DISTORTION RISK MEASURES: PRUDENCE, COHERENCE, AND THE EXPECTED SHORTFALL

MASSIMILIANO AMARANTE

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

FELIX-BENEDIKT LIEBRICH

Workgroup Financial and Insurance Mathematics, Department of Mathematics,
University of Munich, Germany

Abstract. Distortion Risk Measures (DRM) are risk measures that are law invariant and comonotonic additive. The present paper is an extensive inquiry into this class of risk measures in light of new ideas such as qualitative robustness (Cont et al. 2010), prudence and no reward for concentration (Wang and Zitikis 2021), and tail relevance (Liu and Wang 2020). Results include several characterizations of prudent DRMs, a novel representation of coherent DRMs as well as an axiomatization of the Expected Shortfall alternative to the one recently provided by Wang and Zitikis. By linking the two axiomatizations, the paper provides a new perspective on the idea of no reward for concentration. The paper also contains results of independent interest concerning the extension of non-necessarily convex risk measures as well as the core of a general submodular distortion.

1. Introduction

Recent scholarly work showcases an increasing interest toward axiomatic studies of various classes of risk measures. No doubt, this trend has been strongly motivated by the value that such axiomatizations offer, to practitioners in general and to the regulatory debate in particular, in the form of an improved understanding of the scope of the risk measures in questions as well as of the implications associated with their use. For instance, Kou & Peng [27] and Liu & Wang [32] have given axiomatic characterizations of the VaR that unveil its elicitability properties, while He & Peng [23] have focused on the property of surplus invariance. We also refer to He et al. [22] for a review of many axiomatic studies of risk measures. In the wake of this type of work, it is no surprise that attention has shifted to the Expected Shortfall (ES), following its rise to the status of standard measure for market risk as sanctioned by the Basel Accords of 2016. Notably, Wang & Zitikis [40] have provided an axiomatization of the ES, which puts the use of this risk measure in a new perspective. In fact, in their work the ES is characterized by two entirely novel properties, which they term *prudence* and *No Reward for Concentration* (NRC).

E-mail addresses: massimiliano.amarante@umontreal.ca, liebrich@math.lmu.de.

Date: February 7, 2023.

The axiomatics of Wang & Zitikis is rather striking: despite the fact that the ES has long been known to enjoy mathematical properties such as subadditivity, comonotonic additivity and the so-called Lebesgue property, there is no direct axiomatic counterpart to these properties in their work. This is surprising because the ideas that there is value in hedging (subadditivity) and that there are no diversification benefits by combining risks that are increasing transformations of the same source of randomness (comonotonic additivity) are a foundational part of the *modus cogendi* about risk measurement and premium calculation; cf. Wang [41] and Wang et al. [42]. Obviously, the properties of subadditivity, comonotonic additivity and the Lebesgue property are implied by prudence and NRC along with the two other axioms imposed by Wang & Zitikis (law invariance and monotonicity). Yet, it is not immediately clear what each of the Wang & Zitikis' axioms contributes to the more traditional properties and how the combination of the axioms ends up producing them.

The present paper was borne out of our attempt both to understand how the new properties of prudence and NRC relate to the traditional categories employed in the theory of risk measurement and to obtain as a residual, so to speak, what characterizes the ES within these categories. Rapidly, however, the paper has outgrown its original motivation turning into an extensive inquiry into the class of law-invariant, comonotonic additive risk measures; a class that, for brevity, we will refer to as *distortion risk measures* (DRM) in the remainder. What prompted the broadening in our scope has been the recognition that the axiom of prudence has significant financial implications independent of the property of subadditivity. What is more, once these implications are understood in isolation, the combination of prudence and subadditivity pinpoints at once the special place occupied by the ES within the class of DRMs that are, in addition, coherent. This characterization of the ES is remarkably sharp and, thus, contributes substantially, in our opinion, to the debate about the use of the ES.

The structure of the paper reflects the development we have just outlined. Sections 3 and 4 contain two parallel developments: in the former, we study DRMs that are prudent but not necessarily coherent; in the latter, DRMs that are coherent but not necessarily prudent. In Section 3, we show that prudent DRMs are statistically well-behaved. In fact, in a sense weaker than in the original definition in [11], yet reasonable in the context of risk measurement, they are *qualitatively robust*: evaluating the risk of a given distribution by approximating it by means of a consistent estimator does not underestimate the true risk. This property, which is, in fact, equivalent to prudence, is also equivalent to the fact that prudent DRMs have the Fatou property and must dominate a Value-at-Risk benchmark (Theorem 3.1). In Theorem 3.1, we also observe that the qualitative robustness property (in the weaker sense) of prudent DRMs implies that these risk measures are *tail relevant* in the sense of Liu & Wang [32]. As tail relevant risk measures are characterized by limited possibilities to cross-subsidize losses with large speculative gains, our finding has transparent regulatory implications. Yet, to make it operational, one must determine which gains are to be deemed "speculative". To this end, we build an index which, in correspondence of each prudent DRM, identifies the gains that are considered speculative.

In Section 4, we study coherent DRMs. This class has received extensive attention having been identified as a primitive building block for all other law-invariant coherent risk measures (Kusuoka [29]). We are able, nonetheless, to prove a novel representation theorem (Theorem 4.2), where the characterizing properties of coherence and law-invariance appear in an extremely essential form. The returns of this new result are threefold. Firstly, Theorem 4.2 is instrumental to establishing the results of Section 5 where we combine prudence and coherence to obtain our characterization of the ES. Secondly, it casts a new light on both Kusuoka's theorem and all other representation results (notably, Föllmer & Schied [16, Sections 4.6–4.7] and Shapiro [37]) and, by doing so, opens up to novel interpretations. Roughly speaking (see Section 4) Theorem 4.2 allows one to regard the risk evaluation produced by a coherent DRM as the outcome of the interplay of two probabilities: a (finitely additive) subjective probability $\mathfrak{q}$ – which we call the *backbone* probability – and a statistical reference measure $\mathbb{P}$. Given these, a risk X is evaluated by means of the highest $\mathfrak{q}$-expected loss among the risks that have the same $\mathbb{P}$-distribution as X. Thirdly, Theorem 4.2 leads outright to generalizing/re-interpreting some classic results of mathematical finance. In subsection 4.2, we highlight two: one is a generalization of a well-known result on the anticore of submodular distortions (Carlier and Dana [8]); the other, is a characterization of spectral risk measures as the canonical extensions of those coherent DRMs, whose backbone is a countably additive probability.

In Section 5, by seamlessly combining the results of Section 3 and Section 4, we obtain the sought-after characterization of the ES: it is the minimal risk measure in the class of prudent coherent DRMs whose index (the one we define in Section 3) exceeds a certain threshold. Interestingly, this minimality property of the ES holds also in a much larger class, that of *exact distortions* (Proposition 5.4). The upshot is that, in comparison to a large set of alternatives, the ES is the *least conservative* prudent distortion risk measure; that is, the one that puts the least taxing capital requirements on financial institutions.

Notably, this conclusion leads to yet another characterization of prudence in the case of coherent distortions: a coherent DRM is prudent if and only if the backbone $\mathfrak{q}$ and the (statistical) reference measure $\mathbb{P}$ disagree on the possibility of events. In particular, there must be a $\mathbb{P}$-non-null event to which $\mathfrak{q}$ assigns probability zero. Intuitively, the role of such an event in the evaluation of a risk is that of swallowing large speculative gains. By further exploring this characterization, we then obtain (Corollary 5.5) a dual condition for a general convex law-invariant functional to be statistically well-behaved.

In Section 5, our inquiry into the class of DRMs yields as a by-product an axiomatization of the ES founded on traditional ideas such as the value of mergers, comonotonic additivity and, of course, prudence, for which we provide several alternative perspectives. Section 6, so to speak, closes the circle. We go back to Wang & Zitikis' axioms and re-examine them in light of what we have understood in the previous sections. After showing that a law-invariant, monotone functional satisfying NRC is necessarily a distortion risk measure (Corollary 6.4), we realize, with some surprise, that the No Reward for Concentration axiom plays a dual role to prudence. From this angle, one sees that the ES is the risk measure that requires the *maximal* capital buffers among all distortion risk measures satisfying NRC. In this sense, for DRMs, prudence turns out to be a much more conservative property than NRC.

Three appendices complete the exposition. The first digs into an aspect of a rather technical nature – that of the extension of risk measures to a larger domain – whose consideration is, nonetheless, necessary as Wang & Zitikis define prudence for risk measures with domain L^1 rather than L^∞. We show that each prudent DRM 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. Thus, in particular, there is no gap between our setting and that of Wang & Zitikis: A DRM on L^1 is uniquely determined by its values on L^∞. We should like to observe that our extension result is somewhat striking as, unlike parallel results in the literature (e.g., Filipović & Svindland [18], Liebrich & Svindland [31]), it obtains in the absence of subadditivity.

The second appendix focuses on the relation between *weak prudence* – a property that we introduce in Subsection 3.2 – and Wang& Zitikis prudence. The third appendix contains some auxiliary results that are omitted from the main text.

2. Preliminaries

This section is divided into four subsections. The first is entirely standard. For the largest part, its purpose is solely that of setting the notation that we will be using throughout the paper. The second and the fourth – devoted, respectively, to the Choquet integral and to finitely additive set functions – collect some basic definitions and facts that we will be using extensively in our proofs. The third subsection is especially important in the economy of the paper as it formally identifies DRMs, which are the class of risk measures within which our inquiry takes place.

2.1. **Risk measures.** The basic environment is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Throughout the paper, we will assume that $\mathbb{P}$ is nonatomic; equivalently, that there exists a random variable with a uniform distribution over $(0,1)$. The set of all such random variables is denoted by $\mathcal{U}$. As customary, 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 functions 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

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

Given $X, Y \in L^0$, we write $X \overset{d}{=} Y$ if $F_X = F_Y$; i.e., if the two random variables agree in distribution under $\mathbb{P}$.
Throughout the paper, we follow the actuarial convention that random variables model losses net of profits. Thus, positive random variables correspond to pure losses, negative ones to pure gains.
Let $\mathcal{X} \subset L^0$ be a subspace containing all constant random variables (which we identify with $\mathbb{R}$). A *risk measure* is any functional $\rho \colon \mathcal{X} \longrightarrow (-\infty, \infty]$ that is

(a) *proper*: $\rho(X) < \infty$ for some $X \in \mathcal{X}$.
(b) *monotone*: $X \leq Y$ $\mathbb{P}$-a.s. implies $\rho(X) \leq \rho(Y)$.
(c) *cash-additive*: 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 is

(d) positively homogeneous: $\rho(tX) = t\rho(X)$ for all $t \geq 0$.[1]
(e) subadditive: $\rho(X+Y) \leq \rho(X) + \rho(Y)$ for all $X, Y \in \mathcal{X}$.

Three risk measures that appear recurrently in the remainder of the paper are:

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

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

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

(ii) The *Expected Shortfall* at level $p \in [0,1]$, $\mathrm{ES}_p \colon L^1 \longrightarrow (-\infty, \infty]$, is 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) *Spectral risk measures*: Given a nonnegative and nondecreasing function ϕ on $[0,1]$ such that $\int_0^1 \phi(t)\mathrm{d}t = 1$ (ϕ is called a *spectrum*), the associated spectral risk measure is

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

Thus, $\rho\colon L^1 \longrightarrow (-\infty,\infty]$ and it is lower semicontinuous with respect to the L^1-norm (i.e., every norm-convergent sequence $(X_n) \subset L^1$ with limit X satisfies $\rho(X) \le \liminf_{n\to\infty} \rho(X_n)$; see Lemma C.1 in Appendix C). Notice that, for all $p \in [0,1)$, ES_p is a spectral risk measure corresponding to the spectrum $\phi = \frac{1}{1-p}\mathbf{1}_{(p,1)}$.

