Seven Proofs for the Subadditivity of Expected Shortfall

Paul Embrechts* and Ruodu Wang†

October 7, 2015

Abstract

Subadditivity is the key property which distinguishes the popular risk measures Value-at-Risk and Expected Shortfall (ES). In this paper we offer seven proofs of the subadditivity of ES, some found in the literature and some not. One of the main objectives of this paper is to provide a general guideline for instructors to teach the subadditivity of ES in a course. We discuss the merits and suggest appropriate contexts for each proof. With different proofs, different important properties of ES are revealed, such as its dual representation, optimization properties, continuity, consistency with convex order, and natural estimators.

Keywords: Expected Shortfall, TVaR, subadditivity, comonotonicity, Value-at-Risk, risk management, education.

MSC2000 subject classification: Primary: 28A25; secondary: 60E15, 91B06

1 Introduction

In a course on Quantitative Risk Management, an instructor inevitably has to discuss Value-at-Risk (VaR) and Expected Shortfall (ES) as the two standard risk measures to determine capital requirements for a financial institution. The reader is referred to Embrechts et al. (2014) for recent extensive debates on "*VaR versus ES in banking regulation*" as well as McNeil et al. (2015) for broader background material.

*RiskLab, Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland. Email: embrechts@math.ethz.ch.

†Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada. Email: wang@uwaterloo.ca.

Following Artzner et al. (1999), the key property that distinguishes any coherent risk measure, such as ES, from VaR is *subadditivity*. More precisely, a subadditive risk measure ρ satisfies that for any two risks X and Y, $\rho(X+Y) \le \rho(X)+\rho(Y)$ always holds. This property is closely related to questions on portfolio diversification and risk aggregation; more detailed theory and financial interpretation of subadditivity can be found in Delbaen (2012)[1]. Note that here, and throughout the paper, losses are accounted for as positive values.

Despite its relevance, it is somewhat surprising that many academics and risk professionals do not know explicitly how to prove that ES is subadditive, although they are all aware of the validity of the statement. In most of the main-stream textbooks used in actuarial science, quantitative finance, or quantitative risk management, a proof of this property is either (i) split into several disconnected parts, (ii) reliant on advanced results in modern probability or statistics, (iii) too mathematically involved for a typically broad class of students attracted to a course in the above fields, or (iv) even skipped.

Indeed, we shall see that the subadditivity of ES is not a trivial property; it relates to the dependence structure between random variables. Some mathematical proofs found in the literature can be quite involved. In view of the growing importance of ES for regulation (see recent regulatory documents BCBS (2012, 2013, 2014) and IAIS (2014)), it is clear to the authors that concise proofs of this property should be clearly conveyed to academics and practitioners in the quantitative fields of finance and risk management. Moreover, different proofs reveal different properties of ES, each with their own specific relevance for practice.

We first introduce some basic notation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an atomless probability space[2] Throughout, all random variables are defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and all probability measures are defined on $(\Omega, \mathcal{F})$. Let L^0 be the set of all random variables, L^1 the set of all integrable random variables and L^∞ the set of all (essentially) bounded random variables; for a definition of essential supremum, see for instance Billingsley (1995, p.241).

For $p \in (0,1)$, the two risk measures $\mathrm{VaR}_p : L^0 \to \mathbb{R}$ and $\mathrm{ES}_p : L^0 \to \mathbb{R} \cup \{+\infty\}$ are defined as

$$\mathrm{VaR}_p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \le x) \ge p\}, \qquad X \in L^0,$$

[1]Subadditivity as a desirable property of risk measures is also sometimes contested; see for instance Dhaene et al. (2008) and Cont et al. (2010).

[2]A probability space is atomless if there exists a U[0, 1]-distributed random variable in this space. The desired result in this paper holds also in a discrete probability space since one can naturally extend a discrete probability space to an atomless one.

and

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

The main objective is to provide a variety of proofs of the following theorem.

Theorem 1.1. *For $p \in (0,1)$, ES_p is subadditive on L^0. That is,*

$$\mathrm{ES}_p(X+Y) \le \mathrm{ES}_p(X) + \mathrm{ES}_p(Y) \tag{1.1}$$

for all $X, Y \in L^0$ and $p \in (0,1)$.

Remark 1.1. Of course, in practice ES_p is only used on L^1, but (1.1) is trivially true if the right-hand side is infinite; this is reflected in the $\mathbb{R} \cup \{+\infty\}$ above in the definition of ES_p.

Remark 1.2. Throughout the literature there is no unanimity when it comes to definitions and notation for specific risk measures. We already stressed that losses in our case always correspond to positive values. Also, ES is known under different names, occasionally with slight differences in the definitions, including T(ail)VaR, C(onditional)VaR, A(verage)VaR and CTE(Conditional Tail Expectation). Acerbi and Tasche (2002) contains a study on the equivalence of the above concepts with slightly different definitions.

In Section 2 we discuss some general issues and common basic lemmas related to the proofs of Theorem 1.1 and in Section 3 we present seven proofs based on different techniques. Each proof is self-contained, and when necessary, we refer to classic results in a respective field of study. Although most of the intermediate results are known in the literature, we give an elementary proof wherever possible so that our proofs can be directly used by an instructor.

With different proofs, we reveal different important properties of ES such as its dual representation, optimization properties, continuity, consistency with convex order, and natural estimators. We comment on merits of the proofs, and suggest appropriate contexts within which to use them.

The class of spectral risk measures in Acerbi and Tasche (2002) can be written as a continuously weighted average of ES; see Wang et al. (1997) and Kusuoka (2001). Therefore, by showing the subadditivity of ES, one directly obtains the subadditivity of any spectral risk measures.

Remark 1.3. The question "who was the first to show that ES is subadditive?" has no definite answer, since the introduction of ES came long after the theory of *Choquet integrals* (including

ES as a special case) was established. An implicit result close to the subadditivity of ES in a discrete probability space is Proposition 10.2.5 of Huber (1980), based on a lemma dating back to Choquet (1953). The subadditivity theorem of Choquet integrals in a general probability space is given in Chapter 6 of Denneberg (1994).

Throughout, denote $(x)_+ = \max\{x, 0\}$ for $x \in \mathbb{R}$. $\mathbb{N}$ is the set of positive integers. For a random variable X, we denote by F_X the distribution function or simply the distribution of X (under $\mathbb{P}$). Note that for $p \in (0,1)$, $F_X^{-1}(p) = \inf\{x \in \mathbb{R} : F_X(x) \geq p\} = \mathrm{VaR}_p(X)$; both the notation $\mathrm{VaR}_p(X)$ and the notation $F_X^{-1}(p)$ will be used whenever convenient; for detailed properties of the latter, see Embrechts and Hofert (2013). All expectations ("$\mathbb{E}$") are considered under $\mathbb{P}$ unless a superscript indicating another probability measure ("$\mathbb{E}^Q$") is present.

2 General discussion

2.1 Basic properties and lemmas

In this section, we list some basic properties and short lemmas on random variables and ES, which will be used across different proofs in Section 3. All the properties in this section naturally appear in a quantitative course which covers VaR and ES, and hence they create no extra burden in the teaching of such a course.

A risk measure $\rho : L^0 \to \mathbb{R} \cup \{+\infty\}$ is *law-determined* (or *law-invariant*) if $\rho(X) = \rho(Y)$ for any two identically distributed random variables $X, Y \in L^0$; ρ is *monotone* if $\rho(X) \leq \rho(Y)$ for $X, Y \in L^0$, $X \leq Y$ almost surely; ρ is *translation-invariant* if $\rho(X + c) = \rho(X) + c$ for all $X \in L^0$ and $c \in \mathbb{R}$. For $p \in (0,1)$, it is straightforward to check that both the risk measures VaR_p and ES_p are law-determined, monotone and translation-invariant. These properties will be frequently used throughout the paper. The lemma below yields the foundation of many results in probability theory, such as Sklar's theorem in the study of copulas.

