Bayes risk, elicitability, and the Expected Shortfall

Paul Embrechts[*] Tiantian Mao[†] Qiuqi Wang[‡] Ruodu Wang[§]

March 21, 2021

Abstract

Motivated by recent advances on elicitability of risk measures and practical considerations of risk optimization, we introduce the notions of Bayes pairs and Bayes risk measures. Bayes risk measures are the counterpart of elicitable risk measures, extensively studied in the recent literature. The Expected Shortfall (ES) is the most important coherent risk measure in both industry practice and academic research in finance, insurance, risk management, and engineering. One of our central results is that under a continuity condition, ES is the only class of coherent Bayes risk measures. We further show that entropic risk measures are the only risk measures which are both elicitable and Bayes. Several other theoretical properties and open questions on Bayes risk measures are discussed.

Keywords: Bayes risk, quantiles, Expected Shortfall, elicitability, entropic risk measures

---

Dedication

We dedicate our paper to the memory of Mark H. A. Davis who sadly passed away on March 18, 2020. For those who knew Mark personally, he will always be remembered as a trusted friend whose gentle personality charmed during so many encounters. We all will no doubt recall his shining brilliance as a teacher and researcher. He was one of the rare academic lighthouses bringing clarity to the occasional darkness in the field of mathematical finance. The first author (PE) vividly recalls a talk he (PE) gave at Imperial College on March 7, 2013, with title "Model uncertainty and risk aggregation". During this talk, PE mentioned the notion of elicitability as discussed in the paper

[*]RiskLab, Department of Mathematics, and ETH Risk Center, ETH Zurich, Switzerland. ✉ embrechts@math.ethz.ch

[†]School of Management, University of Science and Technology of China, China. ✉ tmao@ustc.edu.cn

[‡]Department of Statistics and Actuarial Science, University of Waterloo, Canada. ✉ q428wang@uwaterloo.ca

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

Gneiting (2011). It was clear that the latter paper would become fundamental to Quantitative Risk Management. In a wonderful paper, put online on October 19, 2013, Davis (2013) compares and contrasts the elicitability approach to the notion of prequential statistics as developed in Dawid (1984). He graciously states, "The author thanks Paul Embrechts for introducing him to this topic." We further quote from the paper "The direct inspiration for this work comes from the exceptionally stimulating paper Dawid and Vovk (1999). It seems surprising that this whole circle of ideas is totally ignored in almost all published discussions of risk management." We very much hope that our paper will renew interest in the above debate. The fourth author (RW) recalls reading Davis (2013) when visiting ETH Zurich in fall 2013. It was precisely that paper, along with Gneiting (2011), which introduced to RW the world of elicitability and statistical forecasting of risk measures, a field to which Mark Davis continued to contribute substantially; see Davis (2016, 2017). We definitely urge the interested reader to also keep Davis (2014) at the corner of her/his desk. Most unfortunately, Mark Davis will not be there to provide the necessary light on the resulting discussions.

1 Introduction

Mainly through Gneiting (2011), the concept of elicitability has drawn considerable interest within the quantitative risk management literature. The concept is fundamental when comparing different forecasting procedures. We refer to the latter paper for an excellent introduction. In this Introduction we explain our motivation, and more detailed definitions are given in Section 2.

Let $\mathcal{X}$ be a linear space of random variables. A set-valued d-dimensional functional $\mathcal{S} : \mathcal{X} \to 2^{\mathbb{R}^d}$ is *elicitable* if there exists a measurable function $L : \mathbb{R}^{d+1} \to \mathbb{R}$ (called a loss function) such that

$$\mathcal{S}(X) = \underset{x\in\mathbb{R}^d}{\arg\min}\, \mathbb{E}[L(x, X)], \quad X \in \mathcal{X}. \tag{1}$$

In the recent literature, a lot of research has been done on characterizing risk measures that are elicitable for $d = 1$.[1] The case of $d \geq 2$ is much more difficult; see Fissler and Ziegel (2016) and Wang and Wei (2020). In sharp contrast to the functional $\mathcal{S}$ in (1), much less attention has been paid to functionals $\mathcal{R}$ of the form

$$\mathcal{R}(X) = \min_{x\in\mathbb{R}^d} \mathbb{E}[L(x, X)], \quad X \in \mathcal{X}, \tag{2}$$

[1]For characterization of elicitability in dimension $d = 1$, see Ziegel (2016) on coherent risk measures, Bellini and Bignozzi (2015) and Delbaen et al. (2016) on convex risk measures, Kou and Peng (2016) and Wang and Ziegel (2015) on Choquet risk measures, and Liu and Wang (2021) on tail risk measures.

which computes the *Bayes risk* of X for the *Bayes estimator* $\mathcal{S}$. As an important example for $d = 1$, the Value-at-Risk (VaR) at level α is elicited by the loss function $L : (x, y) \mapsto x + \frac{1}{1-\alpha}(y - x)_+$, and the Expected Shortfall (ES) at level α is the corresponding Bayes risk (e.g., Rockafellar and Uryasev (2002)); see (4) in Section 2.

The functionals in (1) and (2) are both important in the context of expected loss minimization, and yet only the characterization of (1) is extensively studied in the literature of elicitability. In case $d = 1$, the functional $\mathcal{S}$ in (1) is a *minimizer* and the functional $\mathcal{R}$ in (2) is a *minimum*. The classic interpretation of risk measures, as in Artzner et al. (1999), is the *least amount of capital* needed for a financial loss to be acceptable for the regulator. In other words, an acceptable capital reserve needs to be no smaller than the value of the risk measure. With this interpretation, requiring for a capital reserve to be larger than the minimizer $\mathcal{S}$ does not have a clear financial meaning. On the other hand, the minimum in the Bayes risk $\mathcal{R}$ may be interpreted as a "generalized L-distance" from X to the real line,[2] so that the corresponding capital reserve may be interpreted as a penalty for deviating from constancy, thus for bearing risk.[3]

The main focus of this paper is $(\mathcal{S}, \mathcal{R})$ in (1) and (2), which we call a Bayes pair. After the formal definition of Bayes pairs and Bayes risk measures in Section 2, we derive two main characterization results. In Theorem 2 we show that, under a continuity assumption, an ES is the only Bayes risk measure that is either coherent or Choquet, and in Theorem 4 we pin down entropic risk measures as the only monetary risk measures which are both elicitable and Bayes.

Currently, ES is the standard risk measure in the Fundamental Review of the Trading Book (BCBS (2019)) in banking. Our characterization of ES as the only coherent Bayes risk measure strengthens the unique role of ES from the perspective of elicitability. This result complements the recent finding of Wang and Zitikis (2021) on an axiomatic characterization of ES from the perspective of portfolio risk aggregation. See also Emmer et al. (2015) and Embrechts et al. (2018) for discussions on comparative advantages of VaR and ES as regulatory risk measures.

Our main technical results are closely related to those in Weber (2006), Ziegel (2016) and Wang and Wei (2020) on risk measures with convex level sets, those in Ben-Tal and Teboulle (2007) on optimized certainty equivalents, those in Rockafellar and Uryasev (2013) on risk quadrangles, and those in Frongillo and Kash (2021) on elicitation complexity. More general discussions on elicitabil-

[2] For instance, in the simple case $L(x, y) = (x - y)^2$, the Bayes risk $\min_{x \in \mathbb{R}} \mathbb{E}[(X - x)^2] = \mathrm{var}(X)$, which is the squared $\mathcal{L}^2$-distance from X to the real line. If $\mathcal{R}$ is used as a regulatory risk measure, we typically need to adjust the value by the location of X (that is why we call it a "generalized L-distance"), e.g., using $L(x, y) = x + \lambda(x - y)^2$ for $\lambda > 0$ would give rise to a mean-variance risk measure. The minimum in (2) should not be interpreted as a minimum over economic scenarios; indeed, it is more natural to take a maximum over economic scenarios.

[3] We do acknowledge that elicitability is an important statistical property, especially when comparing competing forecast procedures; see e.g., Fissler and Ziegel (2016). Our paper stresses the important practical difference, even complementarity, between regulatory interpretation (Bayes) and statistical tractability (elicitability).

ity and forecasting risk measures can be found in Davis (2013, 2016) and Nolde and Ziegel (2017). Although we do focus on risk measures, the general theory of elicitability has wide applications outside of finance. For some recent work on interval-valued elicitable functionals, see Fissler et al. (2020) and Brehmer and Gneiting (2021). Elicitability is also closely related to empirical risk minimization; see e.g., Lambert et al. (2008), Steinwart et al. (2014) and Frongillo and Kash (2021) in the context of machine learning.

2 Bayes pairs and Bayes risk measures

2.1 Risk measures

In Definition 1 below, we slightly generalize the standard definition of scalar risk measures in Artzner et al. (1999) and Föllmer and Schied (2002) to interval-valued risk measures such as quantiles. In what follows, equalities and inequalities between intervals are understood as holding for both end-points, and so are addition and scalar multiplication. Denote by $(\Omega, \mathcal{F}, \mathbb{P})$ an atomless probability space. For $p \in [1, \infty)$, let $\mathcal{L}^p$ represent the space of random variables with finite p-th moment, and let $\mathcal{L}^\infty$ represent the set of bounded variables. Throughout, $\mathcal{X}$ is a linear space of random variables containing $\mathcal{L}^\infty$, representing the domain of risk measures. Let $I(\mathbb{R})$ be the set of closed real intervals, including $(-\infty, a]$ and $[a, \infty)$, and the interval $[a, a]$ is identified with its element a (hence, $\mathbb{R}$ is treated as a subset of $I(\mathbb{R})$).

Definition 1. A *risk measure* $\mathcal{S}$ is a mapping from $\mathcal{X}$ to $I(\mathbb{R})$, and it is *scalar* if it maps $\mathcal{X}$ to $\mathbb{R}$. A risk measure $\mathcal{S}$ is *monetary* if it is (i) monotone: $\mathcal{S}(X) \le \mathcal{S}(Y)$ for $X, Y \in \mathcal{X}$ with $X \le Y$, and (ii) translation invariant: $\mathcal{S}(X + c) = \mathcal{S}(X) + c$ for all $X \in \mathcal{X}$ and $c \in \mathbb{R}$. A scalar risk measure $\mathcal{S}$ is *coherent* if it is monetary, (iii) convex: $\mathcal{S}(\lambda X + (1-\lambda)Y) \le \lambda \mathcal{S}(X) + (1-\lambda)\mathcal{S}(Y)$ for all $X, Y \in \mathcal{X}$ and $\lambda \in [0, 1]$, and (iv) positively homogeneous: $\mathcal{S}(\lambda X) = \lambda \mathcal{S}(X)$ for all $X \in \mathcal{X}$ and $\lambda \in (0, \infty)$. A scalar risk measure $\mathcal{S}$ is *Choquet* if it is monetary and (v) comonotonic-additive: $\mathcal{S}(X + Y) = \mathcal{S}(X) + \mathcal{S}(Y)$ for all comonotonic $X, Y \in \mathcal{X}$.[4]

It was shown by Schmeidler (1986) that comonotonic-additivity characterizes Choquet integrals, and hence we use the name Choquet risk measure. Law-invariant Choquet risk measures are also called *distortion risk measures*;[5] Theorem 1 of Wang et al. (2020) gives a characterization of law-invariant comonotonic-additive functionals. Both coherence and comonotonic-additivity are

[4] Two random variables X and Y are said to be *comonotonic*, if there exists a random variable U and two increasing functions f, g such that $X = f(U)$ and $Y = g(U)$ almost surely. Such U can be chosen as $\mathrm{U}[0, 1]$ distributed, and f and g can be chosen as the quantile functions of X and Y, respectively.

[5] A risk measure $\mathcal{S}$ is *law invariant* if $\mathcal{S}(X) = \mathcal{S}(Y)$ whenever X and Y have the same distribution.

argued as desirable properties, and a law-invariant risk measure that is both coherent and Choquet is called a spectral risk measure by Acerbi (2002).

We make an important clarification concerning Definition 1, which also justifies our focus on risk measures in (2). It is well known that with positive homogeneity, convexity is equivalent to subadditivity: $\mathcal{S}(X+Y) \le \mathcal{S}(X) + \mathcal{S}(Y)$ for all $X, Y \in \mathcal{X}$, which is not easy to financially interpret if $\mathcal{S}$ is interval-valued. In view of this, convexity and coherence are suitable properties for risk measures in (2), but it is unclear whether they are suitable for risk measures in (1), unless one additionally assumes uniqueness of the optimizer, or some convention (e.g., using the left end-point) is imposed.

