An Axiomatic Foundation for the Expected Shortfall

Ruodu Wang* Ričardas Zitikis†

12th March 2020‡

Abstract

In the recent Basel Accords, the Expected Shortfall (ES) replaces the Value-at-Risk (VaR) as the standard risk measure for market risk in the banking sector, making it the most popular risk measure in financial regulation. Although ES is – in addition to many other nice properties – a coherent risk measure, it does not yet have an axiomatic foundation. In this paper we put forward four intuitive economic axioms for portfolio risk assessment – monotonicity, law invariance, prudence and no reward for concentration – that uniquely characterize the family of ES. The herein developed results, therefore, provide the first economic foundation for using ES as a globally dominating regulatory risk measure, currently employed in Basel III/IV. Key to the main results, several novel notions such as tail events and risk concentration naturally arise, and we explore them in detail. As a most important feature, ES rewards portfolio diversification and penalizes risk concentration in a special and intuitive way, not shared by any other risk measure.

KEYWORDS: Risk measure, Expected Shortfall, risk concentration, diversification, risk aggregation.

*RUODU WANG: Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L 5A7, Canada. E-mail: wang@uwaterloo.ca

†RIČARDAS ZITIKIS: School of Mathematical and Statistical Sciences, University of Western Ontario, London, Ontario N6A 5B7, Canada. E-mail: rzitikis@uwo.ca

‡The authors thank the Department Editor Manel Baucells, an Associate Editor, three anonymous referees, Michel Baes, Fabio Bellini, Paul Embrechts, Danqiao Guo, Nan Guo, Cosimo Munari, Pablo Koch-Medina, Liyuan Lin, Ludger Rüschendorf, Alexander Schied, Dirk Tasche, Andreas Tsanakas, and Xunyu Zhou for many helpful comments, as well as participants at the 3rd International Congress on Actuarial Science and Quantitative Finance in Manizales (Colombia), the 6th Dependence Modelling workshop in Agistri (Greece), the 2019 INFORMS Annual Meeting in Seattle (USA), the 12th International Conference of the ERCIM WG on Computational and Methodological Statistics in London (UK), and seminars at Peking University, Chongqing University, York University, University of Zurich, University of Wisconsin-Milwaukee, and London School of Economics and Political Science for useful feedback. The authors acknowledge financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada (Grant Nos. RGPIN-2018-03823, RGPAS-2018-522590, RGPIN-2016-427216), and the National Research Organization "Mathematics of Information Technology and Complex Systems" (MITACS) of Canada.

1 Introduction

The Value-at-Risk (VaR) and the Expected Shortfall (ES) – the latter also known as CVaR, TVaR and AVaR – are the most popular risk measures in banking and insurance. They are widely employed for regulatory capital calculation, decision making, performance analysis, and risk management. In particular, both VaR and ES appear in the banking regulation frameworks of Basel III/IV, as well as in the insurance regulation frameworks of Solvency II and the Swiss Solvency Test. A major interpretation of these risk measures is the capital requirement for potential losses faced by financial institutions, and in this paper we take the perspective of banking regulation as in Basel III/IV, although the results obtained are applicable well beyond the regulatory framework of banking.

The Basel Committee on Banking Supervision in their Fundamental Review of the Trading Book (FRTB) (BCBS, 2016) confirmed the replacement of VaR with ES as the standard risk measure for market risk; see also the more recently revised document BCBS (2019). Specifically, the VaR at the probability level $p = 0.99$ is officially replaced by the ES at the level $p = 0.975$ as the standard risk measure for market risk. There are several reasons for this transition (BCBS, 2016, 2019), and the literature on comparative advantages of VaR and ES is abundant (e.g., Embrechts et al., 2014, 2018, and the references therein). Based on this regulatory transition, we may therefore conclude that ES is currently the most important risk measure in banking practice. See also Azenes and Effixis (2018) for the use of ES in the context of insurance regulation.

Naturally, the properties that a risk measure satisfies, or fails to satisfy, as well as their economic implications, play a pivotal role when deciding whether or not it is suitable for practical use. As a dominating class of risk measures in financial applications, ES has many nice theoretical properties. In particular, ES satisfies the four axioms of coherent risk measures (Artzner et al., 1999; Acerbi and Tasche, 2002), and it is also additive for comonotonic risks (Kusuoka, 2001), thus admitting a convex Choquet integral representation (Yaari, 1987; Schmeidler, 1989). In addition to its economically relevant properties, ES also admits a nice representation as the minimum of expected losses (Rockafellar and Uryasev, 2002), which allows for convenient convex optimization.

In addition to its prominent appearance in financial regulation, the family of ES plays a special and important role in decision analysis. For instance, foundational to the notion of risk aversion (Rothschild and Stiglitz, 1970), second-order stochastic dominance is equivalent to the partial order induced by the family of ES.[1] Moreover, the dual utility of Yaari (1987) can be written as a mixture

[1]More precisely, for two random losses X and Y, $X \prec_{\text{ssd}} Y$ if and only if $\mathrm{ES}_p(X) \geqslant \mathrm{ES}_p(Y)$ for all $p \in (0,1)$, where $\prec_{\text{ssd}}$ stands for second-order stochastic dominance between losses (the convention is the smaller the better);

of ES if and only if the corresponding preference is risk averse.[2]

Note, however, that the aforementioned properties, as well as several other ones that we find in the literature, do not (jointly) characterize ES. Indeed, despite its great popularity and practical relevance, the family of ES has not yet been characterized with any set of economic axioms.[3] As a consequence, a number of important questions remain unanswered. For example, is there a set of natural economic axioms suitable for regulatory risk measures that would characterize ES? How special is ES among the class of coherent risk measures? In sharp contrast to ES, the VaR (quantile) has been uniquely characterized using various sets of axioms. How would those axioms for VaR compare with the axioms that characterize ES? Answering these and other questions is the main purpose of the present paper.

Motivated by practical features of portfolio risk assessment, and in particular by penalizing concentrated portfolios and emphasizing tail events, we put forward four axioms. The axioms – monotonicity, law invariance, prudence, and no reward for concentration – naturally characterize ES from an economic perspective. Our characterization theorem, therefore, empowers regulators from the decision-making perspective. Indeed, if the aforementioned axioms correctly reflect the regulator's practical intentions with respect to risk, then the use of ES as a standard regulatory risk measure is justified. If, however, the axioms contradict the regulator's intentions, then it is the right time to discuss whether ES is still the best risk measure to use. Hence, the present paper yields a technical foundation for further practical discussions of the (dis-)advantages of using ES, among all possible alternative risk measures.

The rest of the paper is organized as follows. In Section 2, we introduce four economic axioms designed for portfolio risk assessment, and put forward an axiomatic characterization of ES in the form of two theorems. A deeper understanding as well as proofs of these two theorems rely on several properties and results concerning tail events and risk concentration, which will be discussed in Sections 3–4. In particular, Section 3 offers some properties and characterizations of risk concentration, and its role as a dependence concept. Section 4 illustrates the prominent role of risk concentration in the maximal risk aggregation for ES and VaR. We conclude the paper in Section 5

see e.g., Shaked and Shanthikumar (2007, Theorem 4.A.3).

[2]This is because any L^1-continuous convex distortion risk measure can be written as a mixture of ES; see e.g., McNeil et al. (2015, Proposition 8.18).

[3]In the literature, the family of ES can be identified with some properties based on VaR. For instance, ES is known to be the smallest law-invariant coherent risk measure dominating the corresponding VaR; see Artzner et al. (1999, Proposition 5.4) and Föllmer and Schied (2016, Theorem 4.67). Moreover, Wang and Wei (2018, Theorem 3) show that ES is the only coherent distortion risk measure co-elicitable with the corresponding VaR. These results do not provide an axiomatic characterization of ES as they rely on the other risk measure VaR. Moreover, the above properties do not seem to have a clear interpretation in portfolio capital assessment.

by discussing a number of important implications of our axioms and results. We relegate proofs of all results to several appendices, which also contain auxiliary results of interest to risk modeling, although they do not fit well into the narratives of the main body of the paper.[4]

1.1 Related literature

Various risk-related terms permeate economic literature: ambiguity, preferences, risk perceptions, and knowledge-based (also called subjective) distortions of probability measures, to name a few. The research on axiomatic approaches for these risk functionals has a long history and, by now, makes up a large part of literature dealing with economic (and statistical) decision theories. For a specimen, see for instance the monographs by Gilboa (2009), Wakker (2010), and to the extensive lists of references therein. In particular, axiomatic studies on preference functionals have been prolific in decision theory (e.g., Maccheroni et al., 2006; Gilboa et al., 2010; Cerreia-Vioglio et al., 2011). In a discussion article, Gilboa et al. (2019) share their insights on the usefulness of axiomatic approaches in modern economic questions.[5]

The development of axiomatic approaches for risk measures has also been quite fruitful. In the financial engineering literature, as a prominent piece of work on the axiomatic approach, Artzner et al. (1999) introduced the class of coherent risk measures, with ES being one of them. Furthermore, Föllmer and Schied (2002) and Frittelli and Rosazza Gianin (2002) have axiomatically identified the class of convex risk measures. Chen et al. (2013) studied axioms for systemic risk measures, an important subject in risk management, especially after the 2007-2009 financial crisis. Many axiomatic studies in quantitative finance and insurance find their roots in economic decision theory. The rank-dependent (also called dual) and Choquet utility theories (Quiggin, 1982; Yaari, 1987; Schmeidler, 1989; Denneberg, 1994, and the references therein) are particularly popular, and they have given rise to, e.g., distortion (Wang, 1996) and spectral (Acerbi, 2002) risk measures, extensively explored in the insurance and finance literature. For the case of VaR, different sets of axioms have been given; see Chambers (2009), Kou and Peng (2016), He and Peng (2018), and Liu and Wang (2020).

---

[4] In particular, Lemma A.2 gives a characterization of functionals that are additive for concentrated risks; Lemmas A.3, A.4 and A.6 and Corollary A.1 give several technical properties of tail events; Lemma A.5 characterizes tail events of the sum of concentrated risks; Proposition A.1 gives a relationship between concentration and pair-wise concentration; Lemma A.7 connects ES with conditional expectation on a tail event.

[5] Relevant to the axiomatic approach in this paper, we quote from Gilboa et al. (2019) the following question: "Are we devoting too much time to axiomatic derivations at the expense of developing theories that fit the data?" together with the paper's answer: "[...] our response, namely that axiomatic derivations are powerful rhetorical devices, and outlines several ways that axiomatic derivations of decision rules may be useful for economics, even when the decision models are interpreted descriptively."

The risk measure ES has been extensively studied in financial econometrics. We refer to Scaillet (2004), Chen (2008), Cai and Wang (2008), Patton et al. (2019) and the references therein for the inference for, and modelling of, ES with financial data. Fermanian and Scaillet (2005) analyze sensitivity of ES in the presence of netting and collateral agreements in the banking industry. For recent developments on (variations of) ES and related risk measure in the context of systemic risk, we refer to Acharya et al. (2012, 2017) and Adrian and Brunnermeier (2016). Studies on the backtesting and elicitation of ES as a regulatory risk measure are found in Ziegel (2016), Fissler and Ziegel (2016), and Du and Escanciano (2017).

1.2 Notation and definitions

Next are basic mathematical notation and definitions that we use throughout the paper. First, we fix an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ of all possible (financial) scenarios of interest. The assumption of an atomless probability space is standard in the literature, and it is important for the main results of this paper. A random variable $X : \Omega \to \mathbb{R}$ usually carries the connotation of a portfolio or asset loss in a fixed period of time (e.g. 10 days in the Basel FRTB, BCBS (2019, p. 89)); F_X denotes its cumulative distribution function (cdf). In our sign convention, a positive value of $X \in \mathcal{X}$ represents a loss and a negative value represents a surplus. We write $X \overset{\mathrm{d}}{=} Y$ if two random variables X and Y follow the same cdf. For $q \in (0, \infty)$, L^q denotes the set of all random variables with finite q-th moment, L^0 is the set of all random variables, and L^∞ is the set of essentially bounded random variables.

A risk measure is a mapping from $\mathcal{X}$ to $\mathbb{R}$, where $\mathcal{X}$ is a convex cone of random variables representing losses faced by financial institutions, e.g., $\mathcal{X} = L^q$, $q \geqslant 0$. In particular, the Value-at-Risk (VaR) at level $p \in (0, 1)$ is the functional $\mathrm{VaR}_p : L^0 \to \mathbb{R}$ defined by

$$\mathrm{VaR}_p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant p\},$$

which is the left p-quantile of X, and the Expected Shortfall (ES) at level $p \in (0, 1)$ is the functional $\mathrm{ES}_p : L^1 \to \mathbb{R}$ defined by

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

In this paper, terms such as increasing or decreasing functions are in the non-strict sense.

2 An axiomatic characterization of ES

This section contains the main ideas and results of this paper. First, we introduce and discuss four economic axioms for a suitable regulatory risk measure. In particular, one of the axioms, called Axiom NRC, is the most important one that distinguishes ES from all other risk measures; it specifically addresses risk assessment at the portfolio level. Second, we shall see that the four axioms uniquely characterize the class of ES on L^1, and that no meaningful risk measure can satisfy the axioms on L^q for any $q \in [0, 1)$. Finally, we shall discuss some technical aspects of the axioms as well as certain implications of our results.