Lemma 2.1. *For any random variable X, there exists a* $\mathrm{U}[0,1]$ *random variable U_X such that $X = F_X^{-1}(U_X)$ almost surely.*

Proof. This is a classic result; see Rüschendorf (2013, Proposition 1.3) where the construction is referred to as *distributional transform*. We give the construction of U_X below. If F_X is continuous, taking $U_X = F_X(X)$ would suffice. Generally, let V be a $\mathrm{U}[0,1]$ random variable independent of X, and write

$$U_X = F_X(X-) + V(F_X(X) - F_X(X-)),$$

where $F_X(x-)$ denotes the left-limit of the function F_X at $x \in \mathbb{R}$. The interested reader can check that U_X is U$[0,1]$-distributed and $X = F_X^{-1}(U_X)$ almost surely. $\square$

Throughout the rest of the paper, for any random variable X, let U_X be a U$[0,1]$ random variable such that $X = F_X^{-1}(U_X)$. The two lemmas below give basic ES formulas in terms of VaR. Note that Lemma 2.3 is based on Lemma 2.2, which is further based on Lemma 2.1. For each proof in Section 3, we will indicate whether any of the Lemmas 2.1-2.3 is required.

Lemma 2.2. *For $p \in (0,1)$ and $X \in L^1$,*

$$\mathrm{ES}_p(X) = \mathrm{VaR}_p(X) + \frac{1}{1-p}\mathbb{E}[(X - \mathrm{VaR}_p(X))_+].$$

Proof. By direct calculation,

$$\begin{aligned}
\mathrm{ES}_p(X) &= \frac{1}{1-p}\int_p^1 F_X^{-1}(q)\mathrm{d}q \\
&= F_X^{-1}(p) + \frac{1}{1-p}\int_p^1 (F_X^{-1}(q) - F_X^{-1}(p))\mathrm{d}q \\
&= \mathrm{VaR}_p(X) + \frac{1}{1-p}\mathbb{E}[(F_X^{-1}(U_X) - \mathrm{VaR}_p(X))_+] \\
&= \mathrm{VaR}_p(X) + \frac{1}{1-p}\mathbb{E}[(X - \mathrm{VaR}_p(X))_+].
\end{aligned}$$

$\square$

Lemma 2.3. *For any $X \in L^1$ and $p \in (0,1)$,*

$$(1-p)\mathrm{ES}_p(X) = \mathbb{E}[X\mathrm{I}_{\{X > \mathrm{VaR}_p(X)\}}] + \mathrm{VaR}_p(X)(\mathbb{P}(X \le \mathrm{VaR}_p(X)) - p).$$

Proof. From Lemma 2.2,

$$\begin{aligned}
(1-p)\mathrm{ES}_p(X) &= (1-p)\mathrm{VaR}_p(X) + \mathbb{E}[(X - \mathrm{VaR}_p(X))\mathrm{I}_{\{X > \mathrm{VaR}_p(X)\}}] \\
&= \mathbb{E}[X\mathrm{I}_{\{X > \mathrm{VaR}_p(X)\}}] + \mathrm{VaR}_p(X)(\mathbb{P}(X \le \mathrm{VaR}_p(X)) - p).
\end{aligned}$$

$\square$

Remark 2.1. Lemma 2.2 yields a precise mathematical formulation of the vague statement from practice that "ES captures tail-risk beyond VaR"; see BCBS (2012, p.3). Lemmas 2.2 and 2.3 are well known; see for instance Dhaene et al. (2006, Theorem 2.1).

2.2 Reduction to L^∞

In all proofs in Section 3, we show Theorem 1.1 only for $X, Y \in L^\infty$. This conclusion alone is usually sufficient for a graduate course in a related field. For completeness, below we give a brief argument showing that the case of L^∞ directly implies the case of L^0. When only risks in L^∞ are relevant, the instructor may skip the following argument.

Assume that (1.1) holds for all $X, Y \in L^\infty$.

(i) If $\mathbb{E}[(X)_+] = \infty$, then $\mathrm{ES}_p(X) = \infty$ and (1.1) holds trivially.

(ii) For $X, Y \in L^1$ and bounded from below, without loss of generality we can assume that $X, Y \geq 0$ since ES_p is translation-invariant. Let $X_k = \min\{X, k\}$, $Y_k = \min\{Y, k\}$ and $Z_k = X_k + Y_k$ for $k = 1, 2, \dots$. Hence, for all $k = 1, 2, \dots$, $X_k, Y_k, Z_k \in L^\infty$, implying

$$\mathrm{ES}_p(Z_k) \leq \mathrm{ES}_p(X_k) + \mathrm{ES}_p(Y_k) \leq \mathrm{ES}_p(X) + \mathrm{ES}_p(Y). \tag{2.1}$$

Note that $Z_k = \min\{X + Y, X + k, Y + k, 2k\} \geq \min\{X + Y, k\}$. Hence

$$\begin{aligned}
\mathrm{ES}_p(X+Y) \geq \mathrm{ES}_p(Z_k) &= \frac{1}{1-p}\int_p^1 \mathrm{VaR}_q(Z_k)\mathrm{d}q \\
&\geq \frac{1}{1-p}\int_p^1 \mathrm{VaR}_q(\min\{X+Y, k\})\mathrm{d}q \\
&= \frac{1}{1-p}\int_p^1 \min\{\mathrm{VaR}_q(X+Y), k\}\mathrm{d}q \\
&\to \mathrm{ES}_p(X+Y) \quad \text{as } k \to \infty,
\end{aligned}$$

where the convergence is justified by the Monotone Convergence Theorem. It follows that $\mathrm{ES}_p(Z_k) \to \mathrm{ES}_p(X+Y)$ as $k \to \infty$. Taking the limit as $k \to \infty$ in (2.1) we obtain (1.1).

(iii) For $X, Y \in L^0$, write $\bar{X} = \max\{X, \mathrm{VaR}_p(X)\}$ and $\bar{Y} = \max\{Y, \mathrm{VaR}_p(Y)\}$. It is obvious that $\mathrm{ES}_p(\bar{X}) = \mathrm{ES}_p(X)$, $\mathrm{ES}_p(\bar{Y}) = \mathrm{ES}_p(Y)$ and $X + Y \leq \bar{X} + \bar{Y}$. By (i) and (ii), (1.1) holds for all $X, Y \in L^0$ bounded from below. Therefore

$$\mathrm{ES}_p(X+Y) \leq \mathrm{ES}_p(\bar{X} + \bar{Y}) \leq \mathrm{ES}_p(\bar{X}) + \mathrm{ES}_p(\bar{Y}) = \mathrm{ES}_p(X) + \mathrm{ES}_p(Y).$$

3 Seven proofs of the subadditivity of ES

In the following, seven proofs are ordered by their (perceived or real) level of technical difficulty. Each proof is self-contained and the reader does not need to follow the given order.

The proofs may require some of the Lemmas 2.1-2.3. Proofs 5-7 further require some classic results from probability or statistics. To the best of our knowledge, Proofs 1 and 4 are not clearly given in the literature, whereas the others can be found, be it in slightly different forms.

3.1 A proof based on comonotonicity

This proof requires Lemma 2.1. In the following, denote by B_p the set of Bernoulli$(1-p)$ random variables and write $A_X = \mathrm{I}_{\{U_X \geq p\}} \in B_p$.

Lemma 3.1. $\mathbb{E}[XA_X] \geq \mathbb{E}[XB]$ *for all* $B \in B_p$.

