Risk measures induced by efficient insurance contracts

Qiuqi Wang[*] Ruodu Wang[†] Ričardas Zitikis[‡]

January 8, 2022

Abstract

The Expected Shortfall (ES) is one of the most important regulatory risk measures in finance, insurance, and statistics, which has recently been characterized via sets of axioms from perspectives of portfolio risk management and statistics. Meanwhile, there is large literature on insurance design with ES as an objective or a constraint. A visible gap is to justify the special role of ES in insurance and actuarial science. To fill this gap, we study the characterization of risk measures induced by efficient insurance contracts, i.e., those that are Pareto optimal for the insured and the insurer. One of our major results is that we characterize a mixture of the mean and ES as the risk measure of the insured and the insurer, when contracts with deductibles are efficient. Characterization results of other risk measures, including the mean and distortion risk measures, are also presented by linking them to different sets of contracts.

Keywords: optimal insurance, Expected Shortfall, Pareto optimality, deductible, concentration

1 Introduction

Optimal insurance and reinsurance design problems have been a prevalent topic for both researchers and practitioners in insurance for decades, since the seminal work of Arrow (1963) showing that deductible insurance is optimal for a risk-averse insured when the insurer is risk neutral. As natural extensions, Raviv (1979) studied conditions for optimality of deductible insurances when the insured and the insurer are both risk averse. Schlesinger (1981) examined the optimal choice of a risk-averse insured given that the insurance is of deductible type.

Previous studies on optimal (re)insurance design problems have shown considerations from several different perspectives. The majority of the studies focus on optimization under specific classes of optimization criteria quantifying the risk of decision makers; see e.g., Gollier and Schlesinger (1996)

[*]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

[‡]School of Mathematical and Statistical Sciences, University of Western Ontario, Canada. ✉ rzitikis@uwo.ca

and Schlesinger (1997) for criteria preserving second-order stochastic dominance; Cai and Tan (2007), Cai et al. (2008) and Bernard and Tian (2009) for Value-at-Risk (VaR) and the Expected Shortfall (ES, also called CTE or TVaR in the above literature); Cui et al. (2013) for distortion risk measures or dual utilities (Yaari, 1987); and Braun and Muermann (2004) for regret-theoretical expected utilities. For more recent developments on optimal insurance with risk measures, we refer to Cai and Chi (2020) and the references therein. Moreover, optimal (re)insurance contract design problems are studied under a variety of constraints and formulations. We refer to studies on efficient insurance contracts with background risk (e.g., Gollier (1996) and Dana and Scarsini (2007)) and limited liability (e.g., Cummins and Mahul (2004) and Hofmann et al. (2019)). More recently, Lo et al. (2021) analyzed the set of universally marketable indemnities with risk measures preserving convex orders.

Most of the previous literature aims to derive optimal forms of ceded loss functions under various scenarios and constraints. To the best of our knowledge, there is no relevant research on (re)insurance contract design problems focusing on identifying risk measures adopted by the insured and the insurer. Therefore, we study optimal insurance contract design problems through a distinctive perspective if compared to previous literature. Namely, the main goal of the present paper is to answer the following (converse) question: In order for efficient contracts to be some sets of contracts commonly seen in insurance practice (e.g., of deductible form), which risk measures should the insurer and the insured use as their objectives? Specifically, we characterize different classes of risk measures adopted by the insured and the insurer given different sets of ceded loss functions that are Pareto optimal.

The risk measure ES has been widely applied in the contexts of financial regulation, risk management, and insurance. In particular, ES is the standard measure for market risk in the Fundamental Review of the Trading Book (FRTB) of BCBS (2016, 2019). In the insurance regulation framework of Solvency II, the risk measure Value-at-Risk (VaR) is the dominating risk measure. There is a growing academic literature on various problems using ES in actuarial science (where ES is often called TVaR). Most of these studies motivate the use of ES as a coherent risk measure (Artzner et al., 1999) and its advantages over VaR. Recently, Wang and Zitikis (2021) proposed the axiom called "no reward for concentration" (NRC) which, together with a few other standard axioms, characterizes ES.[1] The main objective of Wang and Zitikis (2021) is to separate ES from other coherent risk measures via the axiom of NRC, thus answering the question of why one uses ES instead of other risk measures from an axiomatic point of view. The interpretation and implication of the NRC axiom in financial regulation have been extensively discussed in Wang and Zitikis (2021) in the context of FRTB; see also an alternative formulation for axiomatizing ES in Han et al. (2021).

[1]The NRC axiom for a risk measure ρ means that the exists a regulatory stress event A such that $\rho(X+Y) = \rho(X)+\rho(Y)$ whenever X and Y both have the tail event A, meaning that X satisfies $X(\omega) \geq X(\omega')$ for almost surely all $\omega \in A$ and $\omega' \in A^c$, and so does Y.

Given the big volume of research with ES in actuarial science, it is of great interest to understand whether ES plays a special role in insurance. The NRC axiom of Wang and Zitikis (2021) does not apply in the insurance context since it is interpreted as a requirement of portfolio risk assessment. To understand the special role of ES in insurance, new insights that are specific to insurance design are therefore needed.

We work mainly within the framework of convex risk measures of Föllmer and Schied (2002), which is a flexible and popular class of risk measures in risk management. As the main contribution of this paper, we show that the set of efficient ceded loss functions of deductible form corresponds to the family of mixtures of ES and the mean (Theorem 4.2). If we further impose lower semicontinuity as in Wang and Zitikis (2021), then we arrive at the family of ES (Lemma A.3). Our work also extends Embrechts et al. (2021), who characterized the mixture of the mean and ES, called an ES/$\mathbb{E}$-mixture, as the only coherent Bayes risk measure from the perspective of statistical inference. In addition, if the set of efficient ceded loss functions is the set of all slowly growing (1-Lipschitz) functions, then the corresponding risk measures are precisely the convex distortion risk measures (Theorem 4.1). Mathematically, our results are based on connecting various risk measures with different additivity forms over the ceded losses and the retained losses.

For illustrative purposes, we take the perspective of an insurance design problem between an insurer and an insured. Our technical results can certainly be applied in the reinsurance setting as well, where risk measures are often encountered.

The rest of the paper is organized as follows. Section 2 contains some preliminaries on insurance losses and risk measures. Section 3 sets up the formulation of the insurance contract design problem and states economic assumptions. Section 4 contains our main characterization results of the risk measures used by the insured and the insurer given different Pareto-optimal sets of ceded loss functions. The results make natural connections between some common sets of ceded loss functions and common classes of risk measures in insurance practice. We also discuss economic implications of these results on the design of insurance menus by the insurer. Appendix A contains proofs of the main results accompanied with relevant technical lemmas.

2 Preliminaries on risk measures

We consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{X}$ be the set of all bounded random variables, and let $\mathcal{X}_+$ be the set of all non-negative random variables in $\mathcal{X}$ representing insurable losses. Let $\mathcal{I}$ be a class of non-negative functions on $[0, \infty)$ which represent possible insurance ceded loss functions. For an insurable loss random variable $X \in \mathcal{X}_+$ and a contract $f \in \mathcal{I}$, $f(X)$ represents the payment to the insured, and $X - f(X)$ represents the retained loss of the insured. Losses are usually quantified by risk measures which are mappings from $\mathcal{X}$ to the set of real numbers, representing riskiness. Below we recall some properties of risk measures ρ, which are commonly encountered in the risk management literature.

Law invariance: $\rho(X) = \rho(Y)$ for all $X, Y \in \mathcal{X}$ such that $X \overset{\mathrm{d}}{=} Y$.[2]

Monotonicity: $\rho(X) \geq \rho(Y)$ for all $X, Y \in \mathcal{X}$ such that $X \geq Y$.

Translation invariance: $\rho(X + d) = \rho(X) + d$ for all $X \in \mathcal{X}$ and $d \in \mathbb{R}$.

Convexity: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$ for all $X, Y \in \mathcal{X}$ and $\lambda \in [0, 1]$.

Positive homogeneity: $\rho(\lambda X) = \lambda\rho(X)$ for all $X \in \mathcal{X}$ and $\lambda \geq 0$.

Following Artzner et al. (1999) and Föllmer and Schied (2016), ρ is a *monetary risk measure* if it is monotone and translation invariant; a monetary risk measure ρ is called a *convex risk measure* if it satisfies convexity, and it is *coherent* if it is also positively homogeneous.[3] For $X \in \mathcal{X}$, a *distortion risk measure* is defined as

$$\rho(X) = \int_0^\infty h(\mathbb{P}(X > x))\,\mathrm{d}x + \int_{-\infty}^0 (h(\mathbb{P}(X > x)) - 1)\,\mathrm{d}x,$$

where $h : [0, 1] \to [0, 1]$ is an increasing function with $h(0) = 0$ and $h(1) = 1$, and h is called the *distortion function* of ρ. Distortion risk measures are always monetary, positively homogeneous, and law invariant, and they are coherent if and only if their distortion functions are concave; see e.g., Yaari (1987) and Wang et al. (1997). For the application of distortion risk measures to insurance premium principle calculation, see Wang et al. (1997). For $X \in \mathcal{X}$ and $p \in (0, 1)$, the *Value-at-Risk* (VaR) is the left-quantile given by

$$\mathrm{VaR}_p(X) = F_X^{-1}(p) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq p\}.$$

For $X \in \mathcal{X}$ and $p \in [0, 1)$, the *Expected Shortfall* (ES) is defined as

$$\mathrm{ES}_p(X) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_t(X)\,\mathrm{d}t.$$

It is well known that ES_p is a convex risk measure while VaR_p is not. Similarly, for $X \in \mathcal{X}$ and $p \in (0, 1]$, the *left-ES* risk measure (see e.g., Embrechts et al. (2015)) is defined by

$$\mathrm{ES}_p^-(X) = \frac{1}{p}\int_0^p \mathrm{VaR}_t(X)\,\mathrm{d}t.$$

[2] We write $X \overset{\mathrm{d}}{=} Y$ when two random variables X and Y follow the same distribution.

