DISENTANGLING DISTORTION RISK MEASURES AND THE EXPECTED SHORTFALL

MASSIMILIANO AMARANTE

Département de Sciences Économiques & CIREQ,
Université de Montréal, Canada

FELIX-BENEDIKT LIEBRICH

Institute of Actuarial and Financial Mathematics & House of Insurance,
Leibniz Universität Hannover, Germany

ABSTRACT. Distortion risk measures are risk measures that are law invariant and comonotonic additive. We characterize notable sub-classes of this wide range of functionals, starting with the property of prudence recently introduced by Wang & Zitikis. Moreover, we develop a new view of *coherent* distortion risk measures and show that they are captured by a single probability charge. By linking our insights into these two properties, we obtain new characterizations of the Expected Shortfall and implications associated with its utilization as standard measure of market risk. Along the route, we obtain some ancillary results of independent interest. These concern: (i) the anticore of a general submodular distortion; (ii) a full characterization of spectral risk measures on integrable random variables; and (iii) a novel proof of the automatic Fatou property of convex, law-invariant risk measures. Finally, we fully close the remaining gap to the Wang-Zitikis axiomatization of the Expected Shortfall and carefully disentangle the interplay of the involved axioms within the large class of distortion risk measures.

1. INTRODUCTION

With the Basel Accords of 2016, the Expected Shortfall (ES) has risen to the status of standard risk measure for market risk, replacing the Value-at-Risk (VaR) in that position. Needless to say, this change has drawn increased attention both on the theoretical properties of the ES and on the implications associated with its utilization. Notably, a recent paper by Wang & Zitikis [43] has provided an axiomatization of the ES, which greatly clarifies the logic of the ES as a measure of market risk. Throughout our investigation, we will return to the Wang-Zitikis (WZ) axiomatization extensively.

In the recent years, axiomatic studies of particular classes of risk measures have enjoyed growing interest and inseminated regulatory debates. For instance, full axiomatic characterizations of VaR by means of elicitability properties have been provided by Kou & Peng [27] and Liu & Wang [31], while the axiomatic focus of He & Peng [23] lies on the property of surplus invariance. On a more structural

level, both the ES and the VaR belong to the family of *distortion risk measures*. These are statistical functionals in the sense that the risk of a loss profile X only depends on its distribution under a reference probability measure $\mathbb{P}$ (a property widely known as *law invariance* in the literature), and have the additional property of being *comonotonic additive*. That is, any pair (X, Y) of risks that are increasing transformations of one and the same source of randomness cannot offer diversification benefits when combined; the risk $\rho(X + Y)$ agrees with the sum $\rho(X) + \rho(Y)$ of individual risks. As a consequence of a classical result of Schmeidler [37], such risk measures ρ are *Choquet integrals* with respect to distortions of the reference probability $\mathbb{P}$, i.e., a capacity that distorts probabilities under $\mathbb{P}$ with a function $T: [0, 1] \longrightarrow [0, 1]$. The literature on functionals of this type and their utilization in insurance pricing (cf. Castagnoli et al. [10], Wang [44], Wang et al. [45]), risk management (cf. Acerbi [1], Acerbi & Tasche [2], Dhaene et al. [16], Föllmer & Schied [18, Sections 4.6–4.7], Wang et al. [42], and the references therein), and decision theory (cf. Carlier & Dana [9], Kadane & Wassermann [25], and Schmeidler [38]) is rich.

In spite of this commonality, it is well known (cf. Embrechts & Wang [17]) that the Expected Shortfall is *subadditive*, i.e., for all integrable risks X, Y,

$$\mathrm{ES}_\bullet(X + Y) \leq \mathrm{ES}_\bullet(X) + \mathrm{ES}_\bullet(Y),$$

a property that VaR lacks. Subadditivity of the Expected Shortfall is, for instance, due to the ES being a special *spectral risk measure* in the sense of Acerbi [1], which makes it a *coherent* risk measure; cf. Artzner et al. [6]. Moreover, the same relation to spectral risk measures shows that the ES has far-reaching continuity properties such as the *Lebesgue property*.

Interestingly, the only axiomatization of the ES to date is Wang & Zitikis [43]. That paper considers functionals ρ defined on the space L^1 of integrable random variables over a rich (i.e., atomless) probability space $(\Omega, \mathcal{F}, \mathbb{P})$. They prove that ES_p for some level $0 < p < 1$ is the only functional $\rho: L^1 \longrightarrow \mathbb{R}$ satisfying $\rho(1) = 1$,

(a) law invariance,
(b) monotonicity,

and two new axioms called

(c) *prudence*, and
(d) *no reward for concentration* (NRC).

Broadly speaking, prudence means that the risk $\rho(X)$ of the almost-sure limit X of a sequence $(X_n)_{n \in \mathbb{N}}$ does not exceed the approximating risks. No reward for concentration identifies a shock event such that risks X and Y producing their largest losses in that event — and must therefore be seen as concentrated — do not offer diversification benefits when combined.

A striking observation is that Wang & Zitikis make no use of subadditivity, comonotonic additivity, or the Lebesgue property of the ES. Evidently, these key building blocks of risk measures must be hidden in the combination of properties (a)–(d) above.

The goal of the present paper is therefore to disentangle the WZ axioms within the huge class of distortion risk measures, thereby unveiling how exactly they imply subadditivity or comonotonic additivity, for instance, and develop a better understanding of the new axioms prudence and (NRC).

Structure and contributions of the paper. In Section 3 we study arbitrary distortion risk measures $\rho: L^\infty \longrightarrow \mathbb{R}$ that are additionally prudent. In Theorem 3.2 we show that prudent distortion

risk measures must treat near full-probability events as full probability events. More specifically, we give a complete characterization of them in terms of

(i) the geometry of the associated distortion function T,
(ii) the fact that ρ must dominate a Value-at-Risk benchmark, and
(iii) the statistical property of lower semicontinuity with respect to convergence in distribution, i.e., computing estimates for the risk of a given distribution approximated with a consistent estimator does not underestimate the true risk.

Remarkably, we do not need additional properties like subadditivity here. Instead, we establish the link to the tail risk measures of Liu & Wang [31] and introduce a simple behavioural *index of nontriviality* to classify prudent risk measures.

In Section 4, we move one step closer to the Expected Shortfall by looking at coherent distortion risk measures (which we refer to as coherent Choquet distortions). They have enjoyed immense interest and been identified as a primitive building block for all other law-invariant coherent risk measures; cf. Kusuoka [28]. Our main result, Theorem 4.3, casts this result and other representation results in a new light, showing that every coherent Choquet distortion obtains in the following two-step procedure:

(i) Take a Bayesian view by selecting a finitely additive probability charge $\mathfrak{q}$ to evaluate expected losses $\int X \, d\mathfrak{q}$. In the terminology of Amarante & Ghossoub [4], this would correspond to a "Bayesian expert".
(ii) Secondly, *enforce* law invariance under the reference measure $\mathbb{P}$ by computing the worst-case expected loss $\int X' \, d\mathfrak{q}$ among all X' sharing the same distribution as X.

This procedure starts with the selection of the *backbone* $\mathfrak{q}$, and our further analysis makes heavy use of its existence.

In Section 5 we then take a detour by exhibiting immediate implications of the aforementioned characterization. Firstly, we generalize a result of independent interest on the anticore of submodular distortions due to Carlier & Dana [9]. Secondly, a new and compact proof of the fundamental "automatic Fatou property" of law-invariant convex risk measures is provided. Thirdly, we show that spectral risk measures on the much larger space L^1 are canonical extensions of a special class of coherent Choquet distortions, namely those for which the backbone is a countably additive probability measure.

In Section 6, we take the final step towards the ES and combine our considerations of prudence and coherence. We prove that the minimal element in the class of prudent and coherent Choquet distortions with index of nontriviality beyond a certain threshold is the Expected Shortfall. The assumption of coherence can be dropped: The same result holds in the much larger class of distortion risk measures with respect to *exact distortions*. The value of this result is to highlight that the ES, in comparison to a huge set of alternatives, is the *least conservative* prudent distortion risk measure, putting the least taxing capital requirements on financial institutions. *Inter alia*, the proof leads to yet another characterization of prudence. A coherent Choquet distortion is prudent if and only if the backbone $\mathfrak{q}$ and the (statistical) reference measure $\mathbb{P}$ disagree on the possibility of events; a $\mathbb{P}$-nontrivial null set of $\mathfrak{q}$ must exist. When computing the worst-case conceivable expected loss of a "rearrangement" of X under $\mathbb{P}$, this null set must swallow large and therefore speculative gains. As an application, we obtain a dual condition for statistical well-behavedness of a general convex law-invariant functional in Corollary 6.6.

Section 7 concludes and closes the gap to the WZ axioms. Surprisingly, the no reward for concentration axiom turns out to play a dual role to prudence. We show that a law-invariant, monotone functional satisfying NRC is necessarily a distortion risk measure. The ES class then turns out to require *maximal* capital buffers among all distortion risk measures satisfying NRC. Even though this is not clear at the outset, prudence turns out to be a much more conservative property than NRC for distortion risk measures. Our novel mathematical approach to the WZ characterization also demonstrates clearly that both prudence and NRC play (at least) a double role in pinning down the Expected Shortfall. Another contribution of independent, but rather technical interest are extension properties of prudent distortion risk measures outlined in Appendix B. Extension results for risk measures are a classical topic of interest; cf., for instance, Filipović & Svindland [20] and Liebrich & Svindland [30]. We show that each prudent distortion risk measure extends uniquely to a proper functional on the space L^0 of all real-valued random variables retaining law invariance, monotonicity, prudence, and lower semicontinuity with respect to convergence in distribution. Such an extension exists without subadditivity, a striking feature when compared to the literature. Crucially, we produce no gap to Wang & Zitikis by focusing our analysis mostly on L^∞ instead of L^1: A comonotonic additive, prudent, and monotone functional on a larger space is uniquely determined by its values on bounded random variables; cf. Corollary B.3.

2. Preliminaries

Basic notation. Throughout the paper, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space. As usual, spaces of equivalence classes up to $\mathbb{P}$-almost sure ($\mathbb{P}$-a.s.) equality of real-valued random variables on Ω are denoted by L^p, $p \in [0, \infty]$. We will be mostly interested in the spaces L^0, L^1, and L^∞ of $\mathcal{F}$-measurable, resp. integrable, resp. $\mathbb{P}$-essentially bounded real-valued function on Ω.

Given $X \in L^0$, its *cumulative distribution function* and its (left-continuous) *quantile function* are denoted by $F_X \colon \mathbb{R} \longrightarrow [0, 1]$ and $F_X^{-1} \colon (0, 1) \longrightarrow \mathbb{R}$, respectively, and defined by

$$F_X(x) := \mathbb{P}(X \le x) \quad \text{and} \quad F_X^{-1}(s) := \inf\{x \in \mathbb{R} \mid F_X(x) \ge s\}.$$

Given $X, Y \in L^0$, we write $X \overset{d}{=} Y$ if $F_X = F_Y$, i.e., the two random variables agree in distribution under $\mathbb{P}$.

Charges. Given a probability measure $\mathbb{P}$ on the measurable space $(\Omega, \mathcal{F})$, $\mathbf{ba}(\mathbb{P}) = \mathbf{ba}$ denotes the space of all charges that are *absolutely continuous* with respect to $\mathbb{P}$, and $\mathbf{ca}(\mathbb{P}) = \mathbf{ca}$ the band of all signed measures in $\mathbf{ba}$. Absolute continuity of a positive charge $\xi \in \mathbf{ba}_+$ means that every $\mathbb{P}$-null set is ξ-null; then note that $\mathbf{ba} = \mathbf{ba}_+ - \mathbf{ba}_+$. A positive charge $\mu \in \mathbf{ba}_+$ is a *pure charge* (also denoted by $\mu \in \mathbf{ca}_+^d$) if necessarily $\nu = 0$ whenever $\nu \in \mathbf{ca}_+$ satisfies $\nu(A) \le \mu(A)$ for all $A \in \mathcal{F}$. An alternative characterisation is given by [8, Theorem 10.3.3]: A positive charge $\xi \in \mathbf{ba}_+$ is a pure charge if and only if there is a vanishing sequence of events $(B_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ such that, for all $n \in \mathbb{N}$, $\xi(B_n^c) = 0$. Each $\xi \in \mathbf{ba}_+$ decomposes uniquely as the sum of a finite measure $\zeta \in \mathbf{ca}_+$ and a pure charge $\tau \in \mathbf{ca}_+^d$. For $\xi \in \mathbf{ba}_+$, we write

$$\frac{d\xi}{d\mathbb{P}} := \frac{d\zeta}{d\mathbb{P}}.$$

At last, note that by $\mathfrak{q}$ we will denote probability charges, positive charges normalized to $\mathfrak{q}(\Omega) = 1$, and their totality makes up the set $\Delta \subset \mathbf{ba}$.

Risk measures. Throughout the paper, we follow the actuarial convention that risk measures are applied to random variables modelling *net losses*. Positive random variables correspond to pure losses,

negative ones to pure gains. A function f defined on a domain $\mathcal{D} \subset L^0$ is *law-invariant* if the value $f(X)$ only depends on the distribution of X under $\mathbb{P}$: For all $X, Y \in \mathcal{D}$,

$$X \stackrel{d}{=} Y \quad \implies \quad f(X) = f(Y).$$

Let $\mathcal{X} \subset L^0$ be a subspace containing all constant random variables (which we identify with $\mathbb{R}$). A functional $\rho \colon \mathcal{X} \longrightarrow (-\infty, \infty]$ is a *risk measure* if it has the following properties:

(a) *properness*, i.e. $\rho(X) < \infty$ for some $X \in \mathcal{X}$.
(b) *monotonicity*, i.e. $X \le Y$ $\mathbb{P}$-a.s. implies $\rho(X) \le \rho(Y)$.
(c) *cash-additivity*, i.e. for all $X \in \mathcal{X}$ and all $m \in \mathbb{R}$, $\rho(X + m) = \rho(X) + m$.

