Convex risk functionals: representation and applications

Fangda Liu* Jun Cai† Christiane Lemieux‡ Ruodu Wang§

This version: October 24, 2019

Abstract

We introduce the family of law-invariant convex risk functionals, which includes a wide majority of practically used convex risk measures and deviation measures. We obtain a unified representation theorem for this family of functionals. Two related optimization problems are studied. In the first application, we determine worst-case values of a law-invariant convex risk functional when the mean and a higher moment such as the variance of a risk are known. Second, we consider its application in optimal reinsurance design for an insurer. With the help of the representation theorem, we can show the existence and the form of optimal solutions.

Key-words: Law-invariant convex risk functional, dual representation, robust evaluation, optimal reinsurance design, budget constraint

1 Law-invariant convex risk functionals

In the last decades, risk measures (Artzner et al. (1999), Föllmer and Schied (2002)) and deviation measures (Rockafellar et al. (2006)) have been popular in banking and finance for various purposes, such as calculating solvency capital reserves, pricing of insurance risks, performance analysis, and internal risk management. Different classes of axioms are proposed for risk measures and deviation measures in the literature. In this paper, we propose a general class of functionals, termed *convex risk functionals*, to unify risk measures and deviation measures in the literature.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an atomless probability space and $\mathcal{L}^p$, $p \in [0, \infty)$ be the set of all random variables with finite p-th moment and $\mathcal{L}^\infty$ be the set of essentially bounded random variables. Each random variable represents a random risk in the future. For $X \in \mathcal{L}^\infty$, its $\mathcal{L}^\infty$-norm is defined as $\|X\|_\infty = \sup\{x \in \mathbb{R} : \mathbb{P}(|X| > x) > 0\}$. We use $\xrightarrow{\mathbf{P}}$ to denote convergence in probability. Our key concept, the convex risk functionals are defined below.

Definition 1.1 (*Law-invariant Convex Risk Functionals*)**.** Fix $p \in [1, \infty]$. A mapping $\rho : \mathcal{L}^p \to \mathbb{R}$ is called a *law-invariant convex risk functional* if it satisfies following properties for any $X, Y \in \mathcal{L}^p$.

(B1) (Translation invariance) $\rho(X + m) = \rho(X) + cm$ for any $m \in \mathbb{R}$, where $c = \rho(1) - \rho(0)$.

*Department of Statistics and Actuarial Science, University of Waterloo. ✉ f27liu@uwaterloo.ca
†Department of Statistics and Actuarial Science, University of Waterloo. ✉ jcai@uwaterloo.ca
‡Department of Statistics and Actuarial Science, University of Waterloo. ✉ clemieux@uwaterloo.ca
§Department of Statistics and Actuarial Science, University of Waterloo. ✉ wang@uwaterloo.ca

(B2) (Convexity) $\rho(\lambda X + (1-\lambda)Y) \le \lambda\rho(X) + (1-\lambda)\rho(Y)$ for any $\lambda \in [0,1]$;

(B3) (Continuity) ρ is continuous with respect to the $\mathcal{L}^p$-norm.

(B4) (Law-invariance) $\rho(X) = \rho(Y)$, if X and Y have the same distribution (denoted by $X \stackrel{d}{=} Y$).

If we interpret random variables as future losses, then in the case of convex and coherent risk measures, the constant c in (B1) is 1, and in the case of generalized deviations, the constant c in (B1) is 0. We sometimes omit "law-invariant" and call ρ in Definition 1.1 simply a convex risk functional, as all functionals we encounter in this paper are law-invariant. The four properties in Definition 1.1 are common and well studied in the literature. We omit a detailed discussion and the reader is referred to classic books Föllmer and Schied (2011), Delbaen (2012) and McNeil et al. (2015). Quite obviously, the class of convex risk functionals includes all coherent risk measures (Artzner et al. (1999)), convex risk measures (Föllmer and Schied (2002)), convex premium principles (Deprez and Gerber (1985); Gerber et al. (2019)), and generalized deviations (Rockafellar et al. (2006); Grechuk et al. (2009)) as special cases.

In the classic literature (e.g. Artzner et al. (1999) and Rockafellar et al. (2006)), risk measures are monotone, satisfying[1]

(A1) (Monotonicity) $\rho(X) \le \rho(Y)$, for $X, Y \in \mathcal{L}^p$, $X \le Y$.