Proof. Since $\mathbb{E}[A_X - B] = 0$, we have $\mathbb{E}[X(A_X - B)] = \mathbb{E}[(X-m)(A_X - B)]$ for all $m \in \mathbb{R}$. Take $m = F_X^{-1}(p)$. If $F_X^{-1}(U_X) > m$, then $U_X > p$, $A_X = 1$ and $\mathbb{E}[(X-m)(A_X-B)] \geq 0$; if $F_X^{-1}(U_X) < m$, then $U_X < p$, $A_X = 0$ and $\mathbb{E}[(X-m)(A_X-B)] \geq 0$; if $F_X^{-1}(U_X) = m$, then $\mathbb{E}[(X-m)(A_X-B)] = 0$. In summary, $\mathbb{E}[X(A_X-B)] = \mathbb{E}[(X-m)(A_X-B)] \geq 0$. □

Theorem 1.1, Proof 1. We have that

$$\mathrm{ES}_p(X) = \frac{1}{1-p}\int_p^1 F_X^{-1}(q)\mathrm{d}q = \frac{1}{1-p}\mathbb{E}\left[F_X^{-1}(U_X)\mathrm{I}_{\{U_X \geq p\}}\right] = \frac{1}{1-p}\mathbb{E}[XA_X].$$

From Lemma 3.1,

$$\mathrm{ES}_p(X) = \frac{1}{1-p}\sup\{\mathbb{E}[XB] : B \in B_p\}, \;\; X \in L^\infty. \tag{3.1}$$

That is, ES_p is the supremum of the additive maps $X \mapsto \frac{1}{1-p}\mathbb{E}[XB]$ over $B \in B_p$, and hence is subadditive. □

Remark 3.1. Lemma 3.1 is implied by the well-known fact that comonotonic random variables (like X and A_X) have the maximum correlation among random variables with the same marginal distributions; see McNeil et al. (2005, Theorem 5.25(2)) and McNeil et al. (2015, Theorem 7.28(2)). Historically, this result dates back to Hoeffding (1940) and Fréchet (1951), and it is now common knowledge in quantitative risk management at the graduate level; if it is taken for granted, then Lemma 3.1 can be omitted and the proof can be further shortened. (3.1) gives a special form of the coherence representation of ES_p along the lines of Artzner et al. (1999); a similar formulation is

$$\mathrm{ES}_p(X) = \sup\{\mathbb{E}[X|A] : A \in \mathcal{F}, \; \mathbb{P}(A) \geq 1-p\}, \;\; X \in L^\infty;$$

see for instance Denuit et al. (2005, Remark 2.4.8). This representation links ES_p to stress testing. Indeed note that in practice p is close to 1 so that $1-p$ is typically very small. In the case of Economic Capital, $p = 0.9997$ whereas for regulatory purposes p typically lies in the range of 0.95 to 0.999.

The merits of Proof 1 include: It is by far the shortest proof; it reveals a coherence representation of ES_p in (3.1), and it connects naturally to comonotonicity and the theory of copulas.

We would recommend Proof 1 in a course where the concepts of copulas, comonotonicity, or the coherence representation of ES are points of interest.

3.2 A proof based on an optimization property of VaR and ES

This proof requires Lemma 2.2. It is based on a few extra lemmas but it requires no additional knowledge of modern probability theory.

Lemma 3.2. *For $p \in (0,1)$ and $X \in L^\infty$,*

$$\mathrm{VaR}_p(X) \in \operatorname*{argmin}_{t\in\mathbb{R}} \left\{ t + \frac{1}{1-p}\mathbb{E}[(X-t)_+] \right\}. \tag{3.2}$$

Proof. Let $f : \mathbb{R} \to \mathbb{R}$, $t \mapsto t + \frac{1}{1-p}\mathbb{E}[(X-t)_+]$, or equivalently,

$$f(t) = t + \frac{1}{1-p}\int_t^\infty (x-t)\mathrm{d}F_X(x) = t + \frac{1}{1-p}\int_t^\infty (1-F_X(x))\mathrm{d}x,\ t \in \mathbb{R},$$

where the last equality is due to integration by parts. Write $t_0 = \mathrm{VaR}_p(X)$. For $t_1 > t_0$, we have

$$f(t_1) - f(t_0) = (t_1 - t_0) - \frac{1}{1-p}\int_{t_0}^{t_1} (1-F_X(x))\mathrm{d}x.$$

From the definition of $\mathrm{VaR}_p(X)$, we have that for $x \in (t_0, t_1)$, $F_X(x) \geq p$ and hence $1 - F_X(x) \leq 1-p$. As a consequence,

$$f(t_1) - f(t_0) \geq (t_1 - t_0) - \frac{1}{1-p}(1-p)(t_1 - t_0) = 0.$$

For $t_2 < t_0$, we have

$$f(t_2) - f(t_0) = \frac{1}{1-p}\int_{t_2}^{t_0} (1-F_X(x))\mathrm{d}x - (t_0 - t_2).$$

From the definition of $\mathrm{VaR}_p(X)$, we have that for $x \in (t_2, t_0)$, $F_X(x) < p$ and hence $1 - F_X(x) > 1-p$. As a consequence,

$$f(t_2) - f(t_0) > \frac{1}{1-p}(1-p)(t_0 - t_2) - (t_0 - t_2) = 0.$$

In summary, $t_0 \in \operatorname{argmin}_{t\in\mathbb{R}} f(t)$, that is, (3.2) holds. $\square$

Lemma 3.3. *For $p \in (0,1)$ and $X \in L^\infty$,*

$$\mathrm{ES}_p(X) = \min_{t\in\mathbb{R}} \left\{ t + \frac{1}{1-p}\mathbb{E}[(X-t)_+] \right\}.$$

Proof. This follows directly from Lemmas 2.2 and 3.2. □

Theorem 1.1, Proof 2. Write $t_1 = \mathrm{VaR}_p(X)$ and $t_2 = \mathrm{VaR}_p(Y)$. By Lemma 2.2, one has

$$\mathrm{ES}_p(X) + \mathrm{ES}_p(Y) = t_1 + t_2 + \frac{1}{1-p}\mathbb{E}[(X-t_1)_+] + \frac{1}{1-p}\mathbb{E}[(Y-t_2)_+].$$

Note that $(x+y)_+ \le (x)_+ + (y)_+$ for all $x, y \in \mathbb{R}$. Therefore, by writing $t_0 = t_1 + t_2$,

$$\begin{aligned}
\mathrm{ES}_p(X) + \mathrm{ES}_p(Y) &= t_0 + \frac{1}{1-p}\mathbb{E}[(X-t_1)_+ + (Y-t_2)_+] \\
&\ge t_0 + \frac{1}{1-p}\mathbb{E}[(X+Y-t_0)_+] \\
&\ge \min_{t\in\mathbb{R}} \left\{ t + \frac{1}{1-p}\mathbb{E}[(X+Y-t)_+] \right\} \\
&= \mathrm{ES}_p(X+Y),
\end{aligned}$$

where the last equality follows from Lemma 3.3. □

Remark 3.2. This proof is used in Denuit et al. (2005, Section 2.4.3) and Kaas et al. (2008, Section 5.6). The optimization properties in Lemmas 3.2 and 3.3 were established in Acerbi and Tasche (2002) and Rockafellar and Uryasev (2002). Based on these optimization properties, a clear interpretation of VaR_p being a cost-efficient threshold and ES_p being the corresponding minimal cost is given in Denuit et al. (2005) and Kaas et al. (2008). A straightforward geometric proof of Lemma 3.3 can be found in Dhaene et al. (2008, Theorem 1). The main idea in Proof 2 is also used to show the subadditivity of the Haezendonck-Goovaerts risk measure; see Goovaerts et al. (2004).

The merits of Proof 2 include: It is real analysis based without involving techniques from modern probability theory; it reveals the important optimization properties of VaR_p and ES_p in Lemmas 3.2 and 3.3, and it is easy to understand for undergraduate students.