Two properties that are key in the study of risk measures are the Fatou property and the Lebesgue property. Let $(X_n) \subset L^\infty$ 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 $\rho(X) \le \liminf_{n\to\infty} \rho(X_n)$; it has the *Lebesgue property* if both ρ and $-\rho$ have the Fatou property. Notice that, under the actuarial convention above, the Lebesgue (Fatou) property that is usually stated (see [16]) as "continuity from below" ("above") now corresponds to "continuity from above" ("below").

2.2. **Capacities and the Choquet integral.** Two random variables X, Y are *comonotonic* if they are nondecreasing transformations, $X = f(Z)$ and $Y = g(Z)$, of one and the same random variable Z. A celebrated result of Schmeidler [35], extended in [33], asserts that, for a continuous functional $\varphi\colon L^\infty \longrightarrow \mathbb{R}$, the following two properties are equivalent:

(i) φ is *comonotonic additive*: X, Y comonotonic imply $\varphi(X+Y) = \varphi(X) + \varphi(Y)$.

(ii) φ is a *Choquet integral* w.r.t. the set function $v_\varphi\colon \mathcal{F} \longrightarrow \mathbb{R}$ defined by $v_\varphi(A) := \varphi(\mathbf{1}_A)$: i.e., for all $X \in L^\infty$,

$$\varphi(X) = \int X \, \mathrm{d}v_\varphi := \int_{[0,\infty)} v_\varphi(X > x)\, \mathrm{d}x + \int_{(-\infty,0)} [1 - v_\varphi(X > x)]\mathrm{d}x,$$

where the integrals in the last expression are Riemann integrals.

Notice that $v_\varphi(\varnothing) = 0$ and that φ is monotone in the $\mathbb{P}$-a.s. order if and only if v_φ is monotone w.r.t. set inclusion.

A set function that is monotone and satisfies $v(\varnothing) = 0$ is called a *capacity*. A capacity is *normalized* if $v(\Omega) = 1$. Most capacities in this paper are normalized and we will often omit mentioning normalization explicitly.

2.3. **Distortion Risk Measures.** A functional φ defined on a domain $\mathcal{D} \subset L^0$ is *law invariant* if the value $\varphi(X)$ only depends on the distribution of X under $\mathbb{P}$: For all $X, Y \in \mathcal{D}$,

$$X \overset{d}{=} Y \quad \Longrightarrow \quad f(X) = f(Y).$$

When $\mathcal{D} = L^\infty$ and the functional φ is comonotonic additive and continuous, the corresponding set function v_φ has the property that $v_\varphi(A) = v_\varphi(B)$ whenever $\mathbb{P}(A) = \mathbb{P}(B)$. Owing to the assumption of non-atomicity of $\mathbb{P}$, there exists a unique function $T_\varphi\colon [0,1] \longrightarrow \mathbb{R}$ such that $v_\varphi = T_\varphi \circ \mathbb{P}$; v_φ is said to be a *distortion* of $\mathbb{P}$ and T_φ is called the distortion function. We say that a distortion v_φ is:

(a) *continuous* if T_φ is continuous.
(b) *submodular* if, for all events $A, B \in \mathcal{F}$, $v_\varphi(A \cup B) + v_\varphi(A \cap B) \leq v_\varphi(A) + v_\varphi(B)$.
(c) *exact* if its *anticore*

$$\text{acore}(v_\varphi) := \{\mu \in \mathbf{ba} \mid \mu(\Omega) = v_\varphi(\Omega) \text{ and } \mu(A) \leq v_\varphi(A) \text{ for all } A \in \mathcal{F}\}$$

is nonempty and satisfies

$$v_\varphi(A) = \sup_{\mu \in \text{acore}(v_\varphi)} \mu(A), \quad A \in \mathcal{F}.$$

If v_φ is submodular, its anticore is necessarily nonempty [34].

A *distortion risk measure* (DRM) is a law-invariant risk measure $\rho \colon L^\infty \longrightarrow \mathbb{R}$ that is comonotonic additive. Notice that in this case the associated set function v_ρ is a normalized capacity and the distortion function T_ρ is nondecreasing.
DRMs that are, in addition, coherent give rise to exact distortions. In such a case, the hypograph of the distortion function T_ρ, $\mathbb{H}(T_\rho) := \{(x, y) \in [0, 1]^2 \mid y \leq T_\rho(x)\}$, is star shaped around $(0, 0)$ and $(1, 1)$ (Kadane & Wassermann [25] and Aouani & Chateauneuf [4], who show that the converse holds as well). A DRM is convex iff v_ρ is submodular iff T_ρ is concave. In the remainder, we will occasionally consider also supermodular capacities and their cores. Both concepts are defined by reversing the inequalities above. Statements about submodular capacities translates automatically into statements about supermodular capacities via the passage to the conjugate capacity, $\bar{v}(A) := v(\Omega) - v(A)$, $A \in \mathcal{F}$.

2.4. **Charges.** A *positive bounded charge* is a function $\mu : \mathcal{F} \longrightarrow \mathbb{R}$ which is finitely additive and such that $\mu(\varnothing) = 0$. Given a probability measure $\mathbb{P}$ on the measurable space $(\Omega, \mathcal{F})$, μ is said to be *absolutely continuous* with respect to $\mathbb{P}$ if every $\mathbb{P}$-null set is μ-null. We denote by $\mathbf{ba}_+(\mathbb{P}) = \mathbf{ba}_+$ the set of all such charges and by $\mathbf{ba} = \mathbf{ba}(\mathbb{P}) = \mathbf{ba}_+ - \mathbf{ba}_+$ the linear space they generate. The band of all signed measures in $\mathbf{ba}$ is denoted by $\mathbf{ca}(\mathbb{P}) = \mathbf{ca}$.
A positive charge $\mu \in \mathbf{ba}_+$ is a *pure charge* if it dominates no non-zero measure. Equivalently ([7, Theorem 10.3.3]), a positive charge $\mu \in \mathbf{ba}_+$ is a pure charge if and only if there is a vanishing sequence of events $(B_n) \subset \mathcal{F}$ such that, for all $n \in \mathbb{N}$, $\mu(B_n^c) = 0$.
Each $\mu \in \mathbf{ba}_+$ decomposes uniquely as the sum of a finite measure $\zeta \in \mathbf{ca}_+$ and a pure charge τ. For $\mu \in \mathbf{ba}_+$, we denote by $D_\mu := \frac{d\zeta}{d\mathbb{P}}$ the Radon-Nikodym derivative w.r.t. $\mathbb{P}$ of its countably additive part. Finally, we denote by $\Delta \subset \mathbf{ba}$ the set of all probability charges.

3. Prudent risk measures

In [40], Wang & Zitikis introduced a new concept, that of prudence of a risk measure. It is defined as follows. Let $\mathcal{D} \subset L^0$ and let $\rho \colon \mathcal{D} \longrightarrow \mathbb{R}$ be a functional:

Prudence (cf. [40]): ρ is prudent if, whenever $(X_n) \subset \mathcal{D}$ is a sequence converging $\mathbb{P}$-a.s. to $X \in \mathcal{D}$ and $\lim_{n\to\infty} \rho(X_n)$ exists in $[-\infty, \infty]$, then

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

[2] Strictly speaking, [40] 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, this difference vanishes as the interest in both papers lies solely in functionals with the property that two random variables agreeing $\mathbb{P}$-a.s. are mapped to the same value.

The scope of this section is to study the concept of prudence and its implications within the class of DRMs. The section's main result is Theorem 3.1 below, which gives several equivalent characterizations of prudence. We comment on Theorem 3.1 in subsections 3.1 and 3.2, where we highlight that prudent DRMs not only have remarkable statistical properties but also that they correspond to a specific regulatory policy. The section concludes with the introduction and the examination of our "index of non-triviality" which serves the purpose, among other things, to make operational our findings concerning the use of prudent DRMs for regulatory purposes.

Theorem 3.1. *For a DRM $\rho\colon L^\infty \longrightarrow \mathbb{R}$, the following are equivalent:*

1. *ρ is prudent.*
2. *ρ is l.s.c. with respect to convergence in distribution.*
3. *Whenever $(X_n) \subset L^\infty$ converges $\mathbb{P}$-a.s. to X and $\lim_{n\to\infty} \rho(X_n)$ exists in $\mathbb{R}$, then*
$$\rho(X) \le \lim_{n\to\infty} \rho(X_n).$$
4. *T_ρ is left-continuous and there is $0 < p < 1$ such that $T_\rho|_{[p,1]}$ is constant.*
5. *ρ has the Fatou property and there is $0 < p < 1$ such that $\rho \ge \mathrm{VaR}_p$.*

Proof. The equivalence of (1) and (2) is shown in Lemma C.2 in Appendix C. (1) clearly implies (3).
(3) implies (4): Let $(p_n) \subset (0,1)$ such that $p_n \uparrow p$. Select an increasing sequence $(A_n) \subset \mathcal{F}$ 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,
$$T_\rho(p) = \rho(\mathbf{1}_A) \le \lim_{n\to\infty} \rho(\mathbf{1}_{A_n}) = \lim_{n\to\infty} T_\rho(p_n),$$
which is left-continuity of T_ρ. 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) 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 (3).
(4) implies (5): As ρ is monotone, the Fatou property is equivalent to continuity from below, which in turn follows with the left-continuity of T_ρ and monotone convergence. Moreover, for $p \in (0,1)$ such that $T_\rho|_{[p,1]} \equiv 1$, we also have
$$T_\rho \ge \mathbf{1}_{(p,1]} = T_{\mathrm{VaR}_{1-p}}.$$
As such a relation between distortion functions transfers to the associated DRMs, it follows that $\mathrm{VaR}_{1-p} \le \rho$.
(5) implies (2): If (5) holds, then ρ is a *tail risk measure*, and one can show that there exists (see [32, Theorem F.1]) $q \in (0,1)$ such that, for all $X \in L^\infty$,
$$\rho(X) = \rho\big(X \vee \mathrm{VaR}_q(X)\big). \tag{3.1}$$
Let $(X_n) \subset L^\infty$ and assume that the sequence converges in distribution to X. Let $U \in \mathcal{U}$ be arbitrary and note that $X_n' := F_{X_n}^{-1}(U) \overset{d}{=} X_n$ converges to $X' = F_X^{-1}(U) \overset{d}{=} X$ 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 $\mathrm{VaR}_q(Y_n) = \mathrm{VaR}_q(Y_n')$, $n \in \mathbb{N}$. Hence,
$$\rho(X) = \rho(X') = \rho(X' \vee \mathrm{VaR}_q(X')) = \rho(Y' \vee \mathrm{VaR}_q(Y')) = \rho(Y').$$

Using (3.1) and the Fatou property for the first estimate, we observe:

$$\rho(X) = \rho(Y') \leq \liminf_{n\to\infty} \rho(Y_n') = \liminf_{n\to\infty} \rho(Y_n) \leq \liminf_{n\to\infty} \rho(X_n') = \liminf_{n\to\infty} \rho(X_n).$$

□

Before commenting on Theorem 3.1, it would be handy to state two of its immediate consequences. The first, which follows from item (4), has already been obtained, via a rather lengthy proof, in [40].