[3] Artzner et al. (1999) defined coherent risk measures via subadditivity instead of convexity. A risk measure ρ is subadditive if $\rho(X + Y) \leq \rho(X) + \rho(Y)$ for all $X, Y \in \mathcal{X}$. Subadditivity and convexity are equivalent when positive homogeneity holds.

Throughout the paper, we write $x \wedge y = \min\{x, y\}$, $x \vee y = \max\{x, y\}$, $x_+ = x \vee 0$ and $x_- = (-x) \vee 0$. For an event $A \in \mathcal{F}$, its complement is denoted by A^c.

3 Optimal insurance contract design

In this section, we explain the optimal insurance design problem. For the economic setting, we make the following assumptions:

(A) The insured and the insurer may hold different attitudes towards risk. The insured adopts the risk measure $\rho : \mathcal{X} \to \mathbb{R}$ while the insurer uses the risk measure $\psi : \mathcal{X} \to \mathbb{R}$. The insured and the insurer do not observe the risk measure of their counterparty.

(B) The premium functional is specified as $\pi : \mathcal{I} \to \mathbb{R}$, which usually does not take negative values. For insurance loss $X \in \mathcal{X}_+$, note that $X - f(X) + \pi(f)$ is the total risk (i.e., total loss random variable) of the insured, and $f(X) - \pi(f)$ is the total risk of the insurer. Thus, the risk values of the insurance loss to the insured and the insurer are $\rho(X - f(X) + \pi(f))$ and $\psi(f(X) - \pi(f))$, respectively.

(C) The insured and the insurer agree on an insurance contract $f \in \mathcal{I}$ that is *Pareto optimal* defined next.

Definition 3.1. For $X \in \mathcal{X}_+$, $\pi : \mathcal{I} \to \mathbb{R}$, and $\rho, \psi : \mathcal{X} \to \mathbb{R}$, an insurance contract $f \in \mathcal{I}$ is called Pareto optimal if there is no $g \in \mathcal{I}$, such that

$$\rho(X - f(X) + \pi(f)) \geq \rho(X - g(X) + \pi(g))$$

and

$$\psi(f(X) - \pi(f)) \geq \psi(g(X) - \pi(g)),$$

with at least one of the two inequalities strict. Pareto optimality is also known as (Pareto) *efficiency*.

A Pareto optimization problem is closely related to the minimization of a convex combination of the objective functionals of all parties, which can be seen in, e.g., Gerber (1974), Barrieu and Scandolo (2008), Cai et al. (2017) and Embrechts et al. (2018). For $X \in \mathcal{X}_+$, $\pi : \mathcal{I} \to \mathbb{R}$, and $\rho, \psi : \mathcal{X} \to \mathbb{R}$, we define the set of minimizers of the sum of the two objectives for the insured and the insurer as

$$\mathcal{I}^X_{\rho,\psi} = \underset{g \in \mathcal{I}}{\arg\min}\{\rho(X - g(X) + \pi(g)) + \psi(g(X) - \pi(g))\}.$$

If we further assume that ρ and ψ are translation invariant, then we have

$$\mathcal{I}^X_{\rho,\psi} = \arg\min_{g\in\mathcal{I}}\{\rho(X - g(X)) + \psi(g(X))\}. \tag{3.1}$$

In this case, the set $\mathcal{I}^X_{\rho,\psi}$ is independent of the choice of the premium functional π. Below we give a characterization of the Pareto-optimal problem in our context as the minimization of the total insurance value of the insured and the insurer.

Proposition 3.1. *For two translation-invariant risk measures $\rho, \psi : \mathcal{X} \to \mathbb{R}$ and $X \in \mathcal{X}_+$, the following are equivalent:*

(i) an insurance contract $f \in \mathcal{I}$ is Pareto optimal for all $\pi : \mathcal{I} \to \mathbb{R}_+$;

(ii) an insurance contract $f \in \mathcal{I}$ is Pareto optimal for $\pi : h \mapsto \psi(h(X))$;[4]

(iii) $f \in \mathcal{I}^X_{\rho,\psi}$.

Proofs of all results in this paper are in Appendix A.

In a similar spirit to Proposition 3.1, a characterization of Pareto optimality in the context of risk sharing problems can be found in Embrechts et al. (2018). Proposition 3.1 ensures that if the objectives ρ and ψ for the two parties are translation invariant, then by (3.1), a Pareto-optimal insurance contract can typically be obtained by solving the following minimization problem:

$$\min_{g\in\mathcal{I}} \left\{\rho(X - g(X)) + \psi(g(X))\right\}. \tag{3.2}$$

A minimizer of (3.2) may not be unique in many situations. Hence, the set $\mathcal{I}^X_{\rho,\psi}$ of efficient ceded loss functions is not a singleton in general. In the literature on optimal insurance design problems, there are many common sets of ceded loss functions. Some notable refinements include:

1. The set $\mathcal{I}_0$ of all non-negative functions f on $[0, \infty)$ satisfying $f(x) \le x$ for $x \ge 0$. This property means that the payment cannot exceed the total loss incurred, and it is a common feature of almost all insurance contracts in practice. In particular, $f(0) = 0$, and thus there is no insurance payment if there is no loss incurred.

2. The set $\mathcal{I}_1$ of all increasing functions in $\mathcal{I}_0$. This property means that larger incurred losses lead to higher payments to the insured.

3. The set $\mathcal{I}_2 = \{f \in \mathcal{I}_1 : f(y) - f(x) \le y - x \text{ for all } y \ge x \ge 0\}$, which is the set of all *slowly growing* increasing functions in $\mathcal{I}_1$. The slowly growing property is commonly assumed

[4]This means $\phi(h) = \psi(h(X))$ for all $h \in \mathcal{I}$.

to avoid the problem of ex-post moral hazard (Huberman et al. (1983)) via the concept of comonotonicity; see Proposition 4.1 below.

4. The set $\mathcal{I}_1^d = \{f \in \mathcal{I}_1 : f(x) \le (x-d)_+ \text{ for all } x \ge 0\}$. Ceded loss functions within this set does not exceed the direct deductible form. Note that

$$\mathcal{I}_1^d = \{f \in \mathcal{I}_1 : f(d) = 0,\ x - f(x) \ge d \text{ for all } x > d\}.$$

   Thus this class includes contract functions with deductible $d \ge 0$. Also, we require that the retained loss of the insured should be at least at the deductible level d, given that the random loss exceeds the deductible level. In particular, we have $\mathcal{I}_1^0 = \mathcal{I}_1$.

Among the above sets, we have

$$\mathcal{I}_2 \subset \mathcal{I}_1 \subset \mathcal{I}_0 \quad \text{and} \quad \mathcal{I}_1^d \subset \mathcal{I}_1 \subset \mathcal{I}_0.$$

Throughout, $\subset$ represents non-strict set inclusion. Contracts of deductible forms within the set $\mathcal{I}_1^d$ are commonly seen in the insurance market. We next give some examples.

Example 3.1 (Deductible insurance with coinsurance)**.** Consider the following ceded loss function:

$$f(x) = \alpha(x-d)_+, \quad x \ge 0,$$

which presents an insurance contract with deductible $d \ge 0$ and coinsurance parameter $\alpha \in [0,1]$. We have $f \in \mathcal{I}_1^d$ since f is bounded from above by $(x-d)_+$. See Figure 3.1 (left-hand panel).

Example 3.2 (Deductible insurance with policy limit)**.** The following ceded loss function

$$f(x) = (x-d)_+ \wedge u, \quad x \ge 0,$$

is also in the set $\mathcal{I}_1^d$. It represents an insurance contract truncated at deductible $d \ge 0$ and censored at the policy upper limit $u \ge 0$. The function is plotted in Figure 3.1 (right-hand panel).

We focus on the above three subsets due to their prominence in real-world insurance contracts. Other subsets of $\mathcal{I}_1$, such as classes of convex functions, piece-wise linear functions, or functions with the Vajda condition, have also been studied in the literature, but they correspond to different practical considerations; see e.g., Vajda (1962), Cai et al. (2008), Chi and Weng (2013) and Chen (2021).