A risk measure is *coherent* if, additionally, it has the properties:

(d) positive homogeneity, i.e. $\rho(tX) = t\rho(X)$ for all $t \ge 0$.[1]
(e) subadditivity, i.e. $\rho(X + Y) \le \rho(X) + \rho(Y)$ for all $X, Y \in \mathcal{X}$.

Let $(X_n)_{n \in \mathbb{N}} \subset L^\infty$ be a sequence of random variables converging $\mathbb{P}$-a.s. to $X \in L^\infty$ and satisfying $\sup_{n \in \mathbb{N}} \|X_n\|_\infty < \infty$. A risk measure $\rho \colon L^\infty \longrightarrow \mathbb{R}$ has the *Fatou property* if, in the preceding situation, $\rho(X) \le \liminf_{n \to \infty} \rho(X_n)$, and the *Lebesgue property* if both ρ and $-\rho$ have the Fatou property. Under the actuarial convention above, the Lebesgue (Fatou) property that is usually stated (see [18]) as "continuity from below" ("above") now corresponds to "continuity from above" ("below"). Three risk measures that appear recurrently in the remainder of the paper are:

(i) The *Value-at-Risk* $\mathrm{VaR}_p \colon L^0 \longrightarrow (-\infty, \infty]$ at level $p \in (0, 1]$ defined by

$$\mathrm{VaR}_p(X) := \begin{cases} F_X^{-1}(p) & p < 1 \\ F_X^{-1}(1-) := \lim_{s \uparrow 1} F_X^{-1}(s) & p = 1; \end{cases}$$

(ii) The *Expected Shortfall* $\mathrm{ES}_p \colon L^1 \longrightarrow (-\infty, \infty]$ at level $p \in [0, 1]$ defined by

$$\mathrm{ES}_p(X) = \begin{cases} \frac{1}{1-p} \int_p^1 F_X^{-1}(s) ds & p < 1 \\ F_X^{-1}(1-) & p = 1; \end{cases} \tag{2.1}$$

(iii) The *spectral risk measures* of Acerbi [1]: Given a nonnegative and nondecreasing function ϕ on $[0, 1]$ such that $\int_0^1 \phi(t) dt = 1$, the associated spectral risk measure defined on L^1 is given by

$$\rho(X) = \int_0^1 \phi(t) F_X^{-1}(t) dt = \int_0^1 \phi(t) F_X^{-1}(t)^+ dt + \int_0^1 \phi(t)(-F_X^{-1}(t)^-) dt. \tag{2.2}$$

This gives a well-defined functional $\rho \colon L^1 \longrightarrow (-\infty, \infty]$ that is additionally lower semicontinuous with respect to the L^1-norm, i.e., every norm-convergent sequence $(X_n)_{n \in \mathbb{N}} \subset L^1$ with limit X satisfies $\rho(X) \le \liminf_{n \to \infty} \rho(X_n)$; cf. Lemma A.1. The function ϕ is called a *spectrum*.

By using $\phi := \frac{1}{1-p} \mathbf{1}_{(p,1)}$, one easily sees that ES_p is spectral for all $p \in (0, 1)$.

Capacities and the Choquet integral. Capacities and the Choquet integral are key tools in the present work. Given a measurable space $(\Omega, \mathcal{F})$, a (normalized) capacity is a set function $v \colon \mathcal{F} \longrightarrow [0, 1]$ that is nondecreasing with respect to set inclusion and satisfies $v(\varnothing) = 1 - v(\Omega) = 0$. Given a probability $\mathbb{P}$ on $\mathcal{F}$, v is $\mathbb{P}$-invariant if $v(A) = v(B)$ whenever $\mathbb{P}(A) = \mathbb{P}(B)$. In such a case, we say that v is a *distortion* of $\mathbb{P}$ and there is a nondecreasing function $T \colon [0, 1] \longrightarrow [0, 1]$ such that $v = T \circ \mathbb{P}$. If $\mathbb{P}$ is atomless, T is unique. All appearing capacities in this paper will be distortions, we therefore

---

[1] Here, and in the following, we employ the convention $0 \cdot \infty = 0$.

restrict our attention to this type. On an atomless probability space, a distortion is *continuous* if the associated distortion function is continuous.

The *Choquet integral* of $X \in L^\infty$ with respect to a distortion $v = T \circ \mathbb{P}$ is defined as

$$\int X \, dv = \int X \, d(T \circ \mathbb{P}) := \int_0^\infty T(\mathbb{P}(X > x)) dx + \int_{-\infty}^0 (1 - T(\mathbb{P}(X > x))) \, dx.$$

The integrals on the right-hand side are Riemann integrals. Notice that the Choquet integral is a positively homogeneous risk measure on L^∞, and we shall therefore call call it *distortion risk measure* (DRM). If instead of the capacity the DRM ρ is given, T_ρ denotes an associated distortion function. Its characterizing property of the Choquet integral is its comonotonic additivity: Whenever two bounded random variables X, Y are *comonotonic*, i.e., nondecreasing transformations $X = f(Z)$ and $Y = g(Z)$ of one and the same random variable Z, then

$$\int (X + Y) dv = \int X \, dv + \int Y \, dv.$$

A capacity v on $\mathcal{F}$ is *submodular* if, for all events $A, B \in \mathcal{F}$,

$$v(A \cup B) + v(A \cap B) \leq v(A) + v(B).$$

As is well known, the submodularity of a distortion $v = T \circ \mathbb{P}$ of an atomless $\mathbb{P}$ is equivalent to the convexity of the associated Choquet integral and concavity of T ([18, Proposition 4.75]).

The *anticore* of a distortion v, acore(v), is the possibly empty set

$$\{\mathfrak{q} \in \mathbf{ba}_+ \mid \mathfrak{q}(A) \leq v(A) \text{ for all } A \in \mathcal{F} \text{ and } \mathfrak{q}(\Omega) = 1\}.$$

If v is submodular, its anticore is necessarily nonempty [36]. Last, a distortion is called *exact* if, for all $A \in \mathcal{F}$, $v(A) = \max_{\mathfrak{q} \in \text{acore}(v)} \mathfrak{q}(A)$. Due to results of Kadane & Wassermann [25] and later Aouani & Chateauneuf [5], a distortion is exact if and only if the hypograph of the distortion function T, $\mathbb{H}(T) := \{(x, y) \in [0, 1]^2 \mid y \leq T(x)\}$, is star shaped around $(0, 0)$ and $(1, 1)$.

3. PRUDENT RISK MEASURES

In this first section, we present a full characterisation of DRMs which satisfy the recently introduced *prudence* axiom below. It goes back to Wang & Zitikis [43].

Prudence (cf. [43]): A risk measure ρ is prudent if, whenever $(X_n)_{n \in \mathbb{N}}$ converges $\mathbb{P}$-a.s. to X and $\lim_{n \to \infty} \rho(X_n)$ exists, then

$$\rho(X) \leq \lim_{n \to \infty} \rho(X_n).^2$$

As we shall see momentarily, prudent DRMs a particularly notable structure on the level of distortion functions. This further justifies interest for this class. What is more, the study of these properties later on casts some new light on the Expected Shortfall. We begin by highlighting a useful implication of prudence in combination with law-invariance. As in most of the literature we assume:

Assumption 3.1. The underlying probability measure $\mathbb{P}$ is nonatomic. In particular, the set $\mathcal{U}$ of random variables with a uniform distribution over $(0, 1)$ is nonempty.

---

[2] Strictly speaking, [43] uses pointwise convergent sequences of random variables in their definition of prudence. This is due to the fact that they do not identify random variables to equivalence classes as we do here. However, their interest lies solely in functionals with the property that two random variable agreeing $\mathbb{P}$-a.s. are mapped to the same value. Clearly, for such functionals, their notion of prudence is equivalent to the definition here.

The following result on DRMs serves as the cornerstone of our further investigation of prudence. Note that we do not require additional properties like subadditivity of the risk measure in question.

Theorem 3.2. *For a DRM $\rho\colon L^\infty \longrightarrow \mathbb{R}$, the following are equivalent:*

(1) *ρ is prudent.*
(2) *T_ρ is left-continuous and there is $0 < p < 1$ such that $T_\rho|_{[p,1]}$ is constant.*
(3) *ρ has the Fatou property and there is $0 < p < 1$ such that $\rho \geq \mathrm{VaR}_p$.*
(4) *ρ is lower semicontinuous with respect to convergence in distribution.*

Proof. (1) implies (2): Let $(p_n)_{n\in\mathbb{N}} \subset (0,1)$ such that $p_n \uparrow p$. Select an increasing sequence $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ such that $\mathbb{P}(A_n) = p_n$, $n \in \mathbb{N}$. For $A := \bigcup_{n=1}^\infty A_n$, $\mathbf{1}_A = \lim_{n\to\infty} \mathbf{1}_{A_n}$ a.s. Hence,

$$T_\rho(p) = \rho(\mathbf{1}_A) \leq \lim_{n\to\infty} \rho(\mathbf{1}_{A_n}) = \lim_{n\to\infty} T_\rho(p_n).$$

This is left-continuity. Now, by the way of contradiction, assume that $k_n := T_\rho(1) - T_\rho(\frac{n-1}{n}) > 0$ for all $n \in \mathbb{N}$. Let $(B_n)_{n\in\mathbb{N}}$ be a decreasing sequence of events such that $\mathbb{P}(B_n) = \frac{1}{n}$. Then $\rho(-k_n^{-1}\mathbf{1}_{B_n}) = \rho(k_n^{-1}\mathbf{1}_{B_n^c}) - k_n^{-1} = -k_n^{-1}\left(1 - T_\rho(\frac{n-1}{n})\right) = -1$ holds for all $n \in \mathbb{N}$. Together with $\lim_{n\to\infty} k_n^{-1}\mathbf{1}_{B_n} = 0$ a.s., this yields a contradiction to prudence.

(2) implies (3): As ρ is monotone, the Fatou property is equivalent to continuity from below, which in turn follows easily with the left-continuity of T_ρ and monotone convergence. Moreover, for $p \in (0,1)$ such that $T|_{[p,1]} \equiv 1$, we also have

$$T \geq \mathbf{1}_{(p,1]} = T_{\mathrm{VaR}_{1-p}}.$$

As such a relation between distortion functions transfers to the associated DRMs, we can infer $\mathrm{VaR}_{1-p} \leq \rho$.

(3) implies (4): Under assertion (3), [31, Theorem F.1] shows that ρ is a so-called *tail-relevant DRM* in the terminology of the aforementioned paper. In particular, there exists $q \in (0,1)$ such that, for all $X \in L^\infty$,

$$\rho(X) = \rho(X \vee \mathrm{VaR}_q(X)). \tag{3.1}$$

Let $(X_n)_{n\in\mathbb{N}} \subset L^\infty$ and assume that the sequence converges in distribution to X. Let $U \in \mathcal{U}$ be arbitrary and note that $X_n' := F_{X_n}^{-1}(U) \stackrel{d}{=} X_n$ converges to $X' = F_X^{-1}(U) \stackrel{d}{=} X$ a.s. Hence, also $Y_n := X_n' \wedge X'$ satisfies $\lim_{n\to\infty} Y_n = X'$. By Skorokhod's Representation Theorem, $F_{Y_n}^{-1} \to F_X^{-1}$ Lebesge-a.e., which allows us to select $0 < r < q$ such that $\lim_{n\to\infty} \mathrm{VaR}_r(Y_n) = \mathrm{VaR}_r(X)$. Let $z := \inf_{n\in\mathbb{N}} \mathrm{VaR}_r(Y_n)$ and set $Y_n' := Y_n \vee z$, $n \in \mathbb{N}$, and $Y' := X' \vee z$. Moreover, observe that $\mathrm{VaR}_q(X') = \mathrm{VaR}_q(Y')$ and that $\mathrm{VaR}_q(Y_n) = \mathrm{VaR}_q(Y_n')$, $n \in \mathbb{N}$. Using (3.1) and the Fatou property for the first estimate, we observe:

$$\begin{aligned}
\rho(X) = \rho(X') &= \rho(X' \vee \mathrm{VaR}_q(X')) \\
&= \rho(Y' \vee \mathrm{VaR}_q(Y')) = \rho(Y') \\
&\leq \liminf_{n\to\infty} \rho(Y_n') = \liminf_{n\to\infty} \rho(Y_n) \\
&\leq \liminf_{n\to\infty} \rho(X_n') = \liminf_{n\to\infty} \rho(X_n).
\end{aligned}$$

(4) implies (1): This implication is due to a.s. convergence implying convergence in distribution. $\qquad\square$

Remark 3.3.

(1) Point (2) in Theorem 3.2 demonstrates that prudence is geometrically characterized by a particular shape of the distortion function T_ρ, which in turn treats near full-probability events as full probability events.

(2) Theorem 3.2(3) makes precise the intuition that prudence is a "propped-up Fatou property" of the risk measure in question.

(3) Item (4) in Theorem 3.2 suggests that prudent DRMs are statistically well behaved: The risk does not exceed approximating risks even under a weak notion of approximation like convergence in distribution. It is an open question to us whether the equivalence between items (1), prudence, and (4) lower semicontinuity with respect to convergence in distribution, holds for general law-invariant functionals. Sufficient conditions beyond DRMs are stated in Proposition B.5.

(4) As mentioned in the context of (3.1) in the proof of Theorem 3.2, prudent DRMs are necessarily *tail relevant* in the terminology of [31]. That paper presents a thorough study of tail risk measures that are characterized by the limited possibilities to cross-subsidize losses with large, but speculative gains, i.e., gains that are not sufficiently likely under $\mathbb{P}$. To a certain degree it is surprising that there is such an intimate link between this property with a clear regulatory interpretation and statistical well-behavedness established in our result. Note also the reminiscence between (3.1) and the property of *surplus invariance*; cf. [23, 26].

Theorem 3.2 turns out to be very useful. For one thing, it immediately leads to the following result, which was proved in [43] in a rather lengthy way:

Corollary 3.4.