2.1 Three natural axioms for portfolio risk assessment

Given a risk measure $\rho : \mathcal{X} \to \mathbb{R}$, the risk value $\rho(X)$ is commonly interpreted as the capital requirement associated with loss $X \in \mathcal{X}$. Next are the first three of the four axioms that we propose in this paper, followed by brief comments on their roles in the context of capital requirement calculations. All the random variables that show up in the axioms are implicitly assumed to be in the domain $\mathcal{X}$ of ρ.

Axiom M (Monotonicity)**.** A surely larger or equal loss leads to a larger or equal risk value, that is, $\rho(X) \leqslant \rho(Y)$ whenever $X \leqslant Y$.

Axiom LI (Law invariance)**.** The risk value depends on the loss via its distribution, that is, $\rho(X) = \rho(Y)$ whenever $X \stackrel{\mathrm{d}}{=} Y$.

Axiom P (Prudence)**.** The risk value is not underestimated by approximations, that is, the bound $\lim_{k\to\infty} \rho(\xi_k) \geqslant \rho(X)$ holds whenever $\xi_k \to X$ (pointwise) and the limit $\lim_{k\to\infty} \rho(\xi_k)$ exists.

Although the above three axioms are standard, we nevertheless briefly discuss them, to show their naturalness in the current context. First, Axiom M means that if the loss increases for all scenarios $\omega \in \Omega$, then the capital requirement should increase as well.

Axiom LI is common for all risk measures used in practice, as it gives sense of objectivity by requiring the capital requirement to depend on the loss distribution, but not on the realized loss outcome. Hence, the axiom facilitates statistical modeling and inference, which are distribution-based methods by their very nature.

Finally, Axiom P is precisely lower semi-continuity,[6] which means that if the loss X is modelled using a truthful approximation (e.g., via the empirical distribution or some other consistent

[6]This can be equivalently rewritten as the requirement that $\liminf_{k\to\infty} \rho(\xi_k) \geqslant \rho(X)$ whenever $\xi_k \to X$ (pointwise), which provides an alternative, and perhaps more mathematically concise, reformulation of Axiom P.

estimator), then the approximated risk model should not underreport the capital requirement as the approximation error reduces to zero. Therefore, all three Axioms M, LI, and P are natural requirements for any reasonable risk measure that is used in practice. Both VaR and ES satisfy these axioms (see Proposition 1 in Section 2.4 below).

Axioms M, LI, and P are arguably the most natural for risk assessment in any area involving statistical modeling, such as insurance, engineering, natural catastrophes and econometrics. In other words, the axioms are not specifically designed for portfolio risk assessment, although the latter is the main interpretation of risk measures in the current paper. In the next subsection, we will formulate the fourth axiom, which is specific to the setting of portfolio risk assessment.

Remark 1. The reader may have noticed that our axioms are formulated in the *weakest* possible form, so that they are *minimal* to require. For instance, along the same line of reasoning, it may be intuitive to replace "$X \leqslant Y$" by "$X \leqslant Y$ a.s." in Axiom M, and to replace "$\xi_k \to X$ (pointwise)" by "$\xi_k \to X$ a.s." or "$\xi_k \to X$ in distribution" in Axiom P. These changes will make the axioms stronger and harder to satisfy (thus harder to defend), and the corresponding characterization theorem will generally become weaker. Therefore, we have chosen to present the axioms in the current weakest form. Nevertheless, it does not hurt to keep the axioms in mind with their possible stronger versions as noted above, since our risk measures of interest actually satisfy these stronger versions (shown in the proof of Proposition 1), and all our results in this section still remain valid with the stronger version.

2.2 Perspective of the regulator and the fourth axiom

We recall that in the context of the Basel FRTB (BCBS, 2016, 2019), regulatory risk measures are used to calculate portfolio capital requirements. Below, we summarize two important features (regulatory considerations) of portfolio risk assessment as reflected in the FRTB.

First, regulators are concerned with *tail events*, which are rare events (i.e., have small probabilities) in which risky positions incur large losses. The consideration of "tail risk" is the official reason that the Basel Committee on Banking Supervision has replaced VaR by ES:

> *A shift from Value-at-Risk (VaR) to an Expected Shortfall (ES) measure of risk under stress.* Use of ES will help to ensure a more prudent capture of "tail risk" and capital adequacy during periods of significant financial market stress. (BCBS, 2016, Executive Summary.)[7]

[7]See also BCBS (2019, 30.20, p. 70) on rare events: "Banks' stress scenarios must cover a range of factors that (i) can create extraordinary losses or gains in trading portfolios, or (ii) make the control of risk in those portfolios very

To reflect this, we naturally define a tail event of a random loss X as an event on which X takes its largest values.

Definition 1 (Tail event)**.** A *tail event* of a random variable X is an event $A \in \mathcal{F}$ with $0 < \mathbb{P}(A) < 1$ such that the inequality $X(\omega) \geqslant X(\omega')$ holds for a.s. all scenarios $\omega \in A$ and $\omega' \in A^c$, where A^c stands for the complement of A.

The second feature concerns diversification and risk concentration. We use a random vector $(X_1, \dots, X_n)$ to represent a portfolio with several risky components. If the portfolio $(X_1, \dots, X_n)$ is "properly diversified" (to be defined precisely later), there should be *diversification benefit*, namely, $\rho\left(\sum_{i=1}^n X_i\right) < \sum_{i=1}^n \rho(X_i)$. In other words, the portfolio receives some reduction in capital requirement because of diversification or hedging, if compared to the sum of the risk values of its components.[8] On the other hand, if the portfolio is *concentrated*, or *non-diversified*, then there is no diversification benefit, namely, $\rho\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \rho(X_i)$. This reflects regulators' intention to penalize risk concentration or unjustifiable diversification.[9]

We put the above two features into a rigorous mathematical framework by the following intuitive argument, which identifies a notion of risk concentration that should not yield diversification benefit. Suppose that A is a *stress event* in which the financial system is under severe stress (an adverse economic scenario, e.g., a financial crisis), identified by the regulator. Consider a portfolio of random losses, whose components all share the stress event A as a tail event; in other words, if A happens, all risk factors in the portfolio are realized as their biggest possible losses altogether. Arguably, this is the most problematic type of risk concentration for the regulator, leading to significant negative impact on the financial system. To reflect this discussion, such a portfolio should not be considered as diversified, and no reward should be given. This observation leads us to the fourth and final axiom.

Axiom NRC (No reward for concentration)**.** There exists an event $A \in \mathcal{F}$ such that $\rho(X+Y) = \rho(X) + \rho(Y)$ holds for all risks X and Y sharing the tail event A.

---

difficult. These factors include low-probability events in all major types of risk, [...]".

[8]On the matter, the Basel Committee on Banking Supervision notes the following (BCBS, 2019, 10.22, p. 11): "Diversification: the reduction in risk at a portfolio level due to holding risk positions in different instruments that are not perfectly correlated with one another." The feature of rewarding diversification is also the key notion in the definition of *coherent risk measures* introduced by Artzner et al. (1999).

[9]In particular, the Basel Committee on Banking Supervision specifies (BCBS, 2019, 30.17(3b), p. 70): "[...] with sufficient consideration given to ensuring: [...] that the models reflect concentration risk that may arise in an undiversified portfolio." and (BCBS, 2019, 22.4, p. 54): "No diversification benefit is recognised between the DRC requirements for: (1) non-securitisations; (2) securitisations (non-CTP); and (3) securitisations (CTP)."

As noted above, Axiom NRC intuitively means that a concentrated portfolio,[10] whose components incur large losses simultaneously in the stress event A, does not receive any capital reduction, thus reflecting the two features of portfolio risk assessment discussed above. The stress event A in Axiom NRC should be interpreted as a symbolic event of regulatory interest. As we shall see later, as long as it exists, which event A is chosen as the stress event is not relevant for the characterization of the desirable risk measure.[11] Moreover, Axiom NRC is a model-free conceptual requirement, as it does not depend on the choice of the probability measure $\mathbb{P}$.[12] This allows for flexibility in the statistical inference of risk models in practice, which comes after the regulatory risk measure is decided.

Quite obviously, a linear functional, such as the expectation $\mathbb{E} : L^1 \to \mathbb{R}$ or the conditional expectation $\mathbb{E}[\cdot|B]$ for some $B \in \mathcal{F}$, satisfies Axiom NRC. Later we shall see that ES also satisfies Axiom NRC, whereas VaR does not. Furthermore, the four axioms that we have introduced are independent (Proposition 3).

Remark 2. Axiom NRC does not specify what happens to $\rho(X+Y)$ if X and Y do not share the tail event A. A reasonable condition for this situation may be to impose $\rho(X+Y) \leqslant \rho(X)+\rho(Y)$ for all (X,Y), which is a requirement for ρ to be a coherent risk measure in Artzner et al. (1999). We do not impose this requirement to formulate our axiom in its weakest form, as explained in Remark 1. However we will see that, with our four axioms, we indeed arrive at coherent risk measures (more precisely, at ES), and hence this extra requirement is satisfied automatically.

2.3 Characterization of ES

The next theorem, which is the main result of this paper, shows that Axioms M, LI, P and NRC characterize the class of ES on $\mathcal{X} = L^1$.

Theorem 1. *A functional $\rho : L^1 \to \mathbb{R}$ with $\rho(1) = 1$ satisfies Axioms M, LI, P and NRC if and only if $\rho = \mathrm{ES}_p$ for some $p \in (0,1)$. Moreover, in the forward direction of the above statement, the value of p is uniquely given by $1 - \mathbb{P}(A)$, where A is any stress event in Axiom NRC.*

A few clarifying remarks follow. First, as implied by the last statement of Theorem 1, although the stress event A in Axiom NRC is generally not unique, its probability is. Second, the requirement

[10] Axiom NRC can be equivalently reformulated using portfolios with n components (Proposition 2 below) thus facilitating practical relevance of the axiom, but again we choose its weakest form with $n = 2$, which is undoubtedly much more convenient to verify mathematically.

[11] Technically, this is due to Axiom LI. As we will see in Theorem 1, only the probability of A is relevant.

[12] More precisely, $\mathbb{P}$ can be replaced by any other equivalent probability measure. Axioms M and P do not depend on $\mathbb{P}$, either.

$\rho(1) = 1$ is simply for the sake of normalizing the functional; without it, we would have $\rho = \lambda \mathrm{ES}_p$ for some $\lambda \geqslant 0$. Third, in contrast to a vast body of literature on risk measures (e.g., Delbaen, 2012; Föllmer and Schied, 2016, and the references therein), in Theorem 1 we do not impose subadditivity or convexity on the risk measure ρ; nevertheless, we arrive at the coherent risk measure ES_p.[13] Finally, in Theorem 1 we specify the set $\mathcal{X} = L^1$ as the domain of risk measures, but we may wonder whether this choice of domain is too restrictive, noting that the above axioms, as well as the concept of tail events, do not rely on integrability. The next theorem shows that the domain L^1 is the most natural (and essentially the largest) choice for any risk measure under Axioms M, LI, P and NRC. A similar result on the domain of convex risk measures is obtained by Delbaen (2007).

Theorem 2. *Let $q \in [0, 1)$. A functional $\rho : L^q \to \mathbb{R}$ satisfies Axioms M, LI, P and NRC if and only if $\rho(X) = 0$ for every $X \in L^q$.*

In other words, Theorem 2 says that no meaningful risk measure satisfying Axioms M, LI, P and NRC is well defined on any L^q space larger than L^1.[14] The restriction to L^1 is certainly not a problem for portfolio risk assessment, as ample empirical evidence shows that portfolio losses have a finite mean (and typically also a finite variance); see McNeil et al. (2015, Chapter 3).

2.4 Technical remarks on the axioms and the ES characterization

In this section, we discuss several technical aspects related to the four axioms that we have proposed above. First, as mentioned previously, both ES_p and VaR_p satisfy the first three axioms.[15]

Proposition 1. *For $p \in (0, 1)$, the two functionals ES_p and VaR_p on $\mathcal{X} = L^1$ satisfy Axioms M, LI, and P.*

The next proposition shows that Axiom NRC can be equivalently reformulated using any number n of losses in a portfolio. While the case $n = 2$ facilitates the verification of the axiom, knowing that it actually holds for every n makes the axiom appealing from the practical point of view, as real portfolios are comprised of many assets.

[13] By definition, a coherent risk measure is monotone, translation invariant, positively homogeneous, and subadditive (hence convex); see Artzner et al. (1999). For every $p \in (0, 1)$, ES_p is a coherent risk measure.

[14] On the other hand, for a smaller domain L^q with $q \in (1, \infty]$, the characterization in Theorem 1 remains true with the same proof.

[15] Convex risk measures such as ES also satisfy other types of continuity properties; see e.g., Gao et al. (2018). The lower semi-continuity in Axiom P is with respect to point-wise (or a.s.) convergence, and it is different from the continuity properties commonly studied in the risk measure literature.

Proposition 2. *Let A be any event. For a functional $\rho : \mathcal{X} \to \mathbb{R}$, if $\rho(X+Y) = \rho(X) + \rho(Y)$ holds for all X and Y sharing the tail event A, then for every $n \in \mathbb{N}$, $\rho(\sum_{i=1}^n X_i) = \sum_{i=1}^n \rho(X_i)$ holds for all $X_1, \dots, X_n$ sharing the tail event A.*

Next, we show that the four axioms M, LI, P and NRC are independent, so that none of them is redundant in the characterization result of Theorem 1.