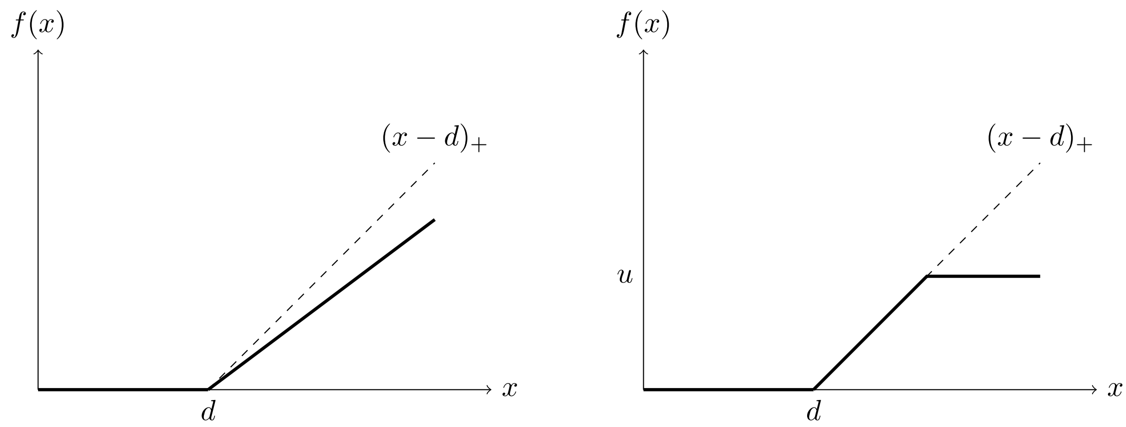

Figure 3.1: Solid lines represent the ceded loss functions of deductible insurance with coinsurance (left-hand panel) and deductible insurance with policy limit (right-hand panel); dashed lines represent ceded loss function with direct deductible

4 Risk measures implied by Pareto-optimal contracts

4.1 Main characterization results

In this section, we characterize measures ρ and ψ for the insured and the insurer in the optimal insurance design problem with different Pareto-optimal sets of ceded loss functions.

We first collect some dependence concepts that will be helpful to distinguish different properties of risk measures in our main results. A random vector $(X,Y)\in\mathcal{X}^2$ is said to be *comonotonic* if $(X(\omega)-X(\omega'))(Y(\omega)-Y(\omega'))\geq 0$ for almost every $\omega,\omega'\in\Omega$; see also Wang and Zitikis (2020). A risk measure $\rho:\mathcal{X}\to\mathbb{R}$ is said to be *comonotonic-additive* if $\rho(X+Y)=\rho(X)+\rho(Y)$ for all comonotonic $(X,Y)\in\mathcal{X}^2$. Following similar definitions as those of Wang and Zitikis (2021), for an event $A\in\mathcal{F}$ with $0<\mathbb{P}(A)<1$, we call A a *tail event* of a random variable $X\in\mathcal{X}$ if $X(\omega)\geq X(\omega')$ for almost surely all $\omega\in A$ and $\omega'\in A^c$. A tail event A is called a *p-tail event* if $\mathbb{P}(A)=1-p$. We say that a random vector $(X_1,\dots,X_n)\in\mathcal{X}^n$ is *p-concentrated* if there exists a common p-tail event of $X_1,\dots,X_n$. For fixed $d\geq 0$ and $p\in[0,1]$, define the sets

$$\mathcal{X}_p^d=\{X\in\mathcal{X}_+:p=\mathbb{P}(X\leq d)\}$$

and

$$\mathcal{X}_p=\{X\in\mathcal{X}:p=\mathbb{P}(X\leq d)\text{ for some }d\geq 0\}\supset\bigcup_{d\geq 0}\mathcal{X}_p^d.$$

We note that $\mathcal{X}_p\supset\mathcal{X}_p^d$ and $\mathcal{X}_p$ contains random variables that may take negative values and may be discrete. The following proposition connects the dependence structure of $(f(X),X-f(X))$ with the function $f \in \mathcal{I}_1$.

Proposition 4.1. *The following statements hold.*

(i) $(f(X), X - f(X))$ *is comonotonic for all* $f \in \mathcal{I}_2$ *and* $X \in \mathcal{X}_+$.

(ii) For fixed $d > 0$ *and* $p \in [0, 1)$, $(f(X), X - f(X))$ *is* p*-concentrated for all* $f \in \mathcal{I}_1^d$ *and* $X \in \mathcal{X}_p^d$.

Following the terminology in Embrechts et al. (2021), for $\lambda \in \mathbb{R}$ and $p \in (0, 1)$, we say that the linear combination

$$\mathrm{ES}_p^\lambda(X) = \lambda \mathrm{ES}_p(X) + (1 - \lambda)\mathbb{E}[X], \quad X \in \mathcal{X}$$

of the mean and ES_p is an ES/$\mathbb{E}$-*mixture*. Note that we allow $\lambda < 0$ in the definition of ES_p^λ, so the ES/$\mathbb{E}$-mixture is not necessarily a monotone risk measure. Define the sets

$$\mathcal{I}_{\rho,\psi} = \bigcap_{X \in \mathcal{X}_+} \mathcal{I}_{\rho,\psi}^X \quad \text{and} \quad \mathcal{I}_{\rho,\psi}^{p,d} = \bigcap_{X \in \mathcal{X}_p^d} \mathcal{I}_{\rho,\psi}^X,$$

which are the intersections of all Pareto optimal contract sets with respect to all models of random losses in $\mathcal{X}_+$ and $\mathcal{X}_p^d$, respectively. Different choices of $\mathcal{I}_{\rho,\psi}$ pin down different forms of ρ and ψ, as we will show below. Obviously, we shall arrive at a narrower class of risk measures as the set of efficient contracts enlarges.

Theorem 4.1. *Suppose that* ρ *and* ψ *are law-invariant convex risk measures. Then:*

(i) $\mathcal{I}_{\rho,\psi} = \mathcal{I}_2$ *if and only if* $\rho = \psi$ *and* ρ *is a convex distortion risk measure on* $\mathcal{X}$*;*

(ii) $\mathcal{I}_{\rho,\psi} = \mathcal{I}_0$ *if and only if* $\rho = \psi = \mathbb{E}$ *on* $\mathcal{X}$.

Our next result, Theorem 4.2, establishes a relationship between deductible contracts and ES, and it is the most sophisticated result of the present paper. The proofs of Theorems 4.1 and 4.2 are technical and rely on additional lemmas, which are presented in Appendix A together with proofs of the theorems.

Theorem 4.2. *Suppose that* ρ *and* ψ *are law-invariant convex risk measures with* $\rho(0) = \psi(0) = 0$. *For any fixed* $d \geq 0$ *and* $p \in [0, 1)$, *we have* $\mathcal{I}_{\rho,\psi}^{p,d} \supset \mathcal{I}_1^d$ *if and only if* $\rho = \psi = \mathrm{ES}_p^\lambda$ *on* $\mathcal{X}_p$ *for some* $\lambda \geq 0$.

We note that, given that the ceded loss functions in the set $\mathcal{I}_1^d$ are Pareto optimal for all insurance losses in the set $\mathcal{X}_p^d$, in Theorem 4.2 we can identify the risk measure adopted by the insured and the insurer as an ES/$\mathbb{E}$-mixture on a larger space of random losses $\mathcal{X}_p$, which does not depend on the deductible level d. In particular, the set $\mathcal{X}_p$ includes all random variables with continuous distributions on bounded supports.

Sets of ceded loss functions		Classes of risk measures
all 1-Lipschitz ceded loss functions	$\Longleftrightarrow$	distortion risk measures
all non-negative ceded loss functions	$\Longleftrightarrow$	the mean
ceded loss functions with deductible form	$\Longleftrightarrow$	an ES/$\mathbb{E}$-mixture

Table 4.1: Connections between sets of ceded loss functions and classes of risk measures

Theorems 4.1 and 4.2 reveal profound connections between common sets of ceded loss functions and common classes of risk measures, as shown in Table 4.1.

As one of the most important economic interpretations of the above results, we show that if the set of Pareto-optimal contracts between the insured and the insurer contains the set $\mathcal{I}_1^d$, then the risk measures of the two parties have to be an ES/$\mathbb{E}$-mixture. Furthermore, if the ES/$\mathbb{E}$-mixture in Theorem 4.2 satisfies lower semicontinuity with respect to almost sure convergence, then it has to be an ES; see Lemma A.3.

If we remove some conditions from the convex risk measures ρ in Theorems 4.1 and 4.2, then we arrive at larger classes of risk measures. For instance, without monotonicity in statement (i) of Theorem 4.1, we expect to arrive at the distortion riskmetrics of Wang et al. (2020a).

4.2 Designing insurance menus

In this section, we discuss economic implications of our characterization results of risk measures. We assume that the risk measures ρ and ψ for the insured and the insurer are coherent throughout this section.

Apart from the link between the common sets of ceded loss functions and the popular classes of risk measures, it is also interesting that all the three sets of Pareto-optimal contracts in Theorems 4.1 and 4.2 lead to the fact that the two risk measures ρ and ψ of the insured and the insurer are the same. In fact, when the risk measures ρ and ψ are coherent, a set of Pareto-optimal contracts with identical risk measures of the two parties is large enough to include all efficient contracts where the insurer is more optimistic than the insured, which can be seen from the next proposition. In this sense, the Pareto-optimal set that we obtain with identical risk measures is the union of Pareto-optimal sets with general risk measures $\rho \geq \psi$.

Proposition 4.2. *We have $\mathcal{I}^X_{\rho,\psi} \subset \mathcal{I}^X_{\psi,\psi}$ for all $X \in \mathcal{X}_+$ and all coherent risk measures ρ and ψ such that $\rho \geq \psi$.*

The relation $\rho \geq \psi$ in Proposition 4.2 indicates that the insured is more pessimistic, or more risk averse, than the insurer in the sense of Pratt (1964). Indeed, the certainty equivalent of any random loss X under the preference described by the coherent risk measure ρ is the risk measure

$\rho(X)$ itself. Therefore, we compare risk aversion of the insured and the insurer through a direct comparison of magnitudes between coherent risk measures ρ and ψ.

