

Article

## Dependence Uncertainty Bounds for the Expectile of a Portfolio

Edgars Jakobsons <sup>1</sup> and Steven Vanduffel <sup>2,\*</sup>

<sup>1</sup> RiskLab, Department of Mathematics, ETH Zurich, 8092 Zürich, Switzerland;

E-Mail: edgars.jakobsons@math.ethz.ch

<sup>2</sup> Faculty of Economics, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Bruxelles, Belgium

\* Author to whom correspondence should be addressed; E-Mail: steven.vanduffel@vub.ac.be;

Tel.: +32-2-6291046.

Academic Editor: Andrea Consiglio

Received: 04 August 2015 / Accepted: 04 December 2015 / Published: 10 December 2015

---

**Abstract:** We study upper and lower bounds on the expectile risk measure of risky portfolios when the joint distribution of the risky components is not fully specified. First, we summarize methods for obtaining bounds when only the marginal distributions of the components are known, but not their interdependence (unconstrained bounds). In particular, we provide the best-possible upper bound and the best-possible lower bound (under some conditions), as well as numerical procedures to compute them. We also derive simple analytic bounds that appear adequate in various situations of interest. Second, we study bounds when some information on interdependence is available (constrained bounds). When the variance of the portfolio is known, a simple-to-compute upper bound is provided, and we illustrate that it may significantly improve the unconstrained upper bound. We also show that the unconstrained lower bound cannot be readily improved using variance information. Next, we derive improved bounds when the bivariate distributions of each of the risky components and a risk factor are known. When the factor induces a positive dependence among the components, it is typically possible to improve the unconstrained lower bound. Finally, the unconstrained dependence uncertainty spreads of expected shortfall, value-at-risk and the expectile are compared.

**Keywords:** expectiles; convex order; elicibility; coherence; dependence

---

## 1. Introduction and Preliminaries

This paper aims to contribute to the broader academic discussion on the properties of risk measures relevant to risk management and regulation in the banking and insurance industry; see [1] and [2] for an overview. The two most well-known risk measures are the value-at-risk (VaR) and expected shortfall (ES),

$$\text{VaR}_\alpha(X) = \inf \{x \in \mathbb{R} : P(X \leq x) \geq \alpha\}, \quad \text{ES}_\beta(X) = \frac{1}{1-\beta} \int_\beta^1 \text{VaR}_q(X) dq$$

where the latter is only defined for random variables (rvs)  $X$  with a finite expectation. While VaR is dominantly used in industry, it lacks the property of subadditivity and is thus not coherent in the sense of [3]. By contrast, ES, which is merely the average of all upper VaRs, is coherent. In fact, it is the smallest coherent risk measure that is more conservative than VaR (see [3]). Recently, [4] brought the issue of elicibility to the foreground. A risk measure is said to be elicitable if it is a minimizer of the expectation of some scoring function, which depends on the point forecast and the true observed loss. The work in [4] showed that VaR is elicitable (if the corresponding quantile is unique), but ES is not. While some authors interpret this to mean that ES cannot be back-tested (e.g., [5,6]), [7] argue that elicibility is relevant for relative comparisons between estimators, but not for absolute significance testing. Moreover, [8] show that the pair (VaR, ES) is jointly elicitable. Nevertheless, the question arises whether there are non-trivial coherent risk measures that are elicitable. In [9–11], it is shown that the only risk measure that is both elicitable and coherent is the expectile. The expectile is introduced in [12] as the minimizer of the expectation of an asymmetric quadratic scoring function,

$$e_\tau(X) = \operatorname{argmin}_{e \in \mathbb{R}} E[(\tau \mathbf{1}_{\{X > e\}} + (1 - \tau) \mathbf{1}_{\{X < e\}})(X - e)^2]$$

It follows that  $e_\tau(X)$  is the unique solution of the equation implied by the first order conditions (however, [13] points out that no differentiability or continuity of the distribution function is required):

$$(1 - \tau) E[(e_\tau - X) \mathbf{1}_{\{X < e_\tau\}}] = \tau E[(X - e_\tau) \mathbf{1}_{\{X > e_\tau\}}] \quad (1)$$

Expectiles are well known in regression analysis [14–16]; they are used for forecasting financial time series [17] and estimating VaR and ES [18]. A penalized least squares approach in portfolio optimization was suggested by [19]; the expectile is a special case when a quadratic downside penalty is used. The expectile is also closely related to the Omega performance measure [20]; see [21], p. 128. The expectile was first explicitly considered as a risk measure in [22], and the authors coined the acronym EVaR. This name was later adopted in other articles ([23,24]). However, even the original authors admit that this acronym was already used for economic-VaR [25] and recently also for entropic-VaR [26]. To avoid confusion, we shall use the notation  $e_\tau$ , as in [2,27] and [28], p. 290.

Throughout, we assume that the random variables represent losses. The expectile-based risk measure  $e_\tau(X)$  is subadditive and thus coherent for  $\tau \in [1/2, 1)$ ; for this property, as well as other features and representations, see [13,27]. A discussion on risk management with expectiles can be found in [24]. In the present paper, we further contribute to this discussion by examining their properties when aggregating risks.

Note that by rearranging Equation (1),

$$e_\tau = E[X] + \theta E[(X - e_\tau)\mathbf{1}_{\{X > e_\tau\}}], \quad \text{where } \theta := \frac{2\tau - 1}{1 - \tau} \geq 0 \quad \text{for } \tau \in [1/2, 1) \quad (2)$$

Thus, the expectile can also be interpreted as the insurance premium using the Dutch premium principle (see [29]), where the insurer buys an excess-of-loss reinsurance contract for any claim above the premium  $e$ , with loading factor  $\theta$  (and applies zero loading for the retained part). From Equation (2), we observe that expectiles are only implicitly defined, and their computation appears cumbersome. However, if the loss distribution  $X \sim F$  is known, the following approach can be used to compute the corresponding expectile. First, define the tail integral,<sup>1</sup> a function which will be useful for shorter notation,

$$TI_X(x) := \int_x^\infty u \, dF(u) = E[X\mathbf{1}_{\{X > x\}}]$$

Analytic expressions for  $TI_X$  are available for many commonly-used distributions, such as Pareto, log-normal, normal, Student  $t$ , exponential, gamma, and other. Next, applying Newton’s method to Equation (1) yields a practical iterative procedure for computing  $e_\tau(X) = \lim_{k \rightarrow \infty} x_k$ , given by:

$$x_{k+1} = \frac{(1 - \tau) E[X] + (2\tau - 1) TI_X(x_k)}{(1 - \tau)p_k + \tau(1 - p_k)} \quad (3)$$

where  $p_k = F(x_k)$ . An analogue for an empirical distribution using iterative reweighting is mentioned in [12] and stated explicitly in [14]. The convergence of this procedure is very fast. It is shown in [24] that for most common distributions, the expectiles are smaller than quantiles at level  $\tau$  (for  $\tau$  high enough), while for heavy-tailed (infinite variance) distributions, the opposite holds true. They coincide exactly (for all  $\tau$ ) for a Student  $t$  distribution with  $\nu = 2$  and asymptotically (as  $\tau \rightarrow 1$ ) for the Pareto(2) distribution. Therefore, initializing the procedure at  $x_0 = VaR_\tau(X)$  appears reasonable.

While the expectile is a familiar object in regression analysis, its properties relevant to risk management are less studied. The focus of this paper is on risk aggregation and measurement under model uncertainty. Often the total (aggregate) loss that a company faces can be expressed as a sum  $S = X_1 + \dots + X_d$ , where the  $X_i$  represent, e.g., the losses of different business lines or risk types. The risks  $X_i$  are typically modeled separately, and little might be known about their interdependence. We will be interested in finding the range of values a risk measure  $\rho(S)$  can take for different aggregate losses  $S \in \mathfrak{S}$ , where  $\mathfrak{S}$  is the so-called admissible class containing all of the aggregate loss distributions that are consistent with the available marginal and dependence information. In particular, define the best-possible upper bound and the best-possible lower bound as:

$$\underline{\rho} = \inf\{\rho(S) : S \in \mathfrak{S}\} \quad \text{and} \quad \bar{\rho} = \sup\{\rho(S) : S \in \mathfrak{S}\}$$

where the risk measure  $\rho$  will be either VaR, ES or the expectile. The idea to assess the impact of (partial) dependence information on risk bounds has been explored in a series of recent papers; see [30–36]. In these papers, the risk measure used was the VaR. In this paper, we will mainly focus on the expectile as a challenger for VaR.

---

<sup>1</sup> Note that for a continuous rv  $X$ ,  $ES_\beta(X) = \frac{1}{1-\beta} TI_X(VaR_\beta(X))$ .

The paper is structured as follows: Section 2 considers the case when only marginal distributions are known. We provide the best-possible upper bound, as well as the best-possible lower bound (under some conditions) and provide numerical procedures to practically compute these bounds. We also provide weaker bounds and show they are close to the best-possible ones in various situations of interest. We study the location-scale family and provide analytical expressions for the best possible bounds in this context. Section 3 gives bounds when the mean and variance of the aggregate loss are known. In Section 4, we consider the availability of dependence information through factor models. We provide various bounds in this context, and the results of this and previous sections are applied in an example using the skew- $t$  distribution. In Section 5, the width of the dependence uncertainty interval for the expectile is compared to that of VaR and ES. Finally, Section 6 summarizes the observations.

## 2. Bounds when Only the Marginal Distributions Are Known

Due to the curse of dimensionality, it is typically easier to statistically fit a one-dimensional distribution function (df) to each  $X_i$  than to fit a multivariate distribution to  $\mathbf{X} = (X_1, \dots, X_d)$ . Under an idealized version of dependence uncertainty (DU), only the marginal distributions  $X_i \sim F_i, i = 1, \dots, d$  are known, while the dependence structure (copula) is completely unknown. Hence, the aggregate loss  $S$  can be any of the elements in the (Fréchet) admissible class  $\mathfrak{S}$ ,

$$\mathfrak{S}(F_1, \dots, F_d) = \{X_1 + \dots + X_d : X_i \sim F_i, i = 1, \dots, d\}$$

The (best-possible) bounds on the expectile are denoted by  $\bar{e}_\tau$  and  $\underline{e}_\tau$ . To determine the bounds, it turns out to be sufficient to find elements in  $\mathfrak{S}(F_1, \dots, F_d)$  that are maximal, respectively, minimal in the sense of convex order.<sup>2</sup> We first recall the definition of this ordering concept and then connect it with  $\bar{e}_\tau$  and  $\underline{e}_\tau$ .

