A Theory for Measures of Tail Risk

Fangda Liu* Ruodu Wang†

Journal accepted version: Feb 19, 2020‡

Abstract

The notion of "tail risk" has been a crucial consideration in modern risk management and financial regulation, as very well documented in the recent regulatory documents. To achieve a comprehensive understanding of the tail risk, we carry out an axiomatic study for risk measures which quantify the tail risk, that is, the behavior of a risk beyond a certain quantile. Such risk measures are referred to as tail risk measures in this paper. The two popular classes of regulatory risk measures in banking and insurance, the Value-at-Risk (VaR) and the Expected Shortfall (ES), are prominent, yet elementary, examples of tail risk measures. We establish a connection between a tail risk measure and a corresponding law-invariant risk measure, called its generator, and investigate their joint properties. A tail risk measure inherits many properties from its generator, but not subadditivity or convexity; nevertheless, a tail risk measure is coherent if and only if its generator is coherent. We explore further relevant issues on tail risk measures, such as bounds, distortion risk measures, risk aggregation, elicitability, and dual representations. In particular, there is no elicitable, tail convex risk measure rather than the essential supremum, and under a continuity condition, the only elicitable and positively homogeneous monetary tail risk measures are the VaRs.

Key-words: Basel III, tail risk, risk aggregation, elicitability, Value-at-Risk

MSC 2020 Classification: Primary: 91G70; secondary: 91B06 **OR/MS subject classification**: Primary: Utility/preference; secondary: Finance

*Department of Statistics and Actuarial Science, University of Waterloo, Canada. ✉ fangda.liu@uwaterloo.ca.

†Department of Statistics and Actuarial Science, University of Waterloo, Canada. ✉ wang@uwaterloo.ca.

‡The authors thank the Editor, an Associate Editor, three referees, Fabio Bellini, Paul Embrechts, Liyuan Lin, Sidney Resnick and Alexander Schied for helpful comments and discussions on an earlier version of the paper. F. Liu acknowledges financial support from Natural Sciences and Engineering Research Council of Canada (RGPIN-2020-04717, DGECR-2020-00340) and the National Natural Science Foundation of China (No. 11601540). R. Wang acknowledges financial support from and the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03823, RGPAS-2018-522590).

1 Introduction

In the past few decades, tail-based risk measures have become the standard metrics for the assessment of risks and regulatory capital calculation in the regulatory frameworks for banking and insurance sectors, such as Basel III and Solvency II (see, for instance, Sandström [50], Cannata and Quagliariello [11] and BCBS [5]). Such risk measures look into the "tail" or "shortfall" of a risk, that is, the behaviour of the risk at or beyond a certain, typically high-level, quantile.

The most popular measures used in banking and insurance practice are the Value-at-Risk (VaR) and the Expected Shortfall (ES, also known as Tail-Value-at-Risk). The risk measures VaR at confidence level $p \in (0,1)$ refer to the left and right p-quantiles of a risk (random variable) X, denoted by $\mathrm{VaR}_p^L(X)$ and $\mathrm{VaR}_p^R(X)$, respectively. The level p here is close to 1 in practice (for instance, typically $p = 0.975$ or $p = 0.99$ in Basel III and Solvency II), thus representing a "tail risk". The risk measure ES is defined as

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

which, roughly speaking, is the mean of the risk X beyond its p-quantile. Formal definitions are given in Section 2. Below we quote the Basel Committee on Banking Supervision, Page 1 of BCBS [5], Executive Summary:

"... 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*."

It is clear from the above quote that the ability of a risk measure to capture "tail risk" is a crucial concern for financial regulation, and this issue is closely related to capital adequacy under financial market stress. Prudent assessment of risks in adverse economic scenarios ("financial market stress") has been an important trend of research in modern risk management; see, for instance, Acharya et. al. [1, 2] and McNeil et al. [42]. For more recent discussions on the recent issues with VaR and ES in regulation, we refer to Embrechts et al. [20] and Föllmer and Weber [25].

As the "tail risk" appears prominent in modern risk management, a systematic study of measures of tail risk is thereby the focus of this paper. The first thing to set straight is the definition of a tail risk. Noting that both VaR and ES are calculated from the tail-part distribution of risks, our definition of measures of tail risk follows naturally. In the sequel, we refer to a risk measure determined solely by the distribution of a random variable beyond its p-quantile as a *p-tail risk measure.*

There are various reasons to develop a theory for tail risk measures which are not limited to VaR and ES. First, for the stability of the financial system, the focal scenario of concern to a regulator is the tail part of a risk which represents big financial losses, instead of the body part of the risk, which typically represents the profit of a financial institution[1]. In light of this, a general theory for tail risk measures is in demand for the purpose of regulation. Second, as the risk measures VaR_p^L and VaR_p^R are p-quantiles and ES_p is the average loss beyond its p-quantile, they merely report simple statistics of the tail risk. None of them captures other important features of the tail risk, such as its variability or distributional shape. Therefore, some other risk measures may be more suitable for the (internal) management of tail risks in specific situations. Third, from a regulatory perspective, as VaR and ES are the standard for solvency capital calculation in the banking and insurance industries, other tail risk measures provide informational support for the regulator to better comprehend the tail risk. This is analogous to using the variance or the skewness of a risk in addition to its mean or median for decision making, albeit now we are looking at the tail risks. Some possible choices of tail risk measures, such as the Gini Shortfall (see Furman et al. [28]), are given in Section 6. Fourth, through the study of other tail risk measures, we understand better the fundamental roles which VaR and ES play among all such risk measures. From a mathematical perspective, tail risk measures exhibit some rather surprising and nice analytical properties, as we shall see from the main results in this paper.

The main feature of p-tail risk measures is that they focus on partial distributional information of the risk. Such a technique is found useful in other applications than regulatory risk assessment. For instance, in a recent study, Dai et al. [14] discuss the relationship between the Gini coefficient and top incomes shares, and propose the *top incomes truncated inequality measure* via an axiomatic approach, which uses a particular quantile range of the income distribution. We refer to Dai et al. [14] and the reference therein for more examples of measures using partial information.

Below we describe the structure and the main contributions of the paper. The first natural question is how to generate tail risk measures. For a fixed probability level p, it is straightforward that one can always obtain a tail risk measure ρ by applying a law-invariant risk measure ρ^* (which we call a *generator*) to the tail distribution of a risk. Moreover, we show that the relationship between a tail risk measure and its generator is one-to-one on the set of random variables bounded from below.

It is however not a trivial task to identify properties of a tail risk measure based on the corresponding properties of its generator. To illustrate this, let us look at the benchmark tail risk measure ES_p. Its generator is the expectation $\mathbb{E}[\cdot]$. It is well known that $\mathbb{E}[\cdot]$ is linear and elicitabile (see Section 5 for definition), but ES_p is neither linear nor elicitable. That is, some properties are not passed on to the tail risk measure from its generator.

In Section 3, we show that monotonicity, translation-invariance, positive homogeneity and comonotonic additivity are passed on from a generator ρ^* to the corresponding tail risk measure ρ.

[1] A different approach to address this concern loss part of a risk is through a surplus-invariant risk measures; see Cont et al. [12], Staum [52] and Koch-Medina et al. [37]. The latter class of risk measures rules out the important regulatory risk measure ES.

However, subadditivity, convexity, $\prec_{\mathrm{cx}}$-monotonicity and elicitability cannot be passed on from ρ^* to ρ in general. Nevertheless, based on a result in risk aggregation, we show that ρ is a coherent risk measure if and only if ρ^* is a coherent risk measure. Thus, subadditivity and convexity can be passed on to ρ when accompanied by other properties (in particular, monotonicity). Another quite interesting finding is that any monetary p-tail risk measure dominates VaR_p, and a coherent tail risk measure always dominates ES_p. In other words, VaR and ES serve as benchmarks for tail risk measures, and in fact they are the smallest tail risk measures with a given probability level p.

We proceed to discuss a few other questions on measures of tail risk. A particularly relevant issue is risk aggregation for tail risk measures under dependence uncertainty, that is, the aggregation of several risks with known marginal distributions and unknown dependence structure; for a stream of research in this direction, we refer to Embrechts et al. [19, 21], Bernard et al. [8] and the references therein. In Section 4, we show that, for monotone risk measures, the worst-case aggregation of a tail risk measure for some given marginal distributions is equivalent to the worst-case aggregation of its generator for the corresponding tail distributions. This result generalizes the existing result in Bernard et al. [8] for VaR and will be useful in showing some important properties of tail risk measures.

Elicitability has drawn an increasing attention in the recent few years due to its connection to statistical backtests and forecasts for risk measures; see Gneiting [30] and the references therein. Existing results in Ziegel [60], Bellini and Bignozzi [6], Delbaen et al. [16] and Kou and Peng [36] suggest that among all convex risk measures, shortfall risk measures are the only elicitable ones, and among all distortion risk measures, VaRs and the expectation are the only elicitable ones. In Section 5, we identify tail shortfall risk measures, all of which turn out to be *surplus-invariant* (see Koch-Medina et al. [37]). Furthermore, all elicitable monetary tail risk measures with a continuity assumption are characterized. We find that the only elicitable, positively homogeneous and monetary tail risk measures are again the VaRs (thus, a new axiomatic characterization of the VaRs), and there are no elicitable tail convex or coherent risk measures except for the essential supremum. Several examples of tail risk measures are presented in Section 6, and some concluding remarks are put in Section 7. Proofs and some related results on tail distortion risk measures and dual representations are put in the appendices.

2 Preliminaries

We work with an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let L^q be the set of all random variables in $(\Omega, \mathcal{F}, \mathbb{P})$ with finite q-th moment, $q \in [0, \infty)$, and let L^∞ be the set of essentially bounded random variables. A positive (resp. negative) value of $X \in L^0$ represents a financial loss (resp. profit) in this paper. Throughout, for any $X \in L^0$, F_X represents the distribution function of X and $F_X^{-1}(p) = \inf\{x \in \mathbb{R} : F_X(x) \geqslant p\}$, $p \in (0,1)$. Let U_X be a uniform random variable on $[0,1]$ such that $F_X^{-1}(U_X) = X$ almost surely for any X; its existence is given, for instance, in Lemma A.28 of Föllmer and Schied [24]. The mappings ess-inf$(\cdot)$ and ess-sup$(\cdot)$ on L^0 stand for the essential infimum and the essential supremum of a random variable, respectively. We denote by $X \overset{\mathrm{d}}{=} Y$ if the random variables X and Y have the same distribution under $\mathbb{P}$. For any set $A \subseteq \Omega$, denoted by $\mathbb{1}_A$ the corresponding indicator function. For $x \in \mathbb{R}$, write $(x)_+ = \max\{x, 0\}$ and denote by δ_x the point-mass probability distribution at x.

Let $\mathcal{X}$ be a convex cone of random variables containing L^∞. Although $\mathcal{X}$ is unspecific in our discussion, it does not hurt to think of $\mathcal{X} = L^\infty$ to better comprehend the main ideas. A *risk measure* ρ is a functional that maps $\mathcal{X}$ to $(-\infty, \infty]$ with $\rho(X) < \infty$ for $X \in L^\infty$. Whenever a risk measure appears in this paper, its domain is $\mathcal{X}$ unless otherwise specified. In Appendix A, we list several standard properties for general risk measures in the literature. In particular, law-invariance (see Appendix A for its definition) is satisfied by all risk measures of this paper, and we shall therefore not mention it specifically.

The two most popular classes of risk measures used in banking and insurance practice are the Value-at-Risk (VaR) and the Expected Shortfall (ES). The VaR at confidence level $p \in (0, 1)$ has two versions, the right p-quantile of X at p, defined as

$$\mathrm{VaR}_p^R(X) = \inf\{x \in \mathbb{R} : F_X(x) > p\} = F_X^{-1}(p+), \quad X \in L^0,$$

and the left p-quantile of X, defined as

$$\mathrm{VaR}_p^L(X) = \inf\{x \in \mathbb{R} : F_X(x) \geqslant p\} = F_X^{-1}(p), \quad X \in L^0.$$

In risk management practice, one often does not distinguish between VaR_p^R and VaR_p^L as they are identical for random variables with an inverse distribution function continuous at p. Both VaR_p^R and VaR_p^L will be referred to as VaRs in this paper. The ES at confidence level $p \in (0, 1)$ is defined as

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

Note that $\mathrm{ES}_p(X)$ may be infinite if X is not integrable. In addition, we also write

$$\mathrm{ES}_1(X) = \mathrm{VaR}_1^R(X) = \mathrm{VaR}_1^L(X) = \text{ess-sup}(X) = \inf\{x \in \mathbb{R} : F_X(x) = 1\}.$$

3 Measures of tail risk

We first give a precise definition of a "tail risk". Intuitively, a tail risk is the behaviour of a risk X at or beyond a certain threshold. In view of this, for any random variable $X \in \mathcal{X}$ and $p \in (0, 1)$, we let X_p be the *tail risk of* X beyond its p-quantile, that is,

$$X_p = F_X^{-1}(p + (1-p)U_X).$$

One can easily check

$$\mathbb{P}(X_p \leqslant x) = \mathbb{P}(X \leqslant x | U_X \geqslant p) = \frac{(\mathbb{P}(X \leqslant x) - p)_+}{1-p}, \quad x \in \mathbb{R}. \tag{2}$$

Clearly, if F_X^{-1} is strictly increasing at p, then X_p follows the conditional distribution of X given $X \geqslant \mathrm{VaR}_p^R(X)$. The distribution of X_p is called the p-tail distribution of X in Rockafellar and Uryasev [48].