On the other hand, deviation measures in Rockafellar et al. (2006), in addition to (B2)-(B3), are required to be subadditive, and to take the value zero on constants, implying (B1) with $c = 0$. Convex risk functionals are not subject to these constraints and this allows us to study many types of deviation measures, such as the variance (not covered in the definition of Rockafellar et al. (2006) since it is not subadditive), the standard deviation, the mean absolute deviation and the Gini mean deviation, and a combination of a risk measure with a deviation measure, such as the standard deviation principle and the variance principle, both widely used in insurance, and the Gini Shortfall which is generally not monotone (Furman et al. (2017)).

Remark 1.1. As a classic result in convex analysis, real-valued law-invariant convex risk measures (they satisfy (A1); see Example 2.3) on $\mathcal{L}^p$, $p \in [1,\infty]$ automatically satisfy continuity (B3). The convex risk functionals in Definition 1.1 are not necessarily monotone, and (B3) is essential in this case.

The main contributions of the paper are three-fold. First, we provide a unified representation theorem for the class of law-invariant convex risk functionals in Section 2 and discuss some examples. Second, we solve the optimization of convex risk functionals under moment constraints in Section 3. Third, we formulate a large class of reinsurance design problems using law-invariant convex risk functionals, and then solve the unconstrained and constrained optimization problems separately in Section 4. Conclusion is given in Section 5, and some proofs are presented in the Appendix.

[1] In this paper, the properties labelled (Bx) are always satisfied by law-invariant convex functionals, and the properties labelled (Ax) are additional properties.

2 Representation theorems and examples

2.1 Preliminaries

In this section we obtain representation theorems for convex risk functionals. For this purpose, we rely on two important examples of risk measures, the Expected Shortfall (ES) and the Value-at-Risk (VaR), both widely used in banking and insurance (see e.g. McNeil et al. (2015)). ES is also known as the Tail Value-at-Risk (TVaR) in the insurance literature.

The *Value-at-Risk (VaR)* of $X \in \mathcal{L}^1$ at level $\alpha \in [0,1)$ is defined as

$$\mathrm{VaR}_\alpha(X) \triangleq \inf\left\{x \in \mathbb{R} : \mathbb{P}(X > x) \le \alpha\right\}.$$

In addition, let $\mathrm{VaR}_1(X) = \inf\left\{x \in \mathbb{R} : \mathbb{P}(X > x) < 1\right\}$. The *Expected Shortfall (ES)* of a random variable $X \in \mathcal{L}^1$ at level $\alpha \in (0,1]$ is defined as

$$\mathrm{ES}_\alpha(X) \triangleq \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_t(X)\mathrm{d}t.$$

Note that we use the "small α" convention in this paper.

For $c \in \mathbb{R}$, denote by Φ_c the set of concave functions h on $[0,1]$ with $h(0) = 0$ and $h(1) = c$, and let $\Phi = \bigcup_{c\in\mathbb{R}} \Phi_c$, which is the set of concave functions h on $[0,1]$ with $h(0) = 0$. For $h \in \Phi$, we define the *signed Choquet integral* with respect to $h \circ \mathbb{P}$ as

$$\rho_h(X) = \int X\mathrm{d}(h\circ\mathbb{P}) = \int_0^\infty h(\mathbb{P}(X > x))\mathrm{d}x + \int_{-\infty}^0 (h(\mathbb{P}(X > x) - h(1))\mathrm{d}x, \quad X \in \mathcal{L}^1. \tag{1}$$

The above integral is always well posed; see Lemma 2.1 below. Law-invariant signed Choquet integrals include many examples, such as the Gini mean deviation, the range, the Gini Shortfall (Furman et al. (2017)), spectral risk measures (Acerbi (2002)), and Wang's premium principles (Wang et al. (1997), see Definition 4.1 below).

For $h \in \Phi$, noting that h is continuous on $(0,1)$, the Riemann-Stieltjes integral $\int f\mathrm{d}h$ for a function f on $[0,1]$ should be interpreted as

$$\int_0^1 f(\alpha)\mathrm{d}h(\alpha) = \int_0^1 f(\alpha)\mathbb{I}_{(0,1)}(\alpha)\mathrm{d}h(\alpha) + h(0+)f(0) + (h(1-) - h(1))f(1),$$

where by convention $0 \times \infty = 0$. Since $h \in \Phi_c$ is concave, its left derivative h' is well defined a.e. For $q \in (1,\infty)$, denote by $||h'||_q$ the q-Lebesgue norm of h', i.e., $||h'||_q = (\int_0^1 |h'(t)|^q\mathrm{d}t)^{1/q}$ if h is continuous on $[0,1]$ and let $||h'||_q = \infty$ if h is not continuous on $[0,1]$. In addition, let $||h'||_\infty = \lim_{q\to\infty} ||h'||_q$ which is the supremum of $|h'|$ on $[0,1]$, and let $||h'||_1$ be the total variation of h' on $[0,1]$, which is always finite for $h \in \Phi$. Below, we use the convention $0^{-1} = \infty$.