**Definition 1** (Convex order). *Let  $X$  and  $Y$  be random variables, such that:*

$$E[\phi(X)] \leq E[\phi(Y)] \quad \text{for all convex functions } \phi : \mathbb{R} \rightarrow \mathbb{R}$$

*provided the expectations exist. Then,  $X$  is said to be smaller than  $Y$  in the convex order ( $X \leq_{\text{cx}} Y$ ).*

Consider the convex functions  $\phi_e(x) = (x - e)\mathbf{1}_{\{x > e\}}$  indexed by  $e \in \mathbb{R}$ . We find that:<sup>3</sup>

$$X \leq_{\text{cx}} Y \Rightarrow E[(X - e)\mathbf{1}_{\{X > e\}}] \leq E[(Y - e)\mathbf{1}_{\{Y > e\}}] \quad \forall e \in \mathbb{R} \Rightarrow e_\tau(X) \leq e_\tau(Y) \quad (4)$$

<sup>2</sup> Likewise, the study of VaR bounds is connected to identifying (in an appropriate admissible class) the elements that are minimum in the sense of convex order, a feature that points to a similarity between the study of bounds on the expectile and the study of bounds on VaR; see Section 2.3 in [30] for these results.

<sup>3</sup> Note that Equation (4) also follows from the more general results in [37,38]. Indeed, [37] has shown that any convex risk measure  $\rho$  with the Fatou property is consistent with the convex order, meaning that  $X \leq_{\text{cx}} Y$  implies  $\rho(X) \leq \rho(Y)$ . Furthermore, [38] shows that law-invariant risk measures have the Fatou property. Since the expectile is convex and law invariant, it is consistent with the convex order. See [39] for further results on the properties of the expectile and other generalized quantiles with respect to various stochastic orders.

In particular, this shows that upper bound  $\bar{e}_\tau$ , resp., lower bound  $\underline{e}_\tau$ , is obtained if one can find the maximum, resp., minimum element, in the convex order sense in the admissible class  $\mathfrak{S}$ . The last implication in Equation (4) comes from the following lemma by taking  $G(e) = E[(Y - e)\mathbf{1}_{\{Y > e\}}]$  and noting that  $E[X] = E[Y]$ . Specifically, the following lemma connects bounds on the stop-loss premium  $E[(S - e)\mathbf{1}_{\{S > e\}}]$  with bounds on  $e_\tau(S)$ .

**Lemma 2.** *Suppose  $G : \mathbb{R} \rightarrow \mathbb{R}$  is a non-increasing function, such that:*

$$E[(S - e)\mathbf{1}_{\{S > e\}}] \leq G(e) \quad \forall e \in \mathbb{R} \tag{5}$$

Then,  $e_\tau(S) \leq e^*$ , where:

$$e^* = \inf \left\{ e \in \mathbb{R} : e \geq E[S] + \frac{2\tau - 1}{1 - \tau} G(e) \right\}$$

Analogously, a lower bound on the stop-loss premium yields a lower bound on  $e_\tau(S)$ .

**Proof.** From Equation (5), it immediately follows that:

$$e \geq E[S] + \frac{2\tau - 1}{1 - \tau} G(e) \quad \Rightarrow \quad e \geq E[S] + \frac{2\tau - 1}{1 - \tau} E[(S - e)\mathbf{1}_{\{S > e\}}] \tag{6}$$

Since for both inequalities, the right-hand side is non-increasing in  $e$ , the solution of Equation (2) (i.e.,  $e_\tau(S)$ ) must be less than or equal to  $e^*$ .  $\square$

### 2.1. Upper Bound with Marginal Information

It is shown in [40] that the comonotonic dependence structure leads to the maximal element (denoted  $S^c$ ) with respect to the convex order in the admissible class  $\mathfrak{S}$ .

$$\forall S \in \mathfrak{S} : S \leq_{cx} S^c := \sum_{i=1}^d F_i^{-1}(U), \quad \text{where } U \sim U(0, 1) \tag{7}$$

Hence, we find that  $\bar{e}_\tau = e_\tau(S^c)$ . In the case of identical margins  $F_i = F_1, i = 2, \dots, d$ , using positive homogeneity, this simplifies to:

$$\bar{e}_\tau = e_\tau(dX_1) = de_\tau(X_1) = \sum_{i=1}^d e_\tau(X_i)$$

In general, however, the expectile is not comonotone additive, and hence, the upper bound  $\bar{e}_\tau$  often needs to be computed numerically. Unfortunately, the df of  $S^c$  is typically not available in an analytical form, so the iterative procedure Equation (3) is more difficult to apply, since it would involve a nested root search. In particular, to compute  $\text{TI}_{S^c}(x_k)$  at each step, one would first need to find a  $p_k$ , such that  $F_{S^c}^{-1}(p_k) = x_k$ , and then sum up the tail integrals  $\text{TI}_i(F_i^{-1}(p_k))$  for the margins.

Since  $S^c$  is defined in Equation (7) in terms of its quantile function  $F_{S^c}^{-1}$ , it is easier to work in terms of the probability level  $p$  corresponding to the expectile. For continuous marginal distributions with densities  $f_i$ , we can again apply Newton’s method by differentiating Equation (1) with respect to

$p$  using the chain rule. This yields an iterative procedure for computing  $e_\tau(S^c) = F_{S^c}^{-1}(p)$  in terms of  $p = \lim_{k \rightarrow \infty} p_k$ , given by:

$$p_{k+1} = p_k - e_k/B(p_k) + \frac{\sum_{i=1}^d ((1 - \tau) E[X_i] + (2\tau - 1) \text{TI}_i(F_i^{-1}(p_k)))}{B(p_k)(p_k(1 - \tau) + (1 - p_k)\tau)} \tag{8}$$

where  $e_k = F_{S^c}^{-1}(p_k)$  and:

$$B(u) = (F_{S^c}^{-1})'(u) = \sum_{i=1}^d (F_i^{-1})'(u) = \sum_{i=1}^d \frac{1}{f_i(F_i^{-1}(u))}, \quad u \in (0, 1)$$

Again, since analytic expressions for the mean, the tail integral and inverse df of parametric marginal distributions are often available, this is a very fast and accurate method. It is possible that Equation (8) yields  $p_{k+1} \geq 1$ , in which case we can take  $p_{k+1} = (p_k + 1)/2$  instead; similarly, if  $p_{k+1} \leq 0$ , set  $p_{k+1} = p_k/2$ . Analogously to Equation (3), it is reasonable to initiate the procedure at  $p_0 = \tau$ .

In general, by subadditivity (recall that we use  $\tau \in [1/2, 1)$ ),

$$e_\tau(S) \leq \sum_{i=1}^d e_\tau(X_i) =: e_\tau^+ \tag{9}$$

so  $e_\tau^+$  is a valid upper bound, too, but it is typically not the best possible. In Section 4.3,  $e_\tau^+$  is computed, as well as the best-possible upper bound in an example with skew- $t$  distributions; we can observe that they are very close in all cases.

### 2.2. Lower Bound with Marginal Information

The analysis of the lower bound is more involved. We first observe that:

$$e_\tau \geq E[S] = \sum_{i=1}^d E[X_i] \tag{10}$$

This can be seen either by applying Jensen’s inequality to the degenerate random variable  $m = E[S]$  to show  $m \leq_{\text{cx}} S, \forall S \in \mathfrak{S}$ , or by noting that  $e_{1/2}(S) = E[S]$  and that  $e_\tau(S)$  is increasing in  $\tau$  (see, e.g., [24] for these and other properties). If the admissible class  $\mathfrak{S}$  contains the constant  $E[S]$ , then this is the smallest element in convex order, and the lower bound Equation (10) is attained (sharp). This situation is achieved when the components  $X_i$  are “compensating” for each other and corresponds to the notion of joint mixability, which was formally introduced in [41] and extends the concept of complete mixability ([42]) to the inhomogeneous case; see also [43] for an overview of these and related concepts. Precisely, a distribution  $F$  is called  $d$ -completely mixable if there exist rvs  $X_i \sim F, i = 1, \dots, d$ , such that  $\sum_{i=1}^d X_i = dc$  a.s.,  $c \in \mathbb{R}$ . Analogously, a  $d$ -tuple of dfs  $(F_1, \dots, F_d)$  is called jointly mixable if there exist rvs  $X_i \sim F_i$ , such that  $\sum_{i=1}^d X_i = k$  a.s.,  $k \in \mathbb{R}$ . Another concept that leads to an explicit smallest element in the Fréchet admissible class is that of mutual exclusivity, which requires that the margins have a large probability mass at an endpoint of the support; see [44].

In general, however, the dependence structure that leads to the smallest element in the convex order is known only for  $d = 2$  (countermonotonicity). If  $d > 2$ , for distributions that are bounded below

(and satisfy some further conditions), in the case of identical margins, one can use the method from [45] (involves solving an integral equation), or for different margins, the method from [46] (requires solving a functional equation). A general, but approximate method is the rearrangement algorithm (RA) (see [47–49]), which is based on a discretization of the margins. In the following, we describe a simple, yet necessary modification of the RA that provides improved approximations for the best-possible lower bounds in the case of expectation-based risk measures, such as ES and the expectile.

### 2.2.1. Rearrangement Algorithm

The most general method currently available for computing lower bounds for the common risk measures is the RA; see [47–49]. While the quantile-based RA performs well when computing bounds on the (quantile-based) VaR, [50] indicates that for heavy-tailed margins, the RA lower bound for ES ([48]) is not sharp, because the tail expectation is underestimated due to discretization. Since also expectiles are defined in terms of the tail expectation (see Equation (2)), it is important to address this issue of the RA. In the following, we first recall the standard discretization of the RA for each margin  $i = 1, \dots, d$  and next provide two modifications that we further investigate.

$$\begin{aligned}
 \text{Standard RA: } & x_{k,i} = F_i^{-1} \left( \frac{k-1}{N} \right), \quad k = 1, \dots, N. \\
 \text{Midpoint RA: } & x_{k,i} = F_i^{-1} \left( \frac{k-1/2}{N} \right), \quad k = 1, \dots, N. \\
 \text{Expectation RA: } & x_{k,i} = F_i^{-1} \left( \frac{k-1/2}{N} \right), \quad k = 2, \dots, N-1, \\
 & x_{N,i} = \text{ES}_{1-1/N}(X_i), \quad x_{1,i} = \text{LES}_{1/N}(X_i) := N \int_0^{1/N} F_i^{-1}(q) \, dq
 \end{aligned}$$

While the standard discretization may seem conservative, it is nonetheless an approximate lower bound, since the RA may stop at a suboptimal rearrangement (see [48]). Moreover, if the distribution is unbounded below, then  $F_i^{-1}(0) = -\infty$ . For high  $N$  (such that  $pN > d$ ), this would still give a finite lower bound for ES; however, it would be undefined for the expectile, since it depends on both the upper and the lower tail of the aggregate distribution. The midpoint RA avoids this problem, but still underestimates tail expectations; the expectation RA should solve both issues. To evaluate and compare the sharpness of the bounds obtained using the different discretizations, we consider the homogeneous case with Pareto marginals. In this case, the exact lower bound on ES can be obtained using the method in [45]. In Table 1, the resulting underestimation errors are listed. Observe that the midpoint RA improves the results considerably, but the errors are still noticeable. The expectation RA, however, gives results that are within 0.1% of the true lower bound. In light of this, we will use the “expectation” version of RA for computing the unconstrained lower bounds on the expectile (*i.e.*,  $e_\tau$ ) in Section 4.3, since it provides more accurate estimates of the tail expectations. One final adjustment concerns the stopping condition for RA. In [48], the RA stops when an iteration reduces the ES by less than a pre-defined  $\varepsilon$ . To match the stopping condition with the objective, we stop the RA when the reduction in  $e_\tau$  becomes smaller than  $\varepsilon$ . This means that the expectile of the current rearrangement needs to be computed at each iteration. However, the Equation expectile of the previous iteration of the RA makes a very good initial guess for Equation (3), so this is not time-consuming. To summarize, the expectation RA only changes the way the

margins are discretized and the stopping condition; the rest of the RA remains as in [48]. For other recent developments on the RA, we refer to <https://sites.google.com/site/RearrangementAlgorithm/> and [51].