Corollary 3.2. *Both* VaR_p *and* ES_p*,* $0 < p \leq 1$*, are prudent.*

Proof. VaR_p is the DRM corresponding to the distortion function $T_{\mathrm{VaR}_p} = \mathbf{1}_{(1-p,1]}$, $0 < p \leq 1$. The latter is left-continuous and constant in a neighbourhood of 1. ES_1 is prudent as $\mathrm{ES}_1 = \mathrm{VaR}_1$. If $0 \leq p < 1$, ES_p is the DRM corresponding to the continuous distortion function T_{ES_p} defined by $T_{\mathrm{ES}_p}(x) = \frac{x}{1-p} \wedge 1$, which satisfies $T_{\mathrm{ES}_p}|_{[1-p,1]} \equiv 1$. □

The next corollary follows at once from item (5) in Theorem 3.1. Given a family of risk measures $\mathcal{R}$, let us say that a family $\mathcal{M} \subset \mathcal{R}$ is a *monotone subset* in $\mathcal{R}$ if for all $\rho_1, \rho_2 \in \mathcal{R}$,

$$\rho_1 \leq \rho_2 \text{ and } \rho_1 \in \mathcal{M} \quad \Longrightarrow \quad \rho_2 \in \mathcal{M},$$

where $\leq$ denotes the usual pointwise order.

Corollary 3.3. *The class of prudent DRMs is a monotone set in the class of all DRMs having the Fatou property.*

We will return to Corollary 3.3 in subsection 3.3, where we show (Corollary 3.7) that, in conjunction with Theorem 3.1 and the concepts introduced in subsection 3.3, it characterizes the VaR family within the class of prudent DRMs.

Remark 3.4. In Theorem 3.1, we considered risk measures with domain L^∞. This might generate the erroneous impression that there is a gap between our analysis and that of [40] as the latter considers risk measures with domain L^1. In Appendix A, we show that this difference is immaterial as prudent DRMs on L^∞ have a unique prudent extension to L^0. Because of this, in the remainder – with the few exceptions that will be highlighted – we will continue to consider DRMs with domain L^∞. For the reader interested in the general problem of extending risk measures defined on L^∞ to larger domains, we would like to signal that Appendix A contains both new results and new methods as, unlike the existing literature, no convexity assumption is made therein.

3.1. **Qualitative robustness, tail relevance and the Fatou property.** Theorem 3.1 provides us with several perspectives to evaluate the property of prudence. It shows that a simple, yet rigid, geometric property (item (4)) corresponds to desirable statistical properties (item (2)) which, in turn, correspond to clear regulatory prescriptions (item (5)). More in detail, item (2) – l.s.c. with respect to convergence in distribution – allows one to evaluate prudent DRMs from the viewpoint of *qualitative robustness*. The study of qualitative robustness of a functional defined on a set of *probability distributions* over the reals was initiated by [21] and has entered the theory of risk measures with the seminal paper [11]. Using the fact that any law-invariant risk measure induces an estimator (on distributions), [11] deems a risk measure qualitatively robust if small deviations in the laws of the sample leads only to small deviations in the laws of the associated estimators, provided the size of the data set is large enough. Notably, [11] show that under the consistency of the historical estimator for the risk measure, this property is equivalent to the continuity of the risk measure with respect to convergence in distribution. Thus, in view of the strong consistency of historical estimators, establishing the qualitative robustness of a risk measure is tantamount to establishing its continuity with respect to convergence in distribution. Our result on prudent DRMs delivers l.s.c. rather than continuity. On the one hand, the property of l.s.c. is still conceptually faithful to the idea of qualitative robustness; on the other hand, being weaker than continuity, it accommodates more risk measures without imposing additional requirements on the structure such as those introduced in [26, 28]. By following the same logic as in [11], the l.s.c. of prudent DRMs practically means that the risk evaluation obtained via the associated consistent estimator does not underestimate the true risk. The combination of items (2) and (4) in Theorem 3.1 highlights that the constancy of the distortion function in a neighborhood of 1 is a critical feature for the qualitative robustness, in our weaker sense, of prudent DRMs. This finding echoes one in [39, Corollary 1], where the same condition appears as part of a set of sufficient conditions guaranteeing the continuity with respect to convergence in distribution of law-invariant, signed Choquet integrals.

Item (5) in Theorem 3.1 shows that prudence is a sort of "boosted" Fatou property. As shown in [32], who introduced the notion of *tail relevance*, this boosted Fatou property implies that DRMs are tail relevant. Below (see Proposition 3.8 and subsequent observation), we will complement [32] by showing that tail relevant risk measures are characterized by limited possibilities to cross-subsidize losses with large, but speculative gains; i.e., gains that are not sufficiently likely under $\mathbb{P}$. In view of the equivalence of items (5) and (2) as well as of the discussion above, it is noteworthy that a statistical property such as qualitative robustness in our weaker sense implies a property with a clear regulatory bearing such as tail relevance.

3.2. **Weak prudence.** A generic functional (not necessarily a DRM) $\varphi\colon L^0 \longrightarrow \mathbb{R}$ with the prudence property obviously satisfies the property in item (3) of Theorem 3.1. The converse may not be true as we may have a $\mathbb{P}$-a.s. convergent sequence $(X_n) \subset \mathcal{D}$ with $\lim_{n\to\infty} \rho(X_n) = -\infty$. That is, generally speaking, the property in item (3) of Theorem 3.1 is weaker than prudence. We shall refer to it as *weak prudence*. Formally, given a domain of definition $\mathcal{D} \subset L^0$ and a functional $\varphi\colon L^0 \longrightarrow \mathbb{R}$, φ is weakly prudent if, for every $\mathbb{P}$-a.s. convergent sequence $(X_n) \subset \mathcal{D}$ with limit $X \in \mathcal{D}$,

$$\lim_{n\to\infty} \varphi(X_n) \text{ exists in } \mathbb{R} \quad \implies \quad \varphi(X) \le \lim_{n\to\infty} \varphi(X_n).$$

Both weak prudence and prudence reflect the intuition that, if loss profiles $(X_n) \subset \mathcal{D}$ are used to approximate some more complex X, then the limit risk ℓ of the risks (X_n) is a safe proxy for the risk $\varphi(X)$ of the limit in the sense that $\varphi(X) \le \ell$.

In [40], there is no distinction between weak prudence and prudence. This is not surprising as the focus of [40] is on monotone functionals ρ measuring financial risk. In this setting, the assumption that a $\mathbb{P}$-a.s. convergent sequence $(X_n) \subset \mathcal{D}$ should not lead to a sequence of risks $(\rho(X_n))$ that diverges to $-\infty$ is sensible. For otherwise, one would allow for sequences featuring larger and larger speculative gains in ever more improbable fortunate scenarios that would outweigh any risk. Outside this setting, however, it is possible that prudence might impose unnecessary demands compared to weak prudence. Because of this, in Appendix B we give sufficient conditions for the equivalence (beyond DRMs) of weak prudence and prudence. Regrettably, we have been unable to answer the question of whether or not weak prudence and prudence/lower semicontinuity with respect to convergence in distribution are generally equivalent properties of law-invariant functionals.

3.3. **The index of nontriviality.** In subsection 3.1, we have anticipated that measuring risk with a prudent DRM limits the possibilities to cross-subsidize losses with large, speculative gains. We now quantify precisely the meaning of "speculative". To this end, we define the *index of nontriviality* of a law-invariant risk measures 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, let us cast ourselves in a behavioral framework and interpret the value $\rho(X)$ as the agent's willingness to pay for risk X. As we identify the arguments of ρ as net losses, 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.
The next proposition shows the index $\mathfrak{nt}(\cdot)$ takes a very special, yet significant, form in the case of DRMs and that it can be used to characterize prudence.

Proposition 3.6. *Let $\rho\colon L^\infty \longrightarrow \mathbb{R}$ be a positively homogeneous law-invariant risk measure.*

(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*
$$\begin{aligned}\mathfrak{nt}(\rho) &= 1 - \inf\{p \in [0,1] \mid T_\rho(p) = 1\}\\ &= \sup\{p \in (0,1) \mid \rho \geq \mathrm{VaR}_p\}.\end{aligned}$$
(3) *If ρ is a DRM with the 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, note that $T_\rho(p) = 1$ holds only if $T_\rho \geq \mathbf{1}_{(q,1]} = T_{\mathrm{VaR}_{1-q}}$ for all $p < q \leq 1$, which in turn means that $T_\rho(q) = 1$ for all $q \in (p, 1]$.
(3) T_ρ is left-continuous by virtue of the Fatou property. By (2) and Theorem 3.1(3), ρ is prudent if and only if $\mathfrak{nt}(\rho) > 0$.

□

By means of the index of nontriviality, we can now unveil what is hidden in the conjunction of Theorem 3.1(3) and Corollary 3.3: the Value-at-Risk family is minimal in the set of prudent DRMs.

Corollary 3.7. *Any prudent DRM ρ satisfies $\rho \geq \mathrm{VaR}_{\mathfrak{nt}(\rho)}$. In particular, for $p \in (0, 1]$, VaR_p is the minimal prudent DRM ρ satisfying $\mathfrak{nt}(\rho) \geq p$.*

We now show that the index of nontriviality quantifies what type of gains are speculative. For each $X \in L^\infty$ and $Y \in -L^\infty_+$ large enough, the financial position $X - Y\mathbf{1}_A$ improves X in that an additional gain of $Y - X$ is made if event A obtains. If the probability of A is too small and betting on this event is too speculative, this should not alter the risk assessment; i.e., $\rho(X) = \rho(X - Y\mathbf{1}_A)$ should hold true. The threshold for assessing if a probability is too small is precisely $\mathfrak{nt}(\rho)$.

Proposition 3.8. *Suppose $\rho\colon L^\infty \longrightarrow \mathbb{R}$ is a DRM. Fix $X \in L^\infty$ arbitrary. Then,*

$$\mathfrak{nt}(\rho) := \inf\{\mathbb{P}(A) \mid A \in \mathcal{F}, \exists Y \in L^\infty_+ : \ \rho(X - Y\mathbf{1}_A) < \rho(X)\}. \tag{3.2}$$

Proof. By ι we abbreviate the right-hand expression in (3.2) and show first that $\mathfrak{nt}(\rho) \geq \iota$. This is clear if $\mathfrak{nt}(\rho) = 1$. Else, Proposition 3.6(2) implies that we can find $A \in \mathcal{F}$ such that $\mathbb{P}(A) \geq \mathfrak{nt}(\rho)$ and such that $T_\rho(\mathbb{P}(A^c)) < 1$. Set $m := \|X\|_\infty$ and fix $k > 2m$. In particular, $Y := k + X \in L^\infty_+$ and $X - Y\mathbf{1}_A = X - k\mathbf{1}_A$. Using comonotonic additivity,

$$\rho(X - k\mathbf{1}_A) = \rho\big((X - k\mathbf{1}_A) \wedge (-m)\big) + \rho\big((X - k\mathbf{1}_A + m)^+\big)$$

We first compute

$$\rho\big((X - k\mathbf{1}_A) \wedge (-m)\big) = \rho(-m - k\mathbf{1}_A) = -k + (k - m)T_\rho(\mathbb{P}(A^c)) < -m.$$

Also,

$$\begin{aligned}\rho\big((X - k\mathbf{1}_A + m)^+\big) &= \rho\big((X + m)^+\mathbf{1}_{A^c} + (X + m - k)^+\mathbf{1}_{A^c}\big) \\ &= \rho\big((X + m)^+\mathbf{1}_{A^c}\big) = \rho((X + m)\mathbf{1}_{A^c}) \\ &\leq \rho(X + m) = \rho(X) + m\end{aligned}$$

In sum, $\rho(X - k\mathbf{1}_A) < \rho(X)$, and $\iota \leq \mathbb{P}(A)$. Letting $\mathbb{P}(A) \downarrow \mathfrak{nt}(\rho)$ yields the claim.
We now turn to the fact that $\mathfrak{nt}(\rho) \leq \iota$. This is clear if $\mathfrak{nt}(\rho) = 0$. Else, let $p > \iota$, $X \in L^\infty$, and select a lower tail event $A \in \mathcal{F}$ for X with probability p.[3] For $k > \|X\|_\infty$ consider $X - k\mathbf{1}_A$ and note that, for a suitable constant $\gamma > k$,

$$\mathbb{P}(X - k\mathbf{1}_A > x) = \mathbf{1}_{(-\infty,k)} + \mathbb{P}(A^c)\mathbf{1}_{[-k,\gamma)}(x) + \mathbb{P}(X > x)\mathbf{1}_{[\gamma,\infty)}(x), \quad x \in \mathbb{R}.$$