In Lemma 2.1 below, we collect some technical facts on signed Choquet integrals, which will be useful for the main representation results. For the proof of Lemma 2.1, see the general properties of signed Choquet integrals in Cerreia-Vioglio et al. (2012, 2015).

Lemma 2.1. *For $h \in \Phi$, the following hold.*

(i) $\rho_h(X) = \int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha) > -\infty$ *for* $X \in \mathcal{L}^1$.

(ii) ρ_h *is a law-invariant convex risk functional.*

(iii) ρ_h *is finite on* $\mathcal{L}^p$ *for* $p \in [1, \infty]$ *if and only if* $||h'||_q < \infty$, *where* $q = (1-1/p)^{-1}$. *In particular,* ρ_h *is always finite on* $\mathcal{L}^\infty$.

2.2 Main representation results

Now we are ready to present the main representation result for convex risk functionals. For $p \in [1, \infty)$ and $c \in \mathbb{R}$, denote by

$$\Phi_c^p = \{h \in \Phi_c : ||h'||_q < \infty, \text{ where } q = (1 - 1/p)^{-1}\},$$

and in addition, let $\Phi_c^\infty = \Phi_c$. By Lemma 2.1, ρ_h in (1) is finite on $\mathcal{L}^p$ for $h \in \Phi_c^p$.

Theorem 2.2. *Fix* $p \in [1, \infty]$. *For a functional* $\rho : \mathcal{L}^p \to \mathbb{R}$, *the following are equivalent:*

1) ρ *is a law-invariant convex risk functional.*

2) There exist $c \in \mathbb{R}$ *and a mapping* $\beta : \Phi_c^p \to (-\infty, \infty]$ *such that*

$$\rho(X) = \sup_{h \in \Phi_c^p} \left\{ \int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha) - \beta(h) \right\}, \qquad X \in \mathcal{L}^p. \tag{2}$$

Proof. "$\Longleftarrow$" Suppose that ρ is given by (2). Let $c = \rho(1) - \rho(0)$. The properties (B1) and (B4) of VaR imply that ρ satisfies (B1) and (B4), respectively. By Lemma 2.1, the mapping $X \mapsto \int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha)$ for $h \in \Phi_c$ is a law-invariant convex risk functional. Hence, ρ, as the supremum of convex functionals, satisfies (B2). To show (B3), from the $\mathcal{L}^p$-continuity of $X \mapsto \int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha)$ for $h \in \Phi_c^p$, we know that ρ is the supremum of $\mathcal{L}^p$-continuous functionals, and hence it is lower semi-continuous with respect to $\mathcal{L}^p$-norm. Recall that real-valued convex functions on Banach spaces are norm-continuous if and only if they are lower semi-continuous with respect to the the norm (see e.g. Theorem 5.3.12 of Kosmol and Müller-Wichards (2011)). Therefore, ρ satisfies (B3).

"$\Longrightarrow$" Suppose that $\rho : \mathcal{L}^p \to \mathbb{R}$ is a law-invariant convex risk functional. We equip $\mathcal{L}^p$ with a topology τ^p such that it has the dual space $\mathcal{L}^q$ where $q \in [1, \infty]$ is the Hölder conjugate of p (satisfying $1/p + 1/q = 1$). If $1 \le p < \infty$, we can take the topology to be the $\mathcal{L}^p$-norm; note that ρ is continuous with respect to τ^p. If $p = \infty$, we take the topology $\tau^\infty = \sigma\left(\mathcal{L}^\infty, \mathcal{L}^1\right)$. By Theorem 30 of Delbaen (2012), the functional ρ satisfying (B2)-(B4) is lower semi-continuous with respect to τ^∞.

