On the coherence of Expected Shortfall

Carlo Acerbi[*]   Dirk Tasche[†]

April 19, 2002

Abstract

Expected Shortfall (ES) in several variants has been proposed as remedy for the deficiencies of Value-at-Risk (VaR) which in general is not a coherent risk measure. In fact, most definitions of ES lead to the same results when applied to continuous loss distributions. Differences may appear when the underlying loss distributions have discontinuities. In this case even the coherence property of ES can get lost unless one took care of the details in its definition. We compare some of the definitions of Expected Shortfall, pointing out that there is one which is robust in the sense of yielding a coherent risk measure regardless of the underlying distributions. Moreover, this Expected Shortfall can be estimated effectively even in cases where the usual estimators for VaR fail.

KEY WORDS: Expected Shortfall; Risk measure; worst conditional expectation; tail conditional expectation; value-at-risk (VaR); conditional value-at-risk (CVaR); tail mean; coherence; quantile; sub-additivity.

1   Introduction

Value-at-Risk (VaR) as a risk measure is heavily criticized for not being sub-additive (see [7] for an overview of the criticism). This means that the risk of a portfolio can be larger than the sum of the stand-alone risks of its components when measured by VaR (cf. [2], [3], [15], or [1]). Hence, managing risk by VaR may fail to stimulate diversification. Moreover, VaR does not take into account the severity of an incurred damage event.

As a response to these deficiencies the notion of *coherent* risk measures was introduced in [2], [3], and [5]. An important example for a risk measure of this kind is the *worst conditional expectation* (WCE) (cf. Definition 5.2 in [3]). This notion is closely related to the *tail conditional expectation* (TCE) from Definition 5.1 in [3], but in general does not coincide with it (see section 5 below). Unfortunately, a somewhat misleading formulation in [2] suggests this coincidence to

be true. Meanwhile, several authors (e.g. [17], [13], or [1]) proposed modifications to TCE, this way increasing confusion since the relation of these modifications to TCE and WCE remained obscure to a certain degree.

The identification of TCE and WCE is to a certain degree a temptation though the authors of [3] actually did their best to warn the reader. WCE is in fact coherent but very useful only in a theoretical setting since it requires the knowledge of the whole underlying probability space while TCE lends itself naturally to practical applications but it is not coherent (see Example 5.4 below). The goal to construct a risk measure which is both coherent and easy to compute and to estimate was however achieved in [1]. The definition of *Expected Shortfall* (ES) at a specified level α in [1] (Definition 2.6 below) is the literal mathematical transcription of the concept "average loss in the worst $100\alpha\%$ cases". We rely on this definition of Expected Shortfall in the present paper, despite the fact that in the literature this term was already used sometimes in another meaning.

With the paper at hand we strive primarily for making transparent the relations between the notions developed in [5], [3], [13], and [1]. We present four characterizations of Expected shortfall: as integral of all the quantiles below the corresponding level (eq. (3.3)), as limit in a tail strong law of large numbers (Proposition 4.1), as minimum of a certain functional introduced in [13] (Corollary 4.3 below), and as maximum of WCEs when the underlying probability space varies (Corollary 6.3). This way, we will show that the ES definition in [1] is complementary and even in some aspects superior to the other notions. Moreover, in a certain sense any law invariant coherent risk measure has a representation with ES as the main building block (see [11]).

Some hints on the organization of the paper:
In section 2 we give precise mathematical definitions to the five notions to be discussed. These are WCE, TCE, CVaR *(conditional value-at-risk)*, ES, and its negative, the so-called α-*tail mean* (TM). Section 3 presents useful properties of α-tail mean and ES, namely the integral representation (3.3), continuity and monotonicity in the level α as well as coherence for ES. In section 4 we show first that α-tail mean arises naturally as limit of the average of the $100\alpha\%$ worst cases in a sample. Then we point out that in fact ES and CVaR are two different names for the same object. Section 5 is devoted to inequalities and examples clarifying the relations between ES, TCE, and WCE. In Section 6 we deal with the question how to state a general representation of ES in terms of WCE. Section 7 concludes the paper.

2   Basic definitions

We have to arrange a minimum set of definitions to be consistent with the notions used in [5], [13], and [1]. Fix for this section some real-valued random variable X on a probability space $(\Omega, \mathcal{A}, P)$. X is considered the random profit or loss of some asset or portfolio. For the purpose of this paper, we are mainly interested in losses, i.e. low values of X. By $E[\ldots]$ we will denote

expectation with respect to P. Fix also some confidence level $\alpha \in (0, 1)$. We will often make use of the *indicator function*

$$(2.1) \qquad \mathbf{1}_A(a) \;=\; \mathbf{1}_A \;=\; \begin{cases} 1\,, \; a \in A \\ 0\,, \; a \notin A\,. \end{cases}$$

Definition 2.1 (Quantiles)
$x_{(\alpha)} = q_\alpha(X) = \inf\{x \in \mathbb{R} : \; P[X \le x] \ge \alpha\}$ *is the* lower α-quantile *of* X,
$x^{(\alpha)} = q^\alpha(X) = \inf\{x \in \mathbb{R} : \; P[X \le x] > \alpha\}$ *is the* upper α-quantile *of* X.
We use the x-notation if the dependence on X is evident, otherwise the q-notion. $\qquad \square$

Note that $x^{(\alpha)} = \sup\{x \in \mathbb{R} : \; P[X \le x] \le \alpha\}$. From $\{x \in \mathbb{R} : \; P[X \le x] > \alpha\} \subset \{x \in \mathbb{R} : \; P[X \le x] \ge \alpha\}$ it is clear that $x_{(\alpha)} \le x^{(\alpha)}$. Moreover, it is easy to see that

$$(2.2) \qquad x_{(\alpha)} \;=\; x^{(\alpha)} \quad \text{if and only if} \quad P[X \le x] \;=\; \alpha \quad \text{for at most one } x\,,$$

and in case $x_{(\alpha)} < x^{(\alpha)}$

$$(2.3) \qquad \{x \in \mathbb{R} : \; \alpha = P[X \le x]\} \;=\; \begin{cases} [x_{(\alpha)}, \, x^{(\alpha)})\,, & P[X = x^{(\alpha)}] > 0 \\ [x_{(\alpha)}, \, x^{(\alpha)}]\,, & P[X = x^{(\alpha)}] = 0\,. \end{cases}$$

(2.2) and (2.3) explain why it is difficult to say that there is an obvious definition for *value-at-risk (VaR)*. We join here [5] taking as VaR^α the smallest value such that the probability of the absolute loss being at most this value is at least $1 - \alpha$. As this is not really comprehensible when said with words here is the formal definition:

Definition 2.2 (Value-at-risk)
$\text{VaR}^\alpha = \text{VaR}^\alpha(X) = -x^{(\alpha)} = q_{1-\alpha}(-X)$ *is the* value-at-risk *at level α of X.* $\qquad \square$

The definition of *tail conditional expectation (TCE)* given in [3], Definition 5.1, depends on the choice of quantile taken for VaR (and of some discount factor we neglect here for reasons of simplicity). But as there is a choice for VaR there is also a choice for TCE. That is why we consider a lower and an upper TCE. Denote the positive part of a number x by $x^+ = \begin{cases} x\,, \; x > 0 \\ 0\,, \; x \le 0\,, \end{cases}$ and its negative part by $x^- = (-x)^+$.

