Risk Functionals with Convex Level Sets

Ruodu Wang* Yunran Wei†

July 30, 2020

Abstract

We analyze the "convex level sets" (CxLS) property of risk functionals, which is a necessary condition for the notions of elicitability, identifiability and backtestability, popular in the recent statistics and risk management literature. We put the CxLS property in the multi-dimensional setting, with a special focus on signed Choquet integrals, a class of risk functionals that are generally not monotone or convex. We obtain two main analytical results in dimension one and dimension two, by characterizing the CxLS property of all one-dimensional signed Choquet integrals, and that of all two-dimensional signed Choquet integrals with a quantile component. Using these results, we proceed to show that under some continuity assumption, a comonotonic-additive coherent risk measure is co-elicitable with Value-at-Risk if and only if it is the corresponding Expected Shortfall. The new findings generalize several results in the recent literature, and partially answer an open question on the characterization of multi-dimensional elicitability.

Keywords: convex level sets, quantiles, Expected Shortfall, elicitability, backtestability

---

1 Introduction

Over the past decade, the concepts of elicitability, identifiability and backtestability have received an increasing attention in the statistics and risk management literature (e.g., Gneiting (2011), Ziegel (2016), Fissler and Ziegel (2016), Kou and Peng (2016), Acerbi and Szekely (2017)). These concepts refer to the assessment of the quality and validity of risk forecasts, and have been a prominent issue in banking regulation and model risk management (see e.g., BCBS (2016)). A

*Department of Statistics and Actuarial Science, University of Waterloo, Canada. Email: wang@uwaterloo.ca.

†Department of Statistics and Actuarial Science, Northern Illinois University, USA. Email: ywei1@niu.edu.

key mathematical property of risk functionals related to the above concepts is the *convex level sets (CxLS)* property of Osband (Osband (1985)). The CxLS property is a necessary condition for the three concepts above, and it is also known to be sufficient under mild regularity conditions for one-dimensional risk functionals (Corollary 9 of Steinwart et al. (2014)). In general, CxLS property is not a sufficient condition for elicitability; some counter-examples are the mode functional provided by Heinrich (2014) and the max-functional provided by Brehmer and Strokorb (2019).

In this paper, we focus on the CxLS property of (possibly multi-dimensional) risk functionals. For the reader who is not familiar with elicitability, identifiability and backtestability, we recommend him/her to start reading the paper from Section 6, where we present the formal definitions of those concepts, their risk management implications, and their relation with the CxLS property.

Below, we give the definition of the CxLS property, the main object of this paper. Denote by $\mathcal{M}_0$ the set of distributions (i.e., Borel probability measures on $\mathbb{R}$).

Definition 1.1. (The CxLS property) For $\mathcal{M} \subset \mathcal{M}_0$, we say a functional $\rho : \mathcal{M} \to \mathbb{R}^d$ has *convex level sets (CxLS)* on $\mathcal{M}$ if $\rho(\lambda F + (1-\lambda)G) = \rho(F)$ for all $\lambda \in [0,1]$ and $F, G \in \mathcal{M}$ satisfying $\rho(F) = \rho(G)$ and $\lambda F + (1-\lambda)G \in \mathcal{M}$.

If $\mathcal{M}$ is convex, the CxLS property of ρ means that the level set $\{F \in \mathcal{M} : \rho(F) = \gamma\}$ is convex for each $\gamma \in \mathbb{R}^d$, thus the name. Although Definition 1.1 does not require $\mathcal{M}$ itself to be convex, common choices of $\mathcal{M}$ are convex sets, such as the set of distributions with bounded support, or the set of distributions with positive densities.

To interpret the CxLS property, it means that if two risk models are assessed as equally risky, then a mixture of the two models should remain at the same risk level. In decision theory, this property is closely related to (slightly weaker than) the axiom of *betweenness*, one of the possible relaxations of the independence axiom of the von Neumann-Morgenstern expected utility theory; see, e.g., Dekel (1986) and Chew (1989). As mentioned above, the recently growing importance of the CxLS property in risk management is mainly due to its close relation with the statistical notions of elicitability, identifiability and backtestability.

In the literature of risk measures, many results on the characterization of elicitable risk measures are obtained via characterizing the CxLS property in dimension one; see Weber (2006), Bellini and Bignozzi (2015) and Delbaen et al. (2016) for convex risk measures, Ziegel (2016) for coherent risk measures, Kou and Peng (2016) and Wang and Ziegel (2015) for distortion risk measures, Liu and Wang (2020) for tail risk measures, and Fissler et al. (2019a,b) for set-valued risk functionals. As far as we are aware of, studies dedicated to CxLS in the multi-dimensional setting are not found in the literature, although multi-dimensional elicitability has been a popular topic; see Lambert et al. (2008), Fissler and Ziegel (2016), Nolde and Ziegel (2017) and Acerbi and Szekely (2017) for multi-dimensional elicitability and their statistical implications. For general statistical functionals, Lambert et al. (2008) and Frongillo and Kash (2018) studied notions of elicitation complexity, which are quantitative measurements of non-elicitability.

In the above literature, characterization results are obtained for one-dimensional increasing or convex risk functionals. Although risk measures are typically increasing functionals, many statistical quantities, such as measures of variability or shape, are not monotone with respect to the natural order on $\mathcal{M}$, and they play an important role in the statistical analysis of risks. In this paper, we study the CxLS property of non-monotone, non-convex, and multi-dimensional functionals, with a particular focus on the class of signed Choquet integrals (Wang et al. (2020)). One-dimensional signed Choquet integrals include many commonly used risk functionals, such as risk measures and variability measures. Moreover, as discussed by Fissler and Ziegel (2016), a two-dimensional signed Choquet integral (Example 4.5) gives the first example of a multi-dimensional elicitable risk functional that is not connected to one-dimensional elicitable ones via a bijection. A characterization of elicitability or CxLS for multi-dimensional signed Choquet integrals is generally an open question, as mentioned by both Kou and Peng (2016, p. 1063) and Fissler and Ziegel (2016, p. 1698).

This paper contains two main contributions to the theory of risk functionals with the CxLS property, and hence to the theory of elicitability. First, in Section 3, we characterize all one-dimensional signed Choquet integrals with the CxLS property (Theorem 3.2). It turns out that the only signed Choquet integrals that have CxLS are the monotone ones with CxLS multiplied by a constant. This result requires some new techniques which we provide in a few lemmas, as non-monotonicity of signed Choquet integrals creates great challenges, which cannot be addressed by the existing methods in the literature. Second, in Sections 4-5, we extend our discussion to the multi-dimensional setting, and relate the CxLS property to risk quadrangles of Rockafellar and Uryasev (2013). Because of the special role of the quantile functionals in the theory of elicitability and backtestability (for instance, quantile functionals and the mean are the only signed Choquet integrals that are elicitable), we characterize the CxLS property of all two-dimensional signed Choquet integrals, whose one component is a quantile (Theorem 5.3). To our knowledge, this result is the first CxLS characterization in the literature beyond the one-dimensional case, and it partially answers the open question of Kou and Peng (2016) and Fissler and Ziegel (2016). To establish this result, we provide some general results on multi-dimensional CxLS property in

Section 4. Since CxLS is a necessary condition of elicitability, identifiability and backtestability, our characterization identifies candidates for risk functionals with these statistical properties. Based on the result in dimension two, we show that under the assumption of lower semi-continuity with respect to weak convergence, the only spectral risk measure (i.e., comonotonic-additive and coherent risk measure) co-elicitable with a Value-at-Risk is the corresponding Expected Shortfall (Theorem 6.9 and Corollary 6.12). To better illustrate the concept of the CxLS property, we present a list of commonly used functionals with or without CxLS in Section 2, and in Section 6 we give an overview on the relationship among the statistical concepts of elicitability, identifiability, backtestability and the CxLS property. The proofs of the technical results are put in the Appendix.

Notation

For $q \in [1, \infty)$, let $\mathcal{M}_q$ be the set of distributions with finite q-th moments. Denote by $\mathcal{M}_\infty$ the set of distributions of bounded random variables, $\mathcal{M}_{\text{con}} = \{F \in \mathcal{M}_\infty : F \text{ has a density}\}$, and $\mathcal{M}_{\text{dis}} = \{F \in \mathcal{M}_\infty : F \text{ is discrete}\}$. Recall that, as in Definition 1.1, $\mathcal{M}_0$ is the set of all distributions on $\mathbb{R}$. For the ease of presentation, we identify distributions in $\mathcal{M}_0$ with the corresponding cumulative distribution functions, that is, for $F \in \mathcal{M}_0$ and $x \in \mathbb{R}$, $F(x)$ is understood as $F((-\infty, x])$. For $F \in \mathcal{M}_0$, define the left-continuous generalized inverse (left-quantile) as

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}, \quad t \in (0, 1],$$

and in addition let $F^{-1}(0) = \sup\{x \in \mathbb{R} : F(x) = 0\}$. For $x \in \mathbb{R}$, δ_x denotes the point mass at x. Throughout this article, we stick to the following convention. In a result, if we do not specify $\mathcal{M}$, then the statements hold for any set $\mathcal{M} \subset \mathcal{M}_0$ such that the risk functional at consideration is finite on $\mathcal{M}$.

2 Examples of risk functionals and their CxLS properties

We first present some common examples of one-dimensional risk functionals with or without CxLS. These examples will help to understand the main concepts in this paper, and they will be referred to repeatedly throughout. We start with a few interesting common quantities that have CxLS. They are, in fact, increasing Choquet integrals; see Section 3. A full characterization of all signed Choquet integrals with CxLS will be given in Theorem 3.2 below.

Example 2.1. (i) The expectation:

$$\mathbb{E}[F] = \int_{-\infty}^{\infty} x \, \mathrm{d}F(x), \quad F \in \mathcal{M}_1.$$

Note that in this paper we define $\mathbb{E}$ on the set of distributions $\mathcal{M}_1$ instead of the set of integrable random variables.

(ii) The left-quantile, or the Value-at-Risk (VaR): For $p \in (0, 1]$, define

$$\mathrm{VaR}_p(F) = \inf\{x \in \mathbb{R} : F(x) \geq p\}, \quad F \in \mathcal{M}_0.$$

In addition, the essential supremum functional is defined as ess-sup $= \mathrm{VaR}_1$.

(iii) The right-quantile: For $p \in [0, 1)$, define

$$\mathrm{VaR}_p^+(F) = \inf\{x \in \mathbb{R} : F(x) > p\}, \quad F \in \mathcal{M}_0.$$

In addition, let ess-inf $= \mathrm{VaR}_0^+$.

(iv) The mixed-quantile: For $p \in (0, 1)$ and $c \in [0, 1]$, define

$$\mathrm{VaR}_p^c = c\mathrm{VaR}_p^+ + (1 - c)\mathrm{VaR}_p.$$

In addition, we include the cases $p = 0$ and $p = 1$ by letting $\mathrm{VaR}_1^c = \mathrm{VaR}_1 =$ ess-sup and $\mathrm{VaR}_0^c = \mathrm{VaR}_0^+ =$ ess-inf for all $c \in [0, 1]$.

(v) The mid-point of range:

$$\text{Mid-range}(F) = \frac{1}{2}\,\text{ess-sup}(F) + \frac{1}{2}\,\text{ess-inf}(F), \quad F \in \mathcal{M}_\infty.$$

Next, let $f : \mathbb{R} \to \mathbb{R}$ be a measurable function and $\rho : \mathcal{M} \to \mathbb{R}$, $\rho(F) = \int f \, \mathrm{d}F$. In this case, ρ has CxLS since it is linear in $F \in \mathcal{M}$, and any one-to-one transform of ρ also has CxLS by checking the definition. Common examples of such functionals include von Neumann-Morgenstern expected utilities, expected loss functionals, and moments.

Example 2.2. (i) A von Neumann-Morgenstern expected utility functional has the form $F \mapsto \int_{-\infty}^{\infty} u(x) \, \mathrm{d}F(x)$ for some increasing utility function $u : \mathbb{R} \to [-\infty, \infty)$.

(ii) Second moment:

$$F \mapsto \int_{-\infty}^{\infty} x^2 \, \mathrm{d}F(x), \quad F \in \mathcal{M}_2.$$

(iii) The entropic risk measure for $\gamma \in (0, \infty)$:

$$E_\gamma(F) = \frac{1}{\gamma} \log \left(\int_{-\infty}^{\infty} e^{-\gamma x} \, \mathrm{d}F(x) \right), \quad F \in \mathcal{M}_\infty.$$

(iv) Excess loss function: for some $k \in \mathbb{R}$,

$$F \mapsto \int_{-\infty}^{\infty} (x-k)_+ \, \mathrm{d}F(x), \quad F \in \mathcal{M}_1.$$

A lot of results on CxLS of risk measures appear in the recent literature. For definitions of these risk measures, see McNeil et al. (2015). A coherent risk measure does not have CxLS on $\mathcal{M}_\infty$ unless it is an expectile (Corollary 4.6 of Ziegel (2016)). A convex risk measure has CxLS on $\mathcal{M}_\infty$ if and only if it is a utility-based shortfall risk measure in Definition 4.112 of Föllmer and Schied (2016) (Theorem 3.10 of Delbaen et al. (2016); this includes Examples 2.2 (iii) and 2.3 (ii)). In particular, the Expected Shortfall (ES)[1] does not have CxLS, as noted by Gneiting (2011).

Example 2.3. (i) The Expected Shortfall (ES) for $p \in (0,1)$, defined as

$$\mathrm{ES}_p(F) = \frac{1}{1-p} \int_p^1 \mathrm{VaR}_t(F) \, \mathrm{d}t, \quad F \in \mathcal{M}_1,$$

is a coherent risk measure, and it does not have CxLS. The following famous ES-VaR relation of Rockafellar and Uryasev (2002) will be used repeatedly in this paper.

$$\begin{cases} [\mathrm{VaR}_p(F), \mathrm{VaR}_p^+(F)] = \underset{x \in \mathbb{R}}{\arg\min} \left\{ x + \dfrac{1}{1-p} \displaystyle\int_{-\infty}^{\infty} (y-x)_+ \, \mathrm{d}F(y) \right\}; \\ \mathrm{ES}_p(F) = \underset{x \in \mathbb{R}}{\min} \left\{ x + \dfrac{1}{1-p} \displaystyle\int_{-\infty}^{\infty} (y-x)_+ \, \mathrm{d}F(y) \right\}. \end{cases} \tag{1}$$

(ii) The expectile (see e.g., Newey and Powell (1987)) for $p \in (0,1)$, defined as

$$e_p(F) = \underset{x \in \mathbb{R}}{\arg\min} \left\{ p \int_x^{\infty} (y-x)^2 \, \mathrm{d}F(y) + (1-p) \int_{-\infty}^{x} (y-x)^2 \, \mathrm{d}F(y) \right\}, \quad F \in \mathcal{M}_2,$$

is a coherent risk measure, and it has CxLS since it is a shortfall risk measure.

A deviation measure, such as the variance or the standard deviation, is a functional D that satisfies $D(\delta_c) = 0$ for a constant c and $D(\mu) > 0$ for $\mu \in \mathcal{M}$ not a point mass. For a general theory on deviation measures and measures of variability, see Bickel and Lehmann (1976), Rockafellar et al. (2006) and Furman et al. (2017). Deviation measures generally cannot have CxLS, as easily seen from the following argument. For any $x, y \in \mathbb{R}$, $x \neq y$, by definition, $\mathcal{D}(\lambda\delta_x + (1-\lambda)\delta_y) > 0 = \mathcal{D}(\delta_x) = \mathcal{D}(\delta_y)$ for $\lambda \in (0,1)$. This implies that $\mathcal{D}$ does not have CxLS on $\mathcal{M}$. We summarize this observation in the following proposition. Nevertheless, later in Section 4 we shall see that deviation measures may have CxLS when they are paired with other risk functionals.

[1] ES is not a utility-based shortfall risk measure, despite its name.

Proposition 2.4. *Let $\mathcal{M}$ be a set that contains all point masses and two-point distributions. A deviation measure does not have CxLS on $\mathcal{M}$.*

Example 2.5. As shown in Proposition 2.4 above, the following deviation measures do not have CxLS on their domains.

(i) The variance $\mathrm{Var}(F) = \int (x - \mathbb{E}[F])^2 \,\mathrm{d}F(x)$, $F \in \mathcal{M}_2$.