The most important examples of risk measures are VaR and ES, widely used in financial regulation. At a probability level $\alpha \in [0,1]$, VaR has two versions, the left- and right-quantiles. We define $\mathrm{VaR}_\alpha : \mathcal{X} \to I(\mathbb{R})$ by $\mathrm{VaR}_\alpha(X) = [\mathrm{VaR}^-_\alpha(X), \mathrm{VaR}^+_\alpha(X)]$, where

$$\begin{cases} \mathrm{VaR}^-_\alpha(X) &= \inf\{x \in \mathbb{R} : \mathbb{P}(X \le x) \ge \alpha\}; \\ \mathrm{VaR}^+_\alpha(X) &= \inf\{x \in \mathbb{R} : \mathbb{P}(X \le x) > \alpha\}. \end{cases} \tag{3}$$

By definition, $\mathrm{VaR}^-_0 = -\infty$ and $\mathrm{VaR}^+_1 = \infty$. The interval-valued risk measure VaR_α is monetary. The ES (also called CVaR, TVaR and AVaR) at a probability level $\alpha \in [0,1)$ is defined as

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}^-_\beta(X)\,\mathrm{d}\beta, \quad X \in \mathcal{X}.$$

It is well known that an ES is both coherent and Choquet. Rockafellar and Uryasev (2002) obtained the following ES-VaR relation (4)

$$\begin{aligned} [\mathrm{VaR}^-_\alpha(X), \mathrm{VaR}^+_\alpha(X)] &= \operatorname*{arg\,min}_{x \in \mathbb{R}} \left\{ x + \frac{1}{1-\alpha}\mathbb{E}[(X-x)_+] \right\}; \\ \mathrm{ES}_\alpha(X) &= \min_{x \in \mathbb{R}} \left\{ x + \frac{1}{1-\alpha}\mathbb{E}[(X-x)_+] \right\}. \end{aligned} \tag{4}$$

The relation (4) will be used repeatedly in this paper.

2.2 Bayes pairs and Bayes risk measures

To define the main objects of the paper, we follow the standard terminology of Bayes estimator and Bayes risk in statistical decision theory. Despite this terminology, our discussion stays purely within the theory of risk measures, and does not require specific knowledge on Bayesian statistics to understand.

Definition 2. A pair of risk measures $(\mathcal{S}, \mathcal{R}) : \mathcal{X}^2 \to I(\mathbb{R}) \times \mathbb{R}$ is a *Bayes pair* if for some Borel function $L : \mathbb{R}^2 \to \mathbb{R}$, called the *loss function*,

$$\mathcal{S}(X) = \operatorname*{arg\,min}_{x \in \mathbb{R}} \mathbb{E}[L(x, X)] \quad \text{and} \quad \mathcal{R}(X) = \min_{x \in \mathbb{R}} \mathbb{E}[L(x, X)], \quad X \in \mathcal{X}. \tag{5}$$

If $\mathcal{S}$ is further translation invariant, then we call $\mathcal{S}$ a *Bayes estimator*, and $\mathcal{R}$ a *Bayes risk measure*.[6]

In a Bayes pair $(\mathcal{S}, \mathcal{R})$, $\mathcal{R}$ is always scalar, whereas $\mathcal{S}$ not necessarily. Therefore, it is appropriate to consider conditions for $\mathcal{R}$, instead of $\mathcal{S}$, to be a coherent risk measure. By (4), for $\alpha \in [0, 1)$, the pair $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ is a Bayes pair, and ES_α is a coherent Bayes risk measure. Obviously, a Bayes pair is always law invariant.

Let us first explain the important requirement of $\mathcal{S}$ being translation invariant in Definition 2. In the next theorem, we show a negative result: if we do not impose any conditions on $\mathcal{S}$, then the interpretation of Bayes estimator and Bayes risk is lost.

Theorem 1. *A risk measure $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ satisfies (5) for some $\mathcal{S} : \mathcal{X} \to 2^{\mathbb{R}}$ and loss function L if and only if there exists a set $\mathcal{A}$ of real Borel functions such that*

$$\mathcal{R}(X) = \min_{\ell \in \mathcal{A}} \mathbb{E}[\ell(X)], \quad X \in \mathcal{X}. \tag{6}$$

Proof. The $\Rightarrow$ implication follows directly by setting $\mathcal{A} = \{y \mapsto L(x, y) : x \in \mathbb{R}\}$. Next we show the $\Leftarrow$ implication. Let ϕ be a one-to-one mapping from $\mathbb{R}$ to the set of real Borel functions on $\mathbb{R}$ (since both sets have the same cardinality), and define $L(x, y) = \phi(x)(y)$ for $x, y \in \mathbb{R}$. Let

$$\mathcal{S}(X) = \phi^{-1}\left(\operatorname*{arg\,min}_{\ell \in \mathcal{A}} \mathbb{E}[\ell(X)]\right) = \operatorname*{arg\,min}_{x \in \mathbb{R}} \mathbb{E}[L(x, X)]. \tag{7}$$

Hence, (5) holds. $\square$

The negative result in Theorem 1 is very simple, but it is important for the motivation behind the concept of Bayes risk measures as in Definition 2. If no property is imposed on $\mathcal{S}$, then we can directly define $\mathcal{R}$ by (6) without introducing $\mathcal{S}$. However, this would be problematic because the Bayes estimator $\mathcal{S}$ is not interpretable as there is nothing to estimate. A similar problem appears in Frongillo and Kash (2021) when they define elicitation complexity.[7] In the context of Bayes

[6] $\mathcal{S}$ is also called a Bayes act; see e.g., Grünwald and Dawid (2004). When we say that a risk measure is Bayes in this paper, we mean that it is a Bayes risk measure (instead of a Bayes estimator).

[7] Frongillo and Kash (2021) argue that, through a one-to-one mapping from $\mathbb{R}$ to the set of real Borel functions on $\mathbb{R}$ like ϕ in the proof of Theorem 1, one arrives at a counter-intuitive statement that all functionals have elicitation complexity 1. Hence, some regularity requirements are needed.

estimation, both $\mathcal{S}$ and $\mathcal{R}$ have a concrete meaning: $\mathcal{S}$ is the estimated parameter and $\mathcal{R}$ is the risk of this estimation. Therefore, directly defining a risk measure $\mathcal{R}$ by (6) cannot be called a Bayes risk measure because it is not the Bayes risk of any interpretable parameter. For this reason, we impose translation invariance on $\mathcal{S}$, which means that the parameter of interest of the unknown financial loss is additive under location shift. This is similar to the consideration of Artzner et al. (1999), where location shift is interpreted as capital injection. Other types of regularization on $\mathcal{S}$ may also be considered, among which translation invariance seems both natural and easy to work with. See Examples 2 and 3 in Section 3 for instances of $\mathcal{R}$ in (5) where translation invariance of $\mathcal{S}$ is not assumed.

For a Bayes pair $(\mathcal{S}, \mathcal{R})$ with loss function L, by defining

$$\mathcal{R}' : X \mapsto \lambda \mathcal{R}(X) + (1-\lambda)\mathbb{E}[f(X)] \tag{8}$$

for any real function f and $\lambda \in (0,1]$, the pair $(\mathcal{S}, \mathcal{R}')$ is also a Bayes pair with loss function $L' : (x, y) \mapsto \lambda L(x, y) + (1-\lambda) f(y)$. Hence, some conditions on $\mathcal{R}$ also need to be imposed to obtain an economically meaningful risk measure. For this, we have plenty of candidates in the literature, notably in the theories of coherent and Choquet risk measures.

Some advantages of Bayes risk measures follow from the definition and results in this paper, and we briefly summarize them below. Bayes risk measures are (i) convenient to optimize due to their form as a minimizer to a linear mapping on distributions;[8] (ii) concave in mixtures and thus correctly measuring randomness (see Section 5); (iii) relatively easy to evaluate forecasts due to their second-order elicitability (Corollary 1); (iv) relatively easy to compute due to their low elicitation complexity (Frongillo and Kash (2021)), which is at most 2.[9]

2.3 Examples of Bayes pairs

We present some common examples of Bayes pairs. Except for the ES/$\mathbb{E}$-mixture, none of the other Bayes risk measures in Example 1 are coherent risk measures; this gives a hint on the unique role of ES/$\mathbb{E}$-mixtures among coherent Bayes risk measures.

[8] A typical optimization problem is to minimize $\mathcal{R}(f(a, Y))$ over $a \in A$ where A is a set of actions, Y is a random vector, and $f : A \times \mathbb{R}^d \to \mathbb{R}$ is a function; this includes the classic problem of portfolio selection with risk measures. If $\mathcal{R}$ is a Bayes risk measure with loss function L, then via the relationship $\min_a \mathcal{R}(f(a, Y)) = \min_x \min_a \mathbb{E}[L(x, f(a, Y))]$, the above optimization problem can be solved by first minimizing an expected loss over $a \in A$, which is well studied. See Rockafellar and Uryasev (2013) and the references therein for optimizing risk measures of the form (2).

[9] Roughly speaking, the elicitation complexity of a functional $\mathcal{R}$ is the lowest dimension of $\mathcal{R}'$ such that (i) $\mathcal{R}'$ is elicitable; (ii) $\mathcal{R}$ is determined by $\mathcal{R}'$; (iii) $\mathcal{R}'$ satisfies some regularity conditions. We omit a detailed definition in this paper since some heavy preparation is needed for a proper definition of elicitation complexity. The interested reader is referred to Frongillo and Kash (2021).

Example 1. In all examples below, $\mathcal{S}$ in a Bayes pair $(\mathcal{S}, \mathcal{R})$ is translation invariant, and hence $\mathcal{R}$ is a Bayes risk measure in Definition 2.

(i) $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha^\lambda)$: As we have seen from (4), for $\alpha \in [0,1)$, $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ is a Bayes pair with loss function $L : (x, y) \mapsto x + \frac{1}{1-\alpha}(y-x)_+$, and ES_α is a coherent Bayes risk measure. Moreover, using (8), the convex combination of ES_α and $\mathbb{E}$, called an ES/$\mathbb{E}$-*mixture* and denoted by $\mathrm{ES}_\alpha^\lambda$, i.e.,

$$\mathrm{ES}_\alpha^\lambda = \lambda \mathrm{ES}_\alpha + (1-\lambda)\mathbb{E}, \quad \lambda \in [0,1],\ \alpha \in (0,1), \tag{9}$$

is a coherent Bayes risk measure with loss function $L : (x, y) \mapsto x + \lambda(y-x) + \frac{1-\lambda}{1-\alpha}(y-x)_+$. Note that $\lambda = 0$ corresponds to the mean, and $\lambda = 1$ corresponds to ES_α.

(ii) $(\mathrm{ER}_\gamma, \mathrm{ER}_\gamma)$: An entropic risk measure (ER) is defined as

$$\mathrm{ER}_\gamma(X) = \frac{1}{\gamma} \log \mathbb{E}[e^{\gamma X}], \quad X \in \mathcal{L}^\infty,$$

for $\gamma \in (0, \infty)$, with the limiting case $\mathrm{ER}_0 = \mathbb{E}$.[10] The entropic risk measure ER_γ is known to be convex but not coherent. Next we see that ER_γ is both Bayes and elicitable for the same loss function $L : (x, y) \mapsto x + (e^{\gamma(y-x)} - 1)/\gamma$. Indeed, by defining

$$\mathcal{R}(X) := \min_{x \in \mathbb{R}} \mathbb{E}[L(x, X)] = \min_{x \in \mathbb{R}} \left\{ x + \frac{1}{\gamma}\mathbb{E}[e^{\gamma(X-x)} - 1] \right\}, \tag{10}$$

one can verify that the minimizer of (10) is $\mathcal{S}(X) = \frac{1}{\gamma} \log \mathbb{E}[e^{\gamma X}] = \mathrm{ER}_\gamma(X)$. Substituting it into (10), we have

$$\mathcal{R}(X) = \frac{1}{\gamma} \log \mathbb{E}[e^{\gamma X}] + \frac{\mathbb{E}[e^{\gamma X}]}{\mathbb{E}[\gamma e^{\gamma X}]} - \frac{1}{\gamma} = \frac{1}{\gamma} \log \mathbb{E}[e^{\gamma X}] = \mathrm{ER}_\gamma(X).$$

(iii) $(\mathbb{E}, \sigma^2)$: The variance

$$\sigma^2(X) := \mathbb{E}[(X - \mathbb{E}[X])^2] = \min_{x \in \mathbb{R}} \mathbb{E}[(X-x)^2], \quad X \in \mathcal{L}^2,$$

is a Bayes risk measure with loss function $L : (x, y) \mapsto (y-x)^2$. The corresponding Bayes estimator is the mean $\mathbb{E}$.

[10] The domain of ER_γ can be enlarged to include random variables with finite exponential moments, such as normal random variables.

