CrossMark

Aggregation-robustness and model uncertainty of regulatory risk measures

Paul Embrechts[1] · Bin Wang[2] · Ruodu Wang[3]

Abstract Research related to aggregation, robustness and model uncertainty of regulatory risk measures, for instance, value-at-risk (VaR) and expected shortfall (ES), is of fundamental importance within quantitative risk management. In risk aggregation, marginal risks and their dependence structure are often modelled separately, leading to uncertainty arising at the level of a joint model. In this paper, we introduce a notion of qualitative robustness for risk measures, concerning the sensitivity of a risk measure to the uncertainty of dependence in risk aggregation. It turns out that coherent risk measures, such as ES, are more robust than VaR according to the new notion of robustness. We also give approximations and inequalities for aggregation and diversification of VaR under dependence uncertainty, and derive an asymptotic equivalence for worst-case VaR and ES under general conditions. We obtain that for a portfolio of a large number of risks, VaR generally has a larger uncertainty spread compared to ES. The results warn that unjustified diversification arguments for VaR used in risk management need to be taken with much care, and they potentially support the use of ES in risk aggregation. This in particular reflects on the discussions in the recent consultative documents by the Basel Committee on Banking Supervision.

✉ R. Wang
wang@uwaterloo.ca

P. Embrechts
embrechts@math.ethz.ch

B. Wang
wang_bin@btbu.edu.cn

[1] Department of Mathematics, RiskLab and SFI, ETH Zurich, 8092 Zurich, Switzerland

[2] Department of Mathematics, Beijing Technology and Business University, Beijing 100048, China

[3] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada

1 Introduction

Risk measurement, with its crucial importance for financial institutions such as banks, insurance companies and investment funds, has drawn a lot of attention in both academia and industry over the past several decades. Although a financial risk, often modelled by a probability distribution, cannot be characterized by a single number, sometimes one needs to assign a number to a risk position. The determination of regulatory capital is one such example, the ranking of risks another. For such purposes, quantitative tools that map risks to numbers were introduced, and they are called *risk measures*.

Over the past three decades, value-at-risk (VaR) became the *benchmark* (Jorion [22]). Expected shortfall (ES), an alternative to VaR which is *coherent* (Artzner et al. [3]), is arguably the second most popular risk measure in use. In two recent consultative documents BCBS [4, 5], the Basel Committee on Banking Supervision proposed to take a move from VaR to ES for the measurement of market risk in banking. Under Solvency 2 and the Swiss Solvency Test, the same discussion takes place within insurance regulation; see, for instance, Sandström [34] and SCOR [35]. As a consequence, there have been extensive debates on issues related to diversification, aggregation, economic interpretation, optimization, extreme behaviour, robustness and backtesting of VaR and ES. We omit a detailed analysis here and refer to Embrechts et al. [15], Emmer et al. [16] and the references therein.

Here are some of the issues raised. VaR is not coherent, but it is *elicitable* (see Gneiting [18, Theorem 9]; that paper also contains some earlier references), easy to backtest and more robust with respect to statistical uncertainty, as argued in Gneiting [18] and Cont et al. [10]. ES is coherent, but not elicitable and difficult to backtest. There have been extensive discussions on the problematic diversification and aggregation issues of VaR due to its lack of subadditivity; see, for example, Embrechts et al. [14]. Daníelsson et al. [11] argue that the violation of subadditivity for VaR is rare in practice. VaR, being a quantile, does not address the crucial "what if" question, i.e. what are the consequences if a particular rare event (measured by VaR) occurs. Whereas this was clear since its introduction within the financial industry around 1994, it took some serious financial crises to bring this issue fully onto the regulatory agenda.

The importance of robustness properties of risk measures has only fairly recently become a focal point of regulatory attention. By now, numerous academic as well as applied papers address the topic. Conflicting views typically result from different notions of robustness; Embrechts et al. [15] contains a brief discussion and some references. In the present paper, the measurement of aggregated risk positions under

uncertainty with respect to the dependence structure of the underlying risk factors will be discussed. We show that ES enjoys a new property of aggregation-robustness which VaR generally does not have.

The mathematical property of (non-)subadditivity of a risk measure becomes relevant upon analysing the aggregate position of a portfolio. As is often the case in practice, the dependence structure among individual risks in a portfolio is difficult to obtain from a statistical point of view, while the marginal distributions of the individual risks (assets) may typically be easier to model; see, for instance, Embrechts et al. [14] and Bernard et al. [7]. Modelling a high-dimensional dependence structure is well known to be data-costly, and dimension reduction techniques such as vine copulas, hierarchical structures and very specific parametric models often have to be implemented. Whereas such simplifying techniques in general create computational and modelling ease, they typically involve considerable model uncertainty. This leads to a notion of *dependence uncertainty* (DU) in risk aggregation, a concept of main interest for this paper.

From a mathematical or statistical point of view, it is clearly better to look at robustness properties of a model at the level of the joint distribution of the risk factors. The main reason for separating the two (marginals, dependence) is because of processes in practice, where indeed the two are often modelled separately. This is particularly true in a stress testing environment.

Hence for this paper, we introduce the notion of *aggregation-robustness* to study properties of risk measures for aggregation in the presence of dependence uncertainty. The new notion is based on the classic notion of robustness for statistical functionals in, e.g. Huber and Ronchetti [21]. However, as opposed to the conclusions in Cont et al. [10], we show that when model uncertainty lies solely at the level of the dependence structure, coherent distortion risk measures (such as ES) are continuous with respect to weak convergence of the underlying distributions, whereas VaR in general is not. This result supports the use of ES for risk aggregation, especially when statistical information on marginal distributions is reliable.

Under DU, the attainable values of a risk measure lie in an interval. This interval can be seen as a measurement of model uncertainty for that risk measure. When a risk measure is used to quantify the riskiness of an aggregate position of a portfolio, the ratio between the risk measure of the aggregate risk and the sum of the risk measures of the marginal risks is called a *diversification ratio*. The diversification ratio measures how good the risks in a portfolio hedge (compensate for) each other. With only models for marginal distributions available, the diversification ratio also takes values in a DU-interval.

To study the DU-interval of VaR and ES and their diversification ratios, one needs to calculate the worst-case and best-case values of VaR and ES under dependence uncertainty. Due to the subadditivity of ES, the worst-case value of ES is the sum of the ES of the marginal risks. However, the other three quantities (best- and worst-case VaR, best-case ES) are in general unknown. Partial results exist. The worst-case value of VaR for $n = 2$ was given in Makarov [27] based on early results in multivariate probability theory. Embrechts and Puccetti [13] gave a dual bound for the worst-case VaR for $n \geqslant 3$ in the homogeneous model, i.e. when all marginal risks have the same distribution. Partial solutions for the worst-case and best-case values of VaR are to

be found in Wang et al. [40], Puccetti and Rüschendorf [30] and Bernard et al. [7], based on the notion of *complete mixability* (CM) introduced in Wang and Wang [37]. A fast algorithm to numerically calculate the worst-case and best-case values of VaR under general conditions was introduced in Embrechts et al. [14]; this is the so-called *rearrangement algorithm* (RA). For the best-case ES, some partial analytical results can be found in Bernard et al. [7] and Cheung and Lo [9], and a numerical procedure was proposed by Puccetti [29].

In most of the existing analytical results, it is assumed that the marginal distributions are identical (homogeneous case), with some extra conditions on the shape of the underlying risk factor densities (assumed to exist). In this paper, we relax the assumptions on the marginal distributions. Instead of explicit values for the worst-case and best-case VaR, we obtain approximations. The new results obtained can be used within a discussion on capital requirement; they, moreover, yield a DU-interval for VaR and its diversification ratio.

Further understanding of the worst-case VaR can be obtained through the asymptotic behaviour as the number of risks in the portfolio grows to infinity, i.e. for a large portfolio regime. In the homogeneous case, Puccetti and Rüschendorf [31] obtained an asymptotic equivalence between the worst-case VaR and the worst-case ES under dependence uncertainty, and this under a strong condition on the identical marginal distributions. The required condition was later weakened by Puccetti et al. [32] (based on further results on complete mixability) and Wang [39] (based on a duality result obtained in Rüschendorf [33]). It was finally removed by Wang and Wang [38] (based on the notion of extreme negative dependence). When the marginal distributions are not identical, Puccetti et al. [32] also obtained the asymptotic equivalence under the assumption that only finitely many different choices of the marginal distributions can appear; this mathematically allows a reduction to the case of identical marginal distributions. In this paper, we give a unifying result on this asymptotic equivalence, by allowing the marginal distributions to be arbitrary. Only weak uniformity conditions on the moments of the marginal distributions are required for our results to hold. These conditions are easily justified in practice and are necessary for the most general equivalence to hold. The new results lead to the asymptotic DU-spread of VaR and ES, and show that VaR in general yields a larger DU-spread compared to ES.

The rest of the paper is organized as follows. In Sect. 2, we introduce the notion of aggregation-robustness and show that ES is aggregation-robust but VaR is not. In Sect. 3, we give new bounds on the diversification ratios under dependence uncertainty, and establish an asymptotic equivalence between VaR and ES under a worst-case scenario. The dependence uncertainty spreads of VaR and of ES are derived and compared in Sect. 4. In Sect. 5, numerical examples are presented to illustrate our results. Section 6 draws some conclusions. All proofs are put in the Appendix.

Throughout the paper, we let $(\Omega, \mathcal{A}, \mathbb{P})$ be an atomless probability space and $L^0 := L^0(\Omega, \mathcal{A}, \mathbb{P})$ the set of all (equivalence classes of) real-valued random variables on that probability space. Elements of L^0 are often referred to as risks. Their distribution functions are simply referred to as distributions. We write $X \sim F$ to denote $F(x) = \mathbb{P}[X \leqslant x]$, $x \in \mathbb{R}$. We also denote the generalized inverse function of F by F^{-1}, that is, $F^{-1}(p) = \inf\{t \in \mathbb{R} : F(t) \geqslant p\}$ for $p \in (0, 1]$, and $F^{-1}(0) = \inf\{t \in \mathbb{R} : F(t) > 0\}$.

2 Robustness of VaR and ES for risk aggregation

2.1 Robustness of risk measures

The robustness of a statistical functional or an estimation procedure describes the sensitivity to underlying model deviations and/or data changes. Different definitions and interpretations of robustness exist in the literature; see, for example, Huber and Ronchetti [21] from a purely statistical perspective, Hansen and Sargent [20] in the context of economic decision making, and Ben-Tal et al. [6] within optimization. In statistics, robustness mainly concerns the so-called *distributional* (or *Hampel–Huber*) *robustness*: the statistical consequences when the shape of the actual underlying distribution deviates slightly from the assumed model.

A risk measure ρ is a function which maps a risk in a set $\mathcal{X}$ to a number, $\rho : \mathcal{X} \to (-\infty, +\infty]$, where $\mathcal{X} \subset L^0$ typically contains L^∞ and is closed under addition and positive scalar multiplication. A risk measure is *law-invariant* if it only depends on the distribution of the risk. We omit the general introduction of risk measures, and refer the interested reader to Föllmer and Schied [17, Chap. 4]. Since law-invariant risk measures are a specific type of statistical functionals, their robustness properties are already extensively studied in the statistical literature; see, e.g. Huber and Ronchetti [21, Chap. 3].