(ii) The standard deviation $\mathrm{SD}(F) = \sqrt{\mathrm{Var}(F)}$, $F \in \mathcal{M}_2$.

(iii) The mean absolute deviation from the median (median deviation, MD):

$$\mathrm{MD}(F) = \min_{x\in\mathbb{R}} \int_{-\infty}^{\infty} |y - x| \,\mathrm{d}F(y) = \int_{-\infty}^{\infty} \left(\left| y - F^{-1}\left(\frac{1}{2}\right) \right| \right) \mathrm{d}F(y), \quad F \in \mathcal{M}_1.$$

(iv) The Gini deviation:

$$\mathrm{Gini}(F) = \int_0^1 F^{-1}(t)(2t-1)\,\mathrm{d}t, \quad F \in \mathcal{M}_1.$$

(v) The range:

$$\mathrm{Range}(F) = \text{ess-sup}(F) - \text{ess-inf}(F), \quad F \in \mathcal{M}_\infty.$$

Finally, we give a few other notable functionals with or without CxLS.

Example 2.6. (i) The mode functional Mod is defined as $\mathrm{Mod}(F) = \arg\max_{x\in\mathbb{R}} \frac{\mathrm{d}}{\mathrm{d}x} F(x)$ on $\mathcal{M}$ which is the set of distributions with a unique mode. Then Mod has CxLS on $\mathcal{M}$ by definition.

(ii) The skewness functional SK is defined as $\mathrm{SK}(F) = \int \left(\frac{x-\mu}{\sigma}\right)^3 \mathrm{d}F(x)$ for $F \in \mathcal{M}_3$ with $\mathrm{SD}(F) > 0$, where $\mu = \mathbb{E}[F]$ and $\sigma = \mathrm{SD}(F)$. By definition, we can calculate $\mathrm{SK}\left(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_0\right) = \mathrm{SK}\left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_2\right) = 0$, and

$$\mathrm{SK}\left(\frac{2}{3}\left(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_0\right) + \frac{1}{3}\left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_2\right)\right) = \mathrm{SK}\left(\frac{1}{3}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{6}\delta_2\right) = 1 \neq 0.$$

The skewness functional does not have CxLS on its domain.

3 Main results in dimension one

3.1 Signed Choquet integrals

We first define signed Choquet integrals, a popular class of functionals in risk management and statistics and the main object of this section. Let

$$\mathcal{H} = \{h : h \text{ maps } [0,1] \text{ to } \mathbb{R},\ h \text{ is of bounded variation and } h(0) = 0\}.$$

Definition 3.1. A *signed Choquet integral* $I_h : \mathcal{M} \to \mathbb{R}$ is defined as

$$I_h(F) = \int_{-\infty}^{0} (h(1 - F(x)) - h(1))\,\mathrm{d}x + \int_{0}^{\infty} h(1 - F(x))\,\mathrm{d}x,$$

where $h \in \mathcal{H}$ and $\mathcal{M}$ is a convex subset of $\mathcal{M}_0$ such that I_h is well-defined. The function h is called the *distortion function* of I_h. An increasing (resp. decreasing) Choquet integral is a signed Choquet integral with an increasing (resp. decreasing) distortion function.

Signed Choquet integrals are studied extensively in the literature; for their axiomatic characterization and economic interpretation, we refer to Yaari (1987), De Waegenaere and Wakker (2001), Kou and Peng (2016) and Wang et al. (2020). A functional on $\mathcal{M}$ is a signed Choquet integral if and only if it is comonotonic-additive and satisfies a continuity condition (Theorem 1 of Wang et al. (2020)); see Proposition 5.2 below for a precise statement of this characterization in multi-dimension. Many examples of signed Choquet integrals are listed in Section 2. In particular, increasing Choquet integrals I_h with $h(1) = 1$ are also known as distortion risk measures, which include the mean, the quantiles, the mid-point of range, and ES. Many measures of variability also belong to the class of signed Choquet integrals, such as the median deviation, the Gini deviation and the range; see Wang et al. (2020) for their distortion functions.

3.2 Characterization of signed Choquet integrals with CxLS

As we have seen in Section 2, some signed Choquet integrals have CxLS whereas some others do not. The main result in this section characterizes all signed Choquet integrals with CxLS. It turns out that the following three forms of h in the subsets $\mathcal{H}_1^*, \mathcal{H}_2^*$ and $\mathcal{H}_3^*$ of $\mathcal{H}$ are important for the CxLS property.

(i) $h \in \mathcal{H}_1^*$: For some $c \in [0, 1]$, $h(t) = ch(1)\mathbb{1}_{\{0<t<1\}} + h(1)\mathbb{1}_{\{t=1\}}$, $t \in [0, 1]$. In this case,

$$I_h(F) = h(1)(c\,\text{ess-sup}(F) + (1 - c)\,\text{ess-inf}(F)),\ \ F \in \mathcal{M}. \tag{2}$$

(ii) $h \in \mathcal{H}_2^*$: $h(t) = th(1)$, $t \in [0, 1]$. In this case,

$$I_h(F) = h(1)\mathbb{E}[F],\ \ F \in \mathcal{M}. \tag{3}$$

(iii) $h \in \mathcal{H}_3^*$: For some $\alpha \in (0, 1)$ and $c \in [0, 1]$, $h(t) = ch(1)\mathbb{1}_{\{t=\alpha\}} + h(1)\mathbb{1}_{\{t>\alpha\}}$, $t \in [0, 1]$. In this case,

$$I_h(F) = h(1)\mathrm{VaR}^c_{1-\alpha}(F),\ \ F \in \mathcal{M}. \tag{4}$$

We also denote by $\mathcal{H}^* = \bigcup_{i=1}^{3} \mathcal{H}_i^*$.

For increasing Choquet integrals I_h with $h(1) = 1$, Kou and Peng (2016) show that only the above three cases are possible for I_h to have CxLS. Without monotonicity, the class of signed Choquet integrals is much larger than the class of increasing ones. Nevertheless, in the next theorem, we show that the only possible signed Choquet integrals with CxLS are still the ones listed above.

Theorem 3.2. *Let $\mathcal{M}$ be a convex set that contains all three-point distributions. A signed Choquet integral I_h has CxLS on $\mathcal{M}$ if and only if $h \in \mathcal{H}^*$.*

Comparing Theorem 3.2 with the characterization of increasing Choquet integrals with CxLS in Kou and Peng (2016), we see that removing monotonicity does not lead to many more choices of functionals with CxLS. More precisely, all functionals in (2)-(4) have either an increasing or decreasing distortion function, and they are monotone with respect to the usual stochastic order (e.g., Lemma 2 of Wang et al. (2020)).

Corollary 3.3. *If I_h has CxLS on $\mathcal{M}_{\mathrm{dis}}$, then it is monotone (i.e., increasing or decreasing).*

Corollary 3.3 is a surprising result, as the CxLS property by definition is not related to monotonicity. For instance, for any measurable function $f : \mathbb{R} \to \mathbb{R}$, the mapping $\rho : F \mapsto \int f(x)\,\mathrm{d}F(x)$ has CxLS; whether f is monotone (i.e., whether ρ is monotone with respect to stochastic order) is irrelevant to the CxLS property. Among the class of signed Choquet integrals, however, the CxLS property surprisingly implies monotonicity. On the other hand, Lemma 14 and Theorem 15 of Steinwart et al. (2014) yield that, under some regularity conditions, CxLS implies quasi-linearity. According to the example of ρ above, quasi-linearity is not directly related to monotonicity.

The proofs for signed Choquet integrals are much more involved than the case of increasing ones, due to the lack of monotonicity. Two new technical lemmas are needed for a proof of Theorem 3.2, which are put in the Appendix.

In risk management practice, one may be only interested in backtestability or elicitability over the set of continuous distribution models. In the following we will show that, when constrained on the set of continuous distributions, the only possible choices of h to allow for the CxLS property of I_h are still the three cases in Theorem 3.2. In particular, we can show that CxLS on $\mathcal{M}_{\mathrm{con}}$ implies CxLS on $\mathcal{M}_{\mathrm{dis}}$ for I_h, and as a consequence of Theorem 3.2, we know $h \in \mathcal{H}^*$.

Proposition 3.4. *For $h \in \mathcal{H}$, if I_h has CxLS on $\mathcal{M}_{\mathrm{con}}$, then I_h has CxLS on $\mathcal{M}_{\mathrm{dis}}$, and, as a consequence, $h \in \mathcal{H}^*$.*

3.3 CxLS under risk transforms

In this section, we discuss the impact of risk transforms on the CxLS property of a functional. We focus on a special class of transforms; for a comprehensive treatment of risk transforms, see Liu et al. (2020). For a measurable function v (which is typically monotone in applications), a v-*transform* maps the distribution of a random variable X to the distribution of $v(X)$. In other words, the probability measure F is transformed to $F \circ v^{-1}$, where v^{-1} is the set-valued inverse of v. For a functional ϕ well defined on $\mathcal{M}^v = \{F \circ v^{-1} : F \in \mathcal{M}\}$, let ϕ_v be defined as

$$\phi_v(F) = \phi(F \circ v^{-1}), \quad F \in \mathcal{M}. \tag{5}$$

In the following result, we show an invariance property of CxLS under v-transforms.

Proposition 3.5. *For any measurable function v and $\mathcal{M} \subset \mathcal{M}_0$, ϕ_v in (5) has CxLS on $\mathcal{M}$ if ϕ has CxLS on $\mathcal{M}^v$, and the converse is true if v is an injection.*

Note that the converse implication in Proposition 3.5 does not hold in general if v is not injective. For instance, if $v(x) = 0$ for all $x \in \mathbb{R}$, then $\phi_v(F) = \phi(\delta_0)$ for all $F \in \mathcal{M}$, and it has CxLS regardless of the choice of ϕ.

Risk functionals based on v-transforms are widely used in finance. For instance, an expected utility, a certainty equivalent, or a rank-dependent expected utility involves v-transforms via utility or loss functions, and a pricing functional for options can be seen as the (risk-neutral) expectation of a transformed asset price distribution. The loss-based risk measures ρ of Cont et al. (2013) are defined via the loss-dependence property: $\rho(F) = \rho(F \circ u^{-1})$ for all $F \in \mathcal{M}$, where $u(x) = x_+$ for all $x \in \mathbb{R}$. Equivalently, $\rho = \phi_u$ for some $\phi : \mathcal{M} \to \mathbb{R}$ (which may be trivially chosen as ρ itself). Many examples of loss-based risk measures in Cont et al. (2013) are obtained by letting $\rho = \phi_u$ and choosing ϕ satisfying some desirable properties such as monotonicity, cash-additivity, and convexity.

Let $\mathcal{M}_\infty^+ \subset \mathcal{M}_\infty$ be the set of compactly supported distributions on $\mathbb{R}_+ = [0, \infty)$. Using Proposition 3.5, we know that if ϕ has CxLS on $\mathcal{M}_\infty^+$, then the loss-based risk measure $\rho = \phi_u$ has CxLS on $\mathcal{M}_\infty$. The converse may not be true since $u : x \mapsto x_+$ is not an injection. However, since u is injective on $\mathbb{R}_+$, we know that if ρ has CxLS on $\mathcal{M}_\infty^+$, then ϕ has CxLS on $\mathcal{M}_\infty^+$. Moreover, if ϕ is cash-additive, then ϕ also has CxLS on $\mathcal{M}_\infty$. Recall that a functional $\rho : \mathcal{M} \to \mathbb{R}$ is cash-additive if $\rho(F \circ t_c) = \rho(F) - c$ for all $c \in \mathbb{R}$, where $t_c : x \mapsto x + c$ is a constant shift function. We summarize these findings in the following corollary.

Corollary 3.6. *Let $u(x) = x_+$ for $x \in \mathbb{R}$ and $\phi : \mathcal{M}_\infty \to \mathbb{R}$ be cash-additive. The loss-based risk measure $\rho = \phi_u$ has CxLS on $\mathcal{M}_\infty$ if and only if ϕ has CxLS on $\mathcal{M}_\infty$.*

Corollary 3.6 can be directly applied to characterize CxLS of many loss-based risk measures together with Theorem 3.2, as most examples of Cont et al. (2013) are generated by choosing a cash-additive ϕ, such as a convex risk measure or an increasing Choquet integral with $\phi(1) = 1$.

Example 3.7 (Certainty equivalents)**.** An example of loss-based risk measures that does not fit in Corollary 3.6 is the class of loss certainty equivalents (Example 2.9 of Cont et al. (2013)). A loss certainty equivalent ρ defined as, for a strictly increasing function $v : \mathbb{R} \to \mathbb{R}$,

$$\rho(F) = v^{-1}\left(\int_0^1 v((F^{-1}(t))_+)\,\mathrm{d}t\right), \quad F \in \mathcal{M}.$$

It is clear that ρ above always has CxLS, since it is a strictly monotone transform of an expected loss in Example 2.2. Similarly, the classic certainty equivalent, defined as the mapping $F \mapsto v^{-1}(\int_0^1 v(F^{-1}(t))\,\mathrm{d}t)$ for a strictly increasing loss function v, also has CxLS.

4 CxLS in multi-dimension

4.1 Some general properties

In this section, we provide some simple results for CxLS in multi-dimension, which will be useful later to show our main characterization result in dimension two. These results are easy to verify, and self-contained proofs are provided in the Appendix for completeness. First we present the straightforward fact that a d-dimensional functional has CxLS if it is an injective function of some one-dimensional functionals with CxLS. Since one-dimensional functionals with CxLS are well studied, this result gives a convenient way to construct simple functionals with CxLS in multi-dimension.

Proposition 4.1. *A functional $\rho : \mathcal{M} \to \mathbb{R}^d$ has CxLS on $\mathcal{M}$ if it is an injective function of some other functionals with CxLS on $\mathcal{M}$.*

As a direct consequence of Proposition 4.1, a functional $\rho : \mathcal{M} \to \mathbb{R}^d$ has CxLS on $\mathcal{M}$ if each of its components has CxLS on $\mathcal{M}$. Another simple implication is that if (ρ_1, ρ_2) and (ρ_1, ρ_3) has CxLS, then so does the combined functional (ρ_1, ρ_2, ρ_3); here each of ρ_1, ρ_2, ρ_3 may be multi-dimensional.

The next proposition establishes a link between a two-dimensional functional with CxLS and its components, assuming one of the component already has CxLS.

Proposition 4.2. *Let ρ_1 and ρ_2 be two functionals from $\mathcal{M}$ to $\mathbb{R}$ such that ρ_2 has CxLS on $\mathcal{M}$. The pair of functionals (ρ_1, ρ_2) has CxLS on $\mathcal{M}$ if and only if ρ_1 has CxLS on $\mathcal{M}(r)$ for all $r \in \mathbb{R}$, where $\mathcal{M}(r) = \{F \in \mathcal{M} : \rho_2(F) = r\}$.*

Proposition 4.2 is very useful to prove or disprove the CxLS property of two-dimensional functionals. For instance, it would justify Propositions 4.3 and 4.4 below, and it is used repeatedly in the proofs of Theorems 5.3 and 6.9.

4.2 Risk quadrangle and the Bayes risk

Many one-dimensional risk functionals are connected via a *risk quadrangle* proposed by Rockafellar and Uryasev (2013). A risk quadrangle consists of five functionals that depend on each other: the statistic $\mathcal{S}$, the risk $\mathcal{R}$, the regret $\mathcal{V}$, the deviation $\mathcal{D}$, and the error $\mathcal{E}$. We briefly explain their relationship below following the setup of Rockafellar and Uryasev (2013), albeit using our convention of defining functionals on a set $\mathcal{M} \subset \mathcal{M}_1$. As in Section 3.3, define the constant shift function $t_c : x \mapsto x + c$ for $c \in \mathbb{R}$; in other words, $F \circ t_c$ is the distribution of F shifted to the left by c. For a given $\mathcal{V} : \mathcal{M} \to \mathbb{R}$, the following functionals on $\mathcal{M}$ satisfy:

(i) $\mathcal{E}(F) = \mathcal{V}(F) - \mathbb{E}[F]$;

(ii) $\mathcal{S}(F) = \arg\min_{c\in\mathbb{R}}\{c + \mathcal{V}(F \circ t_c)\} = \arg\min_{c\in\mathbb{R}}\{\mathcal{E}(F \circ t_c)\}$;

(iii) $\mathcal{R}(F) = \min_{c\in\mathbb{R}}\{c + \mathcal{V}(F \circ t_c)\}$;

(iv) $\mathcal{D}(F) = \mathcal{R}(F) - \mathbb{E}[F]$.