Proposition 3. *For a risk measure $\rho : L^1 \to \mathbb{R}$, any combination of three of Axioms M, LI, P and NRC does not imply the remaining fourth axiom.*

In the following last proposition of this section, we show that if Axiom LI holds, then Axiom NRC can be equivalently formulated via a (stronger) functional property, which we shall simply call *p-additivity*.

Proposition 4. *Suppose $q \in [0, \infty]$ and a functional $\rho : L^q \to \mathbb{R}$ satisfies Axiom LI. Then ρ satisfies Axiom NRC if and only for some $p \in (0, 1)$, ρ is p-additive: $\rho(X + Y) = \rho(X) + \rho(Y)$ for all X and Y sharing a tail event of probability $1 - p$.*

In Section 3, we will formally study the notion of dependence in which the components of a random vector share a common tail event of a specific probability as used in Proposition 4. The new dependence notion and p-additivity in Proposition 4 will become useful technical tools, eventually leading to the proofs of Theorems 1 and 2.

Remark 3. We can compare all of our axioms, and in particular Axiom NRC, with the sets of axioms in the literature that have characterized VaR. In Chambers (2009) and He and Peng (2018), the main axioms for VaR are invariance properties of some risk transforms. In Kou and Peng (2016) and Liu and Wang (2020), the main axioms for VaR rely on the statistical notion of elicitability. To summarize, our axiom NRC is mainly based on the perspective of portfolio risk aggregation, whereas the key axioms for VaR in the literature are mainly based on statistical or mathematical advantages of VaR as a quantile. Therefore, these axioms reflect different considerations, and they suggest that VaR and ES may have their own advantages in different applications.

3 The dependence notion of risk concentration

3.1 Definition and technical properties

We have already seen the pivotal role that risk concentration plays in characterizing ES via Axiom NRC. Moreover, in Proposition 4, a special notion of risk concentration appears, in which random losses share a common tail event of probability $1-p \in (0,1)$. This notion serves as a key step in the proofs of our main theorems in Section 2. We formally define the notion under the name of p-concentration, and we present a few results on verifying and characterizing p-concentration in this section.

Definition 2 (Risk concentration)**.** For $p \in (0,1)$, a random vector $(X_1, \dots, X_n)$ is *p-concentrated* if its components share a common tail event of probability $1-p$. We also call a tail event of probability $1-p$ a *p-tail event*.

The terminology that a p-tail event has probability $1-p$ stems from the regulatory language where e.g., a tail event with probability 1% corresponds to the calculation of a 99% VaR.

We briefly explain some simple facts about p-tail events. For any random variable and $p \in (0,1)$, a p-tail event of any specific probability always exists (Lemma A.3), but it may not be unique (Example A.1), unless X is continuously distributed, in which case a tail event is a.s. unique (Corollary A.1); these results and claims are collected in Appendix A.2.1.

To better understand and to appropriately position p-concentration among dependence concepts, we next recall the classical notion of comonotonicity.

Definition 3 (Comonotonicity)**.** A random vector $(X_1, \dots, X_n)$ is *comonotonic* if there exists a random variable Z and increasing functions $f_1, \dots, f_n$ on $\mathbb{R}$ such that $X_i = f_i(Z)$ a.s. for every $i = 1, \dots, n$.

The following proposition characterizes p-concentration, which can be seen as an analogue to Denneberg's characterization of comonotonicity (Denneberg, 1994). For the notion of copulas, we refer to Nelsen (2006).

Theorem 3. *Let $p \in (0,1)$, and let $(X_1, \dots, X_n)$ be any random vector. The following statements are equivalent, where $S = X_1 + \cdots + X_n$:*

(i) *$(X_1, \dots, X_n)$ is p-concentrated;*

(ii) *$(X_1, \dots, X_n, S)$ is p-concentrated;*

(iii) *$(X_i, S - X_i)$ is p-concentrated for every $i = 1, \dots, n$;*

(iv) *$(f_1(X_1), \dots, f_n(X_n))$ is p-concentrated for all increasing functions $f_1, \dots, f_n$;*

(v) *there is a copula C of $(X_1, \dots, X_n)$ that satisfies $C(p, \dots, p) = p$.*

The implication (iii)⇒(i) is quite remarkable, as it establishes p-concentration of the vector $(X_1, \dots, X_n)$ from p-concentration of the pair $(X_i, S - X_i)$ without any continuity assumption.

Remark 4. Generally, the copula of a p-concentrated random vector $(X_1, \dots, X_n)$ is not necessarily unique, unless the random vector has continuously marginal distributions, as implied by Sklar's Theorem (e.g., Nelsen, 2006). For instance, any copula is a copula of a constant vector $(X_1, \dots, X_n)$, which is p-concentrated. This explains why the property in statement (v) of Theorem 3 holds for some copula of $(X_1, \dots, X_n)$ but not for all copulas.

By its very nature, p-concentration is defined for arbitrary but fixed $p \in (0, 1)$, which is exactly what is needed to characterize ES_p. Nevertheless, it is quite interesting to observe that if a random vector $(X_1, \dots, X_n)$ is p-concentrated simultaneously for all $p \in (0, 1)$, then the vector is comonotonic. This observation relates our present research to the vast literature on comonotonicity, and we therefore formulate it as our next theorem.

Theorem 4. *A random vector $(X_1, \dots, X_n)$ is p-concentrated for all $p \in (0, 1)$ if and only if it is comonotonic.*

In case the vector $(X_1, \dots, X_n)$ has continuous marginal distributions, Theorems 3–4 are naturally connected via the following fact: $(X_1, \dots, X_n)$ is comonotonic if and only if its copula C satisfies $C(u_1, \dots, u_n) = \min(u_1, \dots, u_n)$ (e.g., Dhaene et al., 2002, Theorem 2). Note that the latter condition is equivalent to $C(p, \dots, p) = p$ for all $p \in (0, 1)$.

For a random vector, comonotonicity is equivalent to pairwise comonotonicity of its components (e.g., Dhaene et al., 2002, Theorem 3). Given the connection between p-concentration and comonotonicity, one naturally asks whether this also holds for p-concentration. Obviously, if $(X_1, \dots, X_n)$ is p-concentrated, then each pair (X_i, X_j) is p-concentrated. The converse, however, is generally false (Example A.2), which is in sharp contrast to the case of comonotonicity. Nevertheless, when at least one of the random variables is continuously distributed, p-concentration becomes equivalent to pairwise p-concentration (Proposition A.1). These observations, as well as several other auxiliary results on tail events and p-concentration, are relegated to Appendix A.2.1.

3.2 The role of p-concentration in the comparison of risk measures

We next discuss the role of p-concentration in Axiom NRC when comparing risk measures. Recall that, by Theorem 4, p-concentration is a weaker notion than comonotonicity.

Theorem 1 distinguishes ES from other subadditive risk measures,[16] such as the standard deviation,[17] the MINVAR (or MAXVAR) in Cherny and Madan (2009), the expectiles in Newey

[16] Recall that subadditivity reflects rewarding diversification. Coherent risk measures are subadditive.

[17] Although the standard deviation is not monotone, it is commonly regarded as a metric of diversification.

and Powell (1987) and Bellini et al. (2014), and the Gini Shortfalls in Furman et al. (2017). The most important distinction between these risk measures and the current paper is Axiom NRC. Note that for a subadditive risk measure ρ, the inequality

$$\rho\left(\sum_{i=1}^{n} X_i\right) \leqslant \sum_{i=1}^{n} \rho(X_i) \tag{1}$$

holds for every portfolio $(X_1, \ldots, X_n)$. However, for a fixed p, in contrast to ES_p, inequality (1) is usually a strict inequality even when $(X_1, \ldots, X_n)$ is p-concentrated. More precisely, we can check (excluding some cases leading to $\mathbb{E}$ or ES_p) that for bound (1) to be an equality, it is necessary and sufficient for $(X_1, \ldots, X_n)$ to be:

(a) positively and linearly dependent if ρ is the standard deviation;

(b) comonotonic if ρ is a MINVAR;

(c) p-concentrated and comonotonic on a common tail event if ρ is a coherent Gini Shortfall;

(d) p-concentrated if ρ is ES (Theorem 5).

Therefore, the above risk measures as well as other coherent risk measures except ES yield capital reduction even if large losses in a portfolio occur together, and they typically penalize risk concentration when dependence is more extreme, like in the cases of comonotonicity or linear dependence.[18] This gives a unique advantage for ES when assessing portfolio risk.

Remark 5. The dependence structure of p-concentration may not appear to be very common in real financial data, or classic statistical models, except for some conditional models such as losses conditional on default events (nevertheless, a brief real-data example of p-concentration is presented in Section 5). This is not surprising, because the new notion represents extremely dangerous portfolios, and it is used for conceptual exercises and logical considerations. Hypothetical portfolios are actually very helpful to understand desirable properties of risk measures. For instance, comonotonicity is stronger than p-concentration and hence even less likely to be seen in real data, and yet it is a very popular notion in the axiomatic characterization of risk functionals and preferences (e.g., Yaari, 1987; Schmeidler, 1989; Kusuoka, 2001; Marinacci and Montrucchio, 2004).

[18] A necessary and sufficient condition for (1) to be an equality for coherent expectiles is more complicated and we omit the discussion here.

4 Risk aggregation for ES and VaR

In the proof of Theorem 1, we shall need to verify Axiom NRC and thus p-additivity of ES_p. The latter property is closely related to the maximization of risk aggregation, whose definition we recall next.

Definition 4. Given any $p \in (0,1)$, a random vector $(X_1, \dots, X_n)$ with all the components in L^1 is said to maximize the ES_p aggregation if

$$\mathrm{ES}_p(X_1 + \cdots + X_n) = \max \left\{ \mathrm{ES}_p(X_1' + \cdots + X_n') : X_i' \overset{\mathrm{d}}{=} X_i,\ i = 1, \dots, n \right\}.$$

Risk aggregation of ES and VaR with specified marginal distributions has been extensively studied in the quantitative finance literature (e.g., Wang et al., 2013; Embrechts et al., 2013, 2015). It is known (e.g., McNeil et al., 2015, Section 8.4.4) that if $(X_1, \dots, X_n)$ is comonotonic, then it maximizes the ES_p aggregation. However, comonotonicity is not necessary. Indeed, a necessary and sufficient condition in the case of two continuously distributed random variables was derived by Wang and Zitikis (2020). By generalizing the latter fact to arbitrary dimensions as well as to arbitrary marginal distributions, the next theorem establishes a unique role of p-concentration in risk aggregation for ES.

Theorem 5. *Let $p \in (0,1)$, and let $(X_1, \dots, X_n)$ be any random vector with all its components in L^1. The following statements are equivalent:*

(i) $(X_1, \dots, X_n)$ *is p-concentrated;*

(ii) $(X_1, \dots, X_n)$ *maximizes the ES_p aggregation;*

(iii) $\mathrm{ES}_p\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{ES}_p(X_i)$.

As implied by Theorem 5, ES_p is additive for and only for p-concentrated portfolios, and it satisfies Axiom NRC by Proposition 4. We shall next show that this is generally not the case for VaR_p, although VaR_p is actually *subadditive* for p-concentrated random vectors. To equip the next theorem with generality, in addition to the left p-quantile VaR_p, we also use the right p-quantile

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

Theorem 6. *Let $p \in (0,1)$, and let $(X_1, \dots, X_n)$ be any p-concentrated random vector. We have* *the inequalities*

$$\mathrm{VaR}_p\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \mathrm{VaR}_p(X_i) \leqslant \sum_{i=1}^n \mathrm{VaR}_p^+(X_i) \leqslant \mathrm{VaR}_p^+\left(\sum_{i=1}^n X_i\right). \tag{2}$$

Each inequality in (2) *is generally not an equality, and thus* VaR_p *is not* p*-additive.*

If the quantile function of the aggregate loss $\sum_{i=1}^n X_i$ is continuous at p, then all the inequalities in (2) are equalities. However, this is generally not the case. Namely, for a p-concentrated random vector $(X_1, \ldots, X_n)$, even when all of its marginal quantiles are continuous, the quantile of the sum $\sum_{i=1}^n X_i$ may not be continuous. To see this, we next give an explicit example that illustrates that VaR_p may not satisfy Axiom NRC. This observation will be particularly useful when proving the main characterization result (i.e., Theorem 1) of this paper.

Example 1. Let $U \sim \mathrm{U}[0, 1]$ be a uniform random variable on $[0, 1]$. Fix $p \in (0, 1)$ and let A be an event of probability $\mathbb{P}(A) = 1 - p$ and independent of U. Define $X = U\mathbb{1}_{A^c} + \mathbb{1}_A$ and $Y = (1-U)\mathbb{1}_{A^c}+\mathbb{1}_A$. Clearly, A is a common p-tail event of X and Y. We check $\mathrm{VaR}_p(X) = \mathrm{VaR}_p(Y) = 1$ and $\mathrm{VaR}_p(X + Y) = \mathrm{VaR}_p(\mathbb{1}_{A^c} + 2\mathbb{1}_A) = 1$, which imply $\mathrm{VaR}_p(X + Y) < \mathrm{VaR}_p(X) + \mathrm{VaR}_p(Y)$. Hence, VaR_p is not p-additive.