We shall throughout assume that $X \in \mathcal{X}$ implies $X_p \in \mathcal{X}$. This assumption holds for common choices of $\mathcal{X}$, such as $\mathcal{X} = L^q$, $q \in [0, \infty]$. Now we are ready to define a measure of tail risk.

Definition 1. For $p \in (0,1)$, a risk measure ρ is a *p-tail risk measure* if $\rho(X) = \rho(Y)$ for all $X, Y \in \mathcal{X}$ satisfying $X_p \overset{\mathrm{d}}{=} Y_p$. A risk measure ρ is *tail-relevant* if it is a p-tail risk measure for some $p \in (0,1)$; we will simply call it a *tail risk measure.*

In other words, the value of a p-tail risk measure for a risk X is solely determined by the distribution of X beyond its p-quantile. It is immediate from Definition 1 that VaRs and ES in Section 2 are all tail risk measures, whereas the expectation is not a tail risk measure. The value p here should be chosen according to the specific application or context, similarly to the specification of the confidence level p in using a VaR or ES in banking or insurance regulation. In view of this, p can be close to 1 in risk management practice, nevertheless all results in this paper hold for all $p \in (0,1)$.

Remark 1. Regulators specifically emphasize the importance of "capturing tail risk" in several Basel documents over the past several years (e.g. BCBS [4, 5]). This strongly motivates us to look for a precise definition of risk measures that specifically account for tail risk. Definition 1 seems to be the most natural choice for such a property. Arguably, a risk measure in Definition 1 only takes into account the tail risk, and hence it is not only "capturing the tail risk" but also "solely capturing the tail risk". The feature of a p-tail risk measure is that, in plain words, "for a risky position X, we do not care about how much profit it makes in a good day, but only how much loss it causes in a bad day (the worst outcome with probability $1-p$)." This feature is consistent with modern quantitative risk management (see e.g. McNeil et al. [42]) and the practical choices of risk measures in regulation. As we shall see below, the regulatory risk measures VaR and ES play important roles in the family of tail risk measures.

Remark 2. The feature of a measure that only uses partial information of the underlying distribution finds applications in other fields. For instance, Dai et al. [14] proposed and characterized the *top incomes truncated inequality measure* which excludes top income groups to capture the essential incomes inequality information in a society, and provided an axiomatic framework based on nonlinear expected utility theory. The top incomes truncated inequality measures in Dai et al. [14] and the p-tail risk measures in Definition 1 share similar considerations in the sense that they utilize partial information from a particular quantile range of the underlying distribution.

Obviously, any tail risk measure is law-invariant. From the definition, for $0 < q < p < 1$, a p-tail risk measure is also a q-tail risk measure, as the latter is less restrictive. For $p \in (0,1)$, the risk measures VaR_p^R and ES_p are p-tail risk measures, and VaR_q^L is a p-tail risk measure for $q \in (p,1]$. One may immediately notice the simple relations

$$\mathrm{VaR}_p^R(X) = \text{ess-inf}(X_p) \quad \text{and} \quad \mathrm{ES}_p(X) = \mathbb{E}[X_p], \quad X \in \mathcal{X}.$$

Indeed, for any law-invariant risk measure ρ^* on $\mathcal{X}$, we may define its corresponding p-tail risk measure, for $p \in (0,1)$, via

$$\rho(X) = \rho^*(X_p), \quad X \in \mathcal{X}. \tag{3}$$

If (3) holds, we say that ρ is the p-tail risk measure *generated by* ρ^* and ρ^* is a *p-generator* of ρ. The relation (3) is denoted by an operator $\mathcal{T}_p : R(\mathcal{X}) \to R(\mathcal{X})$ as $\rho = \mathcal{T}_p[\rho^*]$, where $R(\mathcal{X})$ is the set of risk measures on $\mathcal{X}$.

Conversely, in the following we shall see, for any p-tail risk measure, that we can find a p-generator; thus, a risk measure is a p-tail risk measure if and only if it is generated by another risk measure. Denote by $\mathcal{X}^*$ the set of random variables in $\mathcal{X}$ with a finite essential infimum, that is

$$\mathcal{X}^* = \{X \in \mathcal{X} : \text{ess-inf}(X) > -\infty\}.$$

Note that if we take $\mathcal{X} = L^\infty$, then $\mathcal{X}^*$ coincides with $\mathcal{X}$. For any $X \in \mathcal{X}^*$ and $p \in (0,1)$, let $X^{(p)}$ be a random variable with distribution function

$$\mathbb{P}(X^{(p)} \leqslant x) = p\mathbb{1}_{\{x \geqslant \text{ess-inf}(X)\}} + (1-p)\mathbb{P}(X \leqslant x), \quad x \in \mathbb{R}.$$

Equivalently, $X^{(p)}$ is identically distributed as $\text{ess-inf}(X)B + X(1-B)$ where the random variable $B \sim \mathrm{Bern}(p)$ is independent of X. Here and in the sequel a $\mathrm{Bern}(p)$ distribution means $\mathbb{P}(B=1) = p$ and $\mathbb{P}(B=0) = 1-p$. The next proposition gives the uniqueness of the p-generator. We omit its proof since it is straightforward from a simple fact: for $X \in \mathcal{X}^*$, $(X^{(p)})_p \overset{\mathrm{d}}{=} X$ (see Lemma B.1 in the Appendix).

Proposition 1. *The p-generator ρ^* of a p-tail risk measure ρ is unique on $\mathcal{X}^*$, and is given by*

$$\rho^*(X) = \rho(X^{(p)}), \quad X \in \mathcal{X}^*. \tag{4}$$

Remark 3. The reason why ρ^* is only unique on $\mathcal{X}^*$ is because X_p for $p \in (0,1)$ is always bounded from below (thus in $\mathcal{X}^*$); as a consequence ρ^* in (3) can be arbitrary for random variables with infinite essential infimum. If we further assume that ρ^* is continuous from above in the sense that, for $X, X_1, X_2, \dots \in \mathcal{X}$, as $n \to \infty$, $X_n \downarrow X$ a.s. implies $\rho^*(X_n) \to \rho^*(X)$, then ρ^* is uniquely determined on $\mathcal{X}$.

From now on we can treat (ρ, ρ^*) in (3) as a pair of risk measures, and study their joint properties.

Definition 2. For $p \in (0,1)$, a pair of risk measures (ρ, ρ^*) is called a *p-tail pair* if ρ^* is law-invariant and $\rho = \mathcal{T}_p[\rho^*]$.

The domain of ρ^*, $\mathcal{X}$ or $\mathcal{X}^*$, does not affect the relation $\rho = \mathcal{T}_p[\rho^*]$. As such, we do not distinguish between whether ρ^* is defined on $\mathcal{X}$ or $\mathcal{X}^*$. The two distribution-wise transformations

$$X \mapsto X_p, \ X \in \mathcal{X} \ \text{ and } \ X \mapsto X^{(p)}, \ X \in \mathcal{X}^*$$

will repeatedly appear throughout the rest of the paper.

Some simple relations for the p-tail pair (ρ, ρ^*) and the operators $\mathcal{T}_p : R(\mathcal{X}) \to R(\mathcal{X})$ are briefly listed below, which can be verified in a straightforward manner. First, from (3) and Lemma B.1, we have $\rho(c) = \rho^*(c)$ for all $c \in \mathbb{R}$, and $\rho(X) \geqslant \rho^*(X)$ for all $X \in \mathcal{X}^*$ if ρ^* is monotone. Second, the class of operators $\mathcal{T} : R(\mathcal{X}) \to R(\mathcal{X})$ satisfies a composition property: $\mathcal{T}_p \circ \mathcal{T}_q = \mathcal{T}_q \circ \mathcal{T}_p = \mathcal{T}_{p+q-pq}$ for $p, q \in (0,1)$. In particular, the representative classes of tail risk measures VaRs and ES for different probability levels are connected via (i) $\mathcal{T}_p[\mathrm{ES}_q] = \mathcal{T}_q[\mathrm{ES}_p] = \mathrm{ES}_{p+q-pq}$; (ii) $\mathcal{T}_p[\mathrm{VaR}^R_q] = \mathcal{T}_q[\mathrm{VaR}^R_p] = \mathrm{VaR}^R_{p+q-pq}$; (iii) $\mathcal{T}_p[\mathrm{VaR}^L_q] = \mathcal{T}_q[\mathrm{VaR}^L_p] = \mathrm{VaR}^L_{p+q-pq}$.

Next, we study which classic properties are preserved or lost in the transform from ρ^* to ρ for a p-tail pair of risk measures (ρ, ρ^*). Here, we follow the standard terminologies in the risk measure literature; for precise definitions, see properties (A1)-(A7) listed in Appendix A. Before approaching a general result, we first look at a counter example where convexity (A3), subadditivity (A5) and $\prec_{\mathrm{cx}}$-monotonicity (A7) are not inherited by ρ from ρ^*.

Example 1 (Tail standard deviation)**.** The class of *standard deviation risk measures* is defined as, for $\beta > 0$,

$$\mathrm{SD}_\beta(X) = \mathbb{E}[X] + \beta\sqrt{\mathrm{var}(X)}, \quad X \in L^2. \tag{5}$$

It is well known that SD_β is translation-invariant, convex, positively homogeneous, subadditive and $\prec_{\mathrm{cx}}$-monotone, but it is not monotone or comonotonically additive (for its mathematical properties, see Section 5.3 of Kaas et al. [34]). Take $p \in (0,1)$ and let ρ be the p-tail risk measure generated by SD_β, that is,

$$\rho(X) = \mathbb{E}[X_p] + \beta\sqrt{\mathrm{var}(X_p)}, \quad X \in L^2.$$

See Furman and Landsman [27] for more on ρ, called the tail standard deviation. Now, take independent and identically distributed (iid) random variables X and Y such that $\mathbb{P}(X = -1) = p$ and $\mathbb{P}(X = 0) = 1-p$, and write $Z = X+Y$. Note that Z_p is not a constant as $\mathbb{P}(Z = 0) = (1-p)^2$ implies $\mathbb{P}(Z_p = 0) = 1-p$. It follows that $\mathrm{var}(Z_p) > 0$. Therefore, by taking β large enough, we have

$$\rho(X+Y) = \mathbb{E}[Z_p] + \beta\sqrt{\mathrm{var}(Z_p)} > 0.$$

On the other hand, noting that $X_p = Y_p = 0$ almost surely, we have $\rho(X) = \rho(Y) = 0$. Thus, $\rho(X+Y) > \rho(X) + \rho(Y)$, and ρ is not subadditive (and therefore not convex). Moreover, ρ is not

$\prec_{\mathrm{cx}}$-monotone either, which can be seen from $X+Y \prec_{\mathrm{cx}} 2X$ (see, for instance, Theorem 3.5 of Rüschendorf [49]) and $\rho(X+Y) > \rho(2X)$.

The following theorem identifies individual properties that can be passed on to a tail risk measure ρ from its generator ρ^* and the other way around.

Theorem 1. *Suppose that $p \in (0,1)$ and (ρ, ρ^*) is a p-tail pair of risk measures on $\mathcal{X}$ and $\mathcal{X}^*$. The following statements hold.*

(i) ρ is monotone (translation-invariant, positively homogeneous, comonotonically additive) if and only if so is ρ^*.

(ii) If ρ is subadditive (convex, $\prec_{\mathrm{cx}}$-monotone) then so is ρ^*.

(iii) ρ is a coherent (convex, monetary) risk measure if and only if so is ρ^*.

The converse statement of (ii) in Theorem 1 does not hold in general. Indeed, Example 1 shows that ρ is not necessarily subadditive, convex or $\prec_{\mathrm{cx}}$-monotone, even if ρ^* is subadditive, convex and $\prec_{\mathrm{cx}}$-monotone. Although these three properties may not be passed on from ρ^* to ρ, we can see in Theorem 1 (iii) that the whole set of properties for coherent risk measures as well as for convex risk measures can be passed on to ρ. In the proof of Theorem 1 (iii), we show the fact that ρ^* is a coherent (convex) risk measure implies ρ is coherent (convex). The implication of this fact is arguably the most important of all, as it would allow us to generate coherent (convex) tail risk measures by freely choosing generic coherent (convex) risk measures. To establish such a mechanism is one of the initial motivations for the study of tail risk measures. A proof of Theorem 1 relies on a new result on worst-case risk aggregation (Theorem 3) which we present in Section 4.

We conclude this section by establishing the essential importance of VaRs and ES as benchmarks for tail risk measures.

Theorem 2. *Let $p \in (0,1)$. If ρ is a monetary p-tail risk measure with $\rho(0)=0$, then $\rho \geqslant \mathrm{VaR}_p^R$ on $\mathcal{X}$, and if ρ is a coherent p-tail risk measure, then $\rho \geqslant \mathrm{ES}_p$ on L^∞.*

The converse statements to Theorem 2 are not true in general. For instance, take $\rho(X) = \max\{\mathbb{E}[X], \mathrm{VaR}_p^R(X)\}$, $X \in \mathcal{X}$. Then $\rho \geqslant \mathrm{VaR}_p^R$ on $\mathcal{X}$ but ρ is not a p-tail risk measure by definition.