In both cases, the functional ρ is lower semi-continuous with respect to τ^p, and hence we obtain its representation and its conjugate by the Fenchel-Moreau Theorem, given by

$$\rho(X) = \sup_{X' \in \mathcal{L}^q} \left\{\mathbb{E}\left[X'X\right] - \rho'(X)\right\}, \quad X \in \mathcal{L}^p,$$

where

$$\rho'(X') = \sup_{X \in \mathcal{L}^p} \left\{ \mathbb{E}\left[X'X\right] - \rho(X) \right\}, \quad X' \in \mathcal{L}^q. \tag{3}$$

A proof of the above equations can be found, for example, in Theorem 5 in Rockafellar (1974) or Theorem A.62 of Föllmer and Schied (2011). Since ρ satisfies law-invariance (B4), we have

$$\begin{aligned}
\rho(X) = \sup_{Y \in \mathcal{L}^p, Y \stackrel{d}{=} X} \rho(Y) &= \sup_{Y \in \mathcal{L}^p, Y \stackrel{d}{=} X} \sup_{Y' \in \mathcal{L}^q} \left\{ \mathbb{E}\left[Y'Y\right] - \rho'(Y') \right\} \\
&= \sup_{Y' \in \mathcal{L}^q} \sup_{Y \in \mathcal{L}^p, Y \stackrel{d}{=} X} \left\{ \mathbb{E}\left[Y'Y\right] - \rho'(Y') \right\} \\
&= \sup_{Y' \in \mathcal{L}^q} \left\{ \int_0^1 \mathrm{VaR}_t(Y')\mathrm{VaR}_t(X)\mathrm{d}t - \rho'(Y') \right\},
\end{aligned} \tag{4}$$

where the last equality is the Fréchet-Hoeffding inequality; see Lemma 4.60 in Föllmer and Schied (2011). Note that $\mathcal{L}^p = \{Y \in \mathcal{L}^p : Y + m \in \mathcal{L}^p\}$ for any constant $m \in \mathbb{R}$. For all $m \neq 0$, $Y' \in \mathcal{L}^q$, and since ρ satisfies (B1), we have

$$\begin{aligned}
\rho'(Y') &= \sup_{Y \in \mathcal{L}^p} \left\{ \mathbb{E}\left[Y'Y\right] - \rho(Y) \right\} \\
&= \sup_{Y + m \in \mathcal{L}^p} \left\{ \mathbb{E}\left[Y'Y\right] + m\mathbb{E}\left[Y'\right] - \rho(Y + m) \right\} \\
&= m\mathbb{E}\left[Y'\right] + \sup_{Y + m \in \mathcal{L}^p} \left\{ \mathbb{E}\left[Y'Y\right] - \rho(Y) - cm \right\} = m\left(\mathbb{E}\left[Y'\right] - c\right) + \rho'(Y').
\end{aligned}$$

It follows that $\rho'(Y') = \infty$ for all $Y' \in \mathcal{L}^q$ with $\mathbb{E}[Y'] \neq c$. Thus,

$$\rho(X) = \sup_{Y' \in \mathcal{L}^q, \mathbb{E}[Y'] = c} \left\{ \int_0^1 \mathrm{VaR}_t(Y')\mathrm{VaR}_t(X)\mathrm{d}t - \rho'(Y') \right\}. \tag{5}$$

Next we show that ρ' is law-invariant. Note that for $Y' \stackrel{d}{=} X' \in \mathcal{L}^q$, using the same argument leading to (4), we have

$$\begin{aligned}
\rho'(Y') &= \sup_{Y \in \mathcal{L}^p} \left\{ \int_0^1 \mathrm{VaR}_t(Y)\mathrm{VaR}_t(Y')\mathrm{d}t - \rho(Y) \right\} \\
&= \sup_{Y \in \mathcal{L}^p} \left\{ \int_0^1 \mathrm{VaR}_t(Y)\mathrm{VaR}_t(X')\mathrm{d}t - \rho(Y) \right\} = \rho'(X').
\end{aligned}$$

Therefore, ρ' is law-invariant.