Remark 6. The left and right p-quantiles (i.e., VaR_p and VaR_p^+, respectively) differ only when the quantile function $\phi(p) = \mathrm{VaR}_p(X)$ of loss X has a jump at $p \in (0, 1)$. It is known (e.g., Embrechts et al., 2014) that the left and right p-quantiles are additive for every comonotonic random vector. Since p-concentration is weaker than comonotonicity, the quantiles are no longer additive for p-concentrated random vectors. Nevertheless, we see from Theorem 6 that for every p-concentrated random vector, the left p-quantile VaR_p is subadditive and the right p-quantile VaR_p^+ is superadditive.

5 Concluding remarks

In this paper, we have proposed four intuitive axioms for portfolio risk assessment, which jointly characterize the family of ES. The first three axioms, which are not necessarily specific to portfolio risk assessment, are very simple and should be satisfied by every risk measure used in practice. The fourth axiom, on the other hand, is precisely designed for the purpose of capital requirement calculation, within the context of the Basel FRTB.

There have been extensive studies in the literature on comparative advantages of VaR or ES in risk management (e.g., Embrechts et al., 2014; Emmer et al., 2015). In particular, it is argued that VaR has several statistical advantages (e.g., Cont et al., 2010; Gneiting, 2011; Kou and Peng, 2016; Fissler and Ziegel, 2016), whereas ES has advantages in optimization, risk aggregation, and capturing tail risk (e.g., Krätschmer et al., 2014; Embrechts et al., 2015, 2018). It is not, however, the intention of the current paper to jump into a conclusion on which risk measure is better. Nevertheless, our results yield further insights into this important matter, which is relevant for regulators, practitioners and academics. Our main goal is – via the four economic axioms – to show that ES has a unique role in the theory of risk measures, specifically in the context of portfolio capital calculation.

As to the question of whether (and where) ES is indeed the best risk measure to use, the answer surely depends on whether Axiom NRC is the most desirable to require, since the other three axioms are satisfied by both VaR and ES. In the context of banking regulation, and in the spirit of the Basel FRTB features, Axiom NRC seems to fit most naturally: it reflects the regulators' focal consideration of risk concentration and stress events, and at the same time, provides substantial simplicity, tractability and interpretability of the corresponding portfolios and risk measures. Thus, we believe that it is no coincidence that ES is the only coherent risk measure used in current financial regulation.

As discussed in Remark 5, the dependence structure of p-concentration is not common for financial data, and this is not a disadvantage of our theory. Nevertheless, we briefly present a simple empirical example in Figure 1 below, where we plot the daily return/loss data of S&P 500, NASDAQ and the Dow Jones Index in 2009.[19] As we see from the figure, largest losses of the indices occur simultaneously, which constitute a common tail event of probability $1 - p = 6/252 = 2.38\%$ (close to the probability level 2.5% chosen in the FTRB). Thus, a portfolio of these indices (e.g., via index ETFs) is empirically p-concentrated, although not comonotonic.[20] By Theorem 5, using ES_p calibrated to empirical losses in 2009, there is no diversification benefit awarded to this portfolio.[21] More importantly, Theorem 1 offers the powerful converse direction: If one insists no diversification benefit for portfolios that behave like the above one (i.e., big losses with probability $1 - p$ occuring simultaneously), then ES_p is *the only risk measure* to use, assuming the other three natural axioms.

[19]There are 252 daily data points for each index in 2009, publicly available from Yahoo Finance.

[20]We also tried similar data analysis for other periods of time, and the phenomenon of p-concentration is sometimes approximately observed, but not always. The example here is only to show that p-concentration may be empirically observed for a highly correlated portfolio, and a comprehensive empirical study is not the focus of this paper.

[21]Even for market risk evaluation today, the empirical distribution of losses in 2009 is practically important since historical observations dating back to the financial crisis must be used for stressed ES calibration as specified in the FRTB. Quoting BCBS (2019, 33.7, p.90): "The observation horizon for determining the most stressful 12 months must, at a minimum, span back to and include 2007."

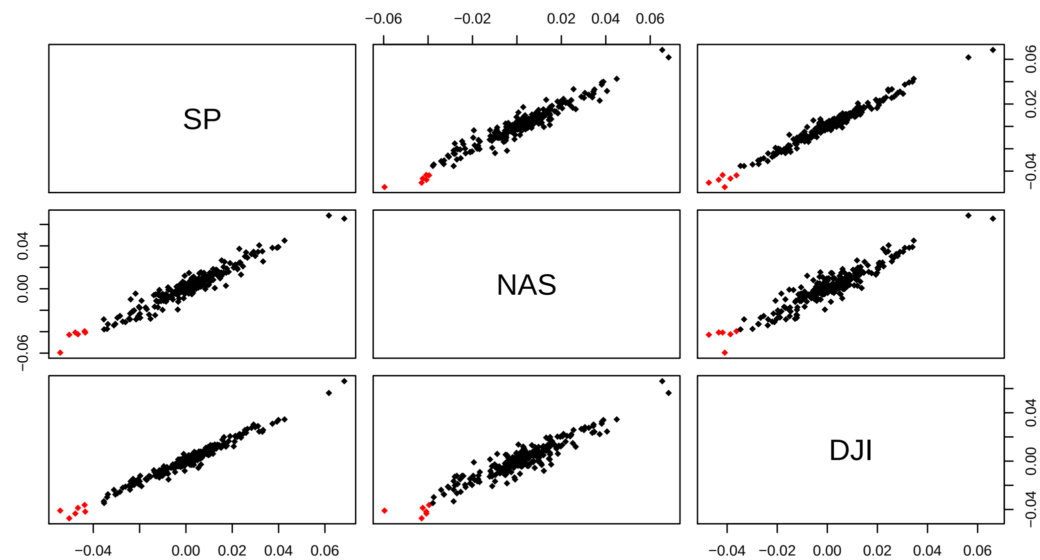

Figure 1: S&P 500, NASDAQ and the Dow Jones Index daily returns in 2009

A Technical appendices

A.1 Proofs of results of Section 2

In this appendix, we prove the main results of Section 2, which are Theorems 1–2 and Propositions 1–4. Some parts of the proofs need technical results of Sections 3–4 (proofs of results in Sections 3–4 are self-contained and do not rely on those in Section 2). With a focus on the characterization results, we choose to start with the proofs of Theorems 1–2. For this purpose, we need the properties in Proposition 1 and 4 as well as later results in Theorems 5–6 which verify that ES satisfies Axiom NRC and VaR does not. As for Propositions 1–4, the proof of Proposition 2 requires Lemma A.5 and that of Proposition 3 requires Theorems 5–6.[22]

A.1.1 Characterization of p-additive functionals

First, we present two additional results to prepare for the proofs of Theorems 1–2. Specifically, Lemma A.1 identifies the form of monotone, law invariant, and linear functionals on L^1_+, which is the space of all non-negative and integrable random variables. Lemma A.2 identifies the form of monotone, law-invariant, and p-additive functionals on L^1. Recall that a functional ρ is called *additive* if $\rho(X+Y)=\rho(X)+\rho(Y)$ for all X and Y in the domain of ρ.

[22]The reader who is interested in the full mathematical development may read the appendices in the following order, which does not involve any references to later results: A.2⇒A.3⇒A.1.3⇒A.1.1⇒A.1.2.

Lemma A.1. *A functional $\rho : L^1_+ \to \mathbb{R}$ is monotone, additive, and law invariant if and only if $\rho = \lambda\mathbb{E}$ on L^1_+ for some $\lambda \geqslant 0$.*

Proof. Part ($\Leftarrow$) is trivial to check, and we thus only prove part ($\Rightarrow$). Let $\lambda = \rho(1)$. Standard arguments based on additivity and monotonicity imply $\rho(x) = \lambda x$ for all $x \geqslant 0$. Define $\widehat{\rho} : L^1 \to \mathbb{R}$ by

$$\widehat{\rho}(X) = \rho(X_+) - \rho(X_-),$$

where $x_+ = \max\{x, 0\}$ and $x_- = \max\{-x, 0\}$ for every $x \in \mathbb{R}$. Since both X_+ and X_- are non-negative, $\widehat{\rho}$ is well defined. If $X \in L^1_+$, then $X_+ = X$ and $X_- = 0$, and $\widehat{\rho}(X) = \rho(X) - \rho(0) = \rho(X)$. Hence, $\widehat{\rho} = \rho$ on L^1_+. It is easy to check that the functional $\widehat{\rho}$ is monotone and law invariant, because ρ is such.

Let $X, Y \in L^1$ and $Z := X + Y$. From $X + Y - Z = 0$, we have the equation $X_+ + Y_+ - Z_+ = X_- + Y_- - Z_-$. Since $X_+ + Y_+ - Z_+ \geqslant 0$ and ρ is additive on L^1_+, we have

$$\rho(X_+) + \rho(Y_+) = \rho(X_+ + Y_+) = \rho(X_+ + Y_+ - Z_+) + \rho(Z_+). \tag{3}$$

Similarly, we have

$$\rho(X_-) + \rho(Y_-) = \rho(X_- + Y_-) = \rho(X_- + Y_- - Z_-) + \rho(Z_-). \tag{4}$$

Since $\rho(X_+ + Y_+ - Z_+) = \rho(X_- + Y_- - Z_-)$, equations (3)–(4) imply

$$\widehat{\rho}(Z) = \rho(Z_+) - \rho(Z_-) = \rho(X_+) + \rho(Y_+) - \rho(X_-) - \rho(Y_-) = \widehat{\rho}(X) + \widehat{\rho}(Y).$$

Therefore, $\widehat{\rho}$ is additive on L^1.

Furthermore, $\widehat{\rho}$ is a finite coherent risk measure (multiplied by a positive constant) on L^1, and hence is L^1-continuous (e.g., Rüschendorf, 2013, Corollary 7.10). Since $\widehat{\rho}$ is L^1-continuous and linear, the risk measure has the representation (e.g., Rüschendorf, 2013, Theorem 7.20)

$$\widehat{\rho}(X) = \int X \mathrm{d}Q$$

for a measure Q on $(\Omega, \mathcal{F})$. Since $\widehat{\rho}$ is law invariant, Q has to be equal to $\mathbb{P}$ multiplied by a constant. The constant must be $\widehat{\rho}(1)$, and since $0 \leqslant \widehat{\rho}(1) = \rho(1) = \lambda$, we arrive at $\widehat{\rho}(X) = \lambda\mathbb{E}[X]$ for all $X \in L^1$. Because $\widehat{\rho} = \rho$ on L^1_+, we have $\rho = \lambda\mathbb{E}$ on L^1_+. This concludes the proof of Lemma A.1. □

The following lemma is the main result of this appendix and plays a pivotal role in establishing Theorems 1–2 in the next appendix. Recall that a functional ρ is called p-additive if $\rho(X+Y) = \rho(X)+\rho(Y)$ for all X and Y sharing a tail event of probability $1-p$.

Lemma A.2. *Let $p \in (0,1)$. A monotone functional $\rho : L^1 \to \mathbb{R}$ is law invariant and p-additive if and only if $\rho = a\mathbb{E} + b\mathrm{ES}_p$ on L^1 for some $a, b \in \mathbb{R}$.*

Proof. ($\Leftarrow$): The proof follows by noting that the functionals $\mathbb{E}$ and ES_p are monotone, law invariant, and p-additive (Theorem 5). Hence, such is their linear combination $\rho = a\mathbb{E} + b\mathrm{ES}_p$.