In practice, the insurer with the risk measure ψ does not know the risk measure ρ of the insured. Thus it is necessary for the insurer to provide a menu of contracts that is large enough to include all possible efficient contracts that might be chosen by the insured who is more pessimistic than the insurer. Specifically, we consider the following process for the design of insurance menus.

1. An insurer adopts the coherent risk measure ψ as her own risk attitude.
2. The insurer does not have exact information about the risk attitudes of her customers. In other words, the insurer does not know the coherent risk measure ρ held by any insured. However, in order to achieve the deal, the insured should be more pessimistic than the insurer (i.e. $\rho \geq \psi$).
3. Due to incomplete information, the insurer provides a menu of contracts $\mathcal{I}^X_{\psi,\psi} = \bigcup_{\rho\geq\psi} \mathcal{I}^X_{\rho,\psi}$ for a random loss $X \in \mathcal{X}_+$. The set $\mathcal{I}^X_{\psi,\psi}$ is large enough so that Pareto optimality can be obtained for any insured that is more pessimistic than the insurer. The deal can be achieved as long as we have $\rho \geq \psi$ since both parties benefit from the final deal.
4. If the insurer aims to design a "universal" menu of contracts so that Pareto optimality can be achieved for a bundle of random losses, the menu is then obtained by taking intersections of $\mathcal{I}^X_{\psi,\psi}$ with respect to a set of random losses. In this case, the insurer must choose specific classes of risk measures ψ, provided that the "universal" menu of contracts contains some common sets of contracts in the insurance market. Specifically, Table 4.2 illustrates our characterization results.

Pareto-optimal menu		Insurer's risk measure ψ
$\mathcal{I}_{\psi,\psi} = \mathcal{I}_2$	$\iff$	ψ is a distortion risk measure
$\mathcal{I}_{\psi,\psi} = \mathcal{I}_0$	$\iff$	$\psi = \mathbb{E}$
$\mathcal{I}^{p,d}_{\psi,\psi} \supset \mathcal{I}^d_1$	$\iff$	$\psi = \mathrm{ES}^\lambda_p$

Table 4.2: Connections between Pareto-optimal sets of contracts and the insurer's risk measures

5 Concluding remarks

In this paper, the optimal insurance design problem is considered in the sense of Pareto optimality. Unlike existing studies, we solved a characterization problem of the risk measures of the insured and the insurer given the form of the Pareto-optimal contracts, and thus this paper is in an opposite direction to the vast majority of the literature on optimal insurance. As our main finding, we are able to link the ES family, the most popular convex risk measures, to the set of ceded loss functions with a deductible form, commonly seen in insurance practice. It is not our intention to assert that ES dominates other convex risk measures in the insurance market, since there are so many other factors that need to be taken into account. Nevertheless, given the large volume of research based on ES in insurance and actuarial science, we hope that the present paper brings in additional insights on why ES is a natural risk measure to use by the insurer when evaluating risks in the insurance market.

We note that our characterization results can be extended to the multi-player case with multiple insurers. This naturally links our study to the characterization of risk measures in risk sharing problems. Another potential application that can be further developed through our characterization results is that insurance companies may wish to evaluate risk attitudes of their customers based on contracts chosen from provided menus. This research direction requires more experimental studies as well as theoretical justifications. As yet another future direction, viewing the insured and the insurer as two economic agents in a competitive game, characterization problems may be explored via game theoretic approaches.

Acknowledgments

We thank two anonymous referees and an editor for helpful comments, Taizhong Hu for technical suggestions, and Mario Ghossoub for useful discussions. Ruodu Wang is supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03823, RGPAS-2018-522590). Ričardas Zitikis would like to acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-04452).

A Proofs of main results and related technical lemmas

In this appendix, we present proofs of our main results as well as several related lemmas.

A.1 Technical lemmas

We first collect technical lemmas that are related to, or are needed for proving, Theorems 4.1 and 4.2. We note in this regard that some parts of the proofs of the main results needed characterizations without assuming translation invariance. Hence, our next lemma characterizes risk measures ρ and ψ without this assumption and is restricted to the space $\mathcal{X}_0^0$. The lemma was used in the proof of Theorem 4.2.

Lemma A.1. *Suppose that risk measures ρ and ψ are law invariant, monotone, convex and uniformly continuous with respect to L^∞-norm. Then we have the following two characterization* *results.*

(i) The inclusion

$$\bigcap_{X\in\mathcal{X}_0^0} \operatorname*{arg\,min}_{g\in\mathcal{I}}\{\rho(X-g(X))+\psi(g(X))\}\supset\mathcal{I}_2$$

holds if and only if

$$\rho(X)=\psi(X)=\int_0^\infty h(\mathbb{P}(X\geq x))\,\mathrm{d}x \tag{A.1}$$

for all $X\in\mathcal{X}_0^0$, where $h:[0,1]\to[0,\infty)$ is an increasing concave function with $h(0)=0$.[5]

(ii) The inclusion

$$\bigcap_{X\in\mathcal{X}_0^0} \operatorname*{arg\,min}_{g\in\mathcal{I}}\{\rho(X-g(X))+\psi(g(X))\}\supset\mathcal{I}_1$$

holds if and only if $\rho=\psi=\lambda\mathbb{E}$ on $\mathcal{X}_0^0$ for some $\lambda\geq 0$.

Proof. (i) For convenience of the proof for Theorem 4.1, we prove the result on space $\mathcal{X}_+$, and the proof of (i) holds by directly changing $\mathcal{X}_+$ to $\mathcal{X}_0^0$.

"$\Rightarrow$": Suppose that $\mathcal{I}_{\rho,\psi}\supset\mathcal{I}_2$. Let $h_0(x)=0$ and $h_1(x)=x$, $x\geq 0$, the constant zero function and the identity, respectively. Since $h_0,h_1\in\mathcal{I}_2$, we have

$$\rho(X)=\psi(X)=\min_{g\in\mathcal{I}}\left\{\rho(X-g(X))+\psi(g(X))\right\},\quad X\in\mathcal{X}_+.$$

Hence $\rho=\psi$ on $\mathcal{X}_+$ and $\rho(X-f(X))+\rho(f(X))=\rho(X)$ for all $f\in\mathcal{I}_2$ and $X\in\mathcal{X}_+$.

By Proposition 4.5 of Denneberg (1994), for any comonotonic $(Y,Z)\in\mathcal{X}_+^2$ with $Y+Z=X$, there exists $f\in\mathcal{I}_2$ such that $Y=f(X)$ and $Z=X-f(X)$. Since X is arbitrary, we therefore have the equation $\rho(Y)+\rho(Z)=\rho(Y+Z)$ for all comonotonic $(Y,Z)\in\mathcal{X}_+^2$. This shows that ρ is comonotonic-additive on $\mathcal{X}_+$. Thus (A.1) holds by Theorems 1 and 3 of Wang et al. (2020b).

"$\Leftarrow$": Suppose that ρ and ψ satisfy (A.1) on $\mathcal{X}_+$. For all $f\in\mathcal{I}_2$ and $X\in\mathcal{X}_+$, we have by Proposition 4.1 that $(f(X),X-f(X))$ is comonotonic. By comonotonic-additivity of ρ, we have $\rho(X-f(X))+\rho(f(X))=\rho(X)$. Furthermore, due to subadditivity of ρ, we have $f\in\mathcal{I}_{\rho,\rho}$. It follows that $\mathcal{I}_2\subset\mathcal{I}_{\rho,\rho}$.

(ii) The "if" part is straightforward by linearity of the mean. We prove the "only if" part. Since $\mathcal{I}_1\supset\mathcal{I}_2$, by (i), we have $\rho(X)=\psi(X)=\int_0^\infty h(\mathbb{P}(X\geq x))\,\mathrm{d}x$ for all $X\in\mathcal{X}_0^0$. By Theorem 5 of Wang et al. (2020a), there is a finite Borel measure μ on $[0,1]$ such that $\rho(X)=\int_0^1 \mathrm{ES}_\alpha(X)\,\mu(\mathrm{d}\alpha)$ for $X\in\mathcal{X}_0^0$. For all $0<\alpha\leq 1$, there exists differentiable $f\in\mathcal{I}_1$ such that $f'(x)\leq 1$ for all $x\in[0,\mathrm{VaR}_\alpha(X))$ and $f'(x)>1$ for all $x\in[\mathrm{VaR}_\alpha(X),\infty)$. Thus $x\mapsto x-f(x)$ is increasing

[5]Functionals of form (A.1) belong to the family of distortion riskmetrics of Wang et al. (2020a) with increasing distortion functions.

on $[0, \mathrm{VaR}_\alpha(X))$ and decreasing on $[\mathrm{VaR}_\alpha(X), \infty)$ in strict sense. According to Lemma A.3 and Lemma A.7 of Wang and Zitikis (2021), we have a p-tail event A of X and $f(X)$ with

$$\{X > \mathrm{VaR}_\alpha(X)\} \subset A \subset \{X \geq \mathrm{VaR}_\alpha(X)\}$$

such that

$$\mathrm{ES}_\alpha(X) = \mathbb{E}[X|A] \text{ and } \mathrm{ES}_\alpha(f(X)) = \mathbb{E}[f(X)|A].$$

On the other hand, for a p-tail event B of $X - f(X)$ satisfying

$$\{X - f(X) > \mathrm{VaR}_\alpha(X - f(X))\} \subset B \subset \{X - f(X) \geq \mathrm{VaR}_\alpha(X - f(X))\},$$

we have

$$\mathrm{ES}_\alpha(X - f(X)) = \mathbb{E}[X - f(X)|B] > \mathbb{E}[X - f(X)|A].$$