The most popular category of risk quadrangles is the expectation quadrangles, in which $\mathcal{V}$ is given by $\mathcal{V}(F) = \int_{\mathbb{R}} v(x)\,\mathrm{d}F(x)$ for a function v on $\mathbb{R}$; see the examples in Section 4.3. It turns out that, although the risk functionals $\mathcal{R}$ and $\mathcal{D}$ in an expectation quadrangle may not have CxLS by themselves, the multi-dimensional CxLS property holds if they are combined with $\mathcal{S}$.

Proposition 4.3. *In an expectation quadrangle, each of the following quantities, if its components are well defined from $\mathcal{M}$ to $\mathbb{R}$, has CxLS: $\mathcal{V}$, $\mathcal{E}$, $\mathcal{S}$, $(\mathcal{S}, \mathcal{R})$, $(\mathcal{S}, \mathcal{D})$, and any combined functional of the above, such as $(\mathcal{S}, \mathcal{R}, \mathcal{D})$ or $(\mathcal{V}, \mathcal{E}, \mathcal{S}, \mathcal{R}, \mathcal{D})$.*

In Proposition 4.3, we need $\mathcal{M} \subset \mathcal{M}_1$ as required in any risk quadrangle. Proposition 4.4 below provides a more general result than Proposition 4.3 by addressing the minimization of $\int_0^1 L(c, x)\,\mathrm{d}F(x)$ for a general function L on $\mathbb{R}^2$, and $\mathcal{M}$ is not necessarily a subset of $\mathcal{M}_1$. This result is closely related to the study of the Bayes risk elicitation by Frongillo and Kash (2018).

Proposition 4.4. *Suppose $\mathcal{M} \subset \mathcal{M}_0$. For some function $L : \mathbb{R}^2 \to \mathbb{R}$, if a pair of functionals $(\gamma, \Gamma) : \mathcal{M} \to \mathbb{R}^2$ is given by*

$$\gamma : F \mapsto \min_{c\in\mathbb{R}} \int_{\mathbb{R}} L(c, x)\,\mathrm{d}F(x) \quad \textit{and} \quad \Gamma : F \mapsto \arg\min_{c\in\mathbb{R}} \int_{\mathbb{R}} L(c, x)\,\mathrm{d}F(x),$$

then Γ *and* (γ,Γ) *both have CxLS.*

In the setting of Proposition 4.4, γ is called the Bayes risk of Γ by Frongillo and Kash (2018). Indeed, Corollary 1 of Frongillo and Kash (2018) shows that (γ,Γ) is elicitable, a sufficient condition for CxLS (see Section 6). Our self-contained proofs of Propositions 4.3-4.4 are obtained from direct arguments based on Proposition 4.2.

4.3 Examples

We present some examples of multi-dimensional functionals with CxLS. Since a functional that has all components with CxLS automatically has CxLS, we only give examples where at least one of the components does not have CxLS. All examples listed in Example 4.5 belong to some risk quadrangle, although (iii) is not in an expectation quadrangle.

Example 4.5 (Two-dimensional examples)**.**

(i) $(\mathrm{VaR}_p,\mathrm{ES}_p)$ for $p\in(0,1)$ has CxLS on $\mathcal{M}_1$. Indeed, $(\mathrm{VaR}_p,\mathrm{ES}_p)=(\mathcal{S},\mathcal{R})$ in the expectation quadrangle with $v(x)=x_+/(1-p)$; see the VaR-ES relationship (1) and Example 2 of Rockafellar and Uryasev (2013).

(ii) $(\mathrm{Median},\mathrm{MD})$ has CxLS on $\mathcal{M}_1$, where $\mathrm{Median}=\mathrm{VaR}_{1/2}$. Indeed, $(\mathrm{Median},\mathrm{MD})=(\mathcal{S},\mathcal{D})$ in the expectation quadrangle with $v(x)=2x_+$; see Example 3 of Rockafellar and Uryasev (2013).

(iii) $(\text{Mid-range},\text{Range})$ has CxLS on $\mathcal{M}_\infty$ by Proposition 4.1, since the pair is a bijection from $(\text{ess-inf},\text{ess-sup})$, which has components with CxLS. We note that $(\text{Mid-range},\text{Range})=(\mathcal{S},\mathcal{D})$ in Example 4 of Rockafellar and Uryasev (2013). This risk quadrangle is not an expectation quadrangle, and one can check that $\mathcal{V}=\mathbb{E}+\max(|\,\text{ess-sup}\,|,|\,\text{ess-inf}\,|)$ does not have CxLS; thus Proposition 4.3 does not apply.

(iv) $(\mathbb{E},\mathrm{Var})$ and $(\mathbb{E},\mathrm{SD})$ both have CxLS on $\mathcal{M}_2$ by Proposition 4.1, since each of them is a bijection from the pair of the first two moments, which has components with CxLS. Moreover, $(\mathbb{E},\mathrm{Var})=(\mathcal{S},\mathcal{D})$ in the expectation quadrangle with $v(x)=x+x^2$; see Examples 1 and 1' of Rockafellar and Uryasev (2013).

(v) The pair (e_p,var_p) for $p\in(0,1)$ has CxLS on $\mathcal{M}_2$, where var_p is the variantile functional defined as

$$\mathrm{var}_p(F)=\min_{c\in\mathbb{R}}\left\{p\int_c^\infty(x-c)^2\,\mathrm{d}F(x)+(1-p)\int_{-\infty}^c(x-c)^2\,\mathrm{d}F(x)\right\},\quad F\in\mathcal{M}_2,$$

which is the Bayes risk for the corresponding expectile e_p, as noted by Frongillo and Kash (2018). Moreover, $(e_p, \mathrm{var}_p) = (\mathcal{S}, \mathcal{D})$ in the expectation quadrangle with $v(x) = x + p(x_+)^2 + (1-p)(x_-)^2$.

Example 4.6 (Three-dimensional examples)**.**

(i) The Range-Value-at-Risk (RVaR) is a signed Choquet integral with distortion function $h(t) = \frac{(t-1+q)_+}{q-p} \wedge 1,\ t \in [0,1]$, where $0 < p < q < 1$; see Cont et al. (2010) and Embrechts et al. (2018). An RVaR does not have CxLS because its distortion function does not belong to the cases in Theorem 3.2. From its definition, we can alternatively write

$$\mathrm{RVaR}_{p,q}(F) = \frac{1}{q-p}((1-p)\mathrm{ES}_p(F) - (1-q)\mathrm{ES}_q(F)),\ F \in \mathcal{M}_1,$$

and therefore the triplet $(\mathrm{VaR}_p, \mathrm{VaR}_q, \mathrm{RVaR}_{p,q})$ has CxLS, since $\mathrm{RVaR}_{p,q}$ is linear on the set of distributions with fixed VaR_p and VaR_q. A study on elicitability of RVaR can be found in a recent paper Fissler and Ziegel (2019).

(ii) Recall that the skewness functional SK in (ii) of Example 2.6 does not have CxLS. The triplet $(\mathbb{E}, \mathrm{SD}, \mathrm{SK})$ has CxLS on $\mathcal{M} = \{F \in \mathcal{M}_3 : \mathrm{SD}(F) > 0\}$, by using Proposition 4.1, since the triplet is a bijection from the triplet of the first three moments on $\mathcal{M}$, which has components with CxLS.

5 Main characterization result in dimension two

5.1 Multi-dimensional signed Choquet integrals

In this section we investigate the CxLS property in higher dimension. It is natural to define the signed Choquet integral in dimension $d \geq 2$ as follows.

Definition 5.1. Let $\mathbf{h} = (h_1, \ldots, h_n)$, where each $h_i \in \mathcal{H},\ i = 1, \ldots, d$. A *signed Choquet integral* $I_{\mathbf{h}} : \mathcal{M} \to \mathbb{R}^d$ is defined as

$$I_{\mathbf{h}}(F) = (I_{h_1}(F), \ldots, I_{h_d}(F))\,, \tag{6}$$

where each $I_{h_i},\ i = 1, \ldots, d$, is the one-dimensional signed Choquet integral defined in Definition 3.1.

Signed Choquet integrals in multi-dimension share a similar characterization via *comonotonic-additivity*. Random variables X and Y defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are said to be *comonotonic* if there exists $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that $\omega, \omega' \in \Omega_0$,

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geq 0.$$

Let $\mathcal{X}_{\mathcal{M}}$ be the set of random variables in $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution in $\mathcal{M}$. For a functional $\rho : \mathcal{M} \to \mathbb{R}^d$, we define $\hat{\rho} : \mathcal{X}_{\mathcal{M}} \to \mathbb{R}^d$ via $\hat{\rho}(X) = \rho(F)$, where F is the distribution of X. We say that ρ is *comonotonic-additive* if $\hat{\rho}(X + Y) = \hat{\rho}(X) + \hat{\rho}(Y)$ for any comonotonic random variables $X, Y \in \mathcal{X}$.

Recall that we use $\mathcal{M}_\infty$ to denote the set of distributions of bounded random variables. We define continuity on $\mathcal{M}_\infty$ via the metric w given by $w(F, G) = \sup_{t \in [0,1]} |F^{-1}(t) - G^{-1}(t)|$, known as the Wasserstein-L^∞ metric. For a sequence $\{F_n\}_{n \in \mathbb{N}} \subset \mathcal{M}_\infty$, we write $F_n \xrightarrow{w} F$ if the sequence $\{F_n\}_{n \in \mathbb{N}}$ converges to F in the metric w. By applying Theorem 1 of Wang et al. (2020) to each component of I, and noting that the w-continuity of ρ is equivalent to L^∞-continuity of $\hat{\rho}$, we obtain the following characterization of multi-dimensional signed Choquet integrals.

Proposition 5.2. *A functional $I : \mathcal{M}_\infty \to \mathbb{R}^d$ is comonotonic-additive and uniformly continuous with respect to w if and only if I is a d-dimensional signed Choquet integral.*

Clearly, for $h_1, h_2 \in \mathcal{H}^*$, the two-dimensional signed Choquet integral (I_{h_1}, I_{h_2}) has CxLS, due to Proposition 4.1. More interestingly, (I_{h_1}, I_{h_2}) may have CxLS even if it is not an injection from one-dimensional signed Choquet integrals with CxLS. A famous example is $(\mathrm{VaR}_p, \mathrm{ES}_p)$ for $p \in (0, 1)$ as in Example 4.5, as shown by Fissler and Ziegel (2016) and Acerbi and Szekely (2017).

Characterization of multi-dimensional signed Choquet integrals with CxLS seems to be an extremely challenging task, which is left as an open question by Fissler and Ziegel (2016) and Kou and Peng (2016). Since VaR_p is a canonical candidate for a one-dimensional signed Choquet integral with CxLS, below we explore for which $h \in \mathcal{H}$, (I_h, VaR_p) has CxLS.

5.2 Characterizing a signed Choquet integral and a VaR with CxLS

Below we will characterize all pairs (I_h, VaR_p) with CxLS. To the best of our knowledge, this is the first result in the literature on characterizing CxLS in multi-dimension. By Proposition 4.1, (I_h, VaR_p) has CxLS if and only if $(I_h + a\mathrm{VaR}_p, \mathrm{VaR}_p)$ has CxLS for one (or all) $a \in \mathbb{R}$. Therefore, adding a constant times VaR_p to I_h does not change the CxLS property of (I_h, VaR_p). This explains why the term $a\mathrm{VaR}_p$ appears in all cases in the following theorem.

Theorem 5.3. *For $p \in (0,1)$ and $h \in \mathcal{H}$, (I_h, VaR_p) has CxLS on $\mathcal{M}_{\mathrm{dis}}$ if and only if I_h is one of the following cases*

(i) $I_h = a\mathrm{VaR}_p + I_{h^*}$ *for some* $a \in \mathbb{R}$ *and* $h^* \in \mathcal{H}^*$*;*

(ii) $I_h = a\mathrm{VaR}_p + b\mathbb{E} + c\mathrm{ES}_p$ *for some constants* $a, b, c \in \mathbb{R}$*;*

(iii) $I_h = a\mathrm{VaR}_p + b\mathrm{VaR}_p^+ + c\,\text{ess-sup}$ *for some constants* $a, b, c \in \mathbb{R}$ *with* $bc > 0$*;*

(iv) $I_h = a\mathrm{VaR}_p + b\mathrm{VaR}_p^+ + c\,\text{ess-inf}$ *for some constants* $a, b, c \in \mathbb{R}$ *with* $bc < 0$*.*

Theorem 5.3 reveals a characterization of all signed Choquet integrals that have CxLS jointly with VaR_p, $p \in (0,1)$. In risk management practice, one usually does not distinguish VaR_p and VaR_p^+ by (sometimes implicitly) assuming a continuous quantile at p; certainly, this assumption does not hold for all distributions in $\mathcal{M}_{\mathrm{dis}}$. If we loosely treat VaR_p and VaR_p^+ as identical, then cases (iii) and (iv) can be combined into case (i), and we are left with the following two cases.

(i) $I_h = a\mathrm{VaR}_p + I_{h^*}$ for some $a \in \mathbb{R}$ and $h^* \in \mathcal{H}^*$;

(ii) $I_h = a\mathrm{VaR}_p + b\mathbb{E} + c\mathrm{ES}_p$ for some constants $a, b, c \in \mathbb{R}$.

From (i) and (ii), if (I_h, VaR_p) is not a bijection from a pair of signed Choquet integrals with CxLS, then I_h is a linear combination of VaR_p, $\mathbb{E}$ and ES_p. Regarding the CxLS property, the class of VaR_p for $p \in (0,1)$ plays a unique role among the class of distortion risk measures (Kou and Peng (2016)) and among positively homogeneous tail risk measures (Liu and Wang (2020)). The above observation shows that ES_p is also very special regarding CxLS. In particular, as the CxLS property is necessary for elicitability (see Section 6), we will see in Theorem 6.9 that a convex combination of ES_p and $\mathbb{E}$ is the only type of comonotonic-additive coherent risk measure that is co-elicitable with VaR_p.

It may be worth noting that the roles of VaR_p and VaR_p^+ are symmetric. To get functionals I_h such that (I_h, VaR_p^+) has CxLS, one simply switches the positions in the pairs $(\mathrm{VaR}_p, \mathrm{VaR}_p^+)$ and $(\text{ess-inf}, \text{ess-sup})$ in Theorem 5.3. This statement is due to the following relation. For any distribution $F \in \mathcal{M}_0$, let $\bar{F}$ be given by $\bar{F}(A) = F(-A)$, where $-A = \{-x : x \in A\}$, $A \in \mathcal{B}(\mathbb{R})$.

Proposition 5.4. *For $p \in (0,1)$ and $h \in \mathcal{H}$, (I_h, VaR_p) has CxLS on $\mathcal{M}$ if and only if $(I_{\bar{h}}, \mathrm{VaR}_{1-p}^+)$ has CxLS on $\bar{\mathcal{M}}$, where $\bar{h} \in \mathcal{H}$ is given by $\bar{h}(t) = h(1-t) - h(1)$, $t \in [0,1]$ and $\bar{\mathcal{M}} = \{\bar{F} : F \in \mathcal{M}\}$.*

Remark 5.5. One may wonder whether the characterization of I_h in Theorem 5.3 also holds if $\mathcal{M}_{\mathrm{dis}}$ is replaced by a different set, such as $\mathcal{M}_{\mathrm{con}}$ in Proposition 3.4, or the set $\mathcal{M}_0^*(p) = \{F \in \mathcal{M}_0 :$

F^{-1} is continuous at $p\}$ on which VaR_p is elicitable (see Section 6). Unfortunately, we do not have a definite answer to the above question, as the current techniques used in Proposition 3.4 cannot be generalized to the case of (I_h, I_g) or (I_h, VaR_p) for general $h, g \in \mathcal{H}$. Nevertheless, as a consequence of Lemma A.5 in the Appendix, if h is additionally assumed increasing and concave, then CxLS of (I_h, VaR_p) on $\mathcal{M}_0^*(p)$ indeed implies CxLS on $\mathcal{M}_{\mathrm{dis}}$. This result will be useful in the proof of Theorem 6.9 below.