(iv) $(\mathbb{E}, \mathbb{E} + \lambda\sigma^2)$: The mean-variance functional, for $\lambda > 0$,

$$\mathbb{E}[X] + \lambda\sigma^2(X), \quad X \in \mathcal{L}^2,$$

is a Bayes risk measure with loss function $L : (x, y) \mapsto y + \lambda(y - x)^2$. The Bayes estimator corresponding to the mean-variance functional is also the mean $\mathbb{E}$.

(v) $(\mathrm{VaR}_{1/2}, \mathrm{MD})$: The mean-median deviation,

$$\mathrm{MD}(X) := \min_{x\in\mathbb{R}} \mathbb{E}[|X - x|], \quad X \in \mathcal{L}^1,$$

is a Bayes risk measure with loss function $L : (x, y) \mapsto |y - x|$, and the corresponding Bayes estimator is the median (interval) $\mathrm{VaR}_{1/2}(X)$. The mean-median deviation is a signed Choquet integral with distortion function $h(t) = \min\{t, 1 - t\}$, $t \in [0, 1]$; see Wang et al. (2020).

(vi) $(\mathrm{ex}_\alpha, \mathrm{var}_\alpha)$: The variantile (e.g., Wang and Wei (2020)),

$$\mathrm{var}_\alpha(X) := \min_{x\in\mathbb{R}} \left\{\alpha\, \mathbb{E}[(X - x)_+^2] + (1 - \alpha)\, \mathbb{E}[(X - x)_-^2]\right\}, \quad X \in \mathcal{L}^2,$$

where $\alpha \in (0, 1)$, is a Bayes risk measure with loss function

$$L : (x, y) \mapsto \alpha(y - x)_+^2 + (1 - \alpha)(y - x)_-^2.$$

The Bayes estimator corresponding to the variantile var_α is the *expectile* at the level α, denoted by ex_α; see Bellini et al. (2014) and Ziegel (2016).

3 Characterizing ES as a Bayes risk measure

We will present below our first main result on the characterization of Bayes risk measures. Recall that the ES/$\mathbb{E}$-mixtures in (9) of Example 1 are coherent and Choquet Bayes risk measures. Theorem 2 below further shows that they are the only possible class of Bayes risk measures which are either coherent or Choquet. This result is illustrated by the Venn diagram in Figure 1.

Below, lower semicontinuity is defined with respect to almost sure convergence. This form of lower semicontinuity is used to formulate the *prudence axiom* of Wang and Zitikis (2021), and the interpretation is that a consistent statistical approximation of the true risk should not underestimate the risk measure.

Theorem 2. *Suppose that $\mathcal{L}^\infty \subset \mathcal{X} \subset \mathcal{L}^1$. For a risk measure $\mathcal{R} : \mathcal{X} \to \mathbb{R}$, the following are equivalent:*

(i) $\mathcal{R}$ is a coherent Bayes risk measure;

(ii) $\mathcal{R}$ is a Choquet Bayes risk measure;

(iii) $\mathcal{R} = \mathrm{ES}_\alpha^\lambda$ for some $\alpha \in (0,1)$ and $\lambda \in [0,1]$.

If $\mathcal{R}$ further satisfies lower semicontinuity, then $\mathcal{R} = \mathrm{ES}_\alpha$ for some $\alpha \in (0,1)$.

Proof. The full proof is presented in Appendix A.1, and we give a sketch of the main steps here. The implication (iii)⇒(ii) is obvious. The implication (ii)⇒(i) is implied by Proposition 3 in Section 5. Below are the main steps for the most important implication (i)⇒(iii). Assume $(\mathcal{S}, \mathcal{R})$ is a Bayes pair in which $\mathcal{S}$ is translation invariant and $\mathcal{R}$ is coherent.

We first show in Lemma 1, that, using the fact that $\mathcal{S}$ and $\mathcal{R}$ are both translation invariant, we can choose a loss function for $(\mathcal{S}, \mathcal{R})$ in the form $(x, y) \mapsto x + v(y - x)$ for some real function v. Thus, we have

$$\mathcal{R}(X) = \inf_{c \in \mathbb{R}} \{c + \mathbb{E}[v(X - c)]\}, \quad X \in \mathcal{X}. \tag{11}$$

Using the monotonicity of $\mathcal{R}$, we proceed to show in Lemma 2 that such v can be replaced by an increasing function $\tilde{v}$ without changing $\mathcal{R}$. Next, using the convexity of $\mathcal{R}$, we show in Lemma 3 that $\tilde{v}$ can be replaced by an increasing convex function $\hat{v}$. Using the positive homogeneity of $\mathcal{R}$, in Lemma 5 we show that $\hat{v}$ can be replaced by the piece-wise linear function $\bar{v}(x) = \lambda x + (\gamma - \lambda)x_+$ for some $\gamma \geq 1$ and $\lambda \in [0,1]$. Finally, with the above loss function, we derive $\mathcal{R} = \mathrm{ES}_\alpha^{1-\lambda}$ where $\alpha = (\gamma - 1)/(\gamma - \lambda) \in (0,1)$.

For the last statement of the theorem, the lower semicontinuity of $\mathrm{ES}_\alpha^{1-\lambda}$ implies $\lambda = 0$ since ES_α is lower semicontinuous and $\mathbb{E}$ is not, as implied by Theorem 1 of Wang and Zitikis (2021). ☐

Remark 1. As we see in the proof of Theorem 2, a key step is to show that a translation-invariant Bayes risk measure $\mathcal{R}$ has the form (11) in Lemma 1. Risk measures directly defined via the form (11) have appeared in the literature, and we make two notable connections.

1. The *optimized certainty equivalent* (OCE) of Ben-Tal and Teboulle (2007) has the form (11) where v is increasing, convex, and satisfying $v(0) = 0$ and $v'(0+) \geq 1$; here it is adapted to our convention that a positive value of X represents a loss. Theorem 3.1 of Ben-Tal and Teboulle (2007), which is closely related to Theorem 2, states that, assuming that v is real-valued, increasing, convex, $v(0) = 0$, $v'(0+) > 0$, and $v(x) > x$ for all $x \neq 0$, the only coherent risk

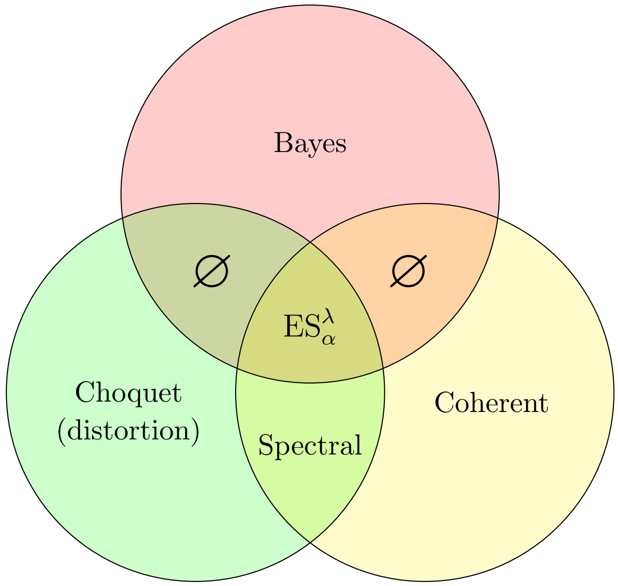

Figure 1: A Venn diagram for three classes of law-invariant risk measures

measure in the OCE class is generated by $v(x) = \lambda x + (\gamma - \lambda)x_+$ for some $\infty > \gamma > 1 > \lambda \geq 0$, which is an ES/$\mathbb{E}$-mixture, similar to Lemma 5 except for the boundary cases of $\gamma = \infty$, $\gamma = 1$ and $\lambda = 1$. Different from Ben-Tal and Teboulle (2007), all our assumptions are made on $\mathcal{R}$ and not on the form of v in (11).

2. The form (11) also appears in the expectation quadrangle of Rockafellar and Uryasev (2013), where $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ also serves as an important example. Our choice of notation, especially $\mathcal{S}$ and $\mathcal{R}$, is consistent with the notation of Rockafellar and Uryasev (2013). Nevertheless, our interpretation of the Bayes pair and our focus on characterization are different from their framework. See also the recent paper Chong et al. (2021) where (11) appears as an optimized objective in the context of capital allocation.

Remark 2. Using Lemmas 2 and 3 in Appendix A.1, we also obtain the forms of monetary and convex Bayes risk measures. A risk measure $\mathcal{R}$ is a monetary (resp. monetary and convex) Bayes risk measure if and only if (11) holds for some increasing (resp. increasing convex) function $v : \mathbb{R} \to \mathbb{R}$.

Below we further present two examples of a coherent risk measure $\mathcal{R}$ satisfying (5), but the corresponding minimizer $\mathcal{S}$ is not translation invariant (indeed, not interpretable). By Theorem 1, these risk measures have the form (6), for a set of loss functions $\mathcal{A}$,

$$\mathcal{R}(X) = \min_{\ell \in \mathcal{A}} \mathbb{E}[\ell(X)], \quad X \in \mathcal{X}.$$

The examples also show that the assumption of translation invariance on $\mathcal{S}$ in Definition 2 is essential for the characterization of the Bayes risk measures in Theorem 2.

Example 2. The first example is a convex combination of ES at different levels. Define a risk measure $\mathcal{R} = \frac{1}{2}\mathrm{ES}_\alpha + \frac{1}{2}\mathrm{ES}_\beta$ for some distinct numbers $\alpha, \beta \in (0,1)$. Clearly, $\mathcal{R}$ is a coherent risk measure. By (4), for $X \in \mathcal{L}^1$,

$$\begin{aligned}\mathcal{R}(X) &= \frac{1}{2}\min_{x\in\mathbb{R}}\left\{x + \frac{1}{1-\alpha}\mathbb{E}[(X-x)_+]\right\} + \frac{1}{2}\min_{x\in\mathbb{R}}\left\{x + \frac{1}{1-\beta}\mathbb{E}[(X-x)_+]\right\}\\ &= \frac{1}{2}\min_{x_1,x_2\in\mathbb{R}}\left\{x_1 + x_2 + \frac{1}{1-\alpha}\mathbb{E}[(X-x_1)_+] + \frac{1}{1-\beta}\mathbb{E}[(X-x_2)_+]\right\}.\end{aligned}$$

By Theorem 1, $\mathcal{R}$ satisfies (5) for some $\mathcal{S}$. Since $\mathcal{R}$ is not an ES/$\mathbb{E}$-mixture, by Theorem 2, $\mathcal{R}$ is not a Bayes risk measure. This implies that any minimizer $\mathcal{S}$ satisfying (5) is not translation invariant. We can also see from this example that $\mathcal{S}(X)$ should be a one-to-one function of the minimizer (x_1, x_2) above, which is difficult to interpret in a financial context (one-to-one mappings from $\mathbb{R}^2$ to $\mathbb{R}$ are usually quite strange).

Example 3. The second example is the coherent entropic risk measure introduced by Föllmer and Knispel (2011), defined, for some $c > 0$, as

$$\mathcal{R}(X) = \min_{\gamma>0}\left\{\frac{1}{\gamma}\log\mathbb{E}[e^{\gamma X}] + \frac{c}{\gamma}\right\}, \qquad X \in \mathcal{L}^\infty.$$

Föllmer and Knispel (2011) showed that $\mathcal{R}$ is a coherent risk measure; it satisfies (5) by Theorem 1. Since $\mathcal{R}$ is not an ES/$\mathbb{E}$-mixture, by Theorem 2, $\mathcal{R}$ is not a Bayes risk measure.

Before ending this section, we show that the Bayes pair $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha^\lambda)$ can be characterized if $\mathcal{R}$ is coherent or Choquet. This result slightly generalizes Theorem 2 which only gives the form of $\mathcal{R}$ but not that of $\mathcal{S}$. A proof of Proposition 1 is put in Appendix A.2.

Proposition 1. *For a Bayes pair $(\mathcal{S}, \mathcal{R})$ with loss function L, the following are equivalent:*

(i) $\mathcal{S}(0) = 0$ *and* $\mathcal{R}$ *is a coherent Bayes risk measure;*

(ii) $\mathcal{S}(0) = 0$ *and* $\mathcal{R}$ *is a Choquet Bayes risk measure;*

(iii) $(\mathcal{S}, \mathcal{R}) = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha^\lambda)$ *for some* $\alpha \in (0,1)$ *and* $\lambda \in [0,1)$*;*

(iv) the loss function can be chosen as $L : (x,y) \mapsto x + (1-\lambda)(y-x) + \frac{\lambda}{1-\alpha}(y-x)_+$ *for some* $\alpha \in (0,1)$ *and* $\lambda \in (0,1]$.

For a given $\alpha \in (0,1)$, Wang and Wei (2020, Theorem 6.9) showed that an ES/$\mathbb{E}$-mixture is the only coherent Choquet risk measure ρ such that $(\rho, \mathrm{VaR}_\alpha)$ is elicitable. This result does not imply, and is not implied by, Theorem 2 and Proposition 1, although the similarity is visible.