4 Risk aggregation

In the presence of model uncertainty, a popular approach risk management is to evaluate the worst-case value of a risk measure over plausible models; see e.g. Natarajan et al. [43] and Zhu and Fukushima [61]. In this section, we study a particular type of model uncertainty in risk aggregation, which has been an active topic recently. A typical question on risk aggregation is to determine the worst-case value of a risk measure for the aggregation of risks with given marginal distributions, due to statistical uncertainty arising from estimating a dependence structure (see Embrechts et al. [20] and the references therein). More precisely, for given univariate distributions $F_1, \dots, F_n$, one calculates

$$\sup\{\rho(S) : S \in \mathcal{S}_n(F_1, \dots, F_n)\}, \tag{6}$$

where $\mathcal{S}_n(F_1, \dots, F_n)$ is the *aggregation set* defined as

$$\mathcal{S}_n(F_1, \dots, F_n) = \{X_1 + \cdots + X_n : X_i \in \mathcal{X},\ X_i \sim F_i,\ i = 1, \dots, n\}.$$

The worst-case value in (6) represents the risk value assuming that the portfolio is non-diversifiable, which is a conservative scenario considered by BCBS [5]; see Remark 5.

All distributions mentioned in this section are assumed to be compatible with random variables in $\mathcal{X}$. A particularly relevant case of (6) is $\rho = \mathrm{VaR}_p^L$ or $\rho = \mathrm{VaR}_p^R$ for some $p \in (0,1)$. This case is extensively studied recently in, for instance, Embrechts et al. [19, 21], and Wang et al. [56]; see also Bernard et al. [9] for the case of partial dependence information, and Wang et al. [55] and Cai et al. [10] for the case of general risk measures. For non-convex risk measures such as the VaRs, an analytical evaluation of (6) is generally unavailable.

For $p \in (0,1)$ and any distribution F, we denote by $F^{[p]}$ its p-tail distribution, that is, the distribution of $F^{-1}(U_p)$ where U_p is a uniform random variable on $[p, 1]$. In other words, for $X \in L^0$, $F_X^{[p]}$ is the distribution of X_p. The following theorem gives the convenient result that the worst-case value of a monotone p-tail risk measure over $\mathcal{S}_n(F_1, \dots, F_n)$ is equal to the worst-case value of its p-generator over $\mathcal{S}_n(F_1^{[p]}, \dots, F_n^{[p]})$. In other words, one may freely translate the risk aggregation problem of a tail risk measure to its generator. This result gives an essential step that completes the proof of Theorem 1 in Section 3.

Theorem 3. *Let $p \in (0,1)$ and (ρ, ρ^*) be a p-tail pair of monotone risk measures. For any univariate distributions $F_1, \dots, F_n$, we have*

$$\sup\{\rho(S) : S \in \mathcal{S}_n(F_1, \dots, F_n)\} = \sup\{\rho^*(T) : T \in \mathcal{S}_n(F_1^{[p]}, \dots, F_n^{[p]})\}. \tag{7}$$

Remark 4. For the cases of VaR and ES, Theorem 3 reduces to some classic results.

(i) If we take $\rho = \mathrm{VaR}_p^R$ in (7), then

$$\sup\{\mathrm{VaR}_p^R(S) : S \in \mathcal{S}_n(F_1, \dots, F_n)\} = \sup\{\text{ess-inf}(T) : T \in \mathcal{S}_n(F_1^{[p]}, \dots, F_n^{[p]})\}.$$

which is Lemma 4.3 of Bernard et al. [8]; see also Proposition 3 of Embrechts et al. [20].

(ii) If we take $\rho = \mathrm{ES}_p$ in (7), then, for any $X_1, \dots, X_n \in L^1$ with respective distributions $F_1, \dots, F_n$,

$$\mathrm{ES}_p(X_1 + \cdots + X_n) \leqslant \sup\{\mathbb{E}[T] : T \in \mathcal{S}_n(F_1^{[p]}, \dots, F_n^{[p]})\} = \sum_{i=1}^n \mathbb{E}[(X_i)_p] = \sum_{i=1}^n \mathrm{ES}_p(X_i),$$

which gives the classic subadditivity of ES_p.

By Theorem 3, to investigate its worst-case value of a monotone p-tail risk measure in risk aggregation, it suffices to consider the tail risk of each marginal distribution. This conclusion is arguably rather intuitive; however the statement is not true for non-monotone risk measures, as illustrated in the following example.

Example 2 (Example 1 continued)**.** Take $p \in (0,1)$, $X, Y \in L^2$, $\beta > 0$ and $\rho = \mathcal{T}_p[\mathrm{SD}_\beta]$ as in Example 1. We have already seen that $\rho(X+Y) > 0$ and $X_p = Y_p = 0$ almost surely. Therefore, we have

$$\sup \left\{ \mathrm{SD}_\beta(T) : T \in \mathcal{S}_2(F_X^{[p]}, F_Y^{[p]}) \right\} = \mathrm{SD}_\beta(0+0) = 0 < \rho(X+Y),$$

thus (7) fails to hold.

Remark 5. The main application of the worst-case risk aggregation is to obtain conservative risk value under the assumption of no diversification (i.e. worst-case dependence), which is a practical approach in banking. In the Fundamental Review of the Trading Book of the Basel Committee, firms are required to use a weighted average of an internally modelled risk value and the *non-diversifiable* risk value in (6) for the ES; see p.63 of BCBS [5]. In case of ES, the worst-case value is precisely the summation of individual ES values, due to subadditivity and comonotonic additivity of ES; this is not the case for generic risk measures such as the VaRs. We refer to Embrechts et al. [20] for more discussions on this issue.

5 Tail shortfall risk measures and elicitability

The notion of elicitability has drawn an increasing interest in risk management recently, due to its connection to comparative backtests and forecasts; see for instance Lambert et al. [40], Gneiting [30], Fissler and Ziegel [22] and Kou and Peng [36]. It is shown in Ziegel [60] and Delbaen et al. [16] that, among all convex risk measures, only shortfall risk measures are elicitable, and among all coherent risk measures, only *expectiles* (including the mean; see Remark 8) are elicitable. On the other hand, Kou and Peng [36] showed that among all distortion risk measures, only the mean and the quantiles are elicitable; see also Wang and Ziegel [57]. This leaves us wondering: Are there tail risk measures, other than the quantiles, which are elicitable? Note that for the p-tail pair $(\mathrm{ES}_p, \mathbb{E})$, ES_p is not elicitable but its generator is elicitable, and hence elicitability cannot be translated to a tail risk measure from its generator.

Elicitability is closely related to the notion of shortfall risk measures that we shall investigate in the sequel. Our findings can be summarized as follows. First, the only tail shortfall risk measures are the ones with a flat loss function on the negative real line. From there, with an additional continuity condition in Weber [58], they are also the only monetary tail risk measures that are elicitable. Further, no tail convex risk measures can be elicitable except for the essential supremum, and the only elicitable and positively homogeneous monetary tail risk measures are the VaRs.

We fix $\mathcal{X} = L^\infty$ in this section, with the exception in Theorem 6 that we generalize our VaR characterization to larger spaces than L^∞.

5.1 Tail shortfall risk measures

A function $\ell : \mathbb{R} \to \mathbb{R}$ is called a *loss function* if it is non-decreasing and $\inf_{x\in\mathbb{R}} \ell(x) < 0 < \sup_{x\in\mathbb{R}} \ell(x)$. For a loss function ℓ, define a risk measure

$$\rho_\ell(X) = \inf\{m \in \mathbb{R} : \mathbb{E}[\ell(X - m)] \leqslant 0\}, \quad X \in \mathcal{X}. \tag{8}$$

The risk measure in (8) is called a *shortfall risk measure induced by* ℓ in the literature. ρ_ℓ is a monetary risk measure, and it is convex if and only if ℓ is convex; see Föllmer and Schied [24, Section 4.9] for more on shortfall risk measures. Note that if ℓ^* is the left-continuous version of ℓ, then $\rho_{\ell^*} = \rho_\ell$ (one may verify that they have the same acceptance set). Hence, we may conveniently take ℓ to be left-continuous. Moreover, it suffices to study ρ with $\rho(0) = 0$, as one can always write $\bar{\ell}(x) = \ell(x - \rho(0))$, $x \in \mathbb{R}$, so that $\rho_{\bar{\ell}}(0) = 0$.

For $p \in (0, 1)$, we say that a risk measure is a *p-tail shortfall (resp. convex) risk measure* if it is both a shortfall (resp. convex) risk measure and a p-tail risk measure. Immediate examples of tail shortfall risk measures are the left-quantiles. For $p \in (0, 1)$, let

$$\ell(x) = \mathbb{1}_{\{x>0\}} - (1 - p), \ x \in \mathbb{R}. \tag{9}$$

Then one can verify that $\rho_\ell = \mathrm{VaR}_p^L$. For the case of the right quantile VaR_p^R one needs to modify (8) slightly; see Remark 7. As characterized in the following theorem, the class of tail shortfall risk measures includes more than just the quantiles.

Theorem 4. *For $p \in (0, 1)$, a shortfall risk measure ρ induced by ℓ with $\rho(0) = 0$ is a p-tail risk measure if and only if*

$$\ell(x) = \ell(-1) \text{ for all } x < 0 \text{ and } p\ell(-1) + (1 - p)\ell(y) > 0 \text{ for all } y > 0. \tag{10}$$

Remark 6. In the case $\rho_\ell = \mathrm{VaR}_p^L$ where $p \in (0, 1)$, the loss function ℓ given in (9) satisfies

$$\ell(x) = \ell(-1) = -(1 - p), \ x < 0 \text{ and } p\ell(-1) + (1 - p)\ell(y) = 0, \ y > 0.$$

For $q < p$, it holds $q\ell(-1) + (1 - q)\ell(y) > 0$, $y > 0$. From there, Theorem 4 confirms that VaR_p^L is a q-tail risk measure for $q \in (0, p)$, a fact we already know.

To interpret Theorem 4, note that a loss function ℓ satisfying (10) can be written as

$$\ell(x) = (\ell(x) - \ell(-1))\mathbb{1}_{\{x\geqslant 0\}} + \ell(-1) = \ell^*(x)\mathbb{1}_{\{x\geqslant 0\}} - c, \ x \in \mathbb{R},$$

where $c = -\ell(-1) > 0$ and $\ell^*(x) = \ell(x) + c > 0$, $x \in \mathbb{R}$. Noting again that ℓ may always be taken as its left-continuous version, the acceptance set of ρ_ℓ satisfies

$$\mathcal{A}_{\rho_\ell} = \{X \in \mathcal{X} : \mathbb{E}[\ell(X)] \leqslant 0\} = \{X \in \mathcal{X} : \mathbb{E}[\ell^*(X_+)] \leqslant c\}.$$

If ρ_ℓ is used as a regulatory capital principle, then the regulator accepts a position according to whether $\mathbb{E}[\ell^*(X_+)]$ exceeds a given number $c > 0$. In order to protect liability holders, only the potential loss (positive part of X) should be a concern to the regulator instead of the potential profit (negative part of X); see related arguments in Cont et al. [12] and Staum [52]. This property is called *surplus-invariance* in Koch-Medina et al. [37] and He and Peng [31]. Thus, by Theorem 4, a tail shortfall risk measure is always surplus-invariant. The fact that all tail shortfall risk measures are surplus-invariant is indeed surprising, as the definition of a tail risk measure is not directly related to surplus-invariance; for instance, ES_p is a tail risk measure which is not surplus-invariant.

Remark 7. One may replace the non-strict inequality in (8) with a strict one, and define a risk measure

$$\rho_\ell^+(X) = \inf\{m \in \mathbb{R} : \mathbb{E}[\ell(X-m)] < 0\}, \quad X \in \mathcal{X}. \tag{11}$$

For ℓ in (9), we have $\rho_\ell^+(X) = \mathrm{VaR}_p^R$. One can similarly show that, for $p \in (0,1)$, ρ_ℓ^+ is a p-tail risk measure if and only if the strict inequality in (10) is replaced by a non-strict one, that is,

$$\ell(x) = \ell(-1) \text{ for all } x < 0 \text{ and } p\ell(-1) + (1-p)\ell(y) \geqslant 0 \text{ for all } y > 0.$$

Other results for risk measures in (11) can be obtained analogously.

Remark 8. Another notable class of risk measures related to the tail risk is the class of *expectiles* (see Newey and Powell [44] and Bellini et al. [7]). Expectiles are shortfall risk measures with loss function $\ell : x \mapsto ax_+ - bx_-$, where $a, b > 0$. Although an expectile with $a > b$ arguably emphasizes the tail part of the risk, its value is not determined solely by the tail distribution (roughly speaking, an expectile is determined by a balance between expected profit and expected loss from a risk), and therefore is not a tail risk measure in our terminology. Expectiles are not used as a regulatory risk measure in practice even though they are the only coherent and elicitable risk measures. The fact that expectiles are not tail risk measures may be accounted as a reason to explain this observation.