**Table 1.** Relative underestimation as a percent of the exact  $\underline{ES}_p$  for  $d$  Pareto( $\theta$ ) distributions, using the rearrangement algorithm (RA) with different discretizations of size  $N = 10^5$  and stopping condition  $\varepsilon = 10^{-4}$ .

Standard RA	$d$	$p = 0.95$						$p = 0.99$					
		1	2	3	4	5	8	1	2	3	4	5	8
$\theta = 5$		0.1	0.2	0.2	0.2	0.3	0.4	0.3	0.5	0.6	0.7	0.8	1.2
$\theta = 3$		0.3	0.4	0.5	0.6	0.7	1.0	0.7	1.1	1.4	1.7	1.9	2.6
$\theta = 2.5$		0.5	0.7	0.9	1.0	1.2	1.5	1.2	1.7	2.2	2.6	2.9	3.8
$\theta = 2$		1.2	1.6	1.9	2.2	2.5	3.1	2.5	3.4	4.2	4.9	5.4	6.9
$\theta = 1.5$		5.3	6.6	7.6	8.3	9.0	10.6	9.0	11.5	13.4	14.9	16.2	19.4
Midpoint RA	$d$	$p = 0.95$						$p = 0.99$					
		1	2	3	4	5	8	1	2	3	4	5	8
$\theta = 5$		0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2
$\theta = 3$		0.1	0.1	0.1	0.2	0.2	0.3	0.2	0.3	0.4	0.5	0.5	0.7
$\theta = 2.5$		0.2	0.2	0.3	0.3	0.4	0.5	0.4	0.6	0.7	0.8	1.0	1.3
$\theta = 2$		0.5	0.7	0.8	0.9	1.0	1.3	1.0	1.4	1.7	2.0	2.2	2.8
$\theta = 1.5$		3.0	3.8	4.3	4.7	5.1	6.0	5.1	6.5	7.5	8.3	9.0	10.6
Expectation RA	$d$	$p = 0.95$						$p = 0.99$					
		1	2	3	4	5	8	1	2	3	4	5	8
$\theta = 5$		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\theta = 3$		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\theta = 2.5$		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
$\theta = 2$		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1
$\theta = 1.5$		0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2

### 2.3. Example: Location-Scale Family

We assume that the  $X_i$  belong to the same location-scale family of dfs, i.e.,  $F_i(\cdot) = F((\cdot - \mu_i)/\sigma_i)$ ,  $i = 1, \dots, d$  for some df  $F$ . Denote also  $\mu = \sum_{i=1}^d \mu_i$  and  $\sigma = \sum_{i=1}^d \sigma_i$ .

#### 2.3.1. Upper Bound

By Equation (7), the convex order-maximal element in  $\mathfrak{S}(F_1, \dots, F_d)$  is given by:

$$S^c = \sum_{i=1}^d F_i^{-1}(U) = \mu + \sigma F^{-1}(U), \text{ where } U \sim U(0, 1)$$

Hence,

$$\bar{e}_\tau = \mu + \sigma e_\tau(F^{-1}(U)) = e_\tau^+$$



which can be computed using the procedure described in Equation (3). This also means that when the margins have the same shape, the bound based on subadditivity Equation (9) is the best possible.

### 2.3.2. Lower Bound

As mentioned before, obtaining an element in the admissible class that is minimum in the sense of convex order is often difficult or not even possible to achieve. However, in case the marginal dfs are from the same location-scale family that is symmetric, we can express the minimal element in the admissible class explicitly and, thus, also obtain the best-possible lower bound on the expectile.

**Theorem 3.** *Let  $X_i \sim F_i(\cdot) = F((\cdot - \mu_i)/\sigma_i)$ ,  $i = 1, \dots, d$  belong to the location-scale family of a symmetric df  $F$ . Suppose without loss of generality that  $\sigma_1 \geq \sigma_i$ ,  $i = 2, \dots, d$ .*

(i) *If  $\sigma_1 \geq \sum_{i=2}^d \sigma_i$ , then a minimal element in  $\mathfrak{S}(F_1, \dots, F_d)$  in convex order is:*

$$S^\ell = F_1^{-1}(U) + \sum_{i=2}^d F_i^{-1}(1 - U) = \sum_{i=1}^d \mu_i + \left(\sigma_1 - \sum_{i=2}^d \sigma_i\right) F^{-1}(U), \text{ where } U \sim U(0, 1)$$

Correspondingly,

$$e_\tau = \sum_{i=1}^d \mu_i + \left(\sigma_1 - \sum_{i=2}^d \sigma_i\right) e_\tau(F^{-1}(U))$$

(ii) *Otherwise, if  $F$  furthermore admits a unimodal density, then the minimal element in the admissible class is the constant  $\mu = \sum_{i=1}^d \mu_i$ , and thus,  $e_\tau = \mu$ .*

**Proof.** (i) Case  $\sigma_1 = \sum_{i=2}^d \sigma_i$  is trivial. If  $\sigma_1 > \sum_{i=2}^d \sigma_i$ , we use the well-known fact that the convex order is consistent with the ordering of expected shortfall (note that  $\text{VaR}_u(X) = F_X^{-1}(u)$ ). In particular, Theorem 3.A.5 in [52] states that:

$$X \leq_{\text{cx}} Y \iff \int_p^1 F_X^{-1}(u) \, du \leq \int_p^1 F_Y^{-1}(u) \, du \quad \forall p \in (0, 1) \quad \text{and} \quad E[X] = E[Y]$$

Clearly,  $X_1^\ell = F_1^{-1}(U)$  and  $X_i^\ell = F_i^{-1}(1 - U)$ ,  $i = 2, \dots, d$  have the required dfs, so  $S^\ell = \sum_{i=1}^d X_i^\ell \in \mathfrak{S}$ . If  $F$  is continuous, then for any  $S = \sum_{i=1}^d X_i \in \mathfrak{S}$  and any  $p \in (0, 1)$ , we have that:

$$\int_p^1 F_{S^\ell}^{-1}(u) \, du = E[S^\ell \mathbf{1}_{\{X_1^\ell > F_1^{-1}(p)\}}] \leq E[S \mathbf{1}_{\{X_1 > F_1^{-1}(p)\}}] \leq \int_p^1 F_S^{-1}(u) \, du \tag{11}$$

The first inequality follows from the fact that  $\{X_1^\ell > F_1^{-1}(p)\} = \{X_i^\ell < F_i^{-1}(1 - p)\}$ ,  $i = 2, \dots, d$  and  $A = \{X_i < F_i^{-1}(1 - p)\}$  minimizes  $E[X_i \mathbf{1}_A]$  over events  $A$  of probability  $1 - p$ . Similarly, the second inequality follows because  $E[S \mathbf{1}_A]$  is maximal when  $A = \{S > F_S^{-1}(p)\}$ .

If  $F$  is not continuous, then the indicators in Equation (11) need to be augmented by adding sets, such as:

$$\{X_1^\ell = F_1^{-1}(p), V \leq (P(X_1^\ell \leq F_1^{-1}(p)) - p)/P(X_1^\ell = F_1^{-1}(p))\}, \text{ where } V \sim U(0, 1) \text{ (independent)}$$

to the first one (and similarly for the others). Since the dfs are symmetric and belong to a location-scale family, the atoms at  $F^{-1}(p)$ ,  $F_1^{-1}(p)$  and  $F_i^{-1}(1 - p)$ ,  $i = 2, \dots, d$  are of the same size.

In Case (ii), the rvs are jointly mixable by Corollary 3.6 in [41], and hence, the result follows.  $\square$

Note that Theorem 3 is of interest beyond the context of expectiles, as the convex order least element also yields lower bounds on, e.g., variance and ES in this admissible class. An early result in this direction was [53], where identical symmetric unimodal distributions with a differentiable density are considered. It is shown that for such a df  $F$ , there exist  $X_i \sim F, i = 1, \dots, d$ , in the form  $X_i = RU_i$ , where  $U_i \sim U(-1, 1)$  and  $R$  is a continuous rv. Then, using that the uniform distribution is completely mixable (an explicit construction is given), it follows that also  $F$  is completely mixable. This model can be considered as a scale mixture of uniform distributions, where  $R$  is the scale factor, common to all margins. In Section 4, more general factor models are considered, and Theorem 3 is applied to find the minimal element in the convex order and to compute exact lower bounds in an example.

### 3. Bounds when the Mean and Variance of the Sum Are Known

We consider the case in which additional to the marginal information, also the variance of  $S$  is known, i.e., we consider the admissible class  $\mathfrak{S}(F_1, \dots, F_d, s^2)$ ,

$$\mathfrak{S}(F_1, \dots, F_d, s^2) = \{X_1 + \dots + X_d : X_i \sim F_i, i = 1, \dots, d \text{ and } \text{Var}(X_1 + \dots + X_d) = s^2\}$$

where  $s^2 > 0$  is a compatible variance constraint. In this setting, the bounds on  $e_\tau$  will be denoted  $\bar{e}_\tau^{s^2}$  and  $\underline{e}_\tau^{s^2}$ . It is not so clear how to determine these best-possible bounds. Instead, we proceed by considering a larger admissible class that is easier to deal with,<sup>4</sup> but gives weaker bounds. Note that in the case of VaR, in [30], it is shown that the weaker bounds are typically close to the best-possible ones. In the following sections, let  $m = \sum_1^d E[X_i]$ , and denote:

$$\mathcal{M}_2(m, s^2) := \{S \in L^2 : E[S] = m, \text{Var}(S) = s^2\} \tag{12}$$

#### 3.1. Upper Bound with Variance Constraint

Since  $\mathfrak{S}(F_1, \dots, F_d, s^2)$  is a subset of  $\mathcal{M}_2(m, s^2)$ , we find that  $\bar{e}_\tau^{s^2} \leq B$ , where:

$$B := \sup\{e_\tau(S) : S \in \mathcal{M}_2(m, s^2)\} \tag{13}$$

It will become apparent that variables supported on two points play an important role in the class  $\mathcal{M}_2$ .