(1) VaR_p *is prudent for all* $0 < p \le 1$.

(2) ES_p *is prudent for all* $0 < p \le 1$.

Proof. (1) VaR_p is the DRM with respect to the distortion function $T_{\text{VaR}_p} = \mathbf{1}_{(1-p,1)}$, $0 < p < 1$. The latter is left-continuous and constant in a neighbourhood of 1.

(2) $\text{ES}_1 = \text{VaR}_1$ is prudent by (1). For $0 < p < 1$, ES_p is the DRM with respect to the continuous distortion function $T_{\text{ES}_p} \colon [0,1] \longrightarrow [0,1]$ defined by $T_{\text{ES}_p}(x) = \frac{x}{1-p} \wedge 1$. The latter satisfies $T_{\text{ES}_p}|_{[1-p,1]} \equiv 1$.

$\square$

In Remark 3.3, we observed that measuring risk with a prudent DRM limits the possibilities to cross-subsidize losses with large, but speculative gains. In the following, we use Theorem 3.2 to put this intuition into a more behavioural context. To this end, we define the *a priori* unrelated index of nontriviality defined as follows.

Definition 3.5. Let ρ be a law-invariant risk measure. The *index of nontriviality* of ρ is

$$\mathfrak{nt}(\rho) := \inf\{\mathbb{P}(A) \mid A \in \mathcal{F}, \ \rho(-\mathbf{1}_A) < 0\}.$$

In order to get some intuition for this definition, recall that we identify the arguments of ρ as net losses. Thus, the financial position $X = -\mathbf{1}_A$ provides a gain of 1 on A and neutrality on A^c. When the event A is such that $0 < \mathbb{P}(A) < 1$, both A and A^c obtain with positive probability, and nontriviality means that the agent strictly prefers $X = -\mathbf{1}_A$ to not receiving any payoff at all. Notice that if ρ is law invariant and $A \in \mathcal{F}$ is nontrivial event (i.e., $0 < \mathbb{P}(A) < 1$) such that $\rho(-\mathbf{1}_A) < 0$, then $\rho(-\mathbf{1}_B) = \rho(-\mathbf{1}_A) < 0$ for all $B \in \mathcal{F}$ such that $\mathbb{P}(B) = \mathbb{P}(A)$. This intuition can be generalized to more general net losses. Moreover, the index of nontriviality can characterize prudent DRMs.

Proposition 3.6. *Let ρ be any positively homogeneous law-invariant risk measure on L^∞.*

(1) *We have*
$$\mathfrak{nt}(\rho) = \inf\{\mathbb{P}(X > 0) \mid X \in L_+^\infty,\ \rho(-X) < 0\}.$$

(2) *If ρ is additionally a DRM, then*
$$\mathfrak{nt}(\rho) = 1 - \inf\{p \in [0,1] \mid T_\rho(p) = 1\}$$
$$= \sup\{p \in (0,1) \mid \rho \geq \mathrm{VaR}_p\}.$$

(3) *If ρ is a DRM with Fatou property, then ρ is prudent if and only if $\mathfrak{nt}(\rho) > 0$.*

Proof.

(1) Suppose $X \in L_+^\infty$ satisfies $\rho(-X) < 0$. Then $X \neq 0$ and monotonicity of ρ together imply $\rho(-\|X\|_\infty \mathbf{1}_{\{X<0\}}) < 0$. By positive homogeneity of ρ,
$$\rho(-\mathbf{1}_{\{X<0\}}) = \frac{\rho(-\|X\|_\infty \mathbf{1}_{\{X<0\}})}{\|X\|_\infty} < 0.$$

Hence, $\mathfrak{nt}(\rho) \leq \inf\{\mathbb{P}(X > 0) \mid X \in L_+^\infty,\ \rho(-X) < 0\}$. The converse inequality obviously holds.

(2) We have $T_\rho(p) = 1$ if and only if, for all $A \in \mathcal{F}$ with $\mathbb{P}(A) = 1 - p$,
$$\rho(-\mathbf{1}_A) = \rho(\mathbf{1}_{A^c}) - 1 = T_\rho(p) - 1 = 0.$$

This is sufficient to prove the first identity. For the second, $T_\rho(p) = 1$ implies that $T_\rho \geq \mathbf{1}_{(p,1)} = T_{\mathrm{VaR}_{1-p}}$. Hence,

(3) T_ρ is left-continuous by virtue of the Fatou property. By (2) and Theorem 3.2(3), ρ is prudent if and only if $\mathfrak{nt}(\rho) > 0$.

$\square$

An immediate corollary is the observation that the Value-at-Risk family is minimal in the family of prudent DRMs.

Corollary 3.7. *Let ρ be a prudent DRM on L^∞. Then $\rho \geq \mathrm{VaR}_{\mathfrak{nt}(\rho)}$. In particular, VaR_p is the minimal prudent DRM ρ with $\mathfrak{nt}(\rho) \geq p$.*

Remark 3.8. As mentioned, one of the present goals is to develop a deeper understanding of the ES axiomatization in [43]. While we consider prudent DRMs defined on L^∞ in this section, prudence in [43] is introduced for maps on the larger model space L^1. As thoroughly discussed in Appendix B, this difference is immaterial for comonotonic additive functionals. Prudent DRMs extend uniquely to L^0 without losing their desirable properties, and comonotonic additive risk functionals on lattices of unbounded random variables are fully captured by their restriction to L^∞.

Our discussion of prudence so far has made no use of additional properties of DRMs like subadditivity. It may therefore not surprise that the Expected Shortfall has not yet made an appearance. In the next section, we will therefore focus on the most immediate property that distinguishes the Expected Shortfall from many other distortion risk measures, most prominently the Value-at Risk: coherence.

4. COHERENT CHOQUET DISTORTIONS

Coherent and/or law-invariant risk measures have been an object of interest for more than two decades. The rise of the ES has drawn attention to those risk measures that satisfy, in addition, the

property of comonotonic additivity. We will refer to risk measures that are law invariant, coherent, and comonotonic additive as *coherent Choquet distortions* (CCD). Summing up more formally, a CCD is any risk measure $\rho\colon \mathcal{X} \subset L^0 \longrightarrow (-\infty, \infty]$ satisfying the following three axioms:

Coherence: ρ has properties (a) to (e) in Section 2.
Law invariance: ρ is $\mathbb{P}$-law invariant.
Comonotonic additivity: If X and Y are comonotonic, then $\rho(X + Y) = \rho(X) + \rho(Y)$

Remark 4.1. Within the realm of properties that distinguish Value-at-Risk and Expected Shortfall, one may also consider *consistency with second-order stochastic dominance*; cf. [32]. Given random variables $X, Y \in L^1$, Y dominates X in second-order stochastic dominance relation ($X \preceq_{ssd} Y$) if, for all nondecreasing and convex test functions $v\colon \mathbb{R} \longrightarrow \mathbb{R}$, $\mathbb{E}[v(X)] \leq \mathbb{E}[v(Y)]$. A DRM is consistent if $X \preceq_{ssd} Y$ implies $\rho(X) \leq \rho(Y)$. It is well known that the Expected Shortfall is consistent while the Value-at-Risk is not. However, this feature does not require separate analysis because a DRM is consistent if and only if it is coherent; see, for instance, [42, Theorem 3].

Before we state Theorem 4.3, the main result of this section, we introduce a piece of notation.

Notation 4.2. Given a probability charge $\mathfrak{q} \in \Delta$, we define the functional $\psi_{\mathfrak{q}}\colon L^\infty \longrightarrow \mathbb{R}$ by

$$\psi_{\mathfrak{q}}(X) = \sup_{X' \overset{d}{=} X} \int X' \, d\mathfrak{q}. \tag{4.1}$$

Theorem 4.3. *For a risk measure $\rho\colon L^\infty \longrightarrow \mathbb{R}$, the following are equivalent:*

(1) *ρ is a CCD.*
(2) *$\rho = \psi_{\mathfrak{q}}$ for some probability charge $\mathfrak{q} \in \Delta$.*

Moreover, two probability charges $\mathfrak{q}, \mathfrak{r}$ satisfy $\psi_{\mathfrak{q}} = \psi_{\mathfrak{r}}$ if and only if $\frac{d\mathfrak{q}}{d\mathbb{P}} \overset{d}{=} \frac{d\mathfrak{r}}{d\mathbb{P}}$.

Proof. (1) implies (2): Consider a CCD $\rho\colon L^\infty \longrightarrow \mathbb{R}$. Combining Theorems 4.93 and 4.70 in [18], ρ is the Choquet integral with respect to a distortion $T \circ \mathbb{P}$ and can be represented as

$$\rho(X) = T(0+)F_X^{-1}(1-) + \int_0^1 F_X^{-1}(t)T'(1-t)\mathrm{d}t, \quad X \in L^\infty, \tag{4.2}$$

where

(a) $T(0+) := \inf_{x \in (0,1]} T(x)$.
(b) T' the a.e. existent derivative of T.

As T is nondecreasing and concave, T' must be nonnegative and nonincreasing. We conclude that the function $(0,1) \ni t \mapsto T'(1-t)$ is nondecreasing and continuous almost everywhere. We can therefore select a random variable Z whose quantile function satisfies $F_Z^{-1}(t) = T'(1-t)$ for almost all $t \in (0,1)$. As $T' \geq 0$ and $\int_0^1 T'(1-t)\mathrm{d}t = 1 - T(0+) < \infty$, we have $Z \in L_+^1$. Thus, Z is the density of a finite measure $\lambda \ll \mathbb{P}$ on $(\Omega, \mathcal{F})$.

Fix any purely additive charge $\xi \in \mathbf{ba}_+$ (whose existence is guaranteed by the infinite dimension of L^∞) additionally satisfying $\xi(\Omega) = T(0+)$. Set $\mathfrak{q} := \lambda + \xi$ and apply [13, Proposition 3.9] to infer that, for arbitrary $X \in L^\infty$,

$$\sup_{X' \overset{d}{=} X} \int X' \, d\mathfrak{q} = T(0+)F_X^{-1}(1-) + \int_0^1 F_X^{-1}(t)F_Z^{-1}(t)\mathrm{d}t$$

$$= T(0+)F_X^{-1}(1-) + \int_0^1 F_X^{-1}(t)T'(1-t)\mathrm{d}t = \rho(X). \tag{4.3}$$

(2) implies (1): Let $\mathfrak{q} \in \Delta$ be a probability charge and define a map ρ on L^∞ as in (4.1). By the Yosida-Hewitt Theorem, $\mathfrak{q}$ decomposes uniquely as the sum of a measure λ and a pure charge ξ. Setting $Z := \frac{\mathrm{d}\mathfrak{q}}{\mathrm{d}\mathbb{P}}$, we obtain from [13, Proposition 3.9] for all $X \in L^\infty$ that

$$\rho(X) = \int_0^1 F_Z^{-1}(t) F_X^{-1}(t) \mathrm{d}t + F_X^{-1}(1-)\mathfrak{q}(\Omega).$$

Comonotonic additivity of the right-hand expression readily follows with [18, Lemma 4.90]. Monotonicity and continuity is due to the quantile function possessing these properties. It remains to prove subadditivity. To this end, let $X, Y \in L^\infty$ be simple functions and abbreviate $Z := X + Y$. Let $Z' \stackrel{d}{=} Z$, $\mathcal{I} := \{z \in \mathbb{R} \mid \mathbb{P}(Z = z)\}$, and $\mathcal{J} := \{(x, y) \in \mathbb{R}^2 \mid \mathbb{P}(X = x, Y = y) > 0\}$. For each $z \in \mathcal{I}$ we can use nonatomicity of $\mathbb{P}$ to partition $\{Z' = z\}$ into events $\{A_{x,y} \mid (x, y) \in \mathcal{J}, \, x + y = z\}$ such that $\mathbb{P}(X = x, Y = y) = \mathbb{P}(A_{x,y})$. Then set

$$X' = \sum_{(x,y)\in\mathcal{J}} x\mathbf{1}_{A_{x,y}} \qquad \text{and} \qquad Y' = \sum_{(x,y)\in\mathcal{J}} y\mathbf{1}_{A_{x,y}},$$

and note that $X' \stackrel{d}{=} X$, $Y' \stackrel{d}{=} Y$, and $X' + Y' = Z'$. In total, we obtain that

$$\int Z' \, \mathrm{d}\mathfrak{q} = \int X' \, \mathrm{d}\mathfrak{q} + \int Y' \, \mathrm{d}\mathfrak{q} \leq \rho(X) + \rho(Y).$$

Taking the supremum over all $Z' \stackrel{d}{=} Z$ yields subadditivity on simple functions. The general case follows by approximation using the continuity of ρ.

It remains to prove the uniqueness statement. Let $\mathfrak{q}$ and $\mathfrak{r}$ be two probability charges such that $\psi_\mathfrak{q} = \psi_\mathfrak{r}$. Let $\xi, \tau \in \mathbf{ca}_+^d$ be the pure charges in the Yosida-Hewitt decomposition of $\mathfrak{q}$ and $\mathfrak{r}$ and $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ be a vanishing sequence of events with positive probability. Then [13, Proposition 3.9] yields

$$\xi(\Omega) = \lim_{n\to\infty} \psi_\mathfrak{q}(\mathbf{1}_{A_n}) = \lim_{n\to\infty} \psi_\mathfrak{r}(\mathbf{1}_{A_n}) = \tau(\Omega).$$

Next, we obtain for all $p \in (0, 1]$ and $A \in \mathcal{F}$ with $\mathbb{P}(A) = p$ that

$$\int_{1-p}^1 F_{\frac{\mathrm{d}\mathfrak{q}}{\mathrm{d}\mathbb{P}}}^{-1}(t)\mathrm{d}t = \psi_\mathfrak{q}(\mathbf{1}_A) - \xi(\Omega) = \psi_\mathfrak{r}(\mathbf{1}_A) - \tau(\Omega) = \int_{1-p}^1 F_{\frac{\mathrm{d}\mathfrak{r}}{\mathrm{d}\mathbb{P}}}^{-1}(t)\mathrm{d}t.$$

By [39, Theorem 3.A.5], this implies equivalence between $\frac{\mathrm{d}\mathfrak{q}}{\mathrm{d}\mathbb{P}}$ and $\frac{\mathrm{d}\mathfrak{r}}{\mathrm{d}\mathbb{P}}$ in the so-called convex order.[3] This is the case if and only if $\frac{\mathrm{d}\mathfrak{q}}{\mathrm{d}\mathbb{P}} \stackrel{d}{=} \frac{\mathrm{d}\mathfrak{r}}{\mathrm{d}\mathbb{P}}$ ([39, Theorem 3.A.43]).