In this paper, we focus on the two most popular risk measures: value-at-risk (VaR) at confidence level p, defined as

$$\mathrm{VaR}_p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}[X \leqslant x] \geqslant p\}, \quad p \in (0, 1), X \in L^0,$$

and the expected shortfall (ES) at confidence level p, defined as

$$\mathrm{ES}_p(X) = \frac{1}{1-p} \int_p^1 \mathrm{VaR}_q(X)\mathrm{d}q, \quad p \in (0, 1), X \in L^0. \tag{2.1}$$

Clearly, $\mathrm{VaR}_p(X) = F^{-1}(p)$ for $p \in (0, 1)$, where $X \sim F$. Though typically it is assumed in (2.1) that $\mathbb{E}[|X|] < \infty$, we may occasionally allow that $\mathrm{ES}_p(X) = \infty$ for some X. On the other hand, $\mathrm{VaR}_p(X)$ is always a finite number for all $X \in L^0$. Both risk measures occur most frequently in the setting of solvency requirements for financial institutions; hence the appearance of "regulatory risk measures" in the title of the paper.

It is often argued in the literature that quantile-based risk measures such as VaR are more robust when compared to mean-based risk measures such as ES; the notion of robustness used most often is Hampel's (Hampel [19]). ES is only robust with respect to stronger metrics (e.g. the Wasserstein distance); arguments of this type can be found for instance in Cont et al. [10], Kou and Peng [23] and Emmer et al. [16]. More general results on continuity of law-invariant risk measures with respect to certain metrics on sets of probability measures are provided in Krätschmer et al. [25]. It is well known that the qualitative robustness of a statistical estimator, as in Hampel [19], is equivalent to the continuity of the corresponding risk measure at the true distribution, with respect to the weak topology. In general, to analyse statistical robustness of a given risk measure, one typically studies continuity properties

of that risk measure in a given metric d, say. It would hence be proper in such a case to talk about d-robustness. Hence we say that a law-invariant risk measure ρ is d-robust at a distribution F if $d(F_n, F) \to 0$ implies that $\rho(X_n) \to \rho(X)$, where $X_n \sim F_n, n = 1, 2, \ldots,$ and $X \sim F$. For example, the *Lévy distance* is used in Cont et al. [10] to measure the difference between any two univariate distributions F and G by

$$d(F, G) := \inf\{\varepsilon > 0 : F(x - \varepsilon) - \varepsilon < G(x) < F(x + \varepsilon) + \varepsilon, \ \forall x \in \mathbb{R}\}.$$

Note that the Lévy distance metrizes the weak topology on the set of distributions. A law-invariant risk measure is said to be robust in Hampel's sense if it is continuous with respect to convergence in distribution. Other metrics can also be used for the analysis of robustness; see Krätschmer et al. [24, 25] and Cambou and Filipović [8]. It is a very classical result that the pth (lower) quantile functional $F \mapsto F^{-1}(p)$ (and so VaR_p) is weakly continuous at each F_0 for which the mapping $s \mapsto F_0^{-1}(s)$ is continuous at $s = p$. A more general result can be found, for instance, in van der Vaart [36, Lemma 21.2]. In Krätschmer et al. [25], it is argued that Hampel's notion of (statistical) robustness is less relevant for risk management. Using a different definition, they introduce a continuous scale of robustness.

In the following, we introduce a new, in our opinion practically relevant notion of robustness for risk aggregation, which turns out to favour ES over VaR.

2.2 Aggregation-robustness

In this section, we show that VaR is more sensitive to model uncertainty at the level of dependence than ES. For single risks $X_i, i = 1, \ldots, n$, the aggregate risk S is simply defined as $S = X_1 + \cdots + X_n$. Often in practice, a joint model of $X_1, \ldots, X_n$ is chosen in two stages: n marginal distributions $F_1, \ldots, F_n$ and a dependence structure (often through a copula C). Whereas the modelling of marginal distributions is fairly standard, the dependence structure can be really difficult to model, statistically estimate and test. Considerable model uncertainty, which is often different in nature from the model uncertainty of marginal distributions, arises from modelling the dependence structure. In the following, we study sensitivity with respect to uncertainty in the dependence structure; for the purpose of this paper, **we assume the marginal distributions $F_1, \ldots, F_n$ are given**.

When the dependence structure between the risks is unknown, the possible distributions of S form a set. We denote the $(F_1, \ldots, F_n)$-*admissible class* as

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

or simply $\mathfrak{S}_n = \mathfrak{S}_n(F_1, \ldots, F_n)$ if $(F_1, \ldots, F_n)$ is clear from the context. $\mathfrak{S}_n$ is the set of all possible aggregate risks. Risk aggregation with dependence uncertainty concerns the probabilistic and statistical behaviour of $S \in \mathfrak{S}_n$; in particular, $\mathfrak{S}_n$ is closed with respect to the weak topology (see Bernard et al. [7]). We say that an admissible class $\mathfrak{S}_n$ is *compatible with* a risk measure $\rho : \mathcal{X} \to (-\infty, +\infty]$ if $X_i \in \mathcal{X}, X_i \sim F_i$ (note that this implies $\mathfrak{S}_n \subset \mathcal{X}$ since $\mathcal{X}$ is closed under addition) and $\rho(X_i) < \infty$, for $i = 1, \ldots, n$.

Definition 2.1 (Aggregation-robustness) We say that a law-invariant risk measure $\rho : \mathcal{X} \to (-\infty, +\infty]$ is *aggregation-robust* if ρ is continuous with respect to weak convergence in each admissible class $\mathfrak{S}_n$ compatible with ρ.

The robustness character of Definition 2.1 in intuitively clear. If the joint distributions of $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ are close according to the Lévy metric, then the distributions of $X_1 + \cdots + X_n$ and $Y_1 + \cdots + Y_n$ are also close according to the Lévy metric. As a consequence, ρ is insensitive to small perturbations of the joint distribution of the underlying risk factors, keeping the marginal distributions of the individual risks fixed. It is clear that Hampel's robustness, as discussed above, without the restriction of risks being in a common admissible class, implies aggregation-robustness. When the dependence structure is modelled by copulas, our definition of robustness implies that a risk measure is insensitive to the copula of the individual risks when the marginal distributions are assumed to be known. The fact that in Definition 2.1 we look at risks in $\mathfrak{S}_n$ reflects our interest in aggregation and diversification. One could, of course, look at other functional-robustness definitions beyond aggregation (summation).

Example 2.2 (VaR is not aggregation-robust) For $t \in [0, 1]$, let X_t and Y_t have the joint distribution C_t given by

$$C_t(x, y) = txy + (1 - t)\big(\max\big\{\min\{x, 1/2\} + \min\{y, 1/2\} - 1/2, 0\big\} \\ + \max\{x + y - 3/2, 0\}\big)$$

for $x, y \in [0, 1]$. It is easy to see that X_t and Y_t are both $\mathrm{U}[0, 1]$-distributed, hence C_t is a copula, for $t \in [0, 1]$. Note that C_t, $t \in (0, 1)$, is a mixture of the independence copula C_1 and another copula

$$C_0 : [0, 1]^2 \to [0, 1],$$
$$(x, y) \mapsto \max\big\{\min\{x, 1/2\} + \min\{y, 1/2\} - 1/2, 0\big\} + \max\{x + y - 3/2, 0\}.$$

C_0 is the *ordinal sum* of two Fréchet lower copulas; see Nelsen [28, Sect. 3.2.2].

It is immediate that the distribution of $X_t + Y_t$ for $t \in (0, 1]$ is symmetric, centred at 1, with positive density on the interval $(1/2, 3/2)$. Thus, $\mathrm{VaR}_{1/2}(X_t + Y_t) = 1$. It is also straightforward that $X_0 + Y_0$ is a discrete random variable on $\{1/2, 3/2\}$ with $\mathrm{VaR}_{1/2}(X_0 + Y_0) = 1/2$. As a consequence,

$$\mathrm{VaR}_{1/2}(X_0 + Y_0) \neq \lim_{t \to 0} \mathrm{VaR}_{1/2}(X_t + Y_t).$$

Based on the simple fact that $X_t + Y_t \to X_0 + Y_0$ weakly as t goes to zero, we conclude that $\mathrm{VaR}_{1/2}$ is not aggregation-robust.

To build an example for VaR_p, $p \in (1/2, 1)$, let A be an event of probability $2 - 2p$, independent of X_t and Y_t, and let $Z_t = \mathrm{I}_A X_t$, $W_t = \mathrm{I}_A Y_t$ for each $t \in [0, 1]$. By construction it is clear that Z_t, W_t, $t \in [0, 1]$, are all identically distributed, and

$$\mathrm{VaR}_p(Z_t + W_t) = \mathrm{VaR}_{1/2}(X_t + Y_t), \quad t \in [0, 1].$$

Analogously to the above argument, we have $d(Z_t + W_t, Z_0 + W_0) \to 0$ as t goes to zero, but $\mathrm{VaR}_p(Z_0 + W_0) \neq \lim_{t \to 0} \mathrm{VaR}_p(Z_t + W_t)$. Putting a negative sign in front of Z_t and W_t, we obtain that VaR_p, $p \in (0, 1/2)$, is also discontinuous in an admissible class. This shows that VaR_p is not aggregation-robust for any $p \in (0, 1)$.

The non-aggregation-robustness of VaR_p essentially comes from the fact that it is not continuous with respect to weak convergence (Hampel's robustness). Suppose that VaR_p is not continuous at some distribution, say F_0. One may find $F_n, n \in \mathbb{N}$, which converge to F_0 weakly, but $F_n^{-1}(p), n \in \mathbb{N}$, do not converge to $F_0^{-1}(p)$; if in addition, such $F_n, n \in \mathbb{N}$, and F_0 lie in the same admissible class, then VaR_p is not aggregation-robust. That leads to the construction in Example 2.2.

In the above example, the joint distribution C_t with a small $t > 0$ can be seen as the joint distribution C_0 influenced by a small perturbation. It is moreover worth noting that in Example 2.2, the marginal distributions of X_t and Y_t are continuous with positive densities. Hence, even if the true marginal distributions are known to have positive densities, VaR can still be discontinuous in aggregation. When one considers absolutely continuous models for a single risk, one has Hampel's robustness for VaR_p; however, considering absolutely continuous marginal models is not sufficient for the aggregation-robustness of VaR_p. On the other hand, we shall see that ES *is* aggregation-robust, although it is well known to be non-robust in Hampel's sense (see Cont et al. [10]) since it is discontinuous at any distribution with respect to the weak topology.

For generality, we study the aggregation-robustness of *distortion risk measures*, defined as

$$\rho(X) = \int_{\mathbb{R}} x \, \mathrm{d}h\big(F(x)\big), \quad X \in \mathcal{X}, X \sim F, \tag{2.2}$$

where $\mathcal{X}$ is a set of random variables such that the integral in (2.2) is properly defined, and $h : [0, 1] \to [0, 1]$ is a nondecreasing function with $h(0) = 0$, $h(1) = 1$; h is called the *distortion function* of ρ. If h has left limits and is right-continuous, i.e. h is a distribution function on $[0, 1]$, then

$$\rho(X) = \int_0^1 F^{-1}(t) \, \mathrm{d}h(t), \quad X \in \mathcal{X}, X \sim F. \tag{2.3}$$

See Wang et al. [41] for distortion risk measures in the context of insurance premium calculations, Kusuoka [26] for their connection with coherent risk measures, and Cont et al. [10] for their robustness properties. A distortion risk measure ρ is coherent if and only if h is convex, in which case ρ is called a *spectral risk measure*; see Acerbi [2]. Distortion risk measures are also closely related to L-statistics; see Huber and Ronchetti [21, Sect. 3.3]. For $p \in (0, 1)$, VaR_p and ES_p are special cases of distortion risk measures, with distortion functions $h(t) = \mathrm{I}_{\{t \geqslant p\}}, t \in [0, 1]$, and $h(t) = \mathrm{I}_{\{t \geqslant p\}}(t - p)/(1 - p), t \in [0, 1]$, respectively.