We would recommend Proof 2 in a course where the target audience is at the undergraduate level, or the optimization properties in Lemmas 3.2 and 3.3 are points of interest.

3.3 A proof based on generalized indicator functions

This proof requires Lemma 2.3. We first need some basic results about generalized indicator functions.

Lemma 3.4. *For any $X \in L^\infty$ and $p \in (0,1)$, $\mathbb{P}(X < \mathrm{VaR}_p(X)) \le p \le \mathbb{P}(X \le \mathrm{VaR}_p(X))$.*

Proof. By definition of VaR_p, $\mathbb{P}(X \le \mathrm{VaR}_p(X)) \ge p$ and $\mathbb{P}(X \le \mathrm{VaR}_p(X) - \epsilon) < p$ for all $\epsilon > 0$. By taking $\epsilon \downarrow 0$ we obtain $\mathbb{P}(X < \mathrm{VaR}_p(X)) \le p$. □

In the following, for $p \in (0,1)$, $X \in L^\infty$ and $x \in \mathbb{R}$, define a generalized indicator function

$$\mathrm{I}^{(p)}_{\{X \ge x\}} = \begin{cases} \mathrm{I}_{\{X>x\}}, & \text{if } \mathbb{P}(X=x)=0; \\ \mathrm{I}_{\{X>x\}} + \frac{\mathbb{P}(X\le x)-p}{\mathbb{P}(X=x)}\mathrm{I}_{\{X=x\}}, & \text{if } \mathbb{P}(X=x)>0. \end{cases}$$

Lemma 3.5. *For $p \in (0,1)$, $X \in L^\infty$ and $x \in \mathbb{R}$, the following hold:*

(i) $0 \le \mathrm{I}^{(p)}_{\{X \ge \mathrm{VaR}_p(X)\}} \le 1$;

(ii) $\mathbb{E}[\mathrm{I}^{(p)}_{\{X \ge \mathrm{VaR}_p(X)\}}] = 1-p$;

(iii) $\mathbb{E}[X\mathrm{I}^{(p)}_{\{X \ge \mathrm{VaR}_p(X)\}}] = (1-p)\mathrm{ES}_p(X)$.

Proof. (i) It suffices to verify that $0 \le \mathbb{P}(X \le \mathrm{VaR}_p(X)) - p \le \mathbb{P}(X = \mathrm{VaR}_p(X))$, which directly follows from Lemma 3.4.

(ii) If $\mathbb{P}(X = \mathrm{VaR}_p(X)) = 0$, then

$$\mathbb{E}[\mathrm{I}^{(p)}_{\{X>\mathrm{VaR}_p(X)\}}] = \mathbb{P}(X > \mathrm{VaR}_p(X)) = 1 - \mathbb{P}(X \le \mathrm{VaR}_p(X)) \le 1-p.$$

On the other hand,

$$1 - \mathbb{P}(X \le \mathrm{VaR}_p(X)) = 1 - \mathbb{P}(X < \mathrm{VaR}_p(X)) \ge 1-p,$$

hence $\mathbb{E}[\mathrm{I}^{(p)}_{\{X \ge \mathrm{VaR}_p(X)\}}] = 1-p$.

If $\mathbb{P}(X = \mathrm{VaR}_p(X)) > 0$, then

$$\mathbb{E}[\mathrm{I}^{(p)}_{\{X\ge\mathrm{VaR}_p(X)\}}] = \mathbb{P}(X > \mathrm{VaR}_p(X)) + \frac{1-\mathbb{P}(X > \mathrm{VaR}_p(X)) - p}{\mathbb{P}(X = \mathrm{VaR}_p(X))}\mathbb{P}(X = \mathrm{VaR}_p(X)) = 1-p.$$

(iii) If $\mathbb{P}(X = \mathrm{VaR}_p(X)) = 0$, then by Lemma 2.3 and noting that $\mathbb{P}(X \leq \mathrm{VaR}_p(X)) = p$,

$$\mathbb{E}[X\mathrm{I}^{(p)}_{\{X\geq \mathrm{VaR}_p(X)\}}] = \mathbb{E}[X\mathrm{I}_{\{X>\mathrm{VaR}_p(X)\}}] = (1-p)\mathrm{ES}_p(X).$$

If $\mathbb{P}(X = \mathrm{VaR}_p(X)) > 0$, then

$$\begin{aligned}\mathbb{E}[X\mathrm{I}^{(p)}_{\{X\geq \mathrm{VaR}_p(X)\}}] &= \mathbb{E}[X\mathrm{I}_{\{X>\mathrm{VaR}_p(X)\}}] + \frac{\mathbb{P}(X \leq \mathrm{VaR}_p(X)) - p}{\mathbb{P}(X = \mathrm{VaR}_p(X))}\mathbb{E}[X\mathrm{I}_{\{X=\mathrm{VaR}_p(X)\}}] \\ &= \mathbb{E}[X\mathrm{I}_{\{X>\mathrm{VaR}_p(X)\}}] + (\mathbb{P}(X \leq \mathrm{VaR}_p(X)) - p)\mathrm{VaR}_p(X) \\ &= (1-p)\mathrm{ES}_p(X).\end{aligned}$$

□

Theorem 1.1, Proof 3. By Lemma 3.5(iii),

$$\begin{aligned}&(1-p)(\mathrm{ES}_p(X) + \mathrm{ES}_p(Y) - \mathrm{ES}_p(X+Y)) \\ &= \mathbb{E}[X\mathrm{I}^{(p)}_{\{X\geq \mathrm{VaR}_p(X)\}}] + \mathbb{E}[Y\mathrm{I}^{(p)}_{\{Y\geq \mathrm{VaR}_p(Y)\}}] - \mathbb{E}[(X+Y)\mathrm{I}^{(p)}_{\{X+Y\geq \mathrm{VaR}_p(X+Y)\}}] \\ &= \mathbb{E}[X(\mathrm{I}^{(p)}_{\{X\geq \mathrm{VaR}_p(X)\}} - \mathrm{I}^{(p)}_{\{X+Y\geq \mathrm{VaR}_p(X+Y)\}})] + \mathbb{E}[Y(\mathrm{I}^{(p)}_{\{Y\geq \mathrm{VaR}_p(Y)\}} - \mathrm{I}^{(p)}_{\{X+Y\geq \mathrm{VaR}_p(X+Y)\}})].\end{aligned}$$

Let

$$M = (X - \mathrm{VaR}_p(X))(\mathrm{I}^{(p)}_{\{X\geq \mathrm{VaR}_p(X)\}} - \mathrm{I}^{(p)}_{\{X+Y\geq \mathrm{VaR}_p(X+Y)\}}). \tag{3.3}$$

We will show that $\mathbb{E}[M]$ is non-negative. Note that by Lemma 3.5(i), $0 \leq \mathrm{I}^{(p)}_{\{X+Y\geq \mathrm{VaR}_p(X+Y)\}} \leq 1$. If $X > \mathrm{VaR}_p(X)$, then $\mathrm{I}^{(p)}_{\{X\geq \mathrm{VaR}_p(X)\}} = 1$ and $M \geq 0$; if $X < \mathrm{VaR}_p(X)$, then $\mathrm{I}^{(p)}_{\{X\geq \mathrm{VaR}_p(X)\}} = 0$ and $M \geq 0$; if $X = \mathrm{VaR}_p(X)$, then $M = 0$. In particular,

$$\mathbb{E}[M] = \mathbb{E}[(X - \mathrm{VaR}_p(X))(\mathrm{I}^{(p)}_{\{X\geq \mathrm{VaR}_p(X)\}} - \mathrm{I}^{(p)}_{\{X+Y\geq \mathrm{VaR}_p(X+Y)\}})] \geq 0.$$

