Complete mixability and asymptotic equivalence of worst-possible VaR and ES estimates

Giovanni Puccetti[a], Bin Wang[b], Ruodu Wang[c,*]

[a] *School of Economics and Management, University of Firenze, 50127 Firenze, Italy.*
[b] *Huawei Technologies Co. Ltd. Beijing R&D Institute, Beijing 100095, China*
[c] *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada.*

Abstract

We give a new sufficient condition for a continuous distribution to be completely mixable, and we use this condition to show that the worst-possible Value-at-Risk for the sum of d inhomogeneous risks is equivalent to the worst-possible Expected Shortfall under the same marginal assumptions, in the limit as $d \to \infty$. Numerical applications show that this equivalence holds also for relatively small dimensions d.

Keywords: Complete mixability, Worst-dependence scenarios, Value-at-Risk, Expected Shortfall, Basel III
AMS 2010 Subject Classification: 60E05, 91B30

1. Motivation of the paper

In the past twenty years Value-at-Risk (VaR) has become the standard risk measure for the calculation of minimum capital requirements under the Advanced Measurement Approach (AMA) within the Basel II (shortly becoming Basel III) agreement. The VaR of an loss random variable L, computed at a probability level $\alpha \in (0,1)$, is the α-quantile of its distribution, defined as

$$\mathrm{VaR}_\alpha(L) := F_L^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F_L(x) > \alpha\}, \qquad (1.1)$$

where $F_L(x) = P(L \leq x)$ is the distribution function of L. In addition, let $F_L^{-1}(1) = \inf\{x \in \mathbb{R} : F_L(x) = 1\}$ be the essential supremum of the support of the distribution F. Probably due to its immediate meaning (the probability that L exceeds $\mathrm{VaR}_\alpha(L)$ is at most $1-\alpha$) and mature statistical estimation, VaR rapidly became the most popular risk measure used in banking and insurance. At the same time, the academic world warned against an irresponsible use of VaR in industry; see for instance the early discussions given in Artzner et al. (1999) and Embrechts et al. (2002). Indeed, a number of disadvantages have been identified with VaR, the most important ones being its non-coherence (in that it fails the subadditivity criterion possibly preventing diversification in a risk portfolio) and its inability to capture the magnitude of extreme losses. Finally, the recent financial crisis raised the question whether VaR is still suitable as the default risk metric. The recent document Basel Committee on Banking Supervision (2012) seems to raise the possibility of a change. In the words of the Committee (see p.41, question 8 in Basel Committee on Banking Supervision (2012)) :

What are the likely operational constraints with moving from VaR *to* ES*, including any challenges in delivering robust backtesting, and how might these be best overcome?*

Thus, Expected Shortfall (ES) seems to be the official candidate to replace VaR in the years to come. If the loss random variable L satisfies $\mathbb{E}[|L|] < \infty$, the ES computed at the confidence level $\alpha \in (0,1)$ is defined as the average of all VaRs above the α-level, i.e.

$$\mathrm{ES}_\alpha(L) := \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_q(L)\,dq.$$

Unlike VaR, ES is a *coherent* risk measure (see Artzner et al. (1999)) and considers both the size and likelihood of losses above the α-quantile. ES is also a more conservative risk measure in the sense that

$$\mathrm{ES}_\alpha(L) \geq \mathrm{VaR}_\alpha(L), \text{ for all } \alpha \in (0,1).$$

However, ES has the big disadvantage of not being computable when the random variable L under study does not possess a finite expectation. This deficiency is particularly relevant for instance in the treatment of Operational Risk, where the underlying loss random variables often follow infinite mean models; see Gourier et al. (2009).

In this paper we consider an aggregate loss random variable L of the form

$$L_d^+ := \sum_{i=1}^d L_i,$$

where $L_1, \ldots, L_d$ correspond to marginal loss random variables held by a bank/insurance company over a fixed time period. For example, each L_i may represent the yearly aggregate loss for a specific risk type/business line or can be seen as the aggregate

*Corresponding author. Address: Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave. W., Waterloo, ON N2L3G1, Canada.

Email addresses: giovanni.puccetti@unifi.it (Giovanni Puccetti), wbcj305@163.com (Bin Wang), wang@uwaterloo.ca (Ruodu Wang)

claim for a single yearly policy within a particular line of insurance.

The computation of $\mathrm{VaR}_\alpha(L_d^+)$ and $\mathrm{ES}_\alpha(L_d^+)$ requires the knowledge of the d-variate joint distribution function H of the risk portfolio $(L_1, \dots L_d)$. In principle, the statistical estimation of H requires a d-variate dataset for the past occurred losses. In practice, such d-dimensional dataset is seldom available and, typically, only the marginal distribution functions F_i of the L_i will be statistically estimated by the bank/insurance company. Throughout the paper, we will therefore assume that the marginal distributions of the marginal losses L_i are given but no dependence information is known about the joint portfolio $(L_1, \dots, L_d)$. Since there exist infinitely many joint models for $(L_1, \dots, L_d)$ which are consistent with a given set of d marginal distributions $F_1, \dots, F_d$, for a fixed $\alpha \in (0,1)$ we define the worst-case VaR and the worst-case ES for the aggregate position L_d^+ as

$$\overline{\mathrm{VaR}}_\alpha(L_d^+) := \sup\left\{\mathrm{VaR}_\alpha(L_1 + \cdots + L_d); L_i \stackrel{d}{=} F_i, 1 \le i \le d\right\}, \tag{1.2}$$

$$\overline{\mathrm{ES}}_\alpha(L_d^+) := \sup\left\{\mathrm{ES}_\alpha(L_1 + \cdots + L_d); L_i \stackrel{d}{=} F_i, 1 \le i \le d\right\}. \tag{1.3}$$

$\overline{\mathrm{VaR}}_\alpha(L_d^+)$ and $\overline{\mathrm{ES}}_\alpha(L_d^+)$ represent the largest (conservative) estimate of $\mathrm{VaR}_\alpha(L_d^+)$ and $\mathrm{ES}_\alpha(L_d^+)$, respectively, if only the marginal distributions of the random variables $L_1, \dots, L_d$ are known. Equivalently stated, if one assumes that the vector $(L_1^C, \dots, L_d^C)$ has marginals $F_1, \dots, F_d$ and a dependence structure in the form of a copula C, the following inequalities will hold for any possible choice of C:

$$\mathrm{VaR}_\alpha(L_1^C + \cdots + L_d^C) \le \overline{\mathrm{VaR}}_\alpha(L_d^+),$$
$$\mathrm{ES}_\alpha(L_1^C + \cdots + L_d^C) \le \overline{\mathrm{ES}}_\alpha(L_d^+).$$

We refer the reader unfamiliar with the concept of a copula to McNeil et al. (2005, Chapter 5). For instance, if we take that the risks $(L_1, \dots, L_d)$ are *comonotonic* as a special case of dependence, we have

$$\overline{\mathrm{VaR}}_\alpha(L_d^+) \ge \mathrm{VaR}_\alpha(L_1 + \cdots + L_d) = \sum_{i=1}^d \mathrm{VaR}_\alpha(L_i); \tag{1.4}$$

where the equality in (1.4) is implied by comononotic additivity of VaR; see McNeil et al. (2005, Proposition 6.15).