4 Elicitability of Bayes risk measures

In this section we study the connection between Bayes pairs and elicitability. Recall that the functional $\mathcal{S}$ is elicitable if there exists a loss function $L : \mathbb{R}^{d+1} \to \mathbb{R}$ such that

$$\mathcal{S}(X) = \underset{y \in \mathbb{R}^d}{\arg\min}\, \mathbb{E}[L(y, X)], \quad X \in \mathcal{X}. \tag{12}$$

The first observation is that a Bayes pair $(\mathcal{S}, \mathcal{R})$ is always elicitable. This was essentially shown in Theorem 1 of Frongillo and Kash (2021) where $\mathcal{S}$ takes scalar values. We present a similar proof which is adapted to our slightly different definitions.

Theorem 3. *Any Bayes pair $(\mathcal{S}, \mathcal{R})$ with loss function L is elicitable by*

$$L^*(x, y, z) = \int_0^y h(t)\,\mathrm{d}t + h(y)(L(x, z) - y), \quad (x, y) \in D,\ z \in \mathbb{R},$$

where h is any positive and strictly decreasing function on $\mathbb{R}$ and D is the range of $(\mathcal{S}, \mathcal{R})$.

Proof. We need to show that L^* elicits $(\mathcal{S}, \mathcal{R})$; that is, for $X \in \mathcal{X}$,

$$(\mathcal{S}, \mathcal{R})(X) = \underset{(x,y) \in D}{\arg\min}\, \mathbb{E}[L^*(x, y, X)] = \underset{(x,y) \in D}{\arg\min} \left\{ \int_0^y h(t)\,\mathrm{d}t + h(y)(\mathbb{E}[L(x, X)] - y) \right\}. \tag{13}$$

First, for a fixed (y, X), the minimizers x^* to (13) are the same as the minimizers of $\mathbb{E}[L(x, X)]$. Therefore, we know that the set of minimizers x^* are precisely $\mathcal{S}(X)$, and $\mathbb{E}[L(x^*, X)] = \mathcal{R}(X)$. Next, we need to find the minimizers for

$$\underset{y \in \mathbb{R}}{\arg\min} \left\{ \int_0^y h(t)\,\mathrm{d}t + h(y)(\mathcal{R}(X) - y) \right\},$$

which gives $y^* = \mathcal{R}(X)$ since h is a strictly decreasing function. □

Remark 3. In Theorem 3, the loss function which elicits $(\mathcal{S}, \mathcal{R})$ is not unique. For instance, if $\mathcal{S}(X)$ is itself elicited by a loss function L', then $(x, y, z) \mapsto L^*(x, y, z) + L'(x, z)$ also elicits $(\mathcal{S}, \mathcal{R})$.

Following the terminology of Emmer et al. (2015) and Fissler and Ziegel (2016), a functional $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is *second-order elicitable* if it is a component of a 2-dimensional elicitable functional, and it is *conditionally elicitable* on another functional $\mathcal{S}$ if for each $r \in \mathbb{R}$ and some loss function L_r, we have

$$\mathcal{R}(X) = \underset{y \in \mathbb{R}^d}{\arg\min}\, \mathbb{E}[L_r(y, X)], \quad X \in \mathcal{X} \text{ with } \mathcal{S}(X) = r.$$

Theorem 3 immediately yields that any Bayes risk measure $\mathcal{R}$ is second-order elicitable as a component of the Bayes pair $(\mathcal{S}, \mathcal{R})$ and conditionally elicitable on $\mathcal{S}$ via $L_r : (y, z) \mapsto L^*(r, y, z)$.

Corollary 1. *Any Bayes risk measure is second-order elicitable and conditionally elicitable.*

Remark 4. Another direct consequence of Theorem 3 is that any Bayes risk measure has elicitation complexity of at most 2 (implied by second-order elicitability), and hence they are relatively simple to estimate via empirical risk minimization; see Frongillo and Kash (2021) for a precise definition and related discussions.

5 Other properties of Bayes risk measures

The following two results of a Bayes risk measure do not require monotonicity. Let $\mathcal{M}$ be the set of distributions of the elements in $\mathcal{X}$. For any scalar law-invariant risk measure $\mathcal{R}$, we write $\widehat{\mathcal{R}} : F \mapsto \mathcal{R}(X)$ where $X \sim F \in \mathcal{M}$. Thus, $\widehat{\mathcal{R}}$ represents the risk measure $\mathcal{R}$ treated as a mapping from $\mathcal{M}$ to $\mathbb{R}$. We say that a scalar risk measure $\mathcal{R}$ is *mixture concave* if $\widehat{\mathcal{R}}$ is concave in $F \in \mathcal{M}$; $\mathcal{R}$ has *convex level sets (CxLS)* if the set $\{F \in \mathcal{M} : \widehat{\mathcal{R}}(F) = r\}$ is convex for each $r \in \mathbb{R}$. Mixture concavity represents that using a mixture of models (i.e., introducing a stochastic factor) increases randomness, and it is a desirable property for both risk and deviation measures. Moreover, for Choquet risk measures, mixture concavity is equivalent to coherence (Theorem 3 of Wang et al. (2020)). The CxLS property is a necessary condition for elicitability (Osband (1985)) and has been widely studied in the risk measure literature (e.g., Weber (2006), Ziegel (2016), Delbaen et al. (2016) and Wang and Wei (2020)). The following two properties are useful in the proofs of Theorems 2 and 4. Moreover, Proposition 2 directly inspires the study in Section 6.

Proposition 2. *A Bayes risk measure is necessarily mixture concave, and a Bayes estimator necessarily has CxLS.*

Proof. By definition, $\widehat{\mathcal{R}} : F \mapsto \inf_{x \in \mathbb{R}}\{\int_{\mathbb{R}} L(x, y)\, \mathrm{d}F(y)\}$ is the infimum of linear functions on $\mathcal{M}$, and hence concave. Thus, $\mathcal{R}$ is mixture concave. The second statement is due to the fact that any Bayes estimator is elicitable, and it is well known that elicitable functionals have CxLS (e.g., Ziegel (2016)). $\square$

Proposition 3. *A Choquet Bayes risk measure is necessarily coherent.*

Proof. Note that a monetary risk measure is uniformly continuous with respect to the L^∞-norm. Using Theorem 1 of Wang et al. (2020), a law-invariant, uniformly L^∞-continuous and comonotonic-additive functional admits a representation as a Choquet integral. Theorem 3 of Wang et al. (2020) further implies that mixture concavity is equivalent to convexity. Therefore, as Choquet risk measures are automatically positively homogeneous, $\mathcal{R}$ is coherent. $\square$

6 Elicitable Bayes risk measures

Any Bayes risk measure is mixture concave (Proposition 2), and any elicitable risk measure, such as the Bayes estimator, has CxLS. We wonder what is the intersection of the two classes of risk measures. This question is not only driven by mathematical curiosity, but also has interesting connections with some classical results in decision theory.

As we have seen above, the mean is both elicitable (with loss function $L(x,y) = (y-x)^2$) and Bayes (with loss function $L(x,y) = x + (y-x)_+$). Moreover, the entropic risk measure ER in Example 1 is mixture concave and has CxLS, since it is both elicitable and Bayes. The next result shows that ER is the only risk measure that is mixture concave and has CxLS under the following continuity assumption (recall that $\widehat{\mathcal{R}}(F) = \mathcal{R}(X)$ where $X \sim F$)

(C) For any $x < y$, the mapping $\alpha \mapsto \widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_y)$ on $[0,1]$ is continuous at $\alpha = 0$, where δ_z is the point-mass at $z \in \mathbb{R}$.

Clearly, continuity (C) is weaker than continuity from above.

Theorem 4. *Let $\mathcal{R}$ be a law-invariant monetary risk measure on $\mathcal{X} = \mathcal{L}^\infty$ satisfying continuity (C) with $\mathcal{R}(0) = 0$. Then $\mathcal{R}$ is mixture concave and has CxLS if and only if it is an entropic risk measure.*

The proof of Theorem 4 is technical and put in Appendix A.3. Below we illustrate some intuition of this result by connecting mixture concavity and CxLS to the notions of betweenness (Chew (1983)) and associativity (Grant et al. (2000)) in decision theory.[11] We say that $\mathcal{R}$ satisfies *associativity* if for any $F, G, H \in \mathcal{M}$ and $\lambda \in (0,1)$,

$$\widehat{\mathcal{R}}(F) = \widehat{\mathcal{R}}(G) \implies \widehat{\mathcal{R}}(\lambda F + (1-\lambda)H) = \widehat{\mathcal{R}}(\lambda G + (1-\lambda)H). \tag{14}$$

Lemma 2 of Grant et al. (2000) shows that associativity holds under the assumption of a suitable continuity condition, mixture concavity and betweenness. The betweenness property is slightly

[11] We thank an anonymous referee for brining up this connection.

stronger than the CxLS property, but they are equivalent under some mild assumption (e.g., Lemma 14 of Steinwart et al. (2014)). If $\widehat{\mathcal{R}}$ satisfies associativity, then by the de Finetti-Kolmogorov-Nagumo Theorem (see e.g., Cifarelli and Regazzini (1996)), $\mathcal{R}$ is a certainty equivalent, that is, there exists a continuous and strictly increasing function $u : \mathbb{R} \to \mathbb{R}$ such that $\mathcal{R}(X) = u^{-1}(\mathbb{E}[u(X)])$, $X \in \mathcal{X}$. Finally, by translation-invariance of $\mathcal{R}$, one can conclude that $u(x) = e^{cx}$ for $c > 0$ or $u(x) = x$, $x \in \mathbb{R}$. As a consequence, $\mathcal{R}$ must be an entropic risk measure. The main gap in the above informal argument is to verify the conclusion of Lemma 2 of Grant et al. (2000) under CxLS and (C), which is a complicated mathematical task although intuitively clear. In Appendix A.3, we provide a full proof without using the results of Grant et al. (2000).

Remark 5. Continuity (C) is not satisfied by the essential supremum $X \mapsto \mathrm{VaR}_1^-(X)$, which is mixture concave and has CxLS. In our proof of Theorem 4, the continuity condition (C) is essential and we were not able to relax it. Nevertheless, we conjecture that by including $\mathrm{VaR}_1^- = \mathrm{ER}_\infty$ as an extended member of the ER family, one may remove or weaken (C) in Theorem 4.

A consequence of Theorem 4 is that a risk measure $\mathcal{R}$ with the form

$$\mathcal{R}(X) = \inf\{x \in \mathbb{R} : \mathbb{E}[g(X - x)] \le z\} \tag{15}$$

for a strictly increasing g cannot be mixture concave unless it is an entropic risk measure. In particular, this implies that expectiles defined in Example 1 (vi) are not mixture concave. This fact is shown by Bellini et al. (2018), and it is (surprisingly) not easy to directly verify.

Corollary 2. *Let $\mathcal{R}$ be defined by (15) for some increasing function g and constant z satisfying $g(-t) < z < g(t)$ for all $t > 0$. Then $\mathcal{R}$ is mixture concave if and only if it is an entropic risk measure.*

Proof. It is clear that the risk measure $\mathcal{R}$ defined by (15) is monetary. Note that the condition $g(-t) < z < g(t)$ for all $t > 0$ implies $\mathcal{R}(0) = 0$. Lemma 7 guarantees (C) from the above condition on g. The rest follows by applying Theorem 4. □

Finally, we obtain a characterization of entropic risk measures as the intersection of Bayes estimators and Bayes risk measures. Moreover, as we see in Example 1 (ii), an entropic risk measure is a Bayes estimator and a Bayes risk measure with the same loss function.

Corollary 3. *A monetary risk measure $\mathcal{R}$ with $\mathcal{R}(0) = 0$ is elicitable and Bayes if and only if it is an entropic risk measure.*

Proof. Note that an elicitable risk measure satisfies CxLS and a Bayes risk measure satisfies mixture concavity. Using Theorem 4, it suffices to verify that a Bayes risk measure $\mathcal{R}(X) = \min_x \mathbb{E}[L(x, X)]$ satisfies continuity (C). That is, for any $x < y$, the function

$$\alpha \mapsto H(\alpha) := \min_{s\in\mathbb{R}}\{(1-\alpha)L(s,x) + \alpha L(s,y)\}$$

is continuous at $\alpha = 0$. Note that there exists s_0 such that $L(s_0, x) = x$. Then

$$x \le \liminf_{\alpha\downarrow 0} H(\alpha) \le \limsup_{\alpha\downarrow 0} H(\alpha) \le \lim_{\alpha\downarrow 0}\{(1-\alpha)L(s_0,x) + \alpha L(s_0,y)\} = L(s_0,x) = x.$$

Hence, we have $H(\alpha)$ is continuous at $\alpha = 0$, which gives the desired condition (C). $\square$

7 Concluding remarks

In this paper, we introduce the concepts of Bayes pairs and Bayes risk measures, and offer some characterization results. In particular, Theorem 2 yields a new characterization of ES in the context of statistical inference and optimization, complementing the ES characterization of Wang and Zitikis (2021) based on portfolio risk aggregation.