Definition 2.3 (Tail conditional expectations)
 Assume $\text{E}[X^-] < \infty$. *Then* $\text{TCE}_\alpha = \text{TCE}_\alpha(X) = -\text{E}[X \,|\, X \le x_{(\alpha)}]$ *is the* lower tail conditional expectation *at level α of X.*
$\text{TCE}^\alpha = \text{TCE}^\alpha(X) = -\text{E}[X \,|\, X \le x^{(\alpha)}]$ *is the* upper tail conditional expectation *at level α of X.* $\qquad \square$

TCE^α is (up to a discount factor) the tail conditional expectation from Definition 5.1 in [3]. "Lower" and "upper" here corresponds to the quantiles used for the definitions, but not to the proportion of the quantities. In fact,

$$(2.4) \qquad \mathrm{TCE}_\alpha \quad \geq \quad \mathrm{TCE}^\alpha$$

is obvious.

As Delbaen says in the proof of Theorem 6.10 in [5], TCE^α in general does not define a sub-additive risk measure (see Example 5.4 below). For this reason, in [3], Definition 5.2, the *worst conditional expectation (WCE)* was introduced. Here is the definition (up to a discount factor) in our terms:

Definition 2.4 (Worst conditional expectation)
Assume $\mathrm{E}[X^-] < \infty$. *Then* $\mathrm{WCE}_\alpha = \mathrm{WCE}_\alpha(X) = -\inf\{\mathrm{E}[X \mid A] : \ A \in \mathcal{A}, \mathrm{P}[A] > \alpha\}$ *is the worst conditional expectation at level α of X.* □

Observe that under the assumption $\mathrm{E}[X^-] < \infty$ the value of WCE_α is always finite since then $\lim_{t\to\infty} \mathrm{P}[X \leq x^{(\alpha)} + t] = 1$ implies that there is some event $A = \{X \leq x^{(\alpha)} + t\}$ with $\mathrm{P}[A] > \alpha$ and $\mathrm{E}[\,|X|\,\mathbf{1}_A] < \infty$. We will see in section 5 that Definition 2.4 has to be treated with care nevertheless because the notion $\mathrm{WCE}_\alpha(X)$ hides the fact that it depends not only on the distribution of X but also on the structure of the underlying probability space. From the definition it is clear that for any random variables X and Y on the same probability space

$$\mathrm{WCE}_\alpha(X + Y) \quad \leq \quad \mathrm{WCE}_\alpha(X) + \mathrm{WCE}_\alpha(Y)\,,$$

i.e. WCE is sub-additive. Moreover, Proposition 5.1 in [3] says $\mathrm{WCE}_\alpha \geq \mathrm{TCE}^\alpha$. Hence WCE_α is a majorant to $\mathrm{TCE}^\alpha \geq \mathrm{VaR}^\alpha$. It is in fact the smallest coherent risk measure dominating VaR^α and only depending on X through its distribution if the underlying probability space is "rich" enough (see Theorem 6.10 in [5] for details).

This is a nice result, but to a certain degree unsatisfactory since the infimum does not seem too handy. This observation might have been the reason for introducing the *conditional value-at-risk (CVaR)* in [17] (see also the references therein) and [13]. CVaR can be used as a base for very efficient optimization procedures. We quote here, up to the sign of the random variable and the corresponding change from α to $1 - \alpha$ (cf. Definition 2.2), equation (1.2) from [13].

Definition 2.5 (Conditional value-at-risk)
Assume $\mathrm{E}[X^-] < \infty$. *Then* $\mathrm{CVaR}^\alpha = \mathrm{CVaR}^\alpha(X) = \inf\left\{ \frac{\mathrm{E}[(X-s)^-]}{\alpha} - s : \ s \in \mathbb{R} \right\}$ *is the* conditional value-at-risk *at level α of X.* □

Note that by Proposition 4.2 and (4.9), CVaR is well-defined. But beware: Pflug states in equation (1.3) of [13] (translated to our setting, i.e. $-X$ instead of Y and $1 - \alpha$ instead of α)

4

the relation $\mathrm{CVaR}^\alpha(X) = \mathrm{TCE}^\alpha(X)$, without any assumption. Corollary 5.3 in connection with Corollary 4.3 shows that this is only true if $\mathrm{P}[X < x_{(\alpha)}] = 0, \mathrm{P}[X = x^{(\alpha)}] = 0$ or $\mathrm{P}[X < x_{(\alpha)}] > 0, \mathrm{P}[X \leq x_{(\alpha)}] = \alpha$ (in particular if the distribution of X is continuous).

The last definition we need is that of α-*tail mean* from [1]. In order to make it comparable to the risk measures defined so far, we define it in two variants: the tail mean which is likely to be negative but appears in a statistical context (cf. Proposition 4.1 below), and the Expected Shortfall representing potential loss as in most cases positive number. The advantage of tail mean is the explicit representation allowing an easy proof of super-additivity (hence sub-additivity for its negative) independent of the distributions of the underlying random variables (cf. the theorem in the appendix of [1]). We will see below (Corollary 4.3) that the Expected Shortfall is in fact identical with CVaR and enjoys properties as coherence and continuity and monotonicity in the confidence level (section 3). Moreover, it is in a specific sense the largest possible value WCE can take (Corollary 6.3).

Definition 2.6 (Tail mean and Expected Shortfall)
Assume $\mathrm{E}[X^-] < \infty$. *Then*
$\bar{x}_{(\alpha)} = \mathrm{TM}_\alpha(X) = \alpha^{-1}\big(\mathrm{E}[X\,\mathbf{1}_{\{X \leq x_{(\alpha)}\}}] + x_{(\alpha)}\,(\alpha - \mathrm{P}[X \leq x_{(\alpha)}])\big)$ *is the* α-tail mean *at level* α *of* X.
$\mathrm{ES}_\alpha = \mathrm{ES}_\alpha(X) = -\bar{x}_{(\alpha)}$ *is the* Expected Shortfall (ES) *at level* α *of* X. $\qquad\qquad \square$

Note that by Corollary 4.3 α-tail mean and ES_α only depend on the distribution of X and the level α but not on a particular definition of quantile.

3   Useful properties of tail mean and Expected Shortfall

The most important property of ES (Definition 2.6) might be its coherence.

Proposition 3.1 (Coherence of ES) *Let* $\alpha \in (0,1)$ *be fixed. Consider a set* V *of real-valued random variables on some probability space* $(\Omega, \mathcal{A}, \mathrm{P})$ *such that* $\mathrm{E}[X^-] < \infty$ *for all* $X \in V$. *Then* $\rho : V \to \mathbb{R}$ *with* $\rho(X) = \mathrm{ES}_\alpha(X)$ *for* $X \in V$ *is a coherent risk measure in the sense of Definition 2.1 in [5], i.e. it is*

    (i) monotonous: $X \in V,\ X \geq 0 \quad \Rightarrow \quad \rho(X) \leq 0$,

    (ii) sub-additive: $X, Y, X + Y \in V \quad \Rightarrow \quad \rho(X + Y) \leq \rho(X) + \rho(Y)$,

    (iii) positively homogeneous: $X \in V,\ h > 0, h\,X \in V \quad \Rightarrow \quad \rho(h\,X) = h\,\rho(X)$, *and*

    (iv) translation invariant: $X \in V,\ a \in \mathbb{R} \quad \Rightarrow \quad \rho(X + a) = \rho(X) - a$.

Proof. See Proposition A.1 in the Appendix for an elementary proof of (ii). To check (i), (iii) and (iv) is an easy exercise (cf. also Proposition 3.2). □

In the financial industry there is a growing necessity to deal with random variables with discontinuous distributions. Examples are portfolios of not-traded loans (purely discrete distributions) or portfolios containing derivatives (mixtures of continuous and discrete distributions). One problem with tail risk measures like VaR, TCE, and WCE, when applied to discontinuous distributions, may be their sensitivity to small changes in the confidence level α. In other words, they are not in general continuous with respect to the confidence level α (see Example 5.4).