Using also coherence of ES, it is easy to prove that the worst-possible ES for L_d^+ is given by

$$\overline{\mathrm{ES}}_\alpha(L_d^+) = \sum_{i=1}^d \mathrm{ES}_\alpha(F_i) = \sum_{i=1}^d \frac{1}{1-\alpha}\int_\alpha^1 F_i^{-1}(q)\,dq, \tag{1.5}$$

where we use the notation $\mathrm{ES}_\alpha(F_i)$ to denote the ES of any random variable having distribution F_i.

Due to incoherence of VaR, the computation of $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ is much more difficult and, for a broad class of risk portfolios, still open. The *analytical* computation of the worst VaR estimate $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ is possible only in the homogeneous case where the L_i's are identically distributed with a continuous distribution having an ultimately decreasing density; see Wang et al. (2013) and Puccetti and Rüschendorf (2013). The *numerical* computation of $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ in the general case of inhomogeneous portfolios can be performed using the Rearrangement Algorithm described in Embrechts et al. (2013) for dimensions d in the several hundreds or possibly thousands. The numerical estimates obtained (via the rearrangement algorithm) in Puccetti (2013, Table 3) for $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ and $\overline{\mathrm{ES}}_\alpha(L_d^+)$ constituted the main motivation to investigate the asymptotic properties of the ratio $\overline{\mathrm{VaR}}_\alpha(L_d^+)/\overline{\mathrm{ES}}_\alpha(L_d^+)$.

The main result presented in this paper is that under weak marginal assumptions for the risk portfolio $(L_1, \dots, L_d)$ the worst-possible VaR estimate $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ is equivalent to the worst-possible ES estimate $\overline{\mathrm{ES}}_\alpha(L_d^+)$, in the limit as $d \to \infty$. Formally we have that

$$\lim_{d\to\infty} \frac{\overline{\mathrm{VaR}}_\alpha(L_d^+)}{\overline{\mathrm{ES}}_\alpha(L_d^+)} = 1. \tag{1.6}$$

This case is of particular interest as internal models built by financial institutions to fulfil the AMA approach in the Basel and Solvency regulatory frameworks typically use a large number d of risk factors. Roughly speaking, under a conservative rule for capital reserving, a VaR-based reserve will be equivalent to a ES-based one for the dimensions d typically used within quantitative risk management.

The limit result (1.6) has been proved in Puccetti and Rüschendorf (2014) for a portfolio of risks having homogeneous marginals with monotone densities satisfying some extra technical assumptions. In this paper we extend (1.6) to inhomogeneous risk portfolios using a new sufficient condition for a continuous distribution to be completely mixable. This new condition, given in Theorem 3.4 below, constitutes a strong theoretical achievement of independent interest in the theory of completely mixable distributions. Based on this new condition, the asymptotic equivalence of worst-VaR and worst-ES estimates holds for all inhomogeneous risk portfolios having a finite number of (possibly bounded) marginals with a continuous and positive density.

The practical consequences deriving from the limit result (1.6) are relevant:

- From a worst-case scenario perspective, a move from the VaR to the ES risk metric seems to be robust when the underlying mathematical models have finite expectations; this gives a partial answer to the highlighted question by the Basel committee.

- In the literature $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ has always been referred as a too much conservative estimate of the capital reserve, $\overline{\mathrm{ES}}_\alpha(L_d^+)$ to be preferred as its natural alternative. The result in (1.6) actually shows that the two estimates are asymptotically equivalent.

- Rewriting (1.6) as

$$\overline{\mathrm{VaR}}_\alpha(L_d^+) \stackrel{d\to\infty}{\simeq} \sum_{i=1}^d \frac{1}{1-\alpha}\int_\alpha^1 F_i^{-1}(q)\,dq,$$

gives a straightforward approximation to $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ which allows to avoid any advanced analytical and numerical techniques for the computation of the worst-possible VaR, even for inhomogeneous risk portfolios.

The facts described above seem to be even more relevant to regulation in banking and insurance as the numerical applications given in Section 5 suggest that the equivalence

$$\overline{\mathrm{VaR}}_\alpha(L_d^+) \overset{d\to\infty}{\simeq} \overline{\mathrm{ES}}_\alpha(L_d^+)$$

holds also for relatively small dimensions d.

Remark 1.1. *In some literature,* $\mathrm{VaR}_\alpha(L)$ *is defined as* $\inf\{x \in \mathbb{R} : F_L(x) \geq \alpha\}$. *In this paper we use* $\inf\{x \in \mathbb{R} : F_L(x) > \alpha\}$ *only for technical ease. The worst-possible value* $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ *for both definitions are equivalent for continuous marginals (for example, see Lemma 4.5 of Bernard, Jiang, and Wang (2013)). Hence, the major results in this paper hold for both definitions.*

2. Some preliminaries on complete mixability

As proved in Wang and Wang (2011), Wang et al. (2013) and Puccetti and Rüschendorf (2013), the concept of *complete mixability* plays a crucial role in the computation of the upper VaR-bound defined in (1.2). First, we give a summary of the existing results on completely mixable distributions which we will use in the remainder. Throughout the paper, we identify probability measures with the corresponding cumulative distribution functions, and we always intend a limit of a sequence of random variables under weak convergence.

Definition 2.1. *A distribution function F on $\mathbb{R}$ is called n-*completely mixable *(n-CM) if there exist n random variables $X_1, \dots, X_n$ identically distributed as F such that*

$$P(X_1 + \cdots + X_n = nk) = 1, \tag{2.1}$$

for some $k \in \mathbb{R}$. Any such k is called a center *of F and any vector $(X_1, \dots, X_n)$ satisfying (2.1) with $X_i \overset{d}{=} F, 1 \leq i \leq n$, is called an n-*complete mix.

If F is n-CM and has finite first moment μ, then its center is unique and equal to μ. We denote by $\mathcal{M}_n(\mu)$ the set of all n-CM distributions with center μ, and by $\mathcal{M}_n = \bigcup_{\mu\in\mathbb{R}} \mathcal{M}_n(\mu)$ the set of all n-CM distributions on $\mathbb{R}$.

Definition 2.2. *If X has distribution F, we say that F is n-CM on the interval $A \subset \mathbb{R}$ if the conditional distribution of $(X \mid X \in A)$ is n-CM.*

Complete mixability is a concept of *negative dependence*. It is easy to see for instance that in dimension $d = 2$ complete mixability implies countermonotonicity; see Embrechts et al. (2013). A completely mixable dependence structure minimizes the variance of the sum of risks with given marginal distributions. In fact, a risk vector $(L_1, \dots, L_d)$ with identically distributed marginals is a d-complete mix if and only if the variance of the sum of its components is equal to zero.

Not all univariate distributions F are d-CM: the set of all n-CM distributions can be completely described only when $n = 1$ or $n = 2$. The class $\mathcal{M}_1$ consist of all degenerate distributions, while the class $\mathcal{M}_2$ of all the symmetric distributions; see Proposition 2.3 in Wang and Wang (2011). The following results are given in Wang and Wang (2011), Wang et al. (2013) and Puccetti et al. (2012). We refer the reader to these references for more properties and several examples of completely mixable distributions.