By Lemma 3.5(ii),

$$\mathbb{E}[\mathrm{VaR}_p(X)(\mathrm{I}^{(p)}_{\{X\geq \mathrm{VaR}_p(X)\}} - \mathrm{I}^{(p)}_{\{X+Y\geq \mathrm{VaR}_p(X+Y)\}})] = \mathrm{VaR}_p(X)((1-p) - (1-p)) = 0,$$

and we obtain

$$\mathbb{E}[X(\mathrm{I}^{(p)}_{\{X\geq \mathrm{VaR}_p(X)\}} - \mathrm{I}^{(p)}_{\{X+Y\geq \mathrm{VaR}_p(X+Y)\}})] \geq 0.$$

Similarly,

$$\mathbb{E}[Y(\mathrm{I}^{(p)}_{\{Y\geq \mathrm{VaR}_p(Y)\}} - \mathrm{I}^{(p)}_{\{X+Y\geq \mathrm{VaR}_p(X+Y)\}})] \geq 0.$$

Therefore, $(1-p)(\mathrm{ES}_p(X) + \mathrm{ES}_p(Y) - \mathrm{ES}_p(X+Y)) \geq 0$. □

Remark 3.3. This proof is essentially due to Acerbi and Tasche (2002), with some modifications. Note that if a random variable X has a continuous distribution, then $\mathrm{I}^{(p)}_{\{X\ge x\}} = \mathrm{I}_{\{X\ge x\}}$ almost surely. Therefore, in the case when X, Y and $X+Y$ all have continuous distributions, Lemmas 3.4 and 3.5 are not necessary and Proof 3 can be simplified. The proof in the case of continuous distributions is used in the book McNeil et al. (2015, Section 2.3). Note that the distribution of $X+Y$ is not necessarily continuous even if the distributions of X and Y are continuous.

The merit of Proof 3 is that it is based on standard real analysis without involving techniques from more advanced probability theory, and hence it is accessible to undergraduate students. An important aspect, also relevant for practice, is the special treatment of the case when the loss random variable X has an atom at $\mathrm{VaR}_p(X)$. In addition, the proof is fairly simple if only the case of continuous distributions is of interest.

We would recommend Proof 3 in a course where the target audience is at the undergraduate level, or the instructor intends only to teach the case of continuous distributions but not a complete proof. Proof 3 shares some similar argument with Proof 1. Though it can be viewed as elementary, it is technically more involved than Proof 1.

3.4 A proof based on discrete approximation

This proof requires Lemma 2.1. In the following, for $n \in \mathbb{N}$, we say that a random variable X is *n-discrete* if it takes values in a set of at most n points each with probability $1/n$ or a multiple of $1/n$. We say that a random vector (X, Y) is *n-discrete* if it takes values in a set of at most n vectors each with probability $1/n$ or a multiple of $1/n$. Note that (X, Y) being n-discrete implies that X and Y are n-discrete but not vice-versa.

This proof contains two steps: we first show that Theorem 1.1 holds for an n-discrete random vector, and then approximate a general random vector by n-discrete random vectors. The second step involves convergence of random variables and it is more technical than the first step.

Lemma 3.6. *Suppose that a random vector (X, Y) is n-discrete for a positive integer n. Then for $p \in (0,1)$, $\mathrm{ES}_p(X+Y) \le \mathrm{ES}_p(X) + \mathrm{ES}_p(Y)$.*

Proof. We consider three cases: (i) p is a multiple of $1/n$; (ii) p is rational, and (iii) p is general.

(i) Suppose that $np \in \mathbb{N}$. Since (X, Y) is n-discrete, we can divide the sample space Ω into a partition $\Omega_1, \dots, \Omega_n$, each with probability $1/n$, such that for $i = 1, \dots, n$, (X, Y) takes a fixed value, denoted by (x_i, y_i), on Ω_i. Note that for $k = 1, \dots, n$ and $q \in ((k-1)/n, k/n]$,

$$\mathrm{VaR}_q(X) = \inf\{x : \mathbb{P}(X \le x) \ge q\} = x_{[n-k+1]},$$

where $x_{[i]}$ is the i-th largest element in the multiset $\{x_1, \dots, x_n\}$. Write $p = 1 - m/n$ for some $m \in \mathbb{N}$. One can directly calculate $\mathrm{ES}_p(X)$, which is the average of the largest m elements in the multiset $\{x_1, \dots, x_n\}$, that is,

$$\mathrm{ES}_p(X) = \frac{1}{m}\sum_{i=1}^{m} x_{[i]} = \frac{1}{m}\max\{x_{i_1} + \cdots + x_{i_m} : (i_1, \dots, i_m) \in A_m^n\},$$

where

$$A_m^n = \{(i_1, \dots, i_m) \in \mathbb{N}^m : 1 \le i_1 < \cdots < i_m \le n\}.$$

It follows that

$$\begin{aligned}
&m\mathrm{ES}_p(X+Y)\\
&= \max\{x_{i_1} + \cdots + x_{i_m} + y_{i_1} + \cdots + y_{i_m} : (i_1, \dots, i_m) \in A_m^n\}\\
&\le \max\{x_{i_1} + \cdots + x_{i_m} + y_{j_1} + \cdots + y_{j_m} : (i_1, \dots, i_m) \in A_m^n,\ (j_1, \dots, j_m) \in A_m^n\}\\
&= \max\{x_{i_1} + \cdots + x_{i_m} : (i_1, \dots, i_m) \in A_m^n\} + \max\{y_{j_1} + \cdots + y_{j_m} : (j_1, \dots, j_m) \in A_m^n\}\\
&= m\mathrm{ES}_p(X) + m\mathrm{ES}_p(Y).
\end{aligned}$$

(ii) Suppose that p is a rational number. Write $p = k/m$ for $k, m \in \mathbb{N}$ and $k < m$. Note that $p = (kn)/(mn)$ and (X, Y) is also mn-discrete. Therefore, from (i), we have that $\mathrm{ES}_p(X+Y) \le \mathrm{ES}_p(X) + \mathrm{ES}_p(Y)$.

(iii) For a general real number p, from the definition of ES_p, it follows immediately that $p \mapsto \mathrm{ES}_p(X)$ is a continuous mapping. Therefore, we can find rational numbers $p_1, p_2, \dots$ such that $p_k \to p$ as $k \to \infty$ and for $k = 1, 2, \dots$, $\mathrm{ES}_{p_k}(X+Y) \le \mathrm{ES}_{p_k}(X) + \mathrm{ES}_{p_k}(Y)$. Taking a limit as $k \to \infty$ we obtain $\mathrm{ES}_p(X+Y) \le \mathrm{ES}_p(X) + \mathrm{ES}_p(Y)$. $\square$

Lemma 3.7. *Suppose that $X, X_1, X_2, \dots \in L^\infty$ and $X_k \uparrow X$ in probability as $k \to \infty$. Then $\mathrm{ES}_p(X_k) \uparrow \mathrm{ES}_p(X)$ as $k \to \infty$.*

Proof. It suffices to note that $\mathrm{VaR}_q(X_k) \uparrow \mathrm{VaR}_q(X)$ for almost every $q \in (0,1)$, a basic property of the quantile function; see Resnick (1987, Proposition 0.1) for a proof. As ES_p is an integral of VaR_q, the Monotone Convergence Theorem implies $\mathrm{ES}_p(X_k) \uparrow \mathrm{ES}_p(X)$ as $k \to \infty$. $\square$

Lemma 3.8. *Suppose that the random variables X and Y are n-discrete for a positive integer n. Then for $p \in (0,1)$, $\mathrm{ES}_p(X+Y) \le \mathrm{ES}_p(X) + \mathrm{ES}_p(Y)$.*