Note that $\mathcal{X}$ has to be closed under addition, hence it may fail to contain all X such that the integral in (2.2) is properly defined. For coherent distortion risk measures, one may consider the set $\mathcal{X}_0$ defined by

$$\mathcal{X}_0 = \{X \in L^0 : \mathbb{E}[X^-] < \infty\} \supset L^1.$$

It is easy to check by the convexity of h that all coherent distortion risk measures are properly defined on $\mathcal{X}_0$. Our main result on aggregation-robustness now becomes:

Theorem 2.3 *All coherent distortion risk measures on $\mathcal{X}_0$ with a continuous distortion function are aggregation-robust.*

As a coherent distortion risk measure has a convex distortion function, assuming continuity only excludes a jump of the distortion function at 1. Theorem 2.3 says that when the model uncertainty lies at the level of dependence but not at the level of the marginal distributions, coherent distortion risk measures such as ES are continuous with respect to weak convergence.

Our result can be interpreted as follows. For a distribution F and a random variable $X \sim F$, even adding constraints on marginal distributions of the aggregation model of X, $F \mapsto \mathrm{VaR}_p(X)$ is still not continuous (with respect to weak convergence), whereas $F \mapsto \mathrm{ES}_p(X)$ is continuous with these constraints. It should not be interpreted as an argument against the classic continuity results of VaR, noting that VaR is continuous at most commonly used distributions in financial risk management.

Remark 2.4 Cont et al. [10] also introduced the notion of $\mathcal{C}$-robustness, where $\mathcal{C}$ is a set of distributions. A risk measure ρ is $\mathcal{C}$-robust if ρ is continuous in $\mathcal{C}$ with respect to the Lévy distance; see Cont et al. [10, Proposition 2]. Using this notion, VaR_p is $\mathcal{C}_p$-robust, where $\mathcal{C}_p$ is the set of distributions F for which F^{-1} is continuous at p. If we denote by $\mathfrak{D}(\mathfrak{S}_n)$ the set of all possible distributions for an admissible class $\mathfrak{S}_n$, then ρ is aggregation-robust if and only if ρ is $\mathfrak{D}(\mathfrak{S}_n)$-robust for all possible choices of $n \in \mathbb{N}$ and $\mathfrak{D}(\mathfrak{S}_n)$, in which $\mathfrak{S}_n$ is compatible with ρ.

In the case $\mathcal{X} = L^\infty$, we obtain that a continuous distortion function is a necessary and sufficient condition for the aggregation-robustness of distortion risk measures.

Theorem 2.5 *A distortion risk measure on $\mathcal{X} = L^\infty$ is aggregation-robust if and only if its distortion function h is continuous on $[0, 1]$.*

Finally, we remark that it would be of much interest to characterize aggregation-robust statistical functionals (risk measures) other than the class of distortion risk measures. Such a characterization is beyond the scope of this paper, and we leave it for future work.

3 Bounds on VaR aggregation

In Sect. 2, we mainly looked at the sensitivity properties of risk measures on aggregated risks under small changes of the underlying dependence assumptions. In this section, for VaR, we concentrate on deviations (possibly) far away from some true underlying, though unknown, dependence structure. Such results can be used to analyse extreme scenarios for risk aggregation and may be helpful in order to determine conservative capital requirements under model (i.e. dependence) uncertainty; for a real life example on this, see Aas and Puccetti [1].

3.1 Aggregation and diversification under dependence uncertainty

We start with the motivating notion of *diversification ratio*, which is closely related to the aggregation of VaR. Given a portfolio consisting of individual risks $X_1, \ldots, X_n$, the diversification ratio of VaR at confidence level $p \in (0, 1)$ is defined as

$$\Delta_n^p = \frac{\mathrm{VaR}_p(X_1 + \cdots + X_n)}{\sum_{i=1}^n \mathrm{VaR}_p(X_i)}.$$

The diversification ratio measures a kind of *diversification benefit*, and is, for instance, widely used in operational risk (see the examples in Embrechts et al. [14]). In the latter context, X_i corresponds to next year's operational risk loss in business line i, $i = 1, \ldots, n$ ($n = 8$, typically); often explicit models for the loss-dependence among business lines are not available. For capital charge purposes, one estimates the total capital requirement for the superposition of the risks in each business line. One then typically adds up the risk measures across all business lines, and multiplies by a factor which is an estimate of Δ_n^p. For this purpose, one needs a joint model of the risks $X_1, \ldots, X_n$.

With a known joint distribution of $(X_1, \ldots, X_n)$, Δ_n^p may be calculated theoretically. If $\Delta_n^p \leqslant 1$, we say there is a *diversification benefit* in the portfolio; if $\Delta_n^p \geqslant 1$, we say there is a *diversification penalty*. When $F_1, \ldots, F_n$ are known and the joint model of $(X_1, \ldots, X_n)$ is unspecified, the *worst diversification ratio* is defined as

$$\overline{\Delta}_n^p = \frac{\sup\{\mathrm{VaR}_p(X_1 + \cdots + X_n) : X_i \sim F_i, i = 1, \ldots, n\}}{\sum_{i=1}^n \mathrm{VaR}_p(X_i)}$$

$$= \frac{\sup\{\mathrm{VaR}_p(S) : S \in \mathfrak{S}_n\}}{\sum_{i=1}^n \mathrm{VaR}_p(X_i)}.$$

By definition, $\overline{\Delta}_n^p \geqslant 1$ if $\sum_{i=1}^n \mathrm{VaR}_p(X_i) > 0$. In the following, we denote the comonotonic VaR by $\mathrm{VaR}_p^+(S_n)$, i.e.

$$\mathrm{VaR}_p^+(S_n) = \sum_{i=1}^n \mathrm{VaR}_p(X_i).$$

Note here that S_n is symbolic and does not represent a particular random variable. The calculation of $\overline{\Delta}_n^p$, as a measure of the worst-case diversification effect of VaR, serves two purposes:

– (Conservative capital requirement) $\overline{\Delta}_n^p \mathrm{VaR}_p^+(S_n)$ can be used as the most conservative capital requirement in the case of given (or estimated) marginal distributions $F_1, \ldots, F_n$ of the individual risks.
– (Measurement of model uncertainty) If $\overline{\Delta}_n^p$ is small, then the model uncertainty is small, and the risk measure VaR is considered as less problematic in risk aggregation; capital requirement principles based on VaR_p^+ become more plausible. If $\overline{\Delta}_n^p$ is large, then the model uncertainty is severe, and arguments of diversification benefit need to be taken with care.

The *best diversification ratio*, replacing the sup by an inf, can be studied similarly. Since we are more interested in the worst case (corresponding to a conservative capital requirement), we omit a discussion of the best diversification ratio.

In the recent literature (see, for instance, Embrechts et al. [15]), it was shown that the value of $\overline{\Delta}_n^p$ is closely related to the risk measure ES. Denote the worst-case ES by $\overline{\mathrm{ES}}_p(S_n) = \sup\{\mathrm{ES}_p(S) : S \in \mathfrak{S}_n\}$; since ES is subadditive and comonotonic additive, we have that

$$\overline{\mathrm{ES}}_p(S_n) = \sum_{i=1}^{n} \mathrm{ES}_p(X_i) = \mathrm{ES}_p^+(S_n),$$

where the latter $+$-notation is in line with the notation used for the comonotonic VaR case. Since VaR is bounded by ES, the worst-case VaR is bounded by the worst-case ES. If $\mathrm{VaR}_p^+(S_n) > 0$, we have for $\overline{\Delta}_n^p$ the direct upper bound

$$1 \leqslant \overline{\Delta}_n^p \leqslant \frac{\mathrm{ES}_p^+(S_n)}{\mathrm{VaR}_p^+(S_n)} = \frac{\overline{\mathrm{ES}}_p(S_n)}{\mathrm{VaR}_p^+(S_n)}. \tag{3.1}$$

See also Embrechts et al. [15] for a discussion on this upper bound. Later in this section, we show that the second inequality in (3.1) is asymptotically sharp as $n \to \infty$.

By definition, calculation of the worst diversification ratio is equivalent to the calculation of the worst-case VaR

$$\overline{\mathrm{VaR}}_p(S_n) := \sup\{\mathrm{VaR}_p(S) : S \in \mathfrak{S}_n\}. \tag{3.2}$$

For the history and a general discussion on problems related to (3.2) from the perspective of quantitative risk management, we refer to Embrechts et al. [15]. When $F_1 = F_2 = \cdots = F_n =: F$, i.e. in the homogeneous case, Wang et al. [40] obtained $\overline{\mathrm{VaR}}_p(S_n)$ for F with a tail-decreasing density. If $F_1, \ldots, F_n$ are not identical, explicit calculations of $\overline{\mathrm{VaR}}_p(S_n)$ and $\overline{\Delta}_n^p$ are not available in general. Embrechts et al. [14] introduced the rearrangement algorithm to numerically calculate $\overline{\mathrm{VaR}}_p(S_n)$ based on a discretized approximation.

Regarding the asymptotic behaviour of $\overline{\mathrm{VaR}}_p(S_n)$ and $\overline{\Delta}_n^p$, Puccetti and Rüschendorf [31] obtained that, as $n \to \infty$,

$$\frac{\overline{\mathrm{VaR}}_p(S_n)}{\overline{\mathrm{ES}}_p(S_n)} \to 1, \tag{3.3}$$

in the homogeneous case under a restrictive condition on the marginal distributions. See also Wang [39] and Wang and Wang [38] for weaker conditions so that (3.3) holds. Puccetti et al. [32] considered the case when there are finitely many different marginal distributions in the sequence $F_1, F_2, \ldots$ and obtained the same equivalence (3.3). A consequence of (3.3) is that

$$\lim_{n \to \infty} \overline{\Delta}_n^p = \lim_{n \to \infty} \frac{\overline{\mathrm{ES}}_p(S_n)}{\mathrm{VaR}_p^+(S_n)}, \tag{3.4}$$

provided that the right-hand limit exists. That is, the second inequality in (3.1) is asymptotically sharp. However, as mentioned above, the existing results only deal with the (almost) homogeneous case, and some specific assumptions on the marginal distributions need to be imposed. Later in this section, we provide analytical approximations for $\overline{\mathrm{VaR}}_p(S_n)$ and $\overline{\Delta}_n^p$. Based on these results, we give a proof of (3.3) and (3.4) under very general conditions and, moreover, obtain a rate of convergence.