Proposition 2.3. *The following statements hold (for weak convergence).*

(i) *The set $\mathcal{M}_n(\mu)$ is convex, i.e. $F, G \in \mathcal{M}_n(\mu)$ implies that $\lambda F + (1-\lambda)G \in \mathcal{M}_n(\mu)$, $\lambda \in [0,1]$. Note instead that the set $\mathcal{M}_n$ is* not *convex – a counterexample is given in Wang and Wang (2011).*
(ii) *If $F \in \mathcal{M}_n$ and $G \in \mathcal{M}_k$, then $\frac{n}{n+k}F + \frac{k}{n+k}G \in \mathcal{M}_{n+k}$. Consequently, $F \in \mathcal{M}_n \cap \mathcal{M}_k$ implies $F \in \mathcal{M}_{n+k}$*
(iii) *The limit of a sequence of n-CM distribution functions (with center μ) is n-CM (with center μ).*
(iv) *Fixed arbitrarily $a, b \in \mathbb{R}$ with $a < b$, the uniform distribution on $[a, b]$ is n-CM for all $n \geq 2$.*
(v) *A n-discrete uniform distribution, that is a distribution giving probability mass $1/n$ to each of the n points in its support, is n-CM.*
(vi) *Any continuous distribution F on $[0,1]$ having a non-decreasing density and satisfying the moderate mean condition*
$$\int_0^1 x dF(x) \leq 1 - 1/n$$
is n-CM.
(vii) *For an arbitrary distribution F on $[0,1]$ having mean μ, a necessary condition to be n-CM is that*
$$\frac{1}{n} \leq \mu \leq 1 - \frac{1}{n}.$$

Since we will use the properties listed in Proposition 2.3 frequently, we will refer to them just using (i)–(vii) throughout the paper.

3. A new sufficient condition to complete mixability

In this section we show that any continuous distribution on a finite interval having a strictly positive density function is n-completely mixable for n large enough.

First, we will need some subsidiary results. Denote by U the uniform distribution on $[0,1]$ and by $\mathcal{G}_k$, $k \geq 1$, the set of all discrete distributions function in $[0,1]$ having exactly k different points in their supports. For $n \geq 3$, define the set of distribution functions

$$\mathcal{N}_n(k) := \left\{ F : F = \frac{3}{n}U + \frac{n-3}{n}G_k, \text{ where } G_k \in \mathcal{G}_k \right\},\ k \geq 1. \tag{3.1}$$

We will prove by induction on k that, for a fixed $n \geq 3$, $\mathcal{N}_n(k) \subset \mathcal{M}_n, k \in \mathbb{N}$. The case $n = 3$ is implied by (iv), thus in the remainder of this section we assume $n \geq 4$. In order to use induction, we first need to prove the cases $k = 1$ and $k = 2$.

Lemma 3.1. $\mathcal{N}_n(1) \subset \mathcal{M}_n$.

Proof. Since $G_1 \in \mathcal{G}_1$ is degenerate, then $G_1 \in \mathcal{M}_{n-3}$. Moreover, $U \in \mathcal{M}_3$ by (iv). Using (ii), we obtain that $\frac{3}{n}U + \frac{n-3}{n}G_1 \in \mathcal{M}_n$. □

Lemma 3.2. $\mathcal{N}_n(2) \subset \mathcal{M}_n$.

Proof. Let δ_a be the degenerate distribution at point $a \in \mathbb{R}$. We can write any $F \in \mathcal{N}_n(2)$ as $F = \frac{3}{n}U + \frac{n-3}{n}G_2$, where $G_2 = p\delta_a + (1-p)\delta_b$, for some $0 \le a < b \le 1$ and $p \in (0,1)$. Denote by μ the mean of F. We have

$$\mu = \frac{3}{2n} + \frac{n-3}{n}(pa + (1-p)b).$$

[CASE 1] Assume that $l := p(n-3) \in \mathbb{N}$. Obviously $0 < l < n-3$. Then we can write G_2 as a uniform distribution on the set $\{x_1, \dots, x_{n-3}\}$ by taking $x_i = a, 1 \le i \le l$, and $x_i = b, l+1 \le i \le n-3$. Therefore, G_2 is a $(n-3)$-discrete distribution and thus $G_2 \in \mathcal{M}_{n-3}$ by (v). The lemma then follows analogously to Lemma 3.1.

[CASE 2] If $p(n-3) \notin \mathbb{N}$, we take $l := \lfloor p(n-3) \rfloor$ and $r := p(n-3) - l$. In this case we can write

$$G_2 = (1-r)H_1 + rH_2 \tag{3.2}$$

as the convex combination of two $(n-3)$-discrete distributions H_1 and H_2. We take $H_1 := \frac{1}{n-3}(l\delta_a + (q+1)\delta_b)$ and $H_2 := \frac{1}{n-3}((l+1)\delta_a + q\delta_b)$, where $q := n-4-l$. Note that we can assume $r \ge 1/2$ wlog. At this point, H_1 can be seen a uniform distribution on the set $\{x_1, \dots, x_{n-3}\}$ by taking $x_i = a, 1 \le i \le l$, and $x_i = b, l+1 \le i \le n-3$. Analogously for H_2. Therefore we have that $H_1, H_2 \in \mathcal{M}_{n-3}$. Note that this does not directly imply by (ii) that G_2 is $(n-3)$-CM as H_1 and H_2 may have different means.

Now, let F_1 be the uniform distribution on the interval $[0, 1 - 2(b-a)r/3]$. $F_1 \in \mathcal{M}_3$ by (iv) while $H_1 \in \mathcal{M}_{n-3}$ as seen above. Using (ii) we conclude that

$$\frac{3}{n}F_1 + \frac{n-3}{n}H_1 \in \mathcal{M}_n. \tag{3.3}$$

Let also

$$F_2 := 1/r\,(U - (1-r)F_1). \tag{3.4}$$

One can easily check that F_2 is a well-posed, continuous distribution function, with a nondecreasing density and mean $1/2 + 1/3(1-r)(b-a) < 2/3$. Therefore, $F_2 \in \mathcal{M}_n$ by (vi). Since $H_2 \in \mathcal{M}_{n-3}$ as seen above, using (ii) we conclude that

$$\frac{3}{n}F_2 + \frac{n-3}{n}H_2 \in \mathcal{M}_n. \tag{3.5}$$

Using (3.2) and (3.4), we can decompose any $F \in \mathcal{N}_n(2)$ as

$$\begin{aligned} F &= \frac{3}{n}U + \frac{n-3}{n}G_2 \\ &= \frac{3}{n}[(1-r)F_1 + rF_2] + \frac{n-3}{n}[(1-r)H_1 + rH_2] \\ &= (1-r)\left(\frac{3}{n}F_1 + \frac{n-3}{n}H_1\right) + r\left(\frac{3}{n}F_2 + \frac{n-3}{n}H_2\right). \end{aligned} \tag{3.6}$$

It is easy to check that $\left(\frac{3}{n}F_1 + \frac{n-3}{n}H_1\right)$ has the same mean μ as F. Therefore, (3.6) implies that also $\left(\frac{3}{n}F_2 + \frac{n-3}{n}H_2\right)$ has mean μ. At this point (3.3) and (3.5) directly implies that

$$\frac{3}{n}F_1 + \frac{n-3}{n}H_1 \in \mathcal{M}_n(\mu) \text{ and } \frac{3}{n}F_2 + \frac{n-3}{n}H_2 \in \mathcal{M}_n(\mu).$$

Therefore, by (3.6), $F \in \mathcal{N}_n(2)$ is the convex combination of distributions in $\mathcal{M}_n(\mu)$, i.e. $F \in \mathcal{M}_n(\mu)$. □

Lemma 3.3. $\mathcal{N}_n(k) \subset \mathcal{M}_n, k \in \mathbb{N}$.