Shifting by a constant, we can assume $\gamma = 0$. Moreover,

$$\rho(X - k\mathbf{1}_A) = \int_{[-k,0)} [1 - T_\rho(\mathbb{P}(A^c))]\mathrm{d}x + \int_{[0,\infty)} T_\rho(\mathbb{P}(X > x)\mathrm{d}x. \tag{3.3}$$

Now, as $p > \iota$, we can select k large enough to guarantee that

$$\int_{[-k,0)} [1 - T_\rho(\mathbb{P}(X > x)]\mathrm{d}x + \int_{[0,\infty)} T_\rho(\mathbb{P}(X > x)\mathrm{d}x = \rho(X) < \rho(X - k\mathbf{1}_A). \tag{3.4}$$

By comparing (3.3) and (3.4), we see that we cannot have $T_\rho(\mathbb{P}(A^c)) = 1$. Thus, $T_\rho(\mathbb{P}(A^c)) < 1$ and $p = \mathbb{P}(A) > \mathfrak{nt}(\rho)$ by Proposition 3.6(2). Letting $p \downarrow \iota$ completes the proof. □

Let us recall that a tail-relevant risk measure evaluates a risk X by ignoring parts of the gains tail of its distribution. We have observed in (3.1) in the proof of Theorem 3.1 that this part is characterized precisely by the threshold $\mathrm{VaR}_p(X)$. Proposition 3.8 makes this procedure even more precise by proving that, equivalently, the risk of X is not altered if sufficiently unlikely speculative gains are added.

[3] I.e., there are constants $\alpha, \beta \in \mathbb{R}$, $\alpha \leq \beta$, such that $X\mathbf{1}_A \leq \alpha\mathbf{1}_A$ and $X\mathbf{1}_{A^c} \geq \beta\mathbf{1}_{A^c}$ $\mathbb{P}$-a.s.

4. Coherent DRMs

The previous section concerned a new class of DRMs identified by the newly introduced concept of prudence. In contrast, the class of DRMs we consider in this section has been at the core of mathematical finance for more than two decades. It is the class of coherent DRMs; that is, risk measures that are coherent, law invariant and comonotonic additive. Fundamental results for this class have been established by, e.g., Kusuoka [29] and Shapiro [37]. We also refer to the state-of-the-art presentation in Föllmer & Schied [16, Chapters 4.6–4.7]. What is more, the class of coherent DRMs is home to the spectral risk measures of Acerbi [1] and, therefore, to the ES.
In this section, we complement the existing literature by giving a novel representation for coherent DRMs, which differs from previous ones. In a nutshell, the risk evaluation procedure is the outcome of the interplay of a finitely additive probability $\mathfrak{q}$ and a statistical probability measure $\mathbb{P}$. The risk of a prospect X is the highest $\mathfrak{q}$-expected loss among the prospects that have the same $\mathbb{P}$-distribution (and therefore the same statistical properties) as X. As we shall see explicitly in the next section, $\mathfrak{q}$ may even discard gains that have strictly positive probability under $\mathbb{P}$. While this representation is interesting in its own right, it also plays a key role in the subsequent developments leading to our characterization of the ES.

4.1. **A novel representation of coherent DRMs.** Let us recall that a $\mathbb{P}$-invariant DRM ρ is represented by a Choquet integral with respect to a distortion $T\circ\mathbb{P}$. When ρ is coherent, its distortion function T is concave on $[0,1]$. Hence, its derivative T' exists a.e. on $(0,1)$. The following lemma obtains by readily combining Theorems 4.93 and 4.70 in [16].

Lemma 4.1. *Let $\rho\colon L^\infty \longrightarrow \mathbb{R}$ be a risk measure. Then the following are equivalent:*

1. *ρ is a coherent DRM.*
2. *For $T=T_\rho$ and all $X\in L^\infty$,*
$$\rho(X) = T(0+)F_X^{-1}(1-) + \int_0^1 F_X^{-1}(t)T'(1-t)\mathrm{d}t, \quad X\in L^\infty, \tag{4.1}$$
where $T(0+) := \inf_{x\in(0,1]} T(x)$.

The section's main result is the following theorem. For a probability charge $\mathfrak{q}\in\Delta$, let $\psi_{\mathfrak{q}}\colon L^\infty \longrightarrow \mathbb{R}$ be the functional
$$\psi_{\mathfrak{q}}(X) = \sup_{X' \overset{d}{=} X} \int X' \,\mathrm{d}\mathfrak{q}. \tag{4.2}$$

Theorem 4.2. *For a risk measure $\rho\colon L^\infty \longrightarrow \mathbb{R}$, the following are equivalent:*

1. *ρ is a coherent DRM.*
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 $D_{\mathfrak{q}} \overset{d}{=} D_{\mathfrak{r}}$.

Proof. (1) implies (2): By Lemma 4.1, ρ has the representation (4.1). As $T=T_\rho$ 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. As the underlying probability space is atomless, we can select a random variable D whose quantile function satisfies $F_D^{-1}(t) = T'(1-t)$ 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 $D\in L^1_+$. Thus, D 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^∞) such that $\xi(\Omega) = T(0+)$. By setting $\mathfrak{q} := \lambda + \xi$ and applying [13, Proposition 3.9], for any $X \in L^\infty$, we have

$$\begin{aligned} \sup_{X' \overset{d}{=} X} \int X' \,\mathrm{d}\mathfrak{q} &= T(0+)F_X^{-1}(1-) + \int_0^1 F_X^{-1}(t)F_D^{-1}(t)\mathrm{d}t \\ &= T(0+)F_X^{-1}(1-) + \int_0^1 F_X^{-1}(t)T'(1-t)\mathrm{d}t = \rho(X). \end{aligned} \tag{4.3}$$

(2) implies (1): Let $\mathfrak{q} \in \Delta$ be a probability charge and define a map ρ on L^∞ as in (4.2). By the Yosida-Hewitt Theorem, $\mathfrak{q}$ decomposes uniquely as the sum of a measure λ and a pure charge ξ. From [13, Proposition 3.9], we obtain that, for all $X \in L^\infty$,

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

Comonotonic additivity of the right-hand expression readily follows with [16, Lemma 4.90]. Monotonicity is due to the quantile function possessing these properties. For subadditivity, we use [16, Lemma A.32] to see that, for all $X, Y \in L^\infty$,

$$\begin{aligned} \rho(X) + \rho(Y) &= \sup_{U,U' \in \mathcal{U}} \int \left[F_X^{-1}(U) + F_Y^{-1}(U')\right]\mathrm{d}\mathfrak{q} \\ &\geq \sup_{U \in \mathcal{U}} \int \left[F_X^{-1}(U) + F_Y^{-1}(U)\right]\mathrm{d}\mathfrak{q} = \sup_{U \in \mathcal{U}} \int F_{X+Y}^{-1}(U)\,\mathrm{d}\mathfrak{q} \\ &= \rho(X+Y). \end{aligned}$$

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 $\xi, \tau \in \mathbf{ca}_+^{\mathrm{d}}$ be the pure charges in the Yosida-Hewitt decomposition of $\mathfrak{q}$ and $\mathfrak{r}$ and let $(A_n) \subset \mathcal{F}$ 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, for all $p \in (0,1]$ and $A \in \mathcal{F}$ with $\mathbb{P}(A) = p$, we have that

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

By [36, Theorem 3.A.5], this implies equivalence between $D_\mathfrak{q}$ and $D_\mathfrak{r}$ in the so-called convex order.[4] This is the case if and only if $D_\mathfrak{q} \overset{d}{=} D_\mathfrak{r}$ ([36, Theorem 3.A.43]).
Conversely, if $D_\mathfrak{q} \overset{d}{=} D_\mathfrak{r}$, we first have $\delta := 1 - \mathbb{E}[D_\mathfrak{q}] = 1 - \mathbb{E}[D_\mathfrak{r}]$. Next, [13, Proposition 3.9] yields that for any $X \in L^\infty$

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

This concludes the proof. □

As probability charges like $\mathfrak{q}$ will play a substantial role in what follows, it is worth naming them.

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

Definition 4.3. Given a coherent DRM $\rho\colon L^\infty \longrightarrow \mathbb{R}$, we call a probability charge $\mathfrak{q}$ satisfying $\rho = \psi_{\mathfrak{q}}$ a *backbone* of ρ.

Let us observe that, in the first part of the proof of Theorem 4.2, the pure charge ξ yielding the decomposition of $\mathfrak{q}$ is chosen completely freely up to normalization. Thus, backbone probabilities are never unique. Nonetheless, as shown in the final part of the proof, whether or not a particular probability charge $\mathfrak{q} \in \Delta$ is a backbone of a given coherent DRM ρ depends entirely on the distribution of $D_{\mathfrak{q}}$. In fact, the argument employed in the proof shows a finer property. Let $\preceq_{cx}$ denote the convex order on L^1 and suppose that two charges $\mathfrak{q}, \mathfrak{r} \in \Delta$ satisfy $D_{\mathfrak{q}} \preceq_{cx} D_{\mathfrak{r}}$. Then $\mathbb{E}[D_{\mathfrak{q}}] = \mathbb{E}[D_{\mathfrak{r}}]$ and $\psi_{\mathfrak{q}} \le \psi_{\mathfrak{r}}$.
Even if it might appear obscure at this point, it is probably worth hinting why Theorem 4.2 is going to play a crucial role in the remainder, when we couple coherence and prudence. Mechanically, in the case a of a coherent DRM, a backbone probability is precisely what is going to determine which gains are to be deemed speculative. In terms of interpretation, however, the order is somewhat reversed as a backbone probability can be thought of as the outcome of the following process. Given a statistical reference measure $\mathbb{P}$, one assesses subjectively which gains are to be considered speculative. These are filtered out, in the sense that they won't be taken into account in the evaluation of the risk. The result is the set of backbone probabilities. Evaluation by means of a backbone probability, however, violates $\mathbb{P}$-law invariance. This is restored by assigning the same evaluation to all risks that have the same $\mathbb{P}$-distribution. In conformity to the requirement of subadditivity, this evaluation is the one corresponding to the highest risk.

4.2. **Implications of Theorem 4.2.** We now highlight two notable implications of Theorem 4.2. The first concerns the anticore (and, by passage to the dual capacity, the core) of a submodular (supermodular) distortion. The second concerns the family of spectral risk measures.

4.2.1. *The anticore of general submodular distortions.* While the study of the core/anticore of a capacity is mostly instrumental in the study of risk measures, it is a subject in its own right in the theory of cooperative games, where the core constitutes one of the most important solution concepts. A well-known result of [8] provides a characterization of the core of a differentiable, supermodular distortion. Corollary 4.4 below, generalizes this characterization by removing altogether the assumption of differentiability. This generalization is substantial owing to the importance, both in theory and applications, of submodular (supermodular) distortions that are not continuous at $\varnothing$ (resp. Ω).