Conversely, if $\frac{\mathrm{d}\mathfrak{q}}{\mathrm{d}\mathbb{P}} \stackrel{d}{=} \frac{\mathrm{d}\mathfrak{r}}{\mathrm{d}\mathbb{P}}$, we first have $\delta := 1 - \mathbb{E}[\frac{\mathrm{d}\mathfrak{q}}{\mathrm{d}\mathbb{P}}] = 1 - \mathbb{E}[\frac{\mathrm{d}\mathfrak{r}}{\mathrm{d}\mathbb{P}}]$. Next, [13, Proposition 3.9] yields for arbitrary $X \in L^\infty$ that

$$\psi_\mathfrak{q}(X) = \int_{1-p}^1 F_{\frac{\mathrm{d}\mathfrak{q}}{\mathrm{d}\mathbb{P}}}^{-1}(t)F_X^{-1}(t)\mathrm{d}t + \delta F_X^{-1}(0+) = \int_{1-p}^1 F_{\frac{\mathrm{d}\mathfrak{r}}{\mathrm{d}\mathbb{P}}}^{-1}(t)F_X^{-1}(t)\mathrm{d}t + \delta F_X^{-1}(0+) = \psi_\mathfrak{r}(X).$$

This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Definition 4.4. Given a CCD $\rho\colon L^\infty \longrightarrow \mathbb{R}$, we call a probability charge $\mathfrak{q}$ satisfying $\rho = \psi_\mathfrak{q}$ a *backbone* of ρ.

Remark 4.5.

---

[3] Denoting the convex order by $\preceq$, $X, Y \in L^1$ satisfy $X \preceq Y$ if and only if $\mathbb{E}[v(X)] \leq \mathbb{E}[v(Y)]$ for all convex functions $v\colon \mathbb{R} \longrightarrow \mathbb{R}$.

(1) Backbones are not unique. Whether or not a particular probability charge $\mathfrak{q} \in \Delta$ is a backbone of a given CCD ρ depends, however, only on the distribution of $\frac{d\mathfrak{q}}{d\mathbb{P}}$. The pure charge part ξ of a backbone can be chosen completely freely up to normalization.

(2) The argument in the proof of Theorem 4.3 shows a finer property. Let $\preceq$ denote the convex order on L^1 and suppose that two charges $\mathfrak{p}, \mathfrak{q} \in \Delta$ satisfy $\frac{d\mathfrak{q}}{d\mathbb{P}} \preceq \frac{d\mathfrak{r}}{d\mathbb{P}}$. Then $\mathbb{E}[\frac{d\mathfrak{q}}{d\mathbb{P}}] = \mathbb{E}[\frac{d\mathfrak{r}}{d\mathbb{P}}]$ and $\psi_{\mathfrak{q}} \leq \psi_{\mathfrak{p}}$.

As a corollary, we immediately recover a representation result for CCDs very similar to others previously obtained by Kusuoka [28] and Föllmer & Schied [18, Sections 4.6–4.7], in particular [18, Corollary 4.80].

Corollary 4.6. *Let* $\rho \colon L^\infty \longrightarrow \mathbb{R}$ *be a risk measure. Then the following are equivalent:*

(1) *ρ is a CCD.*
(2) *There is $\alpha \in [0,1]$ such that, for the concave and continuous distortion function $T_1(x) := T(x) - T(0+)$, $x \in [0,1]$, we have*

$$\rho(X) = \alpha \int_0^1 T_1'(t) F_X^{-1}(t) dt + (1-\alpha) F_X^{-1}(1-).$$

In comparison to [28], Corollary 4.6 removes altogether the assumption that the underlying space be standard and identifies all the functions appearing in the representation. We shall see that it seamlessly leads to the characterization of spectral risk measures given in Proposition 5.4 below.

5. Implications of Theorem 4.3

5.1. **The anticore of general submodular distortions.** A noteworthy byproduct of our previous analysis is Corollary 5.1 characterizing the anticore (and, by passage to the dual capacity, the core) of a general submodular (supermodular) distortion. It generalizes the result of [9] obtained in the differentiable case. As the study of the core of a capacity is central to the theory of cooperative games, the implications of Corollary 5.1 go beyond the setting of the present paper.

Corollary 5.1. *Let $v = T \circ \mathbb{P}$ be a submodular distortion with $\mathbb{P}$ nonatomic. Denote by $\mathcal{D}_v$ the set of all measures $\zeta \ll \mathbb{P}$ with*

$$\frac{d\zeta}{d\mathbb{P}} \in \overline{\text{co}\left\{T'(U) \mid U \in \mathcal{U}\right\}}^{L^1}.$$

Then

$$\text{acore}(v) = \mathcal{D}_v + T(0^+)\Delta.$$

Proof. Consider the distortion function T_1 from Corollary 4.6 and set $T_2 := T - T_1 = \mathbf{1}_{(0,1]} T(0+)$. Both functions are concave, i.e., both distortions $v_i := T_i \circ \mathbb{P}$ are submodular. As the mapping $v \mapsto \text{acore}(v)$ is affine on the cone of submodular capacities (cf. [12]), it suffices to identify $\text{acore}(v_i)$, $i = 1, 2$. One easily sees that $\text{acore}(v_2) = T(0+)\Delta$. By continuity of T_1 and [33, Proposition 4.4], $\text{acore}(v_1) \subset \mathbf{ca}_+$. By Remark 4.5(2) and the proof of Theorem 4.3, $\text{acore}(v_1)$ is identifiable with all measures $\zeta \in L_+^1$ such that $\frac{d\zeta}{d\mathbb{P}} \preceq T'(1 - U^*)$ for some $U^* \in \mathcal{U}$. Using [7, Lemma 3.5] in the first and [18, Lemma A.32] in the second identity,

$$\{Z \in L^1 \mid Z \preceq T'(1 - U^*)\} = \overline{\text{co}\{V \stackrel{d}{=} T'(1 - U^*)\}}^{L^1} = \overline{\text{co}\{T'(U) \mid U \in \mathcal{U}\}}^{L^1}.$$

$\square$

The set of extreme points in Δ is the set of pure probability charges. While evidently not Dirac measures, these behave quite like Dirac measures in that they are $\{0,1\}$-valued and correspond to the muliplicative linear functionals on L^∞. By virtue of the Krein-Milman Theorem, Corollary 5.1 establishes that the anticore of a submodular distortion is the closed convex hull of its differentiable part and the set of positive, normalized multiplicative linear functionals.

5.2. **The automatic Fatou property of convex monetary risk measures.** A fundamental result in the study of law-invariant functionals $f\colon L^\infty \longrightarrow \mathbb{R}$ is the *automatic Fatou property* of those that are monotone and convex. Originally proved in [24] for standard probability spaces, it was extended in [41] to arbitrary atomless probability spaces. In recent years, this result has been studied extensively on very general function spaces encompassing unbounded random variables, cf. [11, 21] and the references cited therein.

The strategy of the proofs of [24, 41] is to take a detour and first establish the so-called dilatation monotonicity of the functionals in question. Perhaps unexpectedly, an immediate application of Theorem 4.3 and Corollary 4.6 yields an alternative and more straightforward proof of this important result.

Lemma 5.2. *Suppose $\mathfrak{q} \in \Delta$ is a probability charge. Then there is a set $\mathcal{D}_{\mathfrak{q}} \subset L^\infty_+$ of bounded probability densities such that*

$$\psi_{\mathfrak{q}}(X) = \sup_{D \in \mathcal{D}_{\mathfrak{q}}} \mathbb{E}[DX], \quad X \in L^\infty.$$

Proof. Abbreviate $T := T_{\psi_{\mathfrak{q}}}$ and define, for $n \in \mathbb{N}$, $T_n\colon [0,1] \longrightarrow [0,1]$ by

$$T_n(x) := \begin{cases} 2^n T(2^{-n})^{-1} x & x \le 2^{-n}, \\ T(x) & x > 2^{-n}. \end{cases}$$

T_n is a continuous distortion function with bounded derivative T'_n. Moreover, T_n can be shown to be concave and, because of concavity of the the hypograph $\mathbb{H}(T)$, $T_n \le T$. By Theorem 4.3, there is a bounded density D_n such that the associated probability measure $\mathbb{Q}_n$ satisfies $\psi_{\mathbb{Q}_n} = \int \cdot \, d(T_n \circ \mathbb{P})$. Moreover, by monotone convergence and the Hardy-Littlewood inequality,

$$\psi_{\mathfrak{q}}(X) = \sup_{n \in \mathbb{N}} \psi_{\mathbb{Q}_n}(X) = \sup_{n \in \mathbb{N}} \sup_{D \overset{d}{=} D_n} \mathbb{E}[DX], \quad X \in L^\infty.$$

At last, set $\mathcal{D}_{\mathfrak{q}} := \bigcup_{n \in \mathbb{N}} \{D \overset{d}{=} D_n\}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Theorem 5.3. *Suppose a convex function $\varphi\colon L^\infty \longrightarrow \mathbb{R}$ is monotone and law invariant. Then φ is $\sigma(L^\infty, L^\infty)$-lower semicontinuous and therefore has the Fatou property.*

Proof. To begin, φ is norm-continuous (see, for instance, [35]). Consider the convex conjugate $\varphi^*\colon \mathbf{ba} \longrightarrow (-\infty, \infty]$ defined by $\varphi^*(\mu) := \sup\{\int X \, d\mu - \varphi(X) \mid X \in L^\infty\}$, and set $\mathrm{dom}(\varphi^*) := \{\varphi^* < \infty\}$. By Fenchel-Moreau,

$$\varphi(X) = \sup_{\mu \in \mathrm{dom}(\varphi^*)} \int X \, d\mu - \varphi^*(\mu), \quad X \in L^\infty.$$

By monotonicity, $\mathrm{dom}(\varphi^*) \subset \mathbf{ba}_+$. Applying Lemma 5.2 if $\mu \in \mathrm{dom}(\varphi^*) \setminus \{0\}$, we find for all $\mu \in \mathrm{dom}(\varphi^*)$ a set $\mathcal{D}_\mu$ of bounded densities such that, for all $X \in L^\infty$,

$$\sup_{X' \stackrel{d}{=} X} \int X' \, d\mu - \varphi^*(\mu) = \sup_{D \in \mathcal{D}_n} \mathbb{E}[DX] - \varphi^*(\mu).$$

Hence, by law invariance of φ, we obtain for arbitrary $X \in L^\infty$ that

$$\varphi(X) = \sup_{\mu \in \mathrm{dom}(\varphi^*)} \sup_{X' \stackrel{d}{=} X} \int X' \, \mathrm{d}\mu - \varphi^*(\mu) = \sup_{\mu \in \mathrm{dom}(\varphi^*)} \sup_{D \in \mathcal{D}_\mu} \left(\mathbb{E}[DX] - \varphi^*(\mu) \right),$$

which shows that φ is l.s.c. with respect to the $\sigma(L^\infty, L^\infty)$-topology and therefore has the Fatou property. $\qquad\square$

5.3. **Spectral risk measures.** Introduced by Acerbi [1], spectral risk measures are defined via formula (2.2) (see Section 2). By Lemma A.1, spectral risk measures are subadditive. In view of the proof of Theorem 4.3, one therefore sees that a spectral risk measure restricted to L^∞ is a CCD which additionally satisfies the following property:

> **Lebesgue property:** For any sequence $(X_n)_{n\in\mathbb{N}} \subset L^\infty$ which converges $\mathbb{P}$-a.s. to $X \in L^\infty$ and satisfies $\sup_{n\in\mathbb{N}} \|X_n\|_\infty < \infty$, $\rho(X) = \lim_{n\to\infty} \rho(X_n)$.[4]

The next proposition follows seamlessly from Theorem 4.3 and Corollary 4.6 and shows that the converse to the above statement is also true; that is, the Lebesgue property characterizes spectral risk measures among the CCDs. Proposition 5.4 is different from preceding results in that the domain of the risk measure is L^1 as required by Acerbi's definition.

Proposition 5.4. *The following are equivalent for a functional $\rho\colon L^1 \longrightarrow (-\infty, \infty]$.*

(1) *ρ is norm-l.s.c., coherent, law invariant, comonotonic additive, and has the Lebesgue property.*
(2) *ρ is a spectral risk measure.*
(3) *There is a probability measure $\mathbb{Q} \ll \mathbb{P}$ such that, for all $X \in L^1$,*

$$\rho(X) = \sup\{\mathbb{E}_\mathbb{Q}[X] \mid X' \stackrel{d}{=} X,\ \mathbb{E}_\mathbb{Q}[X'] \text{ well defined}\}. \tag{5.1}$$

Proof. (3) implies (2): If ρ is defined by (5.1), [18, Theorem A.28] yields

$$\rho(X) = \int_0^1 F^{-1}_{\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}}(t) F^{-1}_X(t) \mathrm{d}t.$$

Thus, ρ is a spectral risk measure.

(2) implies (1): A spectral risk measure is proper, l.s.c., and subadditive by Lemma A.1, law invariant by definition, comonotonic additive by [18, Lemma 4.90], monotone and cash-additive by the respective properties of the quantile function.