Proof. We proceed by induction, thus we assume that $\mathcal{N}_n(k-1) \subset \mathcal{M}_n$, with $k \ge 3$. We can write any $F \in \mathcal{N}_n(k)$ as

$$F = \frac{3}{n}U + \frac{n-3}{n}G_k,$$

with $G_k \in \mathcal{G}_k$. Denote by ξ the mean of G_k. Let B_k be the probability mass function of G_k and a and b be the minimal and maximal, respectively, point in the support of G_k. Since $k \ge 3$, we have $a < b$ and therefore it is possible to find $\lambda \in (0,1)$ such that

$$\xi = (\lambda a + (1-\lambda)b).$$

Note also that, consisting the support of G_k of $k \ge 3$ points, we have that $B_k(a) < \lambda$ and $B_k(b) < 1-\lambda$. For μ, the mean of F, we have

$$\mu = \frac{3}{2n} + \frac{n-3}{n}\xi = \frac{3}{2n} + \frac{n-3}{n}(\lambda a + (1-\lambda)b).$$

Now, define $H := (\lambda\delta_a + (1-\lambda)\delta_b)$ and $F_1 := \frac{3}{n}U + \frac{n-3}{n}H$. Since $F_1 \in \mathcal{N}_n(2)$, by Lemma 3.2 we have that $F_1 \in \mathcal{M}_n(\mu)$. Now let $\zeta = \min\{B_k(a)/\lambda, B_k(b)/(1-\lambda)\}$. Then $Z := \frac{G_k - \zeta H}{1-\zeta}$ is a discrete distribution with exactly $(k-2)$ (if $B_k(a)/\lambda = B_k(b)/(1-\lambda)$) or $(k-1)$ (otherwise) points in its support, i.e. $Z \in \mathcal{N}_n(k-1) \cup \mathcal{N}_n(k-2) \subset \mathcal{M}_n$. By the induction assumption

$$F_2 := \frac{3}{n}U + \frac{n-3}{n}Z \in \mathcal{N}_n(k-1) \cup \mathcal{N}_n(k-2) \subset \mathcal{M}_n.$$

At this point, it is straightforward to check that $F = \zeta F_1 + (1-\zeta)F_2$. Since both F_1 and F have mean μ, also F_2 has mean μ and therefore we obtain that $F \in \mathcal{N}_n(k)$ is the convex combination of distributions in $\mathcal{M}_n(\mu)$, i.e. $F \in \mathcal{M}_n(\mu)$. □

We are now ready to prove the main result of this section.

Theorem 3.4. *Assume $n \ge 3$. Any continuous distribution function F on a bounded interval $[a,b]$, $a < b$, having a density f satisfying*

$$f(x) \ge \frac{3}{n(b-a)}, \text{ for all } x \in [a,b], \tag{3.7}$$

is n-CM.

Proof. Without loss of generality we take $a = 0$ and $b = 1$. Note that any continuous distribution F on $[0,1]$ having a density function $f \ge \frac{3}{n}$ can be written as $F = \frac{3}{n}U + \frac{n-3}{n}G$ where G is a continuous distribution on $[0,1]$. Let $G_k(x) = \sum_{i=1}^k G(\frac{i}{k})\mathbf{1}\{\frac{i-1}{k} < x \le \frac{i}{k}\} \in \mathcal{G}_k$ for $k \in \mathbb{N}$. Then $G_k \to G$ weakly. Thus, F is the limit of a sequence $F_k := \frac{3}{n}U + \frac{n-3}{n}G_k$, $k \in \mathbb{N}$, with $F_k \in \mathcal{N}_n(k)$. Therefore the theorem follows by Lemma 3.3 and Proposition 2.3 (iii). For the general case $a < b$, it is sufficient to take U as the uniform distribution in $[a, b]$ in the definition of $\mathcal{N}_n(k)$ in (3.1). □

Corollary 3.5. *Any continuous distribution on $[a, b]$ having a strictly positive density function is n-CM for n sufficiently large.*

In Section 7 below we discuss the sharpness of the sufficient condition given in (3.7).

4. Asymptotic equivalence of worst-case VaR and ES for a sum

Let $(L_d, d \in \mathbb{N})$ be an infinite sequence of random variables for which the marginal distributions are assumed to be known. In practice, a financial institution has to deal with a inhomogenous risk portfolio consisting of finite types of random variables, i.e. a finite number of different marginal distributions. Thus, we assume that the sequence $(L_d, d \in \mathbb{N})$ can be divided in a finite number of m homogeneous subgroups. Given a set of m marginal distributions $F_1, \dots, F_m$, we assume that for any $i \in \mathbb{N}$ we have

$$L_i \stackrel{d}{=} F_j \text{ for some } j \in M := \{1, \dots, m\}. \tag{4.1}$$

Thus, we set

$$A_j(d) := \left\{ i \le d : L_i \stackrel{d}{=} F_j \right\} \quad \text{and} \quad d_j(d) := \#A_j(d).$$

Given a confidence level $\alpha \in (0, 1)$, we assume that each F_j is continuous with a positive and continuous density f_j on the (possibly unbounded) interval $[F_j^{-1}(\alpha), F_j^{-1}(1))$. Note that these assumptions cover all the continuous distributional models used in quantitative risk management, where one has typically to deal with unbounded loss random variables with a positive and continuous density. In case $F_j^{-1}(1) = \infty$ we will assume further that the mean of F_j is finite in order to guarantee the existence of $\mathrm{ES}_\alpha(F_j)$. The behaviour of the VaR metric in case the individual random losses have infinite first moment has been studied in Puccetti and Rüschendorf (2014). Finally, we assume that for each $j = 1, \dots, m$

$$\mathrm{ES}_\alpha(F_j) > 0, \tag{4.2}$$

i.e. a risk with distribution F_j requires a capital to be reserved.

For $u \in [F_j^{-1}(\alpha), F_j^{-1}(1))$, we denote by $K_j(u)$ the minimal value of the density of F_j in the interval $[F_j^{-1}(\alpha), u]$, i.e.

$$K_j(u) := \inf_{x \in [F_j^{-1}(\alpha), u]} f_j(x). \tag{4.3}$$

For a continuous density f_j the inf in (4.3) is attained and it is positive for any $u \in [F_j^{-1}(\alpha), F_j^{-1}(1))$. Since each $d_j(d)$ is increasing on d, we can also define

$$\mathcal{J} := \{j : \lim_{d\to\infty} d_j(d) = \infty\}, \tag{4.4}$$

$$\overline{\mathcal{J}} := \{j : \lim_{d\to\infty} d_j(d) < \infty\} = M \setminus \mathcal{J}. \tag{4.5}$$

Being m, the number of possible marginal models, a finite constant, it is clear that $\mathcal{J}$ contains at least one index $j \in M$.