It is known that entropic risk measures are the only dynamically consistent law-invariant risk measures (Kupper and Schachermayer (2009)), and they are also the only intersection of the class of optimized certainty equivalents (OCE) and the class of shortfall risk measures (Ben-Tal and Teboulle (2007) and Föllmer and Schied (2016)). Theorem 4 further shows that, under a continuity assumption, the entropic risk measures are the only monetary risk measures satisfying mixture concavity (a property of the OCE) and CxLS (a property of the shortfall risk measures).

Bayes risk measures are closely related to elicitability, and they are second-order elicitable (Theorem 3). There are several open questions on the theory of Bayes pairs and Bayes risk measures which will be explored in the future; we discuss a few of them here.

The first question is regarding the special role of Bayes pairs among elicitable two-dimensional functionals. Almost all examples of elicitable two-dimensional functionals $(\mathcal{S}, \mathcal{R})$ in the literature are one-to-one transforms of either a Bayes pair, such as those in Example 1, or a pair whose components are both elicitable, such as $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta)$, $(\mathrm{ex}_\alpha, \mathrm{ex}_\beta)$, or the modal interval (see Brehmer and Gneiting (2021)). We wonder under what conditions an elicitable two-dimensional functional has to be obtained from a Bayes pair.

Next, we focus on the Bayes risk measure $\mathcal{R}$. We say that a risk measure is genuinely 2-

elicitable if it is second-order elicitable but not elicitable. Since coherent risk measures are not elicitable except for the expectiles (Ziegel (2016)), it is natural to study the class of genuinely 2-elicitable coherent risk measures. A non-elicitable Bayes risk measure is genuinely 2-elicitable (see Corollary 1), but the converse is not true; it is unclear what special role Bayes risk measures play among genuinely 2-elicitable risk measures.

There are at least two very different ways to construct a genuinely 2-elicitable coherent risk measure. The first is to combine two elicitable risk measures, such as a mixture of two different expectiles, $(1-\lambda)\mathrm{ex}_\alpha + \lambda \mathrm{ex}_\beta$, and the second is to use a Bayes risk measure, such as an ES/$\mathbb{E}$-mixture. We conjecture that a coherent Choquet risk measure is genuinely 2-elicitable if and only if it is an ES/$\mathbb{E}$-mixture (except for the mean). We also wonder under what conditions, a genuinely 2-elicitable coherent risk measure has to be a mixture of two expectiles.

Finally, if we replace translation invariance of $\mathcal{S}$ in Definition 2 with another property, the characterization in Theorem 2 may fail to hold, as we see in Examples 2 and 3. A full characterization of coherent $\mathcal{R}$ without translation invariance of $\mathcal{S}$ is open at the moment. A similar question arises in a setting where $\mathcal{S}$ is allowed to be multi-dimensional; these questions are planned for future research.

Acknowledgements

The authors thank the Editors Thaleia Zariphopoulou and Jan Obloj, two anonymous referees, and Jonas Brehmer for helpful comments on an earlier version of the paper. TM would like to thank the financial support from the National Science Foundation of China (Nos. 71871208, 71671176, 71921001). RW acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC RGPIN-2018-03823, RGPAS-2018-522590).

A Lemmas and proofs of several results

A.1 Lemmas in the proof of Theorem 2

Below we derive a few lemmas which lead to the proof of Theorem 2, implication (i)$\Rightarrow$(iii). The other implications are already shown in the proof sketch in Section 3. In all lemmas below, $\mathcal{X}$ is a linear space satisfying $\mathcal{L}^\infty \subset \mathcal{X} \subset \mathcal{L}^1$.

Lemma 1. *Suppose that $(\mathcal{S},\mathcal{R})$ is a Bayes pair with loss function L, and $\mathcal{S}$ and $\mathcal{R}$ are translation invariant. The function $(x,y) \mapsto x + L(0, y-x)$ is also a loss function for $(\mathcal{S},\mathcal{R})$. As consequence,* *there exists a real function v such that*

$$\mathcal{R}(X) = \min_{x\in\mathbb{R}}\{x + \mathbb{E}[v(X-x)]\},\ X \in \mathcal{X}.$$

Proof. Define $L^*(x,y) = x + L(0, y-x),\ (x,y) \in \mathbb{R}^2$, and

$$\mathcal{R}^*(X) = \min_{x\in\mathbb{R}} \mathbb{E}[L^*(x,X)], \quad X \in \mathcal{X}.$$

We aim to show $\mathcal{R}^* = \mathcal{R}$. For a random variable $X \in \mathcal{X}$, denote by $\mathcal{S}^*$ the left end-point of $\mathcal{S}$, that is, $\mathcal{S}^*(X) = \min\{\arg\min_{x\in\mathbb{R}} \mathbb{E}[L(x,X)]\}$, $X \in \mathcal{X}$. By translation invariance of $\mathcal{S}$, we have $\mathcal{S}^*(X+c) = \mathcal{S}^*(X)+c$ for $c \in \mathbb{R}$. Then by translation invariance of $\mathcal{R}$ we have $\mathcal{R}(X) = c+\mathcal{R}(X-c)$, that is,

$$\min_{x\in\mathbb{R}} \mathbb{E}[L(x,X)] = c + \min_{x\in\mathbb{R}} \mathbb{E}[L(x, X-c)].$$

As a consequence,

$$\mathbb{E}[L(\mathcal{S}^*(X),X)] = c + \mathbb{E}[L(\mathcal{S}^*(X-c), X-c)] = c + \mathbb{E}[L(\mathcal{S}^*(X)-c, X-c)], \tag{16}$$

where the last equality follows from $\mathcal{S}^*(X+c) = \mathcal{S}^*(X)+c$. By setting $x^* = \mathcal{S}^*(X)$, we have

$$\begin{aligned}
\mathcal{R}^*(X) &= \min_{x\in\mathbb{R}} \mathbb{E}[L^*(x,X)] = \min_{x\in\mathbb{R}}\{x + \mathbb{E}[L(0, X-x)]\} \\
&\le x^* + \mathbb{E}[L(0, X-x^*)] = x^* + \mathbb{E}[L(\mathcal{S}^*(X)-x^*, X-x^*)] \\
&= x^* + \mathbb{E}[L(\mathcal{S}^*(X-x^*), X-x^*)] = x^* + \min_{x\in\mathbb{R}} \mathbb{E}[L(x, X-x^*)] \\
&= x^* + \mathcal{R}(X-x^*) = \mathcal{R}(X),
\end{aligned}$$

where the last equality is due to the translation invariance of $\mathcal{R}$. Thus we have $\mathcal{R}^* \le \mathcal{R}$. In order to show $\mathcal{R}^* \ge \mathcal{R}$, take $y^* \in \arg\min_{x\in\mathbb{R}}\{x + \mathbb{E}[L(0, X-x)]\}$. By (16), we have

$$\begin{aligned}
\mathcal{R}^*(X) &= y^* + \mathbb{E}[L(0, X-y^*)] \ge y^* + \min_{x\in\mathbb{R}} \mathbb{E}[L(x, X-y^*)] \\
&= y^* + \mathbb{E}[L(\mathcal{S}^*(X-y^*), X-y^*)] = \mathbb{E}[L(\mathcal{S}^*(X),X)] = \mathcal{R}(X).
\end{aligned}$$

Hence, we have $\mathcal{R} = \mathcal{R}^*$. Taking $v(y) = L(0,y)$ gives the last statement. $\square$

Using Lemma 1, we can write

$$\mathcal{R}(X) = \min_{x\in\mathbb{R}}\{x + \mathbb{E}[v(X-x)]\}, \quad \text{where} \;\; v(y) = L(0,y). \tag{17}$$

In the following lemmas, we allow v to take the value ∞, and obtain results that are slightly more general than required, i.e., we will also include ES_1 which is the essential supremum. We define the increasing version of v as

$$\tilde{v}(x) = \inf_{y\geq x} v(y), \quad x \in \mathbb{R}.$$

Note that $\mathcal{R}$ in (17) is real-valued, and

$$\mathcal{R}(x) = \inf_{c\in\mathbb{R}}\{c + v(x-c)\} \leq \inf_{y\geq x} v(y).$$

Hence, $\tilde{v}(x) > -\infty$ for all $x \in \mathbb{R}$. The finiteness of $\mathcal{R}$ also implies that v is not always ∞ on $\mathbb{R}$.

Lemma 2. *Suppose that $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ in (17) is monotone. Then*

$$\mathcal{R}(X) = \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[\tilde{v}(X-c)]\}, \quad X \in \mathcal{X}.$$

Proof. Let us denote by $\widetilde{\mathcal{R}}(X) = \inf_{c\in\mathbb{R}}\{c+\mathbb{E}[\tilde{v}(X-c)]\}$, $X \in \mathcal{X}$. Obviously, $\mathcal{R}(X) \geq \widetilde{\mathcal{R}}(X)$, $X \in \mathcal{X}$. Below we show $\mathcal{R} \leq \widetilde{\mathcal{R}}$. Take $\epsilon > 0$ and $c \in \mathbb{R}$. By definition of $\tilde{v}$, for any $x \in \mathbb{R}$, there exists $y \geq x$ such that $v(y) \leq \tilde{v}(x) + \epsilon$, and such y admits an increasing (hence measurable) selection. As a consequence, there exists $Y \in \mathcal{X}$ such that $Y \geq X$ and $v(Y-c) \leq \tilde{v}(X-c) + \epsilon$. This implies $c + \mathbb{E}[v(Y-c)] \leq c + \mathbb{E}[\tilde{v}(X-c)] + \epsilon$. By monotonicity of $\mathcal{R}$ and $Y \geq X$, we further have

$$\mathcal{R}(X) \leq \mathcal{R}(Y) \leq c + \mathbb{E}[v(Y-c)] \leq c + \mathbb{E}[\tilde{v}(X-c)] + \epsilon.$$

Taking an infimum of the above inequality over $c \in \mathbb{R}$ and $\epsilon > 0$ yields $\mathcal{R}(X) \leq \widetilde{\mathcal{R}}(X)$. □

Next, for an increasing function v, we define the largest convex function dominated by v as

$$\hat{v}(x) = \sup\{g(x) : g \leq v \text{ on } \mathbb{R}, \; g \text{ is convex}\}, \quad x \in \mathbb{R}.$$

By definition, $\hat{v}$ is convex. To state the following lemma, we define

$$\mathcal{U} = \big\{v : \; v \text{ is increasing and convex}, 1 \in \operatorname{int} \partial v(\mathbb{R})\big\}, \tag{18}$$

where $\partial v(\mathbb{R}) = \mathrm{cx}\{v'_-(x), v'_+(x), \; x \in \mathbb{R}\}$, $\operatorname{int} A$ is the interior of a set A, and $\mathrm{cx}(A)$ is the convex hull of A. Here we define the right derivative $v'_+(x) = \infty$ if $v(y) = \infty$ for any $y > x$.

Lemma 3. *Suppose that $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ in (17) is monetary and convex, and v is an increasing function. Then*

$$\mathcal{R}(X) = \inf_{c\in\mathbb{R}} \{c + \mathbb{E}[\hat{v}(X - c)]\}, \quad X \in \mathcal{X}.$$

Specifically, if $\hat{v} \notin \mathcal{U}$, then either $\mathcal{R}(X) \equiv -\infty$ or $\mathcal{R}(X) = \mathbb{E}[X] - \hat{v}^*(1)$, where $\hat{v}^*(x) = \sup_y\{xy - \hat{v}(y)\}$ is the conjugate function of $\hat{v}$.