(1) implies (3): Suppose $\rho\colon L^1 \longrightarrow (\infty, \infty]$ is a l.s.c., coherent, law-invariant, comonotonic additive risk measure with the Lebesgue property. Denote by $\rho^\flat := \rho|_{L^\infty}$. By Theorem 4.3, there is a probability charge $\mathfrak{q} \in \mathbf{ba}_+$ such that $\rho^\flat = \psi_\mathfrak{q}$. For every vanishing sequence $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$, we have $\limsup_{n\to\infty} \mathfrak{q}(A_n) \leq \lim_{n\to\infty} \rho^\flat(\mathbf{1}_{A_n}) = 0$, i.e., $\mathfrak{q}$ is a countably additive probability measure $\mathbb{Q}$. By the implication (3) $\implies$ (2) and Lemma A.1, the extension ρ' of $\rho^\flat$ in the sense of (5.1) is l.s.c., law invariant, and coherent. It is well known that a coherent law-invariant risk measure on L^∞ has a unique l.s.c., law-invariant and convex extension to L^1; cf. [20]. We therefore obtain $\rho = \rho'$. $\qquad\square$

---

[4] Recall, once more, the actuarial convention of Section 2.

Remark 5.5. The equivalence (1) $\Longleftrightarrow$ (2) was already shown, with different methods, in Shapiro [40] for *finite* risk measures on L^p-spaces; cf. [40, Remark 3]. We are able to deal with the general case of spectral risk measures defined on their canonical model space L^1. A further technical difference is that we do not rely on the existence of measure-preserving transformations, allowing us to establish the result on a general atomless space.

6. Minimal prudent DRMs: the Expected Shortfall

We now return to prudence, our goal being to link the results obtained in Sections 3 and 4 to the Expected Shortfall. The first proposition shows that the latter is indeed the *minimal* family in wide classes of prudent DRMs. This has three possible interpretations, depending on the application one has in mind:

(a) The Expected Shortfall leads to the minimal capital requirement $\rho(X)$ needed to raise and inject in the financial position modelled by X among using a prudent DRM ρ.

(b) The Expected Shortfall leads to the most relaxed capital adequacy test $\mathcal{A}_\rho := \{\rho \leq 0\}$, where a net loss is deemed adequately capitalized if it belongs to the so-called accpeptance set $\mathcal{A}_\rho$ of a prudent DRM.

(c) The last interpretation is more in the spirit of *comparative ambiguity* or *risk aversion*. Let ρ_1 and ρ_2 be two DRMs with associated distortion functions T_1 and T_2, respectively, satisfying $T_i \geq id_{[0,1]}$, $i = 1, 2$. We say that ρ_1 *distorts* $\mathbb{P}$ *more than* ρ_2 if $T_1 \geq T_2$ pointwise ...

Proposition 6.1. *Suppose ρ is a DRM on L^∞ whose associated distortion is exact. Then the following are equivalent:*

(1) *ρ is prudent.*

(2) *$\rho \geq \mathrm{ES}_p$ for some $0 < p \leq 1$.*

In that case, we also have

$$\rho \geq \mathrm{ES}_{\mathfrak{nt}(\rho)}.$$

Proof. (1) implies (2): If ρ is prudent, we may invoke Proposition 3.6 to infer that $\mathfrak{nt}(\rho) > 0$ and that, for all $q > 1 - \mathfrak{nt}(\rho)$, $T_\rho(q) = 1$. As the hypograph $\mathbb{H}(T_\rho)$ is star shaped arount $(0, 0)$ and contains the point $(q, 1)$,

$$T_\rho(x) \geq \tfrac{x}{q} \wedge 1, \quad x \in [0, 1],$$

which implies $\rho \geq \mathrm{ES}_{1-q}$. Letting $q \downarrow 1 - \mathfrak{nt}(\rho)$ and using continuity of $\alpha \mapsto \mathrm{ES}_\alpha(X)$, we obtain $\rho \geq \mathrm{ES}_{\mathfrak{nt}(\rho)}$.

(2) implies (1): If $\rho \geq \mathrm{ES}_p$ for some $0 < p \leq 1$, we may infer for T_ρ and $1 - p < q < 1$ that $T_\rho(q) = T_{\mathrm{ES}_p}(q) = 1$. As T_ρ is also left-continuous (Lemma A.3), ρ must be prudent by Theorem 3.2. $\square$

Remark 6.2. Recall from Theorem 3.2 that a DRM is prudent if and only if it is tail-relevant. The characterization nevertheless deviates from and generalizes [31, Theorem 2] because the DRMs considered are not necessarily coherent.

We are now ready to identify large sets of prudent DRMs containing a minimal element which turns out to be an ES, thereby complementing the already mentioned [31, Theorem 2].

Corollary 6.3. *Let $0 < p \leq 1$ and set*

(i) *$\mathcal{R}^e(p)$ to be the set of prudent DRMs with respect to an exact capacity satisfying $\mathfrak{nt}(\rho) \geq p$;*

(ii) *$\mathcal{R}^c(p) \subset \mathcal{R}^e(p)$ the subset of prudent CCDs;*

(iii) $\mathcal{R}^s(p) \subset \mathcal{R}^e(p)$ the subset of spectral risk measures.

Then ES_p is the minimal risk measure in any of the sets $\mathcal{R}^e(p)$, $\mathcal{R}^c(p)$, and $\mathcal{R}^s(p)$.

The next lemma leverages our results to infer another alternative formula for the index of nontriviality. This characterization is dual and identifies $\mathfrak{nt}(\rho)$ to be the maximal $\mathbb{P}$-probability of null events of elements in the anticore of $T_\rho \circ \mathbb{P}$.

Lemma 6.4. *Let $\rho\colon L^\infty \longrightarrow \mathbb{R}$ be a DRM with associated exact capacity. Then*

$$\mathfrak{nt}(\rho) = \sup\{\mathbb{P}(N) \mid N \in \mathcal{F} \text{ and } \mathfrak{q}(N) = 0 \text{ for some } \mathfrak{q} \in \mathrm{acore}(v)\}. \tag{6.1}$$

For a CCD with backbone $\mathfrak{q}^*$, we have

$$\mathfrak{nt}(\psi_{\mathfrak{q}^*}) = \sup\{\mathbb{P}(N) \mid N \in \mathcal{F} \text{ and } \mathfrak{q}^*(N) = 0\}. \tag{6.2}$$

Proof. No matter whether v is exact or ρ is a CCD, let $\mathfrak{q} \in \mathrm{acore}(v)$ and $N \in \mathcal{F}$ be such that $\mathfrak{q}(N) = 0$. Then $1 \geq T_\rho(1 - \mathbb{P}(N)) = v(N^c) \geq \mathfrak{q}(N^c) = 1$, meaning that $1 - \mathfrak{nt}(\rho) \leq 1 - \mathbb{P}(N)$, or equivalently, $\mathbb{P}(N) \leq \mathfrak{nt}(\rho)$. This shows the estimate "$\leq$" in both (6.1) and (6.2).

In order to show the converse inequality in (6.1), recall from Proposition 3.6(2) that

$$\mathfrak{nt}(\rho) = 1 - \inf\{p \in (0,1] \mid T_\rho(p) = 1\}.$$

Let $0 \leq q < \mathfrak{nt}(\rho)$, which entails that $T_\rho(1 - q) = 1$. Select $N \in \mathcal{F}$ with $\mathbb{P}(N) = q$. Then

$$\max_{\mathfrak{q} \in \mathrm{acore}(v)} \mathfrak{q}(N^c) = v(N^c) = T_\rho(1 - q) = 1,$$

implying that $\mathfrak{q}(N^c) = 1$ for some $\mathfrak{q} \in \mathrm{acore}(v)$. In total,

$$q \leq \sup\{\mathbb{P}(N) \mid N \in \mathcal{F} \text{ and } \mathfrak{q}(N) = 0 \text{ for some } \mathfrak{q} \in \mathrm{acore}(v)\}.$$

As this estimate holds for all $q < \mathfrak{nt}(\rho)$, we obtain the same bound for the index of nontriviality. This shows the desired formula (6.1) in case v is exact.

Now consider the coherent case and suppose $\mathfrak{q}^*$ is a backbone of ρ. One readily verifies that $T'_\rho|_{(1-\mathfrak{nt}(\rho),1)} \equiv 0$. Abbreviating $Z := \frac{d\mathfrak{q}^*}{d\mathbb{P}}$, we may select a random variable $U \in \mathcal{U}$ such that $Z = F_Z^{-1}(U) = T'_\rho(1 - U)$ ([18, Lemma A.32]). Moreover, we can find a vanishing sequence $(B_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ of events such that the pure charge ξ in the Yosida-Hewitt decomposition of $\mathfrak{q}^*$ satisfies $\xi(B_n^c) = 0$, $n \in \mathbb{N}$. Then

$$\mathfrak{q}^*(\{U < \mathfrak{nt}(\rho)\} \cap B_n^c) = \mathbb{E}\big[T'_\rho(1-U)\mathbf{1}_{\{U<\mathfrak{nt}(\rho)\}\cap B_n^c}\big] = 0, \quad n \in \mathbb{N}.$$

In total,

$$\sup\{\mathbb{P}(N) \mid \mathfrak{q}^*(N) = 0\} \geq \sup_{n \in \mathbb{N}} \mathbb{P}\big(\{U < \mathfrak{nt}(\rho)\} \cap B_n^c\big) = \mathfrak{nt}(\rho).$$

This concludes the proof of (6.2). $\qquad\square$

Using Lemma 6.4, we arrive at a dual characterization of prudence for large classes of DRMs that complements the results of Section 3.

Corollary 6.5. *Let $\rho\colon L^\infty \longrightarrow \mathbb{R}$ be a DRM with associated exact capacity v. Then the following are equivalent:*

(1) ρ *is prudent.*
(2) $\mathrm{acore}(v)$ *contains some $\mathfrak{q} \not\approx \mathbb{P}$.*
(3) $\mathrm{acore}(v)$ *contains a probability measure $\mathbb{Q} \not\approx \mathbb{P}$.*

If ρ is additionally coherent, then (1)–(3) are equivalent to:

(4) *Each backbone $\mathfrak{q}$ of ρ satisfies $\mathfrak{q} \not\approx \mathbb{P}$.*

Proof. First of all, note that the distortion function T_ρ is left-continuous both if v is exact and submodular. Thus, by Proposition 3.6, ρ is prudent if and only if $\mathfrak{nt}(\rho) > 0$. Together with (6.1), this immediately shows that (1) and (2) are equivalent, implied by (3), and equivalent to (4) if ρ is coherent. Thus, it suffices to show that (2) implies (3) in the following.

Let $\mathfrak{q} \in \mathrm{acore}(v)$ be such that $\mathfrak{q} \not\approx \mathbb{P}$. Then the coherent DRM $\psi_\mathfrak{q}$ is prudent and dominated by ρ: $\psi_\mathfrak{q} \leq \rho$. By Proposition 6.1, there is $0 < p < 1$ such that $\mathrm{ES}_p \leq \psi_\mathfrak{q} \leq \rho$. For any probability measure $\mathbb{Q} \ll \mathbb{P}$ satisfying $\mathbb{P}(\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{1}{1-p}) = 1 - \mathbb{P}(\frac{d\mathbb{Q}}{d\mathbb{P}} = 0)$ (and thereby necessarily also satisfying $\mathbb{Q} \not\approx \mathbb{P}$),

$$\mathbb{E}_\mathbb{Q}[\cdot] \leq \mathrm{ES}_p \leq \rho,$$

which means that $\mathbb{Q} \in \mathrm{acore}(v)$. $\qquad\square$

We conculde this section with an implication our results have for convex law-invariant functionals $\varphi \colon L^\infty \longrightarrow \mathbb{R}$. More precisely, we give a sufficient dual condition under which φ is l.s.c. with respect to convergence in distribution. To this end, recall the definition of the convex conjugate φ^* in the context of Theorem 5.3.

Corollary 6.6. *For a monotone and convex law-invariant functional φ on L^∞ define*

$$\mathcal{M} := \{\mu \in \mathrm{dom}(\varphi^*) \mid \mu \not\approx \mathbb{P}\} \qquad and \qquad \mathcal{M}^\star := \mathcal{M} \cap \mathbf{ca}.$$

Then, for all $X \in L^\infty$,

$$\sup_{\mu \in \mathcal{M}} \Big(\int X \, d\mu - \varphi^*(\mu) \Big) = \sup_{\zeta \in \mathcal{M}^\star} \Big(\int X \, d\zeta - \varphi^*(\zeta) \Big).$$

Moreover, φ is l.s.c. with respect to convergence in distribution at X satisfying

$$\varphi(X) = \sup_{\mu \in \mathcal{M}} \Big(\int X \, d\mu - \varphi^*(\mu) \Big)$$

Proof. In this proof, we define ψ_μ for an arbitrary $\mu \in \mathbf{ba}_+$ in complete analogy with (4.1). By definition, $\mathcal{M}^\star \subset \mathcal{M}$, and the latter set is nonempty if the former is nonempty. In that case, $\sup_{\mu \in \mathcal{M}} \big(\int X \, d\mu - \varphi^*(\mu) \big) \leq \sup_{\zeta \in \mathcal{M}^\star} \big(\int X \, d\zeta - \varphi^*(\zeta) \big)$. Now assume that we can choose $\mu \in \mathcal{M}$. If $\mu = 0$, then also $\mu \in \mathcal{M}^\star$. If $\mu \neq 0$, we may consider the probability charge $\mathfrak{q} := \mu(\Omega)^{-1}\mu$. By Corollary 6.5, $\psi_\mathfrak{q}$ is prudent. Let $(T_n)_{n \in \mathbb{N}}$ be the sequence of continuous distortion functions defined in the proof of Lemma 5.2 approximatiung T_{ψ_μ}. Moreover, there is $n_0 \in \mathbb{N}$ such that $n \geq n_0$ implies $T_n(x) = 1$ for all $1 - \mathfrak{nt}(\varphi) < x \leq 1$. As observed in that proof, the associated DRMs ψ_n satisfy $\psi_n \leq \psi_\mathfrak{q}$ and are generated by countably additive backbones $\mathbb{Q}_n$. By Corollary 6.5, for all $n \geq n_0$, $\mathbb{Q}_n \not\approx \mathbb{P}$. We infer that every probability measure $\mathbb{Q}'$ in the set

$$\mathcal{Q} = \bigcup_{n \geq n_0} \{ \mathbb{Q}' \ll \mathbb{P} \mid \tfrac{d\mathbb{Q}'}{d\mathbb{P}} \overset{d}{=} \tfrac{d\mathbb{Q}_n}{d\mathbb{P}} \}$$

satisfies, $n \in \mathbb{N}$ being appropriately chosen,

$$
\begin{aligned}
\varphi^*(\mu(\Omega)\mathbb{Q}') &= \sup_{X \in L^\infty} \big(\mu(\Omega)\psi_n(X) - \varphi(X) \big) \\
&\leq \sup_{X \in L^\infty} \big(\mu(\Omega)\psi_{\mathfrak{q}}(X) - \varphi(X) \big) \\
&= \sup_{X \in L^\infty} \Big(\int X' \, d\mu - \varphi(X') \Big) = \varphi^*(\mu).
\end{aligned}
$$

