How superadditive can a risk measure be?

Ruodu Wang[*], Valeria Bignozzi[†] and Andreas Tsanakas[‡]

March 13, 2015[§]

Abstract

In this paper, we study the extent to which any risk measure can lead to superadditive risk assessments, implying the potential for penalizing portfolio diversification. For this purpose we introduce the notion of extreme-aggregation risk measures. The extreme-aggregation measure characterizes the most superadditive behavior of a risk measure, by yielding the worst-possible diversification ratio across dependence structures. One of the main contributions is demonstrating that, for a wide range of risk measures, the extreme-aggregation measure corresponds to the smallest dominating coherent risk measure. In our main result, it is shown that the extreme-aggregation measure induced by a distortion risk measure is a coherent distortion risk measure. In the case of convex risk measures, a general robust representation of coherent extreme-aggregation measures is provided. In particular, the extreme-aggregation measure induced by a convex shortfall risk measure is a coherent expectile. These results show that, in the presence of dependence uncertainty, quantification of a coherent risk measure is often necessary, an observation that lends further support to the use of coherent risk measures in portfolio risk management.

Key-words: distortion risk measures; shortfall risk measures; expectiles; dependence uncertainty; risk aggregation; diversification.

AMS Classification: 91B30; 62P05.

[*]Department of Statistics and Actuarial Science, University of Waterloo, Canada. (email: `wang@uwaterloo.ca`)

[†]Corresponding author. School of Economics and Management, University of Florence, 50127 Florence, Italy. (email: `valeria.bignozzi@unifi.it`)

[‡]Cass Business School, City University London, UK. (email: `a.tsanakas.1@city.ac.uk`)

[§]The authors thank an Associate Editor and three anonymous referees for valuable comments, and Paul Embrechts (ETH Zurich), Giovanni Puccetti (U Firenze), Jan Dhaene (KU Leuven), Alexander McNeil (Heriot Watt U), Fabio Bellini (U Milano-Bicocca) and Tiantian Mao (U Waterloo) for discussions on a previous version of the paper. Ruodu Wang acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Forschungsinstitut für Mathematik (FIM) at ETH Zurich during his visits in 2013-2014. Valeria Bignozzi acknowledges RiskLab and the Swiss Finance Institute for financial support.

1 Introduction

Debate on the desirability of alternative sets of properties for risk measures has been raging since the seminal paper of Artzner et al. (1999) on coherent measures of risk. Key to coherence is the property of subadditivity, implying that the merging of risky positions should always yield capital savings. Subadditivity, as well as the alternative notion of convexity, has gained wide acceptance in the literature; see Föllmer and Schied (2011, Chapter 4) for an extensive treatment. However, dissenting voices have persisted. For instance, Dhaene et al. (2008) criticize subadditivity from a regulatory perspective, considering the impact of mergers on shortfall risk. Cont et al. (2010) argue that empirical estimators of coherent risk measures lack classical robustness properties, while Krätschmer et al. (2014) introduce a generalized notion of robustness that allows comparisons between coherent risk measures. A further twist is added by considering the property of elicitability (Gneiting, 2011) of a risk measure. A risk measure is elicitable if it can be written as the unique minimizer of a suitable expected loss function; this representation provides a natural statistic for assessing the performance of statistical procedures used to estimate the risk measure. Such discussions are exemplified by the comparative advantages of the (coherent, non-elicitable, less robust) Expected Shortfall (ES) and the (non-coherent, elicitable, more robust) Value-at-Risk (VaR) measures; see Embrechts et al. (2014) and Emmer et al. (2014) for reviews of such arguments. Related debates are not caged within academia; discussions on a potential transition from VaR to ES in regulation and risk assessment are sought by the Basel Committee on Banking Supervision in two recent consultative documents BCBS (2012, 2013), and by the International Association of Insurance Supervisors in a more recent document IAIS (2014).

In this paper, we focus on subadditivity, the key property which distinguishes coherent risk measures (such as ES) and non-coherent risk measures (such as VaR). For a risk measure $\rho : \mathcal{X} \to [-\infty, \infty]$ where $\mathcal{X}$ is a set of risks (random variables), we consider the *diversification ratio* (see for instance Embrechts et al., 2014 and Emmer et al., 2014) for a portfolio $\mathbf{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$, defined as

$$\Delta^{\mathbf{X}}(\rho) = \frac{\rho(X_1 + \cdots + X_n)}{\rho(X_1) + \cdots + \rho(X_n)},$$

that is, the ratio of portfolio risk over the sum of the risks of individual positions. Lack of subadditivity makes the value of $\Delta^{\mathbf{X}}(\rho)$ potentially greater than one, indicating lack of capital savings from diversification. Considering what the largest possible value of $\Delta^{\mathbf{X}}(\rho)$ can be, leads to the fundamental question we attempt to address in this paper: *How superadditive can a risk measure be?*

To answer this question, we focus on the properties of law-invariant risk measures themselves, rather than those of specific portfolios. For that reason, we consider homogeneous portfolios with identical marginal distributions F of size n and let n vary. We then introduce the law-invariant risk measure

$$\Gamma_{\rho,n} = \frac{1}{n} \sup\{\rho(X_1 + \cdots + X_n) : X_i \in \mathcal{X},\ X_i \sim F,\ i = 1, \dots, n\},$$

that is, the worst-case value of the aggregate risk, scaled by $1/n$, across all homogeneous portfolios with marginal distribution F. The largest possible value of $\Delta^{\mathbf{X}}(\rho)$ can be directly obtained from $\Gamma_{\rho,n}$. Subsequently, dependence on the portfolio size is eliminated by defining the *extreme-aggregation measure* induced by ρ as $\Gamma_\rho = \limsup_{n\to\infty} \Gamma_{\rho,n}$, thus considering worst-case diversification under extreme portfolio aggregations.

Γ_ρ is itself a risk measure with many properties inherited from ρ. It provides risk measurement under the most adverse dependence structure for given marginal distributions, which is of interest in the study of *dependence uncertainty*; see for instance Bernard et al. (2014) and Embrechts et al. (2015). In applications such as operational risk modeling, the dependence structure between risks is typically unknown, with very limited empirical evidence to allow for its estimation. When there is insufficient data to estimate the dependence structure of a portfolio, it is necessary to calculate the risk measure Γ_ρ to derive an upper bound on the portfolio risk, even when the portfolio is inhomogeneous (see discussions in Section 3).

We proceed by deriving explicit expressions for the extreme-aggregation measure induced by common risk measures and find that it is in fact coherent in cases of interest. We start with the class of distortion risk measures (Yaari, 1987; Wang et al., 1997; Acerbi, 2002), originating from early study on non-additive measures (Denneberg, 1990, 1994). Distortion risk measures include both VaR and ES as special cases. The main theorem in this paper shows that the extreme-aggregation measure induced by a distortion risk measure ρ is the smallest coherent distortion risk measure dominating ρ. An asymptotic equivalence (in the sense of Embrechts et al., 2014) of distortion risk measures with their extreme-aggregation measures is established, in the case of inhomogeneous portfolios.

A further class of interest is that of convex risk measures (studied by Föllmer and Schied (2002) and Frittelli and Rosazza Gianin (2002) in financial capital requirements and early by Deprez and Gerber (1985) in insurance pricing). We show that when ρ is a convex risk measure, the corresponding extreme-aggregation measure Γ_ρ is the smallest coherent risk measure dominating ρ, and a robust representation of Γ_ρ is thereby provided for ρ satisfying the Fatou property. In the specific case of shortfall risk measures, we show that Γ_ρ is identified with an *expectile* (Newey and Powell, 1987), which is the only coherent shortfall risk measure (Weber, 2006), as well as the only elicitable coherent risk measure (Ziegel, 2014; Delbaen et al., 2015).

In summary, for a wide range of risk measures, Γ_ρ corresponds to the smallest coherent risk measure dominating ρ. These results show that, in the presence of dependence uncertainty, the worst-possible value of a non-coherent risk measure often equals to the value of a coherent risk measure on the same portfolio, an observation that lends further support to the use of coherent risk measures in portfolio risk management.

The structure of the paper is as follows. In Section 2 we list the definitions and notation used in this paper, and connect the diversification ratio of a portfolio with the notion of an extreme-aggregation measure. In Section 3, distortion risk measures are considered and the form of the induced extreme-aggregation measures is obtained. Section 4 deals with the extreme-aggregation measures of convex risk measures. Brief conclusions are stated in Section 5, while all proofs are collected in the Appendix.

2 Diversification and extreme-aggregation

2.1 Definitions and notation

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be an atomless probability space and $L^p := L^p(\Omega, \mathcal{A}, \mathbb{P})$ be the set of all random variables in the probability space with finite p-th moment, $p \in [0, \infty]$. A positive (negative) value of $X \in L^0$ represents a financial loss (profit).

A *risk measure* $\rho : \mathcal{X} \to [-\infty, +\infty]$ assigns to every financial loss $X \in \mathcal{X}$ a real number (or infinity) $\rho(X)$, where the set $\mathcal{X}$ is a convex cone, and $L^\infty \subset \mathcal{X} \subset L^0$ ($\subset$ is the non-strict set inclusion). We always let $\rho(X) \in \mathbb{R}$ for all $X \in L^\infty$ to avoid triviality. We gather here some of the standard properties often required for risk measures. A risk measure ρ may satisfy, for any $X, Y \in \mathcal{X}$:

(a) *Monotonicity*: if $X \leqslant Y$ $\mathbb{P}$-a.s, then $\rho(X) \leqslant \rho(Y)$; (b) *Cash-invariance*: for any $m \in \mathbb{R}$, $\rho(X - m) = \rho(X) - m$; (c) *Positive homogeneity*: for any $\alpha \geqslant 0$, $\rho(\alpha X) = \alpha\rho(X)$; (d) *Subadditivity*: $\rho(X+Y) \leqslant \rho(X)+\rho(Y)$; (e) *Convexity*: for any $\lambda \in [0, 1]$, $\rho(\lambda X+(1-\lambda)(Y)) \leqslant \lambda\rho(X)+(1-\lambda)\rho(Y)$; (f) *Zero-normalization*: $\rho(0) = 0$, (this is implied by (c)); (g) *Comonotonic additivity*: if X, Y are comonotonic, then $\rho(X + Y) = \rho(X) + \rho(Y)$; (h) *Law-invariance*: if X and Y have the same distribution under $\mathbb{P}$, denoted as $X \overset{\mathrm{d}}{=} Y$, then $\rho(X) = \rho(Y)$.

In the above, we say that X and Y are *comonotonic* if

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0 \text{ for a.s. } (\omega, \omega') \in \Omega \times \Omega,$$

(see for instance Föllmer and Schied, 2011, Definition 4.82).

The properties listed above and many more can be found in different literatures, such as finance (e.g. capital setting; Artzner et al., 1999, Föllmer and Schied, 2002, Frittelli and Rosazza Gianin, 2002), insurance (e.g. premium calculation; Bühlmann, 1970, Gerber, 1974, Goovaerts et al., 1984, Deprez and Gerber, 1985, Wang et al., 1997) and economics (e.g. choice under risk; Yaari, 1987; Schmeidler, 1989) with mathematical representations corresponding to different sets of properties typically provided. Interpretations of risk measure properties vary with applications and a detailed discussion is not given here.

Following the terminology of Föllmer and Schied (2011), a *monetary risk measure* satisfies (a, b), a *convex risk measure* satisfies (a, b, e), and a *coherent risk measure* satisfies (a-d) (or equivalently (a-c, e)). Most commonly studied risk measures also satisfy (h), for practical applications and statistical tractability.

The risk measure most commonly used in banking and insurance for capital setting purposes is *Value-at-Risk* (VaR), defined as

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

which satisfies (a-c, f-h), but not (d) or (e). A coherent alternative to VaR is *Expected Shortfall* (ES),

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

satisfying properties (a-h). A convex but not positively homogeneous risk measure frequently encountered in the literature is the *entropic risk measure* (ER)

$$\mathrm{ER}_\lambda(X) = \frac{1}{\lambda}\log \mathbb{E}[e^{\lambda X}], \quad \lambda > 0, \quad X \in L^0, \tag{2.1}$$

which finds its origins in indifference pricing (Gerber, 1974) and satisfies properties (a, b, e, f, h). VaR and ES belong to the class of *distortion risk measures*, while ER is an example of a *convex shortfall risk measure*.