Proof. Let us denote by

$$\mathcal{R}'(X) = \inf_{c\in\mathbb{R}} \{c + \mathbb{E}[\hat{v}(X - c)]\}, \quad X \in \mathcal{X}.$$

Obviously, $\mathcal{R} \geq \mathcal{R}'$. Below we show $\mathcal{R} \leq \mathcal{R}'$. Take $\epsilon > 0$ and $c \in \mathbb{R}$. Note that a law-invariant convex risk measure is monotonic with respect to convex order (e.g., Proposition 3.2 of Mao and Wang (2020)). For $X \in \mathcal{X}$, we assert that there exists $Y \in \mathcal{X}$ be such that

$$X \prec_{\mathrm{cx}} Y \quad \text{and} \quad \mathbb{E}[v(Y - c)] \leq \mathbb{E}[\hat{v}(X - c)] + \epsilon. \tag{19}$$

To show this assertion, we use Theorem 4.1 of Mao et al. (2018), which gives

$$\mathbb{E}[\hat{v}(X)] = \lim_{n\to\infty} \frac{1}{n} \inf \{\mathbb{E}[v(X_1)] + \cdots + \mathbb{E}[v(X_n)] : X_1 + \cdots + X_n = nX\}.$$

Then for any $\epsilon > 0$, there exist $n \in \mathbb{N}$ and $X_1, \ldots, X_n$ such that $X_1 + \cdots + X_n = nX$ and

$$\frac{1}{n}\big(\mathbb{E}[v(X_1)] + \cdots + \mathbb{E}[v(X_n)]\big) \leq \mathbb{E}[\hat{v}(X)] + \epsilon.$$

Denote by F_i the distribution of X_i, $i = 1, \ldots, n$ and take a random variable Y such that its distribution is $H = \sum_{i=1}^n F_i/n$. We then have

$$\mathbb{E}[v(Y)] = \int_{\mathbb{R}} v(y)\,\mathrm{d}H(y) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[v(X_i)] \leq \mathbb{E}[\hat{v}(X)] + \epsilon.$$

For any convex function ℓ, we have $\mathbb{E}[\ell(Y)] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\ell(X_i)] \geq \mathbb{E}[\ell(X)]$, where the inequality follows convexity. This implies $X \prec_{\mathrm{cx}} Y$, and hence (19) holds. This implies $c + \mathbb{E}[v(Y - c)] \leq c + \mathbb{E}[\hat{v}(X - c)] + \epsilon$. By monotonicity of $\mathcal{R}$ with respect to convex order and $X \prec_{\mathrm{cx}} Y$, we further have

$$\mathcal{R}(X) \leq \mathcal{R}(Y) \leq c + \mathbb{E}[v(Y - c)] \leq c + \mathbb{E}[\hat{v}(X - c)] + \epsilon.$$

Taking an infimum of the above inequality over $c \in \mathbb{R}$ and $\epsilon > 0$ yields $\mathcal{R}(X) \leq \mathcal{R}'(X)$. Therefore,

$\mathcal{R} = \mathcal{R}'$. The last statement is shown in the discussion below Definition 3.1 of Wu et al. (2020). $\square$

By Lemma 3, in order to avoid the trivial cases of $\mathcal{R}$, we only need to consider $v \in \mathcal{U}$.

Lemma 4. *Let v be a function such that $v(x) \geq \overline{v}(x) := \lambda x + (\gamma - \lambda)x_+$, $x \in \mathbb{R}$, and*

$$\min_{x\in\mathbb{R}}\{x + \mathbb{E}[v(X - x)]\} = \min_{x\in\mathbb{R}}\{x + \mathbb{E}[\overline{v}(X - x)]\}, \quad X \in \mathcal{X}.$$

Then $v = \overline{v}$.

Proof. We show the result by contradiction. Suppose that $v(x_0) > \overline{v}(x_0)$ for some $x_0 \in \mathbb{R}$. Define a random variable X such that $1 - \mathbb{P}(X = 0) = \mathbb{P}(X = x_0) = p$, where $p \in (0, 1)$ satisfies $1 - p > \alpha := 1 - (1 - \lambda)/\gamma$. Then we have $\mathrm{VaR}_\alpha(X) = 0$ and thus

$$\mathcal{R}(X) = \min_x\{x + \mathbb{E}[\overline{v}(X - x)]\} = 0 + \mathbb{E}[\overline{v}(X - 0)] = \overline{v}(x_0).$$

Note that 0 is the unique minimizer of the above minimization problem, which implies that

$$x + \mathbb{E}[v(X - x)] \geq x + \mathbb{E}[\overline{v}(X - x)] > \mathcal{R}(X), \quad x \neq 0. \tag{20}$$

For $x = 0$, note that $\mathbb{E}[v(X)] = (1 - p)v(0) + pv(x_0) > p\overline{v}(x_0) = \mathcal{R}(X)$. This combined with (20) yields a contradiction to the fact that $\min_{x\in\mathbb{R}} x + \mathbb{E}[v(X - x)]$ can be attained. Hence, $v = \overline{v}$. $\square$

Theorem 3.1 of Ben-Tal and Teboulle (2007) showed that an OCE in (17) is positively homogeneous if and only if $v(x) = \lambda x + (\gamma - \lambda)x_+$ for some γ, λ. The following lemma gives a similar result under slightly different conditions. It states that $\mathcal{R}$ defined by (17) with $v \in \mathcal{U}$ is positively homogeneous if and only if v can be replaced by $\bar{v}$. For completeness, we give a self-contained proof that is different from Ben-Tal and Teboulle (2007).

Lemma 5. *For $v \in \mathcal{U}$, suppose that $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ in (17) is positively homogeneous and v is an increasing convex function. Then there exist $\gamma \in [1, \infty]$ and $\lambda \in [0, 1]$, such that*

$$\mathcal{R}(X) = \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[\bar{v}(X - c)]\}, \quad X \in \mathcal{X}, \tag{21}$$

where $\bar{v}(x) = \lambda x + (\gamma - \lambda)x_+$ for all $x \neq 0$.

Proof. Since $\mathcal{R}$ is positively homogeneous, we have $\mathcal{R}(0) = \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[v(-c)]\} = 0$. Note that one minimizer of the above infimum is $-\zeta_v$ given by $\zeta_v := \inf\{x : v'_-(x) \geq 1\} \in \mathbb{R}$, where v'_- is the left-derivative of v. One can easily verify that $1 \in [v'_-(\zeta_v), v'_+(\zeta_v)]$ and $v(\zeta_v) = \zeta_v$. Define

$u(x) := v(x+\zeta_v) - \zeta_v,\ x \in \mathbb{R}$. We have $1 \in [u'_-(0), u'_+(0)]$, and thus, $u \in \mathcal{U}$ with $u(0) = 0$. It is obvious that $\mathcal{R}(X) = \inf_{c\in\mathbb{R}}\{c+\mathbb{E}[u(X-c)]\},\ X \in \mathcal{X}$. Hence, without loss of generality, we assume that $v \in \mathcal{U}$ and $v(0) = 0$.

For $\lambda > 0$, denote by $v_\lambda(x) = v(\lambda x)/\lambda$ and $\bar{v}(x) = \inf_{\lambda>0} v_\lambda(x),\ x \in \mathbb{R}$. It is clear that $\bar{v}$ is convex, increasing and positively homogeneous. Since v is convex, we have, for $0 < \lambda \le \gamma$,

$$v(\lambda x) \le \frac{\lambda}{\gamma} v(\gamma x) + \left(1 - \frac{\lambda}{\gamma}\right) v(0) = \frac{\lambda}{\gamma} v(\gamma x).$$

As a consequence, $v_\lambda(x) \le v_\gamma(x)$. Thus we know that $v_\lambda(x)$ is increasing in λ. Hence, $\bar{v}(x) = \lim_{\lambda\downarrow 0} v_\lambda(x)$. Note that if $v(x) = \infty$ for any $x > 0$, then we have $\bar{v}(x) = \infty$ for $x > 0$; if however $v(x) < \infty$ for some $x > 0$, then for any $x \in \mathbb{R}$, there exists $\lambda > 0$ such that $v_\lambda(x) < \infty$. For $X \in \mathcal{X}$ with an upper bound and $c \in \mathbb{R}$, the Monotone Convergence Theorem gives

$$\mathbb{E}[\bar{v}(X-c)] = \mathbb{E}\left[\lim_{\lambda\downarrow 0} v_\lambda(X-c)\right] = \lim_{\lambda\downarrow 0} \mathbb{E}\left[v_\lambda(X-c)\right] = \inf_{\lambda>0} \mathbb{E}\left[v_\lambda(X-c)\right].$$

By definition, for $\lambda > 0$ and $X \in \mathcal{X}$,

$$\mathcal{R}(\lambda X) = \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[v(\lambda X - c)]\} = \inf_{c\in\mathbb{R}}\{\lambda c + \mathbb{E}[v(\lambda(X-c))]\} = \lambda \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[v_\lambda(X-c)]\}.$$

Hence, positive homogeneity of $\mathcal{R}$ implies

$$\mathcal{R}(X) = \frac{\mathcal{R}(\lambda X)}{\lambda} = \inf_{c\in\mathbb{R}}\{c + \mathbb{E}[v_\lambda(X-c)]\}.$$

Taking an infimum over $\lambda > 0$ yields that for any $X \in \mathcal{X}$ with an upper bound

$$\mathcal{R}(X) = \inf_{c\in\mathbb{R}}\left\{c + \inf_{\lambda>0} \mathbb{E}[v_\lambda(X-c)]\right\} = \inf_{c\in\mathbb{R}}\left\{c + \mathbb{E}[\bar{v}(X-c)]\right\},$$

thus showing that (21) holds for $X \in \mathcal{X}$ with an upper bound. By Lemma 4, we have $v = \bar{v}$, and thus, (21) holds for all $X \in \mathcal{X}$. Positive homogeneity and monotonicity of $\bar{v}$ imply that

$$\bar{v}(x) = \gamma x_+ - \lambda x_- = \lambda x + (\gamma - \lambda)x_+, \quad x \ne 0,$$

for some $\gamma \in [0,\infty]$ and $\lambda \in [0,\infty)$. Using Lemma 3, we further know that either $\mathcal{R}(X) = \mathbb{E}[X] + c$ for some constant c or $\bar{v} \in \mathcal{U}$. If $\mathcal{R}(X) = \mathbb{E}[X] + c$, then $c = 0$ due to positive homogeneity of $\mathcal{R}$. In this case, $\bar{v}$ can be chosen as $\bar{v}(x) = x$, corresponding to $\gamma = \lambda = 1$. If $\bar{v} \in \mathcal{U}$, then $\operatorname{int} \partial v(\mathbb{R}) = (\lambda, \gamma)$, which implies $\lambda < 1 < \gamma$. $\square$

Proof of Theorem 2. It remains to prove the last step in the implication (i)⇒(iii). Combining Lemmas 1-5, we know that (17) holds, and v can be chosen as $v(x) = \lambda x + (\gamma - \lambda)x_+$ for some $\gamma \in [1, \infty]$ and $\lambda \in [0, 1]$. If $\lambda < 1$, write $\alpha = (\gamma - 1)/(\gamma - \lambda) \in (0, 1]$. Using (4), including the case $\alpha = 1$, we have

$$\begin{aligned}\mathcal{R}(X) &= \inf_{c \in \mathbb{R}}\{c + \lambda\mathbb{E}[(X - c)] + (\gamma - \lambda)\mathbb{E}[(X - c)_+]\}\\ &= (1 - \lambda)\mathrm{ES}_\alpha(X) + \lambda\mathbb{E}[X].\end{aligned}$$

If $\lambda = 1$, then $\mathcal{R}(X) = \inf_{c \in \mathbb{R}}\{(\gamma - 1)\mathbb{E}[(X - c)_+]\} + \mathbb{E}[X] = \mathbb{E}[X]$. In either case, $\mathcal{R} = \mathrm{ES}_\alpha^{1-\lambda} = (1 - \lambda)\mathrm{ES}_\alpha + \lambda\mathbb{E}$. If, moreover, v is real-valued (it is in the definition of the loss function for a Bayes pair), then we have $\gamma < \infty$ and thus, $\alpha < 1$. $\square$

A.2 Proof of Proposition 1

Proof of Proposition 1. Note that the implications (iii)⇔(iv)⇒(ii) are obvious, and the implication (ii)⇒(i) is implied by Proposition 3 in Section 5. We next show the implication (i)⇒(iii). By Lemma 1, there exists a function $v : \mathbb{R} \to \mathbb{R}$ such that

$$\mathcal{S}(X) = \operatorname*{arg\,min}_{x \in \mathbb{R}}\{x + \mathbb{E}[v(X - x)]\}, \quad \mathcal{R}(X) = \min_{x \in \mathbb{R}}\{x + \mathbb{E}[v(X - x)]\}.$$

On the other hand, by Theorem 2, $\mathcal{R}(X) = \mathrm{ES}_\alpha^\lambda(X)$ for some $\lambda \in [0, 1]$ and $\alpha \in (0, 1)$. That is, there exists $\overline{v}(x) = \lambda' x + (\gamma - \lambda')x_+$ with $\lambda' = 1 - \lambda \in [0, 1]$ and $\gamma = (1 - \alpha\lambda')/(1 - \alpha) \in [1, \infty)$ such that

$$\mathcal{R}(X) = \mathrm{ES}_\alpha^\lambda(X) = \min_{x \in \mathbb{R}}\{x + \mathbb{E}[\overline{v}(X - x)]\}.$$

Denote by $\hat{v}$ the largest increasing convex function dominated by v. That is,

$$\hat{v}(x) = \sup\{g(x) : g \le \tilde{v} \text{ on } \mathbb{R},\ g \text{ is convex}\}, \quad x \in \mathbb{R}, \tag{22}$$

with $\tilde{v}(x) = \inf_{y \ge x} v(y)$. We then show the result by considering the following two cases.