Hence, $\{\mu(\Omega)\mathbb{Q}' \mid \mathbb{Q}' \in \mathcal{Q}\} \subset \mathcal{M}^\star$ and

$$
\sup_{\zeta \in \mathcal{M}^\star} \Big(\int X \, d\zeta - \varphi^*(\zeta) \Big) \geq \sup_{\mathbb{Q}' \in \mathcal{Q}} \big(\mu(\Omega) E_{\mathbb{Q}'}[X] - \varphi^*(\mu(\Omega)\mathbb{Q}') \big) \geq \sup_{n \in \mathbb{N}} \mu(\Omega)\psi_n(X) - \varphi^*(\mu) \geq \int X \, d\mu - \varphi^*(\mu).
$$

Taking the supremum over all $\mu \in \mathcal{M}$ on the right-hand side finishes the proof of the desired identity. Now suppose that $X \in L^\infty$ is such that $\varphi(X) = \sup_{\mathfrak{q} \in \mathcal{M}} \big(\int X \, d\mathfrak{q} - \varphi^*(\mathfrak{q}) \big)$. Applying Corollary 6.5 and Theorem 3.2 if $\mu \neq 0$, each functional

$$
X \longmapsto \sup_{X' \overset{d}{=} X} \int X \, d\mu - \varphi^*(\mu)
$$

is l.s.c. with respect to convergence in distribution. We can thus express $\varphi(X)$ as the evaluation of the hull of a family of functionals that are all l.s.c. with respect to said convergence, which in turn transfers to φ. □

Example 6.7. Suppose ρ is a convex and law-invariant monetary risk measure on L^∞ for which

$$
\big\{ \mathbb{Q} \ll \mathbb{P} \mid \tfrac{d\mathbb{Q}}{d\mathbb{P}} \in L^\infty \big\} \subset \mathrm{dom}(\rho^*)
$$

and for which ρ^* is continuous on the former set with respect to total variation. One may think of the entropic risk measure as an example here. Then Corollary 6.6 applies and ρ is l.s.c. with respect to convergence in distribution.

7. No reward for concentration

The previous section characterizes the Expected Shortfall as the *minimal* element in large classes of prudent DRMs. At first blush, this characterization is very different from the one provided by Wang & Zitikis [43]. The scope of this section is to link the two characterizations, which requires to take into account the fourth axiom considered for the Expected Shortfall in [43]: no reward for concentration. In particular, we shall give an alternative proof of the key part of the Wang-Zitikis axiomatisation of the ES, offering a complementary approach.

Recall that [43] characterizes ES_p, $0 < p < 1$, as the unique functional $\rho \colon L^1 \longrightarrow \mathbb{R}$ satisfying $\rho(1) = 1$, monotonicity, law invariance, prudence, and:

> **No reward for concentration (NRC):** There is an event $A \in \mathcal{F}$ satisfying $0 < \mathbb{P}(A) < 1$ such that, for all $X, Y \in L^1$ sharing the tail event A,
> $$ \rho(X + Y) = \rho(X) + \rho(Y).^5 $$

We first consider the combination of monotonicity, law invariance, and NRC. Functionals with this properties have been characterized in [43] by combining Proposition 4 and Lemma A.2 therein. Our approach is different though and based primarily on the observation that such functionals (up to

---

[5] A being a tail event of X and Y means that there are $\alpha, \beta \in \mathbb{R}$ such that $X\mathbf{1}_A \geq \alpha\mathbf{1}_A$, $X\mathbf{1}_{A^c} \leq \alpha\mathbf{1}_{A^c}$, $Y\mathbf{1}_A \geq \beta\mathbf{1}_A$, and $Y\mathbf{1}_{A^c} \leq \beta\mathbf{1}_{A^c}$.

scaling) are DRMs whose distortion function is of very special shape — it has to be piecewise linear and must consist of two pieces only.

Proposition 7.1. *Suppose a law-invariant, monotone functional $\rho\colon L^\infty \longrightarrow \mathbb{R}$ satisfies $\rho(1) = 1$ and offers NRC with tail event A. Then ρ is a DRM whose distortion function T_ρ is given by*

$$T_\rho(x) = \begin{cases} \gamma x & x \leq \mathbb{P}(A) \\ \gamma\mathbb{P}(A) + \delta(x - \mathbb{P}(A)) & x > \mathbb{P}(A). \end{cases}$$

The constants $\gamma, \delta \geq 0$ satisfy

$$(\gamma - \delta)\mathbb{P}(A) = 1 - \delta. \tag{7.1}$$

Proof. In a first step, we prove ρ is a Choquet integral with respect to a $\mathbb{P}$-distortion. To this effect, it suffices to prove that ρ is comonotonic additive. Let $p := \mathbb{P}(A) \in (0, 1)$ and select random variables $U_1\colon A \longrightarrow \mathbb{R}$ and $U_2\colon A^c \longrightarrow \mathbb{R}$ such that $\mathbb{P}(\cdot|A) \circ U_1^{-1}$ is a uniform distribution over $(0, p)$, while $\mathbb{P}(\cdot|A^c) \circ U_2^{-1}$ is a uniform distribution over $(p, 1)$. The random variable $\widehat{U} := U_2 \mathbf{1}_{A^c} + U_1 \mathbf{1}_A$ is then seen to have a uniform distribution over $(0, 1)$. Let $X, Y \in L^1$ be comonotone. Select nondecreasing functions $f, g\colon \mathbb{R} \longrightarrow \mathbb{R}$ such that $X = f(X + Y)$ and $Y = g(X + Y)$ ([18, Lemma 4.89]). Let $U \in \mathcal{U}$ be such that $X + Y = F_{X+Y}^{-1}(U)$. By using the law-invariance of ρ in the second and fourth equality below and NRC in the third, we have

$$\rho(X + Y) = \rho\big((f \circ F_{X+Y}^{-1})(U) + (g \circ F_{X+Y}^{-1})(U)\big) = \rho\big((f \circ F_{X+Y}^{-1})(\widehat{U}) + (g \circ F_{X+Y}^{-1})(\widehat{U})\big)$$
$$= \rho\big((f \circ F_{X+Y}^{-1})(\widehat{U})\big) + \rho\big((g \circ F_{X+Y}^{-1})(\widehat{U})\big) = \rho(X) + \rho(Y).$$

Now that we have verified that ρ is a DRM, it remains to extract the precise shape of the associated distortion function $T := T_\rho$. Suppose that $x, y \in [0, p]$ are such that $x + y \leq p$. Fix disjoint events $B, C \subset A$ such that $\mathbb{P}(B) = x$ and $\mathbb{P}(C) = y$. As $\mathbf{1}_C$ and $\mathbf{1}_B$ share the tail event A, NRC shows that

$$T(x) + T(y) = \rho(\mathbf{1}_B) + \rho(\mathbf{1}_C) = \rho(\mathbf{1}_{B \cup C}) = T(x + y).$$

Thus, T is additive on $[0, p]$, i.e., $T(x) = \gamma x$ for a suitable $\gamma \geq 0$ and all $x \in [0, p]$. Now let $s, t \in [0, 1]$ such that $p + s + t \leq 1$. Select event $D, E \supset A$ such that

$$\mathbb{P}(D) = p + s, \quad \mathbb{P}(E) = p + t, \quad \mathbb{P}(D \cup E) = p + s + t, \quad D \cap E = A.$$

By using NRC in the second equality below and comonotonic additivity in the third, we have

$$T(p + t) + T(p + s) = \rho(\mathbf{1}_D) + \rho(\mathbf{1}_E) = \rho(\mathbf{1}_{D \cup E} + \mathbf{1}_A) = \rho(\mathbf{1}_{D \cup E}) + \rho(\mathbf{1}_A) = T(p + s + t) + T(p).$$

Hence, for all such choices of s, t,

$$T(p + s + t) - T(p + s) = T(p + t) - T(p).$$

which implies that, on $[p, 1]$, T has shape

$$T(x) = T(p) + \delta(x - p) = \gamma p + \delta(x - p), \quad x \in [p, 1],$$

the constant $\delta \geq 0$ being suitably chosen. As $T(1) = 1$, we infer the normalization condition $\gamma p + \delta(1 - p) = 1$. Rearranging this yields (7.1). $\qquad\square$

Corollary 7.2. *A law-invariant, monotone functional $\rho\colon L^\infty \longrightarrow \mathbb{R}$ which offers NRC is necessarily sub- or superadditive.*

Proof. If $\rho \neq 0$, we can assume without loss of generality that ρ is a DRM. Let $A \in \mathcal{F}$ be the tail event with respect to which ρ offers no reward for concentration. By Proposition 7.1 we can distinguish two cases.

CASE 1: $T_\rho(\mathbb{P}(A)) = \gamma \mathbb{P}(A) \geq \mathbb{P}(A)$. In this case, T_ρ is concave, which means that ρ must be subadditive.

CASE 2: $T_\rho(\mathbb{P}(A)) < 1$. In this case, T_ρ is convex, which means that the function $\overline{T} \colon [0,1] \longrightarrow [0,1]$ defined by $\overline{T}(x) := 1 - T_\rho(1-x)$ is concave. Note that the dual capacity $\overline{v}(A) := 1 - v(A^c)$, $A \in \mathcal{F}$, is given precisely by $\overline{T} \circ \mathbb{P}$. Hence, $\int \cdot \, d\overline{v}$ is convex. Moreover, by [33, Proposition 4.12], we have for all $X \in L^\infty$ that

$$\rho(X) = \int X \, dv = - \int (-X) \, d\overline{v},$$

meaning that ρ must be concave. $\qquad\square$

For the sake of completeness, we state how the result of [43] follows seamlessly from Proposition 7.1 and the characterization of prudence given in Theorem 3.2.

Theorem 7.3. [43] *Suppose a functional $\rho \colon L^1 \longrightarrow \mathbb{R}$ is law invariant, monotone, and satisfies NRC as well as prudence. Then $\rho = \mathrm{ES}_p$ for some $p \in (0,1)$.*

Proof. Let $p := \mathbb{P}(A^c)$ for the tail event A associated with ρ. Then $\rho|_{L^\infty}$ is a distortion risk measure as in Proposition 7.1. By Theorem 3.2, we find $1 - p \leq x < 1$ such that $T_\rho(x) = 1$. In particular, letting γ, δ as in Proposition 7.1, $\gamma(1-p) + \delta(x - 1 + p) = \gamma + p(\delta - \gamma)$, which entails $\delta(x-1) = 0$. This is possible only if $\delta = 0$. By (7.1), $\gamma = \frac{1}{1-p}$, which means that $T_\rho = \frac{x}{1-p} \wedge 1$, and $\rho|_{L^\infty} = \mathrm{ES}_p$. This identity is extended to all of L^1 using prudence; for instance, one may apply the uniqueness result in Corollary B.3 below. $\qquad\square$

While rather evident at this point, it is worth recording once more how the WZ axioms imply the more basic axioms studied in this paper and satisfied by the Expected Shortfall. Both sets of axioms contain monotonicity and law invariance. The two together with NRC imply comonotonic additvity (and thereby also cash-additivity); cf. Proposition 7.1. The same combination of axioms also shows that the associated distortion function satisfies $T_\rho(0^+) = 0$, i.e., the resulting risk measure has the Lebesgue property. At last, the addition of prudence implies the concavity of T_ρ, i.e. the subadditivity of ρ. By Proposition 5.4. Again, prudence pins down the risk measure uniquely on L^∞ given the minimality result in Corollary 6.3. In another step, it allows to transfer all considerations from L^∞ to L^1. In sum, both prudence and NRC play — at least — a twofold role in the axiomatization, and our considerations indeed disentangle them.

At last, note that by virtue of Proposition 7.1, the Expected Shortfall requires *maximal* capital buffers among law-invariant risk measures offering no reward for concentration. This dual observation to the minimality result in Corollary 6.3 demonstrates that prudence and NRC express two different types of caution, the former expressing more conservative risk attitudes. In other words, NRC as an axioms acts dually to prudence.