For all risk measures discussed in this paper law-invariance (h) is assumed and not explicitly stated as a property from now on; in the same sense, we assume that if $X \in \mathcal{X}$ and $Y \overset{\mathrm{d}}{=} X$, then $Y \in \mathcal{X}$. We use $X \sim F$ to indicate that $X \in \mathcal{X}$ and X has distribution F; this, implicitly, assumes that all random variables with distribution F are in the set $\mathcal{X}$ of our interest. Throughout the paper, we denote by X_F any random variable with distribution F on $\mathbb{R}$, that is, $X_F \sim F$. For any distribution function F, we denote the generalized inverse function

$$F^{-1}(t) = \inf\{x : F(x) \geqslant t\}, \ t \in (0,1], \ \text{ and } \ F^{-1}(0) = \sup\{x : F(x) = 0\}.$$

A risk measure may not be well defined on all L^0 random variables. Specific constrains on $\mathcal{X}$ relating to families of risk measures are stated when these risk measures are defined in Sections 3 and 4.

2.2 Diversification ratio of a risk measure

For a risk measure ρ and a portfolio of risks $\mathbf{X} = (X_1, \dots, X_n) \in \mathcal{X}^n$ with $0 < \rho(X_1), \dots, \rho(X_n) < \infty$ we define the *diversification ratio*

$$\Delta^{\mathbf{X}}(\rho) = \frac{\rho(X_1 + \cdots + X_n)}{\rho(X_1) + \cdots + \rho(X_n)}. \tag{2.2}$$

$\Delta^{\mathbf{X}}(\rho)$ is a measure of portfolio diversification; see the discussions of Embrechts et al. (2014) and Emmer et al. (2014), as well as the references therein[1]. When $\Delta^{\mathbf{X}}(\rho) \leqslant 1$, a *diversification benefit* is indicated, a situation always guaranteed by the subadditivity of ρ. When ρ is not subadditive, $\Delta^{\mathbf{X}}(\rho) > 1$ is possible.

For a non-subadditive risk measure ρ, we are interested in the largest possible values for $\Delta^{\mathbf{X}}(\rho)$, characterizing the worst-case diversification scenario. In this paper, we treat $\Delta^{\mathbf{X}}_{\rho}$ as a property of ρ rather than the diversification characteristic of individual portfolios. For that reason, we consider homogeneous portfolios, $X_i \sim F,\ i = 1, \ldots, n$. Denote the set of possible portfolio risks with identical marginal distributions F,

$$\mathfrak{S}_n(F) = \{X_1 + \cdots + X_n : X_i \sim F,\ i = 1, \ldots, n\}, \qquad n \in \mathbb{N}.$$

Assuming (for now) $0 < \rho(X_F) < \infty$, we define the *n-superadditivity ratio*, for $n \in \mathbb{N}$,

$$\Delta^F_n(\rho) = \sup\left\{\frac{\rho(X_1 + \cdots + X_n)}{\rho(X_1) + \cdots + \rho(X_n)} : X_i \sim F,\ i = 1, \ldots, n\right\} = \frac{\sup\{\rho(S) : S \in \mathfrak{S}_n\}}{n\rho(X_F)}. \tag{2.3}$$

Taking the supremum in $\Delta^F_n(\rho)$ serves to reflect the question of "how superadditive" the risk measure ρ can become. If ρ is comonotonic additive or positively homogeneous, then choosing $X_1 = \cdots = X_n$ a.s. leads to $\rho(X_1 + \cdots + X_n) = \rho(X_1) + \cdots + \rho(X_n)$; by the supremum in the definition of $\Delta^F_n(\rho)$ it follows that $\Delta^F_n(\rho) \geqslant 1$. For subadditive risk measures ρ, it is $\Delta^F_n(\rho) \leqslant 1$. Hence for coherent risk measures ρ (that are subadditive and positive homogeneous), $\Delta^F_n(\rho) = 1$.

When ρ is not coherent, the calculation of $\Delta^F_n(\rho)$ is not easy and is known as the *Fréchet problem*; see Embrechts and Puccetti (2006) for a study on $\rho = \mathrm{VaR}_p$, $p \in (0, 1)$. Wang et al. (2013) gave the value of $\sup\{\mathrm{VaR}_p(S) : S \in \mathfrak{S}_n\}$ when F has a tail-decreasing density, leading to the explicit value of $\Delta^F_n(\mathrm{VaR}_p)$. For general risk measures ρ or general marginal distributions F, explicit values of $\Delta^F_n(\rho)$ are not available. Even in the case of VaR, analytical results are very limited; see Embrechts et al. (2013) for a numerical method. We are particularly interested in determining the *overall superadditivity ratio* $\sup_{n\in\mathbb{N}} \Delta^F_n(\rho)$, quantifying the greatest possible n-superadditivity ratio across all portfolio sizes n and thus providing an answer to our question of "how superadditive a risk measure can be", as well as characterizing worst-case diversification.

For the particular case of VaR it has been shown that

$$\sup_{n\in\mathbb{N}} \Delta^F_n(\mathrm{VaR}_p) = \lim_{n\to\infty} \Delta^F_n(\mathrm{VaR}_p) = \frac{\mathrm{ES}_p(X_F)}{\mathrm{VaR}_p(X_F)}, \tag{2.4}$$

such that the worst diversification of VaR can be characterized via the associated ES. Puccetti and Rüschendorf (2014) showed (2.4) under an assumption of complete mixability; Puccetti et al. (2013)

[1]For comonotonic additive ρ, the denominator of (2.2) becomes the risk measure of the sum of comonotonic risks with corresponding marginal distributions; this is typically interpreted as a worst-case dependence scenario (see e.g. Dhaene et al., 2002, 2012).

for a strictly positive density; Wang (2014) for a bounded density; finally, Wang and Wang (2015) for any distribution. In the present paper we consider a much more general version of (2.4).

The assumption $0 < \rho(X_F) < \infty$ guarantees that $\Delta_n^F(\rho)$ remains easily interpretable and corresponds to the practical situation that F is the distribution of a loss. However, this is not a mathematical requirement. One may define the law-invariant risk measure

$$\Gamma_{\rho,n}(X_F) := \frac{1}{n}\sup\{\rho(S) : S \in \mathfrak{S}_n(F)\}, \qquad n \in \mathbb{N},$$

that is, the highest possible risk of the homogeneous portfolio $X_1, \ldots, X_n \sim F$, normalized by $1/n$; in particular $\Gamma_{\rho,1}(X_F) = \rho(X_F)$. It is immediate that, whenever well defined, $\Delta_n^F(\rho)$ can be written as $\Delta_n^F(\rho) = \frac{\Gamma_{\rho,n}(X_F)}{\rho(X_F)}$. With this in mind, in the sequel we do not require $\rho(X_F) > 0$ and work with $\Gamma_{\rho,n}(X_F)$ instead of $\Delta_n^F(\rho)$.

2.3 Extreme-aggregation measures

We now discuss the superadditivity and diversification properties of risk measures through a global version of $\Gamma_{\rho,n}$, which does not depend on the portfolio size n. First we consider risk measures that satisfy comonotonic additivity, positive homogeneity or convexity[2], corresponding to most risk measures encountered in practice. For such risk measures, the limit of $\Gamma_{\rho,n}$ as $n \to \infty$ corresponds to the largest possible value of $\Gamma_{\rho,n}$ among all possible portfolio sizes n. We assume $\rho(X_F) > -\infty$ throughout the rest of the paper to avoid pathological cases without loss of generality.

Proposition 2.1. *If the risk measure ρ is (i) positively homogeneous or (ii) comonotonic additive or (iii) convex and zero-normalized, then the following hold:*

(a) For all $n, k \in \mathbb{N}$, it is $\Gamma_{\rho,n} \leqslant \Gamma_{\rho,kn}$;

(b) $\limsup_{n\to\infty} \Gamma_{\rho,n} = \sup_{n\in\mathbb{N}} \Gamma_{\rho,n}$. In case of (ii) or (iii), we have that $\lim_{n\to\infty} \Gamma_{\rho,n} = \sup_{n\in\mathbb{N}} \Gamma_{\rho,n}$.

For any risk measure ρ, the following holds:

(c) For any subadditive risk measure ρ^+ such that $\rho^+ \geqslant \rho$, it is $\Gamma_{\rho,n} \leqslant \rho^+$ for all $n \in \mathbb{N}$.

Motivated by Proposition 2.1(b), we introduce the notion of *extreme-aggregation (risk) measures*, which characterizes the worst-case risk measure among all homogeneous portfolios, under which the risk measure ρ is "at its most superadditive".

[2]In the presence of monotonicity, it is actually found that comonotonic additivity implies positive homogeneity Föllmer and Schied (2011). However, monotonicity is not required here, allowing for commonly used risk measures, such as the standard deviation.

Definition 2.1. The *extreme-aggregation measure* Γ_ρ induced by a risk measure ρ is defined as

$$\Gamma_\rho : \mathcal{X} \to \mathbb{R} \cup \{-\infty, \infty\}, \quad \Gamma_\rho(X_F) = \limsup_{n\to\infty} \left\{ \frac{1}{n} \sup\{\rho(S) : S \in \mathfrak{S}_n(F)\} \right\}.$$

Consider a risk measure that satisfies the assumptions of Proposition 2.1, such that $\Gamma_\rho(X_F) = \sup_{n\in\mathbb{N}} \Gamma_{\rho,n}(X_F)$ holds. Then, by definition, for $0 < \rho(X_F) < \infty$, it is

$$\sup_{n\in\mathbb{N}} \Delta_n^F(\rho) = \frac{\Gamma_\rho(X_F)}{\rho(X_F)}. \tag{2.5}$$

Some properties of ρ are inherited by Γ_ρ, as summarized below.

Lemma 2.2. *If a risk measure ρ satisfies any of the properties (a-f) in Section 2.1, then Γ_ρ inherits the corresponding properties. Moreover, if ρ is (i) positively homogeneous or (ii) comonotonic additive or (iii) convex and zero-normalized, then $\Gamma_\rho \geqslant \rho$; if ρ is subadditive, then $\Gamma_\rho \leqslant \rho$.*

Considering its relevance to the worst-case superadditivity of a risk measure, the mapping $\Gamma_\rho : \rho \mapsto \Gamma_\rho$ from the set of risk measures to itself is of our primary interest. Distortion risk measures, which are positively homogeneous and comonotonic additive, are treated in detail in Section 3; convex risk measures are discussed in Section 4. Explicit constructions of Γ_ρ for those classes of risk measures are given. In those examples, we observe that in addition to the properties of ρ, Γ_ρ very often "gains" more desirable properties, such as positive homogeneity and subadditivity. When Γ_ρ is subadditive, it becomes the smallest subadditive risk measure dominating ρ.

Corollary 2.3. *If ρ induces a coherent (subadditive) extreme-aggregation measure, and satisfies any of the assumptions of Proposition 2.1, then the smallest coherent (subadditive) risk measure dominating ρ exists and is Γ_ρ.*

Thus, a coherent Γ_ρ provides the closest conservative coherent correction to ρ. Note that in general it is not clear whether such a smallest coherent risk measure dominating ρ exists. Corollary 2.3 is of independent mathematical interest, not contingent on either the definition of a diversification ratio or an assumption of portfolio homogeneity.

3 Distortion risk measures

3.1 Preliminaries on distortion risk measures

For the whole of Section 3 we assume that $\mathcal{X}$ is the set of random variables in L^0 bounded from below, unless otherwise specified. This serves to avoid possibly undefined values of the risk measure and, in the present context, corresponds to having bounded gains.

A *distortion risk measure* $\rho_h : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$\rho_h(X_F) = \int_0^1 F^{-1}(t)\mathrm{d}h(t), \tag{3.1}$$

where h is an increasing, right-continuous and left-limit function, with $h(0) = h(0+) = 0$ and $h(1-) = h(1) = 1$. Equivalently, h is a distribution function supported in $(0,1)$[3]. We refer to h as a *distortion function* and $\delta := \mathrm{d}h/(\mathrm{d}t-)$, the left derivative of h, as a *distortion factor* if the derivative exists. As the most popular class of risk measures, distortion risk measures were introduced in insurance pricing by Wang et al. (1997) (see Deprez and Gerber, 1985; Yaari, 1987 and Denneberg, 1990 for early use of (3.1)) and in banking risk measurement by Acerbi (2002) (who focuses on the case of convex h) under the name *spectral risk measures*. Recently, VaR and the mean are shown to be the only elicitable distortion risk measures in Kou and Peng (2014). Sometimes (see e.g. Kusuoka, 2001), h is allowed to have probability mass on $\{0,1\}$; for example, $h(t) = \mathrm{I}_{\{t=1\}}$, $t \in [0,1]$ leads to $h(X) = \text{ess-sup}(X)$, the essential supremum. In this section, we exclude such special cases.

