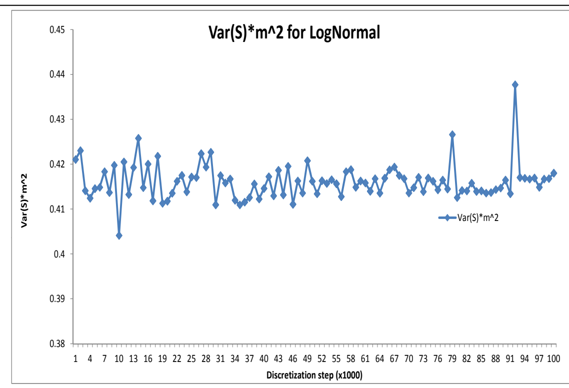




Panel A: $\underline{\text{Var}}(S)_m$ w.r.t. m Panel B: $\underline{\text{Var}}(S)_m \times m^2$ w.r.t. m

Log-Normal(0,1), $n = 3$.

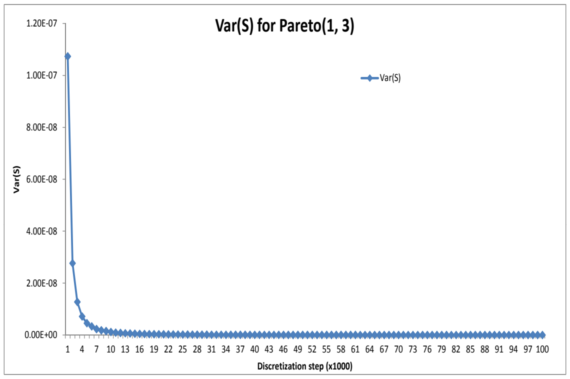




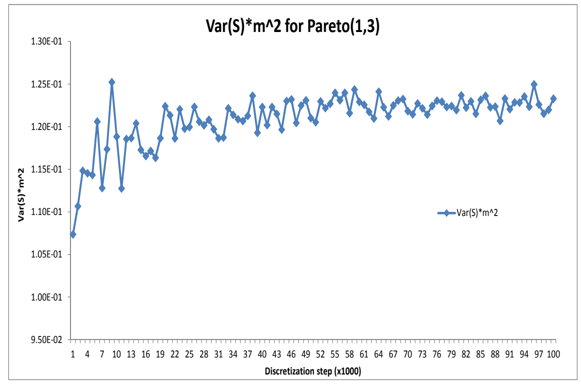
Panel C: $\underline{\text{Var}}(S)_m$ w.r.t. m Panel D: $\underline{\text{Var}}(S)_m \times m^2$ w.r.t. m

Log-Normal(0,1), $n = 10$.

Figure 3: Panels A and C display $\underline{\text{Var}}(S)_m$ as a function of m for a Pareto distribution and Panels B and D illustrate the speed of convergence in $1/m^2$.