Corollary 4.4. *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*

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

Then

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

Proof. Consider the distortion functions $T_1 := (T - T(0+))\mathbf{1}_{(0,1]}$ and $T_2 := T(0+)\mathbf{1}_{(0,1]}$. Both functions are concave; hence, both distortions $v_i := T_i \circ \mathbb{P}$ are submodular. As the mapping $v \mapsto \operatorname{acore}(v)$ is additive on the cone of submodular capacities (cf. [12]), it suffices to identify $\operatorname{acore}(v_i)$, $i = 1, 2$. One easily sees that $\operatorname{acore}(v_2) = T(0+)\Delta$. By continuity of T_1 and [33, Proposition 4.4], $\operatorname{acore}(v_1) \subset \mathbf{ca}_+$. By using the proof of Theorem 4.2, one identifies $\operatorname{acore}(v_1)$ with all measures $\zeta \in \mathbf{ca}_+$ such that

$\frac{\mathrm{d}\zeta}{\mathrm{d}\mathbb{P}} \preceq T'(1-U^*)$ for some $U^* \in \mathcal{U}$. By using [6, Lemma 3.5] in the first and [16, Lemma A.32] in the second identity, it follows that

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

□

When a submodular distortion is continuous at $\varnothing$ or even differentiable, Corollary 4.4 yields that $\mathrm{acore}(v) = \mathcal{D}_v$, which is the representation of [8]. In the general case, one adds the part $T(0^+)\Delta$. The extreme points of Δ, while evidently not Dirac measures, behave quite like Dirac measures. They are $\{0,1\}$-valued pure charges and correspond to the multiplicative linear functionals on L^∞. Thus, by appealing to the Krein-Milman Theorem, one sees that Corollary 4.4 establishes that the anticore of a submodular distortion is the closed convex hull of its differentiable part and the closed convex hull of the set of positive, normalized multiplicative linear functionals.

4.2.2. *Spectral risk measures.* As it is well-known, spectral risk measures are an important class of risk measures which, among other things, is home to the ES. They are defined via formula (2.2) of Section 2) and are subadditive by Lemma C.1 in Appendix C. In view of the proof of Theorem 4.2, one easily sees that a spectral risk measure restricted to L^∞ is a coherent DRM, which additionally satisfies the Lebesgue property; cf. Section 2. Proposition 4.5 below follows readily from Lemma 4.1 and Theorem 4.2. The equivalence between (2) and (3) therein yields a novel representation for spectral risk measures on L^1, which shows, in particular, that spectral risk measures (restricted to L^∞) are precisely those coherent DRMs whose backbone probabilities are measures; i.e., countably additive charges. As measures are continuous at $\varnothing$, it follows at once that the converse to the observation above is also true: the Lebesgue property is precisely what characterizes spectral risk measures among coherent DRMs. The equivalence between (1) and (2) in Proposition 4.5 improves upon an earlier result of Shapiro [37], who established it for *finite* risk measures on a standard measure space (both assumptions are dispensed with below). It should be noted that the domain of the risk measures in Proposition 4.5 is L^1, as required by Acerbi's definition.

Proposition 4.5. *For a proper functional $\rho\colon L^1 \longrightarrow (-\infty,\infty]$, the following are equivalent:*

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

$$\rho(X) = \sup\{\mathbb{E}_{\mathbb{Q}}[X] \mid X' \overset{d}{=} X,\ \mathbb{E}_{\mathbb{Q}}[X'] \text{ well defined}\}. \tag{4.4}$$

Proof. (3) implies (2): Assume that ρ is defined by (4.4) and define $\phi\colon (0,1) \longrightarrow \mathbb{R}_+$ by

$$\phi = F^{-1}_{\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}}.$$

Then, for all $X \in L^1$, [16, Theorem A.28] yields $\rho(X) = \int_0^1 \phi(t) F_X^{-1}(t)\mathrm{d}t$, which means that ρ is a spectral risk measure with spectrum ϕ.
(2) implies (1): A spectral risk measure is proper, l.s.c., subadditive (Lemma C.1), law invariant (by definition), comonotonic additive ([16, 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 4.2, there is a probability charge $\mathfrak{q} \in \mathbf{ba}_+$ such that $\rho^\flat = \psi_{\mathfrak{q}}$. For every vanishing sequence $(A_n) \subset \mathcal{F}$, we have $\limsup_{n\to\infty} \mathfrak{q}(A_n) \le \lim_{n\to\infty} \rho^\flat(\mathbf{1}_{A_n}) = 0$, i.e., $\mathfrak{q}$ is a countably additive probability measure $\mathbb{Q}$. By the implication (3) $\implies$ (2) and Lemma C.1, the extension ρ' of $\rho^\flat$ given by (4.4) is l.s.c., law invariant and coherent. Since, by [18], a coherent law-invariant risk measure on L^∞ has a unique l.s.c., law-invariant and convex extension to L^1, we conclude that $\rho = \rho'$. □

5. Linking prudence and coherence; the Expected Shortfall

In this section, we gradually merge the classes of prudent DRMs and of coherent DRMs. The final result is a characterization of the ES, which offers a wider perspective than those stemming from earlier results of [16, Section 4.5] and [32, Theorem 2]. We begin by introducing the class of *exact* DRMs, which is defined as follows:

Definition 5.1. A risk measure is an exact DRM if it is a DRM and if the associated capacity is exact.

Let us recall that a capacity is exact (see Section 2) if it is the upper envelope of its anticore. Coherent DRMs are exact DRMs, for a coherent DRMs is represented by a convex Choquet integral and this, by [35], is associated to an exact capacity. The converse, however, is not true as the Choquet integral with respect to an exact capacity need not be convex even when the said capacity is an upper envelope. Thus, coherent DRMs are a strict subset of exact DRMs.

Our first result shows that for exact DRMs the index of non-triviality is indeed the maximal $\mathbb{P}$-probability of the null events of the elements in its acore. When specialized to the subset of coherent DRMs, the result makes explicit a hint we gave in Section 4: A backbone probability of a coherent DRM is what determines which gains are to be deemed speculative. Consequently, the identification of the backbone probabilities of a coherent DRM is a key step for evaluating its use for regulatory purposes. Notice that formula (5.2) below is independent of the concrete choice of the backbone probability.

Lemma 5.2. *Let $\rho\colon L^\infty \longrightarrow \mathbb{R}$ be an exact DRM. Then*

$$\mathfrak{nt}(\rho) = \sup\{\mathbb{P}(N) \mid N \in \mathcal{F} \text{ and } \mathfrak{q}(N) = 0 \text{ for some } \mathfrak{q} \in \operatorname{acore}(v)\}. \tag{5.1}$$

If ρ is a coherent DRM with backbone $\mathfrak{r}$, we have

$$\mathfrak{nt}(\psi_{\mathfrak{r}}) = \sup\{\mathbb{P}(N) \mid N \in \mathcal{F} \text{ and } \mathfrak{r}(N) = 0\}. \tag{5.2}$$

Proof. Let $\mathfrak{q} \in \operatorname{acore}(v)$ and $N \in \mathcal{F}$ be such that $\mathfrak{q}(N) = 0$. Then

$$1 = \mathfrak{q}(N^c) \le (T_\rho \circ \mathbb{P})(N^c) = T_\rho(1 - \mathbb{P}(N)).$$

Hence, $1 - \mathfrak{nt}(\rho) \le 1 - \mathbb{P}(N)$ or, equivalently, $\mathbb{P}(N) \le \mathfrak{nt}(\rho)$. This shows the estimate "$\ge$" in both (5.1) and (5.2) and equality if $\mathfrak{nt}(\rho) = 0$.

In order to show the converse inequality in (5.1), assume $\mathfrak{nt}(\rho) > 0$ recall from Proposition 3.6(2) that

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

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

$$\max_{\mathfrak{q}\in\operatorname{acore}(v)} \mathfrak{q}(N^c) = (T_\rho \circ \mathbb{P})(N^c) = T_\rho(1-x) = 1,$$

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

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

As this estimate holds for all $x < \mathfrak{nt}(\rho)$, we obtain the same bound for the index of nontriviality. This shows the desired formula (5.1) for exact DRMs.
Now, suppose that ρ is coherent, that $\mathfrak{nt}(\rho) > 0$, and that $\mathfrak{r}$ is a backbone of $\rho = \psi_{\mathfrak{r}}$. One readily verifies that $T'_\rho|_{(1-\mathfrak{nt}(\rho),1)} \equiv 0$. Let $U \in \mathcal{U}$ be a random variable such that $D_{\mathfrak{r}} = F^{-1}_{D_{\mathfrak{r}}}(U) = T'_\rho(1-U)$ ([16, Lemma A.32]). We can find a vanishing sequence $(B_n) \subset \mathcal{F}$ of events such that the pure charge μ in the Yosida-Hewitt decomposition of $\mathfrak{r}$ satisfies $\mu(B_n^c) = 0$, $n \in \mathbb{N}$. Then

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

Thus,

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

This concludes the proof of (5.2). □

Our next result complements those in Section 3 by giving an additional characterization of prudence within the class of exact DRMs. In particular, the result makes it explicit that, by adding prudence to coherence one further "robustifies" risk assessments: By item (4) in Proposition 5.3 below, a prudent coherent DRM always discards gains that have strictly positive probability under the statistical reference measure $\mathbb{P}$.

Proposition 5.3. *Let $\rho\colon L^\infty \longrightarrow \mathbb{R}$ be an exact DRM with associated capacity $v := T_\rho \circ \mathbb{P}$. 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. By Lemma C.4 in Appendix C, T_ρ is left-continuous. Thus, by Proposition 3.6, ρ is prudent if and only if $\mathfrak{nt}(\rho) > 0$. Together with (5.1), this immediately shows that (1) and (2) are equivalent, implied by (3), and equivalent to (4) if ρ is coherent. It therefore suffices to show that (2) implies (3). To this end, 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 5.4, 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 $\mathbb{P}(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \frac{1}{1-p}) = 1 - \mathbb{P}(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = 0)$ (and thereby necessarily also satisfying $\mathbb{Q} \not\approx \mathbb{P}$), we have

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

that is, $\mathbb{Q} \in \mathrm{acore}(v)$. □

As anticipated, the final result of this section concerns the ES. In [32, Theorem 2], it was shown that ES_p is the minimal coherent p-tail risk measure. Proposition 5.4 below shows that, in fact, such minimality property of the ES family holds in the wide class of *prudent* exact DRMs. As observed above, members of this class need not be coherent. For the purpose of stating Proposition 5.4, we introduce the following notation. By $\mathcal{R}^e(p)$, we denote the set of prudent, exact DRMs satisfying $\mathfrak{nt}(\rho) \geq p$. We also denote by $\mathcal{R}^c(p) \subset \mathcal{R}^e(p)$ the subset of prudent coherent DRMs and by $\mathcal{R}^s(p) \subset \mathcal{R}^e(p)$ the subset of spectral risk measures.

Proposition 5.4. *Suppose ρ is a DRM on L^∞.*

(1) *If ρ is exact, then ρ is prudent if and only if $\rho \geq \mathrm{ES}_p$ for some $0 < p \leq 1$. In this case, we also have $\rho \geq \mathrm{ES}_{\mathfrak{nt}(\rho)}$.*
(2) *For any $0 < p \leq 1$, ES_p is the minimal risk measure in any of the sets $\mathcal{R}^e(p)$, $\mathcal{R}^c(p)$, and $\mathcal{R}^s(p)$.*

Proof. We shall prove only (1) as (2) follows automatically from it. If ρ is a prudent DRM, then – by Proposition 3.6 – $\mathfrak{nt}(\rho) > 0$ and, for all $q > 1 - \mathfrak{nt}(\rho)$, $T_\rho(q) = 1$. As the hypograph $\mathbb{H}(T_\rho) := \{(x, y) \in [0, 1]^2 \mid y \leq T_\rho(x)\}$ is star shaped around $(0, 0)$ and contains the point $(q, 1)$,

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

which implies $\rho \geq \mathrm{ES}_{1-q}$. By letting $q \downarrow 1 - \mathfrak{nt}(\rho)$ and by using continuity of $\alpha \mapsto \mathrm{ES}_\alpha(X)$, we obtain $\rho \geq \mathrm{ES}_{\mathfrak{nt}(\rho)}$.
Conversely, if $\rho \geq \mathrm{ES}_p$ for some $0 < p \leq 1$, we must have that $T_\rho(q) = T_{\mathrm{ES}_p}(q) = 1$ for $1 - p < q < 1$. As T_ρ is also left-continuous (Lemma C.4), ρ must be prudent by Theorem 3.1. □

Transparently, Proposition 5.4 lends itself to several immediate interpretations:

(a) Among the prudent exact DRMs, the Expected Shortfall is the one that imposes the minimal capital requirement $\rho(X)$ for undertaking risk X.
(b) Let us recall that a risk measure ρ induces a *capital adequacy test*, which a net loss X passes if it belongs to the so-called acceptance set $\mathcal{A}_\rho := \{\rho \leq 0\}$. By Proposition 5.4, the Expected Shortfall is the one with the most relaxed capital adequacy test in the class of prudent exact DRMs.
(c) For every prudent exact DRM ρ, the associated distortion function T_ρ distorts $\mathbb{P}$-probabilities *more* than the Expected Shortfall: i.e., $|T_\rho(x) - x| \geq |T_{\mathrm{ES}_\bullet}(x) - x|$ for all $x \in [0, 1]$.