The family of distortion risk measures includes commonly used risk measures, such as VaR_p and ES_p defined in Section 2.1, with distortion functions $h(t) = \mathrm{I}_{\{t \geqslant p\}}$ and $h(t) = \mathrm{I}_{\{t \geqslant p\}}(t-p)/(1-p)$ respectively. In addition to those risk measures, we consider *Range-Value-at-Risk* (RVaR), introduced in Cont et al. (2010) as a robust alternative to ES. $\mathrm{RVaR}_{p,q}$, $p, q \in [0,1)$, $q > p$ is the distortion risk measure with $h(t) = \min\{\mathrm{I}_{\{t \geqslant p\}}(t-p)/(q-p), 1\}$, leading to

$$\mathrm{RVaR}_{p,q}(X_F) = \frac{1}{q-p}\int_p^q \mathrm{VaR}_r(X_F)\mathrm{d}r.$$

Any distortion risk measure ρ_h satisfies properties (a-c, f-h) in Section 2.1. It is shown that a risk measure satisfies (a, g, h) if and only if it is a distortion risk measure up to a scale; see Yaari (1987) and Schmeidler (1989). The risk measure ρ_h is subadditive (d) if and only if h is convex (δ is increasing); this dates back to Yaari (1987, Theorem 2, in the appearance of preserving convex order). A special role of distortion risk measures follows from the Kusuoka (2001) representation, showing that each (law-invariant) coherent risk measure on $\mathcal{X} = L^\infty$ can be written as the supremum over a class of coherent distortion risk measures.

We aim to characterize the extreme-aggregation measures induced by distortion risk measures. Since distortion risk measures are positively homogeneous, by Proposition 2.1, we have $\Gamma_\rho(X_F) = \sup_{n\in\mathbb{N}} \Gamma_{\rho,n}(X_F)$. Motivated by the discussion of Section 2.3, we seek the smallest coherent distortion risk measure ρ_h^+ such that $\rho_h^+ \geqslant \rho_h$. The existence of ρ_h^+ is guaranteed by the following lemma.

[3]A distortion risk measure is often defined as $\rho_h(X_F) = \int_{\mathbb{R}} x\mathrm{d}h(F(x))$ for h increasing but not necessarily right-continuous and left-limit. The two definitions are equivalent for all continuous random variables. In this paper we use (3.1) for its analytical convenience.

Lemma 3.1. *For a given distortion function h, the smallest coherent distortion risk measure ρ_h^+ that dominates ρ_h always exists, and is given by*

$$\rho_h^+(X_F) = \rho_{h^*}(X_F) = \int_0^1 F^{-1}(t)\mathrm{d}h^*(t),$$

where for $t \in [0,1]$,

$$h^*(t) = \sup\{g(t) : g : [0,1] \to [0,1],\ g \leqslant h,\ g \text{ is increasing, and convex on } [0,1]\}. \tag{3.2}$$

In what follows, h^* is defined by (3.2). For a given function h, finding h^* is equivalent to finding the convex hull of the set $\{(x,y) \in [0,1] \times \mathbb{R}^+ : h(x) \leqslant y\}$. Although an analytical formula for h^* may not be available, it can always be computed by approximation (see Section 1.1 in de Berg et al., 2008). Such h^* is referred to as the *largest convex minorant* of h, and it is the most cost-efficient path (in the sense of $||h'||_2$; see Hashorva and Mishura, 2014) from $(0,0)$ to $(1,1)$ dominated by h. Note that $\rho_h^+(X_F)$ is not guaranteed to be finite even if $\rho_h(X_F) < \infty$.

Proposition 2.1 implies that $\rho_h(X_F) \leqslant \Gamma_{\rho_h}(X_F) \leqslant \rho_h^+(X_F)$ for any distortion function h and a distribution F. In the following examples it is seen that for the VaR and RVaR risk measures it actually is $\Gamma_{\rho_h}(X_F) = \rho_h^+(X_F)$, such that in those cases Γ_{ρ_h} is coherent.

Example 3.1 (Value-at-Risk)**.** Let $h(t) = \mathrm{I}_{\{t \geqslant p\}}$, $t \in [0,1]$ for $p \in (0,1)$, such that $\rho_h = \mathrm{VaR}_p$. Then $h^*(t) = \mathrm{I}_{\{t \geqslant p\}}(t-p)/(1-p)$, $t \in [0,1]$, implying $\rho_h^+ = \mathrm{ES}_p$. We have that $\Gamma_{\mathrm{VaR}_p} = \mathrm{ES}_p$ Wang and Wang (see 2015, Corollary 3.7). Hence Γ_{VaR_p} is identified with the smallest dominating coherent distortion risk measure.

Example 3.2 (Range-Value-at-Risk)**.** Let δ, the distortion factor corresponding to h, be a step function

$$\delta(t) = \begin{cases} 0 & t \leqslant p, \\ a & p < t \leqslant q, \\ b & q < t \leqslant 1, \end{cases}$$

where $a > b \geqslant 0$ and such that $h(1) = \int_0^1 \delta(t)\mathrm{d}t = 1$. We can check that $h^*(t) = \mathrm{I}_{\{t \geqslant p\}}(x-p)/(1-p)$, $t \in [0,1]$, such that $\rho_h^+ = \mathrm{ES}_p$. To prove that $\Gamma_{\rho_h} = \mathrm{ES}_p$, observe that $h^*(t) \leqslant h(t) \leqslant \mathrm{I}_{\{t \geqslant p\}}$ and hence $\mathrm{ES}_p \geqslant \rho_h \geqslant \mathrm{VaR}_p$. It follows that $\Gamma_{\mathrm{ES}_p} \geqslant \Gamma_{\rho_h} \geqslant \Gamma_{\mathrm{VaR}_p}$ which leads to $\Gamma_{\rho_h} = \mathrm{ES}_p$ by Example 3.1. In particular, by choosing $b = 0$ and thus $\rho_h = \mathrm{RVaR}_{p,q}$, we have that for $p, q \in (0,1)$ and $q > p$, $\Gamma_{\mathrm{RVaR}_{p,q}} = \mathrm{ES}_p$.

3.2 Extreme-aggregation measures induced by distortion risk measures

Examples 3.1 and 3.2 suggest that for some classes of distortion risk measures the smallest dominating coherent risk measure is again a distortion risk measure and is identical to the extreme-aggregation risk measure. As the main result of this section, it is now shown that the same is true for all distortion risk measures.

Theorem 3.2. *The extreme-aggregation measure induced by any distortion risk measure ρ_h is the smallest coherent risk measure dominating ρ_h, and is given by $\Gamma_{\rho_h} = \rho_h^+$.*

Theorem 3.2 shows that ρ_h^+ characterizes the most superadditive behavior of ρ_h, by providing a sharp upper bound for the n-superadditivity ratios in (2.3). Furthermore, all distortion risk measures induce coherent extreme-aggregation measures which belong to the same class of distortion risk measures. This is a non-trivial conclusion, since the infimum over a set of coherent (resp. distortion) risk measures is not necessarily a coherent (resp. distortion) risk measure in general.

A different interpretation of Theorem 3.2 arises in the context of dependence uncertainty. Recall that one of the stated reasons for introducing non-coherent distortion risk measures is statistical robustness (Cont et al., 2010). If the choice of a non-coherent distortion risk measure involves a trade-off between subadditivity and statistical robustness, the benefits of such a trade-off fade in a portfolio context. Since in the context of dependence uncertainty the calculation of a coherent risk measure becomes necessary, comparisons of robustness among coherent risk measures, as those provided by Krätschmer et al. (2014), are relevant.

The "coherent correction" to the risk measure induced by moving from h to h^* entails the smallest possible reduction in robustness. Example 3.3 illustrates the derivation of such a risk measure, while Example 3.4 deals with the problem of best-case (dual) diversification scenarios.

Example 3.3 (Truncation of convex distortions)**.** Let h be a convex distortion function, representing a decision maker's preferences. The decision maker attempts to "robustify" the coherent risk measure ρ_h, by introducing for some q close to 1 the distortion function

$$g(t) = \begin{cases} h(t), & t \in [0, q) \\ 1, & t \in [q, 1], \end{cases}$$

leading to the risk measure

$$\rho_g(X_F) = \int_0^q \mathrm{VaR}_t(X_F)\mathrm{d}h(t) + (1 - h(q))\mathrm{VaR}_q(X_F).$$

Thus, percentiles with confidence levels beyond q are ignored and the corresponding weight is placed on $\mathrm{VaR}_q(X_F)$.

Now, in the presence of a homogeneous portfolio and dependence uncertainty, risk may be quantified by $\Gamma_{\rho_g} = \rho_g^+ = \rho_{g^*}$. By the convexity of h it follows that

$$g^*(t) = \begin{cases} h(t), & t \in [0, q) \\ h(q) + \big(t - q\big)\frac{1-h(q)}{1-q}, & t \in [q, 1], \end{cases}$$

leading to the risk measure

$$\rho_g^+(X_F) = \int_0^q \mathrm{VaR}_t(X_F)\mathrm{d}h(t) + (1 - h(q))\mathrm{ES}_q(X_F).$$

Thus, the extreme-aggregation risk measure resembles the original ρ_h, with the difference that percentiles with confidence levels beyond q receive a constant weight, leading to an ES-like quantification of extreme risk.

Example 3.4 (Dual bound and best-case scenarios)**.** Let $-\infty < F^{-1}(0) \leqslant F^{-1}(1) < \infty$ and, for simplicity, both F and h be continuous. The relation $\rho_h(X_F) = -\rho_{\bar{h}}(-X_F)$ holds, where $\bar{h}(t) = 1 - h(1-t)$ is the *conjugate distortion* of h. Let $\tilde{F}$ be the distribution of $-X_F$. It follows that the best-case diversification (*least superadditive*) scenario can be quantified by

$$\begin{aligned}\inf_{n\in\mathbb{N}}\left\{\frac{1}{n}\inf_{S\in\mathfrak{S}_n(F)}\rho_h(S)\right\} &= -\sup_{n\in\mathbb{N}}\left\{\frac{1}{n}\sup_{\tilde{S}\in\mathfrak{S}_n(\tilde{F})}\rho_{\bar{h}}(\tilde{S})\right\}\\ &= -\Gamma_{\rho_{\bar{h}}}(X_{\tilde{F}}) = -\rho_{\bar{h}^*}(X_{\tilde{F}}) = \rho_{h_*}(X_F),\end{aligned}$$

where h_* is the conjugate distortion of $\bar{h}^*$ and, by symmetry, is the smallest concave distortion dominating h. This argument generalizes the known best-case VaR bounds; see Embrechts et al. (2014) for a relevant discussion.

As an example, let $\rho_h = \mathrm{RVaR}_{p,q}$, such that $h(t) = \min\{\mathrm{I}_{\{t\geqslant p\}}(t-p)/(q-p), 1\}$, implying

$$\bar{h}(t) = \min\left\{\mathrm{I}_{\{t\geqslant 1-q\}}\frac{t-(1-q)}{q-p}, 1\right\},\quad \bar{h}^*(t) = \mathrm{I}_{\{t\geqslant 1-q\}}\frac{t-(1-q)}{q},\quad h_*(t) = \min\left\{\mathrm{I}_{\{t\leqslant q\}}\frac{t}{q}, 1\right\}.$$

Hence the best-case dependence scenario is characterized by the *superadditive* risk measure

$$\inf_{n\in\mathbb{N}}\left\{\frac{1}{n}\inf_{S\in\mathfrak{S}_n(F)}\rho_h(S)\right\} = \frac{1}{q}\int_0^q \mathrm{VaR}_t(X_F)\mathrm{d}t,$$

sometimes referred to as the *Left-Tail-VaR*.

Remark 3.1. Theorem 3.2 can be easily extended to *generalized distortion risk measures*, defined as

$$\rho_A^G := \sup_{h\in A}\rho_h, \tag{3.3}$$

where A is a set of distortion functions. The extreme-aggregation measure induced by any generalized distortion risk measure ρ_A^G is the smallest coherent risk measure dominating ρ_A^G, and is given by

$$\begin{aligned}\Gamma_{\rho_A^G}(X_F) &= \sup_{n\in\mathbb{N}}\frac{1}{n}\sup\left\{\sup_{h\in A}\rho_h(S) : S\in\mathfrak{S}_n(F)\right\}\\ &= \sup_{n\in\mathbb{N}}\sup_{S\in\mathfrak{S}_n(F)}\sup_{h\in A}\left\{\frac{1}{n}\rho_h(S)\right\}\\ &= \sup_{h\in A}\sup_{n\in\mathbb{N}}\sup_{S\in\mathfrak{S}_n(F)}\left\{\frac{1}{n}\rho_h(S)\right\} = \sup_{h\in A}\rho_h^+(X_F).\end{aligned}$$

Note that (3.3) includes all law-invariant coherent risk measures with the Fatou property (see Kusuoka, 2001; Delbaen, 2012).