5.2 Elicitability and convex level sets

In statistics, a set-valued functional ϕ mapping distributions to subsets of $\mathbb{R}$ is said to be $\mathcal{P}$-*elicitable* for a set $\mathcal{P}$ of distributions, if it can be written as the set of minimizers for the expectation of a score function $S : \mathbb{R}^2 \to \mathbb{R}$. In rigorous terms, there exists a score function S such that

$$\phi : \mathcal{P} \to 2^{\mathbb{R}}, \ \phi(F) = \underset{x \in \mathbb{R}}{\arg\min} \int_{\mathbb{R}} S(x,y)\,\mathrm{d}F(y).$$

The score function S may be required to satisfy some specific conditions in different applications. Typical choices of $S : \mathbb{R}^2 \to \mathbb{R}$ include $S(x,y) = (x-y)^2$, $S(x,y) = |x-y|$ and $S(x,y) = p(x-y)_+ + (1-p)(y-x)_+$, $p \in (0,1)$, and these choices of S correspond to ϕ being the sets of the mean, the medians and the p-quantiles, respectively. For a recent treatment on the application of elicitability and co-elicitability to backtesting and forecasting in risk management, see Nolde and

Ziegel [45]. In this paper, all tail risk measures discussed map a distribution to a single value rather than a set of values. Therefore, instead of using $\mathcal{P}$-elicitability, we adopt the definition in Kou and Peng [36] to define elicitability on single-valued functionals.

Definition 3. A law-invariant risk measure $\rho : \mathcal{X} \to \mathbb{R}$ is *elicitable* if there exists a function $S : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\rho(X) = \min\left\{\arg\min_{x\in\mathbb{R}} \mathbb{E}[S(x,X)]\right\}, \quad X \in \mathcal{X}. \tag{12}$$

To distinguish from statistical functionals, the notion in Definition 3 is referred to as general elicitability in Kou and Peng [36]. Note that the choice of min in (12) is rather artificial; one may choose, for instance, max or mid-point.

Shortfall risk measures play a natural role in the study of elicitability. First, for a shortfall risk measure ρ_ℓ induced by ℓ, by writing

$$S(x,y) = \int_x^0 \ell(y-z)\,\mathrm{d}z, \quad x,y \in \mathbb{R},$$

we have

$$\begin{aligned}\arg\min_{x\in\mathbb{R}} \mathbb{E}[S(x,X)] &= \arg\min_{x\in\mathbb{R}} \int_x^0 \mathbb{E}[\ell(X-z)]\,\mathrm{d}z \\ &= \left[\inf\{z\in\mathbb{R} : \mathbb{E}[\ell(X-z)] \leqslant 0\}, \inf\{z\in\mathbb{R} : \mathbb{E}[\ell(X-z)] < 0\}\right].\end{aligned}$$

Therefore,

$$\rho_\ell(X) = \inf\{z\in\mathbb{R} : \mathbb{E}[\ell(X-z)] \leqslant 0\} = \min\left\{\arg\min_{x\in\mathbb{R}} \mathbb{E}[S(x,X)]\right\}, \quad X \in \mathcal{X}.$$

Thus all shortfall risk measures are elicitable.

On the other hand, one can easily check (see e.g. Osband [46]) that a necessary condition for elicitability is the *convex level sets* (CxLS) property defined below. A law-invariant risk measure ρ is said to have CxLS, if for any $\lambda \in [0,1]$ and $X, Y \in \mathcal{X}$,

$$\rho(X) = \rho(Y) \text{ implies } \rho(Z_\lambda) = \rho(X),$$

where $Z_\lambda \in \mathcal{X}$ is a random variable with distribution $\lambda F_X + (1-\lambda)F_Y$.

Under some continuity assumptions, Weber [58] showed that monetary risk measures satisfying the CxLS property are indeed shortfall risk measures. See Bellini and Bignozzi [6] and Delbaen et al. [16] for more on the CxLS property. Thus, by characterizing tail risk measures that are shortfall risk measures, we can identify all elicitable tail risk measures under the continuity assumption in Weber [58].

For convex risk measures, Theorem 3.10 of Delbaen et al. [16] shows that a convex risk measure with CxLS is either a shortfall risk measure or VaR_1^L. By Proposition D.2, a tail shortfall risk measure cannot be convex, and hence, for any $p \in (0,1)$, the only elicitable p-tail convex risk measure is VaR_1^L, the essential supremum, which is elicitable with the score function $S(x,y) = \mathbb{1}_{\{x<y\}}$.

For non-convex risk measures, we impose a simple semi-continuity condition. A risk measure is said to be *distribution-wise lower-semi-continuous* (DLC), if it satisfies $\liminf_{n\to\infty} \rho(X_n) \geqslant \rho(X)$ for $X, X_1, X_2, \dots \in \mathcal{X}$ with $X_n \to X$ in distribution as $n \to \infty$. In our context, this condition is equivalent to the continuity condition in Weber [58]; see Remark 10. The next theorem shows that a monetary, elicitable, positively homogeneous and DLC p-tail risk measure has to be VaR_q^L for some $q \in (p, 1]$.

Theorem 5. *For $p \in (0,1)$, a monetary and positively homogeneous p-tail risk measure ρ satisfying DLC is elicitable if and only if $\rho = \mathrm{VaR}_q^L$ for some $q \in (p, 1]$.*

Remark 9. To arrive at a symmetric result with VaR_p^R replacing VaR_p^L in Theorem 5, one needs to replace the min in (12) by a max and replace the lower-semi-continuity in DLC by an upper-semi-continuity. See also Remark 7.

Without specifying the value of p, Theorem 5 immediately implies a new characterization of the family of VaR within the class of tail risk measures. We also note that, since the properties elicitability and DLC get stronger as the set $\mathcal{X}$ enlarges, the characterization in Theorem 5 holds for risk measures on any set of random variables containing L^∞. We summarize the above two observations in the following theorem.

Theorem 6. *Suppose that $\mathcal{X}$ is a convex cone containing L^∞. A monetary and positively homogeneous tail risk measure ρ on $\mathcal{X}$ satisfying DLC is elicitable if and only if $\rho = \mathrm{VaR}_q^L$ for some $q \in (0, 1]$.*

Some comparison between existing results on elicitable risk measures are drawn below. There are three main results which characterize a one-parameter family of elicitable and positively homogenous monetary risk measures.

(i) Ziegel [60] additionally assumed convexity (hence coherence), and arrived at expectiles.

(ii) Kou and Peng [36] additionally assumed comonotonic additivity (hence distortion), and arrived at VaRs and the mean.

(iii) We additionally assumed tail-relevance with DLC, and arrived at VaRs.

In addition to the characterizations of VaR given in Theorem 6 and Kou and Peng [36], He and Peng [31] characterized VaRs from surplus-invariance, numéraire-invariance, and truncation-closed acceptance set. To compare the axioms, such as tail-relevance and elicitability in Theorem 6, and the comonotonic independence in Kou and Peng [36], the surplus invariance and the numéraire invariance in He and Peng [31], we illustrate with the following examples.

(i) A tail standard deviation (Example 1) is a tail risk measure, whereas it does not satisfy comonotonic independence in Kou and Peng [36]. The tail standard deviation is quite popular in the insurance literature (e.g. Furman and Landsman [27]);

(ii) An ES is a tail risk measure, whereas it does not satisfy surplus invariance or numéraire invariance in He and Peng [31].

(iii) An expectile (see Remark 8) is elicitable, whereas it does not satisfy surplus invariance or numéraire invariance in He and Peng [31].

Theorem 6, Kou and Peng [36] and He and Peng [31] characterize the class of VaRs by using different sets of conditions, which may represent different practical concerns. Overall, it depends on the specific application which axiom is more convincing. In view of the Fundamental Review of the Trading Book by the Basel Committee which we quote in the Introduction, tail-relevance is one of the main reasons that the Basel Committee chooses VaR and ES as their risk measures, although comonotonic independence, surplus invariance and numéraire invariance are also practically important considerations.

Remark 10. To obtain the characterization in Theorem 5, one may assume instead of DLC that $\mathcal{N}_\rho$ is ψ-weakly closed for some gauge function ψ as in Weber [58, Theorem 3.1]. For general functionals, the DLC property is stronger than the ψ-weakly closedness property. Nevertheless, note that any p-tail shortfall risk measure satisfies DLC; therefore the DLC property is equivalent to the ψ-weakly closedness property in Weber [58] for p-tail risk measures with CxLS.

Remark 11. Via the same arguments used to show Theorem 5, we also conclude, that for $p \in (0,1)$, a monetary p-tail risk measure ρ satisfying DLC, (18) and $\rho(0) = 0$ is elicitable if and only if ρ is a shortfall risk measure induced by ℓ satisfying (10).

6 Examples of tail risk measures

In this section we present several examples of tail risk measures and relate them to various classes of risk or economic functionals in the literature. Throughout this section, $p \in (0,1)$ is a fixed number and (ρ, ρ^*) is a p-tail pair of risk measures on $\mathcal{X}$ and $\mathcal{X}^*$ respectively.

Example 3 (Median Shortfall). Let $\mathcal{X} = L^0$ and ρ^* be the left-median, that is $\rho^* = \mathrm{VaR}^L_{1/2}$. Then

$$\rho(X) = \mathrm{VaR}^L_{1/2}(X_p) = \mathrm{VaR}^L_{(1+p)/2}(X), \quad X \in L^0.$$

The risk measure ρ is called a *Median Shortfall* in Kou et al. [35]. It is clear that, ρ is monetary, positively homogeneous, elicitable and comonotonically additive, but not convex or subadditive.

Example 4 (Gini Shortfall). Let $\mathcal{X} = L^1$ and ρ^* be a *Gini principle* in Denneberg [17], defined as

$$\rho^*(X) = \mathbb{E}[X] + \beta\mathbb{E}[|X' - X''|], \quad X \in L^1,$$

where $\beta > 0$ and X' and X'' are iid copies of X. Then,

$$\rho(X) = \mathrm{ES}_p(X) + \beta \mathbb{E}[|X'_p - X''_p|], \quad X \in L^1,$$

where X'_p and X''_p are iid copies of X_p. The risk measure ρ is called a *Gini Shortfall* in Furman et al. [28]. It is shown in the latter paper that ρ is comonotonically additive, and it is coherent if and only if $\beta \leqslant 1/2$.

Example 5 (Range-VaR)**.** The family of *Range-Value-at-Risk* (RVaR) is introduced by Cont et al. [13]. For $X \in L^1$, a RVaR at level $(p, q) \in [0, 1)^2$ with $p < q$ is defined as

$$\rho(X) = \mathrm{RVaR}_{\alpha,\beta}(X) = \frac{1}{q-p} \int_p^q \mathrm{VaR}_r^R(X) \, \mathrm{d}r, \quad X \in L^1.$$

We can easily see that $\mathrm{RVaR}_{p,q}$ is a p-tail risk measure, and its generator is $\mathrm{RVaR}_{0,(q-p)/(1-p)}$. The family of RVaR includes VaR and ES as its limiting cases, and it has various advantages as compared to VaR and ES. In particular, an RVaR is a robust risk measure (see Cont et al. [13] and Kou et al. [35]), and the class of RVaR is the closure of inf-convolutions of VaR and ES (Embrechts et al. [18]). RVaR is also known as "Spread-VaR" in insurance practice as a simple kernel smoothing method to calculating capital allocations; see for instance Johnson [32]. We refer to Embrechts et al. [18] for more properties of RVaR and its economic implications.

7 Concluding remarks

In this paper, we develop a theory for measures of tail risk. Our main contributions can be summarized as follows. First, we propose a precise definition of measures of tail risk, and discover many of their properties. Second, we establish a simple way to generate tail risk measures with flexible desirable properties from existing non-tail risk measures. Third, we study risk aggregation with dependence uncertainty for tail risk measures, generalizing recent results on risk aggregation. Fourth, we connect tail risk measures with elicitability, and show that a positively homogenous and monetary tail risk measure is elicitable if and only if it is a VaR, leading to a new axiomatic characterization of the VaRs. There is a growing interest on tail risks and extremal events in finance and insurance from both academia and industry. The theory and tools developed in this paper hopefully provide valuable support to a prudent measurement of tail risk, and in particular, the results obtained complement the extensive use of VaR and ES in current regulation and risk assessment.

We believe that the novel concept of tail risk measures will inspire many questions for future research. Replacing a generic risk measure by its tail counterpart is philosophically analogous to replacing the expectation by an ES; many challenges arise in different problems of practical relevance. Some areas of potential applications are portfolio selection, market equilibrium, statistical inference, decision analysis, and optimization.

The economic motivation of a tail risk measure finds some similarity to that of the loss-based (or excess-invariant) risk measures and surplus-invariant risk measures in Cont et al. [12], Koch-Medina et al. [37, 38] and He and Peng [31], but they are essentially different concepts. A tail risk measure looks into the tail distribution of a risk, whereas a loss-based or surplus-invariant risk measure is determined by the loss part of a risk. For instance, an ES is a tail risk measure but not a loss-based or surplus-invariant risk measure.

A Classic properties of risk measures

The following properties have been standard in the theory of coherent risk measures since their introduction by Artzner et al. [3] and Föllmer and Schied [23]. For economic interpretations of these properties one may consult Föllmer and Schied [24] and Delbaen [15]. For representation results of law-invariant risk measures, see Kusuoka [39] and Frittelli and Rossaza Gianin [26].

(A0) *Law-invariance*: if $X \in \mathcal{X}$ and $X \overset{\mathrm{d}}{=} Y$, then $Y \in \mathcal{X}$ and $\rho(X) = \rho(Y)$.

(A1) *Monotonicity*: $\rho(X) \leqslant \rho(Y)$ if $X \leqslant Y$ a.s, $X, Y \in \mathcal{X}$.

(A2) *Translation-invariance*: $\rho(X - m) = \rho(X) - m$ for any $m \in \mathbb{R}$ and $X \in \mathcal{X}$.