Thus we have

$$\mathrm{ES}_\alpha(f(X)) + \mathrm{ES}_\alpha(X - f(X)) > \mathbb{E}[f(X)|A] + \mathbb{E}[X - f(X)|A] = \mathbb{E}[X|A] = \mathrm{ES}_\alpha(X)$$

and so

$$\begin{aligned} \rho(f(X)) + \rho(X - f(X)) &= \int_0^1 \mathrm{ES}_\alpha(f(X)) + \mathrm{ES}_\alpha(X - f(X))\, \mu(\mathrm{d}\alpha) \\ &> \int_0^1 \mathrm{ES}_\alpha(X)\, \mu(\mathrm{d}\alpha) = \rho(X), \end{aligned}$$

which leads to a contradiction. Hence, $\mu((0, 1]) = 0$ and $\rho(X) = \psi(X) = \lambda\mathbb{E}[X]$ for some $\lambda \geq 0$ and for all $X \in \mathcal{X}_0^0$. $\square$

The next lemma characterizes an ES/$\mathbb{E}$-mixture. The lemma implies that a law-invariant convex risk measure dominated by an ES/$\mathbb{E}$-mixture must be the ES/$\mathbb{E}$-mixture itself provided that it coincides with the ES/$\mathbb{E}$-mixture somewhere. We used the lemma when proving Theorem 4.2.

Lemma A.2. *Let $\rho : \mathcal{X} \to \mathbb{R}$ be a law-invariant convex risk measure. Fix $d \geq 0$ and $p \in (0, 1)$. We have $\rho(X) = \rho((X - d)_+) + \rho(X \wedge d)$ and $\rho(X \vee d) = \lambda \mathrm{ES}_p(X) + (1 - \lambda)d$ for all $X \in \mathcal{X}_p^d$ with $\lambda \in \mathbb{R}$ if and only if $\rho(X) = \lambda \mathrm{ES}_p(X) + (1 - \lambda)\mathrm{ES}_p^-(X)$ for all $X \in \mathcal{X}_p^d$ with $\lambda \geq 1 - p$.*

Proof. The "if" part follows immediately from the definitions of ES_p and ES_p^-. Hence, we prove the "only if" part.

Since ρ is a law-invariant convex risk measure, for all $X \in \mathcal{X}_p^d$ we write

$$\rho(X) = \sup_{Z \in \mathcal{Q}} \{\mathbb{E}[ZX] + V(Z)\},$$

where $\mathcal{Q}$ is a set of Radon-Nikodym derivatives and V is a mapping from $\mathcal{Q}$ to $[-\infty, 0]$ (see e.g., Jouini et al. (2006)). We first show that $Z \le \lambda/(1-p)$ for all $Z \in \mathcal{Q}$. Assume for the sake of contradiction that $\mathbb{P}(Z' > \lambda/(1-p)) > 0$ for some $Z' \in \mathcal{Q}$. Take $A \subset \{Z' > \lambda/(1-p)\}$ and $Y = \mathbb{1}_A(d+1)\gamma + \mathbb{1}_B(d+1)$ for $\gamma > 1$, where $\mathbb{P}(A \cup B) = 1-p$ and $A \cap B = \emptyset$. It is clear that $Y \in \mathcal{X}_p^d$. We have

$$\begin{aligned}\sup_{Z \in \mathcal{Q}} \{\mathbb{E}[ZY] + V(Z)\} &\ge \mathbb{E}[Z'(\mathbb{1}_A(d+1)\gamma + \mathbb{1}_B(d+1))] + V(Z') \\ &\ge (d+1)\gamma \mathbb{E}[Z'\mathbb{1}_A] + V(Z') \\ &= \mathbb{E}[Z'|A]\mathbb{E}[\mathbb{1}_A(d+1)\gamma] + V(Z').\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\lambda \mathrm{ES}_p(Y) + (1-\lambda)d &= \lambda \mathrm{ES}_p(\mathbb{1}_A(d+1)\gamma + \mathbb{1}_B(d+1)) + (1-\lambda)d \\ &= \frac{\lambda}{1-p}\mathbb{E}[\mathbb{1}_A(d+1)\gamma + \mathbb{1}_B(d+1)] + (1-\lambda)d.\end{aligned}$$

Since $\mathbb{E}[Z'|A] > \lambda/(1-p)$, we have

$$\lim_{\gamma \to \infty} (\mathbb{E}[Z'|A]\mathbb{E}[\mathbb{1}_A(d+1)\gamma] + V(Z')) > \lim_{\gamma \to \infty} (\lambda \mathrm{ES}_p(Y) + (1-\lambda)d),$$

which contradicts the assumption that $\rho(X) \le \lambda \mathrm{ES}_p(X) + (1-\lambda)d$ for all $X \in \mathcal{X}_p^d$. Therefore, we have $Z \le \lambda/(1-p)$ for all $Z \in \mathcal{Q}$. On the other hand, since $\mathbb{E}[Z] = 1$, we have $\lambda/(1-p) \ge 1$ and thus $\lambda \ge 1-p$.

We next show that $\rho(X) = \lambda \mathrm{ES}_p(X) + (1-\lambda)\mathrm{ES}_p^-(X)$ for all $X \in \mathcal{X}_p^d$. Note that $\{X > d\}$ is a common p-tail event of X and $X \vee d$. We have $\mathrm{ES}_p(X) = \mathrm{ES}_p(X \vee d)$ and

$$d = \frac{1}{p}\mathbb{E}[(X \vee d)\mathbb{1}_{\{X \le d\}}] = \mathrm{ES}_p^-(X \vee d).$$

It follows that

$$\begin{aligned}\sup_{Z \in \mathcal{Q}} \{\mathbb{E}[Z(X \vee d)] + V(Z)\} &= \rho(X \vee d) \\ &= \lambda \mathrm{ES}_p(X) + (1-\lambda)d = \lambda \mathrm{ES}_p(X \vee d) + (1-\lambda)\mathrm{ES}_p^-(X \vee d).\end{aligned}$$

For $X_1, X_2, \ldots \in \mathcal{X}_p^d$ and $X_n \downarrow X$, since Z is non-negative and bounded from above by $1/(1-p)$, the dominated convergence theorem implies

$$\lim_{n \to \infty} \sup_{Z \in \mathcal{Q}} \{\mathbb{E}[ZX_n] + V(Z)\} = \sup_{Z \in \mathcal{Q}} \{\mathbb{E}[ZX] + V(Z)\},$$

which means that ρ is continuous from above. Hence,

$$\rho(X) = \max_{Z \in \mathcal{Q}} \{\mathbb{E}[ZX] + V(Z)\}$$

for all $X \in \mathcal{X}_p^d$; see e.g., Corollary 4.35 of Föllmer and Schied (2016). It follows that there exists $Z_0 \in \mathcal{Q}$ such that

$$\mathbb{E}[Z_0(X \vee d)] + V(Z_0) = \frac{\lambda}{1-p}\mathbb{E}[(X \vee d)\mathbb{1}_{\{X>d\}}] + \frac{1-\lambda}{p}\mathbb{E}[(X \vee d)\mathbb{1}_{\{X \le d\}}]. \tag{A.2}$$

We claim that $Z_0 = \lambda \mathbb{1}_{\{X>d\}}/(1-p) + (1-\lambda)\mathbb{1}_{\{X \le d\}}/p$. Indeed, assume for the sake of contradiction that $Z_0 \neq \lambda \mathbb{1}_{\{X>d\}}/(1-p) + (1-\lambda)\mathbb{1}_{\{X \le d\}}/p$. Since

$$\mathbb{E}[Z_0] = 1 = \mathbb{E}\left[\frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} + \frac{1-\lambda}{p}\mathbb{1}_{\{X \le d\}}\right],$$

we have

$$\mathbb{P}\left(\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X \le d\}}\right)_+ > 0\right) > 0$$

and

$$\mathbb{P}\left(\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X \le d\}}\right)_- > 0\right) > 0.$$

Note that $\lambda/(1-p) \ge 1 \ge (1-\lambda)/p$. Hence,

$$\left\{\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X \le d\}}\right)_+ > 0\right\} \subset \{X \le d\}.$$

We also note that

$$\left\{\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X \le d\}}\right)_- > 0\right\} \cap \{X > d\} \neq \emptyset.$$