In contrast, ES_α is continuous with respect to α. Hence, regardless of the underlying distributions, one can be sure that the risk measured by ES_α will not change dramatically when there is a switch in the confidence level by – say – some base points. We are going to derive this insensitivity property in Corollary 3.3 below as a consequence of an alternative representation of tail mean. This integral representation (Proposition 3.2) – which was already given in [4] for the case of continuous distributions – might be of interest on its own. Another – almost self-evident – important property of ES_α is its monotonicity in α. The smaller the level α the greater is the risk. We show this formally in Proposition 3.4.

Proposition 3.2 *If X is a real-valued random variable on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ with $\mathrm{E}[X^-] < \infty$ and $\alpha \in (0,1)$ is fixed, then*

$$\bar{x}_{(\alpha)} \quad = \quad \alpha^{-1} \int_0^\alpha x_{(u)} \, d\,u \,,$$

with $\bar{x}_{(\alpha)}$ and $x_{(u)}$ as in Definitions 2.1 and 2.6, respectively.

Proof. By switching to another probability space if necessary, we can assume that there is a real random variable U on $(\Omega, \mathcal{A}, \mathrm{P})$ that is uniformly distributed on $(0,1)$, i.e. $\mathrm{P}[U \leq u] = u$, $u \in (0,1)$. It is well-known that then the random variable $Z = x_{(U)}$ has the same distribution as X.

Since $u \mapsto x_{(u)}$ is non-decreasing we have

(3.1) $$\{U \leq \alpha\} \quad \subset \quad \{Z \leq x_{(\alpha)}\} \quad \text{and}$$

(3.2) $$\{U > \alpha\} \cap \{Z \leq x_{(\alpha)}\} \quad \subset \quad \{Z = x_{(\alpha)}\} \,.$$

By (3.1) and (3.2) we obtain

$$
\begin{aligned}
\int_0^\alpha x_{(u)} \, d\,u \quad &= \quad \mathrm{E}[Z \, \mathbf{1}_{\{U \leq \alpha\}}] \\
&= \quad \mathrm{E}[Z \, \mathbf{1}_{\{Z \leq x_{(\alpha)}\}}] - \mathrm{E}[Z \, \mathbf{1}_{\{U > \alpha\} \cap \{Z \leq x_{(\alpha)}\}}] \\
&= \quad \mathrm{E}[X \, \mathbf{1}_{\{X \leq x_{(\alpha)}\}}] + x_{(\alpha)} \left(\alpha - \mathrm{P}[X \leq x_{(\alpha)}] \right).
\end{aligned}
$$

Dividing by α now yields the assertion. □

Note that by definition of Expected Shortfall, Proposition 3.2 implies the representation

$$(3.3) \qquad \mathrm{ES}_\alpha(X) \quad = \quad -\alpha^{-1} \int_0^\alpha q_u(X)\, d\,u\,.$$

Eq. (3.3) shows that ES is the coherent risk measure used in [11] as main building block for the representation of law invariant coherent risk measures.

Corollary 3.3 *If X is a real-valued random variable with $\mathrm{E}[X^-] < \infty$, then the mappings $\alpha \mapsto \bar{x}_\alpha$ and $\alpha \mapsto \mathrm{ES}_\alpha$ are continuous on $(0,1)$.*

Proof. Immediate from Proposition 3.2 and (3.3). □

For some of the results below and in particular the subsequent proposition on monotonicity of the tail mean and ES, a further representation for $\overline{x}_{(\alpha)}$ is useful (cf. Appendix in [1]). Let for $x \in \mathbb{R}$

$$(3.4) \qquad \mathbf{1}_{\{X \le x\}}^{(\alpha)} \quad = \quad \begin{cases} \mathbf{1}_{\{X \le x\}} & , \text{ if } \quad \mathrm{P}[X = x] = 0 \\ \mathbf{1}_{\{X \le x\}} + \frac{\alpha - \mathrm{P}[X \le x]}{\mathrm{P}[X = x]}\, \mathbf{1}_{\{X = x\}}, & \text{ if } \quad \mathrm{P}[X = x] > 0\,. \end{cases}$$

Then a short calculation shows

$$(3.5) \qquad \mathbf{1}_{\{X \le x_{(\alpha)}\}}^{(\alpha)} \quad \in \quad [0,1]\,,$$

$$(3.6) \qquad \mathrm{E}\left[\mathbf{1}_{\{X \le x_{(\alpha)}\}}^{(\alpha)}\right] \quad = \quad \alpha\,, \quad \text{and}$$

$$(3.7) \qquad \alpha^{-1}\mathrm{E}\left[X\, \mathbf{1}_{\{X \le x_{(\alpha)}\}}^{(\alpha)}\right] \quad = \quad \overline{x}_{(\alpha)}\,.$$

Proposition 3.4 *If X is a real-valued random variable with $\mathrm{E}[X^-] < \infty$, then for any $\alpha \in (0,1)$ and any $\epsilon > 0$ with $\alpha + \epsilon < 1$ we have the following inequalities:*

$$\overline{x}_{(\alpha + \epsilon)} \quad \ge \quad \overline{x}_{(\alpha)} \qquad and$$
$$\mathrm{ES}_{\alpha + \epsilon}(X) \quad \le \quad \mathrm{ES}_\alpha(X)\,.$$

Proof. We adopt the representation (3.7). This yields

$$\begin{aligned} \overline{x}_{(\alpha+\epsilon)} - \overline{x}_{(\alpha)} \quad &= \quad \mathrm{E}\left[X\left((\alpha+\epsilon)^{-1}\mathbf{1}_{\{X \le x_{(\alpha+\epsilon)}\}}^{(\alpha+\epsilon)} - \alpha^{-1}\mathbf{1}_{\{X \le x_{(\alpha)}\}}^{(\alpha)}\right)\right] \\ &= \quad (\alpha(\alpha+\epsilon))^{-1}\mathrm{E}\left[X\left(\alpha\,\mathbf{1}_{\{X \le x_{(\alpha+\epsilon)}\}}^{(\alpha+\epsilon)} - (\alpha+\epsilon)\,\mathbf{1}_{\{X \le x_{(\alpha)}\}}^{(\alpha)}\right)\right] \\ &\ge \quad (\alpha(\alpha+\epsilon))^{-1}\mathrm{E}\left[x_{(\alpha)}\left(\alpha\,\mathbf{1}_{\{X \le x_{(\alpha+\epsilon)}\}}^{(\alpha+\epsilon)} - (\alpha+\epsilon)\,\mathbf{1}_{\{X \le x_{(\alpha)}\}}^{(\alpha)}\right)\right] \\ &= \quad \frac{x_{(\alpha)}}{\alpha(\alpha+\epsilon)}\left(\alpha\,\mathrm{E}\left[\mathbf{1}_{\{X \le x_{(\alpha+\epsilon)}\}}^{(\alpha+\epsilon)}\right] - (\alpha+\epsilon)\,\mathrm{E}\left[\mathbf{1}_{\{X \le x_{(\alpha)}\}}^{(\alpha)}\right]\right) \\ &= \quad \frac{x_{(\alpha)}}{\alpha(\alpha+\epsilon)}\left(\alpha\,(\alpha+\epsilon) - (\alpha+\epsilon)\,\alpha\right) \\ &= \quad 0\,. \end{aligned}$$

The inequality is due to the fact that by (3.5)

$$\alpha \, \mathbf{1}_{\{X \le x_{(\alpha+\epsilon)}\}}^{(\alpha+\epsilon)} - (\alpha + \epsilon) \, \mathbf{1}_{\{X \le x_{(\alpha)}\}}^{(\alpha)} \quad \begin{cases} \le \ 0 \,, & \text{if } X < x_{(\alpha)} \\ \ge \ 0 \,, & \text{if } X > x_{(\alpha)} \,. \end{cases} \qquad \square$$