(A3) *Convexity*: $\rho(\lambda X + (1-\lambda)Y) \leqslant \lambda\rho(X) + (1-\lambda)\rho(Y)$ for all $\lambda \in [0,1]$ and $X, Y \in \mathcal{X}$.

(A4) *Positive homogeneity*: $\rho(\lambda X) = \lambda\rho(X)$ for any $\lambda > 0$ and $X \in \mathcal{X}$.

(A5) *Subadditivity*: $\rho(X+Y) \leqslant \rho(X) + \rho(Y)$ for $X, Y \in \mathcal{X}$.

(A6) *Comonotonic additivity*: $\rho(X+Y) = \rho(X) + \rho(Y)$ if $X, Y \in \mathcal{X}$ are comonotonic. Here, two random variables X and Y are *comonotonic* if there exists $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ and $(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0$ for all $\omega, \omega' \in \Omega_0$.

Definition A.1. A *monetary risk measure* is a functional satisfying (A1) and (A2), a *convex risk measure* is a functional satisfying (A1), (A2) and (A3), and a *coherent risk measure* is a functional satisfying (A1), (A2), (A3) and (A4). For a monetary risk measure ρ, its *acceptance set* is defined as $\mathcal{A}_\rho = \{X \in \mathcal{X} : \rho(X) \leqslant 0\}$.

The last two risk measures properties that we introduce are monotonicity with respect to two classic notions of stochastic order.

Definition A.2. For $X, Y \in L^0$ (resp. L^1), we say that X is smaller than Y in stochastic order (resp. convex order), denoted as $X \prec_{\mathrm{st}} Y$ (resp. $X \prec_{\mathrm{cx}} Y$), if $\mathbb{E}[f(X)] \leqslant \mathbb{E}[f(Y)]$ for all increasing (resp. convex) functions f, provided that both expectations exist.

The corresponding risk measure properties are listed below.

(A1') $\prec_{\mathrm{st}}$-*monotonicity*: $\rho(X) \leqslant \rho(Y)$, if $X \prec_{\mathrm{st}} Y$, $X, Y \in \mathcal{X}$.

(A7) $\prec_{\mathrm{cx}}$-*monotonicity*: $\rho(X) \leqslant \rho(Y)$, if $X \prec_{\mathrm{cx}} Y$, $X, Y \in \mathcal{X}$.

The combination of monotonicity (A1) and law-invariance (A0) is equivalent to (A1'), and this simple equivalence is used in this paper. We refer to Shaked and Shanthikumar [51] for more details on stochastic orders, and Mao and Wang [41] for characterization of $\prec_{\mathrm{cx}}$-monotone risk measures.

VaRs and ES belong to the family of *distortion risk measures*, defined as

$$\rho_h(X) = \int_0^\infty (1 - h(F_X(x)))\,\mathrm{d}x - \int_{-\infty}^0 h(F_X(x))\,\mathrm{d}x, \quad X \in \mathcal{X}, \tag{13}$$

where $h : [0,1] \to [0,1]$ is a *distortion function*, that is, h is non-decreasing and $h(0) = 0$ and $h(1) = 1$. The domain $\mathcal{X}$ of ρ_h is such that (13) is properly defined for all $X \in \mathcal{X}$; in general, ρ_h is always well defined on L^∞. See Yaari [59] and Föllmer and Schied [24, Section 4.7] for more on distortion risk measures.

B Proofs in Section 3

In the proofs, for convenience we always assume that we can find a non-constant random variable independent of a given random vector whenever we need. No generality is lost here as we are interested in properties based solely on distributions of random variables.

The following lemma summarizes some simply and useful relations between X, X_p and $X^{(p)}$. The proof is an elementary exercise and is omitted here.

Lemma B.1. *Suppose $p \in (0,1)$.*

(i) For $X \in \mathcal{X}^*$, $(X^{(p)})_p \overset{\mathrm{d}}{=} X$.

(ii) For $X, Y \in \mathcal{X}$, if $X \prec_{\mathrm{st}} Y$, then $X_p \prec_{\mathrm{st}} Y_p$. For $X, Y \in \mathcal{X}^*$, if $X \prec_{\mathrm{st}} Y$, then $X^{(p)} \prec_{\mathrm{st}} Y^{(p)}$.

(iii) For $X \in \mathcal{X}$, $X \prec_{\mathrm{st}} X_p$.

Proof of Theorem 1. We repeatedly make use of (3) and (4), that is,

$$\rho(X) = \rho^*(X_p), \ X \in \mathcal{X}, \ \text{ and } \ \rho^*(X) = \rho(X^{(p)}), \ X \in \mathcal{X}^*.$$

(i) (a) (Monotonicity) We use the equivalence between monotonicity and $\prec_{\mathrm{st}}$-monotonicity. Assume ρ^* is $\prec_{\mathrm{st}}$-monotone and $X \prec_{\mathrm{st}} Y$, $X, Y \in \mathcal{X}$. By Lemma B.1 we have $X_p \prec_{\mathrm{st}} Y_p$, implying $\rho^*(X_p) \leqslant \rho^*(X_p)$ and hence ρ is $\prec_{\mathrm{st}}$-monotone. The converse is analogous.

(b) (Translation-invariance) It suffices to notice that $(X + c)_p \overset{\mathrm{d}}{=} X_p + c$ for $c \in \mathbb{R}$ and $X \in \mathcal{X}$, and $(Y + c)^{(p)} \overset{\mathrm{d}}{=} Y^{(p)} + c$ for $c \in \mathbb{R}$ and $Y \in \mathcal{X}^*$.

(c) (Positive homogeneity) It suffices to notice that $(\lambda X)_p \stackrel{\mathrm{d}}{=} \lambda X_p$ for $\lambda > 0$ and $X \in \mathcal{X}$, and $(\lambda Y)^{(p)} \stackrel{\mathrm{d}}{=} \lambda Y^{(p)}$ for $\lambda > 0$ and $Y \in \mathcal{X}^*$.

(d) (Comonotonic additivity) Assume ρ^* is comonotonically additive and $X, Y \in \mathcal{X}$ are comonotonic. Then $X + Y \stackrel{\mathrm{d}}{=} F_X^{-1}(U) + F_Y^{-1}(U)$ where $U \sim \mathrm{U}[0,1]$, and $(X+Y)_p \stackrel{\mathrm{d}}{=} F_X^{-1}(U_p) + F_Y^{-1}(U_p)$. It follows that

$$\rho(X+Y) = \rho^*((X+Y)_p) = \rho^*(F_X^{-1}(U_p) + F_Y^{-1}(U_p)) = \rho^*(X_p) + \rho^*(Y_p) = \rho(X) + \rho(Y).$$

For the converse, assume ρ is comonotonically additive and $X, Y \in \mathcal{X}^*$ are comonotonic. Let $B \sim \mathrm{Bern}(p)$ be independent of X and Y, and write $x = \text{ess-inf}(X)$ and $y = \text{ess-inf}(Y)$. Then

$$(X+Y)^{(p)} \stackrel{\mathrm{d}}{=} \text{ess-inf}(X+Y)B + (X+Y)(1-B) = xB + X(1-B) + yB + Y(1-B).$$

Note that $xB + X(1-B)$ and $yB + Y(1-B)$ are comonotonic. Therefore,

$$\begin{aligned}\rho^*(X+Y) = \rho((X+Y)^{(p)}) &= \rho(xB + X(1-B) + yB + Y(1-B)) \\ &= \rho(xB + X(1-B)) + \rho(yB + Y(1-B)) \\ &= \rho(X^{(p)}) + \rho(Y^{(p)}) = \rho^*(X) + \rho^*(Y).\end{aligned}$$

Hence ρ^* is comonotonically additive.

(ii) For $X, Y \in \mathcal{X}^*$, let $B \sim \mathrm{Bern}(p)$ be independent of X and Y, and write $x = \text{ess-inf}(X)$, $y = \text{ess-inf}(Y)$, $z = \text{ess-inf}(X+Y)$ and $w = \min\{x, y\}$. We first note that for any $Z \in \mathcal{X}^*$ independent of B and $t \leqslant \text{ess-inf}(Z)$,

$$(tB + Z(1-B))_p \stackrel{\mathrm{d}}{=} (\text{ess-inf}(Z)B + Z(1-B))_p \stackrel{\mathrm{d}}{=} Z. \tag{14}$$

(a) (Subadditivity) Assume ρ is subadditive. By (14) and noting that $z \geqslant x + y$, we have

$$\begin{aligned}\rho^*(X+Y) &= \rho(zB + X(1-B) + Y(1-B)) \\ &= \rho((x+y)B + X(1-B) + Y(1-B)) \\ &\leqslant \rho(xB + X(1-B)) + \rho(yB + Y(1-B)) = \rho^*(X) + \rho^*(Y).\end{aligned}$$

Hence, ρ^* is subadditive.

(b) (Convexity) The proof is analogous to (a).

(c) ($\prec_{\mathrm{cx}}$-monotonicity) Assume ρ is $\prec_{\mathrm{cx}}$-monotone and $X \prec_{\mathrm{cx}} Y$. By (14) and noting that $wB + X(1-B) \prec_{\mathrm{cx}} wB + Y(1-B)$, we have

$$\begin{aligned}\rho^*(X) = \rho(xB + X(1-B)) &= \rho(wB + X(1-B)) \\ &\leqslant \rho(wB + Y(1-B)) \leqslant \rho(yB + Y(1-B)) = \rho^*(Y).\end{aligned}$$

Hence, ρ^* is $\prec_{\mathrm{cx}}$-monotone.

(iii) From (i)-(ii), we know that ρ^* is monetary if and only if ρ is monetary, and ρ is a coherent (convex, monetary) risk measure implies ρ^* is coherent (convex, monetary). Thus, it remains to show that ρ^* is a coherent (convex) risk measure implies ρ is coherent (convex). To this end, we need to use some further result in Theorem 3 on risk aggregation for tail risk measures to show that the convexity of ρ^* implies the convexity of ρ, assuming that ρ^* is monetary. For any $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$, by Theorem 3, we have

$$\begin{aligned}\rho(\lambda X + (1-\lambda)Y) &\leqslant \sup\{\rho^*(\lambda Z + (1-\lambda)W) : Z, W \in \mathcal{X},\ Z \sim F_X^{[p]},\ W \sim F_Y^{[p]}\}\\ &\leqslant \sup\{\lambda\rho^*(Z) + (1-\lambda)\rho^*(W) : Z, W \in \mathcal{X},\ Z \sim F_X^{[p]},\ W \sim F_Y^{[p]}\}\\ &= \lambda\rho^*(X_p) + (1-\lambda)\rho^*(Y_p)\\ &= \lambda\rho(X) + (1-\lambda)\rho(Y).\end{aligned}$$

That is, ρ is convex, and this completes the proof. □

Proof of Theorem 2. Suppose ρ is a monetary p-tail risk measure with $\rho(0) = 0$, then ρ^* is a monetary risk measure on $\mathcal{X}^*$ with $\rho^*(0) = 0$ by using Theorem 1. It follows that for any $X \in \mathcal{X}$

$$\rho(X) = \rho^*(X_p) \geqslant \text{ess-inf}(X_p) = \mathrm{VaR}_p^R(X).$$

For the second assertion, by Theorem 9 of Kusuoka [39], any law-invariant coherent risk measure ρ which dominates VaR_p^R also dominates ES_p on L^∞. □

C Proof in Section 4

Proof of Theorem 3. We first assume $F_i^{-1}(p) > 0$, $i = 1, \dots, n$.

(i) We first show the "$\leqslant$" sign in (7). Take any $S \in \mathcal{S}_n(F_1, \dots, F_n)$ and write $S = X_1 + \cdots + X_n$ where $X_i \sim F_i$, $i = 1, \dots, n$. Then $S = F_1^{-1}(U_{X_1}) + \cdots + F_n^{-1}(U_{X_n})$ almost surely. For $i = 1, \dots, n$, denote by f_i the conditional distribution function of U_{X_i} given $U_S > p$, that is, $f_i(t) = \mathbb{P}(U_{X_i} \leqslant t | U_S > p)$, $t \in [0,1]$. It follows that $\mathbb{P}(f_i(U_{X_i}) \leqslant x | U_S > p) = x$ for $x \in [0,1]$, and thus $f_i(U_{X_i})$ conditionally on $U_S > p$ is uniformly distributed over $[0,1]$.

Let $V_i = p + (1-p) f_i(U_{X_i})$, $i = 1, \dots, n$. Note that for $t \in [0,1]$,

$$p + (1-p) f_i(t) = p + \mathbb{P}(U_{X_i} \leqslant t, U_S > p) \geqslant p + (1 - (1-t) - p) = t.$$

Therefore, $V_i \geqslant U_{X_i}$, and V_i is uniformly distributed over $[p, 1]$ conditionally on $U_S > p$, $i = 1, \dots, n$. Write $S' = (F_1^{-1}(V_1) + \cdots + F_n^{-1}(V_n))\mathbb{1}_{\{U_S > p\}}$. As $F_i^{-1}(p) > 0$ and $V_i \geqslant p$ for $i = 1, \dots, n$, $\mathbb{P}(S' > 0) = \mathbb{P}(U_S > p) = 1 - p$. Therefore, $\mathbb{1}_{\{U_S > p\}} = \mathbb{1}_{\{U_{S'} > p\}}$ a.s. We have $S'\mathbb{1}_{\{U_{S'} > p\}} \geqslant S\mathbb{1}_{\{U_S > p\}}$, which implies $\rho(S') \geqslant \rho(S)$ since ρ is a p-tail risk measure.