Otherwise, we must have $Z_0 = \lambda/(1-p)$ and

$$\mathbb{P}\left(\left(Z_0 - \frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}} - \frac{1-\lambda}{p}\mathbb{1}_{\{X \le d\}}\right)_- > 0\right) = 0,$$

which leads to contradiction. These considerations imply that

$$
\begin{aligned}
&\mathbb{E}\left[\left(Z_0-\frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}}-\frac{1-\lambda}{p}\mathbb{1}_{\{X\le d\}}\right)(X\vee d)\right]\\
&=\mathbb{E}\left[\left(Z_0-\frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}}-\frac{1-\lambda}{p}\mathbb{1}_{\{X\le d\}}\right)_+(X\vee d)\right]\\
&\quad-\mathbb{E}\left[\left(Z_0-\frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}}-\frac{1-\lambda}{p}\mathbb{1}_{\{X\le d\}}\right)_-(X\vee d)\right]\\
&<d\left(\mathbb{E}\left[\left(Z_0-\frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}}-\frac{1-\lambda}{p}\mathbb{1}_{\{X\le d\}}\right)_+\right]-\mathbb{E}\left[\left(Z_0-\frac{\lambda}{1-p}\mathbb{1}_{\{X>d\}}-\frac{1-\lambda}{p}\mathbb{1}_{\{X\le d\}}\right)_-\right]\right)\\
&=0,
\end{aligned}
$$

which contradicts equation (A.2). Therefore, we must have $Z_0=\lambda\mathbb{1}_{\{X>d\}}/(1-p)+(1-\lambda)\mathbb{1}_{\{X\le d\}}/p$. Hence, $Z_0=\lambda\mathbb{1}_{\{X>d\}}/(1-p)+(1-\lambda)\mathbb{1}_{\{X\le d\}}/p\in\mathcal{Q}$ and $V(Z_0)=0$. It follows that

$$
\sup_{Z\in\mathcal{Q}}\{\mathbb{E}[ZX]+V(Z)\}\ge\mathbb{E}\left[\frac{\lambda}{1-p}X\mathbb{1}_{\{X>d\}}+\frac{1-\lambda}{p}X\mathbb{1}_{\{X\le d\}}\right]=\lambda\mathrm{ES}_p(X)+(1-\lambda)\mathrm{ES}_p^-(X).
$$

On the other hand, we have

$$
\rho(X)\le\lambda\mathrm{ES}_p(X)+(1-\lambda)d=\gamma\mathrm{ES}_p(X)+(1-\gamma)\mathrm{ES}_p^-(X),
$$

for some $1-p\le\lambda\le\gamma\le1$ since $\mathrm{ES}_p^-(X)\le d\le\mathrm{ES}_p(X)$. Hence, there exists $\lambda\le\lambda'\le\gamma$, such that $\rho(X)=\lambda'\mathrm{ES}_p(X)+(1-\lambda')\mathrm{ES}_p^-(X)$.

Take $X_m=X\mathbb{1}_{\{X\le d\}}+(X+m)\mathbb{1}_{\{X>d\}}$ for $m>0$. We have $X_m\in\mathcal{X}_p^d$. For some $\lambda_m\in[\lambda,1]$,

$$
\rho(X_m)=\lambda_m\mathrm{ES}_p(X_m)+(1-\lambda_m)\mathrm{ES}_p^-(X_m)=\lambda_m\mathrm{ES}_p(X)+\lambda_m m+(1-\lambda_m)\mathrm{ES}_p^-(X).\tag{A.3}
$$

Since $\rho(X_m\vee d)=\lambda\mathrm{ES}_p(X)+\lambda m+(1-\lambda)d$, this implies that there exists $m>0$ such that $\lambda_m=\lambda$. Indeed, otherwise we can take $m\to\infty$ and have a contradiction to $\rho(X_m)\le\rho(X_m\vee d)$ by monotonicity of ρ. On the other hand, for m such that $\lambda_m=\lambda$, we have

$$
\begin{aligned}
\rho(X_m)=\rho(X_m\vee d)-d+\rho(X_m\wedge d)&=\lambda\mathrm{ES}_p(X)+\lambda m-\lambda d+\rho(X\wedge d)\\
&=\rho(X)+\lambda m=\lambda'\mathrm{ES}_p(X)+(1-\lambda')\mathrm{ES}_p^-(X)+\lambda m.
\end{aligned}\tag{A.4}
$$

Equations (A.3) and (A.4), together with $\lambda_m=\lambda$, yield that $\lambda'=\lambda$ for all $X\in\mathcal{X}_p^d$. This completes the proof. $\square$

Finally, we give a lemma on properties of ES/$\mathbb{E}$-mixtures that can precisely pin down the family of ES within the class of ES/$\mathbb{E}$-mixtures obtained in Theorem 4.2.

Lemma A.3. *For an* $\mathrm{ES}/\mathbb{E}$*-mixture* $\rho = \lambda \mathrm{ES}_p + (1-\lambda)\mathbb{E}$*, we have the following statements:*

(i) ρ *is lower semicontinuous with respect to almost sure convergence if and only if* $\lambda \geq 1$*;*

(ii) ρ *is convex if and only if* $\lambda \geq 0$*;*

(iii) ρ *is monotone if and only if* $\lambda \in [1-1/p, 1]$.

In particular, ρ *is monotone and lower semicontinuous with respect to almost sure convergence if and only if it is* ES_p.

Proof. (i) Suppose that $\lambda < 1$. Let $X_k = -k\mathbb{1}_{\{U<1/k\}}$, where $U \sim \mathrm{U}[0,1]$. Clearly, $X_k \to 0$ almost surely as $k \to \infty$, $\mathbb{E}[X_k] = -1$, and $\mathrm{ES}_p(X_k) = 0$ for $k > 1/p$. Therefore,

$$\liminf_{k\to\infty} \left((1-\lambda)\mathbb{E}[X_k] + \lambda \mathrm{ES}_p(X_k)\right) = -(1-\lambda) < 0 = \rho(0),$$

contradicting lower semicontinuity.

(ii) We note that ρ is a signed Choquet integral of Wang et al. (2020b) with the (not necessarily increasing) distortion function

$$h(t) = \lambda \left(\frac{t}{1-p} \wedge 1\right) + (1-\lambda)t, \quad t \in [0,1].$$

By Theorem 3 of Wang et al. (2020b), ρ is convex if and only if h is concave. It is straightforward to verify that h is concave if and only if $\lambda \geq 0$.

(iii) By Lemma 1 (i) of Wang et al. (2020b), ρ is monotone if and only if h is increasing. Clearly, $\lambda > 1$ implies that h is strictly decreasing on $(1-p, 1]$. For $\lambda \leq 1$, increasing monotonicity of h is equivalent to

$$\frac{\lambda}{1-p} + 1 - \lambda \geq 0 \iff \lambda \geq 1 - \frac{1}{p}.$$

Hence, ρ is monotone if and only if $\lambda \in [1-1/p, 1]$. □

A.2 Proofs of all results

Proof of Proposition 3.1. "(i)⇒(ii)": This is straightforward by taking $\pi : h \mapsto \psi(h(X))$.

"(ii)⇒(iii)": Suppose that $f \in \mathcal{I}$ is Pareto optimal for $\pi : h \mapsto \psi(h(X))$. Assume for the sake of contradiction that $f \notin \mathcal{I}^X_{\rho,\psi}$. It follows that there exists $g \in \mathcal{I}$, such that

$$\rho(X - g(X)) + \psi(g(X)) < \rho(X - f(X)) + \psi(f(X)).$$

By translation invariance of ρ and ψ, we have

$$\begin{aligned}\rho(X-g(X)+\pi(g)) &= \rho(X-g(X))+\psi(g(X)) \\ &< \rho(X-f(X))+\psi(f(X)) = \rho(X-f(X)+\pi(f))\end{aligned}$$

and

$$\psi(g(X)-\pi(g)) = \psi(g(X)-\psi(g(X))) = 0 = \psi(f(X)-\pi(f)),$$

which leads to a contradiction to Pareto optimality of f. Therefore, $f \in \mathcal{I}^X_{\rho,\psi}$.

"(iii)⇒(i)": Suppose that $f \in \mathcal{I}^X_{\rho,\psi}$. Assume for the sake of contradiction that f is not Pareto optimal for some $\pi : \mathcal{I} \to \mathbb{R}$. It follows that there exists $g \in \mathcal{I}$ such that

$$\rho(X-g(X)+\pi(g)) \le \rho(X-f(X)+\pi(f))$$

and

$$\psi(g(X)-\pi(g)) \le \psi(f(X)-\pi(f)),$$

with at least one of the above two inequalities being strict. Hence,

$$\rho(X-g(X)+\pi(g))+\psi(g(X)-\pi(g)) < \rho(X-f(X)+\pi(f))+\psi(f(X)-\pi(f)),$$

which contradicts the fact that $f \in \mathcal{I}^X_{\rho,\psi}$. Therefore, the function f is Pareto optimal for all $\pi : \mathcal{I} \to \mathbb{R}$. □

Proof of Proposition 4.1. (i) Suppose that $f \in \mathcal{I}_2$. Define the function g by $g(x) = x - f(x)$ for $x \in [0,\infty)$. For all $X \in \mathcal{X}_+$, we have $X - f(X) = g(X)$. Since $f \in \mathcal{I}_2$, the function g is increasing and $(f(X), g(X))$ is comonotonic.

(ii) Suppose that $f \in \mathcal{I}^d_1$ for $d > 0$. For all $X \in \mathcal{X}^d_p$, the set $\{X > d\}$ is a common tail event of $f(X)$ and $X - f(X)$ by the definitions of the tail event and the set $\mathcal{I}^d_1$. Also note that $\mathbb{P}(X > d) = 1 - p$. Therefore, $(f(X), X - f(X))$ is p-concentrated. □

Here we present the proof of Theorem 4.2 first because it is useful for the proof of Theorem 4.1.

Proof of Theorem 4.2. "⇐": For all $f \in \mathcal{I}^d_1$, note that $(f(X), X - f(X))$ is p-concentrated for all $X \in \mathcal{X}^d_p$ by Proposition 4.1. By p-additivity of ES_p (see Wang and Zitikis (2021)), we have $\mathrm{ES}_p(X - f(X)) + \mathrm{ES}_p(f(X)) = \mathrm{ES}_p(X)$ and thus $f \in \mathcal{I}^{d,p}_{\mathrm{ES}_p,\mathrm{ES}_p}$. Hence $\mathcal{I}^{d,p}_{\mathrm{ES}_p,\mathrm{ES}_p} \supset \mathcal{I}^d_1$.