4 Motivation for tail mean and Expected Shortfall

Assume that we want to estimate the lower α-quantile $x_{(\alpha)}$ of some random variable X. Let some sample $(X_1, \dots, X_n)$, drawn from independent copies of X, be given. Denote by $X_{1:n} \le \dots \le X_{n:n}$ the components of the ordered n-tuple $(X_1, \dots, X_n)$. Denote by $\lfloor x \rfloor$ the integer part of the number $x \in \mathbb{R}$, hence

$$\lfloor x \rfloor \quad = \quad \max\{n \in \mathbb{Z} : \ n \le x\} \,.$$

Then the order statistic $X_{\lfloor n\alpha \rfloor:n}$ appears as natural estimator for $x_{(\alpha)}$. Nevertheless, it is well known that in case of a non-unique quantile (i.e. $x_{(\alpha)} < x^{(\alpha)}$) the quantity $X_{\lfloor n\alpha \rfloor:n}$ does not converge to $x_{(\alpha)}$. This follows for instance from Theorem 1 in [8] which says that

$$1 \ = \ \mathrm{P}[X_{\lfloor n\alpha \rfloor:n} \le x_{(\alpha)} \text{ infinitely often}] \ = \ \mathrm{P}[X_{\lfloor n\alpha \rfloor:n} \ge x^{(\alpha)} \text{ infinitely often}] \,.$$

Surprisingly, we get a well-determined limit when we replace the single order statistic by an average over the left tail of the sample. Recall the definition (2.1) of an indicator function.

Proposition 4.1 *Let $\alpha \in (0,1)$ be fixed, X a real random variable with $\mathrm{E}[X^-] < \infty$ and $(X_1, X_2, \dots)$ an independent sequence of random variables with the same distribution as X. Then with probability 1*

(4.1)
$$\lim_{n \to \infty} \frac{\sum_{i=1}^{\lfloor n\alpha \rfloor} X_{i:n}}{\lfloor n\alpha \rfloor} \ = \ \bar{x}_{(\alpha)} \,.$$

If X is integrable, then the convergence in (4.1) holds in L_1, too.

Proof. Due to Proposition 3.2, the "with probability 1" part of Proposition 4.1 is essentially a special case of Theorem 3.1 in [18] with $0 = t_0 < \alpha = t_1 < t_2 = 1$, $J(t) = \mathbf{1}_{(0,\alpha]}(t)$, $J_N(t) = \mathbf{1}_{(0, \frac{\lfloor N\alpha \rfloor + 1}{N}]}(t)$, $g(t) = F^{-1}(t)$, and $p_1 = p_2 = \infty$. Concerning the L_1-convergence note that

$$\left| \sum_{i=1}^{\lfloor n\alpha \rfloor} X_{i:n} \right| \ \le \ \sum_{i=1}^{n} |X_i| \,.$$

By the strong law of large numbers $n^{-1} \sum_{i=1}^{n} |X_i|$ converges in L_1. This implies uniform integrability for $n^{-1} \sum_{i=1}^{n} |X_i|$ and for $n^{-1} |\sum_{i=1}^{\lfloor n\alpha \rfloor} X_{i:n}|$. Together with the already proven almost sure convergence this implies the assertion. $\square$

8

To see how a direct proof of the almost sure convergence in Proposition 4.1 would work consider the following heuristic computation. Observe first that

$$\frac{\sum_{i=1}^{\lfloor n\alpha \rfloor} X_{i:n}}{\lfloor n\alpha \rfloor} = \frac{1}{\lfloor n\alpha \rfloor}\left(\sum_{i=1}^{n} X_{i:n}\,\mathbf{1}_{\{X_{i:n}\leq X_{\lfloor n\alpha \rfloor:n}\}} + \sum_{i=1}^{n} X_{i:n}\left(\mathbf{1}_{\{1,\dots,\lfloor n\alpha \rfloor\}}(i) - \mathbf{1}_{\{X_{i:n}\leq X_{\lfloor n\alpha \rfloor:n}\}}\right)\right)$$

$$= \frac{1}{\lfloor n\alpha \rfloor}\left(\sum_{i=1}^{n} X_{i}\,\mathbf{1}_{\{X_{i}\leq X_{\lfloor n\alpha \rfloor:n}\}} + X_{\lfloor n\alpha \rfloor:n}\sum_{i=1}^{n}\left(\mathbf{1}_{\{1,\dots,\lfloor n\alpha \rfloor\}}(i) - \mathbf{1}_{\{X_{i:n}\leq X_{\lfloor n\alpha \rfloor:n}\}}\right)\right)$$

(4.2) $$= \frac{1}{\lfloor n\alpha \rfloor}\left(\sum_{i=1}^{n} X_{i}\,\mathbf{1}_{\{X_{i}\leq X_{\lfloor n\alpha \rfloor:n}\}} + X_{\lfloor n\alpha \rfloor:n}\left(\lfloor n\alpha \rfloor - \sum_{i=1}^{n}\mathbf{1}_{\{X_{i}\leq X_{\lfloor n\alpha \rfloor:n}\}}\right)\right).$$

If we now had

(4.3) $$\lim_{n\to\infty} X_{\lfloor n\alpha \rfloor:n} = x_{(\alpha)},$$

with probability 1, in connection with $\lim_{n\to\infty} n/\lfloor n\alpha \rfloor = 1/\alpha$ it would be plausible to obtain (4.1). Unfortunately (4.3) is not true in general, but only

(4.4) $$\liminf_{n\to\infty} X_{\lfloor n\alpha \rfloor:n} = x_{(\alpha)} \qquad \text{and} \qquad \limsup_{n\to\infty} X_{\lfloor n\alpha \rfloor:n} = x^{(\alpha)}.$$

Nevertheless the proof could be completed on the base of (4.2) by using (4.4) together with the Glivenko-Cantelli theorem and Corollary 4.3 below.

Proposition 4.1 validates the interpretation given to α–tail mean in [1] as *mean of the worst $100\alpha\%$ cases*. This concept, which seems very natural from an insurance or risk management point of view, has so far appeared in the literature by different kinds of *conditional expectation beyond* VaR which is a different concept for discrete distributions. "Tail Conditional expectation", "worst conditional expectation", "conditional value at risk" all bear also in their name the fact that they are conditional expected values of the random variable X (note that concerning CVaR, by Corollary 4.3 below this is a misinterpretation). For TCE_α, for instance, the natural estimator is not given by the one analyzed in (4.1) or its negative, but rather by

(4.5) $$-\frac{\sum_{i=1}^{n} X_{i}\,\mathbf{1}_{\{X_{i}\leq X_{\lfloor n\alpha \rfloor:n}\}}}{\sum_{i=1}^{n}\mathbf{1}_{\{X_{i}\leq X_{\lfloor n\alpha \rfloor:n}\}}}$$

which however has problems of convergence in case $x_{(\alpha)} < x^{(\alpha)}$.

This is the reason why we avoid the term "conditional" in our definition of α–tail mean. In fact, it is not very hard to see (cf. Example 5.4 below) that α–tail mean *does not* admit a general representation in terms of a conditional expectation of X given some event $A \in \sigma(X)$ (i.e. some event only depending on X). Hence it is not possible to give a definition of the type

(4.6) $$\bar{x}_{(\alpha)} = \mathrm{E}[X|A] \qquad \text{for some } A \in \sigma(X),$$

unless the event A is chosen in a σ-algebra $\mathcal{A} \supset \sigma(X)$ on an artificial new probability space (see Corollary 6.2 below).