For each $Y' \in \mathcal{L}^q$ with $\mathbb{E}[Y'] = c$, define $h_{Y'}(\alpha) = \int_0^\alpha \mathrm{VaR}_t(Y')\mathrm{d}t$ for $\alpha \in (0,1]$ with $h_{Y'}(0) = 0$. It is easy to check $h_{Y'}(1) = \mathbb{E}[Y'] = c$ and $h_{Y'} \in \Phi_c^p$. Let $\hat{\Phi}_c^p = \{h_{Y'} : Y' \in \mathcal{L}^q,\ \mathbb{E}[Y'] = c\}$ and note that $\hat{\Phi}_c^p \subset \Phi_c^p$. Since ρ' is law-invariant and $h_{Y'}$ determines the distribution of Y', we can define $\beta : \hat{\Phi}_c^p \to (-\infty, +\infty]$ via $\beta(h_{Y'}) = \rho'(Y')$, $Y' \in \mathcal{L}^q$. Therefore, (5) leads to

$$\rho(X) = \sup_{h \in \hat{\Phi}_c^p} \left\{ \int_0^1 \mathrm{VaR}_t(X)\mathrm{d}h(t) - \beta(h) \right\}.$$

Extending the domain of β via $\beta(h)=\infty$ for $h \in \Phi_c^p \setminus \hat{\Phi}_c^p$ leads to (2). $\square$

Remark 2.3. The mapping β in Theorem 2.2 can be chosen to be convex. More precisely, if ρ is a law-invariant convex risk functional, then it has a representation (2) in which β is convex. To see this, for β in the proof of Theorem 2.2, for any $h_1, h_2 \in \Phi_c$, take a comonotonic pair of random variables $X_1', X_2' \in \mathcal{L}^1$ such that $h_i(q) - h_i(0+) = \int_0^q \mathrm{VaR}_t(X_i')\mathrm{d}t$, $i = 1, 2$. For any $\lambda \in [0,1]$, we have $(\lambda h_1 + (1-\lambda)h_2)(\alpha) = \int_0^\alpha \mathrm{VaR}_t(\lambda X_1' + (1-\lambda)X_2')\mathrm{d}t$ for $\alpha \in [0,1]$, and hence

$$
\begin{aligned}
\beta\left(\lambda h_1 + (1-\lambda)h_2\right) &= \rho'\left(\lambda X_1' + (1-\lambda)X_2'\right) \\
&\le \lambda\rho'\left(X_1'\right) + (1-\lambda)\rho'\left(X_2'\right) = \lambda\beta(h_1) + (1-\lambda)\beta(h_2),
\end{aligned}
$$

which implies the convexity of β.

Next, we present a representation for positively homogeneous convex risk functionals on $\mathcal{L}^p$, namely, those satisfying the following property:

(A2) (Positive homogeneity) $\rho(\lambda X) = \lambda\rho(X)$ for $\lambda > 0$ and $X \in \mathcal{L}^p$.

Theorem 2.4. *Fix $p \in [1,\infty]$. For a functional $\rho : \mathcal{L}^p \to \mathbb{R}$, the following are equivalent:*

1) ρ is a positively homogeneous law-invariant convex risk functional.

2) There exist $c \in \mathbb{R}$ and a set $\Psi_c^p \subseteq \Phi_c^p$ such that

$$
\rho(X) = \sup_{h \in \Psi_c^p} \left\{ \int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha) \right\}, \qquad X \in \mathcal{L}^p. \tag{6}
$$

Proof. It is straightforward to verify that (6) defines a positively homogeneous law-invariant convex risk functional. To show the converse, by (B3) and (A2) of ρ,

$$
\sup_{h \in \Phi_c^p} \{-\beta(h)\} = \rho(0) = \lim_{\lambda \downarrow 0} \rho(\lambda) = \lim_{\lambda \downarrow 0} \lambda\rho(1) = 0.
$$

Therefore, $\beta(h) \ge 0$ for $h \in \Phi_c^p$. Using (A2) again, for $\lambda > 0$, $\rho(X) = \frac{1}{\lambda}\rho(\lambda X)$. Hence,

$$
\rho(X) = \sup_{\lambda>0} \frac{1}{\lambda}\rho(\lambda X) = \sup_{\lambda>0} \sup_{h \in \Phi_c^p} \left\{ \int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha) - \frac{\beta(h)}{\lambda} \right\} = \sup_{h \in \Psi_c^p} \left\{ \int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha) \right\},
$$

where $\Psi_c^p = \{h \in \Phi_c^p : \beta(h) < \infty\}$. $\square$

Remark 2.5. For a set $\Psi_c^p \subset \Phi_c^p$, the representation (6) is a special case of (2) by choosing $\beta = 0$ on Ψ_c^p and $\beta = \infty$ on $\Phi_c^p \setminus \Psi_c^p$.