For $j \in \mathcal{J}$, define the sequence

$$R_j(d) := \sup \Big\{ u \in [F_j^{-1}(\alpha), F_j^{-1}(1)) : K_j(u)\left(u - F_j^{-1}(\alpha)\right) \ge \frac{3}{d_j(d)} \Big\}. \tag{4.6}$$

Denote $\hat{u} = F_j^{-1}(\alpha) + \xi, \xi > 0$. Since $j \in \mathcal{J}$, $d_j(d) \to \infty$ implies that

$$K_j(\hat{u})\left(\hat{u} - F_j^{-1}(\alpha)\right) = K_j(\hat{u})\xi > \frac{3}{d_j(d)}, \tag{4.7}$$

i.e. $R_j(d)$ is well defined, for d large enough. If $F_j^{-1}(1)$ is finite, $R_j(d)$ is also bounded. If $F_j^{-1}(1) = \infty$, we show that $R_j(d) < \infty$ for d sufficiently large. Being f_j the density of a continuous, finite-mean distribution, we have that $\int_{\mathbb{R}} u f_j(u)\, du < \infty$, implying that

$$\lim_{u\to\infty} K_j(u)u \le \lim_{u\to\infty} f_j(u)u = 0, \quad \lim_{u\to\infty} K_j(u) = 0.$$

and

$$\lim_{u\to\infty} K_j(u)\left(u - F_j^{-1}(\alpha)\right) = 0.$$

Recalling (4.7), $R_j(d)$ turns out to be finite for d sufficiently large with

$$\lim_{d\to\infty} R_j(d) = F_j^{-1}(1). \tag{4.8}$$

Before proving the main result of this section, we need the following result.

Lemma 4.1. *Assume F is continuous with a continuous and positive density on $[F^{-1}(\alpha), F^{-1}(1))$. If F is d-CM on the interval $[F^{-1}(\alpha), R]$, $R > F^{-1}(\alpha)$, then it is possible to find a random vector $(L_1^*, \dots, L_d^*)$, such that $L_i^* \stackrel{d}{=} F$ for $i = 1, \dots, d$ and*

$$\mathrm{VaR}_\alpha\left(L_1^* + \cdots + L_d^*\right) = d\mu^*. \tag{4.9}$$

where

$$\mu^* = \frac{1}{F(R) - \alpha} \int_{F^{-1}(\alpha)}^{R} F_j(q)\, dq.$$

Proof. Let H be the distribution of the random variable $\left(X \mid X \in [F^{-1}(\alpha), R]\right)$, where $X \stackrel{d}{=} F$. Under the assumptions of the theorem there exist d random variables $\hat{L}_i, 1 \le i \le d$, such that $\hat{L}_i \stackrel{d}{=} H$ and

$$P\left(\sum_{i=1}^{d} \hat{L}_i = d\mu^*\right) = 1.$$

This implies the existence of a vector $(L_1^*, \dots, L_d^*)$ satisfying (4.9). For instance, $(L_1^*, \dots, L_d^*)$ can be defined as

$$\boldsymbol{L}^* = (X, \dots, X)\, \mathbf{1}\left(X \notin [F^{-1}(\alpha), R]\right) + (\hat{L}_1, \dots, \hat{L}_d)\, \mathbf{1}\left(X \in [F^{-1}(\alpha), R]\right). \square$$

We are now ready to state the main result of this section.

Theorem 4.2. *Let $F_1, \dots, F_m$, m fixed, be a set of m marginal distributions. Assume that each F_j has finite mean and is continuous with a continuous and positive density f_j on $[F_j^{-1}(\alpha), F_j^{-1}(1))$. If $(L_d, d \in \mathbb{N})$ is an infinite sequence of random variables for which* (4.1) *and* (4.2) *holds, then*

$$\lim_{d\to\infty} \frac{\overline{\mathrm{VaR}}_\alpha(L_1 + \dots + L_d)}{\overline{\mathrm{ES}}_\alpha(L_1 + \dots + L_d)} = 1.$$

Proof. Fix a sufficiently large $d \in \mathbb{N}$ such that $R_j(d)$ in (4.6) is well defined for all $j \in \mathcal{J}$. For $j \in J$ we have that

$$f_j(u) \geq K_j(R_j(d)) \geq \frac{3}{d_j(d)(R_j(d) - F_j^{-1}(\alpha))},$$

for all $u \in [F_j^{-1}(\alpha), R_j(d)]$. By Theorem 3.4, the distribution F_j is $d_j(d)$-CM on the interval $[F_j^{-1}(\alpha), R_j(d)]$. By Lemma 4.1 it is then possible to find a d_j-complete mix $\hat{\boldsymbol{L}}^j := \left(\hat{L}_i^j, i \in A_j(d)\right)$, such that $\hat{L}_i^j \overset{d}{=} F_j$ for $i \in A_j(d)$ and

$$\mathrm{VaR}_\alpha\left(\sum_{i\in A_j(d)} \hat{L}_i^j\right) = d_j(d)\mu^j(d), \tag{4.10}$$

where

$$\mu^j(d) = \frac{1}{F_j(R_j(d)) - \alpha} \int_{F_j^{-1}(\alpha)}^{R_j(d)} F_j^{-1}(q)\, dq.$$

For $j \in \mathcal{J}$, (4.8) implies that

$$\lim_{d\to\infty} \mu^j(d) = \mathrm{ES}_\alpha(F_j). \tag{4.11}$$

Using (4.10) and (1.4) we have that

$$\overline{\mathrm{VaR}}_\alpha(L_d^+) \geq \sum_{j\in\mathcal{J}} \mathrm{VaR}_\alpha\left(\sum_{i\in A_j(d)} \hat{L}_i^j\right) + \mathrm{VaR}_\alpha\left(\sum_{j\in\overline{\mathcal{J}}}\sum_{i\in A_j(d)} \hat{L}_i^j\right)$$
$$\geq \sum_{j\in\mathcal{J}} d_j(d)\mu^j(d) + \mathrm{VaR}_\alpha\left(\sum_{j\in\overline{\mathcal{J}}}\sum_{i\in A_j(d)} \hat{L}_i^j\right) =: \mathrm{VaR}_\alpha^*(L_d^+).$$

Thus, the following relationships hold

$$\mathrm{VaR}_\alpha^*(L_d^+) \leq \overline{\mathrm{VaR}}_\alpha(L_d^+) \leq \overline{\mathrm{ES}}_\alpha(L_d^+).$$

implying that ($\overline{\mathrm{ES}}_\alpha(L_d^+) > 0$)

$$\lim_{d\to\infty} \frac{\mathrm{VaR}_\alpha^*(L_d^+)}{\overline{\mathrm{ES}}_\alpha(L_d^+)} \leq \lim_{d\to\infty} \frac{\overline{\mathrm{VaR}}_\alpha(L_d^+)}{\overline{\mathrm{ES}}_\alpha(L_d^+)} \leq 1.$$

In order to prove the theorem it is then sufficient to show that

$$\lim_{d\to\infty} \frac{\mathrm{VaR}_\alpha^*(L_d^+)}{\overline{\mathrm{ES}}_\alpha(L_d^+)} = 1.$$

Since the marginals of the sequence $(L_d, d \in \mathbb{N})$ satisfy (4.1), for any fixed d (1.5) implies

$$\overline{\mathrm{ES}}_\alpha(L_d^+) = \sum_{j\in\mathcal{J}} d_j(d)\mathrm{ES}_\alpha(F_j) + \sum_{j\in\overline{\mathcal{J}}} d_j(d)\mathrm{ES}_\alpha(F_j).$$