($\Rightarrow$): Let $U \sim \mathrm{U}[0,1]$, and let A be an event that has probability $\mathbb{P}(A) = 1-p$ and is independent of U. Define $\phi : L^1_+ \to \mathbb{R}$ by

$$\phi(X) = \rho(F_X^{-1}(U)\mathbb{1}_A).$$

The functional ϕ is monotone and law invariant, since ρ is such. For any two random variables $X, Y \in L^1_+$, take $(\tilde{X}, \tilde{Y}) \overset{\mathrm{d}}{=} (X, Y)$ such that $(\tilde{X}, \tilde{Y})$ is independent of A. Since $\tilde{X}\mathbb{1}_A$ and $\tilde{Y}\mathbb{1}_A$ share the common p-tail event A, we know that $(\tilde{X}\mathbb{1}_A, \tilde{Y}\mathbb{1}_A)$ is p-concentrated. As a consequence, $\rho(\tilde{X}\mathbb{1}_A)+\rho(\tilde{Y}\mathbb{1}_A) = \rho((\tilde{X}+\tilde{Y})\mathbb{1}_A)$. Moreover, noting that $F_Z^{-1}(U)\mathbb{1}_A \overset{\mathrm{d}}{=} Z\mathbb{1}_A$ for $Z \in \{\tilde{X}, \tilde{Y}, \tilde{X}+\tilde{Y}\}$, we have

$$\phi(X) + \phi(Y) = \rho(\tilde{X}\mathbb{1}_A) + \rho(\tilde{Y}\mathbb{1}_A) = \rho((\tilde{X}+\tilde{Y})\mathbb{1}_A) = \phi(X+Y).$$

That is, the functional ϕ is additive on L^1_+. Using Lemma A.1, ϕ has to be equal to $\lambda\mathbb{E}$ for some $\lambda \geqslant 0$. This gives the representation

$$\rho(F_X^{-1}(U)\mathbb{1}_A) = \lambda\mathbb{E}[X], \quad X \in L^1_+. \tag{5}$$

For any random variable $X \in L^1$, let B be a p-tail event of X. With the notation $x_p = \mathrm{VaR}_p(X)$, we see that $X \geqslant x_p$ on B and $X \leqslant x_p$ on B^c. Hence, the random variables $(X-x_p)\mathbb{1}_B$ and $(X-x_p)\mathbb{1}_{B^c}$ share the common p-tail event B, and so the vector $((X-x_p)\mathbb{1}_B, (X-x_p)\mathbb{1}_{B^c})$ is p-concentrated. Moreover, with Y denoting a random variable with the distribution of X conditional on B, that is, $\mathbb{P}(Y \leqslant x) = \mathbb{P}(X \leqslant x|B)$ for all $x \in \mathbb{R}$, we have $\mathbb{E}[Y] = \mathrm{ES}_p(X)$ and

$$(X-x_p)\mathbb{1}_B \overset{\mathrm{d}}{=} (F_Y^{-1}(U) - x_p)\mathbb{1}_A.$$

Combining these facts with equation (5), we have

$$\rho((X-x_p)\mathbb{1}_B) = \rho((F_Y^{-1}(U)-x_p)\mathbb{1}_A) = \lambda\mathbb{E}[Y-x_p] = \lambda(\mathrm{ES}_p(X) - x_p).$$

Analogously, we have

$$\rho((X - x_p)\mathbb{1}_{B^c}) = \gamma(\mathrm{ES}_p^-(X) - x_p)$$

for some $\gamma \geqslant 0$, where the functional ES_p^-, called the Left-ES risk measure (e.g., Embrechts et al., 2015), is defined by

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

for all $Z \in L^1$. Note that

$$p\mathrm{ES}_p^-(Z) + (1-p)\mathrm{ES}_p(Z) = \int_0^1 \mathrm{VaR}_q(Z)\mathrm{d}q = \mathbb{E}[Z].$$

From p-additivity, which implies additivity for constants, we infer that there is a constant κ such that $\rho(x) = \kappa x$ for all $x \in \mathbb{R}$. Putting these observations together, we get the equations

$$\begin{aligned}
\rho(X) &= \rho((X - x_p) + x_p) \\
&= \rho((X - x_p)\mathbb{1}_B + (X - x_p)\mathbb{1}_{B^c}) + \rho(x_p) \\
&= \rho((X - x_p)\mathbb{1}_B) + \rho((X - x_p)\mathbb{1}_{B^c}) + \kappa x_p \\
&= \lambda(\mathrm{ES}_p(X) - x_p) + \gamma(\mathrm{ES}_p^-(X) - x_p) + \kappa x_p \\
&= \lambda\mathrm{ES}_p(X) + \gamma\mathrm{ES}_p^-(X) + (\kappa - \lambda - \gamma)x_p
\end{aligned}$$

for all $X \in L^1$, and thus, in turn,

$$\rho(X) = \lambda\mathrm{ES}_p(X) + \gamma\mathrm{ES}_p^-(X) + c\mathrm{VaR}_p(X) \tag{6}$$

for some constants $\lambda, \gamma, c \in \mathbb{R}$. Using $p\mathrm{ES}_p^- + (1-p)\mathrm{ES}_p = \mathbb{E}$ and substituting $a = \gamma/p$ and $b = \lambda - a(1-p)$, equation (6) turns into

$$\rho = \lambda\mathrm{ES}_p + \frac{\gamma}{p}(\mathbb{E} - (1-p)\mathrm{ES}_p) + c\mathrm{VaR}_p = a\mathbb{E} + b\mathrm{ES}_p + c\mathrm{VaR}_p.$$

By Theorem 5, the functional $\lambda\mathrm{ES}_p$ is p-additive. The expectation $\mathbb{E}$ is also p-additive because it is linear. On the other hand, we have seen from Theorem 6 that VaR_p is not p-additive. Therefore, the constant c must be zero, and so $\rho = a\mathbb{E} + b\mathrm{ES}_p$. This concludes the proof of Lemma A.2. □

A.1.2 Proofs of Theorems 1–2

With Lemma A.2 we are ready to prove Theorem 1, which will in turn lead to the proof of Theorem 2.

Proof of Theorem 1. From Proposition 4 and its proof, we know that, in the presence of Axiom LI, Axiom NRC is equivalent to p-additivity for some $p \in (0,1)$, and $p = 1-\mathbb{P}(A)$ where A is any stress event in Axiom NRC. Hence, it suffices to show that a risk measure is monotone, law invariant, prudent, and p-additive if and only if it is ES_p.

($\Leftarrow$): By Proposition 1 and Theorem 5, the functional ES_p is monotone, law invariant, prudent, and p-additive.

($\Rightarrow$): By Lemma A.2, we know that $\rho = a\mathbb{E} + b\mathrm{ES}_p$ for some $a, b \in \mathbb{R}$. Next, we show that $a = 0$ and $b = 1$. Since $\rho(1) = 1$, we must have $a + b = 1$. If $a < 0$ and $b > 1$, then the functional $a\mathbb{E} + b\mathrm{ES}_p$ is not monotone. To see this, let $X = -\mathbb{1}_A$ for some event A with $\mathbb{P}(A) = p$. Then $\mathbb{E}[X] = -p < 0$ and $\mathrm{ES}_p(X) = 0$. Hence, $a\mathbb{E}[X] + b\mathrm{ES}_p(X) > 0$ but $X \leqslant 0$, and so $a\mathbb{E} + b\mathrm{ES}_p$ cannot be monotone.

Consider now the case $a > 0$. Let $\xi_k = -k\mathbb{1}_{\{U<1/k\}}$, where $U \sim \mathrm{U}[0,1]$. Clearly, $\xi_k \to 0$ almost surely, $\mathbb{E}[\xi_k] = -1$ and $\mathrm{ES}_p(\xi_k) = 0$ for $k > 1/p$. Therefore, $\liminf_{k\to\infty} a\mathbb{E}[\xi_k] + \mathrm{ES}_p(\xi_k) = -a < 0 = \rho(0)$, contradicting prudence.

Hence, $a = 0$ is the only value left, and since $a + b = 1$, we must have $b = 1$. This gives the representation $\rho = \mathrm{ES}_p$ on L^1 and finishes the proof of Theorem 1. □

Proof of Theorem 2. Let $\rho : L^q \to \mathbb{R}$ satisfy Axioms M, LI, P and NRC. Using the result of Theorem 1, we know that when constrained on $L^1 \subset L^q$, the equation $\rho = \lambda\mathrm{ES}_p$ holds with $\lambda = \rho(1) \geqslant 0$.

Let $\widehat{L}^q = L^q \setminus L^1$, and let $\widehat{L}^q_+$ be the set of all non-negative random variables in $\widehat{L}^q$. Both $\widehat{L}^q$ and $\widehat{L}^q_+$ are non-empty. Take $X \in \widehat{L}^q_+$ and let $\xi_k = \min\{X, k\}$ for $k \in \mathbb{N}$. Clearly, $\xi_k \uparrow X$ almost surely, and $\mathrm{ES}_p(\xi_k) \geqslant \mathbb{E}[\xi_k] \to \mathbb{E}[X] = \infty$. Since $X \geqslant \xi_k$, and $\mathrm{ES}_p \geqslant \mathbb{E}$ on L^1, using monotonicity of ρ, we have

$$\rho(X) \geqslant \liminf_{k\to\infty} \rho(\xi_k) = \lambda \liminf_{k\to\infty} \mathrm{ES}_p(\xi_k).$$

If $\lambda > 0$, then $\rho(X) \geqslant \infty$, which violates the assumption that ρ is real-valued. Hence, $\lambda = 0$ and thus $\rho = 0$ on L^1.

To proceed, we utilize the idea of the proof of Lemma A.2. Let $U \sim \mathrm{U}[0,1]$, and let A be an event that has probability $\mathbb{P}(A) = 1 - p$ and is independent of U. Define $\phi : L^q_+ \to \mathbb{R}$ by

$$\phi(X) = \rho(F_X^{-1}(U)\mathbb{1}_A).$$

Using the same arguments as in the proof of Lemma A.2, we know that ϕ is additive and monotone on L^q_+. Next we define $\widehat{\phi} : L^q \to \mathbb{R}$ by $\widehat{\phi}(X) = \phi(X_+) - \phi(X_-)$ and then, using the same arguments as in the proof of Lemma A.1, show that $\widehat{\phi}$ is linear on L^q. It is well known (e.g., Section 1.47 of Rudin (1991) or Delbaen (2007)) that there is no non-zero linear functional on L^q, and hence $\widehat{\phi} = 0$ on L^q, which in turn implies $\phi = 0$ on L^q_+. For any random variable $Y \in L^q_+$, let B_Y be a p-tail event of Y, and let Y' follow the conditional distribution of Y given B_Y. Then $F_{Y'}^{-1}(U)\mathbb{1}_A \overset{\mathrm{d}}{=} Y\mathbb{1}_{B_Y}$ and hence

$$\rho(Y\mathbb{1}_{B_Y}) = \rho(F_{Y'}^{-1}(U)\mathbb{1}_A) = \phi(Y') = 0. \tag{7}$$

Next, we work with the functional $\psi : L^q_- \to \mathbb{R}$ defined by

$$\psi(X) = \rho(F_X^{-1}(U)\mathbb{1}_{A^c})$$

on the set L^q_- of all non-positive random variables in L^q. We arrive at $\psi = 0$ on L^q_-. For any random variable $Z \in L^q_-$, let B_Z be a p-tail event of Z, and let Z' follow the conditional distribution of Z given $(B_Z)^c$. We have $F_{Z'}^{-1}(U)\mathbb{1}_{A^c} \overset{\mathrm{d}}{=} Z\mathbb{1}_{(B_Z)^c}$ and hence

$$\rho(Z\mathbb{1}_{(B_Z)^c}) = \rho(F_{Z'}^{-1}(U)\mathbb{1}_{A^c}) = \psi(Z') = 0. \tag{8}$$

Finally, we take any $X \in L^q$ and let B be a p-tail event of X. With the notations $x_p = \mathrm{VaR}_p(X)$, $Y = (X - x_p)_+$, and $Z = -(X - x_p)_-$, we have $(X - x_p)\mathbb{1}_B = Y\mathbb{1}_B$ and $(X - x_p)\mathbb{1}_{B^c} = Z\mathbb{1}_{B^c}$. Since the random variables $Y\mathbb{1}_B$ and $Z\mathbb{1}_{B^c}$ share the common p-tail event B, using $\rho(x_p) = 0$ and equations (7)–(8), we arrive at

$$\begin{aligned}
\rho(X) &= \rho((X - x_p) + x_p) \\
&= \rho((X - x_p)\mathbb{1}_B + (X - x_p)\mathbb{1}_{B^c}) + \rho(x_p) \\
&= \rho(Y\mathbb{1}_B) + \rho(Z\mathbb{1}_{B^c}) = \phi(Y) + \psi(Z) = 0.
\end{aligned}$$

Hence, $\rho = 0$ on L^q, thus concluding the proof of Theorem 2. □

A.1.3 Proofs of Propositions 1–4

Next, we prove the four propositions of Section 2.

Proof of Proposition 1. Axioms M and LI are straightforward to check. Hence, we are left to verify Axiom P for $\rho = \mathrm{VaR}_p$ and $\rho = \mathrm{ES}_p$. To show that ρ satisfies Axiom P, it suffices to show $\liminf_{k\to\infty} \rho(\xi_k) \geqslant \rho(X)$ for a sequence $(\xi_k)_{k\in\mathbb{N}}$ with all $\xi_k \in L^1$ and such that $\xi_k \to X$ in distribution. This gives a stronger version of Axiom P; see Remark 1.

We start with $\rho = \mathrm{VaR}_p$. Note that $\mathrm{VaR}_p(\xi_k) \to \mathrm{VaR}_p(X)$ for all continuity points $p \in (0,1)$ of the function $\phi(p) = \mathrm{VaR}_p(X)$, which is left-continuous, monotone, and has at most countably many points of discontinuity. If p is a continuity point of ϕ, then $\lim_{k\to\infty} \mathrm{VaR}_p(\xi_k) = \mathrm{VaR}_p(X)$, which verifies Axiom P. If p is a discontinuity point, then, for every $\varepsilon > 0$, we can find a continuity point q of ϕ in the interval $[p-\varepsilon, p]$. Using monotonicity of ϕ, we arrive at

$$\liminf_{k\to\infty} \mathrm{VaR}_p(\xi_k) \geqslant \liminf_{k\to\infty} \mathrm{VaR}_q(\xi_k) = \mathrm{VaR}_q(X).$$

Since ϕ is left-continuous, taking the limit $q \uparrow p$ verifies Axiom P for VaR_p.