In order to make visible the coincidence of CVaR and tail mean, the following proposition collects some facts on quantiles which are well-known in probability theory (cf. Exercise 3 in ch. 1 of [9] or Problem 25.9 of [10] for the here cited version):

Proposition 4.2 *Let X be a real integrable random variable on some probability space $(\Omega, \mathcal{A}, \mathrm{P})$. Fix $\alpha \in (0, 1)$ and define the function $H_\alpha : \mathbb{R} \to [0, \infty)$ by*

$$(4.7) \qquad H_\alpha(s) \quad = \quad \alpha\, \mathrm{E}[(X - s)^+] + (1 - \alpha)\, \mathrm{E}[(X - s)^-].$$

Then the function H_α is convex (and hence continuous) with $\lim_{|s| \to \infty} H_\alpha(s) = \infty$. The set M_α of minimizers to H_α is a compact interval, namely

$$
\begin{aligned}
M_\alpha \quad &= \quad [x_{(\alpha)},\, x^{(\alpha)}] \\
(4.8) \qquad &= \quad \{s \in \mathbb{R} :\ \mathrm{P}[X < s] \le \alpha \le \mathrm{P}[X \le s]\}. \qquad\qquad\qquad \square
\end{aligned}
$$

Note the following equivalent representations for H_α:

$$(4.9) \qquad H_\alpha(s) \quad = \quad \alpha\, \mathrm{E}[X] + \alpha \left(\frac{\mathrm{E}[(X - s)^-]}{\alpha} - s \right)$$

$$(4.10) \qquad\qquad\quad = \quad \alpha\, \mathrm{E}[X] - \alpha \left(\frac{\mathrm{E}[X\, \mathbf{1}_{\{X \le s\}}]}{\alpha} + s\, \frac{\alpha - \mathrm{P}[X \le s]}{\alpha} \right).$$

From Definitions 2.5 and 2.6 for CVaR and ES, respectively, and by (4.10), in connection with Proposition 4.2, we obtain the following corollary to the proposition.

Corollary 4.3 *Let X be a real integrable random variable on some probability space $(\Omega, \mathcal{A}, \mathrm{P})$ and $\alpha \in (0, 1)$ be fixed. Then*

$$
\begin{aligned}
\mathrm{ES}_\alpha(X) \quad &= \quad \mathrm{CVaR}^\alpha(X) \\
(4.11) \qquad &= \quad -\alpha^{-1} \left(\mathrm{E}[X\, \mathbf{1}_{\{X \le s\}}] + s\, (\alpha - \mathrm{P}[X \le s]) \right), \quad s \in [x_{(\alpha)},\, x^{(\alpha)}]. \qquad \square
\end{aligned}
$$

A further representation of ES or CVaR, respectively, as expectation of a suitably modified tail distribution is given in the recent research report [14] (cf. Def. 3 therein).

In Definitions 2.5 and 2.6 only $\mathrm{E}[X^-] < \infty$ is required for X. Indeed, this integrability condition would suffice to guarantee (4.11). We formulated Corollary 4.3 with full integrability of X because we wanted to rely on Proposition 4.2 for the proof.

Note that by a simple calculation one can show that (4.11) is equivalent to

$$(4.12) \qquad \mathrm{ES}_\alpha(X) \quad = \quad -\alpha^{-1} \left(\mathrm{E}[X\, \mathbf{1}_{\{X < s\}}] + s\, (\alpha - \mathrm{P}[X < s]) \right), \quad s \in [x_{(\alpha)},\, x^{(\alpha)}].$$

By (4.12) we see that ES coincides with the coherent risk measure considered in Example 4 of [6] (already mentioned in Example 4.2 of [5]).

5 Inequalities and counter-examples

In this section we compare the Expected Shortfall with the risk measures TCE and WCE defined in section 2. Moreover, we present an example showing that VaR and TCE are not sub-additive in general. By the same example we show that there is not a clear relationship between WCE and lower TCE. We start with a result in the spirit of the Neyman-Pearson lemma.

Proposition 5.1 *Let $\alpha \in (0,1)$ be fixed and X be a real-valued random variable on some probability space $(\Omega, \mathcal{A}, \mathrm{P})$. Suppose that there is some function $f : \mathbb{R} \to \mathbb{R}$ such that $\mathrm{E}[(f \circ X)^-] < \infty$, $f(x) \leq f(x_{(\alpha)})$ for $x < x_{(\alpha)}$, and $f(x) \geq f(x_{(\alpha)})$ for $x > x_{(\alpha)}$. Let $A \in \mathcal{A}$ be an event with $\mathrm{P}[A] \geq \alpha$ and $\mathrm{E}[\,|f \circ X|\,\mathbf{1}_A] < \infty$. Then*

(i) $\mathrm{TM}_\alpha(f \circ X) \ \leq \ \mathrm{E}[f \circ X \,|\, A]\,,$

(ii) $\mathrm{TM}_\alpha(f \circ X) \ = \ \mathrm{E}[f \circ X \,|\, A] \quad$ *if* $\mathrm{P}[A \cap \{X > x_{(\alpha)}\}] = 0$ *and*

$$(5.1) \qquad\qquad \mathrm{P}[X < x_{(\alpha)}] = 0 \quad or$$

$$(5.2) \qquad\qquad \mathrm{P}[X < x_{(\alpha)}] > 0, \ \ \mathrm{P}[\Omega \backslash A \cap \{X < x_{(\alpha)}\}] = 0, \ \ and \quad \mathrm{P}[A] = \alpha\,,$$

(iii) *if* $f(x) < f(x_{(\alpha)})$ *for* $x < x_{(\alpha)}$ *and* $f(x) > f(x_{(\alpha)})$ *for* $x > x_{(\alpha)}$, *then* $\mathrm{TM}_\alpha(f \circ X) = \mathrm{E}[f \circ X \,|\, A]$ *implies* $\mathrm{P}[A \cap \{X > x_{(\alpha)}\}] = 0$ *and either (5.1) or (5.2).*

Proof. Note that by assumption

$$\{X \leq x_{(\alpha)}\} \subset \{f \circ X \leq f(x_{(\alpha)})\} \quad \text{and} \quad \{X < x_{(\alpha)}\} \supset \{f \circ X < f(x_{(\alpha)})\}\,.$$

Hence we see from (4.8) that

$$\mathrm{P}[f \circ X \leq f(x_{(\alpha)})] \geq \alpha \quad \text{and} \quad \mathrm{P}[f \circ X < f(x_{(\alpha)})] \leq \alpha$$

and therefore

$$(5.3) \qquad\qquad q_\alpha(f \circ X) \ \leq \ f(x_{(\alpha)}) \ \leq \ q^\alpha(f \circ X)\,.$$

Moreover, the assumption implies

$$(5.4) \qquad\qquad \{f \circ X \leq f(x_{(\alpha)})\} \backslash \{X \leq x_{(\alpha)}\} \quad \subset \quad \{f \circ X = f(x_{(\alpha)})\}\,.$$