Proof. Without loss of generality we can assume $X, Y \ge 0$ since they are both bounded, and ES_p is translation-invariant. Denote by $\{(x_1, y_1), \dots, (x_m, y_m)\}$ the range of (X,Y); obviously $m \le n^2$. Note that $\mathbb{P}((X,Y) = (x_i, y_i))$ may not be a rational number for some $i = 1, \dots, m$ and hence Lemma 3.6 cannot be directly applied. Denote $A_i = \{(X,Y) = (x_i, y_i)\}$, $i = 1, \dots, m$. Since our probability space is atomless, for $i = 1, \dots, m$, we can find $B_i^k \subseteq A_i$, $k = 1, 2, \dots$, such that $B_i^1 \subseteq B_i^2 \subseteq \dots$, $\mathbb{P}(B_i^k) \in \mathbb{Q}$, and $\mathbb{P}(A_i) - \mathbb{P}(B_i^k) < 1/k$, where $\mathbb{Q}$ is the set of rational numbers. Let

$$X_k = \sum_{i=1}^m X \mathrm{I}_{B_i^k}, \quad Y_k = \sum_{i=1}^m Y \mathrm{I}_{B_i^k}, \quad k = 1, 2, \dots.$$

It is clear that $X_k \uparrow X$ and $Y_k \uparrow Y$ in probability. Moreover, for each $k \in \mathbb{N}$ (X_k, Y_k) is m_k-discrete for some $m_k \in \mathbb{N}$ since the probability mass function of (X_k, Y_k) takes values in $\mathbb{Q}$. By Lemma 3.6 and the monotonicity of ES_p, we have

$$\mathrm{ES}_p(X_k + Y_k) \le \mathrm{ES}_p(X_k) + \mathrm{ES}_p(Y_k) \le \mathrm{ES}_p(X) + \mathrm{ES}_p(Y).$$

Since $X_k + Y_k \uparrow X + Y$ in probability as $k \to \infty$, by taking a limit in $k \to \infty$ and using Lemma 3.7, the above equation yields $\mathrm{ES}_p(X+Y) \le \mathrm{ES}_p(X) + \mathrm{ES}_p(Y)$. $\square$

Theorem 1.1, Proof 4. Let $X_k = F_X^{-1}(V_k)$ and $Y_k = F_Y^{-1}(W_k)$ where $U_k = \lfloor 2^k U_X \rfloor / (2^k)$ and $W_k = \lfloor 2^k U_Y \rfloor / (2^k)$. It is obvious that X_k and Y_k are 2^k-discrete, and $X_k \uparrow X$ and $Y_k \uparrow Y$ in probability. Lemma 3.8 yields

$$\mathrm{ES}_p(X_k + Y_k) \le \mathrm{ES}_p(X_k) + \mathrm{ES}_p(Y_k) \le \mathrm{ES}_p(X) + \mathrm{ES}_p(Y).$$

As a consequence, with Lemma 3.7, we obtain $\mathrm{ES}_p(X+Y) \le \mathrm{ES}_p(X) + \mathrm{ES}_p(Y)$. $\square$

Remark 3.4. The instructor may choose only to show Lemma 3.6 by omitting the more complicated second step of the proof. The result in Lemma 3.6 itself is often sufficient for students to understand the subadditivity of ES.

The merits of Proof 4 include: It is based on standard undergraduate level techniques in probability theory such as discrete approximation and convergence theorems; it reveals an intuitive explanation of ES being subadditive through discrete random variables, and it is easy to understand for students with good combinatorial and analytical skills.

We would recommend Proof 4 in a course where discretization of distribution functions or the continuity of risk measures is a point of interest, or the instructor intends to highlight intuition in the discrete case but not give a complete proof.

3.5 A proof based on the law of large numbers for order statistics

In this section, for a sequence of random variables $X_1, X_2, \dots$, we denote by $X_{[i,n]}$ the i-th largest value in $\{X_1, \dots, X_n\}$, that is, the i-th order statistic up to the n-th observation.

Lemma 3.9. *Let $X_1, X_2, \dots \in L^\infty$ be a sequence of iid random variables. Then*

$$\lim_{n\to\infty} \frac{1}{n(1-p)} \sum_{i=1}^{n\lfloor 1-p \rfloor} X_{[i,n]} = \mathrm{ES}_p(X_1) \quad \text{a.s.} \tag{3.4}$$

where $\lfloor x \rfloor$ is the largest integer not larger than $x \in \mathbb{R}$.

Proof. This lemma is implied by a classic result (strong law of large numbers) for linear combinations of order statistics, originally proved in Van Zwet (1980, Theorem 2.1); see also Wellner (1977, Theorem 3 and Corollary 2). Acerbi and Tasche (2002, Proposition 4.1 and Equations (4.2)-(4.4)) gave some intuitive explanations of the proof. □

Theorem 1.1, Proof 5. For any two positive integers $m \le n$, write

$$A_m^n = \{(i_1, \dots, i_m) \in \mathbb{N}^m : 1 \le i_1 < \dots < i_m \le n\}.$$

For a sequence of random variables, $X_1, X_2, \dots$, note that

$$\sum_{i=1}^{m} X_{[i,n]} = \max\{X_{i_1} + \dots + X_{i_m} : (i_1, \dots, i_m) \in A_m^n\}.$$

For any two random variables $X, Y \in L^\infty$, let $(X_1, Y_1), (X_2, Y_2), \dots$ be a sequence of iid random vectors, identically distributed as (X,Y) and write $Z_i = X_i + Y_i$ for $i = 1, 2, \ldots$. Then

$$\begin{aligned}
\sum_{i=1}^{m} Z_{[i,n]} &= \max\{Z_{i_1} + \cdots + Z_{i_m} : (i_1, \ldots, i_m) \in A_m^n\} \\
&= \max\{X_{i_1} + \cdots + X_{i_m} + Y_{i_1} + \cdots + Y_{i_m} : (i_1, \ldots, i_m) \in A_m^n\} \\
&\le \max\{X_{i_1} + \cdots + X_{i_m} + Y_{j_1} + \cdots + Y_{j_m} : (i_1, \ldots, i_m) \in A_m^n,\ (j_1, \ldots, j_m) \in A_m^n\} \\
&= \max\{X_{i_1} + \cdots + X_{i_m} : (i_1, \ldots, i_m) \in A_m^n\} + \max\{Y_{j_1} + \cdots + Y_{j_m} : (j_1, \ldots, j_m) \in A_m^n\} \\
&= \sum_{i=1}^{m} X_{[i,n]} + \sum_{i=1}^{m} Y_{[i,n]}.
\end{aligned}$$

By setting $m = n\lfloor 1-p \rfloor$, we have

$$\frac{1}{n(1-p)} \sum_{i=1}^{n\lfloor 1-p \rfloor} Z_{[i,n]} \le \frac{1}{n(1-p)} \sum_{i=1}^{n\lfloor 1-p \rfloor} X_{[i,n]} + \frac{1}{n(1-p)} \sum_{i=1}^{n\lfloor 1-p \rfloor} Y_{[i,n]}.$$

Taking $n \to \infty$, by Lemma 3.9, we obtain $\mathrm{ES}_p(X+Y) \le \mathrm{ES}_p(X) + \mathrm{ES}_p(Y)$. □

Remark 3.5. This proof is used in the book McNeil et al. (2005, Section 2.2.4); in the revised version McNeil et al. (2015), two proofs are given (see Remarks 3.3 and 3.7). General results in statistics related to Lemma 3.9 can be found in Huber and Ronchetti (2009, Section 3.3).

The merits of Proof 5 include: It requires the law of large numbers for order statistics (Lemma 3.9 above), which gives also a natural non-parametric estimator of $\mathrm{ES}_p(X)$ in (3.4); in a context where statistical estimation of ES_p is relevant, this proof would fit in naturally.

We would recommend Proof 5 in a course where statistical inference is a point of interest, or the students have a solid statistical background. Proof 5 shares some similar argument with Proof 4. Whereas in Proof 4 the discrete case can be solved fairly elementarily, the general case needs a non-trivial probabilistic limit argument. For Proof 5, the general case can immediately be treated by a more powerful limit theorem from the realm of the theory of linear combinations of order statistics.