**Definition 4.** A random variable is  $X$  called diatomic if  $P(X = a) = p$  and  $P(X = b) = 1 - p$  for some  $a < b$  and  $p \in (0, 1)$ .

The expectile of such a diatomic random variable has a simple expression,

$$e_\tau(X) = \frac{(1 - \tau)pa + \tau(1 - p)b}{(1 - \tau)p + \tau(1 - p)} \tag{14}$$

---

<sup>4</sup> Note indeed that the admissible class  $\mathfrak{S}(F_1, \dots, F_d, s^2)$  reflects  $d + 1$  constraints rendering optimization difficult. By relaxing the  $d$  (infinite dimensional) constraints on the marginal distributions and substituting them by the portfolio mean constraint, we enlarge the class (as there are many marginal distributions that yield the same portfolio mean) and effectively obtain two constraints only, which greatly facilitates the optimization.

**Theorem 5.** The maximum expectile  $B$  defined in Equation (13) is given by:

$$B = m + s\sqrt{\frac{q}{1-q}}$$

where  $q = (2\tau - 1)^2$ . It is attained by a diatomic rv with support  $\{a, b\}$  and mass  $\tau$  at  $a$ , where:

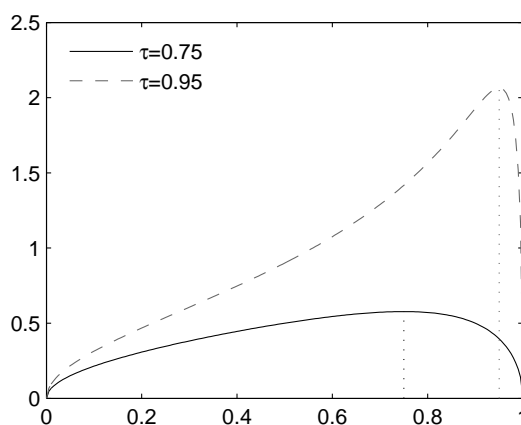
$$\begin{aligned} a &= m - s\sqrt{\frac{1-\tau}{\tau}}, \\ b &= m + s\sqrt{\frac{\tau}{1-\tau}} \end{aligned}$$

**Proof.** Denote by  $\mathcal{A} \subset \mathcal{M}_2(m, s^2)$  the subset of diatomic variables. The proof further consists of two steps. First, we construct a variable  $X_\tau \in \mathcal{A}$  that maximizes  $e_\tau$  on  $\mathcal{A}$ . Next, we show that  $X_\tau$  also provides the solution to Equation (13). Any  $X_p \in \mathcal{A}$  has two support points  $a_p, b_p$  with mass  $p$  at  $a_p$ ,

$$\begin{aligned} a_p &= m - s\sqrt{\frac{1-p}{p}}, \\ b_p &= m + s\sqrt{\frac{p}{1-p}} \end{aligned}$$

where  $0 < p < 1$ . Substituting  $a_p$  and  $b_p$  into Equation (14) yields:

$$e_\tau(X_p) = m + s \frac{(2\tau - 1)\sqrt{(1-p)p}}{(1-p)\tau + (1-\tau)p} \tag{15}$$



**Figure 1.** Expectile  $e_\tau(X_p)$  for a diatomic rv  $X_p$  (standardized to  $m = 0, s = 1$ ), as a function of  $p \in (0, 1)$ . The maximum is attained at  $p = \tau$ ; the minimum is approached as  $p \rightarrow 0$  or  $p \rightarrow 1$ .

Using differentiation with respect to  $p$ , we find that  $e_\tau(X_p)$  attains its maximum on  $(0, 1)$  when  $p = \tau$  (see also Figure 1). Hence, the variable that maximizes  $e_\tau$  on  $\mathcal{A}$  is given by:

$$X_\tau = \begin{cases} m - s\sqrt{\frac{1-\tau}{\tau}} & \text{with probability } \tau, \\ m + s\sqrt{\frac{\tau}{1-\tau}} & \text{with probability } 1 - \tau \end{cases}$$

For this rv, defining  $q = (2\tau - 1)^2$ , Equation (15) simplifies to:

$$e_\tau(X_\tau) = m + s \frac{(\tau - 1/2)}{\sqrt{(1-\tau)\tau}} = m + s\sqrt{\frac{q}{1-q}}$$

Now, consider any  $S \in \mathcal{M}_2(m, s^2)$ . Without loss of generality, we can express  $S = F_S^{-1}(U)$  for some standard uniformly distributed rv  $U$ . Letting  $p = F_S(e_\tau(S))$ , the variable  $S$  can also be written as:

$$S = F_S^{-1}(U)\mathbf{1}_{\{U \leq p\}} + F_S^{-1}(U)\mathbf{1}_{\{U > p\}}$$

Define a diatomic variable  $Y$  such that  $E[Y] = m$  and  $e_\tau(Y) = e_\tau(S)$  by

$$Y = ((m - (1 - p)ES_p(S))/p)\mathbf{1}_{\{U \leq p\}} + ES_p(S)\mathbf{1}_{\{U > p\}}$$

From Jensen’s inequality, it follows that  $\text{Var}(Y) \leq \text{Var}(S) = s^2$ . Since  $Y$  is diatomic and the right-hand side of Equation (15) is increasing in  $s$  (recall that  $\tau \in [1/2, 1)$ ),

$$e_\tau(S) = e_\tau(Y) \leq e_\tau(X_p) \leq e_\tau(X_\tau)$$

which completes the proof.  $\square$

The bound  $B$  does not make use of the specific information on the marginal distributions. When the variance is “too high”, the unconstrained bound  $\bar{e}_\tau$  will be stronger. In the opposite case,  $B$  will dominate  $\bar{e}_\tau$ . We formulate the following corollary.

**Corollary 6.**

$$\bar{e}_\tau^{s^2} \leq \min \{B, \bar{e}_\tau\}$$

**Remarks.**

- (i) A procedure called the extended rearrangement algorithm (ERA) was introduced in [30] and makes it possible to compute an approximation of  $\bar{e}_\tau^{s^2}$  from below, using both the marginal, as well as the variance information. This algorithm will be applied in an example in Section 4.3.
- (ii) Denote by  $C = \sup \{\text{VaR}_\tau(S) : S \in \mathcal{M}_2(m, s^2)\}$  and by  $D = \sup \{\text{ES}_\tau(S) : S \in \mathcal{M}_2(m, s^2)\}$ . A similar proof as in Theorem 5 shows that  $C$  and  $D$  are attained by the same diatomic variable  $X_\tau$  that attains the bound  $B$ ; see also [30]. We find that:

$$C = D = m + s\sqrt{\frac{\tau}{1 - \tau}}$$

and, thus, that  $B < C = D$ . On the other hand, the numerical value of these upper bounds would coincide for  $e_\tau, \text{VaR}_\alpha$  and  $\text{ES}_\beta$ , if we set  $\alpha = \beta = (2\tau - 1)^2$ .

**3.2. Lower Bound with Variance Constraint**

From the proof of Theorem 5, it follows that:

$$A := \inf \{e_\tau(S) : S \in \mathcal{M}_2(m, s^2)\}$$

is given by  $A = m (= E[S])$ . Indeed,  $e_\tau(X_p) \rightarrow m$  as  $p \rightarrow 0$ , respectively  $p \rightarrow 1$ ; see also Figure 1. Moreover, this bound cannot be improved by assuming either an upper bounded support or a lower

bounded support.<sup>5</sup> Hence, we conclude that working in the moment space  $\mathcal{M}_2$  does not make it readily possible to improve on  $e_\tau$ .

#### 4. Bounds for Factor Models

A factor model is introduced in [32] as a way to include additional information on the dependence structure and, hence, reduce the DU. This model considers rvs  $X_i$  and a factor  $W$  for which the bivariate distributions  $H_i$  of  $(X_i, W)$  are known. The aggregate risk  $S = X_1 + \dots + X_d$  then belongs to the factor-constrained admissible class:

$$\mathfrak{S}(H_1, \dots, H_d) = \{X_1 + \dots + X_d : (X_i, W) \sim H_i, i = 1, \dots, d\}$$

In the following, denote by  $F_i$  the marginal distribution of  $X_i$  and by  $F_{i|w}$  the conditional distribution of  $(X_i|W = w)$ ,  $i = 1, \dots, d$  (if defined). The additional information of this factor structure leads to narrower factor-constrained DU bounds,

$$\begin{aligned} \underline{e}_\tau^f &= \inf\{e_\tau(S) : S \in \mathfrak{S}^f(H_1, \dots, H_d)\}, \\ \bar{e}_\tau^f &= \sup\{e_\tau(S) : S \in \mathfrak{S}^f(H_1, \dots, H_d)\} \end{aligned}$$

In the rest of this section, we consider a model where, conditional on a non-negative factor  $W$  with distribution  $G$ , the rvs  $X_i$  belong to the location-scale family of distribution  $F_0$ .

$$X_i = \mu_i + \gamma_i W + \sigma_i \sqrt{W} Z_i, \quad Z_i \sim F_0, Z_i \perp\!\!\!\perp W, i = 1, \dots, d, \text{ and } W \sim G \tag{16}$$

Models of this type are called location-scale mixture models and have a broad range of applications, going back to [55,56], where a particular location-scale mixture family (generalized hyperbolic) is introduced. In the area of financial modeling, [57] show that this family allows a good fit of asset returns; [58,59] apply it for pricing; and [60] apply it in the context of Garch models. Specific consideration has been given in the literature to sub-families of this class; see, for example, [61,62] for the case of the multivariate variance gamma distribution, as well as [63,64] for the case of multivariate skew- $t$  distributions.

##### 4.1. Upper Bound

According to Theorem 4.1 in [32], the largest element in the convex order is achieved in the case when, conditional on the factor, the margins are comonotonic. Thus, computing the upper bound on risk in such a model is as easy as computing  $e_\tau(X_i)$ , since the conditionally comonotonic sum belongs to the same class of location-scale mixtures:

$$S_W^c = \mu + \gamma W + \sigma \sqrt{W} F_0^{-1}(U), \quad \text{where } \mu = \sum_1^d \mu_i, \gamma = \sum_1^d \gamma_i, \sigma = \sum_1^d \sigma_i$$

and  $W \perp\!\!\!\perp U \sim U(0, 1)$ . In particular, we find that  $\bar{e}_\tau^f = e_\tau(S_W^c)$ .