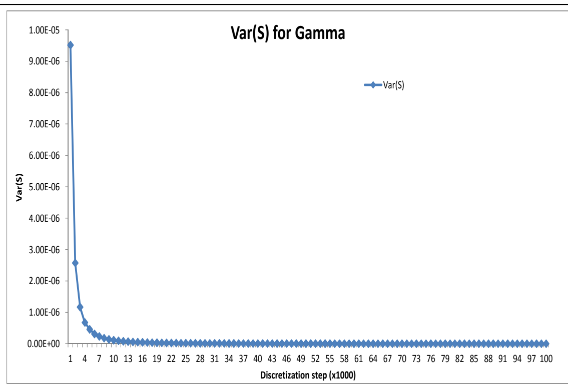


Panel A: $\underline{\mathrm{Var}}(S)_m$ w.r.t. m

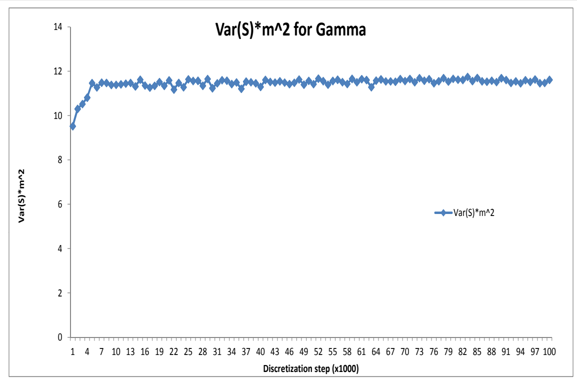


Panel B: $\underline{\mathrm{Var}}(S)_m \times m^2$ w.r.t. m

Pareto(1,3), $n = 4$



Panel C: $\underline{\mathrm{Var}}(S)_m$ w.r.t. m



Panel D: $\underline{\mathrm{Var}}(S)_m \times m^2$ w.r.t. m

Gamma(2,0.5), $n = 3$

Figure 4: Panels A and C display $\underline{\mathrm{Var}}(S)_m$ as a function of m for a Pareto distribution and a Gamma distribution and Panels B and D illustrate the speed of convergence in $1/m^2$.

From Figures 3 and 4, the RA approximations $\underline{\mathrm{Var}}(S)_m$ clearly converge to zero, at a rate of m^{-2}. Based on all the observations in Section 5, we have the following conjecture.

Conjecture 6.1. A Gamma or Log-Normal distribution F is n-CM on the interval $I = [F^{-1}((n-1)c_n), F^{-1}(1-c_n)]$ for any integer n, and the convex ordering bounds in Theorems 3.2 and 3.5 are sharp.

Even if we are not able to prove this conjecture at this moment, the numerical results clearly show that the lower bounds on convex risk measures and convex expectations are sharp enough to apply in practice, for identical or almost identical marginal distributions.

7 Conclusions and Future Work

In this paper, we introduce and investigate the admissible risk class $\mathfrak{S}_n(F_1, \cdots, F_n) = \{X_1 + \cdots + X_n : X_i \sim F_i,\ i = 1, \cdots, n\}$ for given marginal risk distributions $F_1, \cdots, F_n$. We give a new lower bound over $\mathfrak{S}_n(F_1, \cdots, F_n)$. In the homogeneous case, $F_1 = \cdots = F_n$, we give a sufficient condition for the new lower bound to be sharp. The results can be used to find sharp bounds on convex risk measures and other quantities in finance when the dependence information among individual risks is missing. Numerical illustrations suggest that the new lower bound is likely to be sharp for most risk distributions and the conditions used in our main results are usually satisfied.

Some future directions related to this topic include proving Conjecture 6.1. More generally, we expect Conjecture 6.1 to hold for all unimodal densities given some smooth conditions and also for heterogeneous risks under some additional conditions. Recall that the heterogeneous analog of complete mixability is called *joint mixability* and is introduced in Wang et al. (2013). Note that proving Conjecture 6.1 for heterogeneous risks is an open problem even in the case of decreasing densities. Finally, it is of interest to determine conditions under which convex ordering bounds for heterogeneous risks (over $\mathfrak{S}_n(F_1, \cdots, F_n)$) are sharp. We believe that these research directions are all technically challenging and relevant to quantitative risk management.

References

Albrecher, H., Mayer, P., Schoutens, W. (2008). General lower bounds for arithmetic Asian option prices. *Applied Mathematical Finance*, **15**(2), 123–149.

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

d'Aspremont, A., El Ghaoui, L. (2006). Static arbitrage bounds on basket option prices. *Mathematical Programming: Series A and B*, **106**(3), 467–489.

Basel Committee on Banking Supervision (2010). *Developments in modelling risk aggregation*. Basel: Bank for International Settlements.

Basel Committee on Banking Supervision (2012). *Consultative Document May 2012. Fundamental review of the trading book*. 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 Vanduffel S. (2012). Note on Improved Fréchet Bounds. *Journal of Applied Probability, Journal of Applied Probability*, **49**(3), 866–875.

Bernard, C., Liu, Y., MacGillivray, N., Zhang, J. (2013a). Bounds on Capital Requirements For Bivariate Risk with Given Marginals and Partial Information on the Dependence. *Dependence Modelling*, **1**, 37-53.

Bernard, C., Rüschendorf, L. and Vanduffel S. (2013b). Value-at-Risk Bounds with Variance Constraints. *Working paper available at SSRN: http://ssrn.com/abstract=2342068.*

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*, **2**, 103-118.

Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, **16**, 519-547.

Delbaen, F. (2002). Coherent measures of risk on general probability spaces. *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann*, Springer-Verlag, 1–37.

Delbaen, F. (2009). Risk measures for non-integrable random variables. *Mathematical Finance,* **19**(2), 329–333.

Denuit, M., Genest, C. and Marceau, É. (1999). Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, **25**, 85–104.

Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R. and Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics* **31**(1), 3-33.

Dhaene, J., Vanduffel, S., Goovaerts, M.J., Kaas, R., Tang, Q. and Vyncke, D. (2006). Risk measures and comonotonicity: a review. *Stochastic models*, **22**, 573–606.

Dhaene, J., Linders, D., Schoutens, W. and Vyncke, D. (2012). The Herd Behavior Index: A new measure for the implied degree of co-movement in stock markets *Insurance: Mathematics and Economics,* **50**, 357–370.

Embrechts, P., Höing, A. and Juri, A. (2003). Using copulae to bound the value-at-risk for functions of dependent risks. *Finance and Stochastics*, **7**, 145–167.

Embrechts, P. and Puccetti, G. (2006). Bounds for functions of dependent risks. *Finance and Stochastics*, **10**, 341–352.

Embrechts, P. and Puccetti, G. (2010). Risk aggregation. In: *Copula Theory and Its Applications*, P. Jaworski et al. (eds.), Lecture Notes in Statistics **198**, Springer, Berlin/Heidelberg.

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013) Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750-2764.

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

Föllmer, H., Schied, A. (2010). Coherent and convex risk measures, in *Encyclopedia of Quantitative Finance*, Cont, R. (ed.) New York: John Wiley & Sons, 355–363.

Hobson, D., Laurence, P. and Wang, T.-H. (2005). Static-arbitrage upper and lower bounds for the prices of basket options. *Quantitative Finance*, **5**(4), 329–342.

Kaas, R., Laeven, R. and Nelsen, R. (2009). Worst VaR scenarios with given marginals and measures of association. *Insurance: Mathematics and Economics*, **44**, 146–158.

Kaas, R., Dhaene, J. and Goovaerts, M. (2000). Upper and lower bounds for sums of random variables. *Insurance: Mathematics and Economics*, **27**(2), 151–168.

Kaina, M. and Rüschendorf, L. (2009). On convex risk measures on L^p-spaces. *Mathematical Methods of Operations Research*, **69**, 475–495.

KELLER-RESSEL, M. AND G. CLAUSS (2012). Convex order of discrete, continuous and predictable quadratic variation & applications to options on variance. *Working paper available on ArXiv.*

KUSUOKA, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, **3**, 83–95.

MCNEIL, A. J., FREY, R. AND EMBRECHTS, P. (2005). *Quantitative Risk Management: Concepts, Techniques, Tools.* Princeton, NJ: Princeton University Press.

MEILIJSON, I., NADAS, A. (1979). Convex majorization with an application to the length of critical paths. *Journal of Applied Probability*, **16**, 671–677.

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.

PANJER, H. (2006). *Operational Risk: Modeling Analytics*, John Wiley & Sons.

PUCCETTI, G. (2013). Sharp bounds on the expected shortfall for a sum of dependent random variables. *Statistics and Probability Letters,* **83**(4), 1227-1232.

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

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

ROGERS, L.C.G. AND SHI, Z. (1995). The value of an Asian option. *Journal of Applied Probability,* **32**(1), 1077–1088.

RÜSCHENDORF, L. (1980). Inequalities for the expectation of Δ-monotone functions, *Zeitschrift für Wahrscheinlichkeitstheorie*, **54**, 341–349.

RÜSCHENDORF, L. (1983). Solution of a statistical optimization problem by rearrangement methods, *Metrika*, **30**, 55–61.

SVINDLAND, G. (2008). Convex Risk Measures Beyond Bounded Risk. *Ph.D. Thesis,* Ludwig Maximilian Universität München.

Tankov, P. (2011). Improved Fréchet bounds and model-free pricing of multi-asset options. *Journal of Applied Probability*, **48**, 389–403.

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

Valdez, E.A., Dhaene, J., Maj, M. and Vanduffel, S. (2009). Bounds and approximations for sums of dependent log-elliptical random variables. *Insurance: Mathematics and Economics*, **44**(3): 385–397.

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.

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

A Proof of Theorem 2.1

Proof.