3.3 Asymptotic equivalence between worst-case risk measures

Theorems 3.2 and Remark 3.1 imply an asymptotic equivalence between a non-coherent risk measure and its coherent counterpart, under the worst-case scenario of dependence uncertainty. Here we extend the discussion to an inhomogeneous portfolio. For a sequence of distributions $\mathbf{F} = \{F_i,\ i \in \mathbb{N}\}$, we denote

$$\mathfrak{S}_n(\mathbf{F}) := \{X_1 + \cdots + X_n : X_i \sim F_i,\ i = 1, \ldots, n\}.$$

The quantity $\sup_{S \in \mathfrak{S}_n(\mathbf{F})} \rho_A^G(S)$ is the worst-case risk measure ρ_A^G under dependence uncertainty. We establish the asymptotic equivalence for generalized distortion risk measures in the following theorem, based on results in Theorem 3.2 and Remark 3.1.

Theorem 3.3. *Let A be a finite set of distortion functions. Assume that there are only finitely many different distributions in the sequence $\{F_i,\ i \in \mathbb{N}\}$, and $0 < \rho_A^{G+}(X_{F_i}) < \infty$ for each $i \in \mathbb{N}$. Then, as $n \to \infty$,*

$$\frac{\sup_{S \in \mathfrak{S}_n(\mathbf{F})} \rho_A^G(S)}{\sup_{S \in \mathfrak{S}_n(\mathbf{F})} \rho_A^{G+}(S)} \to 1. \tag{3.4}$$

From the perspective of risk management, Theorem 3.3 indicates that when assessing capital conservatively under dependence uncertainty, using ρ_A^G and using ρ_A^{G+} would give roughly the same capital estimates. This suggests that in situations where information on the dependence structure is unavailable, a conservative regulation principle would take the information of the coherent risk measure ρ_A^{G+} into account for quantifying risk aggregation. This ratio is close to 1 even for small numbers n; see Embrechts et al. (2014) for numerics in the case of VaR and ES.

Remark 3.2. The equivalence (3.4) for VaR and ES was studied under various different conditions on the marginal distributions; see Embrechts et al. (2014, 2015) and the references therein.

In the case of infinitely many elements in A, a uniform convergence for $(n, h) \in \mathbb{N} \times A$ is required in the proof of the above theorem for the same equivalence to hold.

4 Convex risk measures

4.1 Extreme-aggregation measures induced by convex risk measures

Convex risk measures, satisfying properties (a), (b), and (e) in Section 2.1, are discussed in detail by Föllmer and Schied (2011, Chapter 4). Since the canonical domain for convex risk measures is L^1 (Filipović and Svindland, 2012), we assume $\mathcal{X} = L^p$, $p \in [1, \infty]$ in this section. Recall that we assume that risk measures are law-invariant throughout.

For any convex risk measure that also satisfies zero-normalization (f), it can be shown that the induced extreme-aggregation measure is once more the smallest dominating coherent risk measure.

Theorem 4.1. *The extreme-aggregation measure induced by any convex risk measure ρ with $\rho(0) = 0$ is the smallest coherent risk measure dominating ρ.*

More can be said when the risk measure ρ satisfies the Fatou property. A law-invariant risk measure ρ on L^p, $p \in [1, \infty)$ satisfies the Fatou property (FP), if

(FP) *Fatou*: $\liminf_{n\to\infty} \rho(X_n) \geqslant \rho(X)$ if $X, X_1, X_2, \dots \in L^p$, $X_n \xrightarrow{L^p} X$ as $n \to \infty$.

A law-invariant convex risk measure on L^p, $p \in [1, \infty)$ satisfying the Fatou property has a dual representation (also called a *robust representation* in Föllmer and Schied, 2011)

$$\rho = \sup_{\mu \in \mathcal{P}} \left\{ \int_0^1 \mathrm{ES}_\alpha \mathrm{d}\mu(\alpha) - v(\mu) \right\}, \tag{4.1}$$

where $\mathcal{P}$ is the set of all probability measures on $[0, 1]$, and $v : \mathcal{P} \to \mathbb{R} \cup \{+\infty\}$ is a *penalty function* of ρ. The representation (4.1) was established in Frittelli and Rosazza Gianin (2005) for convex risk measures on L^∞ (in that case, the L^∞-Fatou property is always guaranteed on an atomless probability space, see Jouini et al., 2006 and Section 5.1 of Delbaen, 2012); for the case of L^p see Svindland (2009, Lemma 2.14). For such risk measures, we have the following characterization for the extreme-aggregation measure of ρ.

Corollary 4.2. *Suppose ρ is a convex risk measure on L^p, $p \in [1, \infty]$ with the Fatou property and a penalty function v in (4.1), then*

$$\Gamma_\rho = \sup_{\mu \in \mathcal{P}_v} \int_0^1 \mathrm{ES}_\alpha \mathrm{d}\mu(\alpha), \tag{4.2}$$

where $\mathcal{P}_v = \{\mu \in \mathcal{P} : v(\mu) < +\infty\}$.

Note that Γ_ρ in (4.2) is a coherent risk measure with the Fatou property. In particular, the robust representation of Γ_ρ in (4.2) reflects directly the corresponding representation of ρ in (4.1), dispensing with the penalty function and focusing on those probability measures that are potentially used in the calculation of ρ.

4.2 Extreme-aggregation measures induced by shortfall risk measures

In this section we focus on risk measures that are derived via loss functions. Here, a *loss function* $\ell : \mathbb{R} \to \mathbb{R}$ is an increasing and convex function that is not identically constant. Following Föllmer and Schied (2011), a *shortfall risk measure* is defined as

$$\tau_{\ell,x_0}(X) = \inf\{x \in \mathbb{R} : \mathbb{E}(\ell(X - x)) \leqslant x_0\}, \quad X \in L^1, \tag{4.3}$$

where x_0 is an interior point of the range of ℓ. Shortfall risk measures belong to the class of convex risk measures. Without essential loss of generality we may assume $x_0 = \ell(0)$ such that τ_{ℓ,x_0} satisfies zero-normalization; in that case we write $\tau_\ell := \tau_{\ell,\ell(0)}$.

The p-expectile, denoted by e_p, is a risk measure of type (4.3) defined for $x_0 = 0$ and the loss function $\ell_p(x) = px_+ - (1-p)x_-$, $p \in (0,1)$, where $x_+ := \max(x,0)$ and $x_- := \max(-x,0)$, $x \in \mathbb{R}$. Equivalently, it satisfies (Newey and Powell, 1987)

$$p\mathbb{E}[(X - e_p(X))_+] = (1-p)\mathbb{E}[(X - e_p(X))_-], \quad X \in L^1.$$

For the purposes of this paper, we extend the definition of expectiles by defining $e_1(X) = \text{ess-sup}(X)$ and $e_0(X) = \text{ess-inf}(X)$. The p-expectile is evidently positively homogeneous. For $p \geqslant 1/2$ the loss function ℓ_p is convex, such that e_p is a shortfall risk measure and is also coherent (Bellini et al., 2014). In fact, $\{e_p,\ p \geqslant 1/2\}$ is the only class of coherent shortfall risk measures (Weber, 2006) and the only class of elicitable coherent risk measures (Delbaen et al., 2015)[4].

Expectiles play a special role in the construction of extreme-aggregation measures induced by shortfall risk measures. Since ℓ is convex and not always a constant, we know that $a_\ell := \lim_{x\to\infty} \ell'_+(x)$ exists in $(0,\infty]$, $b_\ell := \lim_{x\to-\infty} \ell'_+(x)$ exists in $[0,\infty)$, and $b_\ell \leqslant a_\ell$. For each ℓ, define the loss function

$$\ell^*(x) = a_\ell x_+ - b_\ell x_-.$$

By $b_\ell \leqslant a_\ell$, ℓ^* is a convex loss function, derived from ℓ, and giving rise to a coherent risk measure via (4.3). Thus the risk measure τ_{ℓ^*} is a coherent expectile, with $\tau_{\ell^*}(X) = e_{p_\ell}(X)$ for $p_\ell = a_\ell/(a_\ell + b_\ell) \geqslant 1/2$.

Note the analogy with the definition of h^* in Section 3. Indeed, the extreme-aggregation measure induced by a convex shortfall risk measure is an expectile.

Proposition 4.3. *The extreme-aggregation measure induced by any shortfall risk measure τ_{ℓ,x_0} is the smallest coherent expectile dominating τ_{ℓ,x_0}, and is given by $\Gamma_{\tau_{\ell,x_0}} = \tau_{\ell^*} = e_{p_\ell}$.*

For many loss functions used in practice it may be $a_\ell = \infty$ or $b_\ell = 0$. In that case the extreme-aggregation measure is the essential supremum. This implies that for many models used in practice it may be $\Gamma_{\tau_\ell}(X) = \infty$, showing that risk aggregations lead to an explosion of portfolio risk. An example is the entropic risk measure defined in (2.1).

5 Conclusions

We examine the superadditivity properties of general classes of risk measures, corresponding to worst-case diversification scenarios. The introduction of extreme-aggregation measures Γ_ρ allows

[4]See also Ziegel (2014), Bellini and Bignozzi (2014) and Kou and Peng (2014) for characterization of elicitable risk measures.

a systematic study of the behavior of a risk measure ρ "at its most superadditive" and the quantification of worst-case portfolio capital requirements under dependence uncertainty. Furthermore, extreme-aggregation measures, when coherent, allow us to construct for general risk measures their smallest dominating coherent risk measures.

Explicit forms of extreme-aggregation measures are obtained for distortion and convex risk measures and in both cases the induced extreme-aggregation measures are coherent. The main theoretical results in this paper suggest that an extreme-aggregation measure inherits all key properties of a risk measure, and in addition, often "gains" positive homogeneity as well as convexity or subadditivity.

When capital is set using a non-subadditive risk measure ρ, whatever the motivation for this choice, the extreme-coherence of ρ implies that a coherent risk measure Γ_ρ needs to be considered in order to quantify portfolio risk under dependence uncertainty. This is further evidence in favor of coherence (subadditivity, in particular) as a desirable property for risk measures.

A Proofs

Proof of Proposition 2.1

Proof. To prove (a) and (b), first we deal with the case that ρ is positively homogeneous (i) or comonotonic additive (ii).

(a) From the definition of $\Gamma_{\rho,n}$, the inequality $\Gamma_{\rho,n}(X_F) \leqslant \Gamma_{\rho,kn}(X_F)$ is equivalent to

$$\sup\{\rho(S) : S \in \mathfrak{S}_{kn}(F)\} \geqslant k \sup\{\rho(R) : R \in \mathfrak{S}_n(F)\}.$$

Note that for all $R \in \mathfrak{S}_n(F)$, we have $kR \in \mathfrak{S}_{kn}(F)$. Thus

$$\sup\{\rho(S) : S \in \mathfrak{S}_{kn}(F)\} \geqslant \rho(kR) = k\rho(R), \qquad \forall R \in \mathfrak{S}_n(F),$$

where the last equality is implied by that ρ is comonotonic additive or positively homogeneous. In particular this holds for $\sup\{\rho(R) : R \in \mathfrak{S}_n(F)\}$, from which the result follows.

(b) By (a) we can see that for a fixed $m \in \mathbb{N}$ we have that $\Gamma_{\rho,m}(X_F) \leqslant \Gamma_{\rho,km}(X_F)$ for all $k \in \mathbb{N}$, and this directly implies

$$\sup_{m\in\mathbb{N}} \Gamma_{\rho,m}(X_F) \leqslant \limsup_{n\to\infty} \Gamma_{\rho,n}(X_F). \tag{A.1}$$

The opposite inequality follows immediately and thus (A.1) is an equality.