When $\rho = \mathrm{ES}_p$, we choose any continuity point $q \in (0,p)$ of ϕ. Note that

$$\mathrm{ES}_p(\xi_k - \mathrm{VaR}_q(\xi_k)) = \frac{1}{1-p}\int_p^1 (\mathrm{VaR}_t(\xi_k) - \mathrm{VaR}_q(\xi_k))\mathrm{d}t.$$

Using Axiom P for VaR_p and also noting that $\mathrm{VaR}_q(\xi_k) \to \mathrm{VaR}_q(X)$, we have

$$\liminf_{k\to\infty} (\mathrm{VaR}_t(\xi_k) - \mathrm{VaR}_q(\xi_k)) \geqslant \mathrm{VaR}_t(X) - \mathrm{VaR}_q(X)$$

for every $t \in [p,1)$. By Fatou's Lemma,

$$\begin{aligned}
\liminf_{k\to\infty} \mathrm{ES}_p(\xi_k) &= \liminf_{k\to\infty} \mathrm{ES}_p(\xi_k) - \lim_{k\to\infty} \mathrm{VaR}_q(\xi_k) + \mathrm{VaR}_q(X) \\
&= \liminf_{k\to\infty} \mathrm{ES}_p(\xi_k - \mathrm{VaR}_q(\xi_k)) + \mathrm{VaR}_q(X) \\
&\geqslant \frac{1}{1-p}\int_p^1 \liminf_{k\to\infty} (\mathrm{VaR}_t(\xi_k) - \mathrm{VaR}_q(\xi_k))\mathrm{d}t + \mathrm{VaR}_q(X) \\
&\geqslant \frac{1}{1-p}\int_p^1 (\mathrm{VaR}_t(X) - \mathrm{VaR}_q(X))\mathrm{d}t + \mathrm{VaR}_q(X) = \mathrm{ES}_p(X).
\end{aligned}$$

This verifies Axiom P for ES_p and concludes the proof of Proposition 1. □

Proof of Proposition 2. Suppose that $X_1, \dots, X_n$ share the tail event A. By Lemma A.5, we know that A is a tail event of the sum of any two of $X_1, \dots, X_n$. By induction, A is also a tail event of the sum of any of $X_1, \dots, X_n$. Therefore, using the additivity of ρ for risks sharing the tail event A, we get

$$\rho\left(\sum_{i=1}^{n} X_i\right) = \rho\left(X_1 + \sum_{i=2}^{n} X_i\right) = \rho(X_1) + \rho\left(\sum_{i=2}^{n} X_i\right) = \cdots = \rho(X_1) + \cdots + \rho(X_n),$$

and this finishes the proof of Proposition 2. □

Proof of Proposition 3. It suffices to construct four examples, each satisfying a distinct set of three axioms but not the fourth one:

1. For fixed $p \in (0,1)$, VaR_p satisfies all axioms but Axiom NRC.
2. $\mathbb{E}$ satisfies all axioms but Axiom P.
3. For fixed $\omega \in \Omega$, the mapping $X \mapsto X(\omega)$ satisfies all axioms but Axiom LI.
4. For fixed $p \in (0,1)$, the mapping $X \mapsto \mathrm{ES}_p(-X)$ satisfies all axioms but Axiom M.

The above examples are straightforward to check, with the help of Theorems 5–6 and Proposition 1. □

Proof of Proposition 4. The ($\Leftarrow$) implication is straightforward. Take any event A with probability $1-p$, and we can see that ρ satisfies Axiom NRC with A being the stress event.

Below we show the ($\Rightarrow$) implication. Suppose that ρ satisfies Axiom NRC with A being the stress event, and let $p = 1 - \mathbb{P}(A)$. Let X and Y be two L^q random variables which share a tail event B of probability $1-p$. Recall that our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless. As a consequence, the probability spaces $(A, \mathcal{F} \cap A, \mathbb{P}(\cdot|A))$ and $(A^c, \mathcal{F} \cap A^c, \mathbb{P}(\cdot|A^c))$ are again atomless, implying that there exist random vectors on these spaces with any specified joint distribution (this can be verified via e.g., Proposition A.31 of Föllmer and Schied (2016)).

Let (X_1, Y_1) be a random vector on $(A, \mathcal{F} \cap A, \mathbb{P}(\cdot|A))$ with the same distribution as the conditional distribution of (X, Y) on B, and (X_2, Y_2) be a random vector on $(A^c, \mathcal{F} \cap A^c, \mathbb{P}(\cdot|A^c))$ with the same distribution as the conditional distribution of (X, Y) on B^c. Define a random vector (X', Y') on $(\Omega, \mathcal{F}, \mathbb{P})$ by

$$(X', Y')(\omega) = \begin{cases} (X_1, Y_1)(\omega) & \omega \in A, \\ (X_2, Y_2)(\omega) & \omega \in A^c. \end{cases}$$

Since B is a tail event of X, we have $X(\omega) \geqslant X(\omega')$ for $\mathbb{P}$-a.s. $\omega \in B$ and $\omega' \in B^c$. This implies, by construction of X_1 and X_2, $X_1(\omega) \geqslant X_2(\omega')$ for $\mathbb{P}$-a.s. $\omega \in A$ and $\omega' \in A^c$. By definition, we further have $X'(\omega) \geqslant X'(\omega')$ for $\mathbb{P}$-a.s. $\omega \in A$ and $\omega' \in A^c$. That is, A is a tail event of X'. Analogously, A is also a tail event of Y'. By Axiom NRC, we know $\rho(X'+Y') = \rho(X') + \rho(Y')$.

Moreover, using the law of total probability,

$$\begin{aligned}
\mathbb{P}((X',Y') \leqslant (x,y)) &= \mathbb{P}(A)\mathbb{P}((X,Y) \leqslant (x,y)|A) + \mathbb{P}(A^c)\mathbb{P}((X,Y) \leqslant (x,y)|A^c) \\
&= \mathbb{P}(A)\mathbb{P}((X_1,Y_1) \leqslant (x,y)|A) + \mathbb{P}(A^c)\mathbb{P}((X_2,Y_2) \leqslant (x,y)|A^c) \\
&= \mathbb{P}(B)\mathbb{P}((X,Y) \leqslant (x,y)|B) + \mathbb{P}(B^c)\mathbb{P}((X,Y) \leqslant (x,y)|B^c) \\
&= \mathbb{P}((X,Y) \leqslant (x,y)).
\end{aligned}$$

Thus, (X',Y') and (X,Y) are identically distributed, and hence $X',Y' \in L^q$. With Axiom LI, this implies $\rho(X') = \rho(X)$, $\rho(Y') = \rho(Y)$, and $\rho(X'+Y') = \rho(X+Y)$. Therefore, we have

$$\rho(X+Y) = \rho(X'+Y') = \rho(X') + \rho(Y') = \rho(X) + \rho(Y),$$

completing the proof of the statement. □

A.2 Proofs of results of Section 3

A.2.1 Auxiliary results

In this appendix, we provide several auxiliary results on tail events and p-concentration, some of which have been mentioned in the main text of the paper. They will be essential when proving results of Section 3. In the proofs, all statements should be understood in the sense of "a.s.", but we shall omit "a.s." in obvious places for simplicity.

Lemma A.3. *Let $p \in (0,1)$ and denote $x_p = \mathrm{VaR}_p(X)$. For any random variable X, an event A is a p-tail event of X if and only if $\mathbb{P}(A) = 1-p$ and $\{X > x_p\} \subset A \subset \{X \geqslant x_p\}$ a.s. As a consequence, a p-tail event of X always exists.*

Proof. Define $D = \{X > x_p\}$ and $E = \{X \geqslant x_p\}$. Clearly, $D \subset E$.

($\Leftarrow$): Suppose that $\mathbb{P}(A) = 1-p$ and $D \subset A \subset E$. Since $A \subset E$ and $A^c \subset D^c$, we have $X(\omega) \geqslant X(\omega')$ for all $\omega \in A$ and $\omega' \in A^c$. Hence, A is a p-tail event of X.

($\Rightarrow$): Let A be a p-tail event of X. We need to prove the inclusions $D \subset A \subset E$, which we do in two steps, starting with $A \subset E$.

Suppose for the sake of contradiction that $\mathbb{P}(A \setminus E) > 0$, which and the bound $\mathbb{P}(E) \geqslant 1-p = \mathbb{P}(A)$ imply $\mathbb{P}(E \setminus A) > 0$. By definition of A, we have $X(\omega) \geqslant X(\omega')$ for $\omega \in A$ and $\omega' \in A^c$, and thus

$$X(\omega) \geqslant X(\omega') \ \text{ for } \omega \in A \setminus E \text{ and } \omega' \in E \setminus A. \tag{9}$$

By definition of E, we have $X(\omega') \geqslant x_p$ for all $\omega' \in E \setminus A$ and hence for all $\omega \in A \setminus E$ by inequality (9). Therefore, $(A \setminus E) \subset E$, which can only happen if $A \setminus E$ is empty, and thus we have $\mathbb{P}(A \setminus E) = 0$, which contradicts the supposition $\mathbb{P}(A \setminus E) > 0$. Hence, we must have $\mathbb{P}(A \setminus E) = 0$ and thus $A \subset E$.

Now we prove $D \subset A$, which is very similar to the previous case. Namely, suppose that $\mathbb{P}(D \setminus A) > 0$, which and the bound $\mathbb{P}(D) \leqslant 1-p = \mathbb{P}(A)$ imply $\mathbb{P}(A \setminus D) > 0$. By definition of A, we have $X(\omega) \geqslant X(\omega')$ for $\omega \in A$ and $\omega' \in A^c$, and thus

$$X(\omega) \geqslant X(\omega') \ \text{ for } \omega \in A \setminus D \text{ and } \omega' \in D \setminus A. \tag{10}$$

By definition of D, we know that $X(\omega') > x_p$ for all $\omega' \in D \setminus A$ and hence for all $\omega \in A \setminus D$ by inequality (10). Therefore, $(A \setminus D) \subset D$, which implies $\mathbb{P}(A \setminus D) = 0$ and contradicts the supposition $\mathbb{P}(D \setminus A) > 0$. Hence, $\mathbb{P}(D \setminus A) = 0$ and thus $D \subset A$.

We conclude the proof of Lemma A.3 by noting that the existence of a tail event follows directly from the already noted bound $\mathbb{P}(E) \geqslant 1-p \geqslant \mathbb{P}(D)$. Namely, choose any set $B \subset E \setminus D$ of probability $1-p-\mathbb{P}(D)$. Such a choice is possible because the probability space is assumed to be atomless. Finally, we check that $D \cup B$ is a p-tail event of X. This completes the proof of Lemma A.3. □

As a direct consequence of Lemma A.3, we have the following corollary.

Corollary A.1. *If X is continuously distributed, then its p-tail event is a.s. equal to $\{X > \mathrm{VaR}_p(X)\}$ and, therefore, is a.s. unique.*

For discrete random variables, however, their p-tail events may not be unique.

Example A.1. If X is constant, then Lemma A.3 implies that every event $A \subset \Omega$ of probability $1-p$ is a p-tail event of X.

The following lemma provides an additional insight into the structure of p-tail events. The lemma also plays an important role when dealing with copulas in the proof of Theorem 3 below.

Lemma A.4. *Let $p \in (0,1)$, and let X be any random variable. An event A is a p-tail event of X if and only if $A = \{U > p\}$ a.s. for some uniform on $[0,1]$ random variable U satisfying $F_X^{-1}(U) = X$ a.s.*

Proof. ($\Leftarrow$): For any set A which is a.s. equal to $\{U > p\}$, we obviously have $\mathbb{P}(A) = 1 - p$. Furthermore, for a.s. all $\omega \in A$ and $\omega' \in A^c$, we have

$$X(\omega) = F_X^{-1}(U(\omega)) \geqslant F_X^{-1}(p) \geqslant F_X^{-1}(U(\omega')) = X(\omega').$$

Hence, the set A is a p-tail event of X.

($\Rightarrow$): Define the probability space $(A, \mathcal{F}_A, \mathbb{P}_A)$, where $\mathcal{F}_A = \{A \cap C : C \in \mathcal{F}\}$ and $\mathbb{P}_A(C) = \mathbb{P}(C)/\mathbb{P}(A)$ for $C \in \mathcal{F}_A$, and similarly for $B = A^c$, define the probability space $(B, \mathcal{F}_B, \mathbb{P}_B)$. The mapping $X : A \to \mathbb{R}$ (resp. $X : B \to \mathbb{R}$) is a random variable on $(A, \mathcal{F}_A, \mathbb{P}_A)$ (resp. $(B, \mathcal{F}_B, \mathbb{P}_B)$), and we denote its distribution by F_A (resp. F_B). Note that for $u \in (0,1)$,

$$F_A^{-1}(u) = F^{-1}(p + (1-p)u) \quad \text{and} \quad F_B^{-1}(u) = F^{-1}(pu). \tag{11}$$

By Föllmer and Schied (2016, Lemma A.32), there exists a uniform on $[0,1]$ random variable U_A on $(A, \mathcal{F}_A, \mathbb{P}_A)$ such that $F_A^{-1}(U_A) = X$, $\mathbb{P}_A$-a.s., and a uniform on $[0,1]$ random variable U_B on $(B, \mathcal{F}_B, \mathbb{P}_B)$, such that $F_B^{-1}(U_B) = X$, $\mathbb{P}_B$-a.s. Define $U = \mathbb{1}_A(p + (1-p)U_A) + \mathbb{1}_B p U_B$. Clearly, U is a well-defined random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, and it is also clear that $A = \{U > p\}$ a.s. Moreover, U is uniform on $[0,1]$ since its distribution is a mixture of the uniform distribution on $[p,1]$ with probability $1-p$ and the uniform distribution on $[0,p]$ with probability p. Finally, using properties (11), we have

$$\begin{aligned} F_X^{-1}(U) &= \mathbb{1}_A F_X^{-1}((p + (1-p)U_A) + \mathbb{1}_B F_X^{-1}(pU_B) \\ &= \mathbb{1}_A F_A^{-1}(U_A) + \mathbb{1}_B F_B^{-1}(U_B) = X \text{ a.s.} \end{aligned}$$

This completes the proof of Lemma A.4. □

The next lemma on the common tail events of a portfolio vector and its aggregate loss is of profound importance and serves a crucial step in the proofs of Proposition 2, Theorem 3, and Theorem 5.