6 Backtestability, elicitability and identifiability

This section gives formal definitions of elicitability, backtestability and identifiability as studied by Osband (1985), Gneiting (2011) and Acerbi and Szekely (2017), and discusses their relation with the CxLS property. As the main focus of this paper is the CxLS property, this section collects some major relevant facts for the interested reader with self-contained proofs, and we refer to Fissler and Ziegel (2016), Kou and Peng (2016) and Acerbi and Szekely (2017) for excellent summaries and detailed discussions on the implications in statistical inference, risk management, and banking regulation. In addition to the known results, new results in this section include a characterization on co-elicitability of coherent risk measures with VaR_p (Theorem 6.9), and a corresponding score function (Proposition 6.10).

6.1 Definitions and known results

Elicitability refers to the existence of a scoring function for the forecasted value of a risk functional and realized value of future observations, so that the mean of the scoring function attains its minimum value if and only if the value of the risk functional is truly forecasted; see Gneiting (2011) for elicitability in a decision-theoretic framework. Comparative backtests, for which elicitability is a necessary condition, are discussed by Nolde and Ziegel (2017) as an alternative to the traditional backtests.

Definition 6.1 (Elicitability)**.** The functional $\rho : \mathcal{M} \to \mathbb{R}^d$ is $\mathcal{M}$-*elicitable* if there exists a *strictly consistent scoring function* $S : \mathbb{R}^{d+1} \to \mathbb{R}$ such that for all $F \in \mathcal{M}$,

$$\rho(F) = \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^d} \int_{-\infty}^{\infty} S(\mathbf{x}, y)\,\mathrm{d}F(y). \tag{7}$$

We also say that $\rho_1 : \mathcal{M} \to \mathbb{R}$ is *co-elicitable* with $\rho_2 : \mathcal{M} \to \mathbb{R}$ on $\mathcal{M}$ if (ρ_1, ρ_2) is $\mathcal{M}$-elicitable.

In the literature, elicitability is often defined for set-valued risk functionals (e.g., generally, quantiles are interval-valued), as the minimizer to the scoring function is not necessarily a singleton.

In this paper, as we look at risk functionals mapping $\mathcal{M}$ to $\mathbb{R}^d$, we use the slightly narrower definition on $\mathbb{R}^d$-valued functionals. This choice of definition does not affect our discussion.

Next, identifiability refers to the existence of an identification function for the forecasted value of a risk functional and realized value of future observations. The mean of the identification function is zero if and only if the value of the risk functional is truly forecasted. Therefore, the realized average value of the identification function can identify whether a risk forecast is accurate.

Definition 6.2 (Identifiability)**.** A functional $\rho : \mathcal{M} \to \mathbb{R}^d$ is said to be $\mathcal{M}$-*identifiable*, if there exists an *identification function* $I : \mathbb{R}^{d+1} \to \mathbb{R}$ such that for all $F \in \mathcal{M}$ and all $\mathbf{x} \in \mathbb{R}^d$,

$$\int_{-\infty}^{\infty} I(\mathbf{x}, y)\,\mathrm{d}F(y) = 0 \quad \text{if and only if} \quad \mathbf{x} = \rho(F).$$

Finally, we define backtestability as in Acerbi and Szekely (2017). Backtestability refers to the existence of a backtest function, whose mean reports positive value if the risk functional is under-forecasted, and negative value if the risk functional is over-forecasted. Thus, the realized average value of this backtest functional can distinguish between under- and over-estimation in risk forecasts. Moreover, we require the value of the backtest function to be strictly increasing in the prediction to assess the deviation of the prediction from the true value. Due to the lack of a natural order in $\mathbb{R}^d$, one cannot speak of under-estimation or over-estimation for $\mathbb{R}^d$-valued risk functionals. Therefore, the notion of backtestability is suitable for dimension one only (a related notion for multi-dimensional functionals is *ridge backtests*; see Acerbi and Szekely (2017)).

Definition 6.3 (Backtestability)**.** A functional $\rho : \mathcal{M} \to \mathbb{R}$ is said to be $\mathcal{M}$-*backtestable*, if there exists a *backtest function* $Z : \mathbb{R}^2 \to \mathbb{R}$ such that for all $F \in \mathcal{M}$ and all $x \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} Z(x, y)\,\mathrm{d}F(y) = 0 \quad \text{if and only if} \quad x = \rho(F),$$

and $\int Z(x, \cdot)\,\mathrm{d}F$ is strictly increasing in the prediction x.

The three notions introduced above are model-free in the sense that the statements holds for all $F \in \mathcal{M}$, that is, in order to compare scores, to identify forecasts, or to quantify backtests, one does not need to know the underlying distribution that generates the random observations.

In what follows, we illustrate the relationship among the above three concepts and the CxLS property. First, in dimension one, identifiability follows directly from backtestability, and backtestability is generally stronger than elicitability. In any dimension, both elicitability and identifiability imply the CxLS property. Finally, for one-dimensional signed Choquet integrals, CxLS is sufficient for backtestability except for the case of $h \in \mathcal{H}_1^*$. The above statements will be verified (with some conditions) below.

We shall first see that elicitability and backtestability both imply the CxLS property. Suppose that $F, G \in \mathcal{M}$ satisfy $\rho(F) = \rho(G) = \mathbf{x} \in \mathbb{R}^d$. If ρ is $\mathcal{M}$-elicitable, let S be its scoring function in (7). As $\mathbf{x}$ is a minimizer for both $\int_{-\infty}^{\infty} S(\cdot, y)\,\mathrm{d}F(y)$ and $\int_{-\infty}^{\infty} S(\cdot, y)\,\mathrm{d}G(y)$, it must also be a minimizer for $\int_{-\infty}^{\infty} S(\cdot, y)\,\mathrm{d}(\lambda F + (1-\lambda)G)(y)$. By definition of elicitability, $\mathbf{x} = \rho(\lambda F + (1-\lambda)G)$, and ρ has CxLS. Similarly, if ρ is $\mathcal{M}$-identifiable, let $I : \mathbb{R}^{d+1} \to \mathbb{R}$ is its identification function. As $\mathbf{x}$ satisfies $\int_{-\infty}^{\infty} I(\mathbf{x}, y)\,\mathrm{d}F(y) = 0$ and $\int_{-\infty}^{\infty} I(\mathbf{x}, y)\,\mathrm{d}G(y) = 0$, we know $\int_{-\infty}^{\infty} I(\mathbf{x}, y)\,\mathrm{d}(\lambda F + (1-\lambda)G) = 0$ for all $\lambda \in [0, 1]$. By the definition of identifiability, we have $\mathbf{x} = \rho(\lambda F + (1-\lambda)G)$. We summarize these simple arguments in the following proposition, which is known in the literature (e.g., Osband (1985) and Lambert et al. (2008)).

Proposition 6.4. *For $\mathcal{M} \subset \mathcal{M}_0$, if $\rho : \mathcal{M} \to \mathbb{R}^d$ is $\mathcal{M}$-elicitable or $\mathcal{M}$-identifiable, then ρ has CxLS on $\mathcal{M}$.*

On the other hand, CxLS may not be sufficient for elicitability or backtestability even in dimension one. For instance, one can check that the essential supremum functional is not elicitable or backtestable (Proposition 6.7 (i)); see also Heinrich (2014) and Brehmer and Strokorb (2019) for other functionals with CxLS that are not elicitable. Under which conditions CxLS becomes sufficient in high-dimension is an open question; see Fissler and Ziegel (2016, p. 1699).

The next proposition verifies that one-dimensional backtestability implies elicitability, as shown by Acerbi and Szekely (2017).

Proposition 6.5. *If $\rho : \mathcal{M} \to \mathbb{R}$ is $\mathcal{M}$-backtestable with backtest function Z, then ρ is $\mathcal{M}'$-elicitable, where $\mathcal{M}' = \{F \in \mathcal{M} : \int_{-\infty}^{\infty} \int_0^z |Z(x, y)|\,\mathrm{d}x\,\mathrm{d}F(y) < \infty$ for all $z \in \mathbb{R}\}$.*

By Theorem 3.2, if a signed Choquet integral I_h has CxLS, then $h \in \mathcal{H}^*$, belonging to one of the three cases (2)-(4). Hence, it suffices to analyze backtestability in these cases. Note that by Theorem 3.2, a signed Choquet integral I_h with CxLS is either a mean, a mixed-quantile, or a convex combination of ess-sup and ess-inf, multiplied by a constant equal to $h(1)$. The next proposition verifies that the constant multiplier does not affect backtestability as long as it is not zero. This result is similar to the *revelation principle* for scoring functions; see Theorem 4 of Gneiting (2011).

Proposition 6.6. *Suppose that $\rho : \mathcal{M} \to \mathbb{R}$ is backtestable with backtest function Z, then for $c \neq 0$, the functional $c\rho$ is backtestable with backtest function $Z^*(x, y) = cZ(x/c, y)$, $x, y \in \mathbb{R}$.*

For the signed Choquet integral I_h, $h \in \mathcal{H}^*$, if $h(1) = 0$, then $I_h(F) = 0$ for all $F \in \mathcal{M}$. This trivial functional is backtestable with the backtest function $Z(x, y) = x$ for all $x, y \in \mathbb{R}$. If $h(1) \neq 0$, by Proposition 6.6, the backtestability of I_h reduces to that of increasing Choquet integrals studied in Acerbi and Szekely (2017). We list them here for completeness.

Proposition 6.7. *For $h \in \mathcal{H}^*$ with $h(1) \neq 0$,*

(i) if $h \in \mathcal{H}_1^*$, I_h is not $\mathcal{M}_{\text{dis}}$-backtestable;

(ii) if $h \in \mathcal{H}_2^*$, I_h is $\mathcal{M}_1$-backtestable;

(iii) if $h \in \mathcal{H}_3^*$, I_h is $\mathcal{M}_0^*$-backtestable, where $\mathcal{M}_0^* = \{F \in \mathcal{M}_0 : F^{-1}$ is continuous$\}$.

Remark 6.8. For a given $p \in (0, 1)$, VaR_p (or VaR_p^c as in case (iii) of Proposition 6.7) is $\mathcal{M}_0^*(p)$-backtestable (and $\mathcal{M}$-elicitable), where $\mathcal{M}_0^*(p) = \{F \in \mathcal{M}_0 : F^{-1}$ is continuous at $p\}$. The choice of $\mathcal{M} = \mathcal{M}_0^*(p)$ is the biggest such that VaR_p is $\mathcal{M}$-backtestable (or $\mathcal{M}$-elicitable). This is because the minimizers of the expected scoring function for VaR_p fails to be unique if F does not admit a unique p-quantile; see the discussion at the end of Section 2 of Fissler et al. (2019b).

6.2 A new characterization result for co-elicitability

We conclude this paper by a characterization theorem on spectral risk measures that are co-elicitable with a VaR. Let $\mathcal{X}$ be the set of bounded random variables. According to Artzner et al. (1999), a functional $\hat{\rho} : \mathcal{X} \to \mathbb{R}$ is set to be a *coherent risk measure* if it is increasing, cash-additive, convex, and positively homogeneous. Translating this definition into our setting, we say that the functional $\rho : \mathcal{M}_\infty \to \mathbb{R}$ is a coherent risk measure, if $\hat{\rho}$ is a coherent risk measure in the sense of Artzner et al. (1999), where $\hat{\rho} : \mathcal{X} \to \mathbb{R}$ is given by $\hat{\rho}(X) = \rho(F)$ and F is the distribution of X.

We focus on comonotonic-additive and coherent risk measures, a popular class of one-dimensional signed Choquet integrals. Elicitability of comonotonic-additive risk measures is studied by, for instance, Ziegel (2016), Kou and Peng (2016) and Fissler and Ziegel (2016). Using the characterization result of Kusuoka (2001), a functional $\rho : \mathcal{M}_\infty \to \mathbb{R}$ is comonotonic-additive and coherent if and only if it can be written as $\rho = \int_0^1 \mathrm{ES}_p \, \mathrm{d}\mu(p)$ for a Borel probability measure μ on $[0, 1]$, or equivalently (see e.g., Theorem 3 of Wang et al. (2020)), $\rho = I_h$ for a concave and increasing $h \in \mathcal{H}$ satisfying $h(1) = 1$. Therefore, comonotonic-additive and coherent risk measures are also called *spectral risk measures* (Acerbi (2002)).

Since VaR_p is elicitable only on $\mathcal{M}_0^*(p)$, we consider co-elicitability on $\mathcal{M}_\infty^*(p) = \{F \in \mathcal{M}_\infty : F^{-1}$ is continuous at $p\}$. Among the forms of risk measures identified by Theorem 5.3, it is easy to see that the only choice of coherent risk measures is $\rho = a\mathrm{ES}_p + (1-a)\mathbb{E}$ for some $a \in [0,1]$. Therefore, naturally we would expect that the above form of ρ is the only spectral risk measure that is co-elicitable with VaR_p, although some detailed analysis needs to be carried out to translate from the CxLS property on the set $\mathcal{M}_\infty^*(p)$ to that on the set $\mathcal{M}_{\mathrm{dis}}$, in order to utilize Theorem 5.3.

Theorem 6.9. *For $p \in (0,1)$, a spectral risk measure $\rho : \mathcal{M}_\infty \to \mathbb{R}$ is co-elicitable with VaR_p on $\mathcal{M}_\infty^*(p)$ if and only if it is a convex combination of $\mathbb{E}$ and ES_p.*

To make sense of the elicitability of $(\mathrm{VaR}_p, a\mathrm{ES}_p + (1-a)\mathbb{E})$ in Theorem 6.9, we obtain its scoring function in the following proposition.

Proposition 6.10. *For $p \in (0,1)$ and $a \in (0,1]$, let $\rho = a\mathrm{ES}_p + (1-a)\mathbb{E}$. The functional (VaR_p, ρ) is $\mathcal{M}_\infty^*(p)$-elicitable with the strictly consistent scoring function*

$$S(x_1, x_2, y) = g(x_2) + g'(x_2)\left(a\left(x_1 + \frac{1}{1-p}(y - x_1)_+\right) + (1-a)y - x_2\right), \qquad x_1, x_2, y \in \mathbb{R},$$

where g is any differentiable, strictly increasing and strictly concave function on $\mathbb{R}$.

Remark 6.11. The scoring function S in Proposition 6.10 is not the only possible form of scoring functions for (VaR_p, ρ). The purpose here is not to characterize all forms of scoring functions, but to characterize all forms of ρ such that (VaR_p, ρ) is elicitable. For a characterization of all scoring functions for the special case of $(\mathrm{VaR}_p, \mathrm{ES}_p)$, see Fissler and Ziegel (2016, Corollary 5.5).

Finally, we give a simple corollary of Theorem 6.9, where an additional semi-continuity assumption identifies ES_p as the only comonotonic-additive coherent risk measure co-elicitable with VaR_p.

Corollary 6.12. *Suppose that $\rho : \mathcal{M}_\infty \to \mathbb{R}$ is a spectral risk measure that is lower semi-continuous with respect to weak convergence and $p \in (0,1)$. Then ρ is co-elicitable with VaR_p on $\mathcal{M}_\infty^*(p)$ if and only if $\rho = \mathrm{ES}_p$.*

7 Discussions

We provide various results on the CxLS property of one- and multi-dimensional risk functionals, and relate them to risk quadrangles, elicitability, and backtestability. Two major characterization results are established on signed Choquet integrals I_h with CxLS and on (I_h, VaR_p) with CxLS. A particularly elegant message is that the only type of signed Choquet integral that gains CxLS when paired with VaR_p is a linear combination of ES_p and $\mathbb{E}$. Based on these results, we proceed to show that a convex combination of $\mathbb{E}$ and ES_p is the only comonotonic-additive coherent risk measure that is co-elicitable with VaR_p.

It however remains an open question to characterize all two-dimensional signed Choquet integrals (I_h, I_g) with CxLS, or, furthermore, a similar problem in higher dimension. Given the level of technical complexity displayed in the techniques used to show Theorem 5.3, it seems to us that a general conclusion to the above question is far from being reachable with current methods. Even if one assumes that the signed Choquet integrals are increasing as in the risk measure literature, general results in multi-dimension are not available.

Closely related to the above issue, the characterization of elicitable (or identifiable) d-dimensional signed Choquet integrals remains an open problem. As explained in Section 6, the issues of elicitability, identifiability and backtestability are highly relevant for risk management practice, and they all require the CxLS property as a necessary condition. Hence, our study on CxLS provides a useful tool for future studies on the statistical notions above, especially in the multi-dimensional setting.