Using the equation above, we obtain that

$$\lim_{d\to\infty} \frac{\mathrm{VaR}_\alpha^*(L_d^+)}{\overline{\mathrm{ES}}_\alpha(L_d^+)} = \lim_{d\to\infty} \frac{\sum_{j\in\mathcal{J}} d_j(d)\mu^j(d) + \mathrm{VaR}_\alpha\left(\sum_{j\in\overline{\mathcal{J}}}\sum_{i\in A_j(d)} \hat{L}_i^j\right)}{\sum_{j\in\mathcal{J}} d_j(d)\mathrm{ES}_\alpha(F_j) + \sum_{j\in\overline{\mathcal{J}}} d_j(d)\mathrm{ES}_\alpha(F_j)}$$
$$= \lim_{d\to\infty} \frac{\frac{\sum_{j\in\mathcal{J}} d_j(d)\mu^j(d)}{\sum_{j\in\mathcal{J}} d_j(d)\mathrm{ES}_\alpha(F_j)} + \frac{\mathrm{VaR}_\alpha\left(\sum_{j\in\overline{\mathcal{J}}}\sum_{i\in A_j(d)} \hat{L}_i^j\right)}{\sum_{j\in\mathcal{J}} d_j(d)\mathrm{ES}_\alpha(F_j)}}{\left(1 + \frac{\sum_{j\in\overline{\mathcal{J}}} d_j(d)\mathrm{ES}_\alpha(F_j)}{\sum_{j\in\mathcal{J}} d_j(d)\mathrm{ES}_\alpha(F_j)}\right)}. \tag{4.12}$$

Now let $\gamma_j := \lim_{d\to\infty} d_j(d) < \infty$ for $j \in \overline{\mathcal{J}}$. Recalling the definition of the subsets $\mathcal{J}$ and $\overline{\mathcal{J}}$, given in (4.5), it is clear that

$$\lim_{d\to\infty} \mathrm{VaR}_\alpha\left(\sum_{j\in\overline{\mathcal{J}}}\sum_{i\in A_j(d)} \hat{L}_i^j\right) \leq \lim_{d\to\infty} \overline{\mathrm{ES}}_\alpha\left(\sum_{j\in\overline{\mathcal{J}}}\sum_{i\in A_j(d)} \hat{L}_i^j\right)$$
$$\leq \lim_{d\to\infty}\left(\sum_{j\in\overline{\mathcal{J}}} d_j\mathrm{ES}_\alpha(F_j)\right) \leq \sum_{j\in\overline{\mathcal{J}}} \gamma_j\mathrm{ES}_\alpha(F_j), \tag{4.13}$$

while

$$\lim_{d\to\infty} \sum_{j\in\mathcal{J}} d_j\mathrm{ES}_\alpha(F_j) = \infty. \tag{4.14}$$

Using (4.11) and (4.14), we finally obtain from (4.12) that

$$\lim_{d\to\infty} \frac{\mathrm{VaR}_\alpha^*(L_d^+)}{\overline{\mathrm{ES}}_\alpha(L_d^+)} = 1. \quad \square$$

Remark 4.3. We remark the following points about Theorem 4.2.

Optimal couplings. Even if the asymptotic result

$$\overline{\mathrm{VaR}}_\alpha(L_d^+) \overset{d\to\infty}{\simeq} \overline{\mathrm{ES}}_\alpha(L_d^+) \tag{4.15}$$

holds, the structure of dependence attaining $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ and $\overline{\mathrm{ES}}_\alpha(L_d^+)$ for a fixed d might be different. These optimal dependence structure are also called *optimal couplings*. Following Theorem 2.1 in Puccetti and Rüschendorf (2013), it is possible to show that any optimal coupling for the VaR/ES can be described only on the upper part $T_j := \{x \geq F_j^{-1}(\alpha)\}$ of the support of each marginal distribution $F_1, \dots, F_d$ involved. It is well known that the optimal coupling for the ES is given by a *comonotonic* dependence structure in $T := \prod_{j=1}^d T_j$.

For more details about comonotonicity we refer for instance to McNeil et al. (2005). In the homogeneous case when $F_1 = \cdots = F_d = F$, optimal couplings for the VaR are described in the papers Wang and Wang (2011); Wang et al. (2013); Puccetti and Rüschendorf (2013). Whereas the optimal coupling for the ES is the *maximal* element wrt convex order for the Fréchet class of risk having fixed marginals on T (see also Levy and Kroll (1978) on this), the optimal coupling for the VaR is the *minimal* element wrt convex order over the same class of joint distributions on T.

Computation of the worst-possible VaR for a sum with given marginals. Rewriting (4.15) as

$$\overline{\mathrm{VaR}}_\alpha(L_d^+) \overset{d\to\infty}{\simeq} \sum_{i=1}^{d} \frac{1}{1-\alpha} \int_\alpha^1 F_i^{-1}(q)\,dq, \tag{4.16}$$

gives a straightforward approximation to $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ which allows to avoid any advanced analytical and numerical techniques. In Section 5 we show that (4.16) holds for relatively small dimensions d.

Bounded marginal distributions. For $m = 1$ the equality $\overline{\mathrm{VaR}}_\alpha(L_d^+) = \overline{\mathrm{ES}}_\alpha(L_d^+)$ holds for any sufficiently large d when the continuous distribution F_1 has a continuous and positive density over a bounded support. In this case, F_1 will be d-CM on $[F_1^{-1}(\alpha), F_1^{-1}(1)]$ for for d large enough. Analogously to the proof of Lemma 4.1, it is possible to construct a random vector $(L_1^*, \dots, L_d^*)$ with $L_i^* \overset{d}{=} F_1$ such that

$$P\left(\sum_{i=1}^{d} L_i^* = d\mathrm{ES}_\alpha(F_1) \mid \cap_{i=1}^{d} \{L_i^* \in [F_1^{-1}(\alpha), F_1^{-1}(1)]\}\right) = 1,$$

implying that $\overline{\mathrm{VaR}}_\alpha(L_d^+) = d\mathrm{ES}_\alpha(F_1) = \overline{\mathrm{ES}}_\alpha(L_d^+)$. The idea of obtaining sharp bounds for the distribution of the sum of risks under complete mixability goes back to Theorem 2.3 and Remark 2.4 in Puccetti and Rüschendorf (2012). However, finding weak sufficient conditions to complete mixability seems to be very difficult, especially in the inhomogeneous case.

Linear combinations of the L_i. Given a set of m continuous and strictly increasing functions h_j, $j \in M$, any result given for the aggregate random loss L_d^+ can be formulated for the aggregate position $L_d^h = \sum_{j=1}^m \sum_{i\in A_j(d)} h_j(L_i)$ just by scaling the corresponding marginal distributions. This includes any linear combination of the marginal losses with a finite number of weights.

5. Numerical verifications

In this section we compute the ratio

$$\delta_\alpha(d) := \frac{\overline{\mathrm{ES}}_\alpha(L_d^+)}{\overline{\mathrm{VaR}}_\alpha(L_d^+)}$$

for several risk portfolios $(L_1, \dots, L_d)$ of interest in quantitative risk management and for different confidence levels α.