In sum, in this section we have seen that the index of nontriviality takes an especially useful form in the case of exact DRMs. In the special case of coherent DRMs, this form unveils a key role of backbone probabilities: that of determining which gains are deemed speculative. We have seen that coherent DRMs that are, in addition, prudent always identify as speculative some gains that are strictly positive in expectation under the statistical reference measure $\mathbb{P}$. Thus, by adding prudence to coherence one effectively robustifies risk assessments. Finally, we have extended an important minimality property of the ES: it is the risk measure that imposes the least capital requirements within the class of prudent exact DRMs.

5.1. **An aside: Lower semicontinuity with respect to convergence in distribution.** It is noteworthy that Proposition 5.3 allows one to give a sufficient dual condition for a convex law-invariant functional $\varphi\colon L^\infty \longrightarrow \mathbb{R}$ to be l.s.c. with respect to convergence in distribution. To see this, let us recall that the convex conjugate of φ is the function $\varphi^*\colon \mathbf{ba} \longrightarrow (-\infty, \infty]$ defined by

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

Its effective domain $\mathrm{dom}(\varphi^*)$ is the set $\{\mu \in \mathbf{ba} \mid \varphi^*(\mu) < \infty\}$.

Corollary 5.5. *For a monotone and convex law-invariant functional φ on L^∞ define*

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

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

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

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

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

Proof. For $\mu \in \mathbf{ba}_+$, let us define ψ_μ exactly as in (4.2). By definition, $\mathcal{M}^\star \subset \mathcal{M}$, and the latter set is nonempty if the former is nonempty. In such a case, $\sup_{\mu\in\mathcal{M}} \big\{ \int X \,\mathrm{d}\mu - \varphi^*(\mu)\big\} \geq \sup_{\zeta\in\mathcal{M}^\star} \big\{ \int X \,\mathrm{d}\zeta - \varphi^*(\zeta)\big\}$. Now let us pick $\mu \in \mathcal{M}$. If $\mu = 0$, then $\mu \in \mathcal{M}^\star$. If $\mu \neq 0$, set $c := \mu(\Omega)$ and consider the probability charge $\mathfrak{q} := c^{-1}\mu$. By Proposition 5.3, $\psi_{\mathfrak{q}}$ is prudent. Set $T := T_{\psi_\mu}$ and let (T_n) be a sequence of continuous and concave distortion functions defined by $T_n(x) = \min\{2^n T(2^{-n})x, T(x)\}$, $x \in [0,1]$. As $\psi_{\mathfrak{q}}$ is prudent, for all $1 - \mathfrak{nt}(\psi_{\mathfrak{q}}) < x \leq 1$ and all $n \in \mathbb{N}$ large enough, $T_n(x) = 1$. As observed in the proof of Theorem 4.2, the DRMs ψ_n associated with T_n satisfy $\psi_n \leq \psi_{\mathfrak{q}}$ and are generated by countably additive backbones $\mathbb{Q}_n$. By Proposition 5.3, there exists $n_0 \in \mathbb{N}$ such that $n \geq n_0$ implies $\mathbb{Q}_n \not\approx \mathbb{P}$. We infer that every probability measure $\mathbb{Q}'$ in the set

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

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

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

Hence, $\{c\mathbb{Q}' \mid \mathbb{Q}' \in \mathcal{Q}\} \subset \mathcal{M}^\star$ and

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

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 $\varphi(X) = \sup_{\mu\in\mathcal{M}} \big\{ \int X \,\mathrm{d}\mu - \varphi^*(\mu)\big\}$. By applying Proposition 5.3 and Theorem 3.1 if $\mu \neq 0$, we see that each functional $X \longmapsto \psi_\mu(X) - \varphi^*(\mu)$ is l.s.c. with respect to convergence in distribution. Thus, $\varphi(X)$ is the upper envelope of a family of functionals that are all l.s.c. with respect convergence in distribution and it is, therefore, l.s.c. with respect convergence in distribution. □

Example 5.6. As an example, one may consider a convex and law-invariant monetary risk measure ρ on L^∞ for which

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

and for which ρ^* is continuous on the former set with respect to total variation. A concrete example of one such risk measure is the entropic risk measure. Then Corollary 5.5 applies and ρ is l.s.c. with respect to convergence in distribution.

6. Revisiting the Wang-Zitikis axiomatization

We said in the Introduction that one of the developments that triggered the present work was the axiomatization of the ES recently provided by Wang & Zitikis in [40]. In the previous sections, we have examined the class of prudent DRMS, of prudent exact DRMs, of coherent DRMs and, at the end of the path, we have found the ES. The latter has appeared as the minimal element in large classes of prudent DRMs; i.e., as the risk measure that imposes the least costly capital requirements while retaining a certain degree of caution.

Figuratively speaking, Wang & Zitikis did not follow a path like ours but rather went straight to the ES, which they characterize as follows. For $0 < p < 1$, ES_p is the unique functional $\rho\colon L^1 \longrightarrow \mathbb{R}$ with $\rho(1) = 1$ satisfying (a) monotonicity; (b) law invariance; (c) prudence; and (d) *no reward for concentration* (NRC): There is an event $A \in \mathcal{F}$ satisfying $\mathbb{P}(A) = 1 - p$ such that, for all $X, Y \in L^1$ sharing the tail event A,

$$\rho(X + Y) = \rho(X) + \rho(Y).^{5}$$

The difference between the approaches is large as well as evident, and it now suggests that we would relate one approach to the other. Summarily, this relation – which we unveil from Lemma 6.1 to Theorem 6.3, below – is as follows. As it was to be inferred from even a quick glance at the two approaches, the axiom of NRC is "heavily loaded". Not only does it imply comonotonic additivity but also that the distortion function must have a very special form; namely, it must be two-piece linear. Transparently, this implies that (law-invariant) risk measures satisfying NRC can only be either sub-additive or super-additive. By "residuality", the role of prudence is simply that of picking sub-additivity over super-additivity and of further robustifying the risk asessments by discarding speculative gains in the way we saw above. Automatically, this additional robustification, combined with the special form of the distortion function, produces the minimality property we discussed in the previous section.

In detail, we begin with Lemma 6.1 below, which shows (*a posteriori*) that the class of DRMs is the natural one for studying law-invariant DRMs that offer NRC.

Lemma 6.1. *Suppose $\mathcal{X} \subset L^0$ is a law-invariant subspace and $\varphi\colon \mathcal{X} \longrightarrow \mathbb{R}$.*

1. *If φ is law invariant and offers NRC with tail event A, then φ is comonotonic additive.*
2. *If $\mathcal{X} = L^\infty$ and φ is monotone, then φ is a DRM.*

Proof. 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 $X, Y \in \mathcal{X}$ be comonotone. Select nondecreasing functions $f, g\colon \mathbb{R} \longrightarrow \mathbb{R}$ such that $X = f(X + Y)$ and $Y = g(X + Y)$ ([16, Lemma 4.89]). Let $U \in \mathcal{U}$ be such that $X + Y = F^{-1}_{X+Y}(U)$. By using the law-invariance of φ in the second and fourth equality below and NRC in the third, we have

$$\begin{aligned}\varphi(X + Y) &= \varphi\big((f \circ F^{-1}_{X+Y})(U) + (g \circ F^{-1}_{X+Y})(U)\big) = \varphi\big((f \circ F^{-1}_{X+Y})(\widehat{U}) + (g \circ F^{-1}_{X+Y})(\widehat{U})\big) \\ &= \varphi\big((f \circ F^{-1}_{X+Y})(\widehat{U})\big) + \varphi\big((g \circ F^{-1}_{X+Y})(\widehat{U})\big) = \varphi(X) + \varphi(Y).\end{aligned}$$

□

[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 $Y\mathbf{1}_{A^c} \leq \beta\mathbf{1}_{A^c}$.

The next proposition shows the rigidity of the NRC requirement: by combining, as in [40], NRC with monotonicity one forces the distortion function to be two-piece linear. In fact, Proposition 6.2 shows that the impact of NRC is even stronger as it suffices to combine NRC with continuity (which in [40] is implied by monotonicity and cash-invariance) to obtain this conclusion.

Proposition 6.2. *Suppose a law-invariant and continuous functional $\varphi\colon L^\infty \longrightarrow \mathbb{R}$ offers NRC with tail event A. Then φ is a signed Choquet integral whose associated distortion function T is given by*

$$T(x) = \begin{cases} \gamma x & x \le \mathbb{P}(A) \\ \gamma\mathbb{P}(A) + \delta(x - \mathbb{P}(A)) & x > \mathbb{P}(A). \end{cases} \tag{6.1}$$

The constants $\gamma, \delta \in \mathbb{R}$ satisfy

$$(\gamma - \delta)\mathbb{P}(A) = \varphi(1) - \delta. \tag{6.2}$$

In particular, φ is necessarily either subadditive or superadditive (i.e., $-\varphi$ is subadditive).

Proof. Denote by p the $\mathbb{P}$-probability of tail event A. By Lemma 6.1, φ is comonotonic additive. By [35, Proposition 2], φ is a signed Choquet integral whose associated set function satisfies $v = T \circ \mathbb{P}$ for a unique function $T\colon [0,1] \longrightarrow \mathbb{R}$ with $T(0) = 0$. Now, suppose that $x, y \in [0,p]$ are such that $x + y \le p$. 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, by NRC we have that

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

Thus, T satisfies the Cauchy functional equation on $[0,p]$. As T has to be continuous at all but countably many points in that interval, $T(x) = \gamma x$ must hold for a suitable $\gamma \in \mathbb{R}$ and all $x \in [0,p]$. Now let $s, t \in [0,1]$ such that $p + s + t \le 1$. Select events $D, E \supset A$ 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

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

Hence, for all such choices of s and t

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

which implies like above 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 $\delta \in \mathbb{R}$ being suitably chosen. Moreover, $\varphi(1) = T(1) = \gamma p + \delta(1-p)$. Rearranging this yields (6.2).
At last, note that the function T is either concave or convex. In the latter case, the set function $T \circ \mathbb{P}$ is convex in the sense of [33]. By [33, Theorem 4.6], φ is superadditive. If T is concave, subadditivity of φ follows by [39, Theorem 3]. $\square$

By leveraging Proposition 6.2, we can now give a variant of the Wang & Zitikis axiomatization. It casts the ES in a much larger class of not necessarily monotone functionals.