Acknowledgements. The authors thank Tobias Fissler and Rafael Frongillo for many helpful comments on an earlier version of the paper. The authors also thank the editors and the two anonymous referees for insightful comments, which greatly improved the quality of the paper. In particular, Section 3.3 is inspired by a comment from an associate editor, and Section 4.2 is inspired by a comment from a referee. R. Wang acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2018-03823, RGPAS-2018-522590) and from the Center of Actuarial Excellence Research Grant from the Society of Actuaries. Y. Wei is partially supported by funding from NSERC (RGPIN-2016-03654), the Hickman Scholar Program of the Society of Actuaries, and the Munich Reinsurance Company of Canada.

A Proofs of the main results

A.1 Proofs in Section 3

In order to prove Theorem 3.2, we first present two technical lemmas.

Lemma A.1. *If $h \in \mathcal{H}$ satisfies, for all $t, q \in [0,1]$, $0 \le h(t) \le h(1) = 1$ and*

$$h(t) = h(t)h(1-q+qt) + (1-h(t))h(qt), \tag{8}$$

then $h \in \mathcal{H}^*$.

Proof. We first get some smoothness of h using the same trick as in Wang and Ziegel (2015). Integrating both sides of (8) over $q \in [0,1]$, we obtain for $t \in (0,1)$,

$$\begin{aligned} h(t) &= h(t) \int_0^1 h(1-(1-t)q)\,\mathrm{d}q + (1-h(t)) \int_0^1 h(tq)\,\mathrm{d}q \\ &= \frac{h(t)}{1-t} \int_t^1 h(x)\,\mathrm{d}x + \frac{1-h(t)}{t} \int_0^t h(x)\,\mathrm{d}x \\ &= \frac{h(t)}{1-t}(g(1)-g(t)) + \frac{1-h(t)}{t} g(t), \end{aligned}$$

where $g(t) = \int_0^t h(x)\,\mathrm{d}x$. Rearranging terms, we have

$$h(t)\Big(t - \frac{t}{1-t}(g(1)-g(t)) + g(t)\Big) = g(t). \tag{9}$$

Note that the function g is continuous on $(0,1)$. For $t \in (0,1)$, if $g(t) \neq 0$, then (9) implies that h is continuous at t. If $g(t) = 0$ and h is not continuous at t, then $t - \frac{t}{1-t}(g(1)-g(t)) + g(t) = 0$, which implies $t = 1 - g(1)$. To summarize, either h has a jump at $t = 1-g(1)$ and is continuous elsewhere, or h is continuous on $(0,1)$. This fact implies that g is continuously differentiable on $(0,1)$ except for at the point $t = 1-g(1)$. Using the above relation (9) again, we know that h is continuously differentiable on $(0,1)$ except at the point $t = 1-g(1)$.

Differentiating both sides of (8) with respect to q, we get

$$0 = \frac{\mathrm{d}}{\mathrm{d}q}(h(t)h(1-q+qt) + (1-h(t))h(qt)).$$

By the product rule,

$$\frac{\mathrm{d}}{\mathrm{d}q}(h(t)h(1-q+qt) + (1-h(t))h(qt)) = h(t)h'(1-q+qt)(t-1) + (1-h(t))h'(qt)t, \tag{10}$$

assuming the derivatives in the right-hand-side of (10) exist. Plugging in $q = 1$ in (10) and rearranging terms, we have

$$h(t)h'(t) = h'(t)t \quad \text{for all } t \in (0,1) \setminus \{1-g(1)\}.$$

As a consequence,

$$h(t) = t \text{ or } h'(t) = 0 \quad \text{for all } t \in (0,1) \setminus \{1-g(1)\}. \tag{11}$$

Pick any point $t_0 \in (0,1) \setminus \{1-g(1)\}$, and assume that $h'(t_0) \neq 0$ and $h'(t_0) \neq 1$. Using (11), we have $h(t_0) = t_0$. Since $h'(t_0) \neq 1$, there exists a neighbourhood $(t_0 - \epsilon, t_0 + \epsilon)$ such that $h(t) \neq t$ for all $t \in (t_0 - \epsilon, t_0 + \epsilon)$ and $t \neq t_0$. Using (11) again, we know that $h'(t) = 0$ for all

$t \in (t_0 - \epsilon, t_0 + \epsilon)$ and $t \neq t_0$. The continuous differentiability of h at t_0 then implies $h'(t_0) = 0$, a contradiction. Therefore, we conclude

$$h'(t) = 1 \text{ or } h'(t) = 0 \quad \text{for all } t \in (0,1) \setminus \{1 - g(1)\}. \tag{12}$$

First, suppose that h is continuously differentiable on $(0,1)$. In this case, h' cannot switch between 0 and 1. Therefore, we have, either $h'(t) = 0$ on $(0,1)$ or $h'(t) = 1$ on $(0,1)$. This means either $h(t) = c$ on $(0,1)$ for some constant $c \in [0,1]$, or $h(t) = t$ on $(0,1)$. In other words, $h \in \mathcal{H}_1^*$ or $h \in H_2^*$.

Next, suppose that h is not continuously differentiable at $t_0 = 1 - g(1)$. Note that this implies $g(t_0) = 0$, and hence $h(t) = 0$ a.e on $(0, t_0)$. Further, since $1 - t_0 = g(1) = \int_0^1 h(t)\,\mathrm{d}t = \int_{t_0}^1 h(t)\,\mathrm{d}t \leq 1 - t_0$, we know that $h(t) = 1$ a.e. on $(t_0, 1)$. Since h is continuously differentiable on $(0, t_0)$ and $(t_0, 1)$, we know that $h(t) = 0$ on $(0, t_0)$ and $h(t) = 1$ on $(t_0, 1)$. Thus, $h \in \mathcal{H}_3^*$. $\square$

The next lemma gives a sufficient condition for $h(t)$ to have the same sign as $h(1)$. Since h is not necessarily monotone for a signed Choquet integral, it is an important step to verify that $h(t)$ has the same sign as $h(1)$ in order to utilize Lemma A.1.

Lemma A.2. *Fix $h \in \mathcal{H}$ and $t \in (0,1]$, and suppose $h(t) \neq 0$ and $h(1) \neq 0$. For $x, y \in \mathbb{R}$, where x and y satisfy $0 < x < y$ and $y = \left(1 - \frac{h(1)}{h(t)}\right)x + \frac{h(1)}{h(t)}$, if h satisfies*

$$h(1) = I_h(q((1-t)\delta_x + t\delta_y) + (1-q)\delta_1), \tag{13}$$

for all $q \in [0,1]$, then $x < 1 \leq y$ and $\frac{h(1)}{h(t)} \geq 1$.

Proof. Without loss of generality, we assume $h(1) = 1$. Because $y - x = \frac{1}{h(t)}(1-x)$ and $y - 1 = \left(1 - \frac{1}{h(t)}\right)(x-1)$, by the fact that $x < y$, there exist three cases.

(a) $x > 1, y > 1$ and $h(t) < 0$: Equation (13) reduces to

$$h(q)h(t) = h(tq). \tag{14}$$

(b) $x < 1, y < 1$ and $h(t) > 1$: Equation (13) reduces to

$$h(t) = h(1 - q(1-t)) + h(1-q)(h(t) - 1). \tag{15}$$

Then

$$h(t) = \frac{h(1 - q(1-t)) - h(1-q)}{1 - h(1-q)}.$$

(c) $x < 1, y \geq 1$ and $0 < h(t) \leq 1$: Equation (13) reduces to

$$h(t) = h(t)h(1 - q(1-t)) + (1 - h(t))h(tq). \tag{16}$$

We show that cases (a) and (b) above are actually not possible. In other words, a function $h \in \mathcal{H}$ satisfying $h(1) = 1$ and (a)-(c) takes values in $[0, 1]$.

We first show that such a function h is non-negative. Note that (c) implies

$$\text{for any } t \in [0,1], \text{ if } h(t) = 0, \text{ then } h(s) = 0 \text{ for all } s \in [0, t]. \tag{17}$$

Suppose that there exists $t \in [0, 1]$ such that $h(t) < 0$. If $h(\sqrt{t}) < 0$, by (a), we have $h(t) = h(\sqrt{t})h(\sqrt{t}) > 0$, which is a contradiction. Hence, $h(\sqrt{t}) \geq 0$. Note that (14) also holds if $h(t) = 0$, due to (17). By (a), we know that $h(t)h(\sqrt{t}) = h(t\sqrt{t}) \leq 0$. Using (a) again,

$$0 \geq h(t)h(\sqrt{t})h(\sqrt{t}) = h(t\sqrt{t})h(\sqrt{t}) = h(t^2).$$

On the other hand, (a) also gives

$$h(t^2) = h(t)h(t) > 0,$$

a clear contradiction. Therefore, $h(t) \geq 0$ for all $t \in [0, 1]$.

Next we show $h(t) \leq 1$ for all $t \in [0, 1]$. We note the following two useful facts. First, for any $t \in [0, 1]$ such that $h(t) \in (0, 1]$, by (c), we have

$$1 = h(1 - q(1-t)) + \frac{1 - h(t)}{h(t)} h(tq).$$

Using the fact that h is non-negative, we have $h(1 - q(1 - t)) \leq 1$ for all $q \in [0, 1]$. Therefore, we conclude the following statement:

$$\text{For any } t \in [0,1] \text{ such that } h(t) \in (0,1],\ h(s) \in [0,1] \text{ for all } s \in [t, 1]. \tag{18}$$

Second, for any $t \in [0, 1]$ such that $h(t) > 1$, by taking $q = 1 - t$, (15) gives

$$h(t) = h(1 - (1-t)^2) + h(t)(h(t) - 1).$$

Rearranging terms, we have $(h(t) - 1)^2 = 1 - h(1 - (1 - t)^2)$. This implies $h(1 - (1 - t)^2) < 1$. Therefore, we have:

$$\text{For any } t \in [0,1] \text{ such that } h(t) > 1,\ h(1 - (1-t)^2) < 1. \tag{19}$$

Suppose that there exists $t \in (0, 1)$ such that $h(t) > 1$. Let $s = 1 - \sqrt{1 - t}$. Clearly, $s < t$. If $h(s) > 1$, then by (19), we have $h(1 - (1 - s)^2) = h(t) < 1$, a contradiction. If $h(s) \in (0, 1]$, then by (18), $h(t) \in [0, 1]$, another contradiction. Hence, $h(s) = 0$.

Plugging $q = 1 - s = \sqrt{1-t}$ in (15),

$$h(t) = h(1-(1-t)(1-s)) + h(s)(h(t)-1)) = h(1-(1-t)\sqrt{1-t}) > 1. \tag{20}$$

In particular,

$$h(t) = h(1-(1-t)\sqrt{1-t}) > 1, \tag{21}$$

and setting $w = 1-(1-t)\sqrt{1-t}$ in (20), we get

$$h(1-(1-w)\sqrt{1-t}) = h(1-(1-t)^2) > 1.$$

This is a contraction to (19). Combining all cases, there does not exist $t \in (0,1)$ such that $h(t) > 1$. Together with the fact that h is non-negative, we come to the conclusion that $h(t) \in [0,1]$ for all $t \in [0,1]$. Thus only case (c) is possible. $\square$

Proof of Theorem 3.2. It is easy to verify that the three classes of functionals in (2)-(4) have CxLS, and hence $h \in \mathcal{H}^*$ is sufficient for the CxLS property. Below we show the necessity of $h \in \mathcal{H}^*$. Suppose $h(t_0) = 0$ for some fixed $t_0 \in [0,1]$. Observe that $I_h(\delta_0) = 0$ and $I_h((1-t_0)\delta_0 + t_0\delta_1) = h(t_0) = 0$. Since I_h has CxLS, for any $q \in [0,1]$,

$$I_h\left((1-q)\delta_0 + q(1-t_0)\delta_0 + qt_0\delta_1\right) = h(t_0 q) = I_h(\delta_0) = 0. \tag{22}$$

It follows that if $h(1) = 0$, then $h(t) = 0$ on $[0,1]$. This is included in each of cases (i)-(iii). In the following, $h(t) \neq 0$ for any $t \in (0,1]$, and we can assume $h(1) = 1$ without loss of generality, since the set $\mathcal{H}^*$ is invariant under a constant multiplier. For $0 < x < y$ and any fixed $t \in (0,1]$, we have

$$I_h((1-t)\delta_x + t\delta_y) = x + h(t)(y-x).$$

Note that $I_h(\delta_1) = h(1) = 1$. In the following we choose $y = \left(1 - \frac{1}{h(t)}\right)x + \frac{1}{h(t)}$, so that $I_h((1-t)\delta_x + t\delta_y) = 1$. Since I_h has CxLS, for all $q \in [0,1]$,

$$1 = I_h(q((1-t)\delta_x + t\delta_y) + (1-q)\delta_1). \tag{23}$$

By Lemma A.2, we have $x < 1 \leq y$ and $h(t) \in (0,1]$. Hence, (23) reduces to

$$h(t) = h(t)h(1-q+qt) + (1-h(t))h(qt), \tag{24}$$

for all $t \in (0,1]$ with $h(t) \neq 0$ and $q \in [0,1]$. Note that (23) holds for $t = 0$ and if $h(t) = 0$, then (24) holds automatically by (22). Therefore, (24) holds for all $t, q \in [0,1]$. This is precisely (8). By applying Lemma A.1, we obtain $h \in \mathcal{H}^*$. $\square$

The next lemma provides some technical properties of the w-convergence defined in Section 5. The lemma will be useful in bridging the gap between continuous and discrete distributions in the proof of Proposition 3.4, as well as in the characterization of multi-dimensional signed Choquet integrals in Section 5.

Lemma A.3. *(i) Suppose that $F_n, G_n \in \mathcal{M}_\infty$, $n \in \mathbb{N}$, $F_n \overset{w}{\to} F \in \mathcal{M}_\infty$ and $G_n \overset{w}{\to} G \in \mathcal{M}_\infty$. Then $\lambda F_n + (1-\lambda)G_n \overset{w}{\to} \lambda F + (1-\lambda)G$ for all $\lambda \in [0,1]$.*

(ii) For $h \in \mathcal{H}$, I_h is uniformly continuous with respect to w on $\mathcal{M}_\infty$.

Proof. For both statements, we note that $F_n \overset{w}{\to} F$ if and only if $F_n^{-1}(U) \to F^{-1}(U)$ in L^∞ for any U$[0,1]$ random variable U.

(i) Let $A \in \mathcal{F}$ with $\mathbb{P}(A) = \lambda$ and $U \sim$ U$[0,1]$ be independent. It is easy to see that the random variable $\mathbb{1}_A F_n^{-1}(U) + \mathbb{1}_{A^c} G_n^{-1}(U)$ has the distribution $\lambda F_n + (1-\lambda)G_n$, $n \in \mathbb{N}$. Moreover, by the w-convergence of $\{F_n\}_{n\in\mathbb{N}}$ and $\{G_n\}_{n\in\mathbb{N}}$, $\mathbb{1}_A F_n^{-1}(U) + \mathbb{1}_{A^c} G_n^{-1}(U) \to \mathbb{1}_A F^{-1}(U) + \mathbb{1}_{A^c} G^{-1}(U)$ in L^∞. Therefore, $\lambda F_n + (1-\lambda)G_n \overset{w}{\to} \lambda F + (1-\lambda)G$.

(ii) The conclusion follows from the fact that a Signed Choquet integral as a functional on L^∞ is uniformly continuous with respect to the L^∞-norm (Theorem 1 of Wang et al. (2020)). $\square$

Proof of Proposition 3.4. (i) Assume $h(1) = 0$. Let $F_n =$ U$[0, 1/n]$ and $G_n =$ U$[1, 1+1/n]$, $n \in \mathbb{N}$. Note that $F_n \overset{w}{\to} \delta_0$ and $G_n \overset{w}{\to} \delta_1$. By Lemma A.3, we have $(1-\lambda)F_n + \lambda G_n \overset{w}{\to}$ Bernoulli(λ) for $\lambda \in (0,1)$. Also note that from the translation invariance of I_h (e.g., Lemma 2 of Wang et al. (2020)), $I_h(G_n) = I_h(F_n) + h(1) = I_h(F_n)$. The CxLS on $\mathcal{M}_{\text{con}}$ implies $I_h((1-\lambda)F_n + \lambda G_n) = I_h(F_n)$. Therefore, by Lemma A.3,

$$I_h(\text{Bernoulli}(\lambda)) = \lim_{n\to\infty} I_h((1-\lambda)F_n + \lambda G_n) = \lim_{n\to\infty} I_h(F_n) = I_h(\delta_0) = 0.$$

Also note that $I_h(\text{Bernoulli}(\lambda)) = h(\lambda)$. This gives $h(t) = 0$ for $t \in [0,1]$. Hence $I_h(F) = 0$ for all $F \in \mathcal{M}_0$, and it has CxLS on any set of distributions.