Finally, let $\hat{V}_1, \dots, \hat{V}_n$ be uniform random variables on $[p, 1]$ such that $(\hat{V}_1, \dots, \hat{V}_n)$ has joint distribution identical to the conditional distribution of $(V_1, \dots, V_n)$ on $\{U_S > p\}$. It follows that for $x > 0$,

$$\mathbb{P}(S'_p \leqslant x) = \mathbb{P}\left(\sum_{i=1}^n F_i^{-1}(V_i)\mathbb{1}_{\{U_S>p\}} \leqslant x \middle| U_S > p\right) = \mathbb{P}\left(\sum_{i=1}^n F_i^{-1}(\hat{V}_i) \leqslant x\right).$$

Write $T = \sum_{i=1}^n F_i^{-1}(\hat{V}_i)$. As $\hat{V}_i$ is uniformly distributed over $[p, 1]$, $F_i^{-1}(\hat{V}_i) \sim F_i^{[p]}$, $i = 1, \dots, n$. Hence, $T \in \mathcal{S}_n(F_1^{[p]}, \dots, F_n^{[p]})$. Finally, we have $\rho(S) \leqslant \rho(S') = \rho^*(S'_p) = \rho^*(T)$, therefore the "$\leqslant$" sign in (7) holds.

(ii) We proceed to show the "$\geqslant$" sign in (7). Take any $T \in \mathcal{S}_n(F_1^{[p]}, \dots, F_n^{[p]})$ and write $T = Y_1 + \dots + Y_n$ where $Y_i \sim F_i^{[p]}$, $i = 1, \dots, n$. Let V be a uniform $[0, 1]$ random variable independent of $Y_1, \dots, Y_n$. Write $X_i = \mathbb{1}_{\{V>p\}}Y_i + \mathbb{1}_{\{V\leqslant p\}}F_i^{-1}(V)$, $i = 1, \dots, n$ and $S = X_1 + \dots + X_n$. Then we have $S_p \overset{\mathrm{d}}{=} T$ and $\rho(S) = \rho^*(S_p) = \rho^*(T)$, thus the "$\geqslant$" sign in (7) holds.

Now we consider the general case in which $F_1^{-1}(p), \dots, F_n^{-1}(p)$ may not be positive. For $i = 1, \dots, n$, take $X_i \sim F_i$, and let G_i be the distribution of $X_i - F_i^{-1}(p) + 1$. Clearly $G_i^{-1}(p) = 1 > 0$. Let $\tau(X) = \rho(X + \sum_{i=1}^n F_i^{-1}(p) - n)$ and $\tau^* = \rho^*(X + \sum_{i=1}^n F_i^{-1}(p) - n)$, $X \in \mathcal{X}$. Then (τ, τ^*) is also a pair of tail risk measure and the corresponding generator. From the results in (i) and (ii) we have

$$\sup\{\tau(S) : S \in \mathcal{S}_n(G_1, \dots, G_n)\} = \sup\{\tau^*(T) : T \in \mathcal{S}_n(G_1^{[p]}, \dots, G_n^{[p]})\}. \tag{15}$$

Note that $S \in \mathcal{S}_n(F_1, \dots, F_n)$ is equivalent to $S - \sum_{i=1}^n F_i^{-1}(p) + n \in \mathcal{S}_n(G_1, \dots, G_n)$. Therefore, (15) is equivalent to

$$\sup\{\rho(S) : S \in \mathcal{S}_n(F_1, \dots, F_n)\} = \sup\{\rho^*(T) : T \in \mathcal{S}_n(F_1^{[p]}, \dots, F_n^{[p]})\},$$

and the proof is complete. □

D Proofs in Section 5

We first present a lemma on the acceptance set of a monetary tail risk measure.

Lemma D.1. *For $p \in (0, 1)$, a monetary risk measure ρ is a p-tail risk measure if and only if its acceptance set $\mathcal{A}_\rho$ satisfies that, for $X \in \mathcal{A}_\rho$ and $Y \in \mathcal{X}$, $X_p \overset{\mathrm{d}}{=} Y_p$ implies $Y \in \mathcal{A}_\rho$.*

Proof. Proof of Lemma D.1. The implication "$\Rightarrow$" is trivial by definition. To show "$\Leftarrow$", for $X, Y \in \mathcal{X}$ such that $X_p \overset{\mathrm{d}}{=} Y_p$, notice that $(X - \rho(X))_p \overset{\mathrm{d}}{=} (Y - \rho(X))_p$ and $X - \rho(X) \in \mathcal{A}_\rho$. Therefore, $\rho(Y - \rho(X)) \leqslant 0$, which means $\rho(Y) \leqslant \rho(X)$. By symmetry, $\rho(X) = \rho(Y)$. □ □

Proof of Theorem 4. Note that both ρ_ℓ and (10) stay the same if ℓ is replaced by its left-continuous version; thereby we safely assume that ℓ is left-continuous. First, the left-continuity of ℓ implies that for $X \in \mathcal{X}$, $\rho(X) \leqslant 0 \Leftrightarrow \mathbb{E}[\ell(X)] \leqslant 0$; this fact will be used frequently below. From $\rho(0) = 0$ it is easy to verify that $\ell(x) \leqslant 0$ if $x < 0$, and $\ell(x) > 0$ if $x > 0$.

$\Rightarrow$: For $x, y \in \mathbb{R}$, let $Z_{x,y}$ be a random variable with the bi-atomic distribution $p\delta_x + (1-p)\delta_y$. For $x \leqslant y$, noting that $(Z_{x,y})_p \overset{\mathrm{d}}{=} y$, one has $\rho(Z_{x,y}) = \rho(y) = y$ as ρ is a p-tail risk measure. In particular, for any $x < 0 < y$, $\rho(Z_{x,y}) = y > 0$ and hence

$$\mathbb{E}\left[\ell(Z_{x,y})\right] = p\ell(x) + (1-p)\ell(y) > 0. \tag{16}$$

Now suppose $\ell(x) < \ell(z)$ for some $x < z < 0$ and take $0 < y$. It follows that $p\ell(x)+(1-p)\ell(z) < \ell(z) \leqslant 0$. Together with (16), there exists $\lambda \in (0,1)$ such that

$$0 = p\ell(x) + (1-p)(\lambda\ell(z) + (1-\lambda)\ell(y)) < p\ell(z) + (1-p)(\lambda\ell(z) + (1-\lambda)\ell(y)).$$

For $s, t, w \in \mathbb{R}$, let $Z_{s,t,w}$ be a random variable with distribution $p\delta_s+(1-p)\lambda\delta_t+(1-p)(1-\lambda)\delta_w$. Then $\mathbb{E}[\ell(Z_{x,z,y})] = 0$ and $\mathbb{E}[\ell(Z_{z,z,y})] > 0$ imply $\rho(Z_{x,z,y}) \leqslant 0$ and $\rho(Z_{z,z,y}) > 0$ respectively, leading to a contradiction to the fact that $(Z_{x,z,y})_p = (Z_{z,z,y})_p$. Therefore, $\ell(x) = \ell(z) = \ell(-1)$ for all $x < z < 0$.

$\Leftarrow$: Now we assume that a risk measure ρ is induced by a loss function ℓ satisfying condition (10). Let $c = \ell(-1)$. For any $X \in \mathcal{X}$, if $F_X^{-1}(p+) > 0$, then we have $\rho(X) > 0$ because

$$\mathbb{E}[\ell(X)] = \int_0^p \ell(F_X^{-1}(t))\,\mathrm{d}t + \int_p^1 \ell(F_X^{-1}(t))\,\mathrm{d}t \geqslant \frac{1}{1-p}\int_p^1 \left(pc + (1-p)\ell(F_X^{-1}(t))\right)\,\mathrm{d}t > 0;$$

if $F_X^{-1}(p+) \leqslant 0$, we have

$$\mathbb{E}[\ell(X)] = \int_0^p \ell(F_X^{-1}(t))\,\mathrm{d}t + \int_p^1 \ell(F_X^{-1}(t))\,\mathrm{d}t = pc + \int_p^1 \ell(F_X^{-1}(t))\,\mathrm{d}t.$$

To combine both cases, for all $X \in \mathcal{X}$, $\rho(X) \leqslant 0$ if and only if

$$pc + \int_p^1 \ell(F_X^{-1}(t))\,\mathrm{d}t \leqslant 0. \tag{17}$$

Therefore, by Lemma D.1, ρ is a p-tail risk measure. $\square$

Some consequences of Theorem 4 are summarized in the following lemma. In summary, there is no convex tail shortfall risk measure, and all positively homogeneous tail shortfall risk measures are the left-quantiles.

Lemma D.2. *Suppose $p \in (0,1)$ and ρ is a p-tail shortfall risk measure induced by ℓ. Then*

(i) $\ell(\rho(0)-) < 0 < \ell(\rho(0)+)$.

(ii) *ρ is not convex.*

(iii) *If ρ is positively homogeneous, then $\rho = \mathrm{VaR}_q^L$ for some $q \in (p,1)$.*

Proof of Lemma D.2. By Theorem 4, if $\rho(0) = 0$, then the loss function ℓ satisfies (10). Write $c = \ell(-1)$.

(i) Assume $\rho(0) = 0$, and other cases can be obtained via a shift in the argument. By definition of a loss function, $\inf_{x\in\mathbb{R}} \ell(x) < 0$. Therefore, $\ell(0-) = c = \inf_{x\in\mathbb{R}} \ell(x) < 0$. On the other hand, since $pc + (1-p)\ell(y) > 0$ for all $y > 0$, taking $y \to 0+$ we have $\ell(0+) > 0$.

(ii) From (i), ℓ is discontinuous at $\rho(0)$, hence it is not convex on $\mathbb{R}$, and in turn ρ is not a convex risk measure.

(iii) We have $\rho(0) = 0$ from positive homogeneity. Without loss of generality we assume ℓ is left-continuous. For $x < 0 < y$, take $Z_{x,y} \sim q\delta_x + (1-q)\delta_y$ where $q = \ell(y)/(\ell(y) - c) \in (p, 1)$. Then $\mathbb{E}[\ell(Z_{x,y})] = qc + (1-q)\ell(y) = 0$, and hence $\rho(Z_{x,y}) \leqslant 0$. From the positive homogeneity of ρ, we have $\rho(\lambda Z_{x,y}) \leqslant 0$ for all $\lambda > 0$, and hence $qc + (1-q)\ell(\lambda y) \leqslant 0$ implying $\ell(\lambda y) \leqslant \ell(y)$ for all $\lambda > 0$. Noting that ℓ is non-decreasing, we have $\ell(z) = \ell(y)$ for all $z > y$. This means $\ell(y) = \ell(1)$ for all $y > 0$. Therefore, the loss function ℓ satisfies

$$\ell(x) = (\ell(1) - c)\mathbb{1}_{\{x>0\}} + c = (\ell(1) - c)\left(\mathbb{1}_{\{x>0\}} - (1-q)\right).$$

Comparing with (9), ρ and VaR_q^L have the same acceptance set, thus $\rho = \mathrm{VaR}_q^L$. □

Proof of Theorem 5. $\Rightarrow$: Write $\mathcal{N}_\rho = \{F_X : X \in \mathcal{A}_\rho\}$. First, we assume

$$\text{there exists } x \in \mathbb{R} \text{ such that for all } y \in \mathbb{R},\ (1-\lambda)\delta_x + \lambda\delta_y \in \mathcal{N}_\rho \text{ for some } \lambda > 0. \tag{18}$$

Condition (18) implies the assumption in Theorem 3.1 of Weber [58], which states

$$\begin{gathered} \text{there exists } x \in \mathbb{R} \text{ with } \delta_x \in \mathcal{N}_\rho \text{ such that for } y \in \mathbb{R} \text{ and } \delta_y \in \mathcal{N}_\rho^c, \\ (1-\alpha)\delta_x + \alpha\delta_y \in \mathcal{N}_\rho \text{ for sufficiently small } \alpha > 0. \end{gathered} \tag{19}$$

To see that (18) implies (19), we simply need to verify that x in (18) also satisfies (19). First, taking $y = x$ in (18) gives $\delta_x \in \mathcal{N}_\rho$. Note that ρ is a monetary and law-invariant risk measure, and hence it is $\prec_{\mathrm{st}}$-monotone as in property (A1'). For $y \in \mathbb{R}$ with $\delta_y \in \mathcal{N}_\rho^c$, it is clear that $y < x$ by $\prec_{\mathrm{st}}$-monotonicity of ρ and $\delta_x \in \mathcal{N}_\rho$. If $(1-\lambda)\delta_x + \lambda\delta_y \in \mathcal{N}_\rho$ for some $\lambda > 0$, using $\prec_{\mathrm{st}}$-monotonicity of ρ again, and noting that $y < x$, we have $(1-\alpha)\delta_x + \alpha\delta_y \in \mathcal{N}_\rho$ for $\alpha \in (0, \lambda)$. Therefore, x in (18) also satisfies (19).

From DLC, $\mathcal{N}_\rho$ is closed with respect to weak convergence. Note that the CxLS property is equivalent to $\mathcal{N}_\rho$ and $\mathcal{N}_\rho^c$ both being convex. By Theorem 3.1 of Weber [58], the monetary risk measure ρ satisfying DLC, CxLS and condition (18) is necessarily a shortfall risk measure. Finally, by Lemma D.2 (iii), a positive homogeneous p-tail shortfall risk measure is necessarily VaR_q^L for some $q \in (p, 1)$.