11

By Corollary 4.3, (5.3), (5.4), (3.6), and (3.7), we can calculate similarly to the proof of Proposition 3.4

$$
\begin{aligned}
\mathrm{E}[f \circ X | A] - \mathrm{TM}_\alpha(f \circ X) &= \mathrm{E}\left[f \circ X \left(P[A]^{-1} \mathbf{1}_A - \alpha^{-1} \mathbf{1}^{(\alpha)}_{\{f \circ X \leq f(x_{(\alpha)})\}} \right) \right] \\
&= \mathrm{E}\left[f \circ X \left(P[A]^{-1} \mathbf{1}_A - \alpha^{-1} \mathbf{1}^{(\alpha)}_{\{X \leq x_{(\alpha)}\}} \right) \right] \\
&= (\alpha\, P[A])^{-1} \left(f(x_{(\alpha)}) \mathrm{E}\left[\alpha\, \mathbf{1}_A - P[A]\, \mathbf{1}^{(\alpha)}_{\{X \leq x_{(\alpha)}\}} \right] \right. \\
&\qquad + \mathrm{E}\left[(f \circ X - f(x_{(\alpha)})) \left(\alpha\, \mathbf{1}_A - P[A]\, \mathbf{1}^{(\alpha)}_{\{X \leq x_{(\alpha)}\}} \right) \right] \Bigg) \\
&= (\alpha\, P[A])^{-1} \mathrm{E}\left[(f \circ X - f(x_{(\alpha)})) \left(\alpha\, \mathbf{1}_A - P[A]\, \mathbf{1}^{(\alpha)}_{\{X \leq x_{(\alpha)}\}} \right) \right] \\
&\geq 0.
\end{aligned}
$$

(5.5)

Here, we obtain inequality (5.5) from the assumption on f since

(5.6)
$$
\alpha\, \mathbf{1}_A - P[A]\, \mathbf{1}^{(\alpha)}_{\{X \leq x_{(\alpha)}\}} \begin{cases} \leq 0, & \text{if } X < x_{(\alpha)} \\ \geq 0, & \text{if } X > x_{(\alpha)}. \end{cases}
$$

This proves (i). The sufficiency and necessity respectively of the conditions in (ii) and (iii) for equality in (5.5) are easily obtained by careful inspection of (5.5). □

Note that the condition

(5.7)
$$
P[A \cap \{X > x_{(\alpha)}\}] = 0 \quad \text{and} \quad P[\Omega \backslash A \cap \{X < x_{(\alpha)}\}] = 0,
$$

appearing in (ii) and (iii) of Proposition 5.1, means up to set differences of probability 0 that

(5.8)
$$
\{X < x_{(\alpha)}\} \quad \subset \quad A \quad \subset \quad \{X \leq x_{(\alpha)}\}.
$$

In particular, (5.7) is implied by (5.8).

The proof of Proposition 5.1 is the hardest work in this section. Equipped with its result we are in a position to derive without effort a couple of conclusions pointing out the relations between TCE, WCE and ES. Recall $\mathrm{ES}_\alpha = -\mathrm{TM}_\alpha$.

Corollary 5.2 *Let $\alpha \in (0, 1)$ and X a real-valued random variable on some probability space $(\Omega, \mathcal{A}, P)$ with $\mathrm{E}[X^-] < \infty$. Then*

(5.9)
$$
\mathrm{TCE}^\alpha(X) \quad \leq \quad \mathrm{TCE}_\alpha(X) \quad \leq \quad \mathrm{ES}_\alpha(X), \quad \text{and}
$$

(5.10)
$$
\mathrm{TCE}^\alpha(X) \quad \leq \quad \mathrm{WCE}_\alpha(X) \quad \leq \quad \mathrm{ES}_\alpha(X).
$$

Proof. The first inequality in (5.9) is obvious (formally it follows from Lemma 5.1 in [16]). The second follows from Proposition 5.1 (i) by setting $f(x) = x$, $A = \{X \leq x_{(\alpha)}\}$, and observing

12

$P[X \leq x_{(\alpha)}] \geq \alpha$. The first inequality in (5.10) was proven in Proposition 5.1 of [3]. The second follows again from Proposition 5.1 (i) since all the events in the definition of WCE have probabilities $> \alpha$. $\qquad\square$

The following corollary to Proposition 5.1 presents in particular in (i) a first sufficient condition for WCE and ES to coincide, namely continuity of the distribution of X.

Corollary 5.3 *Let α and X be as in Corollary 5.2. Then*

(i) $P[X \leq x^{(\alpha)}] = \alpha$, $P[X < x_{(\alpha)}] > 0$ *or* $P[X \leq x^{(\alpha)}, X \neq x_{(\alpha)}] = 0$ *if and only if*

$$(5.11) \qquad \mathrm{ES}_\alpha(X) \;=\; \mathrm{WCE}_\alpha(X) \;=\; \mathrm{TCE}_\alpha(X) \;=\; \mathrm{TCE}^\alpha(X)\,.$$

In particular, (5.11) holds if the distribution of X is continuous, i.e. $P[X = x] = 0$ for all $x \in \mathbb{R}$.

(ii) $P[X \leq x_{(\alpha)}] = \alpha$ *or* $P[X < x_{(\alpha)}] = 0$ *if and only if* $\mathrm{ES}_\alpha(X) \;=\; \mathrm{TCE}_\alpha(X)$.

Proof. Concerning (i) apply Proposition 5.1 (ii) and (iii) with $A = \{X \leq x^{(\alpha)}\}$ and Corollary 5.2. In order to obtain (ii) apply Proposition 5.1 (ii) and (iii) with $A = \{X \leq x_{(\alpha)}\}$. $\qquad\square$

Corollary 5.2 leaves open the relation between $\mathrm{TCE}_\alpha(X)$ and $\mathrm{WCE}_\alpha(X)$. The implication

$$(5.12) \qquad P[X \leq x_{(\alpha)}] \;>\; \alpha \quad \Rightarrow \quad \mathrm{TCE}_\alpha(X) \leq \mathrm{WCE}_\alpha(X)$$

is obvious. Corollary 5.3 (ii) shows that

$$(5.13) \qquad P[X \leq x_{(\alpha)}] \;=\; \alpha \quad \Rightarrow \quad \mathrm{TCE}_\alpha(X) \geq \mathrm{WCE}_\alpha(X)\,.$$

The following example shows that all the inequalities between TCE, WCE, and ES in (5.9), (5.10), (5.12), and (5.13) can be strict. Moreover, it shows that none of the quantities $-q_\alpha$, VaR^α, TCE_α, or TCE^α defines a sub-additive risk measure in general.

Example 5.4
Consider the probability space $(\Omega, \mathcal{A}, P)$ with $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\mathcal{A}$ the set of all subsets of Ω and P specified by

$$P[\{\omega_1\}] = P[\{\omega_2\}] = p, \quad P[\{\omega_3\}] = 1 - 2\,p\,,$$

and choose $0 < p < \frac{1}{3}$. Fix some positive number N and let X_i, $i = 1, 2$, be two random variables defined on $(\Omega, \mathcal{A}, P)$ with values

$$X_i(\omega_j) \quad = \quad \begin{cases} -N, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases}$$

Choose α such that $0 < \alpha < 2\,p$. Then it is straightforward to obtain Table 1 with the values of the risk measures interesting to us.

Risk Measure	$p < \alpha < 2p$		$p = \alpha$		$p > \alpha$	
	$X_{1,2}$	$X_1 + X_2$	$X_{1,2}$	$X_1 + X_2$	$X_{1,2}$	$X_1 + X_2$
$-q_\alpha$	0	N	N	N	N	N
VaR^α	0	N	0	N	N	N
TCE^α	Np	N	Np	N	N	N
TCE_α	Np	N	N	N	N	N
WCE_α	$N/2$	N	$N/2$	N	N	N
ES_α	Np/α	N	N	N	N	N

Table 1: Values of risk measures for Example 5.4.