APPENDIX A. AUXILIARY RESULTS

Lemma A.1. (2.2) *gives a well-defined map $\rho \colon L^1 \longrightarrow (-\infty, \infty]$ that is additionally norm-l.s.c. and subadditive.*

Proof. Note that Lebesgue-a.e., $(\phi F_X^{-1}) \wedge 0 = \phi(F_X^{-1} \wedge 0) = \phi F_{X \wedge 0}^{-1}$. Hence, for arbitrary $0 < \alpha < 1$, computing the integral of the negative part of ϕF_X^{-1} gives

$$
\begin{aligned}
\int_0^1 (\phi(t) F_X^{-1}(t))^- \mathrm{d}t &= \int_0^1 \phi(t)(-F_{X \wedge 0}^{-1}(t)) \mathrm{d}t \\
&\leq \int_0^\alpha \phi(t)(-F_{X \wedge 0}^{-1}(t)) \mathrm{d}t + \int_\alpha^1 \phi(t)(-F_{X \wedge 0}^{-1}(t)) \mathrm{d}t \\
&\leq \phi(\alpha) \int_0^1 (-F_{X \wedge 0}^{-1}(t)) \mathrm{d}t + |F_{X \wedge 0}^{-1}(\alpha)| \int_0^1 \phi(t) \mathrm{d}t \\
&= \phi(\alpha) \mathbb{E}[X^-] + |F_{X \wedge 0}^{-1}(\alpha)| < \infty.
\end{aligned}
$$

This shows that ρ is well-defined and does not attain the value $-\infty$. Next, we show L^1-l.s.c. Note first that we can instead verify the following property:

$$
(X_n)_{n \in \mathbb{N}} \subset L^1 \text{ satisfies } X_n \uparrow X \in L^1 \implies \rho(X_n) \uparrow \rho(X). \tag{A.1}
$$

Indeed, every norm-l.s.c. monotone functional on L^1 has this property. Conversely, let $(Y_n)_{n \in \mathbb{N}} \subset L^1$ be convergent to $Y \in L^1$. Select a subsequence $(n_k)_{k \in \mathbb{N}}$ with the following properties:

(i) $\lim_{k \to \infty} \rho(Y_{n_k}) = \liminf_{n \to \infty} \rho(Y_n)$.
(ii) $\sum_{k \in \mathbb{N}} \mathbb{E}[|Y_{n_k} - Y|] < \infty$.

In particular, setting $X_m := \inf_{k \geq m} Y_{n_k}$, $m \in \mathbb{N}$, we obtain a sequence in L^1 because

$$
Y_1 \geq X_m \geq |X| - |X_m - X| \geq |X| - \sum_{k \in \mathbb{N}} |Y_{n_k} - X| \in L^1.
$$

Moreover, $X_m \uparrow Y$ as $m \to \infty$. Consequently,

$$
\rho(Y) = \sup_{m \in \mathbb{N}} \rho(X_m) \leq \lim_{k \to \infty} \rho(Y_{n_k}) = \liminf_{n \to \infty} \rho(X_n),
$$

and (A.1) is verified to be equivalent to L^1-lower semicontinuity of ρ. Now, if $(X_n)_{n \in \mathbb{N}} \subset L^1$ satisfies $X_n \uparrow X$, monotone convergence implies that ρ satisfies (A.1).

We now turn to verifying subadditivity of ρ. To this end, let $X, Y \in L^1$ and set $Z := X + Y$. By [19, Proposition 5.1] there are nondecreasing continuous functions $f, g \colon \mathbb{R} \longrightarrow \mathbb{R}$ such that $f + g = id_\mathbb{R}$, X dominates $f(Z)$ in convex order, and Y dominates $g(Z)$ in convex order. In particular, $f(Z), g(Z)$ are comonotone. Using [7, Lemma 3.4] for the first estimate and selecting $n \in \mathbb{N}$ large enough such that the second holds,

$$
\begin{aligned}
\int_0^1 (\phi(t) \wedge n) F_Z^{-1}(t) \mathrm{d}t &= \int_0^1 (\phi(t) \wedge n) F_{f(Z)}^{-1}(t) \mathrm{d}t + \int_0^1 (\phi(t) \wedge n) F_{g(Z)}^{-1}(t) \mathrm{d}t \\
&\leq \int_0^1 (\phi(t) \wedge n) F_X^{-1}(t) \mathrm{d}t + \int_0^1 (\phi(t) \wedge n) F_Y^{-1}(t) \mathrm{d}t \\
&\leq \rho(X) + \rho(Y).
\end{aligned}
$$

Now let $n \to \infty$ on the left-hand side and use monotone convergence. $\square$

Lemma A.2. *Let $v = T \circ \mathbb{P}$ be a distortion of an atomless probability measure $\mathbb{P}$ that satisfies* $\mathrm{acore}(v) \neq \varnothing$. *Then T is continuous at 1. In particular, if v is a submodular distortion, the distortion function T is continuous on $(0, 1]$.*

Proof. By [5, Corollary 3.1], $T \geq id_{[0,1]}$ holds under the assumption of the lemma. For any sequence $(x_n)_{n \in \mathbb{N}} \subset [0,1]$ satisfying $x_n \uparrow 1$ as $n \to \infty$, we thus observe

$$1 \geq \sup_{n \in \mathbb{N}} T(x_n) \geq \lim_{n \to \infty} x_n = 1,$$

which is sufficient for continuity of T at 1. If v is submodular, T is concave and thus continuous on $(0,1]$. $\qquad\square$

Lemma A.3. *Let $v = T \circ \mathbb{P}$ be an exact distortion of an atomless probability measure $\mathbb{P}$. Then T is left-continuous.*

Proof. By [29, Proposition 5.10], we have for the set $\mathcal{Q}$ of countably additive elements in $\mathrm{acore}(v)$ that

$$v(A) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}(A), \quad A \in \mathcal{F}.$$

Let $0 \leq x_n \uparrow x \leq 1$ and suppose $U \in \mathcal{U}$. Then

$$\sup_{n \in \mathbb{N}} T(x_n) = \sup_{n \in \mathbb{N}} v(\{U \leq x_n\}) = \sup_{\mathbb{Q} \in \mathcal{Q}} \sup_{n \in \mathbb{N}} \mathbb{Q}(\{U \leq x_n\}) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}(\{U \leq x\}) = v(\{U \leq x\}) = T(x).$$

This establishes left-continuity of T. $\qquad\square$

Appendix B. Extension properties of prudent DRMs

Suppose $\rho \colon L^\infty \longrightarrow \mathbb{R}$ is a prudent DRM. By the proof of Theorem 3.2, ρ is tail relevant (in the terminology of [31]) and we find $q \in (0,1)$ such that, for all $X \in L^\infty$,

$$\rho(X) = \rho\big(X \vee \mathrm{VaR}_q(X)\big) = \int_0^\infty T_\rho\big(\mathbb{P}(X > x)\big)\mathrm{d}x + \int_{\mathrm{VaR}_q(X) \wedge 0}^0 \big(1 - T_\rho\big(\mathbb{P}(X > x)\big)\big)\mathrm{d}x. \qquad (\mathrm{B}.1)$$

Definition B.1. For a prudent distortion risk measure ρ on L^∞ and q chosen as above, we define $\rho^\sharp \colon L^0 \longrightarrow (-\infty, \infty]$ by

$$\rho^\sharp(X) := \int_0^\infty T_\rho\big(\mathbb{P}(X > x)\big)\mathrm{d}x + \int_{\mathrm{VaR}_q(X) \wedge 0}^0 \big(1 - T_\rho\big(\mathbb{P}(X > x)\big)\big)\mathrm{d}x.$$

The following proposition demonstrates that the preceding definition canonically extends the original ρ to all of L^0 retaining all of its nice properties. This is of independent interest because the question how to define risk measures on L^0 has a long history in the literature and is treated for coherent risk measures in [14, Section 5]. Moreover, most extension results for risk measures in the literature make use of their dual representation and therefore have to assume convexity; cf. [20, 30]. Prudent DRMs are not necessarily convex.

Proposition B.2. *The extension $\rho^\sharp$ is well defined, monotone, law invariant, prudent, l.s.c. with respect to convergence in distribution, and comonotonic additive.*

Proof. STEP 1: $\rho^\sharp$ does not depend on the concrete choice of a feasible $q \in (0,1)$. Indeed, suppose two thresholds $0 < q < r < 1$ satisfy (B.1). Then, for all $t > 1 - r$ and $U \in \mathcal{U}$, $\mathrm{VaR}_r(\mathbf{1}_{\{U \leq t\}}) = 1$. Hence,

$$T_\rho(t) = \rho(\mathbf{1}_{\{U \leq t\}}) = \rho(\mathbf{1}_{\{U \leq t\}} \vee 1) = \rho(1) = 1.$$

Moreover, observe that for $x < \text{VaR}_r(X)$, $\mathbb{P}(X > x) > 1 - r$, which entails for the function $h(x) := 1 - T_\rho(\mathbb{P}(X > x))$, $x \in [0, 1]$, that

$$\int_{\text{VaR}_q(X) \wedge 0}^0 h(x)\,\mathrm{d}x = \int_{\text{VaR}_q(X) \wedge 0}^{\text{VaR}_r(X) \wedge 0} h(x)\,\mathrm{d}x + \int_{\text{VaR}_r(X) \wedge 0}^0 h(x)\,\mathrm{d}x$$

$$= \int_{\text{VaR}_r(X) \wedge 0}^0 h(x)\,\mathrm{d}x.$$

STEP 2: For every $X \in L^0$, the function h defined in Step 1 is nondecreasing, hence its integral over the bounded domain $[\text{VaR}_q(X) \wedge 0, 0]$ is always finite. Thus, $\rho^\sharp(X)$ is well defined.

STEP 3: $\rho^\sharp$ is law invariant and monotone on L^0 by definition. Its prudence is verified once we show the stronger property of $\rho^\sharp$ being l.s.c. with respect to convergence in distribution.

STEP 4: Suppose $(X_n)_{n \in \mathbb{N}} \subset L^0$ is a sequence such that $X_n \xrightarrow{d} X \in L^0$ as $n \to \infty$. By Skorokhod's representation, it suffices to consider the case where $X_n \leq X$ and $X_n \to X$ $\mathbb{P}$-a.s. In particular, left-continuity of T_ρ (Theorem 3.2) ensures that, Lebesgue-a.e.,

$$\lim_{n \to \infty} T_\rho\left(\mathbb{P}(X_n > x)\right) = T_\rho\left(\mathbb{P}(X > x)\right).$$

By [3, Theorem 11.32] and dominated convergence, we obtain for all $k \in \mathbb{N}$ that

$$\liminf_{n \to \infty} \int_0^\infty T_\rho\left(\mathbb{P}(X_n > x)\right)\mathrm{d}x \geq \lim_{n \to \infty} \int_0^k T_\rho\left(\mathbb{P}(X_n > x)\right)\mathrm{d}x = \int_0^k T_\rho\left(\mathbb{P}(X > x)\right)\mathrm{d}x.$$

Moreover, by Step 1 and Skorokhod representation, we can select the parameter $q \in (0, 1)$ such that $\lim_{n \to \infty} \text{VaR}_q(X_n) = \text{VaR}_q(X)$. In particular, the sequence $(\text{VaR}_q(X_n))_{n \in \mathbb{N}}$ is bounded below by a constant $c \leq 0$. By the argument underlying Step 1 we infer for all $Y \in \{X, X_1, X_2, \ldots\}$ that

$$\int_{\text{VaR}_q(Y) \wedge 0}^0 \left(1 - T_\rho\left(\mathbb{P}(Y > x)\right)\right)\mathrm{d}(x) = \int_c^0 \left(1 - T_\rho\left(\mathbb{P}(Y > x)\right)\right)\mathrm{d}(x).$$

[3, Theorem 11.32] and dominated convergence again imply

$$\lim_{n \to \infty} \int_{\text{VaR}_q(X_n) \wedge 0} \left(1 - T_\rho\left(\mathbb{P}(X_n > x)\right)\right)\mathrm{d}(x) = \int_{\text{VaR}_q(X) \wedge 0}^0 \left(1 - T_\rho\left(\mathbb{P}(X > x)\right)\right)\mathrm{d}(x).$$

Using the monotonicity of $\rho^\sharp$ in the last estimate,

$$\rho^\sharp(X) = \sup_{k \in \mathbb{N}} \int_0^k T_\rho\left(\mathbb{P}(X > x)\right)\mathrm{d}x + \int_{\text{VaR}_q(X) \wedge 0}^0 \left(1 - T_\rho\left(\mathbb{P}(X > x)\right)\right)\mathrm{d}(x)$$

$$\leq \liminf_{n \to \infty} \int_0^\infty T_\rho\left(\mathbb{P}(X_n > x)\right)\mathrm{d}x + \liminf_{n \to \infty} \int_{\text{VaR}_q(X_n) \wedge 0}^0 \left(1 - T_\rho\left(\mathbb{P}(X_n > x)\right)\right)\mathrm{d}(x)$$

$$\leq \liminf_{n \to \infty} \rho^\sharp(X_n) \leq \limsup_{n \to \infty} \rho^\sharp(X_n) \leq \rho^\sharp(X).$$

STEP 5: $\rho^\sharp$ is comonotonic additive. Let $X, Y \in L^0$ be comonotonic. By [15, Proposition 4.5] we find continuous functions $f, g \colon \mathbb{R} \to \mathbb{R}$ such that $X = f(X + Y)$ and $Y = g(X + Y)$. By [18, Lemma A.32], there is $U \in \mathcal{U}$ such that $X + Y = F_{X+Y}^{-1}(U)$. Set $f^\star := f \circ F_{X+Y}^{-1}$, $g^\star := g \circ F_{X+Y}^{-1}$, and note that these two functions are left-continuous. Using [18, Lemma A.27] in the penultimate identity

$$f^\star(U \vee q) = f\left(F^{-1}(X + Y) \vee \text{VaR}_q(X + Y)\right) = X \vee \text{VaR}_q(f(X + Y)) = X \vee \text{VaR}_q(X).$$

In complete analogy, $g^\star(U \vee q) = Y \vee \mathrm{VaR}_q(Y)$. Abbreviate $U_n := \big(U \wedge (1 - qe^{-n})\big) \vee q$, $n \in \mathbb{N}$, and observe that

$$\rho^\sharp(X + Y) = \rho^\sharp\big((X + Y) \vee \mathrm{VaR}_q(X + Y)\big) = \rho^\sharp\big((f^\star + g^\star)(U \vee q)\big)$$

Now we can use Step 4 and left-continuity of the involved functions to see that

$$\rho^\sharp\big((f^\star + g^\star)(U \vee q)\big) = \lim_{n \to \infty} \rho\big((f^\star + g^\star)(U_n)\big) = \lim_{n \to \infty} \Big(\rho\big(f^\star(U_n)\big) + \rho\big(g^\star(U_n)\big)\Big).$$

In the last identity we have used comonotonic additivity of ρ. Invoking Step 4 again for $f^\star$,

$$\lim_{n \to \infty} \rho\big(f^\star(U_n)\big) = \rho^\sharp\big(f^\star(U \vee q)\big) = \rho^\sharp\big(X \vee \mathrm{VaR}_q(X)\big) = \rho^\sharp(X).$$

Analogously, $\lim_{n \to \infty} \rho\big(g^\star(U_n)\big) = \rho^\sharp(Y)$ and the proof is complete. $\qquad\square$

As a consequence of the preceding proposition, we may identify situations in which prudent functionals on domains of unbounded random variables are uniquely determined by their values on L^∞. More precisely, the following result holds.