In the following we assume ρ is comonotonic additive. Let $k, m, n \in \mathbb{N}$ and $k = \lfloor \frac{n}{m} \rfloor$ be the inferior integer part of n/m so that $n \geqslant km$. Then

$$\begin{aligned}
\Gamma_{\rho,n}(X_F) &= \frac{km}{n} \frac{\sup\{\rho(S) : S \in \mathfrak{S}_n(F)\}}{km} \\
&\geqslant \frac{km}{n} \frac{\sup\{\rho(S) : S \in \mathfrak{S}_{km}(F)\} + (n-km)\rho(X_F)}{km} \\
&\geqslant \frac{km}{n} \frac{\sup\{\rho(S) : S \in \mathfrak{S}_m(F)\}}{m} + \frac{(n-km)\rho(X_F)}{n} \\
&= \frac{km}{n} \Gamma_{\rho,m}(X_F) + \frac{(n-km)\rho(X_F)}{n},
\end{aligned}$$

where in the last inequality we used part (a). It follows that for m fixed,

$$\liminf_{n\to\infty} \Gamma_{\rho,n}(X_F) \geqslant \lim_{n\to\infty} \frac{km}{n} \Gamma_{\rho,m}(X_F) + \lim_{n\to\infty} \frac{n-km}{n} \rho(X_F) = \Gamma_{\rho,m}(X_F).$$

By taking supremum over m, we obtain

$$\liminf_{n\to\infty} \Gamma_{\rho,n}(X_F) \geqslant \sup_{m\in\mathbb{N}} \Gamma_{\rho,m}(X_F).$$

The opposite inequality follows immediately and we get

$$\lim_{n\to\infty} \Gamma_{\rho,n}(X_F) = \sup_{n\in\mathbb{N}} \Gamma_{\rho,n}(X_F).$$

Now consider the case (iii) where ρ is convex. Then for any $X_1, \dots, X_n \sim F$, it is

$$\rho(X_1 + \cdots + X_n) = \rho\left(\frac{1}{n} nX_1 + \cdots + \frac{1}{n} nX_n\right) \leqslant \frac{1}{n}\rho(nX_1) + \cdots + \frac{1}{n}\rho(nX_n) = \rho(nX_F).$$

Consequently

$$\rho(nX_F) = \sup_{S\in\mathfrak{S}_n(F)} \rho(S) \implies \Gamma_{\rho,n}(X_F) = \frac{1}{n}\rho(nX_F).$$

By the convexity of ρ and the assumption $\rho(0) = 0$, $\Gamma_{\rho,n}(X_F)$ is increasing in n, which implies (a) and (b).

(c) For any $n \in \mathbb{N}$, $\Gamma_{\rho,n}(X_F) \leqslant \frac{1}{n}\sup\{\rho^+(S) : S \in \mathfrak{S}_n(F)\} \leqslant \frac{1}{n} n\rho^+(X_F) = \rho^+(X_F)$.

□

Proof of Lemma 2.2

Proof. The inheritance of (a-c, f) is immediate. For subadditivity (d) let X, Y be two random variables with distributions F and G respectively, then

$$\begin{aligned}
\Gamma_{\rho,n}(X+Y) &\leqslant \frac{1}{n} \sup\{\rho(T+R) : T \in \mathfrak{S}_n(F),\ R \in \mathfrak{S}_n(G)\} \\
&\leqslant \frac{1}{n} \sup\{\rho(T) : T \in \mathfrak{S}_n(F)\} + \frac{1}{n} \sup\{\rho(R) : R \in \mathfrak{S}_n(G)\} \\
&= \Gamma_{\rho,n}(X) + \Gamma_{\rho,n}(Y).
\end{aligned}$$

By taking an upper limit as $n \to \infty$ on both sides, we obtain that Γ_ρ is subadditive. For convexity (f) a similar argument applies. Moreover, it follows from Proposition 2.1 (a) that if ρ is (i) positively homogeneous or (ii) comonotonic additive or (iii) convex and zero-normalized, then $\Gamma_\rho \geqslant \rho$. $\square$

Proof of Lemma 3.1

Before giving the proof of Lemma 3.1, we introduce a useful lemma, which characterizes an ordering of distortion risk measures in terms of their distortion functions. The lemma will be used repeatedly in proofs of later results.

Lemma A.1. *(a) For a distortion risk measure ρ_h and every F such that $\rho_h(X_F) < \infty$,*

$$\rho_h(X_F) = \int_0^1 F^{-1}(t)\mathrm{d}h(t) = F^{-1}(0) + \int_0^1 (1-h(t))\mathrm{d}F^{-1}(t).$$

(b) For two distortion risk measures ρ_{h_1}, ρ_{h_2},

$$h_1(t) \leqslant h_2(t) \ \text{ for all } t \in [0,1] \quad \Leftrightarrow \quad \rho_{h_1}(X_F) \geqslant \rho_{h_2}(X_F) \ \text{ for all distributions } F.$$

Proof. (a) By integration by parts,

$$\begin{aligned}
\int_0^1 F^{-1}(t)\mathrm{d}h(t) &= \int_0^1 F^{-1}(t)\mathrm{d}(-(1-h(t))) \\
&= -F^{-1}(t)(1-h(t))\big|_0^1 + \int_0^1 (1-h(t))\mathrm{d}F^{-1}(t) \\
&= -F^{-1}(t)(1-h(t))\big|_{t\to 1^-} + F^{-1}(0) + \int_0^1 (1-h(t))\mathrm{d}F^{-1}(t).
\end{aligned}$$

To prove that the first term $F^{-1}(t)(1-h(t))$ tends to 0 as $t \to 1$, note that

$$\rho_h(X_F) = \int_0^1 F^{-1}(t)\mathrm{d}h(t) = \mathbb{E}[F^{-1}(Y)] < \infty, \tag{A.2}$$

where Y is a random variable with probability distribution h. From (A.2) it follows that

$$\lim_{u\to 1^-} F^{-1}(u)\mathbb{P}(Y > u) = 0.$$

Note that $\mathbb{P}(Y > u) = 1 - h(u)$ which gives the result.

(b) It follows immediately from the definition of distortion risk measure given in (a).

$\square$

Proof of Lemma 3.1. The set

$$H_h := \{g : [0,1] \to [0,1],\ g \leqslant h,\ g \text{ is increasing, and convex on [0,1]}\} \tag{A.3}$$

is not empty, since $g(t) := 0,\ t \in [0,1]$ is an element of H_h. Also the supremum is finite since everything in H_h is bounded above by h. Hence $h^* = \sup\{g \in H_h\}$ is a well-defined function. It is easy to verify that $h^* \leqslant h$. The supremum of increasing functions is increasing, thus h^* is increasing. Further, because $g(0+) \leqslant h(0+) = 0$ for any $g \in H_h$, $h^*(0+) = 0$. We only need to prove that $h^*(1-) = 1$. For any $\epsilon > 0$, let $y := \inf\{x\ :\ h(1-x) \geqslant 1-\epsilon\}$, note that $y > 0$, and define

$$g_\epsilon(t) = \begin{cases} 0 & \text{if } t \in [0, 1-y) \\ \frac{(t-(1-y))(1-\epsilon)}{y} & \text{if } t \in [1-y, 1], \end{cases}$$

so that $g_\epsilon(1-y) = 0$ and $g_\epsilon(1) = 1-\epsilon$. It is clear that $g_\epsilon \leqslant h$, $g_\epsilon \in H_h$ and $g_\epsilon \leqslant h^*$ for any $\epsilon > 0$. In particular,

$$h^*(1-) = \lim_{x\to 0^+} h^*(1-x) \geqslant \sup_\epsilon \lim_{x\to 0^+} g_\epsilon(1-x) = \sup_\epsilon (1-\epsilon) = 1.$$

It follows that h^* is a distortion function. Thus from Lemma A.1 (b), $\rho_{h^*} \geqslant \rho_h$. Since the supremum of convex functions is still convex, h^* is convex and thus ρ_{h^*} is coherent.

Suppose there is another coherent distortion risk measure ρ_{h_0} such that $\rho_h \leqslant \rho_{h_0}$. Always from Lemma A.1 (b) it follows that $h \geqslant h_0$. Hence $h_0 \in H_h$ and $h_0 \leqslant h^*$ by definition. Thus $\rho_{h_0} \geqslant \rho_{h^*}$. That is, ρ_{h^*} is the smallest coherent distortion risk measure that dominates ρ_h. □

Proof of Theorem 3.2

In the following we report the detailed proof of Theorem 3.2. For the ease of presentation we slightly abuse the notation. We use $\rho_h(X) = \rho_\delta(X)$ to represent the risk measure with distortion factor δ and distortion function h. Both are convenient at different places.

We start with the case when δ is a step function with a finite number (m) of steps. In the following we make this assumption throughout. To be more specific, let m be a positive integer, $0 = b_0 < b_1 < \cdots < b_m = 1$ be a partition of $[0,1]$, and $a_1, \ldots, a_m$ be non-negative numbers, with $a_i \neq a_{i+1}$ for $i = 1, \ldots, m-1$. We suppose that δ has the following form: for $i = 1, \ldots, m$,

$$\delta(t) = a_i,\ \ t \in [b_{i-1}, b_i)\,,$$

and in addition, $\delta(1) = a_m$. It is obvious that for each step function δ, the values of m, $\{b_i\}_{i=0}^m$ and $\{a_i\}_{i=1}^m$ are uniquely determined.

We define the *incoherence index*

$$\#(\delta) = \sum_{i=1}^{m-1} \mathrm{I}_{\{a_{i+1} < a_i\}}.$$

If $a_{i+1} < a_i$ for some i, we say δ is *incoherent* at the i-th step. Suppose δ is a step function and let K be its largest incoherent step, i.e. $\delta(t) = a_{K+1} < a_K = \delta(s)$ for $t \in [b_K, b_{K+1})$, $s \in [b_{K-1}, b_K)$, and δ is increasing on $[b_K, 1]$. Define the operator $\mathcal{L}$ on δ as follows: if $\#(\delta) \geqslant 1$, then

$$\mathcal{L}\delta(t) = \begin{cases} \delta(t) & t \notin [b_{K-1}, b_{K+1}),\ t \in [0,1], \\ \frac{b_K - b_{K-1}}{b_{K+1}-b_{K-1}} a_K + \frac{b_{K+1}-b_K}{b_{K+1}-b_{K-1}} a_{K+1} & t \in [b_{K-1}, b_{K+1}). \end{cases}$$

Since

$$\int_{b_{K-1}}^{b_{K+1}} \mathcal{L}\delta(t)\mathrm{d}t = (b_K - b_{K-1})a_K + (b_{K+1} - b_K)a_{K+1} = \int_{b_{K-1}}^{b_{K+1}} \delta(t)\mathrm{d}t,$$

we have that $\mathcal{L}\delta$ is still a distortion factor. If $\#(\delta) = 0$ (i.e. δ does not have an incoherent step; δ is increasing), let $\mathcal{L}\delta = \delta$. Lemma A.1 (b) implies that $\rho_{\mathcal{L}\delta} \geqslant \rho_\delta$ by noting that $h_{\mathcal{L}\delta}(t) \leqslant h(t)$.

Lemma A.2. *Suppose that δ is a step function, and F has a bounded support. Then*

$$\lim_{n\to\infty} \left(\sup\left\{ \rho_{\mathcal{L}\delta}\left(\frac{S}{n}\right) : S \in \mathfrak{S}_n(F) \right\} - \sup\left\{ \rho_\delta\left(\frac{S}{n}\right) : S \in \mathfrak{S}_n(F) \right\} \right) = 0. \tag{A.4}$$

Proof. First, note that for any $S \in \mathfrak{S}_n(F)$, there exists $R \in \mathfrak{S}_n(F)$, $R \overset{\mathrm{d}}{=} S$ such that R can be written as $R = X_1 + \cdots + X_n$ where $X_i \sim F$, $i = 1, \ldots, n$, and $(X_1, \ldots, X_n)$ is exchangeable. This can be seen from the fact that $\mathfrak{F}_n(F) := \{\text{distribution function of } (Y_1, \ldots, Y_n) : Y_i \sim F,\ i = 1, \ldots, n\}$ is a convex set. Hence, for $S = Y_1 + \cdots + Y_n$, $Y_i \sim F$, $i = 1, \ldots, n$, one can always take the average of the distribution functions of all permutations of $(Y_1, \ldots, Y_n)$ to obtain an exchangeable distribution $\mathbf{F}$. Let $(X_1, \ldots, X_n) \sim \mathbf{F}$ then we have that $X_i \sim F$, $i = 1, \ldots, n$ and $R := X_1 + \cdots + X_n \overset{\mathrm{d}}{=} Y_1 + \ldots, Y_n = S$.