"⇒": It suffices to show that $\rho = \psi = \mathrm{ES}^\lambda_p$ on $\mathcal{X}^d_p$ for some $\lambda \ge 0$, and that $\rho = \psi = \mathrm{ES}^\lambda_p$ on $\mathcal{X}_p$ holds due to translation invariance of ρ and ψ. Write $h_d(x) = (x-d)_+$, $x \ge 0$, for all $d \ge 0$ and recall that $h_0(x) = 0$, $x \geq 0$. Since $h_0, h_d \in \mathcal{I}_1^d$, we have

$$\rho(X) = \rho(X \wedge d) + \psi((X-d)_+) = \min_{g \in \mathcal{I}} \{\rho(X - g(X)) + \psi(g(X))\} \tag{A.5}$$

for all $X \in \mathcal{X}_p^d$.

We first prove the case when $d = p = 0$. We know from Lemma A.1 that $\rho(X) = \psi(X) = \lambda \mathbb{E}[X]$ for some $\lambda \geq 0$ and for all $X \in \mathcal{X}_0^0$. Since ρ is translation invariant and $X + c \in \mathcal{X}_0^0$ for all $X \in \mathcal{X}_0^0$ and $c \geq 0$, we have

$$\lambda \mathbb{E}[X] + c = \rho(X) + c = \rho(X + c) = \lambda \mathbb{E}[X + c] = \lambda \mathbb{E}[X] + \lambda c.$$

It follows that $\lambda = 1$.

We now prove the case when $d = 0$ and $p \in (0, 1)$. We known from statement (A.5) that $\rho = \psi$ on X_p^0. For all $X \in \mathcal{X}_0^0$, we define $\phi(X) = \rho(X\mathbb{1}_A)$ by taking an event A independent of X with $\mathbb{P}(A) = 1 - p$ (a specific choice of A does not matter since ρ is law invariant). It is clear that ϕ is law invariant, monotone, convex and uniformly continuous with respect to L^∞-norm. Note that for all $X \in \mathcal{X}_0^0$ and all events B and C independent of X with $\mathbb{P}(B) = \mathbb{P}(C) = 1 - p$, we have $X\mathbb{1}_B \stackrel{\mathrm{d}}{=} X\mathbb{1}_C$. Hence, $\phi(X) = \rho(X\mathbb{1}_B) = \rho(X\mathbb{1}_C)$ and thus ϕ is well defined. Since $X\mathbb{1}_A \in \mathcal{X}_p^0$ and $\mathcal{I}_{\rho,\psi}^{0,p} \supset \mathcal{I}_1$, we have

$$\begin{aligned}
\phi(f(X)) + \phi(X - f(X)) &= \rho(f(X)\mathbb{1}_A) + \rho((X - f(X))\mathbb{1}_A) \\
&= \rho(f(X\mathbb{1}_A)) + \rho(X\mathbb{1}_A - f(X\mathbb{1}_A)) = \rho(X\mathbb{1}_A) = \phi(X)
\end{aligned}$$

for all $f \in \mathcal{I}_1$ and $X \in \mathcal{X}_0^0$. It follows from Lemma A.1 that $\phi(X) = \lambda \mathbb{E}[X]$ for some $\lambda \geq 0$ and for all $X \in \mathcal{X}_0^0$. For all $X \in \mathcal{X}_p^0$, we take any random variable Y such that $Y \stackrel{\mathrm{d}}{=} X | X > 0$. We have $Y \in \mathcal{X}_0^0$ and $X\mathbb{1}_{\{X>0\}} \stackrel{\mathrm{d}}{=} Y\mathbb{1}_A$. Thus

$$\rho(X\mathbb{1}_{\{X>0\}}) = \rho(Y\mathbb{1}_A) = \lambda \mathbb{E}[Y] = \lambda \mathrm{ES}_p(X).$$

It follows that

$$\rho(X) = \psi(X) = \rho(X\mathbb{1}_{\{X>0\}}) + \rho(X\mathbb{1}_{\{X=0\}}) = \lambda \mathrm{ES}_p(X) \tag{A.6}$$

for all $X \in \mathcal{X}_p^0$. Note that for all $X \in \mathcal{X}_p^0$,

$$\mathrm{ES}_p(X) = \frac{1}{1-p} \mathbb{E}[X\mathbb{1}_{X>0}] = \frac{1}{1-p} \mathbb{E}[X].$$

Hence, we have

$$\rho(X) = \lambda' \mathrm{ES}_p(X) + (1 - \lambda') \mathbb{E}[X]$$

for all $X \in \mathcal{X}_p^0$, where $\lambda' = (\lambda - 1 + p)/p$. By equation (A.6) and Lemma A.2, we have $\lambda \geq 1 - p$ and thus $\lambda' \geq 0$.

Next, we prove the case when $d > 0$ and $p = 0$. For all $X \in \mathcal{X}_0^0$, we have $X + d \in \mathcal{X}_0^d$. We obtain from $\mathcal{I}_{\rho,\psi}^{d,0} \supset \mathcal{I}_1^d$ that

$$\rho(X + d - f(X + d)) + \psi(f(X + d)) = \rho(X + d) \tag{A.7}$$

for all $f \in \mathcal{I}_1^d$. Take those f that are of the form $f(x) = g(x - d)$ for any $g \in \mathcal{I}_1$ and all $x \geq d$. Noting that ρ is translation invariant, we have

$$\rho(X - g(X)) + \psi(g(X)) = \rho(X) \tag{A.8}$$

for all $g \in \mathcal{I}_1$. Hence, $\rho(X) = \psi(X) = \lambda\mathbb{E}[X] = \lambda \mathrm{ES}_0(X)$ for some $\lambda \geq 0$ and for all $X \in \mathcal{X}_0^0$ by Lemma A.1. Since ρ is translation invariant, we have $\lambda = 1$.

We finally prove the case when $d > 0$ and $p \in (0, 1)$. For all $X \in \mathcal{X}_p^0$, we have $X + d \in \mathcal{X}_p^d$. Following similar arguments as those we used to derive equations (A.7) and (A.8), we obtain

$$\rho(X - g(X)) + \psi(g(X)) = \rho(X)$$

for all $g \in \mathcal{I}_1$. Hence, $\rho(X) = \psi(X) = \lambda \mathrm{ES}_p(X)$ for some $\lambda \geq 1 - p$ and for all $X \in \mathcal{X}_p^0$ by equation (A.6). For all $X \in \mathcal{X}_p^d$, we have $(X - d)_+ \in \mathcal{X}_p^0$. Therefore,

$$\rho((X - d)_+) = \psi((X - d)_+) = \lambda \mathrm{ES}_p[(X - d)_+] = \lambda(\mathrm{ES}_p(X) - d).$$

Hence, $\rho(X \vee d) = \rho((X - d)_+ + d) = \lambda \mathrm{ES}_p(X) + (1 - \lambda)d$ and $\psi(X \vee d) = \lambda \mathrm{ES}_p(X) + (1 - \lambda)d$. By Lemma A.2, we have $\rho(X) = \psi(X) = \lambda \mathrm{ES}_p(X) + (1 - \lambda)\mathrm{ES}_p^-(X)$ for all $X \in \mathcal{X}_p^d$. Since

$$(1 - p)\mathrm{ES}_p(X) + p\mathrm{ES}_p^-(X) = \mathbb{E}[X],$$

we have $\rho(X) = \psi(X) = \gamma \mathrm{ES}_p(X) + (1 - \gamma)\mathbb{E}[X]$, where $\gamma = 1 - (1 - \lambda)/p \geq 0$. $\square$

Proof of Theorem 4.1. Let $h_0(x) = 0$ and $h_1(x) = x$, $x \geq 0$, the constant zero function and the identity, respectively.

(i) "$\Rightarrow$": Suppose that $\mathcal{I}_{\rho,\psi} = \mathcal{I}_2$. By the proof of Lemma A.1 (i) and translation invariance of ρ and ψ, we have $\rho = \psi$ on $\mathcal{X}$ and ρ is comonotonic-additive on $\mathcal{X}$. Moreover, we know that ρ is uniformly continuous with respect to L^∞-norm since ρ is monetary, and ρ is law invariant. Hence, ρ is a convex distortion risk measure on $\mathcal{X}$ (see e.g., Kusuoka (2001)).

"$\Leftarrow$": Suppose that $\rho = \psi$ is a convex distortion risk measure on $\mathcal{X}$. We will prove that

$\mathcal{I}_{\rho,\rho} = \mathcal{I}_2$. Since ρ is a convex distortion risk measure, it is also coherent by e.g., Corollary 1 of Wang et al. (2020a); see Acerbi (2002). Following the same logic as the proof of Lemma A.1 (i), we have $\mathcal{I}_2 \subset \mathcal{I}_{\rho,\rho}$.