Next we assume (18) does not hold. By positive homogeneity, we have $\rho(0) = 0$. Let $Z_{y,\lambda} \sim \lambda\delta_0 + (1-\lambda)\delta_y$ for $y \in \mathbb{R}$ and $\lambda \in (0, 1)$. From the opposite of (18), there exists $y_0 > 0$ such that $\rho(Z_{y_0,\lambda}) > 0$ for all $\lambda \in (0,1)$, and by positive homogeneity again, we have $\rho(Z_{y,\lambda}) > 0$ for all $\lambda \in (0,1)$ and all $y > 0$.

Arbitrarily take $y > 0$ and let $k_\lambda = \rho(Z_{y,\lambda})$. Note that $0 \leqslant k_\lambda \leqslant y$. For $\lambda \in (0,p]$, since $(Z_{y,\lambda})_p \overset{\mathrm{d}}{=} y$, we have $k_\lambda = y$. Now, take $\lambda \in (p,1)$, and write $\alpha = p/\lambda$ and $\lambda_0 = 1-\alpha(1-\lambda)$. Let $X \sim (1-\alpha)\delta_0 + \alpha(\lambda\delta_{-k_\lambda} + (1-\lambda)\delta_{y-k_\lambda})$. Then $\rho(X) = \rho(Z_{y,\lambda} - k_\lambda) = \rho(0) = 0$ by the CxLS property. Moreover, $X_p \overset{\mathrm{d}}{=} (Z_{y-k_\lambda,\lambda_0})_p$ implies $\rho(Z_{y-k_\lambda,\lambda_0}) = \rho(X) = 0$. Since $\rho(Z_{y-k_\lambda,\lambda}) > 0$ for all $\lambda \in (0,1)$ and all $y - k_\lambda > 0$, then $k_\lambda = y_0$. Noting that both λ and y are arbitrary here, thus $\rho(Z_{y,\lambda}) = y$ for all $\lambda \in (0,1)$ and $y > 0$.

For any random variable Z taking values in a finite set $\{a_1,\dots,a_n\} \subset \mathbb{R}$, we have $Z \sim \sum_{i=1}^n \beta_i(\lambda_i\delta_{a_i} + (1-\lambda_i)\delta_{\text{ess-sup}(Z)})$, where $\beta_i \geqslant 0$ and $\sum_{i=1}^n \beta_i = 1$ and $\lambda_i \in (0,1)$. Let $Z_i \sim \lambda_i\delta_{a_i} + (1-\lambda_i)\delta_{\text{ess-sup}(Z)}$, $i = 1,\dots,n$. It follows that $\rho(Z_i - a_i) = \text{ess-sup}(Z) - a_i$ and hence $\rho(Z_i) = \text{ess-sup}(Z)$, $i = 1,\dots,n$. By CxLS, we have $\rho(Z) = \text{ess-sup}(Z)$.

For a general random variable $Z \in \mathcal{X}$ with $Z \geqslant 0$, write $M = \text{ess-sup}(Z)$ and let $Z_n = M\sum_{i=0}^{2^n-1} \frac{i}{2^n}\mathbb{1}_{\{Z\in(\frac{i}{2^n},\frac{i+1}{2^n}]\}}$, $n \in \mathbb{N}$. Then $Z_n \uparrow Z$, and hence $\rho(Z) \geqslant \rho(Z_n) = \text{ess-sup}(Z_n) = M\frac{2^n-1}{2^n}$. Therefore, $\rho(Z) \geqslant M$, and together with $\rho(Z) \leqslant M$, we have $\rho(Z) = M$. Finally, since ρ is monetary, we have $\rho(Z) = \text{ess-sup}(Z)$ for all $Z \in \mathcal{X}$.

In summary, $\rho = \mathrm{VaR}_q^L$ for some $q \in (p,1]$.

$\Leftarrow$: One can directly verify that for $q \in (p,1]$, $\rho = \mathrm{VaR}_q^L$ is an elicitable, monetary and positively homogeneous p-tail risk measure satisfying DLC. □

Proof of Theorem 6. The "if" part is straightforward to check and below we show the "only-if" part. Note that the properties of ρ are also satisfied on L^∞. Take $p \in (0,1)$ such that ρ is a p-tail risk measure. By Theorem 5, $\rho = \mathrm{VaR}_q^L$ on L^∞ for some $q \in (p,1]$. For $X \in \mathcal{X}$ and $X \notin L^\infty$, let $Y = \max\{X, \mathrm{VaR}_p^L(X)\}$. Clearly, $X_p \overset{\mathrm{d}}{=} Y_p$ and hence $\rho(X) = \rho(Y)$. Let $Y_n = \min\{Y,n\}$ for $n \in \mathbb{N}$. Then $Y_n \in L^\infty$, and we have $\rho(Y_n) = \mathrm{VaR}_q^L(Y_n)$. Note that for $q \in (p,1]$,

$$\lim_{n\to\infty} \mathrm{VaR}_q^L(Y_n) = \mathrm{VaR}_q^L(Y) = \mathrm{VaR}_q^L(X).$$

By DLC, $\rho(Y) \leqslant \lim_{n\to\infty} \mathrm{VaR}_q^L(Y_n)$. On the other hand, since $Y_n \leqslant Y$ and ρ is monotone, we have $\rho(Y) \geqslant \lim_{n\to\infty} \mathrm{VaR}_q^L(Y_n)$. Combining the above two inequalities, we have

$$\rho(X) = \rho(Y) = \lim_{n\to\infty} \mathrm{VaR}_q^L(Y_n) = \mathrm{VaR}_q^L(X).$$

Thus, $\rho = \mathrm{VaR}_q$ on $\mathcal{X}$. □

E Some other examples of tail risk measures

Example E.1 (Tail entropic risk measure)**.** Let $\mathcal{X} = L^\infty$ and ρ^* be an entropic risk measure in Föllmer and Schied [23] (called an exponential principle in Gerber [29]), defined as

$$\rho^*(X) = \frac{1}{\beta} \log \mathbb{E}[e^{\beta X}], \quad X \in L^\infty,$$

where $\beta > 0$. Then, we have

$$\rho(X) = \frac{1}{\beta} \log \mathbb{E}[e^{\beta X_p}] = \frac{1}{\beta} \mathrm{ES}_p(e^{\beta X}), \quad X \in L^\infty.$$

As ρ^* is a convex risk measure, ρ is also a convex risk measure by Theorem 1. Based on discussions in Section 5, ρ^* is elicitable but ρ is not. The tail entropic risk measure ρ belongs to the class of *distortion-exponential risk measures* in Tsanakas [54].

Example E.2 (Tail distortion risk measure)**.** Let $h : [0, 1] \to [0, 1]$ be a distortion function, and $\mathcal{X}$ be a set of random variables such that the distortion risk measure ρ_h in (13) is well-defined on $\mathcal{X}$. Take $\rho^* = \rho_h$. Then p-tail risk measure ρ generated by ρ^* can be expressed as

$$\rho(X) = \mathrm{VaR}_p^R(X) + \int_{\mathrm{VaR}_p^R(X)}^{\infty} \left(1 - h\left(\frac{F_X(x) - p}{1 - p}\right)\right) \mathrm{d}x = \rho_{h_p}(X), \quad X \in \mathcal{X}. \tag{20}$$

where $h_p : [0, 1] \to [0, 1]$ is a distortion function given by

$$h_p(t) = h\left(\frac{t - p}{1 - p}\right) \mathbb{1}_{\{t \geqslant p\}}, \quad t \in [0, 1].$$

That is, ρ is a distortion risk measure with distortion function h_p which takes value 0 on $[0, p]$. Note that $F_X(x) \geqslant p$ for $x \geqslant \mathrm{VaR}_p^R(X)$, and thus the right quantile $\mathrm{VaR}_p^R(X)$ in (20) cannot be replaced by $\mathrm{VaR}_p^L(X)$. For a distortion risk measure ρ_h, subadditivity is equivalent to the convexity of h. Note that h_p is convex if and only if h is, and therefore ρ is coherent if and only if so is ρ^*, a result also implied by Theorem 1.

Example E.3 (Tail risk measures generated by shortfall risk measures)**.** Let $\mathcal{X} = L^\infty$ and ρ^* be a shortfall risk measure induced by a loss function ℓ, defined in Section 5 as,

$$\rho^*(X) = \inf\{m \in \mathbb{R} : \mathbb{E}[\ell(X - m)] \leqslant 0\}, \quad X \in L^\infty. \tag{21}$$

The special case of $\ell(x) = \exp(\beta x) - 1,\ x \in \mathbb{R}$ for some $\beta > 0$ reduces to Example E.1. With ρ^* in (21), we have

$$\rho(X) = \inf\{m \in \mathbb{R} : \mathrm{ES}_p[\ell(X - m)] \leqslant 0\}, \quad X \in L^\infty, \tag{22}$$

Note that ρ is a risk measure induced by the rank-dependent utility (RDU; see Quiggin [47]) functional $X \mapsto \mathrm{ES}_p(\ell(X))$ via (22). Clearly, ρ is also monetary, and by Theorem 1, the following are equivalent: (i) ρ^* is convex; (ii) ρ is convex; (iii) ℓ is convex.

F p-tail distortion risk measures

In Theorem 2 of the paper, we establish the essential importance of VaRs and ES as benchmarks for tail risk measures by showing that VaR_p^R serves as the smallest monetary p-tail risk measure, and ES_p serves as the smallest coherent p-tail risk measure. For a *distortion risk measure* defined in (13) dominating the corresponding VaR is equivalent to tail-relevance.

Theorem F.1. *Let $p \in (0,1)$. If ρ is a distortion risk measure, then the following are equivalent.*

(a) ρ is a p-tail risk measure.

(b) The distortion function of ρ takes value 0 on $[0,p]$.

(c) $\rho \geqslant \mathrm{VaR}_p^R$ on $\mathcal{X}$.

Moreover, if ρ is coherent, then (a)-(c) are equivalent to

(d) $\rho \geqslant \mathrm{ES}_p$ on $\mathcal{X}$.

Proof of Theorem F.1. The implication (a)⇒(c) is implied by Theorem 2.

(c)⇒(b): Let ρ_h be a distortion risk measure with distortion function h such that $\rho_h(X) \geqslant \mathrm{VaR}_p^R(X)$ for any $X \in \mathcal{X}$. Let $X \in \mathcal{X}$ be a $\mathrm{Bern}(1-p)$ random variable. From the definition in (13), we have

$$\mathrm{VaR}_p^R(X) \leqslant \rho_h(X) = \int_0^\infty (1 - h(F_X(x)))\, \mathrm{d}x = 1 - h(p) = \mathrm{VaR}_p^R(X) - h(p).$$

Therefore, $h(p) = 0$, and as h is non-decreasing, we have $h(t) = 0$ for $t \in [0,p]$.

(b)⇒(a): Let ρ_h be a distortion risk measure with distortion function h and $h(t) = 0$ for any $t \in [0,p]$. It follows that, for $X, Y \in \mathcal{X}$ satisfying $X_p \overset{\mathrm{d}}{=} Y_p$,

$$\begin{aligned}
\rho_h(X) &= \int_{\mathrm{VaR}_p^R(X)}^\infty (1 - h(F_X(x)))\, \mathrm{d}x + \mathrm{VaR}_p^R(X) - \int_{-\infty}^{\mathrm{VaR}_p^R(X)} h(F_X(x))\, \mathrm{d}x \\
&= \int_{\mathrm{VaR}_p^R(X)}^\infty (1 - h(F_X(x)))\, \mathrm{d}x + \mathrm{VaR}_p^R(X) \\
&= \int_{\mathrm{VaR}_p^R(Y)}^\infty (1 - h(F_Y(x)))\, \mathrm{d}x + \mathrm{VaR}_p^R(Y) = \rho_h(Y).
\end{aligned}$$

That is, ρ_h is a p-tail risk measure.

Next we assume ρ is coherent. (d)⇒(c) is straightforward.

(b)⇒(d): ES_p is a distortion risk measure with distortion function

$$h_p(t) = \frac{t-p}{1-p} \mathbb{1}_{\{t \geqslant p\}}, \quad t \in [0,1].$$

Note that h is convex on $[0,1]$ because ρ_h is coherent. As $h(p) = h_p(p) = 0$, h_p is linear on $[p,1]$ and h is convex on $[p,1]$, we have $h(t) \leqslant h_p(t)$ for all $t \in [0,1]$, and as a consequence $\mathrm{ES}_p(X) \leqslant \rho_h(X)$ for all $X \in \mathcal{X}$ by its definition (13). $\square$

As a by-product, Theorem F.1 characterizes all p-tail risk measures which belong to the class of distortion risk measures, equivalently, the class of all comonotonically additive and monetary p-tail risk measures.

G Dual representations

We present dual representation results for tail risk measures, assuming some further functional properties. The following representation results (i)-(iii) are based on the Kusuoka representation for coherent risk measures in Kusuoka [39] and Frittelli and Rossaza Gianin [26]. (iv) is based on the representation for monetary and $\prec_{\mathrm{cx}}$-monotone risk measures in Mao and Wang [41]. The details of the proof are put in the appendix. To make the illustration concise, we assume $\mathcal{X} = L^\infty$ in this section. The case for risk measures on larger spaces is analogous albeit some continuity has to be assumed; see Remark G.1.