Below we list some classic examples of convex risk functionals. The first interesting examples are the standard deviation and the variance, both well known to be convex (Deprez and Gerber (1985)) and they have a representation in Theorem 2.2.

Example 2.1 (Standard deviation)**.** The standard deviation, defined as

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

has the following representation

$$\sigma(X) = \sup_{h \in \Phi_0^2} \left\{ \int_0^1 \mathrm{VaR}_t(X) \mathrm{d}h(t) - \beta(h) \right\}, \quad X \in \mathcal{L}^2, \tag{7}$$

where $\beta(h) = 0$ if $||h'||_2^2 \leq 1$ and $\beta(\phi) = \infty$ otherwise, i.e. $\beta(h) = \infty \times \mathbb{I}_{\{||h'||_2^2 > 1\}}$. Equivalently, it can be written in the form of (6) as

$$\sigma(X) = \sup \left\{ \int_0^1 \mathrm{VaR}_t(X) \mathrm{d}h(t) : h \in \Phi_0, \ ||h'||_2^2 \leq 1 \right\}, \quad X \in \mathcal{L}^2. \tag{8}$$

See Example 2.3 of Wang et al. (2019) for a simple proof of this representation on $\mathcal{L}^\infty$, which also applies to $\mathcal{L}^2$. □

Example 2.2 (Variance and mean-variance)**.** In the insurance context, the mean-variance premium principle is defined as

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

where $c \geq 0$. It has the following representation

$$\rho(X) = \sup_{h \in \Phi_c^2} \left\{ \int_0^1 \mathrm{VaR}_t(X) \mathrm{d}h(t) - \frac{1}{4}(||h'||_2^2 - c^2) \right\}, \quad X \in \mathcal{L}^2. \tag{9}$$

In particular, setting $c = 0$, we get a representation for the variance

$$\sigma^2(X) = \sup_{h \in \Phi_0^2} \left\{ \int_0^1 \mathrm{VaR}_t(X) \mathrm{d}h(t) - \frac{1}{4}||h'||_2^2 \right\}, \quad X \in \mathcal{L}^2, \tag{10}$$

To show (9), write $\beta(h) = \frac{1}{4}(||h'||_2^2 - c^2)$ for $h \in \Phi_c$ and note that for a continuous h,

$$\begin{aligned}
\int_0^1 \mathrm{VaR}_t(X) \mathrm{d}h(t) &= \int_0^1 (\mathrm{VaR}_t(X) - \mathbb{E}[X]) h'(t) \mathrm{d}t + c\mathbb{E}[X] \\
&= \int_0^1 (\mathrm{VaR}_t(X) - \mathbb{E}[X])(h'(t) - c) \mathrm{d}t + c\mathbb{E}[X],
\end{aligned}$$

where the first equality is due to $\int_0^1 \mathrm{d}h(t) = h(1) - h(0) = c$. By Hölder's inequality,

$$\begin{aligned}
\int_0^1 \mathrm{VaR}_t(X) \mathrm{d}h(t) &\leq c\mathbb{E}[X] + \sqrt{\int_0^1 (\mathrm{VaR}_t(X) - \mathbb{E}[X])^2 \, \mathrm{d}t} \sqrt{\int_0^1 (h'(t) - c)^2 \, \mathrm{d}t} \\
&= c\mathbb{E}[X] + 2\sqrt{\sigma^2(X)}\sqrt{\beta(h)} \leq \sigma^2(X) + c\mathbb{E}[X] + \beta(h).
\end{aligned}$$

Therefore,

$$\sup_{h\in\Phi_c^2}\left\{\int_0^1 \mathrm{VaR}_t(X)\mathrm{d}h(t)-\beta(h)\right\}\le \sigma^2(X)+c\mathbb{E}[X].$$

That is, the right-hand-side of (9) is at most $\sigma^2(X)+c\mathbb{E}[X]$. To show that this value can be attained, we simply take h such that $h'(t)=2\left(\mathrm{VaR}_t(X)-\mathbb{E}[X]\right)+c,\ t\in[0,1]$. The above argument shows that (9) and (10) hold for all $X\in\mathcal{L}^2$. □