(i) If $\hat{v}(0) = 0$, then by the proofs of Lemmas 2 to 5, we have $\overline{v} \le v$, and hence $v = \overline{v}$ by Lemma 4. By $\mathcal{S}(0) = 0$, which excludes $\lambda = 1$, we have $\mathcal{S}(X) = \mathrm{VaR}_\alpha(X)$.

(ii) If $\hat{v}(0) > 0$, then by the proof of Lemma 5, there exists $c \in \mathbb{R}$ such that $\hat{v}(c) = c$. Define $v^*(x) = v(x + c) - c$, $x \in \mathbb{R}$ and the corresponding $\hat{v^*}$ of v^* by (22). One can verify that

$\hat{v^*}(x) = \hat{v}(x+c) - c$, which implies $\hat{v^*}(0) = 0$, and

$$\mathcal{R}(X) = \min_{x\in\mathbb{R}}\{x + \mathbb{E}[v^*(X-x)]\}.$$

Similar to Case (i), we have $v^* \geq \overline{v}$ and thus, $v^* = \overline{v}$ by Lemma 4. Since $\mathcal{S}(0) = 0$ excludes $\lambda = 1$, it follows that $\mathcal{S}(X) = \mathrm{VaR}_\alpha(X) - c$. By $\mathcal{S}(0) = 0$ again, we have $c = 0$. This completes the proof. □

A.3 Proof of Theorem 4

We present two lemmas used in the proof of Theorem 4. The first lemma uses a weaker continuity (C') than (C) in Theorem 4. This result is similar to Theorem 3.1 of Weber (2006) which uses a different condition of ψ-weak lower semi-continuity to replace mixture concavity.

(C') There exists $x \leq 0$ such that for any $y > 0$, $\widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_y) \leq 0$ for small enough $\alpha > 0$.

Lemma 6. *Let $\mathcal{R}$ be a monetary risk measure on $\mathcal{X} = \mathcal{L}^\infty$ satisfying continuity (C') with $\mathcal{R}(0) = 0$. If $\mathcal{R}$ is mixture concave and has CxLS, then there exist $z \in \mathbb{R}$ and an increasing and left-continuous g such that*

$$\mathcal{R}(X) = \inf\{x \in \mathbb{R} : \mathbb{E}[g(X-x)] \leq z\}, \quad X \in \mathcal{X}. \tag{23}$$

Proof. Take $x \leq 0$ in assumption (C') and any fixed constant $y > 0$, and let $z \in (0,1)$ be such that $[0,z] = \{\alpha \in [0,1] : \widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_y) \leq 0\}$; the interval is closed since $\alpha \mapsto \mathcal{R}((1-\alpha)\delta_x + \alpha\delta_y)$ is concave. We define the function g as

$$g(t) = \begin{cases} \frac{z-\overline{\alpha}(t)}{1-\overline{\alpha}(t)}, & t \leq 0, \\ \frac{z}{\overline{\alpha}(t)}, & t > 0, \end{cases} \tag{24}$$

where

$$\overline{\alpha}(t) = \begin{cases} \sup\{\alpha \in [0,1] : \widehat{\mathcal{R}}((1-\alpha)\delta_t + \alpha\delta_y) \leq 0\}, & t \leq 0, \\ \sup\{\alpha \in [0,1] : \widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_t) \leq 0\}, & t > 0. \end{cases} \tag{25}$$

This is the same construction as in Eq. (3.5) and (3.7) of Weber (2006). Since for any $t \leq 0$, $\alpha \mapsto \widehat{\mathcal{R}}((1-\alpha)\delta_t + \alpha\delta_y)$ is increasing concave in $\alpha \in [0,1]$ and $\widehat{\mathcal{R}}(\delta_y) = y > 0$, we have $\overline{\alpha}(t) < 1$ for $t \leq 0$. By (C'), we have $\overline{\alpha}(t) > 0$ for $t > 0$. Hence, g is well defined. By monotonicity of $\mathcal{R}$, one can verify that g is increasing and satisfies $g(0) = z$. Next we show (23) and the left-continuity of g.

1. Denote by $\mathcal{S}$ the convex hull of $\{\delta_{x_i},\ i=1,\dots,n\}$, that is,

$$\mathcal{S}=\left\{\sum_{i=1}^n \alpha_i\delta_{x_i}:\alpha_i\geq 0,\ i=1,\dots,n,\ \sum_{i=1}^n\alpha_i=1\right\},$$

where $x_1=x$, $x_2=y$ are those fixed above. By mixture concavity of $\mathcal{R}$, one can verify that $(\alpha_1,\dots,\alpha_n)\mapsto\widehat{\mathcal{R}}(\sum_{i=1}^n\alpha_i\delta_{x_i})$ is a concave function, and thus is lower-semicontinuous by Theorem 10.2 of Rockafellar (1970). It follows that $\mathcal{N}:=\{(\alpha_1,\dots,\alpha_n):\widehat{\mathcal{R}}(\sum_{i=1}^n\alpha_i\delta_{x_i})\leq 0\}$ is closed sets in the Euclidean topology. By CxLS of $\mathcal{R}$, we have $\mathcal{N}$ and $\mathcal{S}\setminus\mathcal{N}$ are both convex sets. Using similar arguments of Weber (2006), one can verify that (23) holds for random variable X which has distribution in $\mathcal{S}$.

2. We next show g is left-continuous. Since g is increasing, it suffices to show $\lim_{s\uparrow t}g(s)=g(t)$ for $t\in\mathbb{R}$, which is equivalent to $\lim_{s\uparrow t}\overline{\alpha}(s)\leq\overline{\alpha}(t)$ as $\overline{\alpha}$ is decreasing.

   (a) For $t>0$, if $\overline{\alpha}(t)=1$, then $\lim_{s\uparrow t}\overline{\alpha}(s)\leq\overline{\alpha}(t)$ holds trivially as $\overline{\alpha}(s)\leq 1$ for any s.

   (b) For $t>0$ with $\overline{\alpha}(t)<1$, by definition of $\overline{\alpha}(t)$, we have for any $\epsilon\in(0,1-\overline{\alpha}(t))$, $\widehat{\mathcal{R}}((1-\overline{\alpha}(t)-\epsilon)\delta_x+(\overline{\alpha}(t)+\epsilon)\delta_t)>0$. Since $\mathcal{R}$ is monetary, $\widehat{\mathcal{R}}((1-\alpha)\delta_x+\alpha\delta_t)$ is continuous in $t\in\mathbb{R}$, and hence there exists $s_0<t$ such that $\widehat{\mathcal{R}}((1-\overline{\alpha}(t)-\epsilon)\delta_x+(\overline{\alpha}(t)+\epsilon)\delta_{s_0})>0$. That is, $\overline{\alpha}(s_0)\leq\overline{\alpha}(t)+\epsilon$. Therefore, $\lim_{s\uparrow t}\overline{\alpha}(s)\leq\overline{\alpha}(s_0)\leq\overline{\alpha}(t)+\epsilon$. As ϵ is arbitrary, we have $\lim_{s\uparrow t}\overline{\alpha}(s)\leq\overline{\alpha}(t)$.

   (c) For $t\leq 0$, we have $\overline{\alpha}(t)<1$. Similar arguments as in (b) yield $\lim_{s\uparrow t}\overline{\alpha}(s)\leq\overline{\alpha}(t)$.

3. Next we show that (23) holds for any $X\in\mathcal{X}$. Since $\mathcal{R}$ is monetary, there exist X_n, $n\in\mathbb{N}$, each taking values in a finite set, such that $X_n\uparrow X$ and $\lim_{n\to\infty}\mathcal{R}(X_n)=\mathcal{R}(X)$. Since g is left-continuous, we have $g(X_n-x)\uparrow g(X-x)$ and by the Monotone Convergence Theorem, we obtain $\lim_{n\to\infty}\mathbb{E}[g(X_n-x)]=\mathbb{E}[g(X-x)]$ for any $x\in\mathbb{R}$. This implies

$$\lim_{n\to\infty}\inf\{x\in\mathbb{R}:\mathbb{E}[g(X_n-x)]\leq z\}=\inf\{x\in\mathbb{R}:\mathbb{E}[g(X-x)]\leq z\}.$$

It then follows from $\lim_{n\to\infty}\mathcal{R}(X_n)=\mathcal{R}(X)$ and $\mathcal{R}(X_n)=\inf\{x\in\mathbb{R}:\mathbb{E}[g(X_n-x)]\leq z\}$ that (23) holds for any $X\in\mathcal{X}$. □

Lemma 7. *Let $\mathcal{R}$ be defined by (23) for an increasing function g and $z\in\mathbb{R}$ satisfying $\mathcal{R}(0)=0$. Then $\mathcal{R}$ satisfies (C) if and only if $g(t)<z$ for all $t<0$. Moreover, if $\mathcal{R}$ is mixture concave and satisfies (C), then g in (23) can be chosen continuous and strictly increasing on either $(-\infty,0)$ or $(0,\infty)$, and $\mathcal{R}(X)$ is the unique solution x to the equation $\mathbb{E}[g(X-x)]=z$.*

Proof. 1. We first show that $\mathcal{R}$ satisfies (C) if and only if $g(t) < z$ for $t < 0$. To see the "if" statement, by translation invariance, it suffices to show that $\widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_y)$ is continuous at $\alpha = 0$ for $x \le 0 < y$. For this, we will verify $\lim_{\alpha\downarrow 0} \widehat{\mathcal{R}}((1-\alpha)\delta_x+\alpha\delta_y) \le x$. For any $\epsilon \in (0, y-x)$, we have

$$\lim_{\alpha\downarrow 0}(1-\alpha)g(x-x-\epsilon) + \alpha g(y-x-\epsilon) = g(-\epsilon) < z.$$

Hence, there exists $\alpha_0 \in (0,1)$ such that $(1-\alpha_0)g(x-x-\epsilon) + \alpha_0 g(y-x-\epsilon) < z$, implying $\widehat{\mathcal{R}}((1-\alpha_0)\delta_x + \alpha_0\delta_y) \le x+\epsilon$. By monotonicity of $\mathcal{R}$, we have $\lim_{\alpha\downarrow 0} \widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_y) \le \widehat{\mathcal{R}}((1-\alpha_0)\delta_x + \alpha_0\delta_y) \le x+\epsilon$. As ϵ is arbitrary, we have $\lim_{\alpha\downarrow 0} \widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_y) \le x$.

To see the "only if" statement, suppose that there exists $\epsilon > 0$ such that $g(t) = z$ for $z \in (-\epsilon, 0)$. By $\mathcal{R}(0) = 0$, we have $z < g(t)$ for any $t > 0$. It follows that $\widehat{\mathcal{R}}((1-\alpha)\delta_0 + \alpha\delta_{\epsilon/2}) \ge \epsilon/2$ for any $\alpha > 0$. This contradicts (C).

2. In what follows, we take g from (24) in the proof of Lemma 6. We will show that g is continuous and strictly increasing on either $(-\infty, 0)$ or $(0,\infty)$ by contradiction. We have seen from Step 1 above that (C) and $\mathcal{R}(0) = 0$ together imply that g is strictly increasing at 0. Suppose that there exist $a < b < 0 < c < d$ such that $[a, b] = \{x : g(x) = g(a)\}$ and $[c, d] = \{x : g(x) = g(c)\}$. As $g(a) < z < g(c)$, there exists $\alpha_0 \in [0, 1)$ such that $(1-\alpha_0)g(a) + \alpha_0 g(c) = g(0)$, and hence

$$(1-\alpha)g(a) + \alpha g(c) > z \quad \text{for any } \alpha > \alpha_0. \tag{26}$$

Since $g(t) < z$ for $t < 0$, there exist $\alpha_1 > \alpha_0$ and $\epsilon \in (0, \min\{(b-a)/2, (d-c)/2\})$ such that

$$\frac{1-\alpha_1}{2}g(a) + \frac{\alpha_1}{2}g(c) + \frac{1}{2}g(-\epsilon) < z. \tag{27}$$

Define the distribution $F_{x,y} = (1-\alpha_1)\delta_x+\alpha_1\delta_y$ with $x \in [a+2\epsilon, b]$ and $y \in [c+2\epsilon, d]$. Then by (26), we have $\widehat{\mathcal{R}}(F_{x,y}) > 2\epsilon$. By letting $G = \frac{1}{2}F_{x,y}+\frac{1}{2}\delta_0$, (27) implies $\widehat{\mathcal{R}}(G) \le \epsilon < \widehat{\mathcal{R}}(F_{x,y})/2+\widehat{\mathcal{R}}(\delta_0)/2$, yielding a contradiction. Hence, g is strictly increasing on either $(-\infty, 0)$ or $(0,\infty)$.