Theorem 6.3. *For a functional $\varphi\colon L^\infty \longrightarrow \mathbb{R}$ with $\varphi(1) = 1$, the following are equivalent:*

(1) *For some $0 < p < 1$, $\varphi = \mathrm{ES}_p$.*
(2) *φ has the following properties:*

(i) *Law invariance.*
(ii) φ *offers NRC with tail event* A.
(iii) *Prudence.*
(iv) *Continuity.*
(v) *The tail event* A *satisfies* $|\varphi(\mathbf{1}_A)| \leq 1$.

Proof. It is straightforward to verify that ES_p has properties (i)–(v). In the converse direction, by (i), (ii), (iv) and Proposition 6.2, φ is the Choquet integral with respect to the set function $T \circ \mathbb{P}$, T being given as in (6.1). If the constant γ were negative, every sequence (A_n) of vanishing events would lead to a sequence $X_n := \frac{1}{|\gamma|\mathbb{P}(A_n)}\mathbf{1}_{A_n}$, $n \in \mathbb{N}$, that converges to 0 $\mathbb{P}$-a.s. and satisfies $\lim_{n\to\infty} \varphi(X_n) = -1$. Hence, $\varphi(X) > \lim_{n\to\infty} \varphi(X_n)$, contradicting prudence. If we had $\gamma = 0$, then from $T(1) = \varphi(1) = 1$ and (6.2) it would follow that $\delta > 0$ and φ would be a DRM as T is nondecreasing. But, prudence of φ and Theorem 3.1 would then yield $\delta = 0$, again a contradiction.
Now set $q := \mathbb{P}(A)$. As $\gamma > 0$, assumption (iii) yields $0 < \gamma q = T(q) = \varphi(\mathbf{1}_A) \leq 1$. From (6.2), it follows that $\delta \geq 0$; i.e., T is nondecreasing and φ is a DRM. It remains only to observe that Theorem 3.1 implies that $\delta = 0$, whence $\gamma = \frac{1}{q}$ follows and $T = T_{\mathrm{ES}_{1-q}}$. □

We conclude the section with two corollaries. The first follows by combining the above with our general study of extension properties of prudent DRMs in Appendix A. It yields and refines the uniqueness principle formulated in [40, Theorem 2].

Corollary 6.4. *Suppose* $L^\infty \subset \mathcal{X} \subset L^0$ *is a law-invariant lattice and that a functional* $\varphi\colon \mathcal{X} \to \mathbb{R}$ *is law-invariant, monotone, prudent and offers NRC. Then one of the following cases must hold:*

(1) $\varphi = 0$.
(2) $\mathcal{X} \subset L^1$ *and* $\varphi = \varphi(1)\,\mathrm{ES}_p$ *for some* $0 < p < 1$.

Proof. Suppose that (1) does not hold. By comonotonic additivity (Lemma 6.1) and monotonicity, this entails $\varphi(1) > 0$. Without loss, assume $\varphi(1) = 1$ and note that $\rho := \varphi|_{L^\infty} = \mathrm{ES}_p$ for some $p \in (0,1)$ by Theorem 6.3. By Corollary A.4, $\varphi = (\mathrm{ES}_p)^\sharp|_{\mathcal{X}}$. However, $\mathrm{ES}_p^\sharp(X) = \infty$ for all $X \in L^0_+ \setminus L^1$, which means that $\mathcal{X}_+ \subset L^1$ must hold. This suffices to prove $\mathcal{X} \subset L^1$. □

The second corollary follows at once by combining Corollary 3.3 and Proposition 6.2. It shows that prudent DRMs locally require greater capital buffers than law-invariant risk measures offering NRC unless one deals with an Expected Shortfall, in which case both prudence and NRC are satisfied. In this sense, we can think of prudence as a more conservative property than NRC.

Corollary 6.5. *Suppose* $\rho_1, \rho_2\colon L^\infty \longrightarrow \mathbb{R}$ *are DRMs such that* ρ_1 *offers NRC and* ρ_2 *is prudent. Then one of the following alternatives holds:*

(1) *For some* $X_0 \in L^\infty$, $\rho_1(X_0) < \rho_2(X_0)$.
(2) $\rho_1 = \mathrm{ES}_p$ *for some* $p \in (0,1]$.

In fact, Corollary 6.5 allows us to regard prudence and NRC as dual properties. By Proposition 5.4, ES_p is minimal in the family $\mathcal{R}^e(p)$ of prudent exact DRMs with $\mathfrak{nt}(\rho) \geq p$. By the conjunction of Proposition 6.2 with Corollary 6.5, however, ES_p is maximal among DRMs offering NRC with a tail event A satisfying $\mathbb{P}(A) \geq 1 - p$.

Appendix A. Extension properties of prudent DRMs

As indicated in Remark 3.4, in this Appendix we are going to show that prudent DRMs defined on L^∞ have a unique prudent extension to L^1. For the immediate concerns of the present paper, this shows that there is no gap between our analysis and that of [40], who consider risk measures with domain L^1.

We should like to point out, however, that the results contained in this Appendix go well beyond this point and are of interest in their own right. The problem of extending risk measures to the domain L^0 of all measurable functions on the measurable space under consideration has been a long outstanding problem (see [14, Section 5]). Most extension results obtained thus far rely, nonetheless, on convexity assumptions as they make heavy use of the dual representation of the risk measures in question. As prudent DRMs are not necessarily convex, our methods are different and the results not covered by the earlier literature.

The building block of this Appendix is Proposition A.2 below, where we show that prudent DRMs on L^∞ have a canonical extension to L^0, which preserves all of its defining properties (including l.s.c. with respect to convergence in distribution). From this, we show that any *finite* valued, *weakly* prudent risk measure has a unique extension to L^0, which preserves all of its properties. When the risk measure is not only weakly prudent but also prudent as in [40], we can drop the condition that the risk measure be finite valued.

Given a prudent DRM $\rho\colon L^\infty \longrightarrow \mathbb{R}$, we begin by defining a canonical extension to L^0. By (3.1) in the proof of Theorem 3.1, ρ is tail relevant (in the terminology of [32]) and we find $q \in (0,1)$ such that, for all $X \in L^\infty$,

$$\begin{aligned}\rho(X) &= \rho\big(X \vee \mathrm{VaR}_q(X)\big) = \rho\big(X \vee \mathrm{VaR}_q(X) - \mathrm{VaR}_q(X)\big) + \mathrm{VaR}_q(X)\\ &= \int_{[0,\infty)} T_\rho\big(\mathbb{P}(X > x + \mathrm{VaR}_q(X))\big)\mathrm{d}x + \mathrm{VaR}_q(X)\\ &= \int_{[\mathrm{VaR}_q(X),\infty)} T_\rho\big(\mathbb{P}(X > x)\big)\mathrm{d}x + \mathrm{VaR}_q(X).\end{aligned} \tag{A.1}$$

Note that the integral in the last expression in (A.1) is well defined, whether or not X is bounded.

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

$$\rho^\sharp(X) := \int_{[\mathrm{VaR}_q(X),\infty)} T_\rho\big(\mathbb{P}(X > x)\big)\mathrm{d}x + \mathrm{VaR}_q(X).$$

Proposition A.2 shows that this extension retains all the defining properties of ρ.

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

Proof. We have already observed that $\rho^\sharp$ is well defined.

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 $\rho(Y \vee \mathrm{VaR}_q(Y)) = \rho(Y) = \rho(Y \vee \mathrm{VaR}_r(Y))$ for all $Y \in L^\infty$. Then, for all $t > 1 - r$ and $U \in \mathcal{U}$, $\mathrm{VaR}_r(\mathbf{1}_{\{U\le t\}}) = 1$. Hence,

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

Moreover, observe that for $X \in L^0$ and $x < \mathrm{VaR}_r(X)$, $\mathbb{P}(X > x) > 1 - r$, which entails

$$\mathrm{VaR}_q(X) + \int_{[\mathrm{VaR}_q(X),\infty)} T_\rho\big(\mathbb{P}(X > x)\big)\mathrm{d}x = \mathrm{VaR}_r(X) + \int_{[\mathrm{VaR}_r(X),\infty)} T_\rho\big(\mathbb{P}(X > x)\big)\mathrm{d}x.$$

Step 2: $\rho^\sharp$ is law invariant and monotone on L^0 by definition. Its (weak) prudence is verified once we show the stronger property of $\rho^\sharp$ being l.s.c. with respect to convergence in distribution.

Step 3: Suppose $(X_n) \subset L^0$ 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 $X_n \le X$ and $X_n \to X$ $\mathbb{P}$-a.s. In particular, the left-continuity of T_ρ (Theorem 3.1) ensures that, Lebesgue-a.e.,

$$\lim_{n\to\infty} T_\rho\left(\mathbb{P}(X_n > x)\right) = T_\rho\left(\mathbb{P}(X > x)\right). \tag{A.2}$$

By Step 1 and Skorokhod representation, we can assume that the parameter $q \in (0,1)$ satisfies $\lim_{n\to\infty} \mathrm{VaR}_q(X_n) = \mathrm{VaR}_q(X)$. In particular, we can select a constant $c \in \mathbb{R}$ such that, for all $Y \in \{X, X_1, X_2, ...\}$, $c \le \mathrm{VaR}_q(Y)$ and thus

$$\int_{[\mathrm{VaR}_q(Y),\infty)} T_\rho\big(\mathbb{P}(Y > x)\big)\mathrm{d}x = \int_{[c,\infty)} T_\rho\big(\mathbb{P}(Y > x)\big)\mathrm{d}x + \mathrm{VaR}_q(Y) - c.$$

By [3, Theorem 11.32], equation (A.2) and dominated convergence, for all $k > 0$ large enough we have that

$$\begin{aligned}\liminf_{n\to\infty} \int_{[\mathrm{VaR}_q(X_n),\infty)} T_\rho\left(\mathbb{P}(X_n > x)\right)\mathrm{d}x &= \liminf_{n\to\infty} \Big[\int_{[c,\infty)} T_\rho\left(\mathbb{P}(X_n > x)\right)\mathrm{d}x + \mathrm{VaR}_q(X_n) - c\Big] \\ &\ge \lim_{n\to\infty} \Big[\int_{[c,k]} T_\rho\left(\mathbb{P}(X_n > x)\right)\mathrm{d}x + \mathrm{VaR}_q(X_n) - c\Big] \\ &= \Big[\int_{[c,k]} T_\rho\left(\mathbb{P}(X > x)\right)\mathrm{d}x + \mathrm{VaR}_q(X) - c\Big].\end{aligned}$$

By letting $k \to \infty$, we see that

$$\begin{aligned}\liminf_{n\to\infty} \int_{[\mathrm{VaR}_q(X_n),\infty)} T_\rho\left(\mathbb{P}(X_n > x)\right)\mathrm{d}x &= \int_{[c,\infty)} T_\rho\left(\mathbb{P}(X > x)\right)\mathrm{d}x + \mathrm{VaR}_q(X) - c \\ &= \int_{[\mathrm{VaR}_q(X),\infty)} T_\rho\left(\mathbb{P}(X > x)\right)\mathrm{d}x.\end{aligned}$$

Putting the pieces together, we have shown the desired lower semicontinuity.

Step 4: $\rho^\sharp$ is comonotonic additive. Let $X, Y \in L^0$ be comonotonic. By [15, Proposition 4.5] we find continuous functions $f, g\colon \mathbb{R} \longrightarrow \mathbb{R}$ such that $X = f(X+Y)$ and $Y = g(X+Y)$. By [16, Lemma A.32], there is $U \in \mathcal{U}$ such that $X + Y = F^{-1}_{X+Y}(U)$. Set $f^\star := f \circ F^{-1}_{X+Y}$, $g^\star := g \circ F^{-1}_{X+Y}$, and note that these two functions are left-continuous. Using [16, Lemma A.27] in the penultimate identity

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

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

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

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

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

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

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

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

The next corollary gives sufficient conditions for the extension $\rho^\sharp$ to be uniquely determined by the values taken by the risk measure on L^∞.