In case $p < \alpha < 2\,p$ we see from Table 1 that

$$
\begin{aligned}
-q_\alpha(X_1) - q_\alpha(X_2) &< -q_\alpha(X_1 + X_2) \\
\mathrm{VaR}^\alpha(X_1) + \mathrm{VaR}^\alpha(X_2) &< \mathrm{VaR}^\alpha(X_1 + X_2) \\
\mathrm{TCE}_\alpha(X_1) + \mathrm{TCE}_\alpha(X_2) &< \mathrm{TCE}_\alpha(X_1 + X_2) \\
\mathrm{TCE}^\alpha(X_1) + \mathrm{TCE}^\alpha(X_2) &< \mathrm{TCE}^\alpha(X_1 + X_2).
\end{aligned}
$$

These inequalities show that none of the notions $-q_\alpha$, VaR^α, TCE_α, or TCE^α can be used to define a sub-additive risk measure. In case $p < \alpha < 2\,p$ we have also

$$
\begin{aligned}
\mathrm{TCE}_\alpha(X_1) &< \mathrm{ES}_\alpha(X_1) \\
\mathrm{TCE}^\alpha(X_1) = \mathrm{TCE}_\alpha(X_1) &< \mathrm{WCE}_\alpha(X_1) \\
\mathrm{WCE}_\alpha(X_1) &< \mathrm{ES}_\alpha(X_1).
\end{aligned}
$$

(5.14)

Hence the second inequalities in (5.9), (5.10), and (5.12) may be strict, as can be the first inequality in (5.10). In case $p = \alpha$ we have from Table 1 that

$$
\begin{aligned}
\mathrm{TCE}^\alpha(X_1) &< \mathrm{TCE}_\alpha(X_1) \quad \text{and} \\
\mathrm{TCE}_\alpha(X_1) &> \mathrm{WCE}_\alpha(X_1).
\end{aligned}
$$

Thus, also the first inequality in (5.9) and the inequality in (5.13) can be strict. In particular, we see that there is not any clear relationship between TCE_α and WCE. Beside the inequalities, from the comparison with the results in the region $p > \alpha$, we get an example for the fact that all the measures but ES may have discontinuities in α. Moreover, in case $p < \alpha$ we have a stronger version of (5.14), namely

$$
-\inf\{\mathrm{E}[X_1 \,|\, A] : \ A \in \mathcal{A}, \mathrm{P}[A] \geq \alpha\} \quad < \quad \mathrm{ES}_\alpha(X_1),
$$

which shows that even if one replaces ">" by "$\geq$" in Definition 2.4, strict inequality may appear in the relation between WCE and ES. □

We finally observe that Example 5.4 is not so academic as it may seem at first glance since the X_i's may be figured out as two risky bonds of nominal N with non–overlapping default states ω_i of probability p.

6 Representing ES in terms of WCE

By Example 5.4 we know that WCE and ES may differ in general. Nevertheless, we are going to show in the last part of the paper that this phenomenon can only occur when the underlying probability space is too "small" in the sense of not allowing a suitable representation of the random variable under consideration as function of a continuous random variable. Moreover, as long as only finitely many random variables are under consideration it is always possible to switch to a "larger" probability space in order to make WCE and ES coincide. Finally, we state a general representation of ES in terms of related WCEs.

Proposition 6.1 *Let X and Y be a real-valued random variables on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ such that $\mathrm{E}[Y^-] < \infty$. Fix some $\alpha \in (0,1)$. Assume that Y is given by $Y = f \circ X$ where f satisfies $f(x) \leq f(x_{(\alpha)})$ for $x < x_{(\alpha)}$, and $f(x) \geq f(x_{(\alpha)})$ for $x > x_{(\alpha)}$.*

(i) If $\mathrm{P}[X \leq x_{(\alpha)}] = \alpha$ then

$$\mathrm{ES}_\alpha(Y) \quad = \quad - \inf_{A \in \mathcal{A}, \, \mathrm{P}[A] \geq \alpha} \mathrm{E}[Y \mid A] \,.$$

(ii) If the distribution function of X is continuous then also

$$\mathrm{ES}_\alpha(Y) \quad = \quad \mathrm{WCE}_\alpha(Y) \,.$$

Proof. Concerning (i), by Proposition 5.1 (i) we only have to show

$$(6.1) \qquad\qquad\qquad \mathrm{TM}_\alpha(Y) \quad = \quad \mathrm{E}[Y \mid X \leq x_{(\alpha)}] \,.$$

With the choice $A = \{X \leq x_{(\alpha)}\}$ this follows from Proposition 5.1 (ii).

Concerning Proposition 6.1 (ii), by (6.1), we have to show that there is a sequence $(A_n)_{n \in \mathbb{N}}$ in $\mathcal{A}$ with $\mathrm{P}[A_n] > \alpha$ for all $n \in \mathbb{N}$ such that

$$\lim_{n \to \infty} \mathrm{E}[Y \mid A_n] = \mathrm{E}[Y \mid X \leq x_{(\alpha)}] \,.$$

By continuity of the distribution of X and integrability of Y^- we obtain such a sequence with the definition $A_n = \{X \leq x^{(\alpha)} + 1/n\}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\Box$

Corollary 6.2 *Let $(X_1, \ldots, X_d)$ be an $\mathbb{R}^d$-valued random vector on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ such that $\mathrm{E}[X_i^-] < \infty, i = 1, \ldots, d$. Fix $\alpha \in (0,1)$. Then there is a random vector $(X_1', \ldots, X_d')$ on some probability space $(\Omega', \mathcal{A}', \mathrm{P}')$ with the following two properties:*

(i) The distributions of $(X_1, \ldots, X_d)$ and $(X_1', \ldots, X_d')$ are equal, i.e.

$$\mathrm{P}[X_1 \leq x_1, \ldots, X_d \leq x_d] \quad = \quad \mathrm{P}'[X_1' \leq x_1, \ldots, X_d' \leq x_d] \quad \text{for all } (x_1, \ldots, x_d) \in \mathbb{R} \,.$$

15

(ii) Worst conditional expectation and Expected Shortfall coincide for all $i = 1, \ldots, d$, i.e.

$$\mathrm{WCE}_\alpha(X_i') \; = \; \mathrm{ES}_\alpha(X_i') \,, \;\; i = 1, \ldots, d \,. \hspace{3cm} \Box$$

Proof. By Sklar's theorem (cf. Theorem 2.10.9 in [12]) we get the existence of a random vector $(U_1, \ldots, U_d)$ where each U_i is uniformly distributed on $(0, 1)$ such that (i) holds with $X_i' = q_{U_i}(X_i), i = 1, \ldots, d$. Since q_α is non-decreasing in α the assertion now follows from Proposition 6.1. $\hspace{4cm} \Box$

Corollary 6.2 yields another proof for the sub-additivity of Expected Shortfall: in order to prove $\mathrm{ES}_\alpha(X) + \mathrm{ES}_\alpha(Y) \geq \mathrm{ES}_\alpha(X + Y)$ apply the corollary to the underlying random vector $(X, Y, X + Y)$.

As a final consequence of Corollary 5.2 and Corollary 6.2 we note:

Corollary 6.3 *Let X be a real-valued random variable on some probability space $(\Omega, \mathcal{A}, \mathrm{P})$ with $\mathrm{E}[X^-] < \infty$. Fix $\alpha \in (0, 1)$. Then*

$$\mathrm{ES}_\alpha(X) \;\; = \;\; \max \Big\{ \mathrm{WCE}_\alpha(X') : \; X' \text{ random variable on } (\Omega', \mathcal{A}', \mathrm{P}') \text{ with}$$

$$\mathrm{P}'[X' \leq x] = \mathrm{P}[X \leq x] \text{ for all } x \in \mathbb{R} \Big\},$$

where the maximum is taken over all random variables X' on probability spaces $(\Omega', \mathcal{A}', \mathrm{P}')$ such that the distributions of X and X' are equal. $\hspace{4cm} \Box$

Corollary 6.3 shows that Expected Shortfall in the sense of Definition 2.6 may be considered a robust version of worst conditional expectation (Definition 2.4), making the latter insensitive to the underlying probability space.