(ii) Assume $h(1) \neq 0$. Take $F, G \in \mathcal{M}_{\text{dis}}$ such that $I_h(F) = I_h(G)$. Write $c = h(1)$ and $b = I_h(F)$. Take two sequences of distributions $\{F_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_{\text{con}}$ and $\{G_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_{\text{con}}$ such that $F_n \overset{w}{\to} F$ and $G_n \overset{w}{\to} G$. Such sequences are easy to construct by, e.g., replacing each point mass with a uniform on a small interval of length $1/n$. Let U be a U$[0,1]$ random variable. For $n \in \mathbb{N}$, let F_n^* be the distribution of $F_n^{-1}(U) + \frac{1}{c}(b - I_h(F_n))$. We have $F_n^* \overset{w}{\to} F$, because $F_n \overset{w}{\to} F$ and $I_h(F_n) \to b$ by Lemma A.3. Moreover, by the translation invariance of I_h again, $I_h(F_n^*) = I_h(F_n) + b - I_h(F_n) = b$. Similarly, let G_n^* be the distribution of $G_n^{-1}(U) + \frac{1}{c}(b - I_h(G_n))$, and we have $G_n^* \xrightarrow{w} G$ and $I_h(G_n^*) = b$. By Lemma A.3, we have $\lambda F_n^* + (1-\lambda)G_n^* \xrightarrow{w} \lambda F + (1-\lambda)G$. Finally, noting that $F_n^*, G_n^* \in \mathcal{M}_{\text{con}}$, the CxLS on $\mathcal{M}_{\text{con}}$ implies that $I_h(\lambda F_n^* + (1-\lambda)G_n^*) = I_h(F_n^*) = b$. Therefore, by Lemma A.3 again,

$$I_h(\lambda F + (1-\lambda)G) = \lim_{n\to\infty} I_h(\lambda F_n^* + (1-\lambda)G_n^*) = b = I_h(F) = I_h(G).$$

Hence, I_h has CxLS on $\mathcal{M}_{\text{dis}}$. □

Proof of Proposition 3.5. First, we note the key observation that, for any $\lambda \in (0,1)$ and $F, G \in \mathcal{M}$,

$$(\lambda F + (1-\lambda)G) \circ v^{-1} = (\lambda F) \circ v^{-1} + ((1-\lambda)G) \circ v^{-1} = \lambda(F \circ v^{-1}) + (1-\lambda)(G \circ v^{-1}). \quad (25)$$

This observation follows directly from the definition of probability measures. Suppose that $\phi_v(F) = \phi_v(G)$. By (5), we have $\phi(F\circ v^{-1}) = \phi(G\circ v^{-1})$. Using the CxLS property of ϕ and (25), for $\lambda \in (0,1)$ such that $\lambda F + (1-\lambda)G \in \mathcal{M}$, we have

$$\begin{aligned}\phi_v(\lambda F + (1-\lambda)G) &= \phi((\lambda F + (1-\lambda)G) \circ v^{-1}) \\ &= \phi(\lambda(F \circ v^{-1}) + (1-\lambda)(G \circ v^{-1})) = \phi(F \circ v^{-1}) = \phi_v(F).\end{aligned}$$

Thus ϕ_v has CxLS. If v is injective, then the converse statement holds true by noting that $\phi = (\phi_v)_{v^{-1}}$. □

A.2 Proofs in Section 4

Proof of Proposition 4.1. Fix an $m \in \mathbb{N}$. The functional ρ can be written as $\rho(\cdot) = h(f_1(\cdot), \ldots, f_m(\cdot))$, where $h : \mathbb{R}^m \to \mathbb{R}^d$ is an injective function and each $f_i : \mathcal{M} \to \mathbb{R}^{d_i}$ has CxLS. For any $F, G \in \mathcal{M}$ that satisfy $\rho(F) = \rho(G)$, we have $(f_1(F), \ldots, f_m(F)) = (f_1(G), \ldots, f_m(G))$ by the fact that h is an injection. Because each f_i has CxLS, $(f_1(\lambda F+(1-\lambda)G), \ldots, f_m(\lambda F+(1-\lambda)G)) = (f_1(F), \ldots, f_m(F))$ for any $\lambda \in [0,1]$. Hence, $\rho(\lambda F + (1-\lambda)G) = h(f_1(\lambda F + (1-\lambda)G), \ldots, f_m(\lambda F + (1-\lambda)G)) = h(f_1(F), \ldots, f_m(F)) = \rho(F) = \rho(G)$ for any $\lambda \in [0,1]$. □

Proof of Proposition 4.2. We only need to show the "if" direction. For any $F, G \in \mathcal{M}$, if $(\rho_1(F), \rho_2(F)) = (\rho_1(G), \rho_2(G)) = (r_1, r_2)$, then $\rho_2(\lambda F + (1-\lambda)G) = r_2$ for any $\lambda \in [0,1]$. Since ρ_1 has CxLS on $\mathcal{M}(r_2)$, we have $\rho_1(\lambda F + (1-\lambda)G) = r_1$ for any $\lambda \in [0,1]$, implying that (ρ_1, ρ_2) has CxLS on $\mathcal{M}$. □

Proof of Proposition 4.3. Obviously, $\mathcal{V}$ and $\mathcal{E}$ have CxLS since they are linear on $\mathcal{M}$ as in Example 2.2. Next, assume that $\mathcal{S}$ is real-valued. The CxLS property of $\mathcal{S}$ is implied by the fact that it is the minimizer of an expected score (thus elicitable; see Proposition 6.4). Next, we discuss the case of $(\mathcal{S},\mathcal{R})$. For $F\in\mathcal{M}(r)$ where $\mathcal{M}(r)=\{F\in\mathcal{M}_1:\mathcal{S}(F)=r\}$, $r\in\mathbb{R}$, using the relationship $\mathcal{R}(F)=r+\mathcal{V}(F\circ t_r)$ and $\mathcal{D}(F)=\mathcal{R}(F)-\mathbb{E}[F]$, we have $\mathcal{R}(F)=\int_{\mathbb{R}}v(x-r)\,\mathrm{d}F(x)+r$, and $\mathcal{D}(F)=\int_{\mathbb{R}}(v(x-r)-x)\,\mathrm{d}F(x)+r$. Since both $\mathcal{R}$ and $\mathcal{D}$ are linear in $F\in\mathcal{M}(r)$, Proposition 4.2 implies that $(\mathcal{S},\mathcal{R})$ and $(\mathcal{S},\mathcal{D})$ both have CxLS. Finally, any combination of functionals with CxLS still has CxLS because of the injective relation in Proposition 4.1. Hence the last statement of the proposition holds true. □

Proof of Proposition 4.4. Since Γ is the minimizer of an expected loss function, it has CxLS (see also Proposition 6.4). On the set $\mathcal{M}(r)=\{F\in\mathcal{M}_0:\Gamma(F)=r\}$, $r\in\mathbb{R}$, γ is linear in $F\in\mathcal{M}(r)$, and hence (γ,Γ) has CxLS by Proposition 4.2. □

A.3 Proofs in Section 5

In order to prove Theorem 5.3, we need the following technical lemma, which connects the problem on dimension two with the result in dimension one.

Lemma A.4. *For $p\in(0,1)$ and $h\in\mathcal{H}$, if (I_h,VaR_p) has CxLS on $\mathcal{M}_{\mathrm{dis}}$, then*

$$h(t)=h_1\left(\frac{t}{1-p}\right)\mathbb{1}_{\{t\le 1-p\}}+\left(h_2\left(\frac{1-t}{p}\right)+h(1)\right)\mathbb{1}_{\{t>1-p\}},\tag{26}$$

for some h_1 and $h_2\in\mathcal{H}^*$.

Proof. Define $h_1,h_2:[0,1]\to\mathbb{R}$ by $h_1(t)=h(t(1-p))$, $t\in[0,1]$ and $h_2(t)=h(1-tp)-h(1)$, $t\in[0,1)$. Further, let $h_2(1)=\lim_{t\uparrow 1}h_2(t)$, so that h_2 is continuous at $t=1$. Clearly, $h_1,h_2\in\mathcal{H}$ and (26) holds. We shall show $h_1,h_2\in\mathcal{H}^*$ below.

Let $\mathcal{M}_\infty^+=\{F\in\mathcal{M}_{\mathrm{dis}}:F((-\infty,0))=0\}$ which is the set of distributions supported on $[0,\infty)$. Further, let $\mathcal{M}^*=\{p\delta_0+(1-p)F:F\in\mathcal{M}_\infty^+\}$. Note that $\mathrm{VaR}_p(F)=0$ for $F\in\mathcal{M}^*$ by definition. Therefore, by Proposition 4.2, I_h has CxLS on $\mathcal{M}^*$. By definition of I_h, for $F\in\mathcal{M}_\infty^+$,

$$I_h(p\delta_0+(1-p)F)=\int_0^\infty h(1-(p+(1-p)F(x)))\,\mathrm{d}x=\int_0^\infty h_1(1-F(x))\,\mathrm{d}x=I_{h_1}(F).$$

Note that in the above equation we treat the measure F as the corresponding cumulative distribution function. Since I_h has CxLS on $\mathcal{M}^*$, and there is a linear mapping between $\mathcal{M}^*$ and $\mathcal{M}_\infty^+$, we know that I_{h_1} has CxLS on $\mathcal{M}_\infty^+$.

For $G \in \mathcal{M}_{\text{dis}}$, we can write $G = T_x F$ for some $x \in \mathbb{R}$ and $F \in \mathcal{M}_\infty^+$, where T_x is the operator of left-shift by $x \in \mathbb{R}$, that is, $T_x F(y) = F(y+x)$ for $y \in \mathbb{R}$. By definition of a signed Choquet integral (or, see Lemma 2 of Wang et al. (2020)), we have $I_{h_1}(G) = I_{h_1}(T_x F) = I_{h_1}(F) - x h_1(1)$. Therefore, if $G_1, G_2 \in \mathcal{M}_{\text{dis}}$ satisfy $I_{h_1}(G_1) = I_{h_2}(G_2)$, then for some $x \in \mathbb{R}$ such that $T_{-x}G_1, T_{-x}G_2 \in \mathcal{M}_\infty^+$, we know $I_{h_1}(T_{-x}G_1) = I_{h_2}(T_{-x}G_2)$. Since I_{h_1} has CxLS on $\mathcal{M}_\infty^+$,

$$\begin{aligned}I_{h_1}(\lambda G_1 + (1-\lambda)G_2) &= I_{h_1}(\lambda T_{-x}G_1 + (1-\lambda)T_{-x}G_2) + x h_1(1)\\ &= I_{h_1}(T_{-x}G_1) + x h_1(1) = I_{h_1}(G_1).\end{aligned}$$

As a consequence, I_{h_1} has CxLS on $\mathcal{M}_{\text{dis}}$. By Theorem 3.2, we know $h_1 \in \mathcal{H}^*$.

The statement for h_2 is somehow more complicated as it is not symmetric to the case of h_1. Fix $q \in (0,p)$, and let $\mathcal{M}_- = \{F \in \mathcal{M}_{\text{dis}} : F((-\infty, 0]) = 1\}$, $g_q(t) = h(1-tq) - h(1)$, $t \in [0,1]$, and $\mathcal{M}_q^* = \{(1-q)\delta_0 + qF : F \in \mathcal{M}_-\}$. Note that $\mathrm{VaR}_p(F) = 0$ for $F \in \mathcal{M}_q^*$ by definition. Therefore, by Proposition 4.2, I_h has CxLS on $\mathcal{M}_q^*$. Let $\bar{g}_q(t) = g_q(1-t) - g_q(1)$, $t \in [0,1]$. Note that $\bar{g}_q \in \mathcal{H}$ and $\bar{g}_q(1) = -g_q(1)$. By definition of I_h, for $F \in \mathcal{M}_-$,

$$\begin{aligned}I_h((1-q)\delta_0 + qF) &= \int_{-\infty}^0 (h(1-qF(x)) - h(1))\,\mathrm{d}x\\ &= \int_{-\infty}^0 g_q(F(x))\,\mathrm{d}x = \int_{-\infty}^0 (\bar{g}_q(1-F(x)) - \bar{g}_q(1))\,\mathrm{d}x = I_{\bar{g}_q}(F).\end{aligned}$$

Since I_h has CxLS on $\mathcal{M}_q^*$, and there is a linear mapping between $\mathcal{M}_q^*$ and $\mathcal{M}_-$, we know that I_{g_q} has CxLS on $\mathcal{M}_-$. Following the similar arguments for h_1, we obtain $\bar{g}_q \in \mathcal{H}^*$. Checking the three forms of functions in $\mathcal{H}^*$, we know that $g_q \in \mathcal{H}^*$. Note that q is arbitrarily chosen in $(0,p)$, and $h_2(t) = g_q(pt/q) = g_{tp}(1)$ for $t \le q/p$. If $g_q(1) = 0$ for all $q \in (0,1)$, then h_2 is zero on $(0,1)$, thus $h_2 \in \mathcal{H}_1^*$. Next, assume that there exists $q_0 \in (0,p)$ such that $g_{q_0}(1) = c \neq 0$. There are four cases to analyze. If $g_{q_0} \in \mathcal{H}_1^*$ and $g_{q_0}(1-) = 0$, i.e., h_2 is zero on $(0, q_0/p)$, then, by letting q vary in (q_0, p), constrained by $g_q \in \mathcal{H}^*$, h_2 must be equal to a constant d on $(q_0/p, 1)$ with $dc > 0$ and $|d| \ge |c|$, thus $h_2 \in \mathcal{H}_3^*$. If $g_{q_0} \in \mathcal{H}_1^*$ and $g_{q_0}(1-) \neq 0$, i.e., h_2 is a non-zero constant on $(0, q_0/p)$, then, by letting q vary in (q_0,p), constrained by $g_q \in \mathcal{H}^*$, h_2 must be equal to c on $(q_0/p, 1)$, thus $h_2 \in \mathcal{H}_1^*$. If $g_{q_0} \in \mathcal{H}_2^*$, i.e., h_2 is linear on $(0, q_0/p)$, then, by letting q vary in (q_0,p), constrained by $g_q \in \mathcal{H}^*$, h_2 must be linear on $(0,1)$, thus $h_2 \in \mathcal{H}_2^*$. If $g_{q_0} \in \mathcal{H}_3^*$, i.e., there is a jump of h_2 in $(0, q_0/p)$, then, by letting q vary in (q_0,p), constrained by $g_q \in \mathcal{H}^*$, h_2 must be equal to c on $(q_0/p, 1)$, thus $h_2 \in \mathcal{H}_3^*$. In all cases, $h_2 \in \mathcal{H}^*$. □

Proof of Theorem 5.3. For $r \in \mathbb{R}$, denote by $\mathcal{M}(r) = \{F \in \mathcal{M}_{\text{dis}} : \mathrm{VaR}_p(F) = r\}$. We first verify (I_h, VaR_p) in cases (i)-(iv) indeed has CxLS.

(i) Both VaR_p and I_{h^*} have CxLS by Theorem 3.2. Hence, (I_h, VaR_p) has CxLS as justified by Proposition 4.1.

(ii) For $F \in \mathcal{M}(r)$, using the ES-VaR formula of Rockafellar and Uryasev (2002),

$$I_h(F) = ar + b\int_{-\infty}^{\infty} x\,\mathrm{d}F(x) + c\left(r + \frac{1}{1-p}\int_r^{\infty}(x-r)\,\mathrm{d}F(x)\right).$$

Hence, I_h is affine for $F \in \mathcal{M}(r)$. So I_h has CxLS on $\mathcal{M}(r)$. By Proposition 4.2, we know that (I_h, VaR_p) has CxLS.