(i) Suppose $S_1 \in \mathfrak{S}_n(F_1, \cdots, F_n)$. Write

$$S_1 = X_1 + \cdots + X_n, \quad X_i \sim F_i, \ i = 1, \cdots, n.$$

Denote the joint distribution of $(X_1, \cdots, X_n, S_1)$ by H. Since the probability space is rich enough, there exists a random vector $(Y_1, \cdots, Y_n, S_2)$ with distribution H, and with S_2 as its last component. It follows immediately that $S_2 = Y_1 + \cdots + Y_n$ and $Y_i \sim F_i$, $i = 1, \cdots, n$. That is $S_2 \in \mathfrak{S}_n(F_1, \cdots, F_n)$.

(ii) Write $S_1 = \mathbf{X}\mathbf{1}_n$ and $S_2 = \mathbf{Y}\mathbf{1}_n$ where $\mathbf{X} \in \mathfrak{F}_n(\mathbf{F})$ and $\mathbf{Y} \in \mathfrak{F}_n(\mathbf{G})$. Let $B \in \mathcal{A}$ be independent of $\mathbf{X}$ and $\mathbf{Y}$, and $\mathbb{P}(B) = \mathbb{P}(A)$ (this is always possible in an atomless space). It is easy to check that $\mathrm{I}_A S_1 + (1 - \mathrm{I}_A) S_2 \stackrel{d}{=} \mathrm{I}_B S_1 + (1 - \mathrm{I}_B) S_2$. Note that $\mathrm{I}_B S_1 + (1 - \mathrm{I}_B) S_2 = (\mathrm{I}_B X_1 + (1 - \mathrm{I}_B) Y_1) + \cdots + (\mathrm{I}_B X_n + (1 - \mathrm{I}_B) Y_n) \in \mathfrak{S}_n(\mathbb{P}(B)\mathbf{F} + (1 - \mathbb{P}(B))\mathbf{G})$. It follows that $\mathrm{I}_A S_1 + (1 - \mathrm{I}_A) S_2 \in \mathfrak{S}_n(\mathbb{P}(A)\mathbf{F} + (1 - \mathbb{P}(A))\mathbf{G})$.

(iii) Write $S_1 = \mathbf{X}\mathbf{1}_n$ and $S_2 = \mathbf{Y}\mathbf{1}_n$ where $\mathbf{X} \in \mathfrak{F}_n(\mathbf{F})$ and $\mathbf{Y} \in \mathfrak{F}_n(\mathbf{G})$. Let $\mathbf{Z} \in \mathfrak{F}_n(\mathbf{G})$ be independent of $\mathbf{X}$ and $\mathbf{Z} \stackrel{d}{=} \mathbf{Y}$. It is easy to check that $S_1 + S_2 \stackrel{d}{=} \mathbf{X}\mathbf{1}_n + \mathbf{Z}\mathbf{1}_n$. Note that $\mathbf{X}\mathbf{1}_n + \mathbf{Z}\mathbf{1}_n = (\mathbf{X} + \mathbf{Z})\mathbf{1}_n \in \mathfrak{S}_n(F_1 \otimes G_1, \cdots, F_n \otimes G_n)$. It follows that $S_1 + S_2 \in \mathfrak{S}_n(F_1 \otimes G_1, \cdots, F_n \otimes G_n)$.

(iv)-(vi) Trivial.

(vii) (a) Write $S = \mathbf{X}\mathbf{1}_n$, where $\mathbf{X} \in \mathfrak{F}_n(\mathbf{F})$ and let C be the copula of $\mathbf{X}$. Let $S_k = \mathbf{X}_k\mathbf{1}_n$, where $\mathbf{X}_k \in \mathfrak{F}_n(F_1^{(k)}, \cdots, F_n^{(k)})$ with copula C. It is obvious that $S_k \xrightarrow{d} S$.

(b) Write $S_k = \mathbf{X}_k\mathbf{1}_n$, where $\mathbf{X}_k \in \mathfrak{F}_n(F_1^{(k)}, \cdots, F_n^{(k)})$ with copula C_k. Note that the space of n-copulas is a compact space. Hence, there is a subsequence C_{k_i} of C_k such that C_{k_i} has a limit. Then the subsequence $S_{k_i} \xrightarrow{d} \mathbf{X}\mathbf{1}_n$ where $\mathbf{X} \in \mathfrak{F}_n(\mathbf{F})$ with copula C as the limit of C_{k_i}. Since $S_k \xrightarrow{d} S$, we have $S \stackrel{d}{=} \mathbf{X}\mathbf{1}_n \in \mathfrak{S}_n(\mathbf{F})$.

(c) This is a special case of (vi)(b).

□