3.2 Bounds on VaR aggregation for a finite number of risks

In this section, we give inequalities for the worst-case and best-case VaR and its diversification ratio. For a distribution F_i, define

$$\mu_{p,q}^{(i)} = \frac{1}{q-p} \int_p^q F_i^{-1}(t) \mathrm{d}t,$$

for $1 \geqslant q > p \geqslant 0$, $i = 1, \ldots, n$. Note that $\mu_{0,q}^{(i)}$ and $\mu_{p,1}^{(i)}$ might be infinite. Using the above notation, it is immediate that

$$\overline{\mathrm{ES}}_p(S_n) = \sum_{i=1}^n \mathrm{ES}_p(X_i) = \sum_{i=1}^n \mu_{p,1}^{(i)}.$$

For future discussion, we also denote the best-case VaR by $\underline{\mathrm{VaR}}_p(S_n)$, that is,

$$\underline{\mathrm{VaR}}_p(S_n) = \inf_{S \in \mathbb{S}_n} \mathrm{VaR}_p(S),$$

and the best-case ES by $\underline{\mathrm{ES}}_p(S_n)$, that is,

$$\underline{\mathrm{ES}}_p(S_n) = \inf_{S \in \mathbb{S}_n} \mathrm{ES}_p(S).$$

Analytical formulas for each of $\overline{\mathrm{VaR}}_p(S_n)$, $\underline{\mathrm{VaR}}_p(S_n)$ and $\underline{\mathrm{ES}}_p(S_n)$ are not available under general assumptions on the marginal distributions; see Bernard et al. [7] and Embrechts et al. [15] for existing results.

The following theorem contains our main result regarding approximations of $\overline{\mathrm{VaR}}_p(S_n)$ and $\underline{\mathrm{VaR}}_p(S_n)$.

Theorem 3.1 *For any distributions $F_1, \ldots, F_n$, we have for $p \in (0,1)$ that*

$$\sup_{q \in (p,1]} \left\{ \sum_{i=1}^n \mu_{p,q}^{(i)} - \max_{i=1,\ldots,n} \left(F_i^{-1}(q) - F_i^{-1}(p) \right) \right\} \leqslant \overline{\mathrm{VaR}}_p(S_n) \leqslant \overline{\mathrm{ES}}_p(S_n), \quad (3.5)$$

and

$$\sum_{i=1}^n \mu_{0,p}^{(i)} \leqslant \underline{\mathrm{VaR}}_p(S_n) \leqslant \inf_{q \in [0,p)} \left\{ \sum_{i=1}^n \mu_{q,p}^{(i)} + \max_{i=1,\ldots,n} \left(F_i^{-1}(q) - F_i^{-1}(p) \right) \right\}. \quad (3.6)$$

In particular, if $F_1, \ldots, F_n$ are supported on $[a,b]$, $a < b$, $a, b \in \mathbb{R}$, then

$$\overline{\mathrm{ES}}_p(S_n) - (b-a) \leqslant \overline{\mathrm{VaR}}_p(S_n) \leqslant \overline{\mathrm{ES}}_p(S_n).$$

Note that in the case when all marginal distributions are bounded, $\overline{\mathrm{VaR}}_p(S_n)$ and $\overline{\mathrm{ES}}_p(S_n)$ differ by at most a constant which does not depend on n. Theorem 3.1 can also be formulated for the worst diversification ratio of VaR.

Corollary 3.2 *For any distributions $F_1, \ldots, F_n$, suppose that $\mathrm{VaR}_p^+(S_n) > 0$. We have for $p \in (0, 1)$ that*

$$\sup_{q \in (p, 1]} \frac{\sum_{i=1}^n \mu_{p,q}^{(i)} - \max_{i=1,\ldots,n}(F_i^{-1}(q) - F_i^{-1}(p))}{\mathrm{VaR}_p^+(S_n)} \leqslant \overline{\Delta}_n^p \leqslant \frac{\overline{\mathrm{ES}}_p(S_n)}{\mathrm{VaR}_p^+(S_n)}. \quad (3.7)$$

In particular, if $F_1, \ldots, F_n$ are supported in $[a, b]$, $a < b$, $a, b \in \mathbb{R}$, then

$$\frac{\overline{\mathrm{ES}}_p(S_n)}{\mathrm{VaR}_p^+(S_n)} - \frac{b - a}{\mathrm{VaR}_p^+(S_n)} \leqslant \overline{\Delta}_n^p \leqslant \frac{\overline{\mathrm{ES}}_p(S_n)}{\mathrm{VaR}_p^+(S_n)}. \quad (3.8)$$

In the homogeneous case $F := F_1 = F_2 = \cdots$, the left- and right-hand sides of (3.8) both converge to $\frac{\mathrm{ES}_p(X)}{\mathrm{VaR}_p(X)}$ as $n \to \infty$, where $X \sim F$, assuming $\mathrm{VaR}_p(X) \neq 0$. In the following, we study the limit, as n goes to infinity, of the worst- and best-case VaR and its diversification ratio under general marginal assumptions.

3.3 Asymptotic equivalence and limit of the worst diversification ratio

Based on Theorem 3.1, we now derive the asymptotic equivalence between the worst-case VaR and the worst-case ES under very weak general conditions. For an asymptotic analysis, some uniformity conditions on F_i, $i \in \mathbb{N}$, need to be imposed. In what follows, X_i is any random variable with distribution F_i, $i \in \mathbb{N}$. Define the following conditions, for some $p \in (0, 1)$ and $k > 1$:

$$\mathbb{E}\left[|X_i - \mathbb{E}[X_i]|^k\right] < M \quad \text{for some } M > 0, \quad (3.9)$$

$$\liminf_{n \to \infty} n^{-1/k} \sum_{i=1}^n \mathrm{ES}_p(X_i) = +\infty, \quad (3.10)$$

$$C_0 := \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \mathrm{ES}_p(X_i) > 0. \quad (3.11)$$

The above conditions only concern the moments of F_i, $i \in \mathbb{N}$, and they are quite weak and commonly satisfied. Condition (3.9) is a uniform boundedness condition, ensuring that the aggregate portfolio S_n does not contain a single risk with a too heavy tail that dominates the other risks. Condition (3.10) is imposed to guarantee that the average ES of the sequence of risks does not vanish to zero too fast. The condition (3.11) is a stronger version of (3.10). In particular, in the homogeneous case when F_i, $i \in \mathbb{N}$, are identical, $\mathrm{ES}_p(X_1) > 0$ implies (3.11) and hence it also implies (3.10). We also remark that condition (3.12) below is stronger than condition (3.9):

$$\mathbb{E}[|X_i|^k] \text{ is uniformly bounded in } i. \quad (3.12)$$

Theorem 3.3 *Suppose that the distributions $F_i, i \in \mathbb{N}$, satisfy (3.9) and (3.10) for some $p \in (0, 1)$ and $k > 1$. Then*

$$\lim_{n\to\infty} \frac{\overline{\mathrm{VaR}}_p(S_n)}{\overline{\mathrm{ES}}_p(S_n)} = 1. \tag{3.13}$$

If in addition (3.10) is replaced by (3.11), then for sufficiently large n,

$$1 \geqslant \frac{\overline{\mathrm{VaR}}_p(S_n)}{\overline{\mathrm{ES}}_p(S_n)} \geqslant 1 - Cn^{-1+1/k}, \tag{3.14}$$

where

$$C = \left(\frac{1}{1-p} \frac{k}{k-1} + 1 \right) \frac{M^{1/k}}{C_0} + \varepsilon > 0,$$

M is given in (3.9), C_0 is given in (3.11), and ε is any fixed positive real number.

Theorem 3.3 establishes the asymptotic equivalence of the worst-case ES and the worst-case VaR for risk aggregation for general, possibly inhomogeneous portfolios. As mentioned in Sect. 3.1, homogeneous or almost-homogeneous cases for which (3.13) holds were previously obtained in the literature. While existing methods of proof were mainly based on the theory of complete mixability, an extension using the same techniques to arbitrarily many different marginal distributions was not possible.

Similarly to Theorem 3.3, we can obtain the limit of the best-case VaR bounds. In the following, we define the left-tail ES (LES) as

$$\mathrm{LES}_p(X) = \frac{1}{p} \int_0^p \mathrm{VaR}_q(X)\mathrm{d}q = -\mathrm{ES}_{1-p}(-X),$$

and denote its best-case value under dependence uncertainty by

$$\underline{\mathrm{LES}}_p(S_n) := \inf_{S\in\mathfrak{S}_n} \mathrm{LES}_p(S) = \sum_{i=1}^n \mathrm{LES}_p(X_i) = \sum_{i=1}^n \mu_{0,p}^{(i)},$$

where the second equality can be seen from the symmetry between ES and LES. For the best-case VaR bounds, we use a slightly different set of conditions. For some $p \in (0, 1)$ and $k > 1$, we introduce

$$\liminf_{n\to\infty} n^{-1/k} \sum_{i=1}^n \mathrm{LES}_p(X_i) = +\infty, \tag{3.15}$$

$$C_0 := \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \mathrm{LES}_p(X_i) > 0. \tag{3.16}$$

The following corollary is obtained from Theorem 3.3 by symmetry:

Corollary 3.4 *Suppose that the distributions F_i, $i \in \mathbb{N}$, satisfy (3.9) and (3.15) for some $p \in (0, 1)$ and $k > 1$, then*

$$\lim_{n \to \infty} \frac{\mathrm{VaR}_p(S_n)}{\underline{\mathrm{LES}}_p(S_n)} = 1.$$

If, in addition, (3.15) is replaced by (3.16), then for sufficiently large n,

$$1 \geqslant \frac{\mathrm{VaR}_p(S_n)}{\underline{\mathrm{LES}}_p(S_n)} \geqslant 1 - Cn^{-1+1/k},$$

where

$$C = \left(\frac{1}{1-p} \frac{k}{k-1} + 1 \right) \frac{M^{1/k}}{C_0} + \varepsilon > 0,$$

M is given in (3.9), C_0 is given in (3.16), and ε is any fixed positive real number.

Remark 3.5 The conditions (3.15) and (3.16) are slightly stronger than (3.10) and (3.11), respectively, and this asymmetry is due to the fact that we mainly consider the cases when the aggregate risk measures LES and ES are positive. The asymmetry can be trivially removed by assuming

$$\liminf_{n \to \infty} \left| \frac{1}{n} \sum_{i=1}^{n} \mathrm{LES}_p(X_i) \right| > C_0$$

instead of (3.15).

Finally, we remark that the limit of $\overline{\Delta}_n^p$ as $n \to \infty$ can be obtained directly from Theorem 3.3. Suppose the continuous distributions F_i, $i \in \mathbb{N}$, satisfy (3.9) and (3.10) for some $p \in (0, 1)$ and $k > 1$. Then, as $n \to \infty$,

$$\overline{\Delta}_n^p \frac{\mathrm{VaR}_p^+(S_n)}{\overline{\mathrm{ES}}_p(S_n)} \to 1.$$

If in addition $R_p := \lim_{n \to \infty} \frac{\overline{\mathrm{ES}}_p(S_n)}{\mathrm{VaR}_p^+(S_n)}$ exists in $[1, \infty]$, then $\Delta_n^p \to R_p$ as $n \to \infty$.

4 Uncertainty spread of VaR and ES

In addition to the distribution-wise continuity as discussed in Sect. 2, in this section, based on results obtained in Sect. 3, we study the uncertainty spread of VaR and ES when the dependence structure is unspecified. This quantifies the magnitude of dependence uncertainty in a model for risk aggregation. We show that VaR generally exhibits a larger spread compared to ES. This result suggests that VaR is more

sensitive to dependence uncertainty compared to ES. For $p \in (0, 1)$, we define the dependence uncertainty spread (DU-spread) of VaR_p as

$$\overline{\text{VaR}}_p(S_n) - \underline{\text{VaR}}_p(S_n),$$

and of ES_p as

$$\overline{\text{ES}}_p(S_n) - \underline{\text{ES}}_p(S_n).$$

See Embrechts et al. [15] for a discussion on the DU-spread of VaR and its relevance in risk management.