---

<sup>5</sup> Assuming a compact support would improve the lower bound, but we do not elaborate on this case here and refer to [54].

### 4.2. Lower Bound

The minimal element in the convex order sense is given by the following result (counterpart to Theorem 4.1 in [32]).

**Theorem 7.** *If  $S_w^\ell$  is a convex order-minimal element in  $\mathfrak{S}_w = \mathfrak{S}(F_{1|w}, \dots, F_{d|w})$  for each  $w$ , then  $S_W^\ell$  is a minimal element in  $\mathfrak{S}^f(H_1, \dots, H_d)$  and  $\underline{e}_\tau^f = e_\tau(S_W^\ell)$ .*

**Proof.** Since  $S_w^\ell \in \mathfrak{S}_w$ , it can be written as  $\sum_{i=1}^d X_{i,w}$  for some  $X_{i,w} \sim F_{i|w}$ ,  $i = 1, \dots, d$ . Thus,  $(X_{i,W}, W)$  have the required bivariate distributions, and  $S_W^\ell \in \mathfrak{S}^f$ . To show it is minimal, consider any  $T \in \mathfrak{S}^f(H_1, \dots, H_d)$ , and denote  $T_w = T|(W = w) \in \mathfrak{S}_w$ . By the definition of the convex order,  $E[\phi(S_w^\ell)] \leq E[\phi(T_w)]$  for any convex function  $\phi$ . Using monotonicity and the tower property, we obtain:

$$E[\phi(S_W^\ell)] \leq E[\phi(T)]$$

which completes the proof.  $\square$

Let  $\sigma^\ell = \max\{0, 2 \max_{i=1}^d \{\sigma_i\} - \sum_{j=1}^d \sigma_j\}$ . By Theorem 3, if the df  $F_0$  in model Equation (16) is symmetric and unimodal, then a minimal element in  $\mathfrak{S}_w$  is:

$$S_w^\ell = \mu + \gamma w + \sigma^\ell \sqrt{w} F_0^{-1}(U), \quad U \sim U(0, 1)$$

Thus, by Theorem 7,  $S_W^\ell = \mu + \gamma W + \sigma^\ell \sqrt{W} F_0^{-1}(U)$  is a convex order minimal element in the factor-constrained admissible class. Moreover, since  $U$  is independent of  $W$ ,  $S_W^\ell$  also belongs to the same mixture family as the margins, so the corresponding lower bound  $\underline{e}_\tau^f = e_\tau(S_W^\ell)$  can be computed as easily as  $e_\tau(X_i)$ . Note that the assumption that  $F_0$  is symmetric is natural, since the location-mixing term  $\gamma_i W$  can be used to add asymmetry to the df of  $X_i$ .

### 4.3. Example: Skewed Student $t$ Distribution

The results in Sections 4.1 and 4.2 apply for general choices of dfs  $F_0$  and  $G$  in Equation (16). The most well-known location-scale mixture class is that of normal mean-variance mixtures, *i.e.*, the case when  $F_0 = \Phi$ . If, in addition,  $W$  follows the generalized inverse Gaussian (GIG) distribution, then the family of generalized hyperbolic (GH) distributions is obtained; see Section 6.2.3 in [28]. This is a flexible class of distributions that exhibits skewness and heavy tails and is therefore useful for modeling financial data. Moreover, it can also be extended to the multivariate GH distribution,

$$\mathbf{X} = \boldsymbol{\mu} + \gamma W + \sqrt{W} \mathbf{A} \mathbf{Z}, \quad \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_d) \tag{17}$$

where vectors in  $\mathbb{R}^d$  are written in bold,  $I_d$  is the identity matrix and  $\Sigma = \mathbf{A} \mathbf{A}^\top$  is the Cholesky decomposition of the scale matrix. The multivariate GH class is closed under linear operations, so it has the portfolio property, which is useful for applications. In this section, a particular subclass of GH is considered: the skew- $t$  distributions.

The hyperbolic skewed Student  $t$  distribution is a special case of normal mean-variance mixtures, where the mixing distribution is the inverse-gamma df; see [64]. The inverse-gamma distribution  $\Pi(\alpha, \beta)$  has density:

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right)$$

Setting  $F_0 = \Phi$  and  $G = \text{IG}(\nu/2, \nu/2)$  in Equation (16) results in  $X_i \sim \text{Skew-}t_\nu(\mu_i, \gamma_i, \sigma_i^2)$ . The multivariate skew- $t$  subclass of the GH distribution Equation (17) is also closed under linear transformations. Using Theorem 4.1 in [32], the factor-constrained worst-case dependence structure is achieved using  $A$  with  $\sigma = (\sigma_1, \dots, \sigma_d)^\top$  as the first column and zeros in the others (conditional comonotonicity), resulting in a degenerate matrix  $\Sigma$ . The corresponding aggregate risk is:

$$S_W^c \sim \text{Skew-}t_\nu(\mu, \gamma, \sigma), \quad \text{where} \quad \mu = \sum_{i=1}^d \mu_i, \gamma = \sum_{i=1}^d \gamma_i, \sigma = \sum_{i=1}^d \sigma_i$$

Applying Theorems 3 and 7, the factor-constrained lower bounds in the “dominated” case  $\sigma^\ell = 2 \max_{i=1}^d \{\sigma_i\} - \sum_{j=1}^d \sigma_j \geq 0$  (see Case (i) of Theorem 3) can be attained using  $A$  with  $\sigma$  as the first column, except  $-\sigma_i$  in row  $i$  corresponding to the largest  $\sigma_i$  and zeros in other columns (conditional countermonotonicity with respect to the  $i$ -th margin). The corresponding aggregate risk is then:

$$S_W^\ell \sim \text{Skew-}t_\nu(\mu, \gamma, \sigma^\ell)$$

The inverse df and tail integral (as well as the df and density) of a skew- $t$  distribution can be computed using the methods in [65], which rely on the use of a Bessel function, numerical integration and root search (and are computationally intensive). Hence, we can apply the iterative algorithm Equation (3) (using the df and tail integral) to compute its expectile. Thus, we have a method to obtain the upper, respectively lower, factor-constrained DU bound. The unconstrained upper bound on the expectile can be computed using the iterative procedure Equation (8) (based on the tail integral, inverse df and density) and the unconstrained lower bound using the “expectation” version of RA introduced in Section 2.2.1.

Note that the conditionally jointly mixable case cannot be attained using a multivariate GH dependence structure. In this case,  $S_W^\ell = \mu + \gamma W$  follows a scaled and translated inverse-gamma distribution. In order to apply the iterative procedure Equation (3), we need the df and tail integral (TI) for inverse-gamma. For a general  $W \sim \text{IG}(\alpha, \beta)$ , we calculate, using the substitution  $u = \beta/t$ ,

$$F_W(x) = \int_0^x f_W(t)dt = \frac{1}{\Gamma(\alpha)} \int_{\beta/x}^\infty u^{\alpha-1} e^{-u} du$$

which is the (normalized) incomplete gamma function. In MATLAB, this can be computed using the function `gammainc(b./x,a,‘upper’)`. Similarly, we have:

$$\text{TI}_W(x) = \left( \frac{\beta}{\alpha - 1} \right) \frac{1}{\Gamma(\alpha - 1)} \int_0^{\beta/x} u^{\alpha-2} e^{-u} du$$

which is given by `b/(a-1)*gammainc(b./x,a-1,‘lower’)` in MATLAB.

In Table 2, the expectile bounds for two examples of a skew- $t$  distribution are listed. The parameters were selected to be in the range observed when fitting skew- $t$  to daily stock returns (scaled by a factor of 250) of companies in the S&P100 index. Model A is a conditionally jointly-mixable case, and Model B is a “dominated” case. First, notice that the approximate upper bound  $e_\tau^+$  is very close to the best-possible bound  $\bar{e}_\tau$  in all cases. Next, observe that due to the positive dependence the factor model induces, the value of the factor-constrained upper bound is similar to the unconstrained one, whereas the factor structure noticeably improves the lower bound; this is in agreement with the observations in [32]. In

Model A, the unconstrained lower bound is close to the mean, so the margins are “almost” jointly mixable.

**Table 2.** Upper and lower dependence uncertainty (DU) bounds for  $e_\tau(S)$  for two skew- $t$  examples with  $d = 8$ . Column  $e_\tau$  lists values for a multivariate skew- $t$  distributed  $\mathbf{X}$  with a diagonal  $\Sigma$  matrix.

<b>Model A.</b> $E[S] = 1.24, \nu = 4.5, \boldsymbol{\mu} = (-0.2, -0.15, \dots, 0.15),$ $\boldsymbol{\gamma} = (-0.25, -0.15, \dots, 0.45), \boldsymbol{\sigma} = (4.5, 5, \dots, 8)$						
$\tau$	$\underline{e}_\tau$	$\underline{e}_\tau^f$	$e_\tau$	$\bar{e}_\tau^f$	$\bar{e}_\tau$	$e_\tau^+$
0.8	1.24	2.16	13.70	35.58	35.62	35.63
0.9	1.24	3.02	21.63	57.14	57.21	57.22
0.95	1.24	4.14	29.65	78.73	78.85	78.87
0.99	1.25	8.44	51.18	135.63	135.98	136.02
0.999	1.30	23.30	96.78	251.11	252.65	252.84
<b>Model B.</b> $E[S] = 1.13, \nu = 5, \boldsymbol{\mu} = (-0.2, -0.15, \dots, 0.15),$ $\boldsymbol{\gamma} = (-0.25, -0.15, \dots, 0.45), \boldsymbol{\sigma} = (3.5, 3.5, \dots, 3.5, 25.5)$						
$\tau$	$\underline{e}_\tau$	$\underline{e}_\tau^f$	$e_\tau$	$\bar{e}_\tau^f$	$\bar{e}_\tau$	$e_\tau^+$
0.8	1.91	2.18	19.34	34.58	34.61	34.62
0.9	2.50	3.01	30.68	55.29	55.36	55.37
0.95	3.15	3.99	41.90	75.74	75.84	75.86
0.99	5.15	7.34	70.80	128.00	128.28	128.31
0.999	10.05	17.51	126.92	228.06	229.15	229.29