7 Conclusion

In the paper at hand we have shown that simply taking a conditional expectation of losses beyond VaR can fail to yield a coherent risk measure when there are discontinuities in the loss distributions. Already existing definitions for some kind of expected shortfall, redressing this drawback, as those in [3] or [13], did not provide representations suitable for efficient computation and estimation in the general case. We have clarified the relations between these definitions and the explicit one from [1], thereby pointing out that it is the definition which is most appropriate for practical purposes.

A   Appendix: Subadditivity of Expected Shortfall

We give here for the sake of completeness the proof of subadditivity for expected shortfall which was originally given in Appendix A of [1].

For the proof it is convenient to adopt the representation of eq. (3.7) for the Tail Mean and write the Expected Shortfall as

$$(A.1) \qquad \mathrm{ES}_\alpha(Y) \;=\; -\frac{1}{\alpha}\,\mathrm{E}[X\,\mathbf{1}^{(\alpha)}_{\{X\leq x_{(\alpha)}\}}]$$

with the function $\mathbf{1}^{(\alpha)}_{\{X\leq s\}}$ defined in (3.4).

Proposition A.1 (Subadditivity of Expected Shortfall) *Given two random variables X and Y with $\mathrm{E}[X^-]<\infty$ and $\mathrm{E}[Y^-]<\infty$ the following inequality holds:*

$$(A.2) \qquad \mathrm{ES}_\alpha(X+Y) \;\leq\; \mathrm{ES}_\alpha(X) + \mathrm{ES}_\alpha(Y)$$

for any $\alpha \in (0,1]$

Proof. Defining $Z = X+Y$, we obtain by virtue of (3.6)

$$(A.3) \qquad \alpha\left(\mathrm{ES}_\alpha(X)+\mathrm{ES}_\alpha(Y)-\mathrm{ES}_\alpha(Z)\right) \;=$$

$$= \mathrm{E}\left[Z\,\mathbf{1}^{(\alpha)}_{\{Z\leq z_{(\alpha)}\}} - X\,\mathbf{1}^{(\alpha)}_{\{X\leq x_{(\alpha)}\}} - Y\,\mathbf{1}^{(\alpha)}_{\{Y\leq y_{(\alpha)}\}}\right]$$

$$= \mathrm{E}\left[X\left(\mathbf{1}^{(\alpha)}_{\{Z\leq z_{(\alpha)}\}} - \mathbf{1}^{(\alpha)}_{\{X\leq x_{(\alpha)}\}}\right) + Y\left(\mathbf{1}^{(\alpha)}_{\{Z\leq z_{(\alpha)}\}} - \mathbf{1}^{(\alpha)}_{\{Y\leq y_{(\alpha)}\}}\right)\right]$$

$$\geq x_{(\alpha)}\,\mathrm{E}\left[\mathbf{1}^{(\alpha)}_{\{Z\leq z_{(\alpha)}\}} - \mathbf{1}^{(\alpha)}_{\{X\leq x_{(\alpha)}\}}\right] + y_{(\alpha)}\,\mathrm{E}\left[\mathbf{1}^{(\alpha)}_{\{Z\leq z_{(\alpha)}\}} - \mathbf{1}^{(\alpha)}_{\{Y\leq y_{(\alpha)}\}}\right]$$

$$= x_{(\alpha)}\,(\alpha-\alpha) + y_{(\alpha)}\,(\alpha-\alpha) = 0$$

which proves the thesis. In the inequality above we used the fact that

$$(A.4) \qquad \begin{cases} \mathbf{1}^{(\alpha)}_{\{Z\leq z_{(\alpha)}\}} - \mathbf{1}^{(\alpha)}_{\{X\leq x_{(\alpha)}\}} \geq 0 & \text{if} \quad X > x_{(\alpha)} \\[2mm] \mathbf{1}^{(\alpha)}_{\{Z\leq z_{(\alpha)}\}} - \mathbf{1}^{(\alpha)}_{\{X\leq x_{(\alpha)}\}} \leq 0 & \text{if} \quad X < x_{(\alpha)} \end{cases}$$

which in turn is a consequence of (3.4) and (3.5) $\qquad\qquad\qquad\qquad\qquad\qquad \square$

References

[1] ACERBI, C., NORDIO, C., SIRTORI, C. (2001) *Expected Shortfall as a Tool for Financial Risk Management.* Working paper. `http://www.gloriamundi.org/var/wps.html`

[2] ARTZNER, P., DELBAEN, F., EBER, J.-M., HEATH, D. (1997) *Thinking coherently.* RISK **10**(11).

[3] ARTZNER, P., DELBAEN, F., EBER, J.-M., HEATH, D. (1999) *Coherent measures of risk.* Math. Fin. **9**(3), 203–228.

[4] BERTSIMAS, D., LAUPRETE, G.J., SAMAROV, A. (2000) *Shortfall as a risk measure: properties, optimization and applications.* Working paper, Sloan School of Management, MIT, Cambridge.

[5] DELBAEN, F. (1998) *Coherent risk measures on general probability spaces.* Working paper, ETH Zürich. `http://www.math.ethz.ch/∼delbaen/`

[6] DELBAEN, F. (2001) *Coherent risk measures.* Lecture Notes, Scuola Normale Superiore di Pisa.

[7] EMBRECHTS, P. (2000) *Extreme Value Theory: Potential and Limitations as an Integrated Risk Management Tool.* Working paper, ETH Zürich.
`http://www.math.ethz.ch/∼embrechts/`

[8] FELDMAN, D., TUCKER, H.G. (1966) *Estimation of non-unique quantiles.* Ann. Math. Stat. **37**, 451–457.

[9] FERGUSON, T.G. (1967) Mathematical Statistics: a Decision Theoretic Approach. *Academic Press: London.*

[10] HINDERER, K. (1972) Grundbegriffe der Wahrscheinlichkeitstheorie (Fundamental terms of probability theory). *Springer: Berlin.*

[11] KUSUOKA, S. (2001) On law invariant coherent risk measures. In *Advances in Mathematical Economics*, volume 3, pages 83–95. *Springer: Tokyo.*

[12] NELSEN, R.B. (1999) An introduction to copulas. *Springer: New York.*

[13] PFLUG, G. (2000) Some remarks on the value-at-risk and the conditional value-at-risk. In, Uryasev, S. (Editor). 2000. Probabilistic Constrained Optimization: Methodology and Applications. *Kluwer Academic Publishers.* `http://www.gloriamundi.org/var/pub.html`

[14] ROCKAFELLAR, R.T., URYASEV, S. (2001) *Conditional value-at-risk for general loss distributions.* Research report 2001-5, ISE Dept., University of Florida.
`http://www.ise.ufl.edu/uryasev/pubs.html`

[15] ROOTZÉN, H., KLÜPPELBERG, C. (1999) *A single number can't hedge against economic catastrophes.* Ambio **28**(6), 550–555. Royal Swedish Academy of Sciences.

[16] TASCHE, D. (2000) *Conditional expectation as quantile derivative.* Working paper, TU München. `http://www.ma.tum.de/stat/`

[17] URYASEV, S. (2000) *Conditional Value-at-Risk: Optimization Algorithms and Applications.* Financial Engineering News **2** (3).
`http://www.gloriamundi.org/var/pub.html`

[18] VAN ZWET, W.R. (1980) *A strong law for linear functions of order statistics.* Ann. Probab. **8**, 986–990.