Denote

$$M = F^{-1}(1) < \infty, \quad \text{and} \quad r_n = \sup\left\{ \rho_{\mathcal{L}\delta}\left(\frac{S}{n}\right) : S \in \mathfrak{S}_n(F) \right\}. \tag{A.5}$$

Note that $\sup\{\rho_\delta(S/n) : S \in \mathfrak{S}_n(F)\} \leqslant r_n < \infty$ since $\rho_{\mathcal{L}\delta} \geqslant \rho_\delta$ and F has bounded support. By definition of r_n, for any $\epsilon > 0$, there exists $R \in \mathfrak{S}_n(F)$ such that $\rho_{\mathcal{L}\delta}(R/n) \geqslant r_n - \epsilon$ and one can write $R = X_1 + \cdots + X_n$ where $X_i \sim F$, $i = 1, \ldots, n$, and $(X_1, \ldots, X_n)$ is exchangeable. Let F_R be the distribution function of R. Define the random event

$$A = \{R \in [F_R^{-1}(b_{K-1}), F_R^{-1}(b_{K+1}))\}. \tag{A.6}$$

In the following we discuss different cases:

(a) Suppose $\mathbb{P}(A) = 0$, then $F_R^{-1}(b_{K-1}) = F_R^{-1}(b_{K+1})$ and

$$\begin{aligned} \rho_{\mathcal{L}\delta}(R) - \rho_\delta(R) &= \int_0^1 F_R^{-1}(t)(\mathcal{L}\delta(t) - \delta(t))\mathrm{d}t \\ &= \int_{b_{K-1}}^{b_{K+1}} F_R^{-1}(t)(\mathcal{L}\delta(t) - \delta(t))\mathrm{d}t \\ &= F_R^{-1}(b_{K-1}) \int_{b_{K-1}}^{b_{K+1}} (\mathcal{L}\delta(t) - \delta(t))\mathrm{d}t = 0. \end{aligned}$$

(b) Suppose $\mathbb{P}(A) > 0$. Let F_A be the distribution function of $X_1|A$. Since the distribution of $(X_1, \dots, X_n)$ is exchangeable, F_A is also the distribution function of $X_i|A$, $i = 2, \dots, n$. We can calculate the mean of F_A:

$$\mathbb{E}[X_1|A] = \frac{1}{n}\mathbb{E}[R|A]. \tag{A.7}$$

(b1) Suppose that $\mathbb{E}[R|A] = F_R^{-1}(b_{K-1})$ or $\mathbb{E}[R|A] = F_R^{-1}(b_{K+1})$. Then R is a constant on A, and

$$\rho_{\mathcal{L}\delta}(R) - \rho_\delta(R) = \mathbb{E}[R|A] \int_{b_{K-1}}^{b_{K+1}} (\mathcal{L}\delta(t) - \delta(t))\mathrm{d}t = 0.$$

(b2) Suppose that $F_R^{-1}(b_{K-1}) < \mathbb{E}[R|A] < F_R^{-1}(b_{K+1})$. Denote for $i = 1, \dots, n$,

$$W_i = X_i(1 - \mathrm{I}_A) + \frac{1}{n}\mathbb{E}[R|A]\mathrm{I}_A,$$

and $T = W_1 + \dots + W_n$. Denote by F_T the distribution function of T. It is easy to see that $T = R$ a.s. on A^c, and since we assume $F_R^{-1}(b_{K-1}) < \mathbb{E}[R|A] < F_R^{-1}(b_{K+1})$, $F_T^{-1}(t) = F_R^{-1}(t)$ for $t \notin [b_{K-1}, b_{K+1})$.

We can check that

$$\begin{aligned}
&\rho_{\mathcal{L}\delta}(R) - \rho_\delta(T)\\
&= \int_0^1 (F_R^{-1}(t)\mathcal{L}\delta(t) - F_T^{-1}(t)\delta(t))\mathrm{d}t\\
&= \int_{b_{K-1}}^{b_{K+1}} (F_R^{-1}(t)\mathcal{L}\delta(t) - F_T^{-1}(t)\delta(t))\mathrm{d}t\\
&= \left(\frac{b_K - b_{K-1}}{b_{K+1} - b_{K-1}}a_K + \frac{b_{K+1} - b_K}{b_{K+1} - b_{K-1}}a_{K+1}\right) \int_{b_{K-1}}^{b_{K+1}} F_R^{-1}(t)\mathrm{d}t - \mathbb{E}[R|A] \int_{b_{K-1}}^{b_{K+1}} \delta(t)\mathrm{d}t\\
&= ((b_K - b_{K-1})a_K + (b_{K+1} - b_K)a_{K+1})\mathbb{E}[R|A] - \mathbb{E}[R|A]((b_K - b_{K-1})a_K + (b_{K+1} - b_K)a_{K+1})\\
&= 0.
\end{aligned}$$

To continue analyzing the case (b2), we will use the following lemma.

Lemma A.3 (Corollary 3.1 of Wang and Wang, 2015)**.** *Suppose G is any distribution. If the support of G is contained in $[a, b]$, $a < b$, $a, b \in \mathbb{R}$, then there exists $T \in \mathfrak{S}_n(G)$ such that $|T - \mathbb{E}[T]| \leqslant b - a$.*

In the following we use M given in (A.5) and A given in (A.6). Lemma A.3 tells us that there exist random variables $Y_1, \dots, Y_n$ with a common distribution F_A such that

$$|Y_1 + \dots + Y_n - \mathbb{E}[Y_1 + \dots + Y_n]| \leqslant M. \tag{A.8}$$

We choose $Y_1, \dots, Y_n$ such that they are independent of A (this is always possible since (A.8) only concerns the distribution of $(Y_1, \dots, Y_n)$). Note that by (A.7), $\mathbb{E}[Y_1 + \cdots + Y_n] = \mathbb{E}[R|A]$. Hence

$$|Y_1 + \cdots + Y_n - \mathbb{E}[R|A]| \leqslant M.$$

Denote for $i = 1, \dots, n$,

$$Z_i = X_i(1 - \mathrm{I}_A) + Y_i \mathrm{I}_A.$$

For $x \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{P}(Z_i \leqslant x) &= \mathbb{P}(X_i \leqslant x, A^c) + \mathbb{P}(Y_i \leqslant x, A) \\
&= \mathbb{P}(X_i \leqslant x, A^c) + \mathbb{P}(Y_i \leqslant x)\mathbb{P}(A) \\
&= \mathbb{P}(X_i \leqslant x, A^c) + \mathbb{P}(X_i \leqslant x|A)\mathbb{P}(A) \\
&= \mathbb{P}(X_i \leqslant x) = F(x).
\end{aligned}$$

Thus, $Z_i \sim F$ for $i = 1, \dots, n$ and $\hat{R} := Z_1 + \cdots + Z_n$ is in $\mathfrak{S}_n(F)$. It is easy to see that

$$|\hat{R} - T| = |Y_1 + \cdots + Y_n - \mathbb{E}[R|A]|\mathrm{I}_A \leqslant M.$$

Finally, $|\rho_\delta(T) - \rho_\delta(\hat{R})| \leqslant M$ since monetary risk measures are Lipschitz continuous. It follows that $|\rho_{\mathcal{L}\delta}(R) - \rho_\delta(\hat{R})| \leqslant M$.

We have either (a, b1), $|\rho_{\mathcal{L}\delta}(R) - \rho_\delta(R)| = 0$ or (b2), $|\rho_{\mathcal{L}\delta}(R) - \rho_\delta(\hat{R})| \leqslant M$. Recall that $\rho_{\mathcal{L}\delta}(R/n) \geqslant r_n - \epsilon$. Hence, we have that in both cases,

$$\sup\left\{\rho_\delta\left(\frac{S}{n}\right) : S \in \mathfrak{S}_n(F)\right\} \geqslant \rho_{\mathcal{L}\delta}\left(\frac{R}{n}\right) - \frac{M}{n} \geqslant r_n - \epsilon - \frac{M}{n}.$$

Since $\epsilon > 0$ is arbitrary, we conclude that

$$r_n \geqslant \sup\left\{\rho_\delta\left(\frac{S}{n}\right) : S \in \mathfrak{S}_n(F)\right\} \geqslant r_n - \frac{M}{n},$$

and Lemma (A.2) follows. □

Lemma A.4. *Suppose that the distortion factor δ is a step function with m steps and F has a bounded support. We have that*

$$\lim_{n\to\infty} \frac{1}{n} \sup\left\{\rho_\delta(S) : S \in \mathfrak{S}_n(F)\right\} = \rho_{\mathcal{L}^m\delta}(X_F),$$

and moreover, $\rho_{\mathcal{L}^m\delta} = \rho_\delta^+$.

Proof. Since the support of F is bounded, we have $\rho_{\mathcal{L}^k\delta}(X_F) < \infty$ for all $k \in \mathbb{N}$. By applying Lemma A.2 iteratively, we have

$$\lim_{n\to\infty}\left(\sup\left\{\rho_{\mathcal{L}^k\delta}\left(\frac{S}{n}\right): S \in \mathfrak{S}_n(F)\right\} - \sup\left\{\rho_\delta\left(\frac{S}{n}\right): S \in \mathfrak{S}_n(F)\right\}\right) = 0$$

for any $k \in \mathbb{N}$. Note that the operator $\mathcal{L}: \delta \to \mathcal{L}\delta$ either reduces the number of steps in δ by one (if $\#(\delta) > 0$) or $\mathcal{L}\delta = \delta$ (if $\#(\delta) = 0$). Since the number of steps in δ is m, we have that $\#(\delta) \leqslant m$ and $\mathcal{L}^m\delta$ is an increasing function. It follows that $\rho_{\mathcal{L}^m\delta}$ is a coherent risk measure and hence

$$\sup\left\{\rho_{\mathcal{L}^m\delta}(S): S \in \mathfrak{S}_n(F)\right\} = n\rho_{\mathcal{L}^m\delta}(X_F).$$

Thus,

$$\lim_{n\to\infty}\frac{1}{n}\sup\left\{\rho_\delta(S): S \in \mathfrak{S}_n(F)\right\} = \rho_{\mathcal{L}^m\delta}(X_F).$$

Since $\#(\mathcal{L}^m\delta) = 0$, $\rho_{\mathcal{L}^m\delta}$ is a coherent distortion risk measure, $\rho_{\mathcal{L}^m\delta} \geqslant \rho_\delta$, and hence $\rho_{\mathcal{L}^m\delta} \geqslant \rho^+$ by Lemma 3.1. In addition, $\rho_{\mathcal{L}^m\delta}(X_F) \leqslant \rho^+(X_F)$ for $X_F \in L^\infty$ by Proposition 2.1 (c). Thus, the two distortion risk measures $\rho_{\mathcal{L}^m\delta}$ and ρ^+ agree on L^∞, and hence they agree on $\mathcal{X}$. $\square$

Recall that in Lemma 3.1, for any distortion function h, its largest dominated convex distortion function is given by $h^* := \sup\{g : g \in H_h\}$, where $H_{(\cdot)}$ is defined in (A.3).

Lemma A.5. *Let f, g and f_m, $m \in \mathbb{N}$ be distortion functions.*

(a) If $f_m \to f$ weakly as $m \to \infty$, then $f_m^* \to f^*$ uniformly, and $|\rho^+_{f_m}(X) - \rho^+_f(X)| \to 0$ as $m \to \infty$ for all $X \in L^\infty$.

(b) Let $\epsilon > 0$ be a real number. If $|f - g| \leqslant \epsilon$ on $[0,1]$, then $|\rho_f(X) - \rho_g(X)| \leqslant 2\epsilon||X||_\infty$ and $|\Gamma_{\rho_f}(X) - \Gamma_{\rho_g}(X)| \leqslant 2\epsilon||X||_\infty$ for all $X \in L^\infty$.

Proof. (a) We will use the Lévy distance between distribution functions, defined as

$$d(F,G) := \inf\{\epsilon > 0 : F(x-\epsilon) - \epsilon < G(x) < F(x+\epsilon) + \epsilon, \quad \forall x \in \mathbb{R}\}. \tag{A.9}$$

Note that the Lévy distance metricizes the weak topology on the set of distributions on $\mathbb{R}$. For any distortion functions (treated as distribution functions on $\mathbb{R}$) f and g, suppose $d(f,g) < \epsilon$. Then for each $f_0 \in H_f$, let $g_0(t) = \max\{0, f_0(t-\epsilon) - \epsilon\}$. It follows that g_0 is also convex and $g_0 \leqslant g$, hence $g_0 \in H_g$. Note that $d(f_0, g_0) \leqslant \epsilon$, and since f_0 is arbitrary we have that $d(f^*, g^*) \leqslant \epsilon$. This shows that $d(f_m^*, f^*) \to 0$ if $d(f_m, f) \to 0$. As f^* is convex with $f^*(1-) = f^*(1)$, we have that f^* is continuous. Therefore, the weak convergence $f_m^* \to f^*$ is uniform; see e.g. Chow and Teicher (2003, p.281). Recall that from Lemma A.1 (a), we have that

$$\rho_f(X_F) - \rho_g(X_F) = \int_0^1 (g(t) - f(t))\mathrm{d}F^{-1}(t). \tag{A.10}$$

Therefore, by the uniform convergence $f_m^* \to f^*$, we obtain that $|\rho^+_{f_m}(X) - \rho^+_f(X)| \to 0$.