3.6 A proof based on convex order

For $X, Y \in L^1$, we say that X is smaller than Y *in convex order*, denoted by $X \prec_{\mathrm{cx}} Y$, if for all convex functions f,

$$\mathbb{E}[f(X)] \le \mathbb{E}[f(Y)], \tag{3.5}$$

whenever both sides of (3.5) are well-defined.

Lemma 3.10. *For two random variables $X, Y \in L^\infty$, if $X \prec_{\mathrm{cx}} Y$, then $\mathrm{ES}_p(X) \le \mathrm{ES}_p(Y)$ for $p \in (0,1)$.*

Proof. This is a classic result in convex order; see Shaked and Shanthikumar (2007, Theorem 3.A.5) for a proof. Indeed, the latter result states that, for any $X, Y \in L^1$, $X \prec_{\mathrm{cx}} Y$ if and only if $\mathrm{ES}_p(X) \le \mathrm{ES}_p(Y)$ for all $p \in (0,1)$. $\square$

Lemma 3.11. *For any two random variables $X, Y \in L^\infty$ and a $\mathrm{U}[0,1]$-distributed random variable U, it holds that $X + Y \prec_{\mathrm{cx}} F_X^{-1}(U) + F_Y^{-1}(U)$.*

Proof. This is a classic result on comonotonicity; see for instance Dhaene et al. (2002, Theorem 7) or Rüschendorf (2013, Theorem 3.5) for a proof. $\square$

Lemma 3.12. *For any two non-decreasing functions f, g and a random variable $Z \in L^\infty$, it holds that $\mathrm{ES}_p(f(Z)+g(Z)) = \mathrm{ES}_p(f(Z)) + \mathrm{ES}_p(g(Z))$. That is, ES_p is comonotonic additive.*

Proof. This is another classic result on comonotonicity. First, note that for any non-decreasing function h and a $\mathrm{U}[0,1]$-distributed random variable U, $\mathrm{VaR}_p(h(U)) = h(p)$. Then

$$
\begin{aligned}
\mathrm{ES}_p(f(Z)+g(Z)) &= \mathrm{ES}_p(f(F_Z^{-1}(U_Z)) + g(F_Z^{-1}(U_Z))) \\
&= \frac{1}{1-p}\int_p^1 \left(f(F_Z^{-1}(u)) + g(F_Z^{-1}(u))\right) \mathrm{d}u \\
&= \frac{1}{1-p}\int_p^1 \mathrm{VaR}_u(f(Z))\mathrm{d}u + \frac{1}{1-p}\int_p^1 \mathrm{VaR}_u(g(Z))\mathrm{d}u \\
&= \mathrm{ES}_p(f(Z)) + \mathrm{ES}_p(g(Z)),
\end{aligned}
$$

where the second-last equality comes from the fact that $f(\mathrm{VaR}_u(Z)) = \mathrm{VaR}_u(f(Z))$ for almost every $u \in [0,1]$ since f is non-decreasing. $\square$

Theorem 1.1, Proof 6. We obtain $\mathrm{ES}_p(X+Y) \le \mathrm{ES}_p(F_X^{-1}(U)+F_Y^{-1}(U)) = \mathrm{ES}_p(X)+\mathrm{ES}_p(Y)$ by combining Lemmas 3.10-3.12. $\square$

Remark 3.6. The idea of this proof is presented in Wang and Dhaene (1998) and is also used in the review paper Dhaene et al. (2006). Lemma 3.10 dates back to Levy and Kroll (1978) in the context of stochastic dominance. Lemma 3.11 was first shown in Meilijson and Nádas (1979). The fact that ES_p is comonotone additive is part of the properties of *Choquet integrals*, and it can be found in Yaari (1987) and Denneberg (1994). The current form of Lemma 3.12 is given in Kusuoka (2001, Proposition 20).

The merit of Proof 6 is that it naturally connects to the concepts of convex order, comonotonicity and comonotonic additivity, all of which are important modern concepts in quantitative risk management. This proof requires additional techniques in probability theory, and it may not be suitable for an audience without the corresponding knowledge.

We would recommend Proof 6 in an advanced course where convex order, comonotonicity and comonotonic additivity are points of interest or available as preliminaries.

3.7 A proof based on the coherence representation of ES

This proof requires Lemma 2.3 and knowledge on Radon-Nikodym derivatives of probability measures. It is probably the most mathematically advanced among all proofs in this paper.

Lemma 3.13 (Neyman-Pearson Lemma). *Let P and Q be two probability measures such that $\phi = \mathrm{d}P/\mathrm{d}Q$ is finite. For any $\alpha \in (0,1)$, let $c = \inf\{x \in \mathbb{R} : \mathbb{E}^Q[\mathrm{I}_{\{\phi \le x\}}] \ge 1-\alpha\}$ and $\psi^0 = \mathrm{I}_{\{\phi > c\}} + \kappa \mathrm{I}_{\{\phi = c\}}$ where κ is a constant such that $\mathbb{E}^Q[\psi^0] = \alpha$. Then, for any $\psi \in L^\infty$, $0 \le \psi \le 1$, $\mathbb{E}^Q[\psi] \le \alpha$, one has*

$$\mathbb{E}^P[\psi] \le \mathbb{E}^P[\psi^0].$$

Proof. First, one can easily check the existence of κ. Note that by definition of c, $\mathbb{E}^Q[\mathrm{I}_{\{\phi \ge c\}}] \ge \alpha$ and $\mathbb{E}^Q[\mathrm{I}_{\{\phi > c\}}] \le \alpha$, and hence either $\kappa \in [0,1]$ or $\mathbb{E}[\mathrm{I}_{\{\phi = c\}}] = 0$.

For any $\psi \in L^\infty$, $0 \le \psi \le 1$, $\mathbb{E}^Q[\psi] \le \alpha$, by definition of ψ^0, one has $(\psi^0 - \psi)(\phi - c) \ge 0$ almost surely. Therefore,

$$0 \le \mathbb{E}^Q[(\psi^0 - \psi)(\phi - c)] = \mathbb{E}^P[\psi^0 - \psi] - c\mathbb{E}^Q[\psi^0 - \psi] \le \mathbb{E}^P[\psi^0 - \psi]. \qquad \square$$

Lemma 3.14. *For $p \in (0,1)$, ES_p has the representation*

$$\mathrm{ES}_p(X) = \sup_{Q \in \mathcal{Q}_p} \mathbb{E}^Q[X], \quad X \in L^\infty,$$

where $\mathcal{Q}_p$ is the set of probability measures Q on $(\Omega, \mathcal{F})$ with Radon-Nikodym derivative $\mathrm{d}Q/\mathrm{d}\mathbb{P} \le 1/(1-p)$.