Corollary A.3. *Suppose $\mathcal{X} \subset L^0$ is a lattice containing L^∞. Suppose that $\varphi\colon \mathcal{X} \longrightarrow \mathbb{R}$ is monotone, law invariant, weakly prudent, comonotonic additive, and satisfies $\varphi(1) = 1$. Then the following assertions hold:*

(1) *$\rho := \varphi|_{L^\infty}$ is a prudent DRM.*
(2) *$\varphi = \rho^\sharp|_{\mathcal{X}}$.*
(3) *φ is prudent.*

Proof. ρ is finite, monotone, comonotonic additive, law invariant and prudent; that is, it is a prudent DRM. This is item (1).
Turning to (2), let $\rho^\sharp$ denote the extension of ρ to all of L^0 from Proposition A.2. Let $X \in \mathcal{X}$, $m \in \mathbb{N}$, and set $Y := X \vee (-m)$. Observe that $Y \wedge n \uparrow Y$ a.s. and that $\lim_{n\to\infty} \rho(Y \wedge n)$ exists in $\mathbb{R}$, whence $\lim_{n\to\infty} \rho(Y \wedge n) = \varphi(Y)$ follows with weak prudence. In additition, the argument from Step 4 in the proof of Proposition A.2 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, $\lim_{m\to\infty} \varphi(Z_m) = 0 = \lim_{m\to\infty} \rho^\sharp(Z_m)$. Using comonotonic additivity,

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

Last, for (3), prudence of φ in inherited from the prudence of $\rho^\sharp$ as recorded in Proposition A.2. □

One potential drawback of Corollary A.3 is that the functional must be finite-valued. If φ is not only weakly prudent, but also prudent, 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 A.4. *Suppose $\mathcal{X} \subset L^0$ is a lattice containing L^∞. Suppose that $\varphi\colon \mathcal{X} \longrightarrow (-\infty, \infty]$ satisfies $\varphi(0) = 0$, monotonicity, law invariance, prudence, and comonotonic additivity. Set $\rho := \varphi|_{L^\infty}$. Then ρ is a prudent DRM and $\varphi = \rho^\sharp|_{\mathcal{X}}$.*

Proof. First note that ρ only takes finite values. Indeed, for all $X \in L^\infty$, monotonicity of ρ yields

$$\rho(X) \leq \rho(\|X\|_\infty) = \rho(-\|X\|_\infty) - \rho(-2\|X\|_\infty) \leq -\rho(-2\|X\|_\infty),$$

and the latter is a finite number as $\rho(y) \leq 0$ for all $y \leq 0$. Hence, ρ is a prudent DRM. The remainder of the proof is identical to the one of Corollary A.3. □

Appendix B. Weak prudence vs prudence

In this Appendix, we give some sufficient conditions for the equivalence between weak prudence (see subsection 3.2) and prudence in the case of law-invariant, monotone functionals that are not necessarily DRMs. As anticipated in Section 3.2, whether or not this equivalence always holds remains an open question to us.

Proposition B.1. *Suppose $\varphi\colon L^\infty \longrightarrow (-\infty,\infty]$ is law invariant, monotone and weakly prudent. Moreover, assume 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 $\zeta \ll \mathbb{P}$ and $c \in \mathbb{R}$ such that $\zeta \not\approx \mathbb{P}$ and*
$$\varphi \geq \int \cdot\, d\zeta + c. \tag{B.1}$$

Then φ is also prudent.

Proof. Suppose a sequence $(X_n) \subset L^\infty$ has a.s. limit $X \in L^\infty$ and that $\lim_{n\to\infty}\varphi(X_n) \in [-\infty,\infty]$ exists. We need to show that $\varphi(X) \leq \lim_{n\to\infty}\varphi(X_n)$. By law invariance, we can assume the existence of $U \in \mathcal{U}$ such that $X_n = F_{X_n}^{-1}(U) \longrightarrow F_X^{-1}(U) = X$ a.s. If $\lim_{n\to\infty}\varphi(X_n) > -\infty$, then the inequality follows either trivially or by invoking weak prudence. In the following, we verify that each of the conditions (i)–(iii) prevents the possibility that $\lim_{n\to\infty}\varphi(X_n) = -\infty$.
Set $m := \|X\|_\infty + 1$, $A_n := \{X_n > -m\}$, $n \in \mathbb{N}$, and
$$Y_n := -m\mathbf{1}_{A_n} - \|X_n\|_\infty \mathbf{1}_{A_n^c}. \tag{B.2}$$
Then $Y_n \leq X_n$, i.e., $\lim_{n\to\infty}\varphi(Y_n) = -\infty$. Suppose now that condition (i) holds and choose a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $\sum_{k=1}^\infty \mathbb{P}(A_{n_k}^c) < \infty$. Then
$$Y := -\sum_{k=1}^\infty \|X_{n_k}\|_\infty \mathbf{1}_{A_{n_k}}$$
is a well-defined random variable in $-L^0_+$. Moreover, $Y \wedge (-m) \leq Y_{n_k}$ for all $k \in \mathbb{N}$. Hence, by properness of the extension $\varphi^\sharp$ of φ, $-\infty < \varphi^\sharp\big(Y \wedge (-m)\big) \leq \inf_{k\in\mathbb{N}}\varphi(Y_{n_k}) = -\infty$, a contradiction.
Now suppose that (ii) holds, and define once again a sequence (Y_n) by (B.2). Then, for all n large enough,
$$-a\|X\|_\infty - a + b = a\mathrm{VaR}_p(Y_n) + b \leq \varphi(Y_n),$$
and we have a contradiction with properness of φ.
At last, suppose that (iii) holds. If $\zeta = 0$, then $\varphi \geq c$ and a sequence constructed as above cannot exist. Else, let $\mathbb{Q} := \zeta(\Omega)^{-1}\zeta$ and note that – by law invariance of φ – (B.1) holds if and only if $\varphi \geq \zeta(\Omega)^{-1}\psi_{\mathbb{Q}} + c$. As $\mathbb{Q} \not\approx \mathbb{P}$, $\psi_{\mathbb{Q}}$ is prudent by Corollary 5.3. By Theorem 3.1, there is $p \in (0,1)$ such that $\psi_{\mathbb{Q}} \geq \mathrm{VaR}_p$. Hence,
$$\varphi \geq \zeta(\Omega)^{-1}\mathrm{VaR}_p + c,$$
and condition (ii) is satisfied. □

Example B.2. 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 $\rho^*(\mathfrak{q}) < \infty$. By Corollary 5.3, ρ satisfies alternative (ii) in Proposition B.1 and is weakly prudent if and only if it is prudent.

Appendix C. Auxiliary results

Lemma C.1. *Let ρ be defined as in equation (2.2). Then, ρ is a well-defined map $\rho\colon L^1 \longrightarrow (-\infty,\infty]$ that is norm-l.s.c. and subadditive.*

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

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

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

$$(X_n) \subset L^1 \text{ satisfies } X_n \uparrow X \in L^1 \quad \Longrightarrow \quad \rho(X_n) \uparrow \rho(X). \tag{C.1}$$

Indeed, every norm-l.s.c. monotone functional on L^1 has this property. Conversely, let $(Y_n) \subset L^1$ be a sequence 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

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

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 (C.1) is verified to be equivalent to L^1-lower semicontinuity of ρ. Now, if $(X_n) \subset L^1$ satisfies $X_n \uparrow X$, monotone convergence implies that ρ satisfies (C.1).
We now turn to the subadditivity of ρ. To this end, let $X, Y \in L^1$ and set $Z := X + Y$. By [17, 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 [6, Lemma 3.4] for the first estimate and selecting $n \in \mathbb{N}$ large enough such that the second holds,

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

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

The proof of the following easy lemma is omitted.

Lemma C.2. *Suppose $\mathcal{D} \subset L^0$ is law invariant and $\rho\colon \mathcal{D} \longrightarrow \mathbb{R}$ is a law-invariant functional. Then ρ is prudent if and only if ρ is lower semicontinuous (l.s.c.) with respect to convergence in distribution: If $X, X_1, X_2, \ldots$ are random variables in $\mathcal{D}$, then*

$$X_n \xrightarrow{d} X \text{ as } n \to \infty \quad \Longrightarrow \quad \rho(X) \leq \liminf_{n\to\infty} \rho(X_n).$$

Lemma C.3. *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 [4, Corollary 3.1], $T \geq id_{[0,1]}$ holds under the assumption of the lemma. For any convergent sequence $(x_n) \subset [0, 1]$ with $\lim_{n\to\infty} x_n = 1$, 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 $(0, 1]$. □

Lemma C.4. *Let $v = T \circ \mathbb{P}$ be an exact distortion of an atomless probability measure $\mathbb{P}$. Then T is left-continuous.*

Proof. By [30, 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

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

This establishes left-continuity of T. □

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

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

[6] 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.

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

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

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

[10] 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.

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

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

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

[18] Filipović, D., and G. Svindland (2012), The canonical model space of law invariant risk measures is L^1. *Mathematical Finance* 22(3):585–589.
[19] 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.
[20] Ghirardato, P., and M. Marinacci (2002), Ambiguity made precise: a comparative foundation. *Journal of Economic Theory* 102:251–289.
[21] Hampel, F. R. (1971), A general qualitative definition of robustness. *Annals of Mathematical Statistics* 42:1887–1896.
[22] He, X. D., S. Kou, and X. Peng (2022), Risk measures: Robustness, elicitability, and backtesting. *Annual Review of Statistics and Its Applications* 9:141–166.
[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., and C. Munari (2014), Law-invariant risk measures: Extension properties and qualitative robustness. *Statistics & Risk Modeling* 31:215–236.
[27] Kou, S., and X. Peng (2016), On the measurement of economic tail risk. *Operations Research* 64(5):1056–1072.
[28] Krätschmer, V., A. Schied, and H. Zähle (2014), Comparative and qualitative robustness for law-invariant risk measures. *Finance and Stochastics* 18(2):271–295.
[29] Kusuoka, S. (2001), On law invariant coherent risk measures. *Advances in Mathematical Economics* 3:83–95.
[30] Liebrich, F.-B., and C. Munari (2022), Law-invariant functionals that collapse to the mean: Beyond convexity. *Mathematics and Financial Economics* 16:447–480.
[31] Liebrich, F.-B., and G. Svindland (2017), Model spaces for risk measures. *Insurance: Mathematics and Economics* 77:150–165.
[32] Liu, F., and R. Wang (2020), A theory for measures of tail risk. *Mathematics of Operations Research* 46(3):1109–1128.
[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] Schmeidler, D. (1972), Cores of exact games I, *Journal of Mathematical Analysis and Applications* 40 (1): 214-225.
[35] Schmeidler, D. (1986), Integral representation without additivity. *Proceedings of the American Mathematical Society* 97:255–261.
[36] Shaked, M., and J. G. Shanthikumar (2007), *Stochastic Orders*. Springer.
[37] Shapiro, A. (2013), On Kusuoka representation of law invariant risk measures. *Mathematics of Operations Research* 38(1):142–152.
[38] Svindland, G. (2010), Continuity properties of law-invariant (quasi-)convex risk functions on L^∞. *Mathematics and Financial Economics* 3(1):39–43.
[39] Wang, R., Y. Wei, and G. Willmot (2020), Characterization, robustness and aggregation of signed Choquet integrals. *Mathematics of Operations Research* 45(3):993–1015.
[40] Wang, R., and R. Zitikis (2021), An axiomatic foundation for the Expected Shortfall. *Management Science* 67(3):1413–1429.
[41] Wang, S. (1996), Premium calculation by transforming the layer premium density. *ASTIN Bulletin* 26(1):71–92.
[42] Wang, S., V. Young, and H. Panjer (1997), Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics* 21(2):173–183.