3. Using results in Steps 1 and 2, for any $X \in \mathcal{X}$, we have $\mathbb{E}[g(X-x)]$ is strictly decreasing for x in the range of X, and thus, the equation $\mathbb{E}[g(X-x)] = z$ has a unique solution x.

4. Finally we show that g is continuous. Using (C), $\overline{\alpha}$ defined by (25) satisfies that $\overline{\alpha}(t)$ is the unique solution to the equation in α, $\widehat{\mathcal{R}}((1-\alpha)\delta_t + \alpha\delta_y) = 0$, if $t \le 0$, and it is the unique solution α to $\widehat{\mathcal{R}}((1-\alpha)\delta_x + \alpha\delta_t) = 0$ if $t > 0$. Since $\mathcal{R}$ is monetary and hence L^∞-continuous, we have that $\widehat{\mathcal{R}}((1-\alpha)\delta_{x_1} + \alpha\delta_{x_2})$ is continuous in $(x_1, x_2) \in \mathbb{R}^2$. Hence, $\overline{\alpha}$ is continuous and

thus, g is continuous. $\square$

Proof of Theorem 4. The "if" statement is argued in (ii) of Example 1, where we see that the entropic risk measure is both a Bayes risk measure and a Bayes estimator. Hence, it is mixture concave and has CxLS. To prove the "only if" statement, first note that by using Lemmas 6 and 7, we have (15) holds with g and z satisfying that g is continuous, and strictly increasing on either $(-\infty, 0)$ or $(0, \infty)$, and the equation $\mathbb{E}[g(X-x)] = z$ always has a unique solution. Further, by Lemma 7 and $\mathcal{R}(0) = 0$, we have $g(-t) < z < g(t)$ for all $t > 0$. We then employ the following steps to show that $\mathcal{R}$ must be an entropic risk measure.

1. Note that a monotone function g has derivatives almost everywhere. Let t be a point such that $g(t) < z$ and g is differentiable at t. Take arbitrary $x, y > 0$. Since $g(t) < z < \min\{g(x), g(y)\}$, for each $\epsilon \in (0, \min\{x, y\})$, there exist unique $\lambda_1(\epsilon) \in (0,1)$ and $\lambda_2(\epsilon) \in (0,1)$ such that

$$\lambda_1(\epsilon) g(x-\epsilon) + (1-\lambda_1(\epsilon)) g(t) = z \quad \text{and} \quad \lambda_2(\epsilon) g(y-\epsilon) + (1-\lambda_2(\epsilon)) g(t) = z. \tag{28}$$

   As g is increasing, $\lambda_i(\epsilon)$ is decreasing in ϵ, $i = 1, 2$. Let $\lambda_i^0 = \lim_{\epsilon \downarrow 0} \lambda_i(\epsilon) \in (0,1)$ for $i = 1, 2$, and we have

$$\lambda_1^0 g(x) + (1-\lambda_1^0) g(t) = z \quad \text{and} \quad \lambda_2^0 g(y) + (1-\lambda_2^0) g(t) = z. \tag{29}$$

2. Let a random variable X be given by $\mathbb{P}(X = x) = \lambda_1^0$ and $\mathbb{P}(X = t) = 1 - \lambda_1^0$, and Y_ϵ be given by $\mathbb{P}(Y_\epsilon = y + \epsilon) = \lambda_2(\epsilon)$ and $\mathbb{P}(Y_\epsilon = t + 2\epsilon) = 1 - \lambda_2(\epsilon)$ for $\epsilon > 0$.

3. Since g is strictly increasing at either t or x, the first equation of (29) gives the inequality $\lambda_1^0 g(x+\delta) + (1-\lambda_1^0) g(t+\delta) > z$ for any $\delta > 0$. This implies $\mathcal{R}(X) = 0$. Similarly, we have $\mathcal{R}(Y_\epsilon) = 2\epsilon$.

4. Let Z have a distribution with is a mixture of the distributions of X and Y_ϵ with weight $1/2$ each. Using mixture concavity, we have $\mathcal{R}(Z) \geq \epsilon$, meaning that $\mathbb{E}[g(Z-\epsilon)] \geq z$. Hence, we have

$$\lambda_1^0 g(x-\epsilon) + (1-\lambda_1^0) g(t-\epsilon) + \lambda_2(\epsilon) g(y) + (1-\lambda_2(\epsilon)) g(t+\epsilon) \geq 2z.$$

   Subtracting the second equality in (28) and the first equality in (29) from the above equation, we get

$$\lambda_1^0 \left(g(x-\epsilon) - g(x)\right) + \lambda_2(\epsilon) \left(g(y) - g(y-\epsilon)\right)$$
$$+ (1-\lambda_1^0) \left(g(t-\epsilon) - g(t)\right) + (1-\lambda_2(\epsilon)) \left(g(t+\epsilon) - g(t)\right) \geq 0.$$

Divide the above equation by ϵ, and letting $\epsilon \downarrow 0$, we obtain

$$\lambda_2^0 \liminf_{\epsilon \downarrow 0} \frac{g(y) - g(y-\epsilon)}{\epsilon} - \lambda_1^0 \limsup_{\epsilon \downarrow 0} \frac{g(x) - g(x-\epsilon)}{\epsilon} \geq (\lambda_2^0 - \lambda_1^0) g'(t), \tag{30}$$

where we use $\lambda_2^0 = \lim_{\epsilon \downarrow 0} \lambda_2(\epsilon)$. Since the positions of (x, λ_1^0) and (y, λ_2^0) are symmetric, we also have

$$\lambda_1^0 \liminf_{\epsilon \downarrow 0} \frac{g(x) - g(x-\epsilon)}{\epsilon} - \lambda_2^0 \limsup_{\epsilon \downarrow 0} \frac{g(y) - g(y-\epsilon)}{\epsilon} \geq (\lambda_1^0 - \lambda_2^0) g'(t), \tag{31}$$

Combining (30) and (31), we conclude that the two inequalities in (30) and (31) are both equalities, and because $\lambda_1^0, \lambda_2^0 > 0$, we have

$$\liminf_{\epsilon \downarrow 0} \frac{g(x) - g(x-\epsilon)}{\epsilon} = \limsup_{\epsilon \downarrow 0} \frac{g(x) - g(x-\epsilon)}{\epsilon}.$$

That is, g has a left-derivative at x and y. Similarly, we can show that it also has a right-derivative at x and y, so that

$$\lambda_2^0 g'(y) - \lambda_1^0 g'(x) = (\lambda_2^0 - \lambda_1^0) g'(t), \tag{32}$$

for all $x, y > 0$. By (29), we can write $\lambda_1^0 = \frac{z-g(t)}{g(x)-g(t)}$ and $\lambda_2^0 = \frac{z-g(t)}{g(y)-g(t)}$. Substituting them into (32) yields

$$(g'(x) - g'(t))(g(y) - g(t)) = (g'(y) - g'(t))(g(x) - g(t)). \tag{33}$$

5. By fixing t and y and noting that $g(y) > g(t)$, (33) can be rewritten as

$$g'(x) - b g(x) = d \tag{34}$$

for some constants b and d. Solving (34), we obtain that, on $(0, \infty)$, either g is linear or $g(x) = ae^{bx} + c$ for some constants a, b, c. Similarly, by fixing x and y, we have that, almost everywhere on $(-\infty, 0)$, either g is linear or $g(x) = a'e^{b'x} + c'$ for some constants a', b', c'; continuity of g now implies that the above for holds on $(-\infty, 0)$.

6. From the previous step, g indeed has a positive derivative at any point $t < 0$. Hence, (33) holds for all $x, y > 0$ and $t < 0$. Note that (33) and the continuity of g imply that g' is continuous at 0. The forms of g on $(-\infty, 0)$ and on $(0, \infty)$ have three parameters each (the linear case corresponds to the limit of $b \to 0$ after normalization). We obtain three equations from $g'(x)(g(y) - c') = g'(y)(g(x) - c')$ (obtained by letting $t \to -\infty$) and the continuity of g and g' at 0, and these three equations give $a = a'$, $b = b'$ and $c = c'$. Hence, we conclude that either g is linear or $g(x) = ae^{bx} + c$ on $\mathbb{R}$.

7. If g is linear, then $\mathcal{R} = \mathrm{ER}_0 = \mathbb{E}$. If g is not linear, then $ab > 0$ since $\mathcal{R}$ is monotone. Moreover, (15) implies
$$\mathcal{R}(X) = \frac{1}{b} \log \mathbb{E}[e^{bX}], \quad X \in \mathcal{X}.$$
Since log is a concave function, mixture concavity does not hold if $b < 0$ (in this case, $\mathcal{R}$ is mixture convex). Hence, $b > 0$, and $\mathcal{R} = \mathrm{ER}_b$. □

References

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

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

BCBS (2019). *Minimum Capital Requirements for Market Risk. February 2019.* 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.

Bellini, F., Bignozzi, V. and Puccetti, G. (2018). Conditional expectiles, time consistency and mixture convexity properties. *Insurance: Mathematics and Economics*, **82**, 117–123.

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

Ben Tal, A. and Teboulle, M. (2007). An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, **17**(3), 449–476.

Brehmer, J. and Gneiting, T. (2021). Scoring interval forecasts: Equal-tailed, shortest, and modal interval. *Bernoulli*, forthcoming.

Chong, W. F., Feng, R. and Jin, L. (2021). Holistic principle for risk aggregation and capital allocation. *Annals of Operations Research*, published online.

Chew, S. H. (1983). A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the allais paradox, *Econometrica*, **51**, 1065–1092.

Cifarelli, D. M. and Regazzini, E. (1996). De Finetti's contribution to probability and statistics. *Statistical Science*, 253–282.

Davis, M. H. A. (2013). Consistency of risk measure estimates. *Risk Management & Analysis in Financial Institutions eJournal. SSRN*: 2342279.

Davis, M. H. A. (2014). The Beaufort scale of predictability. In Podolskij, M., Stelzer, R., Thorbj rnsen, S. and Almut E. D. Veraart, A.E.D. (Eds.) (2014). *The Fascination of Probability, Statistics and their Applications. In honour of Ole E. Barndorff-Nielsen*. Springer, 419–434.

Davis, M. H. A. (2016). Verification of internal risk measure estimates. *Statistics and Risk Modeling*, **33**(3-4), 67–93.

Davis, M. H. A. (2017). Discussion of "elicitability and backtesting: Perspectives for banking regulation" *Annals of Applied Statistics*, **11**(4), 1886–1887.

Dawid, A. P. (1984). Present position and potential developments: some personal views. Statistical theory: the prequential approach (with discussion). *Journal of the Royal Statistical Society: Series A*, **147**, 278–292.

Dawid, A. P. and Vovk, V. (1999). Prequential probability: principles and properties. *Bernoulli*, **5**, 125–162.

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

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

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

Fissler, T., Hlavinová, J. and Rudloff, B. (2020). Forecast evaluation of quantiles, prediction intervals, and other set-valued functionals. *arXiv*: 1910.07912v2.

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

Föllmer, H. and Knispel, T. (2011). Entropic risk measures: Coherence vs. convexity, model ambiguity and robust large deviations. *Stochastics and Dynamics*, **11**, 333–351.

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

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

Frongillo, R. and Kash, I. A. (2021). Elicitation complexity of statistical properties. *Biometrika*, published online.

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

Grant, S., Kajii, A. and Polak, B. (2000). Temporal resolution of uncertainty and recursive non-expected utility models. *Econometrica*, **68**(2), 425–434.

Grünwald, P. D. and Dawid, A. P. (2004). Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Annals of Statistics*, **32**(4), 1367–1433.

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

Kupper, M. and Schachermayer, W. (2009). Representation results for law invariant time consistent functions, *Mathematics and Financial Economics*, **2**, 189–210.

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. (2021). A theory for measures of tail risk. *Mathematics of Operations Research*, published online.

Mao, T., Hu, J. and Liu, H. (2018). The average risk sharing problem under risk measure and expected utility theory. *Insurance: Mathematics and Economics*, **83**, 170–179.

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

Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *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. (1970). *Convex Analysis.* Princeton University Press.

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.

Schmeidler, D. (1986). Integral representation without additivity. *Proceedings of the American Mathematical Society*, **97**(2), 255–261.

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

Wang, R. and Wei, Y. (2020). Risk functionals with convex level sets. *Mathematical Finance*, **30**(4), 1337–1367.

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

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

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

Wu, Q., Mao, T. and Hu, T. (2020). Optimized certainty equivalent with applications in the rank-dependent expected utility model. *SSRN*: 3694866.

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