By definition, $\text{ES}_p(X) \geqslant \text{VaR}_p(X)$ for any risk X, and the inequality is strict when X has a continuous distribution. Naturally, when switching from VaR to ES for the purpose of capital requirement, one should consider a lower confidence level for ES. In the most recent consultative document BCBS [5], it was proposed that for internal risk models, $\text{VaR}_{0.99}$ should be replaced by $\text{ES}_{0.975}$ which often yields a similar value to $\text{VaR}_{0.99}$ for light-tailed risks. Under the Swiss Solvency Test (SST), $\text{VaR}_{0.995}$ is used to compare with $\text{ES}_{0.99}$ to calculate the capital requirement for the change in the risk bearing capital (RBC) over a one-year period; see EIOPA [12, p. 32]. Kou and Peng [23] also proposed that in order to compare with ES_p, one could use the corresponding median shortfall (MS), which is the median of the conditional tail distribution above VaR_p and hence satisfies

$$\text{MS}_p(X) = \text{VaR}_{(p+1)/2}(X);$$

thus it is consistent with the SST regime. Hence it may be useful to compare the DU-spread of VaR_q and that of ES_p for $q \geqslant p$. The following proposition compares the DU-spread of VaR_q and that of ES_p in the asymptotic sense. In what follows, we denote by μ_n the summation of the means of $F_1, \ldots, F_n$, assumed to exist. We need an additional condition to avoid degenerate cases: for some $p \in (0, 1)$,

$$\liminf_{n \to \infty} \frac{1}{\mu_n} \sum_{i=1}^{n} \text{ES}_p(X_i) > 1. \tag{4.1}$$

Theorem 4.1 *Suppose* $1 > q \geqslant p > 0$.

(i) *Suppose that the distributions* $F_i, i \in \mathbb{N}$, *satisfy* (3.9), (3.15) *and* (4.1). *Then*

$$\liminf_{n \to \infty} \frac{\overline{\text{VaR}}_q(S_n) - \underline{\text{VaR}}_q(S_n)}{\overline{\text{ES}}_p(S_n) - \underline{\text{ES}}_p(S_n)} = \liminf_{n \to \infty} \frac{\overline{\text{ES}}_q(S_n) - \underline{\text{LES}}_q(S_n)}{\overline{\text{ES}}_p(S_n) - \underline{\text{ES}}_p(S_n)}$$

$$\geqslant \liminf_{n \to \infty} \frac{\overline{\text{ES}}_q(S_n) - \mu_n}{\overline{\text{ES}}_p(S_n) - \mu_n} \geqslant 1. \tag{4.2}$$

(ii) *Suppose that the distributions* $F_i, i \in \mathbb{N}$, *are identical and equal to a non-degenerate distribution* F, *and* $\mathbb{E}[|X|^k] < \infty$ *for some* $k > 1$, *where* $X \sim F$. *Then*

$$\liminf_{n \to \infty} \frac{\overline{\text{VaR}}_q(S_n) - \underline{\text{VaR}}_q(S_n)}{\overline{\text{ES}}_p(S_n) - \underline{\text{ES}}_p(S_n)} \geqslant \frac{\text{ES}_q(X) - \text{LES}_q(X)}{\text{ES}_p(X) - \mathbb{E}[X]} \geqslant 1. \tag{4.3}$$

Theorem 4.1 suggests that VaR is overall more sensitive to dependence uncertainty for large n, compared to ES. Numerical evidence of the comparison of the DU-spreads for VaR and ES at the same level can be found in Sect. 5, even for small values of n. Note that although the DU-spread of ES is smaller than that of VaR asymptotically, both risk measures have a rather large uncertainty spread in general, suggesting that dependence uncertainty in risk aggregation must be taken with care, no matter whether ES or VaR is chosen as the underlying risk measure; see Aas and Puccetti [1] for values in the context of a real life example.

Remark 4.2 In the homogeneous case, for any continuous distribution F, the limit of the DU-spread ratio in (4.3) is strictly greater than 1 since $\mathrm{LES}_q(X) < \mathbb{E}[X]$ and $\mathrm{ES}_q(X) > \mathrm{ES}_p(X)$. In the case $q = p$, we note that for light-tailed risks X, $\mathrm{LES}_p(X)$ is slightly smaller than $\mathbb{E}[X]$; for heavy-tailed risks X, $\mathrm{LES}_p(X)$ can be significantly smaller than $\mathbb{E}[X]$, leading to a much larger DU-spread of VaR. From Theorem 4.1, we can also see that, approximately, the VaR_q interval under DU is $[\sum_{i=1}^n \mathrm{LES}_q(X_i), \ \sum_{i=1}^n \mathrm{ES}_q(X_i)]$ and the ES_p interval under DU is given by $[\sum_{i=1}^n \mathbb{E}[X_i], \ \sum_{i=1}^n \mathrm{ES}_p(X_i)]$.

In the following, we give a result for finite n, in the case of bounded risks. A proof can be directly obtained from Theorem 3.1.

Corollary 4.3 *Suppose that $1 > q \geqslant p > 0$, the distributions $F_1, \ldots, F_n$ are supported in $[a, b]$, $a < b$, $a, b \in \mathbb{R}$, and*

$$\sum_{i=1}^n \big(\mathrm{ES}_q(X_i) + \mathbb{E}[X_i] - \mathrm{ES}_p(X_i) - \mathrm{LES}_q(X_i)\big) > 2(b - a), \qquad (4.4)$$

where $X_i \sim F_i$, $i = 1, \ldots, n$. Then

$$\frac{\overline{\mathrm{VaR}}_q(S_n) - \underline{\mathrm{VaR}}_q(S_n)}{\overline{\mathrm{ES}}_p(S_n) - \underline{\mathrm{ES}}_p(S_n)} > 1.$$

Note that in Corollary 4.3, since $\mathrm{ES}_q(X_i) \geqslant \mathrm{ES}_p(X_i)$ and $\mathbb{E}[X_i] \geqslant \mathrm{LES}_q(X_i)$, the left-hand side of (4.4) is the summation of n nonnegative terms, while the right-hand side of (4.4) is a constant. Hence (4.4) holds for n sufficiently large as long as the summation of the left-hand side of (4.4) diverges as $n \to \infty$.

We remark that it remains theoretically unclear under what conditions the DU-spread of VaR_q is larger than (or equal to) that of ES_p for finite n and $q \geqslant p$. In all our numerical examples (see Sect. 5 below), VaR_q always has a larger DU-spread than ES_p.

5 Numerical examples

As suggested by BCBS [5], the risk measure $\mathrm{ES}_{0.975}$ is a candidate proposed to replace $\mathrm{VaR}_{0.99}$. The SST (see EIOPA [12]) used $\mathrm{VaR}_{(1+p)/2}$ to compare with ES_p. Based on such considerations, we provide in this section the worst- and best-case values of $\mathrm{VaR}_{0.99}$, $\mathrm{VaR}_{0.9875}$, $\mathrm{VaR}_{0.975}$ and $\mathrm{ES}_{0.975}$ for different portfolios under de-

Table 1 Bounds obtained with RA ($\Delta x = 10^{-6}$), model (A), mixed portfolio

	$n = 5$			$n = 10$			$n = 20$		
	best	worst	spread	best	worst	spread	best	worst	spread
$\mathrm{ES}_{0.975}(S_n)$	22.48	44.88	22.40	22.52	55.59	33.07	29.15	102.35	73.20
$\mathrm{VaR}_{0.975}(S_n)$	9.79	41.46	31.67	10.04	52.67	42.63	21.44	100.65	79.21
$\mathrm{VaR}_{0.9875}(S_n)$	12.06	56.21	44.16	12.06	69.03	56.98	22.12	126.63	104.51
$\mathrm{VaR}_{0.99}(S_n)$	12.96	62.01	49.05	12.96	75.34	62.38	22.29	136.30	114.01
$\dfrac{\overline{\mathrm{ES}}_{0.975}(S_n)}{\overline{\mathrm{VaR}}_{0.975}(S_n)}$		1.08			1.06			1.02	

Table 2 Bounds obtained with RA ($\Delta x = 10^{-6}$), model (B), light-tailed portfolio

	$n = 5$			$n = 10$			$n = 20$		
	best	worst	spread	best	worst	spread	best	worst	spread
$\mathrm{ES}_{0.975}(S_n)$	4.72	10.71	5.99	24.55	63.19	38.63	31.33	126.38	95.04
$\mathrm{VaR}_{0.975}(S_n)$	3.69	10.57	6.88	13.61	61.41	47.81	13.61	125.73	112.13
$\mathrm{VaR}_{0.9875}(S_n)$	4.38	12.15	7.77	19.20	78.75	59.55	19.20	160.75	141.55
$\mathrm{VaR}_{0.99}(S_n)$	4.61	12.66	8.05	21.21	84.80	63.59	21.21	172.96	151.75
$\dfrac{\overline{\mathrm{ES}}_{0.975}(S_n)}{\overline{\mathrm{VaR}}_{0.975}(S_n)}$		1.01			1.03			1.01	

Table 3 Bounds obtained with RA ($\Delta x = 10^{-6}$), model (C), Pareto portfolio

	$n = 5$			$n = 10$			$n = 20$		
	best	worst	spread	best	worst	spread	best	worst	spread
$\mathrm{ES}_{0.975}(S_n)$	103.8	172.6	68.8	166.2	345.1	178.9	266.2	690.3	424.1
$\mathrm{VaR}_{0.975}(S_n)$	15.7	130.6	114.9	21.8	291.3	269.5	43.5	620.8	577.3
$\mathrm{VaR}_{0.9875}(S_n)$	22.6	207.3	184.7	27.6	462.4	434.8	46.7	985.5	938.8
$\mathrm{VaR}_{0.99}(S_n)$	25.5	240.5	215.0	30.5	536.5	506.0	47.5	1143.6	1096.0
$\dfrac{\overline{\mathrm{ES}}_{0.975}(S_n)}{\overline{\mathrm{VaR}}_{0.975}(S_n)}$		1.32			1.19			1.11	

pendence uncertainty. We compare the DU-spread of VaR and ES in each model, and also look at the influence of the number n of risks in the portfolio. The numerical calculation is carried out through the rearrangement algorithm (RA) described in Embrechts et al. [14], with discretization step $\Delta x = 10^{-6}$. The following three models are considered, and the results for $n = 5, 10, 20$ are reported in Tables 1, 2, 3.

(A) (Mixed portfolio) $S_n = X_1 + \cdots + X_n$, where $X_i \sim \mathrm{Pareto}(2 + 0.1i)$, $i = 1, \ldots, 5$, and $X_i \sim \mathrm{Exp}(i - 5)$, $i = 6, \ldots, 10$, and $X_i \sim \mathrm{lognormal}(0, (0.1(i - 10))^2)$, $i = 11, \ldots, 20$.

(B) (Light-tailed portfolio) $S_n = Y_1 + \cdots + Y_n$, where $Y_i \sim \mathrm{Exp}(i)$, $i = 1, \ldots, 5$; $Y_i \sim \mathrm{Weibull}(i - 5, 1/2)$, $i = 6, \ldots, 10$; $Y_i \overset{d}{=} Y_{i-10}$, $i = 11, \ldots, 20$.