Corollary B.3. *Suppose $\mathcal{X} \subset L^0$ is a lattice containing L^∞. Suppose that $\varphi\colon \mathcal{X} \to \mathbb{R}$ is monotone, law invariant, prudent, and comonotonic additive. Set $\rho := \varphi|_{L^\infty}$. Then ρ is a prudent DRM and $\varphi = \rho^\sharp|_{\mathcal{X}}$.*

Proof. ρ is finite, monotone, comonotonic additive, law invariant, and prudent, that is, a prudent DRM. As such, $\rho^\sharp$ is well defined on L^0. Let $X \in \mathcal{X}$, $m \in \mathbb{N}$, and set $Y := X \vee (-m)$. Observe that $Y \wedge n \uparrow Y$ a.s. Hence, $\lim_{n \to \infty} \rho(Y \wedge n) = \varphi(Y)$ must hold by prudence. However, the argument from Step 4 in the proof of the preceding proposition also shows that $\lim_{n \to \infty} \rho(Y \wedge n) = \rho^\sharp(Y)$. Hence, for all $(X, m) \in \mathcal{X} \times \mathbb{N}$, $\varphi(X \vee (-m)) = \rho^\sharp(X \vee (-m))$. Now observe that $Z_m := (X + m)\mathbf{1}_{\{X < -m\}} \leq 0$ for all $m \in \mathbb{N}$ and $Z_m \uparrow 0$ as $m \to \infty$. By prudence, $\lim_{n \to \infty} \varphi(Z_m) = 0 = \lim_{m \to \infty} \rho^\sharp(Z_m)$. Using comonotonic additivity,

$$\varphi(X) = \varphi(X) - \lim_{m \to \infty} \varphi(Z_m) = \lim_{m \to \infty} \rho^\sharp(X \vee (-m)) = \rho^\sharp(X) - \lim_{m \to \infty} \rho^\sharp(Z_m) = \rho^\sharp(X).$$

$\qquad\square$

One potential drawback of Corollary B.3 is that the functional must be finite-valued. Assuming that φ is not only prudent, but l.s.c. with respect to convergence in distribution, we can drop this assumption. *Inter alia*, we show that $\rho^\sharp$ is the *unique* extension of a prudent DRM ρ to L^0 retaining all of its nice properties.

Corollary B.4. *Suppose $\mathcal{X} \subset L^0$ is a lattice containing L^∞. Suppose that $\varphi\colon \mathcal{X} \to (-\infty, \infty]$ satisfies $\varphi(0) = 0$, is monotone, law invariant, prudent, and comonotonic additive. Set $\rho := \varphi|_{L^\infty}$. Then ρ is a prudent DRM and $\varphi = \rho^\sharp|_{\mathcal{X}}$.*

Proof. We first show that ρ only takes finite values. By monotonicity, it suffices to show that $\varphi(x) \in \mathbb{R}$ holds for all $x \in \mathbb{R}$. Indeed, for $x \leq 0$, $\rho(x) \leq \rho(0) = 0$. For $x > 0$, comonotonic additivity implies

$$\rho(x) = \rho(x) + \rho(-x) - \rho(-x) = \rho(0) - \rho(-x) = -\rho(-x) < \infty.$$

Thus, ρ is finite-valued, monotone, comonotonic additive, law invariant, and prudent, and therefore a prudent DRM. The remainder of the proof is identical to the one of Corollary B.3 up to the observation that $\lim_{n \to \infty} \rho(Y \wedge n) = \varphi(Y)$ is established for Y bounded below by virtue of lower semicontinuity with respect to convergence in distribution, not prudence. $\qquad\square$

We conclude this appendix with a more general observation, establishing sufficient conditions for prudence of a law-invariant monotone functional being equivalent to lower semicontinuity with respect to convergence in distribution. As anticipated in Remark 3.3, the general relation is an open question to us.

Proposition B.5. *Suppose $\varphi\colon L^\infty \longrightarrow (-\infty,\infty]$ is law invariant, monotone, and prudent, and that one of the following additional conditions holds:*

(i) *φ has a monotone and proper extension $\varphi^\sharp$ to all of L^0.*

(ii) *There are $a \geq 0$, $b \in \mathbb{R}$, and $p \in (0,1)$ such that*
$$\varphi \geq a\mathrm{VaR}_p + b.$$

(iii) *There is a finite measure $\zeta \ll \mathbb{P}$ such that $\zeta \not\approx \mathbb{P}$ and $\beta \in \mathbb{R}$ satisfying*
$$\varphi \geq \int \cdot \, d\zeta + \beta. \tag{B.2}$$

Then φ is also lower semicontinuous with respect to convergence in distribution.

Proof. Suppose $(X_n)_{n\in\mathbb{N}} \subset L^\infty$ is a sequence satisfying $X_n \overset{d}{\to} X \in L^\infty$ as $n \to \infty$. We need to show that $\varphi(X) \leq \liminf_{n\to\infty} \varphi(X_n)$. By law invariance, we can assume the existence of $U \in \mathcal{U}$ such that $X_n = F_{X_n}^{-1}(U) \longrightarrow F_X(U) = X$ a.s. If $\lim_{n\to\infty} \varphi(X_n) \in (-\infty,\infty]$ exists, the assertion follows either trivially or invoking prudence. In the following, we prove that $\liminf_{n\to\infty} \varphi(X_n) = -\infty$ is impossible. Without loss, we may assume $\varphi(X_n) \downarrow -\infty$ and $\|X\|_\infty \leq \|X_n\|_\infty - 1 \to \infty$ for $n \to \infty$. Set $A_n := \{X_n > -\|X\|_\infty - 1\}$, $n \in \mathbb{N}$, and
$$Y_n := -(\|X\|_\infty + 1)\mathbf{1}_{A_n} - \|X_n\|\mathbf{1}_{A_n^c}.$$

Then $Y_n \leq X_n$, i.e., $\varphi(Y_n) \leq \varphi(X_n) \downarrow -\infty$ as $n \to \infty$. If additional condition (i) holds, we can choose a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $\sum_{k=1}^\infty \mathbb{P}(A_{n_k}^c) < \infty$. Thus,
$$Y := \sum_{k=1}^\infty Y_{n_k}$$

is a well-defined random variable in $-L_+^0$ satisfying $Y \leq Y_{n_k}$ for all $k \in \mathbb{N}$. Hence, $-\infty < \varphi(Y) \leq \inf_{k\in\mathbb{N}} \varphi(Y_{n_k}) = -\infty$, a contradiction.

If alternative (ii) holds, the sequence $(Y_n)_{n\in\mathbb{N}}$ satisfies
$$-a\|X\|_\infty - a + b = a\mathrm{VaR}_p(Y_n) + b \leq \varphi(Y_n)$$

for all n large enough. Hence, the assumption that $\varphi(X_n) \downarrow -\infty$ has to be absurd in this case as well. At last, suppose that (iii) holds. If $\zeta = 0$, then $\varphi \geq 0$ and a sequence constructed as above cannot exist. The equivalence of prudence and lower semicontinuity with respect to convergence in distribution is clear. Else, let $\mathbb{Q} := \zeta(\Omega)^{-1}\zeta$ and note that — by law invariance of φ — (B.2) holds if and only if $\varphi \geq \zeta(\Omega)^{-1}\psi_\mathbb{Q} + \beta$. As $\mathbb{Q} \not\approx \mathbb{P}$, $\psi_\mathbb{Q}$ is prudent by Corollary 6.5. By Theorem 3.2, there is $p \in (0,1)$ such that $\psi_\mathbb{Q} \geq \mathrm{VaR}_p$. Hence,
$$\varphi \geq \zeta(\Omega)^{-1}\mathrm{VaR}_p + \beta,$$

and condition (ii) is satisfied. □

Example B.6. Suppose that for a convex law-invariant risk measure ρ on L^∞ there is $\mathfrak{q} \in \Delta$ such that $\mathfrak{q} \not\approx \mathbb{P}$ and $\rho^*(\mathfrak{q}) < \infty$. Then ρ satisfies alternative (ii) in Proposition B.5 and is prudent if and only if l.s.c. with respect to convergence in distribution.

References

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

[2] Acerbi, C., and D. Tasche (2002), Expected Shortfall: a natural coherent alternative to value at risk. *Economic Notes* 31(2):379–388.

[3] Aliprantis, C. D., and K. C. Border (2006), *Infinite Dimensional Analysis: A Hitchhiker's Guide*. 3rd edition, Springer.

[4] Amarante, M., & M. Ghossoub (2021+), Aggregation of opinions and risk measures. Forthcoming in *Journal of Economic Theory*, https://doi.org/10.1016/j.jet.2021.105310.

[5] Aouani, Z., and A. Chateauneuf (2008), Exact capacities and star-shaped distorted probabilities. *Mathematical Social Sciences*, 56(2):185–194.

[6] Artzner, P., F. Delbaen, J. M. Eber, and D. Heath (1999), Coherent Measures of Risk. *Mathematical Finance* 9(3):203–228.

[7] Bellini, F., P. Koch-Medina, C. Munari, and G. Svindland (2021), Law-invariant functionals on general spaces of random variables. *SIAM Journal on Financial Mathematics*, 12(1):318–341.

[8] Bhaskara Rao, K. P. S., and M. Bhaskara Rao (1983), *Theory of Charges: A Study of Finitely Additive Measures*. Academic Press.

[9] Carlier, G., and R. A. Dana (2003), Core of convex distortions of a probability. *Journal of Economic Theory* 113(2):199–222.

[10] Castagnoli, E., F. Maccheroni, and M. Marinacci (2004), Choquet insurance pricing: A caveat. *Mathematical Finance*, 14:481–485.

[11] Chen, S., N. Gao, D. Leung, and L. Li (2022), Automatic Fatou property of law-invariant risk measures. *Insurance: Mathematics and Economics* 105:41–53.

[12] Danilov, V. I. and G.A. Koshevoy (2000) Cores of cooperative games, superdifferentials of functions, and Minkowski difference of sets. *Journal of Mathematical Analysis and Applications* 247:1–14.

[13] Day, P. W. (1973), Decreasing rearrangements and doubly stochastic operators. *Transactions of the American Mathematical Society* 178:383–392.

[14] Delbaen, F. (2002), Coherent risk measures on general probability spaces. In: Sandmann, K., Schönbucher, P.J. (eds.), *Advances in Finance and Stochastics*. Springer.

[15] Denneberg, D. (1994), *Non-Additive Measure and Integral*. Springer.

[16] Dhaene, J., A. Kukush, D. Linders, and Q. Tang (2012), Remarks on quantiles and distortion risk measures. *European Actuarial Journal* 2(2):319–328.

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

[18] Föllmer, H., and A. Schied (2016), *Stochastic Finance: An Introduction in Discrete Time*. 4th edition, De Gruyter.

[19] Filipović, D., and G. Svindland (2008), Optimal capital and risk allocations for law- and cash-invariant convex functions. *Finance and Stochastics* 12:423–439.

[20] Filipović, D., and G. Svindland (2012), The canonical model space of law invariant risk measures is L^1. *Mathematical Finance* 22(3):585–589.

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

[22] Ghirardato, P., and M. Marinacci (2002), Ambiguity made precise: a comparative foundation. *Journal of Economic Theory* 102:251–289.

[23] He, X. D., and X. Peng (2018), Surplus-invariant, law-invariant, and conic acceptance sets must be the sets induced by Value at Risk. *Operations Research* 66(5):1268–1275.

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

[25] Kadane, J. B., and L. Wasserman (1996), Symmetric, coherent, Choquet capacities. *The Annals of Statistics* 24(3):1250–1264.

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

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

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

[29] Liebrich, F.-B., and C. Munari (2022), Law-invariant functionals that collapse to the mean: Beyond convexity. *Mathematics and Financial Economics* 16:447–480.

[30] Liebrich, F.-B., and G. Svindland (2017), Model spaces for risk measures. *Insurance: Mathematics and Economics* 77:150–165.

[31] Liu, F., and R. Wang (2020), A theory for measures of tail risk. *Mathematics of Operations Research* 46(3):1109–1128.

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

[33] Marinacci, M., and L. Montrucchio (2004), Introduction to the Mathematics of Ambiguity. In: *Uncertainty in Economic Theory: a Collection of Essays in Honor of David Schmeidler's 65th Birthday*.

[34] Mirsky, L. (1963), Results and problems in the theory of doubly-stochastic matrices. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 1:319–334.

[35] Ruszczyński, A., and A. Shapiro (2006), Optimization of convex risk functions. *Mathematics of Operations Research* 31(3):433–452.

[36] Schmeidler, D. (1972), Cores of exact games I, *Journal of Mathematical Analysis and Applications* 40 (1): 214-225.

[37] Schmeidler, D. (1986), Integral representation without additivity. *Proceedings of the American Mathematical Society* 97:255–261.

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

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

[40] Shapiro, A. (2013), On Kusuoka representation of law invariant risk measures. *Mathematics of Operations Research* 38(1):142–152.

[41] Svindland, G. (2010), Continuity properties of law-invariant (quasi-)convex risk functions on L^∞. *Mathematics and Financial Economics* 3(1):39–43.

[42] Wang, R., Y. Wei, and G. Willmot (2020), Characterization, robustness and aggregation of signed Choquet integrals. *Mathematics of Operations Research* 45(3):993–1015.

[43] Wang, R., and R. Zitikis (2021), An axiomatic foundation for the Expected Shortfall. *Management Science* 67(3):1413–1429.

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

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