Lemma A.5. *Let $p \in (0,1)$, and let (X, Y) be any p-concentrated random pair. An event A is a p-tail event of $X + Y$ if and only if A is a p-tail event of both X and Y.*

Proof. (⇐): Let A be a common p-tail event of X and Y. The existence (by assumption) of such an event gives the inequality $X(\omega) + Y(\omega) \geqslant X(\omega') + Y(\omega')$ for all $\omega \in A$ and $\omega' \in A^c$, and thus proves that A is a p-tail event of $X + Y$.

(⇒): By Lemma A.3, we have the inclusions $\{X > x_p\} \subset A \subset \{X \geqslant x_p\}$ and $\{Y > y_p\} \subset A \subset \{Y \geqslant y_p\}$, where $x_p = \mathrm{VaR}_p(X)$, $y_p = \mathrm{VaR}_p(Y)$, and A is a common p-tail event of X and Y. Consequently,

$$\big(\{X > x_p\} \cup \{Y > y_p\}\big) \subset A \subset \big(\{X \geqslant x_p\} \cap \{Y \geqslant y_p\}\big). \tag{12}$$

Take now any p-tail event B of $X + Y$. By inclusion (12), we have

$$\mathbb{P}(X + Y \geqslant x_p + y_p) \geqslant \mathbb{P}(\{X \geqslant x_p\} \cap \{Y \geqslant y_p\}) \geqslant \mathbb{P}(A) = 1 - p$$

and

$$\mathbb{P}(X + Y > x_p + y_p) \leqslant \mathbb{P}(\{X > x_p\} \cup \{Y > y_p\}) \leqslant \mathbb{P}(A) = 1 - p.$$

Hence, since B is a p-tail event of $X+Y$, we obtain $\{X+Y > x_p+y_p\} \subset B \subset \{X+Y \geqslant x_p+y_p\}$. Using inclusion (12) again, we have

$$\begin{aligned}\{X + Y \geqslant x_p + y_p\} &\subset (\{X = x_p, Y = y_p\} \cup \{X > x_p\} \cup \{Y > y_p\}) \\ &\subset (\{X = x_p, Y = y_p\} \cup \{X > x_p\} \cup \{X \geqslant x_p\}) = \{X \geqslant x_p\},\end{aligned}$$

and similarly $\{X + Y \geqslant x_p + y_p\} \subset \{Y \geqslant y_p\}$. Therefore, $B \subset (\{X \geqslant x_p\} \cap \{Y \geqslant y_p\})$. On the other hand, inclusion (12) also implies

$$\{X > x_p\} = (\{Y \geqslant y_p\} \cap \{X > x_p\}) \subset \{X + Y > x_p + y_p\}$$

and similarly $\{Y > y_p\} \subset \{X + Y > x_p + y_p\}$. Therefore, $(\{X > x_p\} \cup \{Y > y_p\}) \subset B$. By Lemma A.3, B is a p-tail event of both X and Y. This finishes the proof of Lemma A.5. □

The next lemma is an immediate consequence of the definition of p-tail events, and it will be used for proving Theorems 3–4.

Lemma A.6. *Let $f : \mathbb{R} \to \mathbb{R}$ be any increasing function, and let X be any random variable. A p-tail event of X is a p-tail event of $f(X)$, and the converse is also true when the function f is strictly increasing.*

The next example illustrates that pairwise p-concentration does not necessarily imply p-concentration.

Example A.2. Let A_1, A_2, A_3 be three disjoint events, each of probability $p = 1/3$, and let $X_i = \mathbb{1}_{A_i}$ for every $i = 1, 2, 3$. Every pair (X_i, X_j) is p-concentrated because X_i and X_j share the same p-tail event $A_i \cup A_j$. However, the triplet (X_1, X_2, X_3) cannot be p-concentrated because for an event to be a p-tail event of all three random variables, the event has to contain all the three sets A_1, A_2, A_3, as implied by Lemma A.3, but this is impossible.

As we have seen from Example A.2, for a random vector $(X_1, \ldots, X_n)$, pairwise p-concentration is generally not sufficient for p-concentration of the vector. The next proposition illustrates the simple fact that pairwise p-concentration becomes sufficient for having p-concentration of the entire vector if at least one of the random components is continuously distributed.

Proposition A.1. *Let $p \in (0, 1)$, and let $(X_1, \ldots, X_n)$ be any random vector with continuously distributed X_1. The following statements are equivalent:*

(i) *$(X_1, \ldots, X_n)$ is p-concentrated;*

(ii) *(X_i, X_j) is p-concentrated for every pair $i, j = 1, \ldots, n$;*

(iii) *(X_1, X_j) is p-concentrated for every $j = 2, \ldots, n$.*

Proof. The implications (i)⇒(ii)⇒(iii) are trivial. To show (iii)⇒(i), note that p-tail events of X_1 are a.s. unique (Corollary A.1) and hence can be chosen as common p-tail events of $X_1, \ldots, X_n$. This finishes the proof of Proposition A.1. □

A direct consequence of Proposition A.1 is that p-concentration is transitive for continuous random variables. Namely, given a random triplet (X, Y, Z) with Y continuously distributed, if both (X, Y) and (Y, Z) are p-concentrated, then (X, Z) is also p-concentrated.

A.2.2 Proofs of Theorems 3–4

With the technical lemmas collected in Appendix A.2.1, we are now ready to prove the main results of Section 3.

Proof of Theorem 3. The equivalence (i)⇔(iv) follows immediately from Lemma A.6. We complete the rest of the proof by first establishing (i)⇒(ii)⇒(iii)⇒(i) and then (i)⇒(v)⇒(i).

(i)⇒(ii): Let A be a common p-tail event of $X_1, \ldots, X_n$. This immediately implies the inequality $\sum_{i=1}^n X_i(\omega) \geqslant \sum_{i=1}^n X_i(\omega')$ for all $\omega \in A$ and $\omega' \in A^c$. Hence, A is a p-tail event of $S = X_1 + \cdots + X_n$, and so $(X_1, \ldots, X_n, S)$ is p-concentrated.

(ii)⇒(iii): With the same A as above, we have, for every $i = 1, \ldots, n$,

$$S(\omega) - X_i(\omega) = \sum_{j=1, j\neq i}^{n} X_j(\omega) \geqslant \sum_{j=1, j\neq i}^{n} X_j(\omega') = S(\omega') - X_i(\omega')$$

for all $\omega \in A$ and $\omega' \in A^c$. Hence, A is a p-tail event of $S - X_i$. Consequently, the pair $(X_i, S - X_i)$ is p-concentrated, and this is true for every $i = 1, \ldots, n$.

(iii)⇒(i): Using Lemma A.5, we know that every p-tail event of the sum S is also a p-tail event of each $X_1, \ldots, X_n$. Hence, $(X_1, \ldots, X_n)$ is p-concentrated.

(i)⇒(v): Since $(X_1, \ldots, X_n)$ is p-concentrated, there is a common p-tail event A. Since it is a p-tail event for every X_i, by Lemma A.4 we can find a uniform on $[0, 1]$ random variable U_i such that $F_{X_i}^{-1}(U_i) = X_i$ a.s. and $A = \{U_i > p\}$ a.s. Given these uniform on $[0, 1]$ random variables $U_1, \ldots, U_n$, let the copula $C : [0, 1]^n \to [0, 1]$ be defined by

$$C(u_1, \ldots, u_n) = \mathbb{P}\big(U_1 \leqslant u_1, \ldots, U_n \leqslant u_n\big).$$

We have $C(p, \ldots, p) = \mathbb{P}(A^c) = p$ and complete the proof of (i)⇒(v).

(v)⇒(i): Since C is a copula of $X_1, \ldots, X_n$, there are uniform on $[0, 1]$ random variables $U_1, \ldots, U_n$ such that $C(u_1, \ldots, u_n) = \mathbb{P}\big(U_1 \leqslant u_1, \ldots, U_n \leqslant u_n\big)$, and $F_{X_i}^{-1}(U_i) = X_i$ a.s. for $i = 1, \ldots, n$. Write $B_i = \{U_i \leqslant p\}$. By Lemma A.4, we know that B_i^c is a p-tail event of X_i, and this is true for every $i = 1, \ldots, n$. Since $C(p, \ldots, p) = p$ by assumption, we have

$$p = C(p, \ldots, p) = \mathbb{P}\big(U_1 \leqslant p, \ldots, U_n \leqslant p\big) = \mathbb{P}\left(\bigcap_{i=1}^{n} B_i\right) = p. \tag{13}$$

Furthermore, since $U_1, \ldots, U_n$ are uniform on $[0, 1]$, we have

$$\mathbb{P}(B_i) = p \quad \text{for every} \quad i = 1, \ldots, n. \tag{14}$$

Statements (13) and (14) imply $B_1 = \cdots = B_n$ a.s. Hence, B_1^c is a common p-tail event of every $X_1, \ldots, X_n$. This finishes the proof of (v)⇒(i) and concludes the proof of Theorem 3. □

Proof of Theorem 4. (⇒): By comonotonicity of $X_1, \ldots, X_n$, we can find a random variable Z and increasing functions $f_1, \ldots, f_n$ such that $X_i = f_i(Z)$ a.s. for every $i = 1, \ldots, n$. By Lemma A.6, every p-tail event of Z is also a p-tail event of $X_1, \ldots, X_n$, implying that $(X_1, \ldots, X_n)$ is p-concentrated.

($\Leftarrow$): Consider first the case $n = 2$. That is, we shall show that if a random pair (X, Y) is p-concentrated for every $p \in (0, 1)$, then (X, Y) is comonotonic. Let A_p be a p-tail event of $X + Y$. We can assume $A_q \subset A_p$ for $1 > q > p > 0$ because p-tail events, as characterized by Lemma A.3, need to only satisfy $\mathbb{P}(A_p) = 1 - p$ and $\{X + Y > s_p\} \subset A_p \subset \{X + Y \geqslant s_p\}$, where $s_p = \mathrm{VaR}_p(X + Y)$.

By Lemma A.5, we know that A_p is also a p-tail event of both X and Y. Define $P : \Omega \to [0, 1]$ by $P(\omega) = \sup\{q \in (0, 1) : \omega \in A_q\}$, with $\sup \varnothing = 0$ by usual convention. Note that for $q \in (0, 1]$, we have $\{P \geqslant q\} = \bigcap_{s<q} A_s$, and hence P is a well-defined random variable on $(\Omega, \mathcal{F})$. Furthermore, $\{P > q\} \subset A_q \subset \{P \geqslant q\}$. Noting that $\mathbb{P}(A_q) = 1 - q$, we have $P \sim \mathrm{U}[0, 1]$, that is, P is a uniformly on $[0, 1]$ distributed random variable. Let $\mathbb{Q} = \mathbb{P} \times \mathbb{P}$ be the probability product measure on Ω^2. We have

$$\mathbb{Q}(\{(\omega, \omega') \in \Omega^2 : P(\omega) = P(\omega')\}) = 0.$$

For $(\omega, \omega') \in \Omega^2$, if $P(\omega) > P(\omega')$, then there exists $q \in (P(\omega'), P(\omega))$ such that $\omega \in A_q$ and $\omega' \in A_q^c$. Since A_q is a q-tail event of both X and Y, we have $(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0$. Similarly, if $P(\omega) < P(\omega')$, we have $(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0$. Therefore,

$$\mathbb{Q}(\{(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0\}) \geqslant \mathbb{Q}(\{P(\omega) \neq P(\omega')\}) = 1.$$

This is an equivalent reformulation of comonotonicity (e.g., Rüschendorf, 2013, Theorem 2.14). In other words, (X, Y) is comonotonic.

For a general n, suppose that $(X_1, \dots, X_n)$ is p-concentrated for every $p \in (0, 1)$. With the notation $S = X_1 + \dots + X_n$ and for every $i = 1, \dots, n$, Theorem 3 implies that $(X_i, S - X_i)$ is p-concentrated for every $p \in (0, 1)$. This in turn implies that $(X_i, S - X_i)$ is comonotonic, by the above established case $n = 2$. By Denneberg's Lemma (Denneberg, 1994, Proposition 4.5), we know that $X_i = f_i(S)$ for some increasing function f_i, and this holds for every $i = 1, \dots, n$. Hence, $(X_1, \dots, X_n)$ is comonotonic. □

A.3 Proofs of results of Section 4

The following lemma is a key step in the study of risk aggregation for ES.