We next prove that $\mathcal{I}_{\rho,\rho} \subset \mathcal{I}_2$. For each $f \notin \mathcal{I}_2$, we will show that there exists $X \in \mathcal{X}_+$ such that $\rho(X - f(X)) + \rho(f(X)) > \rho(X)$. Indeed, there exists $0 \le x < y$, such that $|f(y) - f(x)| > y - x$. It is clear that $f(x) \neq f(y)$. Since ρ is a coherent distortion risk measure, there exists a Borel measure μ on $[0,1]$ such that $\rho = \int_0^1 \mathrm{ES}_t \,\mathrm{d}\mu(t)$ on $\mathcal{X}$. Take $X = x\mathbb{1}_A + y\mathbb{1}_{A^c}$ where $\mathbb{P}(A) = 1/2$. If $f(x) < f(y)$, then

$$\mathrm{ES}_t(X) = \begin{cases} \frac{(1-2t)x+y}{2-2t}, & 0 \le t \le 1/2, \\ y, & 1/2 < t < 1, \end{cases}$$

$$\mathrm{ES}_t(f(X)) = \begin{cases} \frac{(1-2t)f(x)+f(y)}{2-2t}, & 0 \le t \le 1/2, \\ f(y), & 1/2 < t < 1, \end{cases}$$

$$\mathrm{ES}_t(X - f(X)) = \begin{cases} \frac{x-f(x)+(1-2t)(y-f(y))}{2-2t}, & 0 \le t \le 1/2, \\ x - f(x), & 1/2 < t < 1. \end{cases}$$

Hence,

$$\begin{aligned} &\frac{\rho(X - f(X)) + \rho(f(X)) - \rho(X)}{y - x} \\ &= \int_0^{1/2} \frac{t}{1-t}\left(\frac{f(y) - f(x)}{y - x} - 1\right) \mathrm{d}\mu(t) + \int_{1/2}^1 \frac{f(y) - f(x)}{y - x} - 1 \,\mathrm{d}\mu(t) > 0. \end{aligned}$$

Similarly, if $f(x) > f(y)$, then we have

$$\mathrm{ES}_t(X) = \begin{cases} \frac{(1-2t)x+y}{2-2t}, & 0 \le t \le 1/2, \\ y, & 1/2 < t < 1, \end{cases}$$

$$\mathrm{ES}_t(f(X)) = \begin{cases} \frac{f(x)+(1-2t)f(y)}{2-2t}, & 0 \le t \le 1/2, \\ f(x), & 1/2 < t < 1, \end{cases}$$

$$\mathrm{ES}_t(X - f(X)) = \begin{cases} \frac{(1-2t)(x-f(x))+y-f(y)}{2-2t}, & 0 \le t \le 1/2, \\ y - f(y), & 1/2 < t < 1, \end{cases}$$

and thus

$$\begin{aligned} &\rho(X - f(X)) + \rho(f(X)) - \rho(X) \\ &= \int_0^{1/2} \frac{t}{1-t}\,(f(x) - f(y))\,\mathrm{d}\mu(t) + \int_{1/2}^1 f(x) - f(y)\,\mathrm{d}\mu(t) > 0. \end{aligned}$$

Therefore, $\mathcal{I}_{\rho,\rho} \subset \mathcal{I}_2$ and thus $\mathcal{I}_{\rho,\rho} = \mathcal{I}_2$.

(ii) The "if" part is straightforward by linearity of the mean. Hence, we prove the "only if" part. Similar to (i), since $h_0, h_1 \in \mathcal{I}_0$, we have by translation invariance of ρ and ψ that $\rho = \psi$ on

$\mathcal{X}$. Since $\mathcal{I}_1 \subset \mathcal{I}_0$ and $\mathcal{X}_0^0 \subset \mathcal{X}_+$, we know from Theorem 4.2 that $\rho(X) = \mathbb{E}[X]$ for all $X \in \mathcal{X}_0^0$. Since $X \in \mathcal{X}$ is bounded, we take $c > 0$ such that $X + c \in \mathcal{X}_0^0$. It follows that $\rho(X + c) = \mathbb{E}[X + c]$. Hence, translation invariance of ρ implies $\rho(X) = \mathbb{E}[X]$. $\square$

Proof of Proposition 4.2. Take any $X \in \mathcal{X}_+$ and coherent risk measures $\rho, \psi : \mathcal{X} \to \mathbb{R}$ with $\rho \geq \psi$. For all $f \notin \mathcal{I}_{\psi,\psi}^X$, we have

$$\rho(X - f(X)) + \psi(f(X)) \geq \psi(X - f(X)) + \psi(f(X)) > \psi(X),$$

where the last inequality is due to subadditivity of ψ. With $h_1(x) = x$, $x \geq 0$, which belongs to $\mathcal{I}$, we have $\rho(X - h_1(X)) + \psi(h_1(X)) = \psi(X)$ and thus

$$\min_{g \in \mathcal{I}} \{\rho(X - g(X)) + \psi(g(X))\} \leq \psi(X).$$

It follows that $f \notin \mathcal{I}_{\rho,\psi}^X$ and therefore $\mathcal{I}_{\rho,\psi}^X \subset \mathcal{I}_{\psi,\psi}^X$. $\square$

References

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

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, **53**(5), 941–973.

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

Barrieu, P. and Scandolo, G. (2008). General Pareto optimal allocations and applications to multi-period risks. *ASTIN Bulletin*, **38**(1), 105–136.

BCBS (2016). Basel Committee on Banking Supervision (Jan. 2016). *Minimum Capital Requirements for Market Risk.* Bank for International Settlements. https://www.bis.org/bcbs/publ/d352.pdf

BCBS (2019). Basel Committee on Banking Supervision (Jan. 2019; revised Feb. 2019). *Minimum Capital Requirements for Market Risk.* Bank for International Settlements. https://www.bis.org/bcbs/publ/d457.pdf

Bernard, C. and Tian, W. (2009). Optimal reinsurance arrangements under tail risk measures. *Journal of Risk and Insurance*, **76**(3), 709–725.

Braun, M. and Muermann, A. (2004). The impact of regret on the demand for insurance. *Journal of Risk and Insurance*, **71**(4), 737–767.

Cai, J. and Chi, Y. (2020). Optimal reinsurance designs based on risk measures: A review. *Statistical Theory and Related Fields*, **4**(1), 1–13.

Cai, J., Liu, H. and Wang, R. (2017). Pareto-optimal reinsurance arrangements under general model settings. *Insurance: Mathematics and Economics*, **77**, 24–37.

Cai, J. and Tan, K. S. (2007). Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measures. *ASTIN Bulletin*, **37**(1), 93–112.

Cai, J., Tan, K. S., Weng, C. and Zhang, Y. (2008). Optimal reinsurance under VaR and CTE risk measures. *Insurance: Mathematics and Economics*, **43**(1), 185–196.

Chen, Y. (2021). Optimal reinsurance from the viewpoints of both an insurer and a reinsurer under the CVaR risk measure and Vajda condition. *ASTIN Bulletin*, **51**(2), 631–659.

Chi, Y. and Weng, C. (2013). Optimal reinsurance subject to Vajda condition. *Insurance: Mathematics and Economics*, **53**(1), 179–189.

Cui, W., Yang, J. and Wu, L. (2013). Optimal reinsurance minimizing the distortion risk measure under general reinsurance premium principles. *Insurance: Mathematics and Economics*, **53**(1), 74–85.

Cummins, J. D. and Mahul, O. (2004). The demand for insurance with an upper limit on coverage. *Journal of Risk and Insurance*, **71**(2), 253–264.

Dana, R. A. and Scarsini, M. (2007). Optimal risk sharing with background risk. *Journal of Economic Theory*, **133**(1), 152–176.

Denneberg, D. (1994). *Non-additive Measure and Integral.* Springer Science & Business Media.

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

Embrechts, P., Mao, T., Wang, Q. and Wang, R. (2021). Bayes risk, elicitability, and the Expected Shortfall. *Mathematical Finance*, **31**, 1190–1217.

Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, **19**(4), 763–790.

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. Fourth Edition.

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

Gollier, C. (1996). Optimum insurance of approximate losses. *Journal of Risk and Insurance*, **63**(3), 369–380.

Gollier, C. and Schlesinger, H. (1996). Arrow's theorem on the optimality of deductibles: A stochastic dominance approach. *Economic Theory*, **7**(2), 359–363.

Han, X., Wang, B., Wang, R. and Wu, Q. (2021). Risk concentration and the mean-Expected Shortfall criterion. *arXiv*: 2108.05066.

Hofmann, A., Häfen, O. V. and Nell, M. (2019). Optimal insurance policy indemnity schedules with policyholders' limited liability and background risk. *Journal of Risk and Insurance*, **86**(4), 973–988.

Huberman, G., Mayers, D. and Smith Jr, C. W. (1983). Optimal insurance policy indemnity schedules. *The Bell Journal of Economics*, 415–426.

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

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

Lo, A., Tang, Q. and Tang, Z. (2021). Universally marketable insurance under multivariate mixtures. *ASTIN Bulletin*, **51**(1), 221–243.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, **32**, 122–136.

Raviv, A. (1979). The design of an optimal insurance policy. *American Economic Review*, **69**(1), 84–96.

Schlesinger, H. (1981). The optimal level of deductibility in insurance contracts. *Journal of Risk and Insurance*, **48**(3), 465–481.

Schlesinger, H. (1997). Insurance demand without the expected-utility paradigm. *Journal of Risk and Insurance*, **64**(1), 19–39.

Vajda, S. (1962). Minimum variance reinsurance. *ASTIN Bulletin*, **2**(2), 257–260.

Wang, Q., Wang, R. and Wei, Y. (2020a). Distortion riskmetrics on general spaces. *ASTIN Bulletin*, **50**(4), 827–851.

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

Wang, R. and Zitikis, R. (2020). Weak comonotonicity. *European Journal of Operational Research*, **282**(1), 386–397.

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

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

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