(b) By (A.10), we have that $|\rho_f(X) - \rho_g(X)| \leqslant 2\epsilon||X||_\infty$ for $X \in L^\infty$. Similarly, we have that

$$\begin{aligned}|\Gamma_{\rho_f}(X_F) - \Gamma_{\rho_g}(X_F)| &\leqslant \lim_{n\to\infty}\left|\sup\left\{\frac{1}{n}\rho_f(S) : S \in \mathfrak{S}_n(F)\right\} - \sup\left\{\frac{1}{n}\rho_g(S) : S \in \mathfrak{S}_n(F)\right\}\right| \\ &\leqslant \lim_{n\to\infty}\sup\left\{\frac{1}{n}|\rho_f(S) - \rho_g(S)| : S \in \mathfrak{S}_n(F)\right\} \\ &\leqslant 2\epsilon||X_F||_\infty.\end{aligned}$$

□

Lemma A.6. *For any distortion function h and $X \in L^\infty$, we have that*

$$\Gamma_{\rho_h}(X) = \rho_h^+(X).$$

Proof. Let G_m denote the set of distortion functions with an m-step distortion factor. For $X \in L^\infty$ and $h_m \in G_m$, we have shown in Lemmas 3.1 and A.4 that

$$\Gamma_{\rho_{h_m}}(X) = \rho_{h_m}^+(X).$$

For any $\epsilon > 0$, denote

$$h^\epsilon(t) = \begin{cases} 0, & h(t) < \epsilon; \\ 1, & h(t) \geqslant 1-\epsilon; \\ h(t), & \text{otherwise.}\end{cases}$$

It is obvious that h^ϵ is a distortion function and $h^\epsilon(t) = h(t)$ on an interval $I = [a, b)$, $a > 0$ and $b < 1$, since $h(0+) = 0$ and $h(1-) = 1$. We can take two sequences of distortion functions $f_m \in G_m$ and $g_m \in G_m$, $m \in \mathbb{N}$, such that $f_m \nearrow h^\epsilon$ and $g_m \searrow h^\epsilon$ weakly as $m \to \infty$. By Lemma A.1 (b), we have that $\rho_{g_m} \leqslant \rho_{h^\epsilon} \leqslant \rho_{f_m}$ and hence

$$\rho_{g_m}^+(X) = \Gamma_{\rho_{g_m}}(X) \leqslant \Gamma_{\rho_{h^\epsilon}}(X) \leqslant \Gamma_{\rho_{f_m}}(X) = \rho_{f_m}^+(X). \tag{A.11}$$

It follows from Lemma A.5 (a) that $\rho_{f_m}^+(X) \to \rho_{h^\epsilon}^+(X)$ and $\rho_{g_m}^+(X) \to \rho_{h^\epsilon}^+(X)$ as $m \to \infty$. Therefore, taking limits on both sides of (A.11) leads to $\Gamma_{\rho_{h^\epsilon}}(X) = \rho_{h^\epsilon}^+(X)$. Since $|h - h^\epsilon| < \epsilon$, we have that by Lemma A.5 (b),

$$|\Gamma_{\rho_h}(X) - \rho_{h^\epsilon}^+(X)| = |\Gamma_{\rho_h}(X) - \Gamma_{\rho_{h^\epsilon}}(X)| \leqslant 2\epsilon||X||_\infty. \tag{A.12}$$

Note that $h^\epsilon \to h$ uniformly as $\epsilon \to 0$. Applying Lemma A.5 (a) again, we have that $\rho_{h^\epsilon}^+(X) \to \rho_h^+(X)$ as $\epsilon \to 0$. Finally, we obtain $\Gamma_{\rho_h}(X) = \rho_h^+(X)$ from (A.12) by taking $\epsilon \to 0$. □

Proof of Theorem 3.2. Lemma A.6 implies that $\Gamma_{\rho_h} = \rho_h^+$ on L^∞, and recall that $\Gamma_{\rho_h} \leqslant \rho_h^+$ on $\mathcal{X}$ by Proposition 2.1 (c). For $X \in \mathcal{X}$, let X_m, $m \in \mathbb{N}$ be a sequence of random variables in L^∞ such that

$X_m \nearrow X$ as $m \to \infty$. By the monotone convergence theorem, we have that $\rho_h^+(X_m) \to \rho_h^+(X)$. On the other hand, from the monotonicity of Γ_{ρ_h} we have that

$$\rho_h^+(X) = \lim_{m\to\infty} \rho_h^+(X_m) = \lim_{m\to\infty} \Gamma_{\rho_h}(X_m) \leqslant \Gamma_{\rho_h}(X) \leqslant \rho_h^+(X).$$

Therefore, $\Gamma_{\rho_h}(X) = \rho_h^+(X)$ for all $X \in \mathcal{X}$. $\square$

Proof of Theorem 3.3

Proof. Let $G_j,\ j = 1, \dots, K$ be the K different distributions in the sequence $\mathbf{F}$, $K < \infty$. Denote

$$H_j = \{i \in \mathbb{N} : F_i = G_j\}.$$

for $j = 1, \dots, K$. It is obvious that $\bigcup_{j=1}^K H_j = \mathbb{N}$. Define for $j = 1, \dots, K$,

$$\mathfrak{S}_j^n = \{X_1 + \cdots + X_m : X_i \sim G_j,\ m = \#\{i \in H_j, i \leqslant n\}\}.$$

Suppose that $S_j \in \mathfrak{S}_j^n,\ j = 1, \dots, K$. Then $\sum_{j=1}^K S_j \in \mathfrak{S}_n(\mathbf{F})$. Vice versa, for each $S \in \mathfrak{S}_n(\mathbf{F})$, it can be written as $S = \sum_{j=1}^K S_j$ where $S_j \in \mathfrak{S}_j^n$ for $j = 1, \dots, K$. It follows that

$$\sup_{S\in\mathfrak{S}_n(\mathbf{F})} \rho_h(S) \geqslant \sup_{S_j\in\mathfrak{S}_j^n,\ j=1,\dots,K} \rho_h\left(\sum_{j=1}^K S_j\right). \tag{A.13}$$

Note that

$$\{S_j \in \mathfrak{S}_j^n : j = 1, \dots, K\} \supset \{S_j \in \mathfrak{S}_j^n : j = 1, \dots, K,\ S_j \text{ are comonotonic}\}. \tag{A.14}$$

It follows from (A.13)-(A.14) and the comonotonic additivity of ρ_h that

$$\sup_{S\in\mathfrak{S}_n(\mathbf{F})} \rho_h(S) \geqslant \sup_{S_j\in\mathfrak{S}_j^n,\ j=1,\dots,K} \rho_h\left(\sum_{j=1}^K S_j\right) \geqslant \sum_{j=1}^K \sup_{S_j\in\mathfrak{S}_j^n} \rho_h(S_j). \tag{A.15}$$

Also note that

$$\sup_{S\in\mathfrak{S}_n(\mathbf{F})} \rho_h^+(S) = \sum_{j=1}^K \sup_{S_j\in\mathfrak{S}_j^n} \rho_h^+(S_j), \tag{A.16}$$

by the comonotonic additivity and subadditivity of ρ_h^+. For each $j = 1, \dots, K$, if $\#(H_j) = \infty$, by Theorem 3.2, as $n \to \infty$, we have that

$$\frac{\sup_{S_j\in\mathfrak{S}_j^n} \rho_h(S_j)}{\sup_{S_j\in\mathfrak{S}_j^n} \rho_h^+(S_j)} \to 1, \quad \text{and} \quad \sup_{S_j\in\mathfrak{S}_j^n} \rho_h^+(S_j) \to \infty. \tag{A.17}$$

If $\#(H_j) < \infty$, then

$$\sup_{S_j\in\mathfrak{S}_j^n} \rho_h(S_j) \leqslant \sup_{S_j\in\mathfrak{S}_j^n} \rho_h^+(S_j) = \#(H_j)\rho_h^+(X_{G_j}) < \infty,$$

i.e. both $\sup_{S_j\in\mathfrak{S}_j^n}\rho_h(S_j)$ and $\sup_{S_j\in\mathfrak{S}_j^n}\rho_h^+(S_j)$ are finite, and they do not affect the limit of (3.4) as $n\to\infty$, since the denominator of (3.4) goes to infinity by (A.17). In summary, by (A.15)-(A.17),

$$1 \geqslant \frac{\sup_{S\in\mathfrak{S}_n(\mathbf{F})}\rho_h(S)}{\sup_{S\in\mathfrak{S}_n(\mathbf{F})}\rho_h^+(S)} \geqslant \frac{\sum_{j=1}^K \sup_{S_j\in\mathfrak{S}_j^n}\rho_h(S_j)}{\sum_{j=1}^K \sup_{S_j\in\mathfrak{S}_j^n}\rho_h^+(S_j)} \to 1.$$

As a consequence,

$$\frac{\sup_{S\in\mathfrak{S}_n(\mathbf{F})}\rho_h(S)}{\sup_{S\in\mathfrak{S}_n(\mathbf{F})}\rho_h^+(S)} \to 1. \tag{A.18}$$

This shows that the theorem holds for any distortion risk measure ρ_h.

Now, consider the case for ρ_A^G. Since A is a finite set, by (A.18), we have that

$$\frac{\sup_{S\in\mathfrak{S}_n(\mathbf{F})}\rho_A^G(S)}{\sup_{S\in\mathfrak{S}_n(\mathbf{F})}\rho_A^+(S)} = \frac{\sup_{S\in\mathfrak{S}_n(\mathbf{F})}\max_{h\in A}\rho_h(S)}{\sup_{S\in\mathfrak{S}_n(\mathbf{F})}\max_{h\in A}\rho_h^+(S)} \geqslant \min_{h\in A}\left\{\frac{\sup_{S\in\mathfrak{S}_n(\mathbf{F})}\rho_h(S)}{\sup_{S\in\mathfrak{S}_n(\mathbf{F})}\rho_h^+(S)}\right\} \to 1. \tag{A.19}$$

$\square$

Proof of Theorem 4.1

By Corollary 2.3, it suffices to show that Γ_ρ is coherent. This is implied immediately by the following lemma.

Lemma A.7. *For any risk measure ρ, the following hold:*

(a) $\Gamma_\rho(0) = 0$.

(b) For $k\in\mathbb{N}$, $\Gamma_\rho(kX_F) \leqslant k\Gamma_\rho(X_F)$.

(c) If ρ is convex, then Γ_ρ is subadditive and positive homogeneous.

Proof. (a) Recall that $|\rho(0)| < \infty$ is assumed throughout. Then $\Gamma_\rho(0) = \limsup_{n\to\infty}\frac{1}{n}\rho(0) = 0$.

(b) Denote by F_k the distribution of $X_{F_k} := kX_F$. It is obvious that $X_{F_k}\in\mathfrak{S}_k(F)$ by taking $X_1 = \cdots = X_k = X_F$. As a consequence, $\mathfrak{S}_n(F_k)\subset\mathfrak{S}_{kn}(F)$ since each element in $\mathfrak{S}_n(F_k)$ can be written as an element in $\mathfrak{S}_{kn}(F)$. It follows that

$$\begin{aligned}
\Gamma_\rho(kX_F) &= \limsup_{n\to\infty}\frac{1}{n}\sup\{\rho(S): S\in\mathfrak{S}_n(F_k)\}\\
&\leqslant \limsup_{n\to\infty}\frac{1}{n}\sup\{\rho(S): S\in\mathfrak{S}_{kn}(F)\}\\
&= k\limsup_{n\to\infty}\frac{1}{kn}\sup\{\rho(S): S\in\mathfrak{S}_{kn}(F)\}\\
&\leqslant k\limsup_{n\to\infty}\frac{1}{n}\sup\{\rho(S): S\in\mathfrak{S}_n(F)\} = k\Gamma_\rho(X_F).
\end{aligned}$$

(c) By Proposition 2.2, Γ_ρ is convex if ρ is convex. For any $X, Y \in \mathcal{X}$, using (b),

$$\Gamma_\rho(X+Y) \leqslant 2\Gamma_\rho\left(\frac{1}{2}X + \frac{1}{2}Y\right) \leqslant 2\left(\frac{1}{2}\Gamma_\rho(X) + \frac{1}{2}\Gamma_\rho(Y)\right) = \Gamma_\rho(X) + \Gamma_\rho(Y).$$

Thus we have the subadditivity. The positive homogeneity is implied by subadditivity, convexity and zero-normalization, via Deprez and Gerber (1985, Theorem 2).