(C) (Pareto portfolio) $S_n = Z_1 + \cdots + Z_n$, where $Z_i \sim \mathrm{Pareto}(1.5)$, $i = 1, \ldots, 20$.

From Tables 1–3, we have the following observations:

 (i) The worst-case VaR at level 0.975 and the worst-case ES at level 0.975 are very close, even for small values of n, in all models considered (cf. Theorem 3.3, (3.13)).

 (ii) The ratio between the worst-case VaR at level 0.975 and the worst-case ES at level 0.975 goes to 1 as n grows large. In the heavy-tailed model (C), the convergence is relatively slow (cf. Theorem 3.3, (3.14)).

(iii) The DU-spreads of $\mathrm{VaR}_{0.99}$, $\mathrm{VaR}_{0.985}$ and $\mathrm{VaR}_{0.975}$ are larger than those of $\mathrm{ES}_{0.975}$ in all considered models (cf. Theorem 4.1).

(iv) In the heavy-tailed model (C), the DU-spreads of VaR are significantly larger than those of ES (cf. Remark 4.2).

6 Conclusion

In this paper, we have considered the risk measures VaR and ES under dependence uncertainty. We have introduced the notion of aggregation-robustness and have shown that all coherent distortion risk measures, including ES, are aggregation-robust, but VaR is not. We have also derived bounds for the worst- and best-case VaR in aggregation and its diversification ratio under dependence uncertainty. An asymptotic equivalence between VaR and ES for inhomogeneous portfolios under the weakest so far known conditions on the marginal distributions has been established. It has been shown that when the number of risks in aggregation is large, VaR generally exhibits a larger uncertainty spread compared to ES at the same or a lower confidence level. Numerical examples have been provided to support our theoretical results. The main results in this paper suggest that ES is less sensitive with respect to dependence uncertainty in aggregation, and it typically has a smaller uncertainty spread compared to VaR.

Appendix: Proofs

A.1 A useful lemma

Before presenting the main proofs, we first state a lemma that is essential in proving the main results in Sects. 3 and 4 in this paper. Recall the definitions of the essential supremum and the essential infimum of random variables: for any random variable S,

$$\text{ess sup } S = \sup\{t : \mathbb{P}[S \leqslant t] < 1\},$$
$$\text{ess inf } S = \inf\{t : \mathbb{P}[S \leqslant t] > 0\}.$$

We denote $S_n = X_1 + \cdots + X_n$ in the following. We remind the reader that this S_n is different from the symbolic one in the notation of $\overline{\mathrm{VaR}}_p(S_n)$. We hope this will not lead to notational confusion.

Lemma A.1 *Suppose that $F_i, i \in \mathbb{N}$, is a sequence of distributions on $[0, 1]$. Then there exist $X_i \sim F_i$, $i \in \mathbb{N}$, such that for each $n \in \mathbb{N}$,*

$$\text{ess sup } S_n - \text{ess inf } S_n \leqslant 1.$$

Proof We first show that if X and Y are countermonotonic and both take values in $[0, 1]$, then ess sup $(X + Y) -$ ess inf $(X + Y) \leqslant 1$. Since X and Y are countermonotonic, there exists $U \sim U[0, 1]$ such that $X = F^{-1}(U)$ and $Y = G^{-1}(1 - U)$, where F and G are the distributions of X and Y, respectively. For $u, v \in (0, 1)$, one of $F^{-1}(u) - F^{-1}(v)$ and $G^{-1}(1 - u) - G^{-1}(1 - v)$ is nonpositive. Hence,

$$F^{-1}(u) + G^{-1}(1 - u) - \left(F^{-1}(v) + G^{-1}(1 - v)\right)$$
$$= \left(F^{-1}(u) - F^{-1}(v)\right) + \left(G^{-1}(1 - u) - G^{-1}(1 - v)\right)$$
$$\leqslant \max\{F^{-1}(u) - F^{-1}(v), G^{-1}(1 - u) - G^{-1}(1 - v)\}$$
$$\leqslant 1.$$

Thus,

$$\text{ess sup } (X + Y) - \text{ess inf } (X + Y)$$
$$= \sup_{u \in (0,1)} \{F^{-1}(u) + G^{-1}(1 - u)\} - \inf_{v \in (0,1)} \{F^{-1}(v) + G^{-1}(1 - v)\} \leqslant 1.$$

Let $X_1 \sim F_1$. For $k \geqslant 2$, choose X_k to be countermonotonic with S_{k-1}. Since ess sup $(X_1) -$ ess inf $(X_1) \leqslant 1$, by induction we get that

$$\text{ess sup } (S_k) - \text{ess inf } (S_k) = \text{ess sup } (S_{k-1} + X_k) - \text{ess inf } (S_{k-1} + X_k) \leqslant 1$$

for all $k \geqslant 2$. $\qquad\square$

Remark A.2 Lemma A.1 is of independent interest in the theory of negative dependence. Indeed, it shows that an *extremely negatively dependent* sequence always exists for uniformly bounded marginal distributions. The definition of and details on extremely negative dependence can be found in Wang and Wang [38]. In the latter paper, it was shown that an extremely negatively dependent sequence always exists for identical marginal L_1-distributions. Lemma A.1, as a new contribution, confirms that the same statement holds for inhomogeneous marginal distributions if we assume uniform boundedness.

The following useful corollary is directly implied by Lemma A.1.

Corollary A.3 *Suppose that $F_i, i \in \mathbb{N}$, is a sequence of distributions with bounded support. Then there exist $X_i \sim F_i$, $i \in \mathbb{N}$, such that for each $n \in \mathbb{N}$,*

$$|S_n - \mathbb{E}[S_n]| \leqslant L_n.$$

where L_n is the largest length of the support of F_i, $i = 1, \ldots, n$, that is,

$$L_n = \max\{\text{ess sup } X_i - \text{ess inf } X_i : X_i \sim F_i, i = 1, \ldots, n\}.$$

A.2 Proof of Theorem 2.3

Proof Suppose ρ is a coherent distortion risk measure with distortion function h. Since h is increasing and convex on $(0,1)$, its has a left derivative on $(0, 1)$, denoted as

$$\delta(t) := \lim_{x \to 0+} \frac{h(t) - h(t - x)}{x}, \quad t \in (0, 1).$$

It follows from (2.3) that $\rho(X) = \int_0^1 \mathrm{VaR}_t(X)\mathrm{d}h(t) = \int_0^1 \mathrm{VaR}_t(X)\delta(t)\mathrm{d}t$. Note that since $\mathfrak{S}_n$ is compatible with a coherent risk measure ρ, we have that $\mathbb{E}[|X_i|] < \infty$, $X_i \sim F_i$, $i = 1, \ldots, n$. For $q \in (1/2, 1)$, define

$$\tilde{\rho}_q(X) = \frac{1}{1 - h(q)} \int_q^1 \mathrm{VaR}_t(X)\delta(t)\mathrm{d}t, \quad X \in \mathcal{X}_0.$$

We can easily check that $\tilde{\rho}_q$ is also a coherent distortion risk measure.

For any $S \in \mathfrak{S}_n(F_1, \ldots, F_n)$, write $S = X_1 + \cdots + X_n$, where $X_i \sim F_i$, $i = 1, \ldots, n$. For $q \in (1/2, 1)$, we have that

$$\left| \rho(S) - \int_{1-q}^q \mathrm{VaR}_t(S)\delta(t)\mathrm{d}t \right| = \left| \int_0^{1-q} \mathrm{VaR}_t(S)\delta(t)\mathrm{d}t + \int_q^1 \mathrm{VaR}_t(S)\delta(t)\mathrm{d}t \right|$$

$$\leqslant \left| \int_0^{1-q} \mathrm{VaR}_t(S)\delta(t)\mathrm{d}t \right| + \left| (1 - h(q))\tilde{\rho}_q(S) \right|$$

$$\leqslant \delta(1 - q) \int_0^{1-q} |\mathrm{VaR}_t(S)|\mathrm{d}t + \left| (1 - h(q))\tilde{\rho}_q(S) \right|.$$

Note that

$$\left| (1 - h(q))\tilde{\rho}_q(S) \right| \leqslant \left| (1 - h(q)) \sum_{i=1}^n \tilde{\rho}_q(X_i) \right| = \left| \sum_{i=1}^n \int_q^1 \mathrm{VaR}_t(X_i)\delta(t)\mathrm{d}t \right|.$$

On the other hand, by the comonotonic additivity of VaR_t, $t \in (0, 1)$, we have that

$$\int_0^{1-q} |\mathrm{VaR}_t(S)|\mathrm{d}t = \int_0^{1-q} |\mathrm{VaR}_t(S\mathrm{I}_{\{S \geqslant 0\}}) + \mathrm{VaR}_t(S\mathrm{I}_{\{S < 0\}})|\mathrm{d}t$$

$$\leqslant \int_0^{1-q} \mathrm{VaR}_t(S\mathrm{I}_{\{S \geqslant 0\}})\mathrm{d}t + \int_0^{1-q} \mathrm{VaR}_{1-t}(-S\mathrm{I}_{\{S < 0\}})\mathrm{d}t$$

$$\leqslant \int_0^{1-q} \mathrm{VaR}_t(|S|)\mathrm{d}t + \int_0^{1-q} \mathrm{VaR}_{1-t}(|S|)\mathrm{d}t$$

$$\leqslant 2(1 - q)\mathrm{ES}_q(|S|)$$

$$\leqslant 2(1 - q) \sum_{i=1}^n \mathrm{ES}_q(|X_i|)$$

$$= 2 \sum_{i=1}^n \int_q^1 \mathrm{VaR}_t(|X_i|)\mathrm{d}t.$$

Note that for $i = 1, \ldots, n$, $\rho(X_i) < \infty$ implies that $\int_q^1 \mathrm{VaR}_t(X_i)\delta(t)\mathrm{d}t \to 0$ as $q \to 1$, and that $\mathbb{E}[|X_i|] < \infty$ implies that $\int_q^1 \mathrm{VaR}_t(|X_i|)\mathrm{d}t \to 0$ as $q \to 1$. As a consequence, as $q \to 1$,

$$\eta(q) := \left| \rho(S) - \int_{1-q}^q \mathrm{VaR}_t(S)\delta(t)\mathrm{d}t \right| \longrightarrow 0$$

uniformly in $S \in \mathfrak{S}_n$. Therefore, for each $\varepsilon > 0$, there exists $1/2 < q < 1$ such that $\eta(q) < \varepsilon/3$. By Theorem 1 of Cont et al. [10], the distortion risk measure

$$\hat{\rho}_q(X) := \frac{1}{2q - 1} \int_{1-q}^q \mathrm{VaR}_t(X)\delta(t)\mathrm{d}t, \quad X \in \mathcal{X}_0$$

is continuous at all distributions with respect to weak convergence. As a consequence, for fixed $q \in (1/2, 1)$ and $S, S_1, S_2, \ldots \in \mathfrak{S}_n$ with $S_k \to S$ weakly as $k \to \infty$, we have that there exists $K_0 \in \mathbb{N}$ such that for $k \geqslant K_0$, $|\hat{\rho}_q(S_k) - \hat{\rho}_q(S)| < \varepsilon/3$. Therefore, as $k \to \infty$,

$$|\rho(S_k) - \rho(S)| \leqslant (2q - 1)|\hat{\rho}_q(S_k) - \hat{\rho}_q(S)| + 2\eta(q) < \varepsilon.$$

Since ε is arbitrary, we conclude that ρ is aggregation-robust. $\qquad\square$

A.3 Proof of Theorem 2.5

Proof We first show that distortion risk measures with a continuous distortion function on $[0, 1]$ are aggregation-robust. Since $\mathcal{X} = L^\infty$, we suppose for some $M > 0$ that $|X_i| \leqslant M$ a.s., $X_i \sim F_i$, for all $i = 1, 2, \ldots$ For $q \in (1/2, 1)$, we have that

$$\begin{aligned}
\eta(q) &:= \left| \rho(S) - \int_{1-q}^q \mathrm{VaR}_t(S)\mathrm{d}h(t) \right| \\
&= \left| \int_0^{1-q} \mathrm{VaR}_t(S)\mathrm{d}h(t) + \int_q^1 \mathrm{VaR}_t(S)\mathrm{d}h(t) \right| \\
&\leqslant nMh(1-q) + nM\big(h(1) - h(q)\big) \longrightarrow 0,
\end{aligned}$$

uniformly in $S \in \mathfrak{S}_n$. The rest of the proof is similar to the proof of Theorem 2.3.