First, we assume that the L_i's can be divided into $m = 3$ homogeneous subgroups, each one consisting of k random variables, i.e. $d = 3k$. Within the j-subgroup, the k random variables are assumed to be identically distributed as F_j, for some given marginal distributions F_1, F_2, F_3. Under these assumptions we study two portfolios of interest:

- the *Pareto Portfolio*, in which the marginal distributions $F_i = 1 - (1+x)^{-\theta_i}$ are of Pareto type with tail coefficients $\theta_i = i + 1, 1 \le i \le 3$. The variance of the sum of the components of this risk portfolio is infinite.
- the Mixed Portfolio, in which $F_1 = \mathrm{Pareto}(4)$, $F_2 = \mathrm{LogN}(0, \sigma^2)$, $F_3 = \mathrm{Exp}(3/\sqrt{2})$. The parameter σ^2 is set so that all the marginal components of this portfolio have the same variance.

In Table 1 we give estimates of $\delta_\alpha(d)$ for the two portfolios above, for different values of k and confidence levels α. Whereas estimates for $\overline{\mathrm{ES}}_\alpha(L_d^+)$ are available analytically via (1.5), in order to compute estimates for $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ we use the Rearrangement Algorithm as described in Embrechts et al. (2013).

As a second example, we assume $m = 9$ homogeneous subgroups of k random variables, i.e. $d = 9k$. We study two portfolios of interest:

- the *Pareto Portfolio*, in which the marginal distributions $F_i = 1 - (1+x)^{-\theta_i}$ are of Pareto type with tail coefficients $\theta_i = i + 1, 1 \le i \le 9$. The variance of the sum of the components of this risk portfolio is infinite.
- the Mixed Portfolio, in which $F_1 = \mathrm{Pareto}(4)$, $F_2 = \mathrm{LogN}(0, \sigma^2)$, $F_3 = \mathrm{Exp}(3/\sqrt{2})$, $F_4 = \mathrm{N}(0, \sqrt{2}/3)$, $F_5 = \mathrm{Gamma}(1/9, \sqrt{2})$, $F_6 = \mathrm{Weibull}(w, 1)$. The parameters σ^2 and w are set so that all the marginal components of this portfolios have the same variance.

In Table 2 we give estimates of $\delta_\alpha(d)$ for the two portfolios above, for different values of k and confidence levels α.

The figures in Tables 1–2 confirm the validity of the limit result given in Theorem 4.2. The convergence of the sequence $\delta_\alpha(d)$ to the limit value 1 is evident for all portfolio dimensions and confidence levels considered. For the mixed portfolios described above, the approximation

$$\overline{\mathrm{VaR}}_\alpha(L_d^+) \simeq \overline{\mathrm{ES}}_\alpha(L_d^+) = \sum_{i=1}^{d} \frac{1}{1-\alpha} \int_\alpha^1 F_i^{-1}(q)\,dq$$

is accurate also for the case $k = 1$, where one has a small number $d \le 6$ of completely inhomogeneous random variables. For the more heavy-tailed Pareto portfolio, the convergence of the sequence $\delta_\alpha(d)$ is slower but still visible for dimensions $d \ge 10$.

6. Conclusions

When the risk portfolio $(L_1, \dots, L_d)$ is inhomogeneous with a finite number m of different finite-mean marginal models, we

$\alpha = 99.9\%$	$k = 1$	$k = 5$	$k = 10$	$k = 20$
Pareto portfolio	1.3821	1.0672	1.0325	1.0160
Mixed portfolio	1.0639	1.0032	1.0005	1.0000

$\alpha = 99.5\%$	$k = 1$	$k = 5$	$k = 10$	$k = 20$
Pareto portfolio	1.3333	1.0590	1.0287	1.0142
Mixed portfolio	1.0498	1.0022	1.0004	1.0000

$\alpha = 99.0\%$	$k = 1$	$k = 5$	$k = 10$	$k = 20$
Pareto portfolio	1.3163	1.0558	1.0272	1.0134
Mixed portfolio	1.0453	1.0018	1.0003	1.0000

$\alpha = 50.0\%$	$k = 1$	$k = 5$	$k = 10$	$k = 20$
Pareto portfolio	1.3668	1.0528	1.0256	1.0126
Mixed portfolio	1.0495	1.0005	1.0001	1.0000

$\alpha = 20.0\%$	$k = 1$	$k = 5$	$k = 10$	$k = 20$
Pareto portfolio	1.4488	1.0594	1.0286	1.0140
Mixed portfolio	1.0565	1.0004	1.0000	1.0000

Table 1: Estimates of $\delta_\alpha(d)$ for two inhomogeneous portfolios divided into $m = 3$ homogeneous subgroups having Pareto or mixed marginal distributions. The total number of random variables is $d = 3k$.

$\alpha = 99.9\%$	$k = 1$	$k = 2$	$k = 5$	$k = 10$
Pareto portfolio	1.2950	1.1477	1.0565	1.0276
Mixed portfolio	1.0199	1.0051	1.0006	1.0000

$\alpha = 99.5\%$	$k = 1$	$k = 2$	$k = 5$	$k = 10$
Pareto portfolio	1.2345	1.1180	1.0461	1.0227
Mixed portfolio	1.0149	1.0030	1.0003	1.0000

$\alpha = 99.0\%$	$k = 1$	$k = 2$	$k = 5$	$k = 10$
Pareto portfolio	1.2122	1.1067	1.0419	1.0207
Mixed portfolio	1.0137	1.0024	1.0002	1.0000

$\alpha = 50.0\%$	$k = 1$	$k = 2$	$k = 5$	$k = 10$
Pareto portfolio	1.1651	1.0774	1.0298	1.0147
Mixed portfolio	1.0089	1.0012	1.0000	1.0000

$\alpha = 20.0\%$	$k = 1$	$k = 2$	$k = 5$	$k = 10$
Pareto portfolio	1.1857	1.0847	1.0322	1.0158
Mixed portfolio	1.0063	1.0008	1.0000	1.0000

Table 2: Estimates of $\delta_\alpha(d)$ for two inhomogeneous portfolios divided into $m = 6$ homogeneous subgroups having Pareto or mixed marginal distributions. The total number of random variables is $d = 9k$.

show that the worst-possible VaR estimate $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ is equivalent to the worst-possible ES estimate $\overline{\mathrm{ES}}_\alpha(L_d^+)$, in the limit as $d \to \infty$. Formally we have that

$$\lim_{d\to\infty} \frac{\overline{\mathrm{VaR}}_\alpha(L_d^+)}{\overline{\mathrm{ES}}_\alpha(L_d^+)} = 1.$$

Roughly speaking, under a conservative rule for capital reserving, a VaR-based reserve will be equivalent to a ES-based one for the dimensions d typically used within quantitative risk management. The numerical applications given in Section 5 shows that the above limit is evident also for relatively small dimensions d. This implies that, from a worst-case scenario perspective, a possible move from the VaR to the ES risk metric seems to be robust when the underlying mathematical models have finite expectations. Moreover, the limit result $\overline{\mathrm{VaR}}_\alpha(L_d^+) \overset{d\to\infty}{\simeq} \overline{\mathrm{ES}}_\alpha(L_d^+)$ gives a straightforward approximation to $\overline{\mathrm{VaR}}_\alpha(L_d^+)$ which allows to avoid any advanced analytical and numerical techniques for the computation of the worst-possible VaR, even in the case of inhomogeneous risk portfolios.

In order to prove our main result, we use a new sufficient condition for a continuous distribution to be completely mixable. This new condition constitutes a strong theoretical achievement of independent interest in the theory of completely mixable distributions.