Theorem G.1. *Suppose $p \in (0,1)$, $\mathcal{D}_p$ is the set of distributions functions on $[p,1]$ and ρ is a functional mapping $\mathcal{X} = L^\infty$ to $\mathbb{R}$.*

(i) ρ is a comonotonically additive and coherent p-tail risk measure if and only if there exists $g \in \mathcal{D}_p$ such that $\rho(X) = \int_p^1 \mathrm{ES}_q(X)\,\mathrm{d}g(q),\ X \in \mathcal{X}$.

(ii) ρ is a coherent p-tail risk measure if and only if there exists a set $\mathcal{G} \subset \mathcal{D}_p$ such that $\rho(X) = \sup_{g\in\mathcal{G}} \int_p^1 \mathrm{ES}_q(X)\,\mathrm{d}g(q),\ X \in \mathcal{X}$.

(iii) ρ is a convex and monetary p-tail risk measure if and only if there exists a function $v : \mathcal{D}_p \to \mathbb{R}$ such that $\rho(X) = \sup_{g\in\mathcal{D}_p}\{\int_p^1 \mathrm{ES}_q(X)\,\mathrm{d}g(q) - v(g)\},\ X \in \mathcal{X}$.

(iv) ρ is a $\prec_{\mathrm{cx}}$-monotone and monetary p-tail risk measure if and only if there exists a set $\mathcal{H}$ of functions mapping $[p,1]$ to $\mathbb{R}$ such that $\rho(X) = \inf_{\alpha\in\mathcal{H}} \sup_{q\in[p,1]}\{\mathrm{ES}_q(X) - \alpha(q)\},\ X \in \mathcal{X}$.

Remark G.1. If $\mathcal{X} = L^p$ for $p \in [1,\infty)$, the representation results (i)-(iii) in Theorem G.1 hold if one assumes that ρ is finite and satisfies the *Fatou property* (or lower semi-continuity), namely, $\liminf_{n\to\infty} \rho(X_n) \geqslant \rho(X)$ if $X, X_1, X_2, \dots \in \mathcal{X} = L^p$ and $X_n \xrightarrow{p} X$ as $n \to \infty$. The proof is the same as above by noting that the Fatou property guarantees the corresponding representation results; see Lemma 2.14 of Svindland [53]. For the representation result (iv) to hold, the Fatou property is not necessary, see Theorem B.2 and Example B.1 of Mao and Wang [41].

Proof of Theorem G.1. The "if" statements are all straightforward to verify. We only show the "only-if" statements of the theorem. First, note that all law-invariant convex risk measures ρ on L^∞ have the so-called Fatou property (Theorem 30 of Delbaen [15]; cf. Jouini et al. [33]). This property is used to validate the representation results in (i)-(iii).

(i) From Theorem 1, ρ^* is a comonotonically additive, coherent and law-invariant risk measure. By Theorem 7 of Kusuoka [39], it has a representation

$$\rho^*(X) = \int_0^1 \mathrm{ES}_q(X)\,\mathrm{d}h(q), \quad X \in \mathcal{X},$$

where h is a distribution function on $[0,1]$. Note that, as seen in Section 3,

$$\mathrm{ES}_q(X_p) = \mathrm{ES}_{p+q-pq}(X), \quad X \in \mathcal{X}.$$

It follows that,

$$\int_0^1 \mathrm{ES}_q(X_p)\,\mathrm{d}h(q) = \int_0^1 \mathrm{ES}_{p+q-pq}(X)\,\mathrm{d}h(q) = \int_p^1 \mathrm{ES}_r(X)\,\mathrm{d}g(r), \quad X \in \mathcal{X}, \tag{23}$$

where $g(r) = h\left((r-p)/(1-p)\right)$, $r \in [p,1]$, and thus $g \in \mathcal{D}_p$. Therefore, $\rho(X) = \rho^*(X_p) = \int_p^1 \mathrm{ES}_q(X)\,\mathrm{d}g(q)$, $X \in \mathcal{X}$.

(ii) From Theorem 1, ρ^* is a coherent and law-invariant risk measure, and by Theorem 4 of Kusuoka [39], it has a representation

$$\rho^*(X) = \sup_{h \in \mathcal{G}_0} \int_0^1 \mathrm{ES}_q(X)\,\mathrm{d}h(q), \quad X \in \mathcal{X},$$

where $\mathcal{G}_0$ is a set of distribution functions on $[0,1]$. The rest of the proof is straightforward by using (23).

(iii) From Theorem 1, ρ^* is a law-invariant convex risk measure, and by Theorem 7 of Frittelli and Rossaza Gianin [26], it has a representation

$$\rho^*(X) = \sup_{h \in \mathcal{D}_0} \left\{ \int_0^1 \mathrm{ES}_q(X)\,\mathrm{d}h(q) - u(h) \right\}, \quad X \in \mathcal{X},$$

where $u : \mathcal{D}_0 \to \mathbb{R}$ is a function. The rest of the proof is straightforward by using (23).

(iv) Since ρ is a $\prec_{\mathrm{cx}}$-monotone and monetary risk measure, and by Theorem 3.1 of Mao and Wang [41], it has a representation

$$\rho(X) = \inf_{\beta \in \mathcal{H}_0} \sup_{q \in [0,1]} \{\mathrm{ES}_q(X) - \beta(q)\}, \quad X \in \mathcal{X}.$$

where $\mathcal{H}_0$ is a set of functions mapping $[0,1]$ to $(-\infty,\infty]$. Fix $X \in \mathcal{X}$. For $m > 0$, write

$$X^{[m]} = X\mathbb{1}_{\{U_X \geqslant p\}} + (X-m)\mathbb{1}_{\{U_X < p\}}.$$

It is obvious that $(X^{[m]})_p \overset{\mathrm{d}}{=} X_p$, and hence $\rho(X^{[m]}) = \rho(X)$ for all $m > 0$. Moreover, for $q \in [0,p)$,

$$\mathrm{ES}_q(X) - \mathrm{ES}_q(X^{[m]}) = \frac{1}{1-q}\int_q^p m\,\mathrm{d}t = \frac{m(p-q)}{1-q}.$$

Therefore,

$$\rho(X) = \rho(X^{[m]}) = \inf_{\beta\in\mathcal{H}_0} \sup_{q\in[0,1]} \left\{ \mathrm{ES}_q(X) - \beta(q) - \frac{m(p-q)}{1-q} \mathbb{1}_{\{q\in[0,p)\}} \right\}. \tag{24}$$

Write

$$\varepsilon = \rho(X) - \inf_{\beta\in\mathcal{H}_0} \sup_{q\in[p,1]} \{\mathrm{ES}_q(X) - \beta(q)\}.$$

Suppose, for the purpose of contradiction, that $\varepsilon > 0$. There exists $\beta_0 \in \mathcal{H}_0$ such that

$$\sup_{q\in[p,1]} \{\mathrm{ES}_q(X) - \beta_0(q)\} < \inf_{\beta\in\mathcal{H}_0} \sup_{q\in[p,1]} \{\mathrm{ES}_q(X) - \beta(q)\} - \frac{\varepsilon}{2}$$

On the other hand, by (24), for any $m > 0$,

$$\sup_{q\in[0,1]} \left\{ \mathrm{ES}_q(X) - \beta_0(q) - \frac{m(p-q)}{1-q} \mathbb{1}_{\{q\in[0,p)\}} \right\} \geqslant \rho(X) = \varepsilon + \inf_{\beta\in\mathcal{H}_0} \sup_{q\in[p,1]} \{\mathrm{ES}_q(X) - \beta(q)\}.$$

By the choice of β_0, this leads to

$$\sup_{q\in[0,1]} \left\{ \mathrm{ES}_q(X) - \beta_0(q) - \frac{m(p-q)}{1-q} \mathbb{1}_{\{q\in[0,p)\}} \right\} - \sup_{q\in[p,1]} \{\mathrm{ES}_q(X) - \beta_0(q)\} > \frac{\varepsilon}{2}.$$

Letting $m \to \infty$, we arrive at a contradiction. Therefore, $\varepsilon = 0$, and

$$\rho(X) = \inf_{\beta\in\mathcal{H}_0} \sup_{q\in[p,1]} \{\mathrm{ES}_q(X) - \beta(q)\},$$

thus the desired result. □

References

[1] Acharya, V.V., Cooley, T. and Richardson, M. (2010). *Manufacturing Tail Risk: A Perspective on the Financial Crisis of 2007-2009.* Now Publishers Inc.

[2] 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.

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

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

[5] BCBS (2016). *Standards. Minimum Capital Requirements for Market Risk. January 2016.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements.

[6] Bellini, F. and Bignozzi, V. (2015). On elicitable risk measures. *Quantitative Finance*, **15**(5), 725–733.

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

[8] Bernard, C., Jiang, X. and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108.

[9] Bernard, C., Rüschendorf, L. and Vanduffel, S. (2017). VaR bounds with variance constraint. *Journal of Risk and Insurance*, **84**(3), 923–959 .

[10] Cai, J., Liu, H. and Wang, R. (2018). Asymptotic equivalence of risk measures under dependence uncertainty. *Mathematical Finance*, **28**(1), 29–49.

[11] Cannata, F. and Quagliariello, M. (2011). *Basel III and Beyond.* Risk Books, London.

[12] Cont, R., Deguest, R. and He, X. D. (2013). Loss-based risk measures. *Statistics and Risk Modeling*, **30**(2), 133–167.

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

[14] Dai, M., Kou, S. and Shao, H. (2018). Top incomes and income inequality indices: A unified framework based on inequality index curves. *Preprint*, SSRN: 3014513.

[15] Delbaen, F. (2012): *Monetary Utility Functions*. Osaka University Press, Osaka.

[16] Delbaen, F., Bellini, F., Bignozzi, V. and Ziegel, J. (2016). Risk measures with the CxLS property. *Finance and Stochastics*, **20**(2), 433–453.

[17] Denneberg, D. (1990). Premium calculation: why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin*, **20**, 181–190.

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

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

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

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

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

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

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

[25] Föllmer, H. and Weber, S. (2015). The axiomatic approach to risk measures for capital determination. *The Annual Review of Financial Economics*, **7**, 301–337.

[26] Frittelli, M. and Rossaza Gianin, E. (2005). Law invariant convex risk measures. *Advances in Mathematical Economics*, **7**, 33–46.

[27] Furman, E. and Landsman, Z. (2006). Tail variance premium with applications for elliptical portfolio of risks. *ASTIN Bulletin*, **36**, 433–462.

[28] 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.

[29] Gerber, H. U. (1974). On additive premium calculation principles. *ASTIN Bulletin*, **7**(3), 215–222.

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

[31] 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.

[32] Johnson, P. (2015). Capital ideas: multiple approaches to capital allocation bring multifarious insights. *The Actuary*, 06 August 2015.

[33] Jouini, E., Schachermayer, W. and Touzi, N. (2006). Law invariant risk measures have the Fatou property. *Advances in Mathematical Economics*, **9**, 49–71.

[34] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R.* Springer Science & Business Media.

[35] Kou, S., Peng, X. and Heyde, C. C. (2013). External risk measures and Basel accords. *Mathematics of Operations Research*, **38**(3), 393–417.

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

[37] Koch-Medina, P., Moreno-Bromberg, S. and Munari, C. (2015). Capital adequacy tests and limited liability of financial institutions. *Journal of Banking and Finance*, **51**, 93-102.

[38] Koch-Medina, P., Munari, C. and Šikić, M. (2017). Diversification, protection of liability holders and regulatory arbitrage, *Mathematics and Financial Economics*, **11**(1), 63–83.

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

[40] Lambert, N. S., Pennock, D. M. and Shoham, Y. (2008). Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, 129-138. ACM.

[41] Mao, T. and Wang, R. (2020). Risk aversion in regulatory capital principles. *SIAM Journal on Financial Mathematics*, **11**(1), 169–200.

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

[43] Natarajan, K., Pachamanova, D. and Sim, M. (2008). Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, **54**(3), 573–585.

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

[45] Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *Annals of Applied Statistics*, **11**(4), 1833–1874.

[46] Osband, K. H. (1985). Providing incentives for better cost forecasting. *Ph.D. thesis*, University of California, Berkeley.

[47] Quiggin, J. (1993). *Generalized Expected Utility Theory: The Rank-dependent Model.* Kluwer, the Netherlands.

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

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

[50] Sandström, A. (2010). *Handbook of Solvency for Actuaries and Risk Managers: Theory and Practice*. Taylor & Francis US, Boca Raton, Florida.

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

[52] Staum, J. (2013). Excess invariance and shortfall risk measures. *Operations Research Letters*. **41**(1), 47–53.

[53] Svindland, G. (2008). Convex risk measures beyond bounded risk. *Ph.D. Thesis,* Ludwig Maximilian Universität München.

[54] Tsanakas, A. (2009). To split or not to split: Capital allocation with convex risk measures. *Insurance: Mathematics and Economics*, **44**(2), 268–277.

[55] Wang, R., Bignozzi, V. and Tsakanas, A. (2015). How superadditive can a risk measure be? *SIAM Jounral on Financial Mathematics*, **6**, 776–803.

[56] 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.

[57] Wang, R. and Ziegel, J. (2015). Elicitable distortion risk measures: A concise proof. *Statistics and Probability Letters*, **100**, 172–175.

[58] Weber, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance*, **16**, 419–441.

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

[60] Ziegel, J. (2016). Coherence and elicitability. *Mathematical Finance*, **26**, 901–918.

[61] Zhu, S. and Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, **57**(5), 1155-1168.