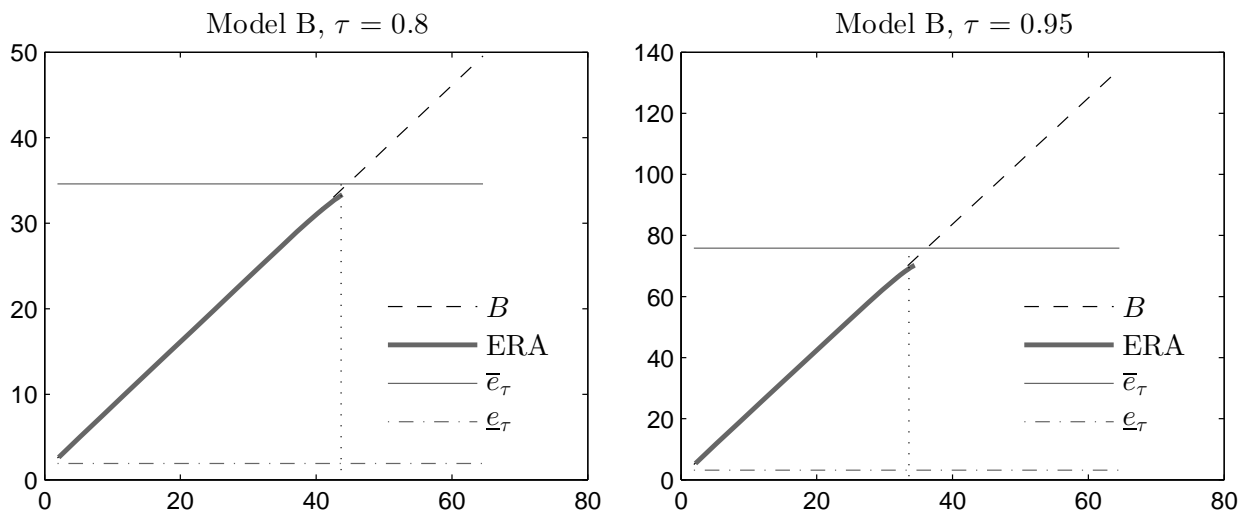
In order to illustrate the influence of variance information on the bounds, we first need to find the feasible range for  $\text{Var}(S) = s^2$ . The law of total variance yields:

$$\begin{aligned} \text{Var}(S) &= E[\text{Var}(S|W)] + \text{Var}(E[S|W]) \\ &= E[W \text{Var}(\sum \sigma_i Z_i)] + \gamma^2 \frac{2\nu^2}{(\nu - 4)(\nu - 2)^2} \end{aligned}$$

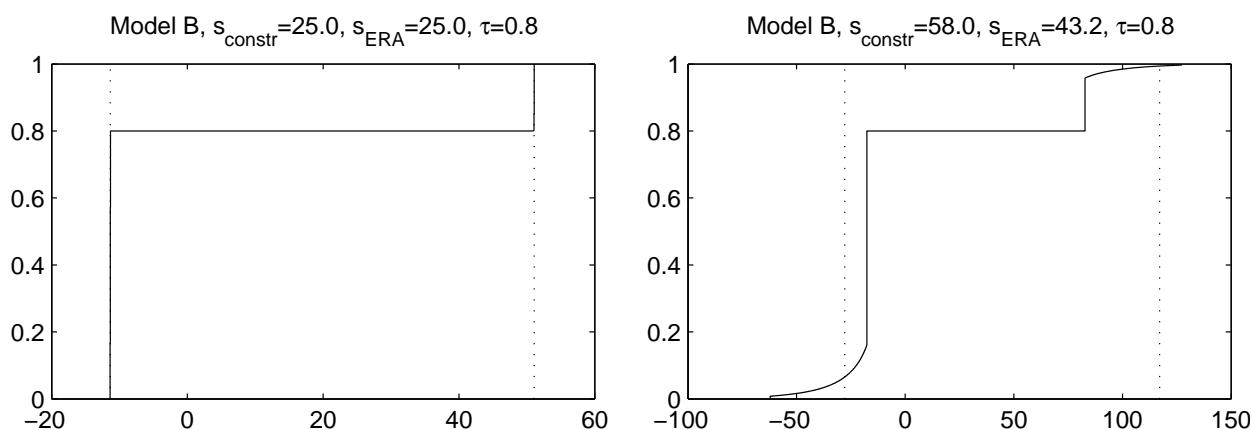
as long as  $\nu > 4$ . The first term lies in the range  $[0, \sigma^2\nu/(\nu - 2)]$ , corresponding to conditional joint mixability up to conditional comonotonicity. In Figure 2, we plot for Model B the variance-constrained bound  $B$  from Theorem 5 (based on the sole knowledge of the first two moments of the sum  $S$ ), and we compare it with the unconstrained upper bound  $\bar{e}$ . We also plot the variance-constrained bound obtained by means of the extended rearrangement algorithm (ERA), which in addition to the first two moments of  $S$ , also takes into account the marginal distributions of the components  $X_i$  (see [30] for a description of this algorithm). Variance constraints  $s^2$  are taken in the range corresponding to standard deviation  $s \in [1.9, 64.6]$ . We observe that variance information yields a considerably reduced upper bound  $B$ , as long as  $s^2$  is small enough. As the parameter  $\tau$  increases, the bound  $B$  becomes weaker and is relevant on a smaller range of  $s^2$ . The approximate bound computed using ERA is very close to the bound  $B$ , indicating that  $B$  can nearly be attained by constructing the appropriate dependence among the random variables (with the given marginal distributions). This dependence yields a sum  $S$  that has a



(nearly) diatomic structure, *i.e.*,  $S$  becomes distributed as the random variable in Theorem 5. However, the highest variance that  $S$  can possibly attain, under the constraint that it is diatomic and consistent with the marginal distributions, occurs when its upper atom is given by  $b = \sum_{i=1}^d \text{ES}_\tau(X_i)$ . Therefore, when the variance constraint is too high, we cannot expect ERA to return a diatomic distribution for  $S$ ; see also Figure 3, where the distribution function of  $S$  obtained using ERA is plotted for different variance constraints.



**Figure 2.** Moment space upper bound  $B$  on the expectile, and an approximation of  $\bar{e}_\tau^{s^2}$  computed using the extended rearrangement algorithm (ERA), as a function of the standard deviation constraint  $s$  on the horizontal axis. The unconstrained expectile bounds are also plotted for the sake of comparison. The dotted vertical line is the maximum standard deviation of a diatomic random variable, which is consistent with the marginal upper- and lower-tail expectations.



**Figure 3.** Distribution function of  $S$  computed using ERA for Model B,  $\tau = 0.8$ . In the left panel, a standard deviation constraint  $s_{\text{constr}} = 25$  is applied and attained by ERA. In the right panel, a constraint  $s_{\text{constr}} = 58$  is attempted, but cannot be attained, resulting in a lower actual standard deviation  $s_{\text{ERA}} = 43.2$ ; moreover, the distribution is not diatomic. The dotted lines are the optimal locations of the atoms from Theorem 5.

Remarks.

- (i) More generally, the GIG and, hence, (non-skew) hyperbolic distributions are infinitely divisible (see [66]), so methods from [67] can be applied to compute the inverse df and the tail integral. In turn, the iterative procedure Equation (3) can be applied to compute the expectile.
- (ii) The most time-consuming quantity to compute was the unconstrained lower bound on the expectile, because the RA requires a discretization of the margins, *i.e.*, calculating the skew- $t$  inverse df  $d \cdot N$  times (each margin took about 10 min on an Intel i5 2.5 GHz desktop with  $N = 10^4$ ). A similar calculation with Pareto dfs as in Table 1 was done for this discretization size. The maximum error using the expectation RA was 0.4% for  $p = 0.99$  and 1.5% for  $p = 0.999$ ; hence, this discretization size was deemed sufficient for our purposes.
- (iii) Due to the mixture form of GH distributions, a faster method for discretizing the margins could be using a Monte Carlo sample. Since GH dfs can have heavy tails, a similar approach to the “expectation” discretization for RA was considered, specifically, rejecting any sample points that lie below  $F_i^{-1}(1/N)$  or above  $F_i^{-1}(1 - 1/N)$  and adding two points equal to the expectations over the corresponding intervals. However, this method resulted in a large variance over repeated trials, so the obtained bounds were not used.

#### 4.4. Adding Variance Information

In this section, we consider factor models with additional variance information. We define the admissible class:

$$\mathfrak{S}(H_i, \dots, H_d, s_W^2) = \{S \in \mathfrak{S}(H_i, \dots, H_d) : \text{Var}(S|W) = s_W^2\}$$

where the conditional variance is known for each outcome of  $W$ . Consider the problem:

$$\bar{e}_\tau^{f, s^2} = \sup \{e_\tau(S) : S \in \mathfrak{S}(H_i, \dots, H_d, s_W^2)\}$$

**Theorem 8.** Let  $e^*$  be given by:

$$e^* = m + (2\tau - 1) E \left[ \sqrt{s_W^2 + (m_W - e^*)^2} \right]$$

where  $m := E[S]$  and  $m_W := E[S|W]$ . Then:

$$\bar{e}_\tau^{f, s^2} \leq e^*$$

**Proof.** Let  $S_W \in \mathfrak{S}(H_i, \dots, H_d, s_W^2)$ . By [54] (Case C<sub>13</sub>), the upper bound<sup>6</sup> on the stop-loss premium over  $S_w \in \mathcal{M}_2(m_w, s_w^2)$  (recall the definition of  $\mathcal{M}_2$  in Equation (12)) is given by:

$$E[(S_w - e)\mathbf{1}_{\{S_w > e\}}] \leq \frac{1}{2} \left( m_w - e + \sqrt{s_w^2 + (m_w - e)^2} \right) \tag{18}$$

---

<sup>6</sup> This upper bound can also be derived using the reasoning in the proof of Theorem 5. Indeed, one shows that the upper bound is attained by a diatomic variable  $X_p$  (with mean  $m_w$  and variance  $s_w^2$ ). Next, one optimizes over  $p \in (0, 1)$  to obtain Equation (18).

which holds for any  $e \in \mathbb{R}$ . Using monotonicity and the tower property, we obtain an upper bound for the unconditional stop-loss premium  $E[(S_W - e)\mathbf{1}_{\{S_W > e\}}]$ . Writing  $\theta = (2\tau - 1)/(1 - \tau)$  and invoking Lemma 2, we find that  $e_\tau(S_W) \leq e^*$ , where  $e^*$  satisfies:

$$\begin{aligned} e^* &= m + \frac{\theta}{2} E \left[ m_W - e^* + \sqrt{s_W^2 + (m_W - e^*)^2} \right] \\ &= m + \frac{\theta}{2}(m - e^*) + \frac{\theta}{2} E \left[ \sqrt{s_W^2 + (m_W - e^*)^2} \right] \end{aligned}$$

The stated equation for  $e^*$  follows by rearranging.  $\square$

Remark. If the conditional variances  $s_w^2$  are not known, but the total variance  $s^2 := \text{Var}(S)$  is available, then we still have that  $\bar{e}_\tau^{f,s^2} \leq \min \{ \bar{e}_\tau^f, B \}$ .