(iii) For $F, G \in \mathcal{M}(r)$, if $I_h(F) = I_h(G)$, then $b\mathrm{VaR}_p^+(F) + c\,\text{ess-sup}(F) = b\mathrm{VaR}_p^+(G) + c\,\text{ess-sup}(G)$. Without loss of generality, assume $\mathrm{VaR}_p^+(F) \geq \mathrm{VaR}_p^+(G)$, which implies $\text{ess-sup}(F) \leq \text{ess-sup}(G)$ since $bc > 0$. Note that for $\lambda \in (0,1)$, since $\mathrm{VaR}_p(F) = \mathrm{VaR}_p(G) = \mathrm{VaR}_p(\lambda F + (1-\lambda G)) = r$, we have $\mathrm{VaR}_p^+(\lambda F + (1-\lambda)G) = \mathrm{VaR}_p^+(G)$. Therefore, for $\lambda \in (0,1)$.

$$b\mathrm{VaR}_p^+(\lambda F + (1-\lambda)G) + c\,\text{ess-sup}(\lambda F + (1-\lambda)G) = b\mathrm{VaR}_p^+(G) + c\,\text{ess-sup}(G).$$

Hence, I_h has CxLS on $\mathcal{M}(r)$. By Proposition 4.2, we know that (I_h, VaR_p) has CxLS.

(iv) For $F, G \in \mathcal{M}(r)$, if $I_h(F) = I_h(G)$, then $b\mathrm{VaR}_p^+(F) + c\,\text{ess-inf}(F) = b\mathrm{VaR}_p^+(G) + c\,\text{ess-inf}(G)$. Without loss of generality, assume $\mathrm{VaR}_p^+(F) \geq \mathrm{VaR}_p^+(G)$, which implies $\text{ess-inf}(F) \geq \text{ess-inf}(G)$ since $bc < 0$. Note that for $\lambda \in (0,1)$, since $\mathrm{VaR}_p(F) = \mathrm{VaR}_p(G) = \mathrm{VaR}_p(\lambda F + (1-\lambda G)) = r$, we have $\mathrm{VaR}_p^+(\lambda F + (1-\lambda)G) = \mathrm{VaR}_p^+(G)$. Therefore, for $\lambda \in (0,1)$.

$$b\mathrm{VaR}_p^+(\lambda F + (1-\lambda)G) + c\,\text{ess-inf}(\lambda F + (1-\lambda)G) = b\mathrm{VaR}_p^+(G) + c\,\text{ess-inf}(G).$$

Hence, I_h has CxLS on $\mathcal{M}(r)$. By Proposition 4.2, we know that (I_h, VaR_p) has CxLS.

Next, suppose that (I_h, VaR_p) has CxLS, and we show that it has to be one of the cases (i)-(iv). To simplify notation, for $p \in (0,1)$ and $c \in [0,1]$, let

$$\mathrm{ES}_p^-(F) = \frac{1}{p}\int_0^p \mathrm{VaR}_t(F)\,\mathrm{d}t, \quad F \in \mathcal{M}_\infty,$$

and

$$Q_p^c(F) = c\,\text{ess-sup}(F) + (1-c)\mathrm{VaR}_p^+(F), \quad F \in \mathcal{M}_\infty.$$

Note that $Q_p^1 = \text{ess-sup}$ and $Q_p^0 = \mathrm{VaR}_p^+$.

By Lemma A.4, h satisfies (26), that is, for some h_1 and $h_2 \in \mathcal{H}^*$,

$$h(t) = h_1\left(\frac{t}{1-p}\right)\mathbb{1}_{\{t\leq 1-p\}} + \left(h_2\left(\frac{1-t}{p}\right) + h(1)\right)\mathbb{1}_{\{t>1-p\}} = g_1(t) + g_2(t),$$

where $g_1(t) = h_1(\frac{t}{1-p})\mathbb{1}_{\{t\le 1-p\}}$ and $g_2(t) = (h_2(\frac{1-t}{p}) + h(1))\mathbb{1}_{\{t>1-p\}}$, $t \in [0,1]$. Analyzing all possible forms of I_{g_1} and I_{g_2}, we have

$$I_h = I_{g_1} + I_{g_2} = a\mathrm{VaR}_p + a_1 I + a_2 J,$$

where $I \in \{\mathrm{VaR}_\alpha^{c_1}, \mathrm{ES}_p^-\}$ and $J \in \{\mathrm{ES}_p, \mathrm{VaR}_\beta^{c_2}, Q_p^c\}$, $a, a_1, a_2 \in \mathbb{R}$, $\alpha \in [0,p)$, $\beta \in (p,1]$ and $c_1, c_2, c \in [0,1]$. Since (I_h, VaR_p) has CxLS if and only if $(I_h - a\mathrm{VaR}_p, \mathrm{VaR}_p)$ has CxLS, we can freely set $a = 0$. Hence, we can write $I_h = a_1 I + a_2 J$. Without loss of generality we assume $a_1 \ge 0$; otherwise we can replace I_h by $-I_h$. There are a few cases to analyze. Below, we use the fact that ES_p^- is a linear combination of ES_p and $\mathbb{E}$ via $p\mathrm{ES}_p^- + (1-p)\mathrm{ES}_p = \mathbb{E}$.

(a) $a_1 = 0$. The case $J \in \{\mathrm{VaR}_\beta^{c_2}, \mathrm{VaR}_p^+, \text{ess-sup}\}$ is included in case (i); the case $J = \mathrm{ES}_p$ is included in case (ii), and the case $J = Q_p^c$ for $c \in (0,1)$ is included in case (iii).

(b) $a_2 = 0$. The case $I = \mathrm{VaR}_\alpha^{c_1}$ is included in case (i) and $I = \mathrm{ES}_p^-$ is included in case (ii).

(c) $a_1 > 0, a_2 > 0$. We claim that if I_h has CxLS on $\mathcal{M}(0)$, then either $(I, J) = (\text{ess-inf}, \text{ess-sup})$, included in case (i), or $(I, J) = (\mathrm{ES}_p^-, \mathrm{ES}_p)$, included in case (ii). Below we show our assertion.

First, suppose that $I = \mathrm{VaR}_\alpha^{c_1}$. For $\epsilon \in (\alpha, p)$, let $F = \epsilon\delta_{-a_2} + (p-\epsilon)\delta_0 + (1-p)\delta_{a_1}$ and $G = \delta_0$. We can easily calculate $\mathrm{VaR}_p(F) = \mathrm{VaR}_p(G) = 0$, $I(F) = -a_2$, $J(F) = a_1$ and $I(G) = J(G) = 0$. Therefore, $I_h(F) = I_h(G) = 0$. For $\lambda \in [0,1]$,

$$\lambda F + (1-\lambda)G = \lambda\epsilon\delta_{-a_2} + (1 - \lambda + \lambda(p-\epsilon))\delta_0 + \lambda(1-p)\delta_{a_1}.$$

If I_h has CxLS on $\mathcal{M}(0)$, then $I_h(\lambda F + (1-\lambda)G) = 0$ for all $\lambda \in [0,1]$ and all $\epsilon \in (\alpha, p)$. Note that the function $\lambda \mapsto I(\lambda F + (1-\lambda)G)$ has a jump at $\lambda_1 = \alpha/\epsilon \in [0,1)$, and the function $\lambda \mapsto J(\lambda F + (1-\lambda)G)$ either has no jump ($J = \mathrm{ES}_p$), a jump at $\lambda_2 = 1$ ($J = Q_p^c$ for $c \neq 0$), a jump at $\lambda_2 = 0$ ($J = \text{ess-sup} = Q_p^0$), or a jump at $\lambda_2 = (1-\beta)/(1-p)$ ($J = \mathrm{VaR}_\beta^c$). Note that $\lambda_1 \neq \lambda_2$ for almost every $\epsilon \in (\alpha, p)$, except for the case $(\alpha, c) = (0,0)$. Hence, except for $(I, J) = (\text{ess-inf}, \text{ess-sup})$, the function $\lambda \mapsto I_h(\lambda F + (1-\lambda)G)$ does not take a constant value on $[0,1]$, and I_h cannot have CxLS on $\mathcal{M}(0)$.

Next, suppose that $I = \mathrm{ES}_p^-$. For $\epsilon \in (0,p)$, let $F = \epsilon\delta_{-a_3} + (p-\epsilon)\delta_0 + (1-p)\delta_{a_1}$ and $G = \delta_0$, where $a_3 = pa_2/\epsilon$. We can easily calculate $\mathrm{VaR}_p(F) = \mathrm{VaR}_p(G) = 0$, $I(F) = -\epsilon a_3/p = -a_2$, $J(F) = a_1$ and $I(G) = J(G) = 0$. Therefore, $I_h(F) = I_h(G) = 0$. For $\lambda \in [0,1]$,

$$\lambda F + (1-\lambda)G = \lambda\epsilon\delta_{-a_3} + (1 - \lambda + \lambda(p-\epsilon))\delta_0 + \lambda(1-p)\delta_{a_1}.$$

If I_h has CxLS on $\mathcal{M}(0)$, then $I_h(\lambda F + (1-\lambda)G) = 0$ for all $\lambda \in [0,1]$. Note that the function $\lambda \mapsto I(\lambda F + (1-\lambda)G)$ has no jump whereas the function $\lambda \mapsto J(\lambda G + (1-\lambda)F)$ has a jump on $[0,1]$ except for $J = \mathrm{ES}_p$. Hence, except for $J = \mathrm{ES}_p$, the function $\lambda \mapsto I_h(\lambda F + (1-\lambda)G)$ does not take a constant value on $[0,1]$, and I_h cannot have CxLS on $\mathcal{M}(0)$.

(d) $a_1 > 0, a_2 < 0$. We claim that if I_h has CxLS on $\mathcal{M}(0)$, then either $(I, J) = (\text{ess-inf}, \mathrm{VaR}_p^+)$, included in case (iv), or $(I, J) = (\mathrm{ES}_p^-, \mathrm{ES}_p)$, included in case (ii). Below we show our assertion.

First, suppose that $I = \mathrm{VaR}_\alpha^{c_1}$. For $\epsilon \in (\alpha, p)$, let $F = \epsilon\delta_{a_2} + (1-\epsilon)\delta_0$ and $G = p\delta_0 + (1-p)\delta_{a_1}$. We can easily calculate $\mathrm{VaR}_p(F) = \mathrm{VaR}_p(G) = 0$, $I(F) = a_2$, $J(F) = 0$, $I(G) = 0$ and $J(G) = a_1$. Therefore, $I_h(F) = I_h(G) = a_1 a_2$. Note that, for $\lambda \in [0,1]$,

$$\lambda F + (1-\lambda)G = \lambda\epsilon\delta_{a_2} + (\lambda(1-\epsilon) + (1-\lambda)p)\delta_0 + (1-\lambda)(1-p)\delta_{a_1}.$$

If I_h has CxLS on $\mathcal{M}(0)$, then $I_h(\lambda F + (1-\lambda)G) = a_1 a_2$ for all $\lambda \in [0,1]$ and all $\epsilon \in (\alpha, p)$. Note that the function $\lambda \mapsto I(\lambda F + (1-\lambda)G)$ has a jump at $\lambda_1 = \alpha/\epsilon \in [0,1)$, and the function $\lambda \mapsto J(\lambda G + (1-\lambda)F)$ either has no jump ($J = \mathrm{ES}_p$), a jump at $\lambda_2 = 1$ ($J = Q_p^c$ for $c \neq 1$), a jump at $\lambda_2 = 0$ ($J = \mathrm{VaR}_p^+ = Q_p^1$), or a jump at $\lambda_2 = 1 - (1-\beta)/(1-p)$ ($J = \mathrm{VaR}_\beta^c$). Note that $\lambda_1 \neq \lambda_2$ for almost every $\epsilon \in (\alpha, p)$, except for the case $(\alpha, c) = (0, 1)$. Hence, except for $(I, J) = (\text{ess-inf}, \mathrm{VaR}_p^+)$, the function $\lambda \mapsto I_h(\lambda F + (1-\lambda)G)$ does not take a constant value on $[0,1]$, and I_h cannot have CxLS on $\mathcal{M}(0)$.

Next, suppose that $I = \mathrm{ES}_p^-$. For $\epsilon \in (0, p)$, let $F = \epsilon\delta_{a_3} + (1-\epsilon)\delta_0$ and $G = p\delta_0 + (1-p)\delta_{a_1}$, where $a_3 = pa_2/\epsilon$. We can easily calculate $\mathrm{VaR}_p(F) = \mathrm{VaR}_p(G) = 0$, $I(F) = \epsilon a_3/p = a_2$, $J(F) = 0$, $I(G) = 0$ and $J(G) = a_1$. Therefore, $I_h(F) = I_h(G) = a_1 a_2$. For $\lambda \in [0,1]$,

$$\lambda F + (1-\lambda)G = \lambda\epsilon\delta_{a_3} + (\lambda(1-\epsilon) + (1-\lambda)p)\delta_0 + (1-\lambda)(1-p)\delta_{a_1}.$$

If I_h has CxLS on $\mathcal{M}(0)$, then $I_h(\lambda F + (1-\lambda)G) = 0$ for all $\lambda \in [0,1]$. Note that the function $\lambda \mapsto I(\lambda F + (1-\lambda)G)$ has no jump whereas the function $\lambda \mapsto J(\lambda G + (1-\lambda)F)$ has a jump except for $J = \mathrm{ES}_p$. Hence, except for $J = \mathrm{ES}_p$, the function $\lambda \mapsto I_h(\lambda F + (1-\lambda)G)$ does not take a constant value on $[0,1]$, and I_h cannot have CxLS on $\mathcal{M}(0)$.

To summarize our findings in (a)-(d), all (I_h, VaR_p) with CxLS are included in cases (i)-(iv). $\square$

Proof of Proposition 5.4. By definition, it is easy to verify that $I_h(F) = I_{\bar{h}}(\bar{F})$ and $\mathrm{VaR}_p(F) = -\mathrm{VaR}_{1-p}^+(\bar{F})$ for all $F \in \mathcal{M}$ (or, see Lemma 2 of Wang et al. (2020)). Therefore, $F, G \in \mathcal{M}$ satisfy $(I_h(F), \mathrm{VaR}_p(F)) = (I_h(G), \mathrm{VaR}_p(G))$ if and only if $\bar{F}, \bar{G} \in \bar{\mathcal{M}}$ satisfy $(I_{\bar{h}}(\bar{F}), \mathrm{VaR}_{1-p}^+(\bar{F})) =$

$(I_{\bar{h}}(\bar{G}), \mathrm{VaR}^+_{1-p}(\bar{G}))$. Hence, the CxLS property of (I_h, VaR_p) on $\mathcal{M}$ and that of $(I_{\bar{h}}, \mathrm{VaR}^+_{1-p})$ on $\bar{\mathcal{M}}$ are equivalent. $\square$

A.4 Proofs in Section 6

Proof of Proposition 6.5. We define $S : \mathbb{R}^2 \to \mathbb{R}$ by letting $S(z, y) = \int_0^z Z(x, y)\,\mathrm{d}x$. Observe that for any $F \in \mathcal{M}'$, Fubini's Theorem gives

$$\int_{-\infty}^{\infty} S(z, y)\,\mathrm{d}F(y) = \int_{-\infty}^{\infty}\int_0^z Z(x, y)\,\mathrm{d}x\,\mathrm{d}F(y) = \int_0^z \int_{-\infty}^{\infty} Z(x, y)\,\mathrm{d}F(y)\,\mathrm{d}x. \tag{27}$$

Note that the integral $\int_0^z \int_{-\infty}^{\infty} Z(x, y)\,\mathrm{d}F(y)\,\mathrm{d}x$ is almost everywhere differentiable with respect to z. Differentiating (27) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\int_0^z \int_{-\infty}^{\infty} Z(x, y)\,\mathrm{d}F(y)\,\mathrm{d}x\right) = \int_{-\infty}^{\infty} Z(z, y)\,\mathrm{d}F(y).$$

Since Z is a backtest function of ρ,

$$\int_{-\infty}^{\infty} Z(z, y)\,\mathrm{d}F(y) = 0 \text{ if and only if } \rho(F) = z,$$

and the following two inequalities hold,

$$\int_{-\infty}^{\infty} Z(z, y)\,\mathrm{d}F(y) < \int_{-\infty}^{\infty} Z(\rho(F), y)\,\mathrm{d}F(y) = 0 \text{ for } z < \rho(F),$$

and

$$\int_{-\infty}^{\infty} Z(z, y)\,\mathrm{d}F(y) > \int_{-\infty}^{\infty} Z(\rho(F), y)\,\mathrm{d}F(y) = 0 \text{ for } z > \rho(F).$$

Thus, $\int_{-\infty}^{\infty} S(z, y)\,\mathrm{d}F(y)$ achieves the global minimum at and only at $\rho(F) = z$. Hence S is strictly consistent for ρ and ρ is $\mathcal{M}'$-elicitable. $\square$

Proof of Proposition 6.6. For $F \in \mathcal{M}$, let $z = c\rho(F)$, we have

$$\int_{-\infty}^{\infty} Z^*(z, y)\,\mathrm{d}F(y) = \int_{-\infty}^{\infty} cZ(z/c, y)\,\mathrm{d}F(y) = \int_{-\infty}^{\infty} cZ(\rho(F), y)\,\mathrm{d}F(y) = 0.$$

For $z_1 < z_2$, if $c > 0$, we have $z_1/c < z_2/c$, and by the fact that Z is a backtest function for ρ, we have

$$\int_{-\infty}^{\infty} cZ(z_1/c, y)\,\mathrm{d}F(y) < \int_{-\infty}^{\infty} cZ(z_2/c, y)\,\mathrm{d}F(y).$$

If $c < 0$, we have $z_1/c > z_2/c$, and in this case,

$$\int_{-\infty}^{\infty} cZ(z_1/c, y)\,\mathrm{d}F(y) < \int_{-\infty}^{\infty} cZ(z_2/c, y)\,\mathrm{d}F(y).$$

In both cases, the function $\int Z^*(z, y)\,\mathrm{d}F(y)$ is strictly increasing in z. Therefore, Z^* is a backtest function for $c\rho$. $\square$

Proof of Proposition 6.7. Without loss of generality, we assume $h(1) = 1$. If $h \in \mathcal{H}_1^*$, $I_h = c\,\text{ess-sup} + (1-c)\,\text{ess-inf}$. The non-elicitability of I_h follows essentially from Theorem 3.3 in Brehmer and Strokorb (2019), and we provide a simple argument below. Suppose I_h is backtestable and Z is a backtest function. For any $u, v \in \mathbb{R}$, $u \neq v$ and $p \in (0,1)$, let $G = p\delta_u + (1-p)\delta_v$. Note that

$$0 = \int_{-\infty}^{\infty} Z(I_h(G), y)\,\mathrm{d}G(y) = pZ(I_h(G), u) + (1-p)Z(I_h(G), v).$$

Since $I_h(G)$ does not depend on $p \in (0,1)$, we have $Z(I_h(G), u) = Z(I_h(G), v) = 0$, which implies $I_h(G) = u = v$, a contradiction. Hence I_h is not backtestable.