Proof. Define a mapping $\rho : L^\infty \to \mathbb{R}$ by

$$\rho(X) = \sup_{Q \in \mathcal{Q}_p} \mathbb{E}^Q[X], \quad X \in L^\infty.$$

We aim to show that $\mathrm{ES}_p = \rho$. First, for $X > 0$, define a probability measure P such that $\mathrm{d}P/\mathrm{d}\mathbb{P} = X/\mathbb{E}[X]$. Then for $X \in L^\infty$,

$$\begin{aligned}
\rho(X) &= \sup\left\{\mathbb{E}\left[\frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}}X\right] : Q \in \mathcal{Q}_p\right\} \\
&= \sup\left\{\mathbb{E}[X]\mathbb{E}^P\left[\frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}}\right] : Q \in \mathcal{Q}_p\right\} \\
&= \mathbb{E}[X]\sup\left\{\mathbb{E}^P\left[Z\right] : Z \in L^1,\ 0 \le Z \le \frac{1}{1-p},\ \mathbb{E}[Z] = 1\right\} \\
&= \frac{\mathbb{E}[X]}{1-p}\sup\left\{\mathbb{E}^P\left[W\right] : W \in L^1,\ 0 \le W \le 1,\ \mathbb{E}[W] \le 1-p\right\}
\end{aligned}$$

By Lemma 3.13, the above supremum is attained by

$$W_0 = \mathrm{I}_{\{X > \mathrm{VaR}_p(X)\}} + \kappa \mathrm{I}_{\{X = \mathrm{VaR}_p(X)\}},$$

where $\kappa \in [0,1]$ is such that $\mathbb{E}[W_0] = 1-p$, that is $\kappa = \frac{\mathbb{P}(X \le \mathrm{VaR}_p(X)) - p}{\mathbb{P}(X = \mathrm{VaR}_p(X))}$ if $\mathbb{P}(X = \mathrm{VaR}_p(X)) > 0$, and κ can take any value in $[0,1]$ if $\mathbb{P}(X = \mathrm{VaR}_p(X)) = 0$. Therefore,

$$\begin{aligned}
\rho(X) &= \frac{\mathbb{E}[X]}{1-p}\mathbb{E}\left[\frac{X}{\mathbb{E}[X]}(\mathrm{I}_{\{X > \mathrm{VaR}_p(X)\}} + \kappa \mathrm{I}_{\{X = \mathrm{VaR}_p(X)\}})\right] \\
&= \frac{1}{1-p}\mathbb{E}[X\mathrm{I}_{\{X > \mathrm{VaR}_p(X)\}} + X\kappa \mathrm{I}_{\{X = \mathrm{VaR}_p(X)\}}] \\
&= \frac{1}{1-p}\mathbb{E}[X\mathrm{I}_{\{X > \mathrm{VaR}_p(X)\}}] + \frac{\mathbb{P}(X \le \mathrm{VaR}_p(X)) - p}{1-p}\mathrm{VaR}_p(X)
\end{aligned}$$

From Lemma 2.3 we have $\rho(X) = \mathrm{ES}_p(X)$. For arbitrary $X \in L^\infty$, $\rho(X) = \mathrm{ES}_p(X)$ follows from the above result by noting that both ρ and ES_p are translation-invariant. □

Theorem 1.1, Proof 7. ES_p is the supremum of the additive maps $X \mapsto \mathbb{E}^Q[X]$ over $Q \in \mathcal{Q}_p$, and hence is subadditive. □

Remark 3.7. This proof is used in the book McNeil et al. (2015, Theorem 8.14). Lemma 3.13 is a general form of the Neyman-Pearson lemma for optimal simple tests; see Föllmer and Schied (2011, Theorem A.31 and Remark A.32) and McNeil et al. (2015, Remark 8.15). Lemma 3.14 is found in classic literature on coherent risk measures; see for instance Föllmer and Schied (2011, Theorem 4.52).

The merit of Proof 7 is that Lemma 3.14 reveals the coherence representation of ES_p, a fundamental property of ES_p, connecting to the dual representation of coherent risk measures in the most general form. This proof requires additional techniques in modern probability theory and statistics, and it may not suitable for an audience without the corresponding knowledge. A further merit is that the link between the Neyman-Pearson lemma and mathematical finance can also be found in the concept of quantile hedging in incomplete markets; see Föllmer and Schied (2011, Section 8.1).

We would recommend Proof 7 in an advanced course where the axiomatic theory of coherent risk measures is a point of interest. Proof 7 can be viewed as a more comprehensive and advanced version of Proof 1.

4 Overall comments

From the seven different proofs we offer in this paper, it becomes clear that the subadditivity question for ES is not a trivial one. It is both mathematically challenging as well as practically relevant. Each proof has its own merit and suitable context. In summary, Proof 1 is the shortest and it reveals a special form of the coherence representation of ES; Proofs 2 and 3 require the least knowledge on probability theory, with Proof 2 being shorter and connected to an optimization property of VaR and ES, and Proof 3 being convenient if only the case of continuous distributions is of interest; Proof 4 explains intuitively the subadditivity of ES in discrete cases and reveals a basic continuity of ES; Proofs 5, 6 and 7 require specialized knowledge and they respectively reveal natural estimators of ES, its consistency with convex order, and the coherence presentation of coherent risk measures. When teaching the subadditivity of ES in a course, the instructor is advised to choose a proof which fits best the knowledge of the audience and the content of the course.

Acknowledgement

The authors would like to thank the Editor and three anonymous referees for their valuable comments on an earlier version of this paper. Paul Embrechts would like to thank the Swiss Finance Institute for financial support. Ruodu Wang would like to thank FIM at ETH Zurich for supporting his visit in 2015, and he acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (RGPIN-435844-2013).

References

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, **26**(7), 1505–1518.

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, **26**(7), 1487–1503.

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

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

BCBS (2013). *Consultative Document October 2013. Fundamental review of the trading book: A revised market risk framework.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements.

BCBS (2014). *Consultative Document December 2014. Fundamental review of the trading book: Outstanding issues.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements.

Billingsley, P. (1995). *Probability and Measure*. Third Edition. Wiley.

Choquet, G. (1953). Theory of capacities. *Annales de L'Institut Fourier*, **5**, 121–293.

Cont, R., Deguest, R. and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, **10**(6), 593–606.

Delbaen, F. (2012). *Monetary Utility Functions*. Osaka University Press.

Denneberg, D. (1994). *Non-additive Measure and Integral.* Springer.

Denuit, M., Dhaene, J., Goovaerts, M.J. and Kaas, R. (2005). *Actuarial Theory for Dependent Risks.* Wiley.

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

Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R., Tang, Q. and Vynche, D. (2006). Risk measures and comonotonicity: a review. *Stochastic Models*, **22**, 573–606.

Dhaene, J., Laeven, R. J., Vanduffel, S., Darkiewicz, G. and Goovaerts, M. J. (2008). Can a coherent risk measure be too subadditive? *Journal of Risk and Insurance*, **75**(2), 365–386.

Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research*, **77**(3), 423–432.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25-48.

Föllmer, H. and Schied, A. (2011). *Stochastic Finance: An Introduction in Discrete Time.* Walter de Gruyter, Third Edition.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon. Sect. A.* **14**, 53–77.

Goovaerts, M. J., Kaas, R., Dhaene, J. and Tang, Q. (2004). Some new classes of consistent risk measures. *Insurance: Mathematics and Economics*, **34**(3), 505–516.

Hoeffding, W. (1940). Massstabinvariante Korrelationstheorie. *Schriften Math. Inst. Univ. Berlin* **5**(5), 181–233.

Huber, P. J. (1980). *Robust Statistics.* First ed., Wiley Series in Probability and Statistics. Wiley, New Jersey.

Huber, P. J. and Ronchetti E. M. (2009). *Robust Statistics.* Second ed., Wiley Series in Probability and Statistics. Wiley, New Jersey. First ed.: Huber, P. (1980).

IAIS (2014). *Consultation Document December 2014. Risk-based global insurance capital standard*. International Association of Insurance Supervisors.

Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R.* Springer.

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

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

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

McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton University Press.

Meilijson, I. and A. Nádas (1979). Convex majorization with an application to the length of critical paths. *Journal of Applied Probability* **16**(3), 671–677.

Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, **26**(7), 1443–1471.

Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer.

Van Zwet, W.R. (1980). A strong law for linear functions of order statistics. *Annals of Probablity*, **8**, 986–990.

Wang, S. and Dhaene, J. (1998). Comonotonicity, correlation order and premium principles. *Insurance: Mathematics and Economics*, **22**(3), 235–242.

Wang, S., Young, V. R. and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, **21**(2), 173–183.

Wellner, J. A. (1977). A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. *Annals of Statistics*, **5**(3), 473–480.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.