It would be interesting to find sufficient conditions for the VaR/ES asymptotic equivalence to hold also when the number of different marginal models m is allowed to depend on d. This would be for instance in the case of an infinite sequence of random variables $(L_d, d \in \mathbb{N})$ having all different marginal distributions. In general it is easy to show that for a infinite sequence of random variables having rapidly increasing variance, the limit result above does not hold. We will investigate this topic in future research.

7. Appendix: on the sharpness of condition (3.7)

In this section we investigate whether the sufficient condition (3.7) in Theorem 3.4 can be weakened to

$$f(x) \geq \frac{\lambda}{n(b-a)}, \tag{7.1}$$

for some positive $\lambda < 3$. Define

$$\mathcal{N}_n^\lambda(k) := \left\{F : F = \frac{\lambda}{n}U + \frac{n-\lambda}{n}G_k, \text{ where } G_k \in \mathcal{G}_k\right\}.$$

In Lemma 3.3 we showed that $\mathcal{N}_n^3(k) = \mathcal{N}_n(k) \subset \mathcal{M}_n$ for any integer k, and this result directly implied Theorem 3.4. It is immediate to show that $\mathcal{N}_n^\lambda(k) \not\subset \mathcal{M}_n$ when $\lambda < 2$. Define

$$F := \left(\frac{\lambda}{n}U + \frac{n-\lambda}{n}\,\delta_0\right) \in \mathcal{N}_n^\lambda(1).$$

Clearly, F has mean $\mu = \frac{\lambda}{2n} < \frac{1}{n}$ and does not satisfy the necessary condition in (vii). Thus $F \notin \mathcal{M}_n$ for $\lambda < 2$. A similar construction shows that $\mathcal{N}_n^\lambda(k) \not\subset \mathcal{M}_n$ for any integer k.

As a consequence, we need at least $\lambda \geq 2$ in (7.1). In Example 7.2 we show that we actually need $\lambda \geq 1 + \sqrt{2}$. First, we give a new necessary condition for a distribution to n-CM.

Lemma 7.1. *If F is a nonnegative, n-CM distribution with finite first moment μ, then*

$$\overline{F}\left(\frac{n\mu}{k}\right) := 1 - F\left(\frac{n\mu}{k}\right) \leq \frac{k-1}{n}, \textit{ for all } k = 1, \dots, n.$$

Proof. Since F is n-CM, there exist a n-complete mix $(X_1, \dots, X_n)$ such that $X_i \overset{d}{=} F, i = 1, \dots, n$, and

$$P(X_1 + \cdots + X_n = n\mu) = 1.$$

As a consequence, at most $k-1$ random variables X_i can be strictly larger than $n\mu/k$, i.e.

$$P(\mathbf{1}\{X_1 > n\mu/k\} + \cdots + \mathbf{1}\{X_n > n\mu/k\} \leq k-1) = 1,$$

implying that

$$\mathbb{E}\left[\sum_{i=1}^n \mathbf{1}\{X_i > n\mu/k\}\right] = n\overline{F}\left(\frac{n\mu}{k}\right) \leq k-1. \;\square$$

Example 7.2. For $\lambda := (1+\sqrt{2})(1-\epsilon), \epsilon > 0$ we construct a distribution in $\mathcal{N}_n^\lambda(2)$ which is not n-CM for any $n \geq 3$. Assume $n \geq 3$ define the two-point distribution G_2 as

$$G_2 := (1-m)\delta_0 + m\delta_{\sqrt{2}/2},$$

where

$$m := \frac{1}{n-\lambda}\left(\frac{2-\sqrt{2}}{2} + \frac{1+\sqrt{2}}{2}\epsilon\right).$$

Since $n - \lambda > 2 - \sqrt{2}$, it is easy to see that $m \in (0,1)$ when $\epsilon > 0$ is sufficiently small.

Now let the distribution F be defined as

$$F := \left(\frac{\lambda}{n}U + \frac{n-\lambda}{n}G_2\right) \in \mathcal{N}_n^\lambda(2).$$

If $X \overset{d}{=} F$, elementary calculations give

$$P(X \geq \sqrt{2}/2) = \frac{1+\epsilon/2}{n} \quad \text{and} \quad \mathbb{E}[X] = \frac{2-\epsilon/2}{n\sqrt{2}}.$$

Therefore we have that

$$\begin{aligned}\overline{F}\left(\frac{n\mathbb{E}[X]}{2}\right) &= \overline{F}\left(\frac{\sqrt{2}}{2} - \frac{\epsilon}{4\sqrt{2}}\right)\\ &\geq P(X \geq \sqrt{2}/2) = \frac{1+\epsilon/2}{n} > 1/n,\end{aligned}$$

which contradicts the necessary condition given in Lemma 7.1 for $k = 2$.

To conclude, we need at least $\lambda \geq 1 + \sqrt{2}$ in (7.1). It remains open the search for the largest possible $\lambda \in [1+\sqrt{2}, 3]$ for which the condition (7.1) is sufficient to guarantee the n-complete mixability of a distribution F having the continuous function f as its density.

Acknowledgment

Giovanni Puccetti would like to thank the Department of Statistics and Actuarial Science at the University of Waterloo, Canada, for financial support and kind hospitality during the writing of this paper.

Ruodu Wang acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC).

The authors thank Ludger Rüschendorf and the referees for some useful comments on a previous version of the paper.

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

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

Bernard, C., X. Jiang and R. Wang (2013). Risk aggregation with dependence uncertainty. *Preprint*, University of Waterloo.

Embrechts, P., A. J. McNeil, and D. Straumann (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk management: value at risk and beyond*, pp. 176–223. Cambridge: Cambridge Univ. Press.

Embrechts, P., G. Puccetti, and L. Rüschendorf (2013). Model uncertainty and var aggregation. *J. Bank. Financ. 37*(8), 2750 – 2764.

Gourier, E., W. Farkas, and D. Abbate (2009). Operational risk quantification using extreme value theory and copulas: from theory to practice. *J. Operational Risk 4*(3), 3–26.

Levy, H. and Y. Kroll (1978). Ordering uncertain options with borrowing and lending. *J. Finance 33*(2), 553–574.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techiniques, Tools*. Princeton, NJ: Princeton University Press.

Puccetti, G. (2013). Sharp bounds on the expected shortfall for a sum of dependent random variables. *Stat. Probabil. Lett. 83*(4), 1227–1232.

Puccetti, G. and L. Rüschendorf (2012). Bounds for joint portfolios of dependent risks. *Statistics & Risk Modeling 29*, 107–132.

Puccetti, G. and L. Rüschendorf (2013). Sharp bounds for sums of dependent risks. *J. Appl. Probab. 50*(1), 42–53.

Puccetti, G. and L. Rüschendorf (2014). Asymptotic equivalence of conservative VaR- and ES-based capital charges. *Journal of Risk*, to appear.

Puccetti, G., B. Wang, and R. Wang (2012). Advances in complete mixability. *J. Appl. Probab. 49*(2), 430–440.

Wang, B. and R. Wang (2011). The complete mixability and convex minimization problems with monotone marginal densities. *J. Multivariate Anal. 102*(10), 1344–1360.

Wang, R., L. Peng, and J. Yang (2013). Bounds for the sum of dependent risks and worst value-at-risk with monotone marginal densities. *Finance Stoch. 17*(2), 395–417.