Now suppose that h is discontinuous at $p \in (0, 1)$. Using the same argument as in Example 2.2, we can see that ρ is not aggregation-robust. The case where h is discontinuous at $p = 0$ or $p = 1$ can be obtained with similar counterexamples. $\quad\square$

A.4 Proof of Theorem 3.1

We use the following lemma, where an alternative definition of VaR is used, namely

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

The following lemma is analogous to Lemma 4.3 of [7], with the continuity condition on the marginal distributions removed. In the following, we set $\mathfrak{S}_n = \mathfrak{S}_n(F_1, \ldots, F_n)$.

Lemma A.4 *For $p \in (0, 1)$,*

$$\sup_{S \in \mathfrak{S}_n} \mathrm{VaR}_p^*(S) = \sup\{\mathrm{ess\,inf}\, S : S \in \mathfrak{S}_n(F_{p,1}, \ldots, F_{p,n})\}, \tag{A.1}$$

$$\inf_{S \in \mathfrak{S}_n} \mathrm{VaR}_p(S) = \inf\{\mathrm{ess\,sup}\, S : S \in \mathfrak{S}_n(F_1^p, \ldots, F_n^p)\}, \tag{A.2}$$

where $F_{p,i}$ is the distribution of $F_i^{-1}(p + (1 - p)U)$, and F_i^p is the distribution of $F_i^{-1}(pU)$, $i = 1, \ldots, n$, for a random variable U uniformly distributed on $[0, 1]$.

Proof We only need to show (A.1), as (A.2) is symmetric to (A.1). First, we show that

$$\sup_{S \in \mathfrak{S}_n} \mathrm{VaR}_p^*(S) \leqslant \sup\{\mathrm{ess\,inf}\, S : S \in \mathfrak{S}_n(F_{p,1}, \ldots, F_{p,n})\} =: a_0.$$

For any $T \in \mathfrak{S}_n$, denote its distribution by F_T. Let $U \sim \mathrm{U}[0, 1]$ be such that $T = F_T^{-1}(U)$, and define $A_0 = \{U \geqslant p\}$. Write $T = X_1 + \cdots + X_n$ where $X_i \sim F_i$, $i = 1, \ldots, n$. Clearly, the conditional random variable $T|A_0 = X_1|A_0 + \cdots + X_n|A_0$ is dominated (in stochastic order) by some $S_0 \in \mathfrak{S}_n(F_{p,1}, \ldots, F_{p,n})$ since each $X_i|A_0$ is dominated by some $\hat{X}_i \sim F_{p,i}$. This implies that $\mathrm{ess\,sup}\, T_0 \leqslant a_0$, where T_0 is distributed as $T|A_0$. Therefore, $\mathrm{VaR}_p^*(T) = \mathrm{ess\,sup}\, T_0 \leqslant a_0$.

Next we show that

$$\sup_{S \in \mathfrak{S}_n} \mathrm{VaR}_p^*(S) \geqslant a_0.$$

Note that by Lemma 4.2 of Bernard et al. [7], there exists $S_0 \in \mathfrak{S}_n(F_{p,1}, \ldots, F_{p,n})$ such that $\mathrm{ess\,inf}\, S_0 = a_0$. Let U_0 be a $\mathrm{U}[0, 1]$ random variable, independent of S_0. Write

$$T_1 = \sum_{i=1}^n F_i^{-1}(U_0)\mathrm{I}_{\{U_0 < p\}} + S_0\mathrm{I}_{\{U_0 \geqslant p\}}.$$

It is easy to check that $T_1 \in \mathfrak{S}_n$. As a consequence, $\mathrm{VaR}_p^*(T_1) \geqslant \mathrm{ess\,sup}\, S_0 = a_0$. $\square$

Proof of Theorem 3.1 We first show that for $p \in (0, 1)$ and $q \in (p, 1]$,

$$\sup\{\mathrm{ess\,inf}\, S : S \in \mathfrak{S}_n(F_{p,1}, \ldots, F_{p,n})\}$$

$$\geqslant \sum_{i=1}^n \mu_{p,q}^{(i)} - \max_{i=1,\ldots,n} \left(F_i^{-1}(q) - F_i^{-1}(p)\right). \tag{A.3}$$

Since the case when $F_i^{-1}(q) = \infty$ for some i is trivial, we suppose that $F_i^{-1}(q) < \infty$ for all $i = 1, \ldots, n$.

Let $F_{p,q}^{(i)}$ be the distribution of $W_i = F_i^{-1}(p + (q - p)U)$ for $0 < p < q \leqslant 1$. By Corollary A.3, there exist random variables $X_i \sim F_{p,q}^{(i)}$, $i = 1, \ldots, n$, such that

$$X_1 + \cdots + X_n \geqslant \sum_{i=1}^n \mu_{p,q}^{(i)} - \max_{i=1,\ldots,n} \left(F^{-1}(q) - F^{-1}(p)\right).$$

Let Z_i, $i = 1, \ldots, n$, be any random variables with distribution $F_{q,i}$, and let C be a set independent of $X_1, \ldots, X_n, Z_1, \ldots, Z_n$, for which $\mathbb{P}[C] = (q - p)/(1 - p)$. Define $Y_i = X_i I_C + Z_i(1 - I_C)$ for $i = 1, \ldots, n$. It is straightforward to check that Y_i has distribution $F_{p,i}$ and

$$Y_1 + \cdots + Y_n \geqslant X_1 + \cdots + X_n \geqslant \sum_{i=1}^{n} \mu_{p,q}^{(i)} - \max_{i=1,\ldots,n} \left(F_i^{-1}(q) - F_i^{-1}(p) \right).$$

Thus

$$\operatorname{ess\,inf} (Y_1 + \cdots + Y_n) \geqslant \sum_{i=1}^{n} \mu_{p,q}^{(i)} - \max_{i=1,\ldots,n} \left(F_i^{-1}(q) - F_i^{-1}(p) \right),$$

and we obtain (A.3). Since $\operatorname{VaR}_p(X) \geqslant \operatorname{VaR}_r^*(X)$ for any $r < p$ and any random variable X, we have that

$$\overline{\operatorname{VaR}}_p(S_n) \geqslant \lim_{r \to p-} \sup_{S \in \mathfrak{S}_n} \operatorname{VaR}_r^*(S)$$

$$\geqslant \lim_{r \to p-} \left(\sum_{i=1}^{n} \mu_{r,q}^{(i)} - \max_{i=1,\ldots,n} \left(F_i^{-1}(q) - F_i^{-1}(r) \right) \right)$$

$$= \sum_{i=1}^{n} \mu_{p,q}^{(i)} - \max_{i=1,\ldots,n} \left(F_i^{-1}(q) - F_i^{-1}(p) \right).$$

Note that here we use the fact that F_i^{-1} is left-continuous for each i. Now, we have (A.3), and with Lemma A.4, we obtain the first inequality in (3.5). On the other hand,

$$\overline{\operatorname{VaR}}_p(S_n) \leqslant \sup_{S \in \mathfrak{S}_n} \operatorname{VaR}_p^*(S) = \sup\{\operatorname{ess\,inf} S : S \in \mathfrak{S}_n(F_{p,1}, \ldots, F_{p,n})\} \leqslant \sum_{i=1}^{n} \mu_{p,1}^{(i)}$$

always holds. Thus we obtain (3.5). We can show (3.6) similarly. □

A.5 Proof of Theorem 3.3

Proof First, let us assume that $\mathbb{E}[X_i] = 0$ for all $i \in \mathbb{N}$. Note that

$$\overline{\operatorname{ES}}_p(S_n) = \sum_{i=1}^{n} \operatorname{ES}_p(X_i) = \sum_{i=1}^{n} \mu_{p,1}^{(i)}$$

for $X_i \sim F_i$. We use (3.5) and take $q_n = 1 - n^{-1}$ for n large enough such that $q_n > p$. By (3.10), we have $\sum_{i=1}^{n} \mu_{p,1}^{(i)} > 0$ for large n.

Note that by (3.9), $\mathbb{E}[|X_i|^k] \leqslant M$ uniformly. Therefore, $[F_i^{-1}(t)]^k (1 - t) \leqslant M$ for $t \in (0, 1)$, and we have

$$F_i^{-1}(t) \leqslant \left(\frac{M}{1 - t} \right)^{1/k}, \quad t \in (0, 1), i \in \mathbb{N}.$$

Note that for $X_i \sim F_i$,

$$\mu_{p,1}^{(i)} - \mu_{p,q_n}^{(i)} = \frac{1}{1-p} \mathbb{E}\big[X_i \mathrm{I}_{\{X_i \geqslant F_i^{-1}(p)\}}\big] - \frac{1}{q_n - p} \mathbb{E}\big[X_i \mathrm{I}_{\{F_i^{-1}(q_n) \geqslant X_i \geqslant F_i^{-1}(p)\}}\big]$$

$$\leqslant \frac{1}{1-p} \mathbb{E}\big[X_i \mathrm{I}_{\{X_i \geqslant F^{-1}(q_n)\}}\big]$$

$$= \frac{1}{1-p} \int_{q_n}^1 F_i^{-1}(t) \mathrm{d}t$$

$$\leqslant \frac{1}{1-p} \int_{q_n}^1 \left(\frac{M}{1-t}\right)^{1/k} \mathrm{d}t$$

$$= \frac{1}{1-p} \frac{1}{1-1/k} M^{1/k} (1-q_n)^{1-1/k}.$$

As a consequence, we have

$$\sup_{S \in \mathfrak{S}_n} \mathrm{VaR}_p(S) \geqslant \sum_{i=1}^n \mu_{p,q_n}^{(i)} - \max_{i=1,\ldots,n} (F_i^{-1}(q_n) - F_i^{-1}(p))$$

$$\geqslant \sum_{i=1}^n \mu_{p,1}^{(i)} - \sum_{i=1}^n (\mu_{p,1}^{(i)} - \mu_{p,q_n}^{(i)}) - \max_{i=1,\ldots,n} F_i^{-1}(q_n)$$

$$\geqslant \sum_{i=1}^n \mu_{p,1}^{(i)} - \sum_{i=1}^n \frac{1}{1-p} \frac{1}{1-1/k} M^{1/k} (1-q_n)^{1-1/k}$$

$$- \left(\frac{M}{1-q_n}\right)^{1/k}$$

$$= \sum_{i=1}^n \mu_{p,1}^{(i)} - \frac{1}{1-p} \frac{1}{1-1/k} M^{1/k} n^{1/k} - M^{1/k} n^{1/k}$$

$$= \sum_{i=1}^n \mu_{p,1}^{(i)} - O(n^{1/k}). \tag{A.4}$$

By (3.10), it follows that

$$1 \geqslant \frac{\overline{\mathrm{VaR}}_p(S_n)}{\sum_{i=1}^n \mu_{p,1}^{(i)}} \geqslant 1 - \frac{O(n^{1/k})}{\sum_{i=1}^n \mu_{p,1}^{(i)}} \longrightarrow 1 \quad \text{as } n \to \infty,$$

hence we obtain (3.13).