□

Proof of Corollary 4.2

Proof. Without loss of generality, we can assume $\rho(0) = 0$; otherwise one can work with the convex risk measure $\hat{\rho} = \rho - \rho(0)$, and easily check that $\Gamma_\rho = \Gamma_{\hat{\rho}}$ since $\Gamma_{\rho,n} - \Gamma_{\hat{\rho},n} = \frac{1}{n}\rho(0) \to 0$. As a consequence, $v \geqslant 0$ on $\mathcal{P}$. Note that by Proposition 2.1 and the fact that a convex risk measure with the Fatou property preserves convex order, we have that for each distribution F, the worst dependence structure is the comonotonic one $X_1 = \ldots = X_n = X_F$, hence

$$\begin{aligned}
\Gamma_\rho(X_F) &= \sup_{n\in\mathbb{N}} \frac{1}{n} \sup_{\mu\in\mathcal{P}} \left\{ \int_0^1 \mathrm{ES}_\alpha(nX_F)\mathrm{d}\mu(\alpha) - v(\mu) \right\} \\
&= \sup_{\mu\in\mathcal{P}} \left\{ \int_0^1 \mathrm{ES}_\alpha(X_F)\mathrm{d}\mu(\alpha) - \inf_{n\in\mathbb{N}} \frac{1}{n} v(\mu) \right\} \\
&= \sup_{\mu\in\mathcal{P}_v} \left\{ \int_0^1 \mathrm{ES}_\alpha(X_F)\mathrm{d}\mu(\alpha) \right\},
\end{aligned}$$

where the second equality is obtained exchanging the two suprema. □

Proof of Proposition 4.3

Proof. First assume $0 < b_\ell \leqslant a_\ell < \infty$. The case when $b_\ell = 0$ or $a_\ell = \infty$ will be commented on at the end of the proof. Without loss of generality we can assume $\ell(0) = x_0 = 0$. Note that

$$\begin{aligned}
&\sup\{\inf\{x \in \mathbb{R} : \mathbb{E}[\ell(S-x)] \leqslant 0)\} : S \in \mathfrak{S}_n(F)\} \\
&= \inf\{x \in \mathbb{R} : \mathbb{E}[\ell(S-x)] \leqslant 0, \text{ for all } S \in \mathfrak{S}_n(F)\} \\
&= \inf\{x \in \mathbb{R} : \sup\{\mathbb{E}[\ell(S-x)] : S \in \mathfrak{S}_n(F)\} \leqslant 0\}.
\end{aligned}$$

We have, by the convexity of ℓ, that

$$\sup\{\mathbb{E}[\ell(S-x)] : S \in \mathfrak{S}_n(F)\} = \mathbb{E}[\ell(nX_F - x)].$$

It follows that

$$\begin{aligned}\Gamma_{\tau_\ell^S}(X_F) &= \limsup_{n\to\infty}\frac{1}{n}\sup\{\inf\{x\in\mathbb{R}:\mathbb{E}[\ell(S-x)]\leqslant 0)\}:S\in\mathfrak{S}_n(F)\}\\ &= \limsup_{n\to\infty}\frac{1}{n}\inf\{x\in\mathbb{R}:\mathbb{E}[\ell(nX_F-x)]\leqslant 0\}\\ &= \limsup_{n\to\infty}\inf\{t\in\mathbb{R}:\mathbb{E}[\ell(nX_F-nt)]\leqslant 0\}.\end{aligned}$$

Let $t_n^* = \frac{1}{n}\tau_\ell^S(nX_F)$ be the unique solution to $\mathbb{E}[\ell(nX_F - nt)] = 0$ (the existence of t_n^* is implied by the convexity of ℓ). It follows that

$$\frac{1}{n}\mathbb{E}[\ell(nX_F-nt_n^*)] = \frac{1}{n}\mathbb{E}[\ell(n(X_F-t_n^*))\mathrm{I}_{\{X_F>t_n^*\}}] + \frac{1}{n}\mathbb{E}[\ell(n(X_F-t_n^*))\mathrm{I}_{\{X_F\leqslant t_n^*\}}] = 0. \tag{A.20}$$

Let $t_0 = \limsup_{n\to\infty} t_n^*$ and $t_1 = \liminf_{n\to\infty} t_n^*$. Note that $\ell(nx)/n \leqslant a_\ell x$ for all $x > 0$. We have that

$$\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[\ell(n(X_F-t_n^*))\mathrm{I}_{\{X_F>t_n^*\}}] \leqslant \limsup_{n\to\infty}\mathbb{E}[a_\ell(X_F-t_n^*)_+] = a_\ell\mathbb{E}[(X_F-t_1)_+].$$

Let $\{t_{n_k}^*\}$ be a subsequence of $\{t_n^*\}$ which converges to t_1. Since

$$\frac{1}{n_k}\mathbb{E}[(\ell(n_kX_F-t_{n_k}^*))_+] \leqslant a_\ell\mathbb{E}[(X_F-t_{n_k}^*)_+] = a_\ell\mathbb{E}[(X_F-t_1+o(1))_+] < \infty,$$

by dominated convergence theorem, we have that

$$\begin{aligned}\lim_{k\to\infty}\frac{1}{n_k}\mathbb{E}[\ell(n_k(X_F-t_{n_k}^*))\mathrm{I}_{\{X_F>t_{n_k}^*\}}] &= \mathbb{E}\left[\lim_{k\to\infty}\frac{1}{n_k}\ell(n_k(X_F-t_{n_k}^*))\mathrm{I}_{\{X_F>t_{n_k}^*\}}\right]\\ &= \mathbb{E}[a_\ell(X_F-t_1)_+].\end{aligned}$$

In summary, we have that

$$\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[\ell(n(X_F-t_n^*))\mathrm{I}_{\{X_F>t_n^*\}}] = a_\ell\mathbb{E}[(X_F-t_1)_+],$$

and similarly we obtain that

$$\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[\ell(n(X_F-t_n^*))\mathrm{I}_{\{X_F\leqslant t_n^*\}}] = -b_\ell\mathbb{E}[(t_0-X_F)_+].$$

It follows from (A.20) that

$$a_\ell\mathbb{E}[(X_F-t_1)_+] = b_\ell\mathbb{E}[(t_0-X_F)_+]. \tag{A.21}$$

Similarly, we have

$$\liminf_{n\to\infty}\frac{1}{n}\mathbb{E}[\ell(n(X_F-t_n^*))\mathrm{I}_{\{X_F>t_n^*\}}] = a_\ell\mathbb{E}[(X_F-t_0)_+],$$

and

$$\liminf_{n\to\infty}\frac{1}{n}\mathbb{E}[\ell(n(X_F-t_n^*))\mathrm{I}_{\{X_F\leqslant t_n^*\}}] = -b_\ell\mathbb{E}[(t_1-X_F)_+].$$

Again, it follows from (A.20) that

$$a_\ell \mathbb{E}[(X_F - t_0)_+] = b_\ell \mathbb{E}[(t_1 - X_F)_+]. \tag{A.22}$$

(A.21)-(A.22) imply that $t_0 = t_1$ and $t^* := \lim_{n\to\infty} t_n^*$ is the unique solution to

$$a_\ell \mathbb{E}[(X_F - t^*)_+] = b_\ell \mathbb{E}[(t^* - X_F)_+].$$

When $b_\ell = 0$ or $a_\ell = \infty$, (A.20) implies that $\limsup_{n\to\infty} \mathbb{E}[\ell(X_F - t_n^*)\mathrm{I}_{\{X_F > t_n^*\}}] = 0$. In this case, $t^* = \text{ess-sup}(X_F)$. $\square$

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.

Bäuerle, N. and Müller, A. (2006). Stochastic orders and risk measures: Consistency and bounds. *Insurance: Mathematics and Economics*, **38**(1), 132–148.

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

BCBS (2013). *Consultative Document October 2013. Fundamental review of the trading book: A revised market risk framework.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements.

Bellini, F. and Bignozzi, V. (2014). On elicitable risk measures. *Quantitative Finance,* forthcoming.

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

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

Bühlmann, H.(1970). *Mathematical methods in risk theory,* **172**. Springer Science & Business Media.

Burgert, C. and Rüschendorf, L. (2006). Consistent risk measures for portfolio vectors. *Insurance: Mathematics and Economics*, **38**(2), 289–297.

Chow, Y. S. and Teicher, H. (2003). *Probability theory: independence, interchangeability, martingales*. Third Edition. Springer, New York.

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

de Berg, M., van Kreveld, M., Overmars, M. and Schwarzkopf O. (2008). *Computational geometry: Algorithms and applications.* Third Edition. Springer, Berlin.

Delbaen, F. (2012). *Monetary utility functions.* Osaka University Press, Osaka.

Delbaen, F., Bellini, F., Bignozzi, V., Ziegel, J. F. (2015). On convex risk measures with the CxLS property. *Preprint*, ETH Zurich.

Denneberg, D. (1990). Distorted probabilities and insurance premiums. *Methods of Operations Research*, **63**, 3–5.

Denneberg, D. (1994). *Non-additive measure and integral,* **27**. Springer Science & Business Media.

Deprez, O. and Gerber, H. U. (1985). On convex principles of premium calculation. *Insurance: Mathematics and Economics*, **4**(3), 179–189.

Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R. and Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, **31**(1), 3–33.

Dhaene, J., Laeven, R. J. A., Vanduffel, S., Darkiewicz, G. and Goovaerts, M. J. (2008). Can a coherent risk measure bo too subadditive? *Journal of Risk and Insurance*, **75**(2), 365–386.

Dhaene, J., Linders, D., Schoutens, W. and Vyncke, D. (2012). The Herd Behavior Index: A new measure for the implied degree of co-movement in stock markets *Insurance: Mathematics and Economics,* **50**(3), 357–370.

Embrechts, P. and Puccetti, G. (2006). Bounds for functions of dependent risks. *Finance and Stochastics*, **10**(3), 341–352.

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

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

Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, forthcoming.

Emmer, S., Kratz, M. and Tasche, D. (2014). What is the best risk measure in practice? A comparison of standard measures. *Preprint*, ESSEC Business School.

Filipović, D. and Svindland, G. (2012). The canonical model space for law-invariant convex risk measures is L^1. *Mathematical Finance*, **22**(3), 585–589.

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. (2011). *Stochastic Finance: An Introduction in Discrete Time.* Third Edition. Walter de Gruyter, Berlin.

Frittelli M. and Rosazza Gianin E. (2002). Putting order in risk measures. *Journal of Banking and Finance*, **26**(7), 1473–1486.

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

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

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

Goovaerts, M., De Vylder, F. and Haezendonck, J. (1984). *Insurance premiums,* 1–398. North-Holland.

Hashorva, E. and Mishura, Y. (2014). Boundary non-crossings of additive Wiener fields. *Lithuanian Mathemathematical Journal*, **54**(3), 277–289.

IAIS (2014). *Consultation Document December 2014. Risk-based global insurance capital standard.* International Association of Insurance Supervisors.

Jouini, E. and Schachermayer, W. and Touzi, N. (2006). Law invariant risk measures have the Fatou property. *Advances in mathematical economics*, **9**, 49–71.

Kou, S. and Peng, X. (2014). On the measurement of economic tail risk. *Preprint*, Hong Kong University of Science and Technology.

Krätschmer, V., Schied, A. and Zähle, H. (2014). Comparative and qualitative robustness for law-invariant risk measures. *Finance and Stochastics*, **18**(2), 271–295.

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

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

Puccetti, G. and Rüschendorf, L. (2014). Asymptotic equivalence of conservative VaR- and ES-based capital charges. *Journal of Risk*, **16**(3), 3–22.

Puccetti, G., Wang, B. and Wang, R. (2013). Complete mixability and asymptotic equivalence of worst-possible VaR and ES estimates. *Insurance: Mathematics and Economics*, **53**(3), 821–828.

Schmeidler, D. (1989). Subject probability and expected utility without additivity. *Econometrica*, **57**(3), 571–587.

Svindland, G. (2009). Convex risk measures beyond bounded risks. *Doctoral dissertation*, Ludwig Maximilian Univeristy.

Wang, R. (2014). Asymptotic bounds for the distribution of the sum of dependent random variables. *Journal of Applied Probability*, **51**(3), 780–798.

Wang, B. and Wang, R. (2015). Extreme negative dependence and risk aggregation. *Journal of Multivariate Analysis*, **136**, 12–25.

Wang, R., Peng, L. and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, **17**(2), 395–417.

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

Weber, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance*, **16**(2), 419–441.

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

Ziegel, J. (2014). Coherence and elicitability. *Mathematical Finance*, forthcoming.