Lemma A.7. *Let $p \in (0, 1)$ and $X \in L^1$. For any event A of probability $1 - p$, the equation $\mathrm{ES}_p(X) = \mathbb{E}[X|A]$ holds if and only if A is a p-tail event of X.*

Proof. We start with the general note (e.g., Embrechts and Wang, 2015, equation (3.1)) that ES_p

has a dual representation in the form

$$\mathrm{ES}_p(X) = \max\Big\{\mathbb{E}[X|B] : B \in \mathcal{F},\ \mathbb{P}(B) = 1-p\Big\}. \tag{15}$$

($\Leftarrow$): For any p-tail event A of X, the pair $(\mathbb{1}_A, X)$ is comonotonic. By the Fréchet-Hoeffding inequality, $\mathbb{E}[X\mathbb{1}_A] \geqslant \mathbb{E}[X\mathbb{1}_B]$ for every event B of probability $1-p$. Hence, by equation (15), we have $\mathrm{ES}_p(X) = \mathbb{E}[X|A]$.

($\Rightarrow$): Take any event A of probability $1-p$ and such that $\mathbb{E}[X|A] = \mathrm{ES}_p(Z)$. In addition, take any event B of the same probability $1-p$. By equation (15), we have $\mathbb{E}[X|A] \geqslant \mathbb{E}[X|B]$ and thus

$$\mathbb{E}[X\mathbb{1}_{A\setminus B}] + \mathbb{E}[X\mathbb{1}_{A\cap B}] = \mathbb{E}[X\mathbb{1}_A] \geqslant \mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[X\mathbb{1}_{B\setminus A}] + \mathbb{E}[X\mathbb{1}_{A\cap B}].$$

Since $\mathbb{P}(A \setminus B) = \mathbb{P}(B \setminus A)$, we have $\mathbb{E}[X|A \setminus B] \geqslant \mathbb{E}[X|B \setminus A]$, and since B is arbitrary, the relationship $X(\omega) \geqslant X(\omega')$ holds for a.s. all $\omega \in A$ and $\omega' \in A^c$. Hence, A is a p-tail event of X. This finishes the proof of Lemma A.7. □

Proof of Theorem 5. We shall rely on the well-known facts that ES_p is always subadditive, and that it is additive when $X_1, \dots, X_n$ are comonotonic (e.g., Embrechts and Wang, 2015).

(i)$\Rightarrow$(ii): Suppose that $X_1, \dots, X_n$ are p-concentrated random variables. Denote by A their common p-tail event, which is also a p-tail event of $X_1 + \cdots + X_n$ due to Lemma A.5. Using Lemma A.7 and subadditivity of ES_p, we have

$$\begin{aligned}\mathrm{ES}_p\left(\sum_{i=1}^n X_i\right) = \mathbb{E}\left[\sum_{i=1}^n X_i \middle| A\right] &= \sum_{i=1}^n \mathbb{E}\left[X_i|A\right] \\ &= \sum_{i=1}^n \mathrm{ES}_p(X_i) = \sum_{i=1}^n \mathrm{ES}_p(X_i') \geqslant \mathrm{ES}_p\left(\sum_{i=1}^n X_i'\right),\end{aligned}$$

where $X_i' \overset{\mathrm{d}}{=} X_i$, $i = 1, \dots, n$. Hence, the vector $(X_1, \dots, X_n)$ maximizes the ES_p aggregation, as per Definition 4.

(ii)$\Rightarrow$(iii): Since ES_p is always subadditive, and it is additive for comonotonic random variables, we have the equation

$$\sum_{i=1}^n \mathrm{ES}_p(X_i) = \max\left\{\mathrm{ES}_p\left(\sum_{i=1}^n X_i'\right) : X_i' \overset{\mathrm{d}}{=} X_i,\ i = 1, \dots, n\right\}, \tag{16}$$

which establishes implication (ii)$\Rightarrow$(iii).

(iii)⇒(i): Using Lemma A.7, there exists a p-tail event A of $X_1 + \cdots + X_n$ such that

$$\sum_{i=1}^{n} \mathrm{ES}_p(X_i) = \mathrm{ES}_p\left(\sum_{i=1}^{n} X_i\right) = \mathbb{E}\left[\sum_{i=1}^{n} X_i \middle| A\right] = \sum_{i=1}^{n} \mathbb{E}[X_i|A]. \tag{17}$$

Since $\mathbb{E}[X_i|A] \leqslant \mathrm{ES}_p(X_i)$ for every $i = 1, \ldots, n$, equations (17) imply $\mathbb{E}[X_i|A] = \mathrm{ES}_p(X_i)$ for every $i = 1, \ldots, n$. Using Lemma A.7 again, we conclude that A is a p-tail event of every X_i, $i = 1, \ldots, n$. Therefore, $(X_1, \ldots, X_n)$ is p-concentrated. This completes the proof of Theorem 5. □

Proof of Theorem 6. The middle inequality of (2) is trivial, and so we only need to check the two remaining ones. We start with two auxiliary definitions. Namely, for any random variable X and event B with $\mathbb{P}(B) > 0$, the essential infimum and supremum of X conditioned on B are defined by

$$\text{ess-inf}(X|B) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x|B) > 0\},$$
$$\text{ess-sup}(X|B) = \sup\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x|B) < 1\},$$

respectively. The bounds $\text{ess-inf}(X+Y|B) \geqslant \text{ess-inf}(X|B)+\text{ess-inf}(Y|B)$ and $\text{ess-sup}(X+Y|B) \leqslant \text{ess-sup}(X|B) + \text{ess-sup}(Y|B)$ always hold.

By Lemma A.3, any p-tail event A of X satisfies $\{X > x_p\} \subset A$ a.s. If $\mathbb{P}(X \leqslant x) > p$, then $A^c \subset \{X \leqslant x_p\} \subset \{X \leqslant x\}$ a.s., and hence

$$\begin{aligned}\mathbb{P}(X \leqslant x) &= \mathbb{P}(\{X \leqslant x\} \cap A) + \mathbb{P}(\{X \leqslant x\} \cap A^c) \\ &= \mathbb{P}(\{X \leqslant x\} \cap A) + \mathbb{P}(A^c) = \mathbb{P}(\{X \leqslant x\} \cap A) + p.\end{aligned}$$

Therefore,

$$\mathrm{VaR}_p^+(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(\{X \leqslant x\} \cap A) > 0\} = \text{ess-inf}(X|A).$$

Similarly, we get the equation $\mathrm{VaR}_p(X) = \text{ess-sup}(X|A^c)$.

By Theorem 3, we can find a common p-tail event A of $X_1, \ldots, X_n, S$, where $S = X_1+\cdots+X_n$. We get

$$\sum_{i=1}^{n} \mathrm{VaR}_p^+(X_i) = \sum_{i=1}^{n} \text{ess-inf}(X_i|A) \leqslant \text{ess-inf}(S|A) = \mathrm{VaR}_p^+(S)$$

and

$$\sum_{i=1}^{n} \mathrm{VaR}_p(X_i) = \sum_{i=1}^{n} \text{ess-sup}(X_i|A^c) \geqslant \text{ess-sup}(S|A^c) = \mathrm{VaR}_p(S).$$

As seen from Example 1, the first inequality in (2) is not always an equality, and VaR_p is not p-additive. By symmetry, the last inequality in (2) is not always an equality either. For the second inequality to hold strictly, it suffices that one of $X_1, \ldots, X_n$ has a quantile with a jump at p. This finishes the proof of Theorem 6. □

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.

Acharya, V., Engle, R. and Richardson, M. (2012). Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review*, **102**(3), 59–64.

Acharya, V., Pedersen, L.H., Philippon, T., Richardson, M. (2017). Measuring systemic risk. *Review of Financial Studies.* **30**(1), 2–47,

Adrian, T. and Brunnermeier, M. K. (2016). CoVaR. *American Economic Review*, **106**(7), 1705–1741.

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

Azenes and Effixis (2018). *Finma's Swiss Solvency Test 2019: Major Changes in the Market Risk Standard Model.* `https://effixis.ch/downloads/2018-10-31-finma-new-sst.pdf`

BCBS (2016). *Minimum Capital Requirements for Market Risk. January 2016.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements. `https://www.bis.org/bcbs/publ/d352.htm`

BCBS (2019). *Minimum Capital Requirements for Market Risk. February 2019.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements. `https://www.bis.org/bcbs/publ/d457.htm`

Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, **54**(1), 41–48.

Cai, Z. and Wang, X. (2008). Nonparametric estimation of conditional VaR and expected shortfall. *Journal of Econometrics*, **147**, 120–130

Chambers, C. P. (2009). An axiomatization of quantiles on the domain of distribution functions. *Mathematical Finance*, **19**(2), 335–342.

Chen, C., Iyengar, G. and Moallemi, C. C. (2013). An axiomatic approach to systemic risk. *Management Science*, **59**(6), 1373–1388.

Chen, S. X. (2008). Nonparametric estimation of Expected Shortfall. *Journal of Financial Econometrics*, **6**, 87–107.

Cherny, A. S. and Madan, D. (2009). New measures for performance evaluation. *Review of Financial Studies*, **22**(7), 2571–2606.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. and Montrucchio, L. (2011). Uncertainty averse preferences. *Journal of Economic Theory*, **146**(4), 1275–1330.

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

Delbaen, F. (2007). Risk measures for non-integrable random variables. *Mathematical Finance*, **19**(2), 329–333.

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

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

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.

Du, Z. and Escanciano, J. C. (2017). Backtesting expected shortfall: Accounting for tail risk. *Management Science*, **63**(4), 940–958.

Embrechts, P., Liu, H. and Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, **66**(4), 936–949.

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750–2764.

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

Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, **19**(4), 763–790.

Embrechts, P. and Wang, R. (2015). Seven proofs for the subadditivity of Expected Shortfall. *Dependence Modeling*, **3**(1), 126–140.

Emmer, S., Kratz, M. and Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, **18**(2), 31–60.

Fermanian, J. D. and Scaillet, O. (2005). Sensitivity analysis of VaR and Expected Shortfall for portfolios under netting agreements. *Journal of Banking and Finance*, **29**, 927–958.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics*, **44**(4), 1680–1707.

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints *Finance and Stochastics*, **6**(4), 429–447.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time.* (Fourth Edition.) Walter de Gruyter, Berlin.

Frittelli, M. and Rosazza Gianin, E. (2002). Putting order in risk measures *Journal of Banking and Finance*, **26**(7), 1473–1486.

Furman, E., Wang, R. and Zitikis, R. (2017). Gini-type measures of risk and variability: Gini shortfall, capital allocation and heavy-tailed risks. *Journal of Banking and Finance*, **83**, 70–84.

Gao, N. and Leung, D. H. and Munari, C. and Xanthos, F. (2018). Fatou Property, representations, and extensions of law-invariant risk measures on general Orlicz spaces. *Finance and Stochastics*, **22**(2), 395–415.

Gilboa, I. (2009). *Theory of Decision under Uncertainty.* Cambridge University Press, Cambridge.

Gilboa, I. Maccheroni, F., Marinacci, M. and Schmeidler, D. (2010). Objective and subjective rationality in a multiple prior model. *Econometrica*, **78**(2), 755–770.

Gilboa, I., Postlewaite, A. and Samuelson, L. (2019). What are axiomatizations good for? *Theory and*

Decision, **86**(3–4), 339–359.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, **106**(494), 746–762.

He, X. and Peng, X. (2018). Surplus-invariant, law-invariant, and conic acceptance sets must be the sets induced by Value-at-Risk. *Operations Research*, **66**(5), 1268–1276.

Kou, S. and Peng, X. (2016). On the measurement of economic tail risk. *Operations Research*, **64**(5), 1056–1072.

Krätschmer, V., Schied, A. and Zähle, H. (2014). Comparative and quantitiative robustness for law-invariant risk measures. *Finance and Stochastics*, **18**(2), 271–295.

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

Liu, F. and Wang, R. (2020). A theory for measures of tail risk. *Mathematics of Operations Research*, forthcoming.

Maccheroni, F., Marinacci, M. and Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, **74**(6), 1447–1498.

Marinacci, M. and Montrucchio L. (2004). Introduction to the mathematics of ambiguity. In *Uncertainty in Economic Theory*, I. Gilboa, Ed. Routledge, New York, NY, USA. 46–107.

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

Nelsen, R. (2006). *An Introduction to Copulas.* (Second Edition.) Springer, New York.

Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*, **55**(4), 819–847.

Patton, A. J., Ziegel, J. F. and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and Value-at-Risk). *Journal of Econometrics*, **211**(2), 388–413.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, **3**(4), 323–343.

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

Rothschild, M. and Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic Theory*, **2**(3), 225–243.

Rudin, W. (1991). *Functional Analysis.* (Second Edition.) McGraw-Hill, New York.

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

Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of Expected Shortfall. *Mathematical Finance*, **14**(1), 115–129.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, **57**(3), 571–587.

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

Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity.* Cambridge University Press, Cambridge.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, **26**(1), 71–92.

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

Wang, R. and Wei, Y. (2018). Risk functionals with convex level sets. SSRN: 3292661.

Wang, R. and Zitikis, R. (2020). Weak comonotonicity. *European Journal of Operational Research*, **282**, 386–397.

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

Ziegel, J.F. Coherence and elicitability. *Mathematical Finance*, **26**(4), 901–918.