Now for the case that $\mathbb{E}[X_i] \neq 0$ for some $i \in \mathbb{N}$, we denote by F_i^* the distribution of $X_i - \mathbb{E}[X_i]$ and set

$$\mathfrak{S}_n^* = \{Y_1 + \cdots + Y_n : Y_i \sim F_i^*, i = 1, \ldots, n\}.$$

Then by (A.4), with $\mathfrak{S}_n$ replaced by $\mathfrak{S}_n^*$, we have

$$
\sup_{S \in \mathfrak{S}_n} \mathrm{VaR}_p(S) = \sup_{S \in \mathfrak{S}_n^*} \mathrm{VaR}_p(S) + \sum_{i=1}^{n} \mathbb{E}[X_i]
$$

$$
= \sum_{i=1}^{n} (\mu_{p,1}^{(i)} - \mathbb{E}[X_i]) - O(n^{1/k}) + \sum_{i=1}^{n} \mathbb{E}[X_i]
$$

$$
= \sum_{i=1}^{n} \mu_{p,1}^{(i)} - O(n^{1/k}).
$$

Thus, (A.4) still holds for $\mathfrak{S}_n$ in the case $\mathbb{E}[X_i] \neq 0$ for some i.

When (3.11) holds, by (A.4), we have that

$$
1 \geqslant \frac{\overline{\mathrm{VaR}}_p(S_n)}{\sum_{i=1}^{n} \mu_{p,1}^{(i)}} \geqslant 1 - \frac{(\frac{1}{1-p}\frac{k}{k-1}+1)M^{1/k}(n^{1/k})}{\sum_{i=1}^{n} \mu_{p,1}^{(i)}} \geqslant 1 - Cn^{-1+1/k},
$$

for n sufficiently large. This leads to (3.14) and completes the proof of the theorem. $\square$

A.6 Proof of Theorem 4.1

Proof (i) Let us introduce $a_n = \overline{\mathrm{VaR}}_q(S_n)$, $b_n = \underline{\mathrm{VaR}}_q(S_n)$, $c_n = \overline{\mathrm{ES}}_q(S_n)$ and $d_n = \underline{\mathrm{LES}}_q(S_n)$. We have that

$$
\liminf_{n \to \infty} \frac{a_n - b_n}{c_n - d_n} = \liminf_{n \to \infty} \frac{a_n/c_n - b_n/c_n}{1 - d_n/c_n} = \liminf_{n \to \infty} \frac{a_n/c_n - (b_n/d_n)(d_n/c_n)}{1 - d_n/c_n}.
$$

Note that by (4.1), we have $\limsup_{n \to \infty} d_n/c_n < 1$. Further, by Theorem 3.3 and Corollary 3.4, we have $a_n/c_n \to 1$ and $b_n/d_n \to 1$. As a consequence,

$$
\liminf_{n \to \infty} \frac{a_n - b_n}{c_n - d_n} \geqslant 1.
$$

Since $c_n \geqslant a_n \geqslant b_n \geqslant d_n$, we have that

$$
\frac{a_n - b_n}{c_n - d_n} \leqslant 1 \quad \Longrightarrow \quad \lim_{n \to \infty} \frac{a_n - b_n}{c_n - d_n} = 1.
$$

Write

$$
\frac{\overline{\mathrm{ES}}_q(S_n) - \underline{\mathrm{LES}}_q(S_n)}{\overline{\mathrm{ES}}_p(S_n) - \underline{\mathrm{ES}}_p(S_n)} = \frac{\overline{\mathrm{VaR}}_q(S_n) - \underline{\mathrm{VaR}}_q(S_n)}{\overline{\mathrm{ES}}_p(S_n) - \underline{\mathrm{ES}}_p(S_n)} \frac{a_n - b_n}{c_n - d_n},
$$

and we obtain the first equality in (4.2). The rest of (4.2) follows by noting that $\mathrm{ES}_q(X) \geqslant \mathrm{ES}_p(X) \geqslant \mathbb{E}[X] \geqslant \mathrm{LES}_q(X)$ for any random variable X and any $0 < p \leqslant q < 1$.

(ii) This can be obtained from part (i) by noting that (3.9), (3.15) and (4.1) are all satisfied by the distribution of $X + c$, where c is some constant. $\square$

References

1. Aas, K., Puccetti, G.: Bounds for total economic capital: the DNB case study. Extremes **17**(4), 693–715 (2014)
2. Acerbi, C.: Spectral measures of risk: a coherent representation of subjective risk aversion. J. Bank. Finance **26**, 1505–1518 (2002)
3. Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. Math. Finance **9**, 203–228 (1999)
4. BCBS: Consultative Document May 2012. Fundamental review of the trading book. Basel Committee on Banking Supervision. Bank for International Settlements, Basel (2012). Available online http://www.bis.org/publ/bcbs219.htm
5. BCBS: Consultative Document October 2013. Fundamental review of the trading book: a revised market risk framework. Basel Committee on Banking Supervision. Bank for International Settlements, Basel (2013). Available online http://www.bis.org/publ/bcbs265.htm
6. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust Optimization. Princeton University Press, Princeton (2009)
7. Bernard, C., Jiang, X., Wang, R.: Risk aggregation with dependence uncertainty. Insur. Math. Econ. **54**, 93–108 (2014)
8. Cambou, M., Filipović, D.: Model uncertainty and scenario aggregation. Math. Finance (2015, to appear). doi:10.1111/mafi.12097
9. Cheung, K.C., Lo, A.: General lower bounds on convex functionals of aggregate sums. Insur. Math. Econ. **53**, 884–896 (2013)
10. Cont, R., Deguest, R., Scandolo, G.: Robustness and sensitivity analysis of risk measurement procedures. Quant. Finance **10**, 593–606 (2010)
11. Daníelsson, J., Jorgensen, B.N., Mandira, S., Samorodnitsky, G., de Vries, C.G.: Subadditivity reexamined: the case for Value-at-Risk. Discussion paper, 549. Financial Markets Group, London School of Economics and Political Science. Available online http://eprints.lse.ac.uk/24668
12. EIOPA: Equivalence assessment of the Swiss supervisory system in relation to articles 172, 227 and 260 of the Solvency II Directive, EIOPA-BoS-11-028 (2011). Available online https://eiopa.europa.eu/Pages/SearchResults.aspx?k=filename:EIOPA-BoS-11-028-Swiss-Equivalence-advice.pdf
13. Embrechts, P., Puccetti, G.: Bounds for functions of dependent risks. Finance Stoch. **10**, 341–352 (2006)
14. Embrechts, P., Puccetti, G., Rüschendorf, L.: Model uncertainty and VaR aggregation. J. Bank. Finance **37**, 2750–2764 (2013)
15. Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., Beleraj, A.: An academic response to Basel 3.5. Risks **2**, 25–48 (2014)
16. Emmer, S., Kratz, M., Tasche, D.: What is the best risk measure in practice? A comparison of standard measures. Preprint, ESSEC Business School (2014). Available online http://ssrn.com/abstract=2370378
17. Föllmer, H., Schied, A.: Stochastic Finance: An Introduction in Discrete Time, 3rd edn. de Gruyter, Berlin (2011)
18. Gneiting, T.: Making and evaluating point forecasts. J. Am. Stat. Assoc. **106**, 746–762 (2011)
19. Hampel, F.: A general qualitative definition of robustness. Ann. Math. Stat. **42**, 1887–1896 (1971)
20. Hansen, L.P., Sargent, T.J.: Robustness. Princeton University Press, Princeton (2007)
21. Huber, P.J., Ronchetti, E.M.: Robust Statistics, 2nd edn. Wiley Series in Probability and Statistics. Wiley, New York (2009)
22. Jorion, P.: Value at Risk: The New Benchmark for Managing Financial Risk, 3rd edn. McGraw-Hill, New York (2006)
23. Kou, S., Peng, X.: On the measurement of economic tail risk. Preprint (2014). Available online arXiv:1401.4787
24. Krätschmer, V., Schied, A., Zähle, H.: Qualitative and infinitesimal robustness of tail-dependent statistical functionals. J. Multivar. Anal. **103**, 35–47 (2012)
25. Krätschmer, V., Schied, A., Zähle, H.: Comparative and quantitative robustness for law-invariant risk measures. Finance Stoch. **18**, 271–295 (2014)
26. Kusuoka, S.: On law invariant coherent risk measures. Adv. Math. Econ. **3**, 83–95 (2001)
27. Makarov, G.D.: Estimates for the distribution function of the sum of two random variables with given marginal distributions. Theory Probab. Appl. **26**, 803–806 (1981)
28. Nelsen, R.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)

29. Puccetti, G.: Sharp bounds on the expected shortfall for a sum of dependent random variables. Stat. Probab. Lett. **83**, 1227–1232 (2013)
30. Puccetti, G., Rüschendorf, L.: Sharp bounds for sums of dependent risks. J. Appl. Probab. **50**, 42–53 (2013)
31. Puccetti, G., Rüschendorf, L.: Asymptotic equivalence of conservative value-at-risk- and expected shortfall-based capital charges. J. Risk **16**(3), 1–19 (2014)
32. Puccetti, G., Wang, B., Wang, R.: Complete mixability and asymptotic equivalence of worst-possible VaR and ES estimates. Insur. Math. Econ. **53**, 821–828 (2013)
33. Rüschendorf, L.: Random variables with maximum sums. Adv. Appl. Probab. **14**, 623–632 (1982)
34. Sandström, A.: Handbook of Solvency for Actuaries and Risk Managers: Theory and Practice. Taylor & Francis, Florida (2010)
35. SCOR: From Principle-Based Risk Management to Solvency Requirements, 2nd edn. SCOR, Zurich (2008). SCOR Switzerland AG. Available online http://www.scor.com/images/stories/pdf/scorpapers/sstbook_second_edition_final.pdf
36. van der Vaart, A.W.: Asymptotic Statistics. Cambridge University Press, Cambridge (1998)
37. Wang, B., Wang, R.: The complete mixability and convex minimization problems for monotone marginal distributions. J. Multivar. Anal. **102**, 1344–1360 (2011)
38. Wang, B., Wang, R.: Extreme negative dependence and risk aggregation. J. Multivar. Anal. **136**, 12–25 (2015).
39. Wang, R.: Asymptotic bounds for the distribution of the sum of dependent random variables. J. Appl. Probab. **51**, 780–798 (2014)
40. Wang, R., Peng, L., Yang, J.: Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. Finance Stoch. **17**, 395–417 (2013)
41. Wang, S.S., Young, V.R., Panjer, H.H.: Axiomatic characterization of insurance prices. Insur. Math. Econ. **21**, 173–183 (1997)