### 5. Dependence Uncertainty Spread Comparison

In this section, the dependence uncertainty (DU) spreads of VaR, ES and the expectile are compared, where the DU spread for a risk measure  $\rho$  is defined as:

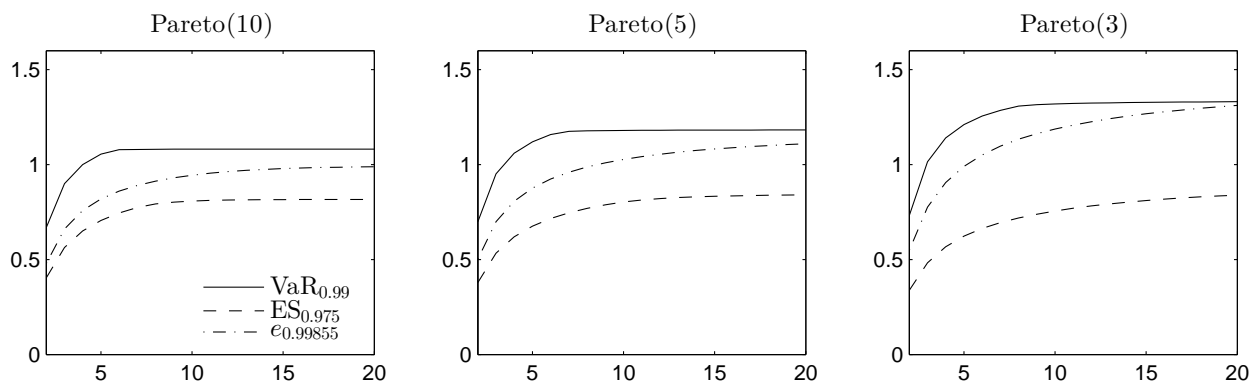
$$\bar{\rho}(S) - \underline{\rho}(S)$$

Here, we focus on the Fréchet admissible class (only marginal dfs known); see Section 2. The behavior of DU spreads of VaR and ES for large-dimensional portfolios is discussed in [68]. In order to make the resulting capital requirements similar under the different risk measures, one could, for example, use the same level  $\alpha = \beta = \tau$  for all three, but multiply by different scaling factors.

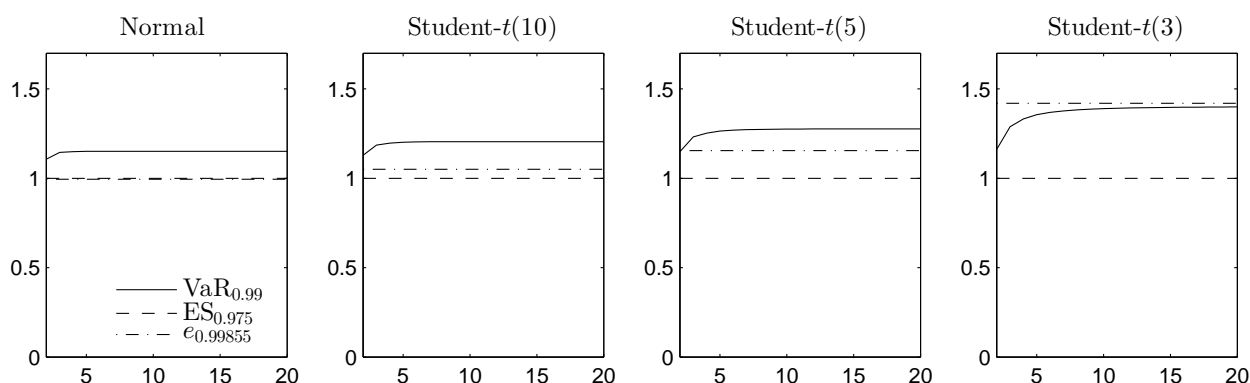
The approach taken by the Basel Committee on Banking Supervision [69], when moving from  $\text{VaR}_{0.99}$  as the risk measure for the trading book capital requirements to ES, consists of adjusting the confidence level, apparently so that the numerical value of  $\text{ES}_\beta(X)$  for a normally-distributed rv  $X$  matches  $\text{VaR}_{0.99}(X) = 2.3263$ . Doing so yields  $\beta \approx 0.97423$ , which gets rounded to  $\beta = 0.975$ . Similarly, [24] suggest using a parameter  $\tau$ , such that  $e_\tau(X) = \text{VaR}_{0.99}(X)$  for  $X \sim \Phi$ ; this yields  $\tau \approx 0.99855$ . Note that rounding to 0.999 would give  $e_{0.999} = 2.4358$ ; therefore, five significant digits  $\tau = 0.99855$  will be used in this section when comparing the expectile to the other risk measures. ES is not as sensitive to the level  $\beta$ , and  $\text{ES}_{0.975} = 2.3378$  is close enough to  $\text{VaR}_{0.99}$ .

In Figures 4 and 5, the DU spreads are plotted in the homogeneous case for different Pareto and Student  $t$  distributions, respectively, as functions of the dimension  $d$ . For the Pareto example,  $\underline{\text{VaR}}$ ,  $\overline{\text{VaR}}$ ,  $\underline{\text{ES}}$  and  $\underline{e}$  are computed from the minimal elements in the convex order, obtained using the methods from [45].  $\overline{\text{ES}}$  and  $\bar{e}$  are obtained using the comonotonic dependence structure.

Since the Student  $t$  distribution is symmetric and unimodal, it is completely mixable [53], so the lower bounds on ES and the expectile are equal to the mean. To compute the lower bound on VaR, we apply RA. As the Student  $t$  density is decreasing from the median, [45] can again be applied for the upper bound on VaR. While VaR and ES focus only on the losses, the expectile also takes the gains into account. Student  $t$  has two infinite tails, which leads to a larger DU spread for the expectile, especially in the most heavy-tailed case. Overall, the results indicate that for the chosen adjusted significance levels, the DU spread is typically the smallest for ES.



**Figure 4.** DU spread for VaR, expected shortfall (ES) and the expectile of the sum of  $d$  Pareto( $\theta$ ) distributed margins, with  $d$  on the horizontal axis. The DU spread is given relative to  $ES_{\beta}^{+}(S)$ .



**Figure 5.** The DU spread for VaR, ES and the expectile of the sum of  $d$  Gaussian or Student  $t(\nu)$  distributed margins, with  $d$  on the horizontal axis. The DU spread is given relative to  $ES_{\beta}^{+}(S)$ .

### 6. Final Remarks

In the statistics literature, the expectile functional and its properties related to regressions are well known. Recently, in the context of risk measurement, the expectile risk measure has also been shown to have appealing theoretical properties. We contribute to the analysis of this new risk measure by focusing on its properties under dependence uncertainty. We first summarize and provide improved methods for computing bounds on the expectile of a portfolio in the case of no information on dependence (unconstrained bounds) and prove analytic bounds for a location-scale family. Next, we discuss the influence of dependence information on these unconstrained bounds.

In this regard, we provide simple-to-compute bounds under an additional constraint on the portfolio variance and show that the upper bound can be considerably improved. By contrast, the unconstrained lower bound cannot be improved by only using the information on the first two moments.

Furthermore, we provide bounds in the factor-constrained case. A family of commonly-used distributions, the normal mean-variance mixtures, is considered as a special case. These models are particularly tractable, and we state the conditional best- and worst-case dependence structures explicitly. We note that due to the restriction on dependence that such a factor model induces, the lower  $e_{\tau}$  bounds

were significantly improved (for high values of  $\tau$ ). The upper bounds are only slightly reduced, and the simple (unconstrained) upper bound based on subadditivity remains adequate for practical purposes.

We compare the dependence uncertainty spread of the expectile (*i.e.*, the difference between the maximum and minimum possible value of the risk measure when only marginal information is used) with that of VaR and ES. We observe that the results are not favorable to the expectile. While the expectile has been proposed as the elicitable counterpart to ES, it is not clear that this property is indeed crucial for back-testing, and evidence exists to the contrary (e.g., [70]). Hence, gaining elicibility may not justify the increase in the dependence uncertainty spread. However, alternative mathematical approaches exist to “provide a broadly similar level of risk capture” ([71], p. 18) when moving to another risk measure (or even sticking with VaR), such as scaling. Although this makes the interpretation less clear (it was not clear for the expectile to begin with), it would allow reducing the confidence level, hence making statistical analysis more feasible and also reducing model uncertainty (see Table 2).

### Acknowledgments

The authors thank Paul Embrechts and three anonymous referees for helpful comments and suggestions. E. Jakobsons thanks the Swiss Finance Institute for financial support. S. Vanduffel acknowledges the financial support from the Flemish Science Foundation (FWO).

### Author Contributions

E.J. contributed mainly to the bounds with marginal information, examples and computations. S.V. contributed mainly to the bounds with variance information. The article was written in close collaboration.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. Embrechts, P.; Puccetti, G.; Rüschendorf, L.; Wang, R.; Beleraj, A. An academic response to Basel 3.5. *Risks* **2014**, *2*, 25–48.
2. Emmer, S.; Kratz, M.; Tasche, D. What is the best risk measure in practice? A comparison of standard measures. *J. Risk* **2015**, *18*, 31–60.
3. Artzner, P.; Delbaen, F.; Eber, J.M.; Heath, D. Coherent measures of risk. *Math. Financ.* **1999**, *9*, 203–228.
4. Gneiting, T. Making and evaluating point forecasts. *J. Am. Stat. Assoc.* **2011**, *106*, 746–762.
5. Kou, S.; Peng, X. Expected shortfall or median shortfall. *J. Financ. Eng.* **2014**, *1*, 1450007.
6. Chen, J.M. Measuring Market Risk Under the Basel Accords: VaR, Stressed VaR, and Expected Shortfall. *Aestimatio IEB Int. J. Financ.* **2014**, *8*, 184–201.
7. Acerbi, C.; Szekely, B. Back-testing expected shortfall. *Risk* **2014**, *27*, 76–81.