If $h \in \mathcal{H}_2^*$, one can easily check that it is backtestable with backtest function $Z(x,y) = x - y$, $x, y \in \mathbb{R}$. If $h \in \mathcal{H}_3^*$, then for some $\alpha \in (0,1)$, $I_h = \mathrm{VaR}_{1-\alpha}$ on $\mathcal{M}_{\mathrm{con}}^*$. One can easily check that it is backtestable with backtest function $Z(x,y) = \alpha \mathbb{1}_{\{y>x\}} + (1-\alpha)\mathbb{1}_{\{y<x\}}$, $x, y \in \mathbb{R}$. The above two backtest functions can be found in Table 3 of Acerbi and Szekely (2017). Finally, using Proposition 6.6 we get the backtest functions for the signed Choquet integral I_h. $\square$

In order to show Theorem 6.9, we need to use the following lemma, as well as Proposition 6.10.

Lemma A.5. *For a comonotonic-additive coherent risk measure $\rho : \mathcal{M}_\infty \to \mathbb{R}$ and $p \in (0,1)$, if (ρ, VaR_p) has CxLS on $\mathcal{M}_\infty^*(p)$, then it has CxLS on $\mathcal{M}_{\mathrm{dis}}$.*

Proof. Write $\rho = I_h$ where $h \in \mathcal{H}$ is concave and increasing with $h(1) = 1$. We first assume h is continuous. If h is not continuous, it can only have a jump at 0, and we will comment on that case at the end of the proof.

For $r \in \mathbb{R}$, denote by $\mathcal{M}_{\mathrm{dis}}(r) = \{F \in \mathcal{M}_{\mathrm{dis}} : \mathrm{VaR}_p(F) = r\}$ and $\mathcal{M}_\infty^*(p,r) = \{F \in \mathcal{M}_\infty^*(p) : \mathrm{VaR}_p(F) = r\}$. By Proposition 4.2, to show that (ρ, VaR_p) has CxLS on $\mathcal{M}_{\mathrm{dis}}$, it suffices to show that ρ has CxLS on $\mathcal{M}_{\mathrm{dis}}(r)$ for all $r \in \mathbb{R}$.

Fix $r \in \mathbb{R}$ and take $F, G \in \mathcal{M}_{\mathrm{dis}}(r)$ such that $\rho(F) = \rho(G)$. We construct two sequences of distributions $\{F_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_\infty^*(p)$ and $\{G_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_\infty^*(p)$ as follows. Let $\epsilon_n = (1-p)/n$, $n \in \mathbb{N}$. For $n \in \mathbb{N}$, let $\hat{F}_n^{-1}(t) = F^{-1}(t)$ for $t \in [0,p] \cup [p+\epsilon_n, 1]$, and $\hat{F}_n^{-1}$ is linear on $[p, p+\epsilon_n]$. Similarly, let $\hat{G}_n^{-1}(t) = G^{-1}(t)$ for $t \in [0,p] \cup [p+\epsilon_n, 1]$, and $\hat{G}_n^{-1}$ is linear on $[p, p+\epsilon_n]$. Note that the quantile function is always left-continuous by definition. Next, let $F_n^{-1}(t) = \min\{\hat{F}_n^{-1}(t), F^{-1}(t)\}$, $t \in [0,1]$ and $G_n^{-1}(t) = \min\{\hat{G}_n^{-1}(t), G^{-1}(t)\}$, $t \in [0,1]$. Since F^{-1} and G^{-1} may only have an up-side jump at p, and $\hat{F}_n$ and $\hat{G}_n$ have continuous quantiles at p, we know that F_n and G_n both have continuous quantiles at p, i.e., $\{F_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_\infty^*(p,r)$ and $\{G_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_\infty^*(p,r)$. It is also easy to see that $F_n \to F$ and $G_n \to G$ weakly.

Note that $F_n^{-1}(t) = F^{-1}(t)$ for t close to 0 and 1, and hence $\{F_n^{-1}\}_{n\in\mathbb{N}}$ is bounded above and below. Therefore, the uniform integrability condition in Theorem 4 of Wang et al. (2020) is satisfied, and we have $\rho(F_n) \to \rho(F)$ as $n \to \infty$. Similarly, $\rho(G_n) \to \rho(G)$ as $n \to \infty$.

Let B be the Bernoulli distribution with mean $c = (1-p)/2$. Since $\rho \geq \mathbb{E}$, we know that $\rho(B) \geq c > 0$. Further, since $F_n^{-1}(t) \leq F^{-1}(t)$ for all $t \in [0,1]$, we know $\rho(F_n) \leq \rho(F)$ since a coherent risk measure is monotone with respect to stochastic order. For $n \in \mathbb{N}$, let F_n^* be given by

$$(F_n^*)^{-1}(t) = F_n^{-1}(t) + \mathbb{1}_{\{t>1-c\}} \frac{\rho(F) - \rho(F_n)}{\rho(B)}, \quad t \in [0,1].$$

Note that $(F_n^*)^{-1}$ is increasing and left-continuous, thus a well-defined quantile function. We can calculate, using the comonotonic-additivity of ρ, that

$$\rho(F_n^*) = \rho(F_n) + \rho(B)\frac{\rho(F) - \rho(F_n)}{\rho(B)} = \rho(F).$$

Moreover, $(F_n^*)^{-1}(p) = F_n^{-1}(p) = F^{-1}(p) = r$. Therefore, $\{F_n^*\}_{n\in\mathbb{N}} \subset \mathcal{M}_\infty^*(p,r)$. On the other hand, since $\rho(F_n) \to \rho(F)$ as $n \to \infty$, we have $F_n^* \to F$ weakly. Similarly, we can construct $\{G_n^*\}_{n\in\mathbb{N}} \subset \mathcal{M}_\infty^*(p,r)$ such that $\rho(G_n^*) = \rho(G)$, $n \in \mathbb{N}$, and $G_n^* \to G$ weakly.

It is clear that $\lambda F_n^* + (1-\lambda)G_n^* \to \lambda F + (1-\lambda)G$ weakly for $\lambda \in [0,1]$. By noting again the uniform integrability in the sense of Theorem 4 of Wang et al. (2020) is satisfied by $\{\lambda F_n^* + (1-\lambda)G_n^*\}_{n\in\mathbb{N}}$, we have $\rho(\lambda F_n^* + (1-\lambda)G_n^*) \to \rho(\lambda F + (1-\lambda)G)$. Thus,

$$\rho(F) = \rho(\lambda F_n^* + (1-\lambda)G_n^*) \to \rho(\lambda F + (1-\lambda)G),$$

and hence ρ has CxLS on $\mathcal{M}_{\rm dis}(r)$. This shows that (ρ, VaR_p) has CxLS on $\mathcal{M}_{\rm dis}$.

If h has a jump at 0, then ρ can be decomposed into a convex combination of ess-sup and I_g for a continuous and concave $g \in \mathcal{H}$. Since $F_n^{-1}(1) = F^{-1}(1)$ and $G_n^{-1}(1) = G^{-1}(1)$, this does not affect the arguments that $\rho(F_n) \to \rho(F)$ and $\rho(G_n) \to \rho(G)$ as $n \to \infty$, or the construction of F_n^* and G_n^*. □

Proof of Proposition 6.10. Fix $F \in \mathcal{M}_\infty^*(p)$. For $x_1, x_2 \in \mathbb{R}$, let $H(x_1,x_2) = \int_{-\infty}^{\infty} S(x_1,x_2,y)\,\mathrm{d}F(y)$, and $G(x_1) = \int_{-\infty}^{\infty}(x_1 + \frac{1}{1-p}(y-x_1)_+)\,\mathrm{d}F(y)$. Both H and G are clearly $\mathbb{R}$-valued. Obviously,

$$H(x_1,x_2) = g(x_2) + g'(x_2)\left(aG(x_1) + (1-a)\mathbb{E}[F] - x_2\right).$$

Using the ES-VaR relation (1), and noting that F has a continuous quantile at p, we have $\mathrm{VaR}_p(F) = \arg\min_{x_1\in\mathbb{R}} G(x_1)$ and $\mathrm{ES}_p(F) = \min_{x_1\in\mathbb{R}} G(x_1)$. Hence, for $x_2 \in \mathbb{R}$, since $ag'(x_2) > 0$, a minimizer of $H(x_1,x_2)$ satisfies $x_1 = \mathrm{VaR}_p(F)$. Note that

$$H(\mathrm{VaR}_p(F), x_2) = g(x_2) + g'(x_2)(a\mathrm{ES}_p(F) + (1-a)\mathbb{E}(F) - x_2) = g(x_2) + g'(x_2)(\rho(F) - x_2).$$

Since g is strictly concave, we have $g(x_2) + g'(x_2)(\rho(F) - x_2) \geq g(\rho(F))$ for $x_2 \in \mathbb{R}$ and the equality is attained at $x_2 = \rho(F)$. Therefore,

$$(\mathrm{VaR}_p(F), \rho(F)) = \underset{(x_1,x_2)\in\mathbb{R}^2}{\arg\min} \int_{-\infty}^{\infty} S(x_1, x_2, y)\, \mathrm{d}F(y),$$

which shows that S is a strictly consistent function for (VaR_p, ρ). □

Proof of Theorem 6.9. To show the "only-if" part, note that the elicitability of (ρ, VaR_p) implies that it has CxLS on $\mathcal{M}^*_\infty(p)$, and ρ is an increasing Choquet integral. By Lemma A.5, we know that (ρ, VaR_p) has CxLS on $\mathcal{M}^*_{\mathrm{dis}}$. Hence, we know that ρ is one of the four cases in Theorem 5.3. Clearly, case (ii) gives a possible coherent risk measure ρ of the form $\rho = a\mathrm{ES}_p + (1-a)\mathbb{E}$ for $a \in [0, 1]$, and all other forms of ρ in Theorem 5.3 are not coherent. The "if" part for $a \in (0, 1]$ follows from Proposition 6.10, and the case $a = 0$ is due to Proposition 4.1 since both $\mathbb{E}$ and VaR_p are $\mathcal{M}^*_\infty(p)$-elicitable. □

Proof of Corollary 6.12. We need to show that ES_p is the only lower semi-continuous risk measure within the class of risk measures identified in Theorem 6.9. For this purpose, we verify two things.

First, ES_p is lower semi-continuous with respect to weak convergence. This is implied by Proposition 2.1 and Remark 2.1 of Wang and Zitikis (2020).

Second, $a\mathrm{ES}_p + (1-a)\mathbb{E}$ is not lower semi-continuous for $a \in [0, 1)$. This fact is indeed shown by a counter-example in Wang and Zitikis (2020), which we give below because it is very simple. Let F_k be the distribution of $-kX_k$, where $X_k \sim \mathrm{Bernoulli}(1/k)$. Clearly, $F_k \to \delta_0$ weakly, $\mathbb{E}[F_k] = -1$ and $\mathrm{ES}_p(F_k) = 0$ for $k > 1/(1-p)$. Therefore, by lower semi-continuity of ρ,

$$0 = \rho(\delta_0) \leq \liminf_{k\to\infty} (a\mathrm{ES}_p(F_k) + (1-a)\mathbb{E}[F_k]) = -(1-a),$$

showing that $a \geq 1$. □

References

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

Acerbi, C. and Szekely, B. (2017). General properties of backtestable statistics. *SSRN*: 2905109.

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

BCBS (2016). *Minimum capital requirements for Market Risk. January 2016.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements.

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

Bickel, P. J. and Lehmann, E. L. (1976). Descriptive statistics for nonparametric models. III. Dispersion. *The Annals of Statistics*. **4**(6). 1139–1158.

Brehmer, J. R. and Strokorb, K. (2019). Why scoring functions cannot assess tail properties. *Electronic Journal of Statistics*, **13**(2), 4015–4034.

Chew, S.H. (1989). Axiomatic utility theories with the betweenness property. *Annals of Operations Research*, **19**, 273–298.

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

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

Dekel, E. (1986). An axiomatic characterization of preferences under uncertainty: weakening the independence axiom. *Journal of Economic Theory*, **40**, 304–318.

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

De Waegenaere, A. and Wakker, P. P. (2001). Nonmonotonic Choquet integrals. *Journal of Mathematical Economics*, **36**, 45–60.

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

Fissler, T., Hlavinová, J. and Rudloff, B. (2019a). Elicitability and identifiability of systemic risk measures. *arXiv*: 1907.01306v2.

Fissler, T., Hlavinová, J. and Rudloff, B. (2019b). Forecast evaluation of set-valued functionals. *arXiv*: 1910.07912.

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

Fissler, T. and Ziegel, J. F. (2019). Elicitability of Range Value at Risk. *arXiv*: 1902.04489.

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

Frongillo, R. and Kash, I. A. (2018). Elicitation complexity of statistical properties. *arXiv:* 1506.07212v2.

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

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

Heinrich, C. (2014). The mode functional is not elicitable. *Biometrika*, **101**(1), 245–251.

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

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

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

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

Liu, P. and Schied, A. and Wang, R. (2020). Distributional transforms, probability distortions, and their applications. *Mathematics of Operations Research*, forthcoming.

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

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

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

Osband, K. (1985). *Providing Incentives for Better Cost Forecasting*. Ph.D. Thesis, University of California, Berkeley.

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

Rockafellar, R. T. and Uryasev, S. (2013). The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, **18**(1-2), 33–53.

Rockafellar, R. T., Uryasev, S. and Zabarankin, M. (2006). Generalized deviation in risk analysis. *Finance and Stochastics*, **10**, 51–74.

Steinwart, I., Pasin, C., Williamson, R. and Zhang, S. (2014). Elicitation and identification of properties. *JMLR: Workshop and Conference Proceedings*, **35**, 1–45.

Wang, R., Wei, Y. and Willmot, G. E. (2020). Characterization, robustness and aggregation of signed Choquet integrals. *Mathematics of Operations Research*, forthcoming.

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

Wang, R. and Zitikis, R. (2020). An axiomatic foundation for the Expected Shortfall. *Management Science*, forthcoming.

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

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

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