8. Fissler, T.; Ziegel, J.F. Higher order elicibility and Osband's principle. Available online: [arxiv.org/abs/1503.08123](http://arxiv.org/abs/1503.08123) (accessed on 30 September 2015).
9. Ziegel, J.F. Coherence and elicibility. *Math. Financ.* **2014**, doi:10.1111/mafi.12080.
10. Bellini, F.; Bignozzi, V. On elicitable risk measures. *Quant. Financ.* **2015**, *15*, 725–733.
11. Delbaen, F.; Bellini, F.; Bignozzi, V.; Ziegel, J.F. Risk measures with the CxLS property. *Financ. Stoch.* **2015**, doi:10.1007/s00780-015-0279-6.
12. Newey, W.K.; Powell, J.L. Asymmetric least squares estimation and testing. *Econom.: J. Econom. Soc.* **1987**, *55*, 819–847.
13. Delbaen, F. A remark on the structure of expectiles. Available online: [arxiv.org/abs/1307.5881](http://arxiv.org/abs/1307.5881) (accessed on 22 July 2013).
14. Efron, B. Regression percentiles using asymmetric squared error loss. *Stat. Sin.* **1991**, *1*, 93–125.
15. Yao, Q.; Tong, H. Asymmetric least squares regression estimation: A nonparametric approach. *J. Nonparametr. Stat.* **1996**, *6*, 273–292.
16. De Rossi, G.; Harvey, A. Quantiles, expectiles and splines. *J. Econom.* **2009**, *152*, 179–185.
17. Granger, C.W.; Sin, C.Y. Modelling the absolute returns of different stock indices: exploring the forecastability of an alternative measure of risk. *J. Forecast.* **2000**, *19*, 277–298.
18. Taylor, J.W. Estimating value at risk and expected shortfall using expectiles. *J. Financ. Econom.* **2008**, *6*, 231–252.
19. Manganelli, S. Asset allocation by penalized least squares. Available online: [www.ecb.europa.eu/pub/pdf/scpwps/ecbwp723.pdf](http://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp723.pdf) (accessed on 6 February 2007).
20. Keating, C.; Shadwick, W.F. A universal performance measure. *J. Perform. Meas.* **2002**, *6*, 59–84.
21. Rémillard, B. In *Statistical Methods for Financial Engineering*; CRC Press: Boca Raton, FL, USA, 2013.
22. Kuan, C.M.; Yeh, J.H.; Hsu, Y.C. Assessing value at risk with CARE, the conditional autoregressive expectile models. *J. Econom.* **2009**, *150*, 261–270.
23. De Rossi, G. Staying ahead on downside risk. In *Optimizing Optimization: The Next Generation of Optimization Applications and Theory*; Satchell, S., Ed.; Academic Press: Waltham, MA, USA, 2009; pp. 143–160.
24. Bellini, F.; di Bernardino, E. Risk management with expectiles. *Eur. J. Financ.* **2015**, doi:10.1080/1351847X.2015.1052150.
25. Ait-Sahalia, Y.; Lo, A.W. Nonparametric risk management and implied risk aversion. *J. Econom.* **2000**, *94*, 9–51.
26. Ahmadi-Javid, A. Entropic value-at-risk: A new coherent risk measure. *J. Optim. Theory Appl.* **2012**, *155*, 1105–1123.
27. Bellini, F.; Klar, B.; Müller, A.; Gianin, E.R. Generalized quantiles as risk measures. *Insur.: Math. Econ.* **2014**, *54*, 41–48.
28. McNeil, A.J.; Frey, R.; Embrechts, P. *Quantitative Risk Management: Concepts, Techniques and Tools*, 2nd ed.; Princeton University Press: Princeton, NJ, USA, 2015.
29. Van Heerwaarden, A.; Kaas, R. The Dutch premium principle. *Insur.: Math. Econ.* **1992**, *11*, 129–133.

30. Bernard, C.; Rüschendorf, L.; Vanduffel, S. Value-at-Risk bounds with variance constraints. *J. Risk Insur.* **2015**, forthcoming.
31. Bignozzi, V.; Puccetti, G.; Rüschendorf, L. Reducing model risk via positive and negative dependence assumptions. *Insur.: Math. Econ.* **2015**, *61*, 17–26.
32. Bernard, C.; Rüschendorf, L.; Vanduffel, S.; Wang, R. Risk bounds for factor models. Available online: [papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2572508](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2572508) (accessed on 26 February 2015).
33. Puccetti, G.; Rüschendorf, L.; Small, D.; Vanduffel, S. Reduction of Value-at-Risk bounds via independence and variance information. *Scand. Actuar. J.* **2015**, forthcoming.
34. Bernard, C.; Vanduffel, S. Quantile of a mixture with application to model risk assessment. *Depend. Model.* **2015**, *3*, 172–181.
35. Bernard, C.; Rüschendorf, L.; Vanduffel, S.; Yao, J. How robust is the value-at-risk of credit risk portfolios? *Eur. J. Financ.* **2015**, doi:10.1080/1351847X.2015.1104370.
36. Bernard, C.; Vanduffel, S. A new approach to assessing model risk in high dimensions. *J. Bank. Financ.* **2015**, *58*, 166–178.
37. Bäuerle, N.; Müller, A. Stochastic orders and risk measures: consistency and bounds. *Insur.: Math. Econ.* **2006**, *38*, 132–148.
38. Jouini, E.; Schachermayer, W.; Touzi, N. Law invariant risk measures have the Fatou property. In *Advances in Mathematical Economics*; Kusuoka, S., Yamazaki, A., Eds.; Springer: Berlin, Germany, 2006; pp. 49–71.
39. Bellini, F. Isotonicity properties of generalized quantiles. *Stat. Probab. Lett.* **2012**, *82*, 2017–2024.
40. Tchen, A.H. Inequalities for distributions with given marginals. *Ann. Probab.* **1980**, *8*, 814–827.
41. Wang, B.; Wang, R. Joint mixability. *Math. Oper. Res.* **2015**, forthcoming.
42. Wang, B.; Wang, R. The complete mixability and convex minimization problems with monotone marginal densities. *J. Multivar. Anal.* **2011**, *102*, 1344–1360.
43. Puccetti, G.; Wang, R. Extremal dependence concepts. *Stat. Sci.* **2015**, forthcoming.
44. Dhaene, J.; Denuit, M. The safest dependence structure among risks. *Insur.: Math. Econ.* **1999**, *25*, 11–21.
45. Bernard, C.; Jiang, X.; Wang, R. Risk aggregation with dependence uncertainty. *Insur.: Math. Econ.* **2014**, *54*, 93–108.
46. Jakobsons, E.; Han, X.; Wang, R. General convex order on risk aggregation. *Scand. Actuar. J.* **2015**, doi:10.1080/03461238.2015.1012223.
47. Embrechts, P.; Puccetti, G.; Rüschendorf, L. Model uncertainty and VaR aggregation. *J. Bank. Financ.* **2013**, *37*, 2750–2764.
48. Puccetti, G. Sharp bounds on the expected shortfall for a sum of dependent random variables. *Stat. Probab. Lett.* **2013**, *83*, 1227–1232.
49. Puccetti, G.; Rüschendorf, L. Computation of sharp bounds on the expected value of a supermodular function of risks with given marginals. *Commun. Stat.—Simul. Comput.* **2015**, *44*, 705–718.

50. Embrechts, P.; Jakobsons, E. Dependence uncertainty for aggregate risk: Examples and simple bounds. In *The Fascination of Probability, Statistics and their Applications: In Honour of Ole E. Barndorff-Nielsen*; Podolskij, M., Stelzer, R., Thorbjørnsen, S., Veraart, A., Eds.; Springer: Berlin, Germany, 2016.
51. Hofert, M.; Memartoluie, A.; Sunders, D.; Wirjanto, T. Improved algorithms for computing worst Value-at-Risk: Numerical challenges and the Adaptive Rearrangement Algorithm. Available online: [arxiv.org/abs/1505.02281](http://arxiv.org/abs/1505.02281) (accessed on 9 May 2015).
52. Shaked, M.; Shanthikumar, J.G. *Stochastic Orders*; Springer: Berlin, Germany, 2007.
53. Rüschenendorf, L.; Uckelmann, L. Variance minimization and random variables with constant sum. In *Distributions with Given Marginals and Statistical Modelling*; Springer: Berlin, Germany, 2002; pp. 211–222.
54. De Vylder, F.; Goovaerts, M.J. Analytical best upper bounds on stop-loss premiums. *Insur.: Math. Econ.* **1982**, *1*, 163–175.
55. Barndorff-Nielsen, O.E. Exponentially decreasing distributions for the logarithm of particle size. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 1977; Volume 353, pp. 401–419.
56. Barndorff-Nielsen, O.E. Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Stat.* **1978**, *5*, 151–157.
57. Barndorff-Nielsen, O.E. Normal inverse Gaussian distributions and stochastic volatility modeling. *Scand. J. Stat.* **1997**, *24*, 1–13.
58. Eberlein, E.; Keller, U. Hyperbolic distributions in finance. *Bernoulli* **1995**, *1*, 281–299.
59. Eberlein, E.; Keller, U.; Prause, K. New insights into smile, mispricing, and value at risk: The hyperbolic model. *J. Bus.* **1998**, *71*, 371–405.
60. Aas, K.; Hobæk Haff, I.; Dimakos, X.K. Risk estimation using the multivariate normal inverse Gaussian distribution. *J. Risk* **2005**, *8*, 39–60.
61. Madan, D.B.; Seneta, E. The variance gamma (VG) model for share market returns. *J. Bus.* **1990**, *63*, 511–524.
62. Madan, D.B.; Carr, P.P.; Chang, E.C. The variance gamma process and option pricing. *Eur. Financ. Rev.* **1998**, *2*, 79–105.
63. Demarta, S.; McNeil, A.J. The t copula and related copulas. *Int. Stat. Rev.* **2005**, *73*, 111–129.
64. Aas, K.; Hobæk Haff, I. The generalized hyperbolic skew Student's t-distribution. *J. Financ. Econom.* **2006**, *4*, 275–309.
65. Dokov, S.; Stoyanov, S.V.; Rachev, S. Computing VaR and AVaR of skewed-t distribution. *J. Appl. Funct. Anal.* **2008**, *3*, 189–209.
66. Barndorff-Nielsen, O.E.; Halgreen, C. Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Probab. Theory Relat. Fields* **1977**, *38*, 309–311.
67. Kim, Y.S.; Rachev, S.T.; Bianchi, M.L.; Fabozzi, F.J. Computing VaR and AVaR in infinitely divisible distributions. *Probab. Math. Stat.* **2010**, *30*, 223–245.
68. Embrechts, P.; Wang, B.; Wang, R. Aggregation-robustness and model uncertainty of regulatory risk measures. *Financ. Stoch.* **2015**, *19*, 763–790.



69. Basel Committee on Banking Supervision. *Fundamental Review of the Trading Book*; Bank of International Settlements: Basel, Switzerland, 2012.
70. Costanzino, N.; Curran, M. Backtesting general spectral risk measures with application to Expected Shortfall. Available online: [papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2514403](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2514403) (accessed on 21 February 2015).
71. Basel Committee on Banking Supervision. *Fundamental Review of the Trading Book: A Revised Market Risk Framework*; Bank of International Settlements: Basel, Switzerland, 2013.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).