Example 2.3 (Convex and coherent risk measures)**.** As mentioned in Section 1, risk measures in the literature are typically monotone, i.e. satisfying (A1). Using the terminology in McNeil et al. (2015), a functional $\rho:\mathcal{L}^\infty\to\mathbb{R}$ is called a *convex risk measure* if it satisfies axioms (B1) with $c=1$, (B2) and (A1), and it is called a *coherent risk measure* if it is a convex risk measure and satisfies (A2). We focus on risk measures that are law-invariant, i.e., satisfying (B4). Convex and coherent risk measures automatically satisfy the continuity (B3) on $\mathcal{L}^p$ (see Remark 1.1), and hence they are convex risk functionals in Definition 1.1. To arrive at a convex risk measure from the representation (2), we simply set $c=1$ in (2) and require $\beta(h)=\infty$ for all h that is not an increasing function. Furthermore, a Fatou-continuous law-invariant convex risk measure has an ES-based representation:

$$\rho(X)=\sup_{\mu\in\mathcal{P}([0,1])}\left(\int_0^1 \mathrm{ES}_\alpha(X)\mu(\mathrm{d}\alpha)-\tilde{\beta}(\mu)\right), \tag{11}$$

for some function $\tilde{\beta}:\mathcal{P}\left([0,1]\right)\to[0,\infty]$ which is lower semi-continuous and convex, where $\mathcal{P}\left([0,1]\right)$ is the set of all Borel probability measures on $[0,1]$. A proof of (11) can be found in Frittelli and Rosazza Gianin (2005) and Föllmer and Schied (2011), or directly obtained from (2) by using the relation $(\mathrm{d}\mu/\mathrm{d}h)(s)=s,\ 0<s<1$ (see (8.26) of McNeil et al. (2015)). A similar representation holds for coherent risk measures, known as the Kusuoka representation (Kusuoka (2001)). □

Example 2.4 (Generalized deviation measures)**.** Law-invariant generalized deviation measures are studied by Rockafellar et al. (2006) and Grechuk et al. (2009). They are functionals ρ that satisfy (B2)-(B4), (A2), and $\rho(m)=0$ for all constants $m\in\mathbb{R}$. Note that (B2) and (A2) together implies subadditivity. A simple exercise shows that a generalized deviation measure satisfies (B1) with $c=0$. Thus, they are a special type of convex risk functionals according to our definition. □

Remark 2.6. If a risk functional ρ satisfies the properties (B1)-(B4), then the mapping $X\mapsto\rho(X)-(\rho(1)-\rho(0))\mathbb{E}[X]$ satisfies the properties (B1)-(B4) with $c=0$, which can be seen as a *variability measure* (which is more general than the deviation measures in Example 2.4). Hence, any law-invariant convex risk functional can be seen as a combination of a variability measure and a constant times the mean.

2.3 Quasi-convexity

In this section, we discuss the notion of quasi-convexity and its relation to convexity (B2).

(B2') (quasi-convexity) $\rho(\lambda X+(1-\lambda)Y)\le\max\{\rho(X),\rho(Y)\}$ for any $\lambda\in[0,1]$ and $X,Y\in\mathcal{L}^p$.

In the mathematical finance literature (see e.g. Föllmer and Schied (2011) and Frittelli and Maggis (2011)), quasi-convexity is argued as a plausible notion to capture the diversification effect. Clearly, (B2') is weaker than (B2), namely (B2) $\implies$ (B2'). A natural question is whether (B2) may be replaced by (B2') in the main representation results. First, in the following proposition we show that (B2) is equivalent to (B2') if $\rho(1) - \rho(0) \neq 0$.

Proposition 2.7. *Fix $p \in [1, \infty]$. For a mapping $\rho : \mathcal{L}^p \to \mathbb{R}$ satisfying (B1) and $\rho(1) - \rho(0) \neq 0$, the properties (B2) and (B2') are equivalent.*

Proof. It suffices to show (B2') $\implies$ (B2). Take $X, Y \in \mathcal{L}^p$ and $\lambda \in [0, 1]$. Let $c = \rho(1) - \rho(0)$ and $b = \rho(Y) - \rho(X)$. Then, by (B1) and (B2'), we have

$$\begin{aligned}
\rho(\lambda X + (1-\lambda)Y) &= \rho(\lambda(X + b/c) + (1-\lambda)Y) - \lambda b \\
&\leq \max\{\rho(X + b/c), \rho(Y)\} - \lambda b \\
&= \max\{\rho(X) + b, \rho(Y)\} - \lambda b \\
&= \rho(Y) - \lambda b = \lambda\rho(X) + (1-\lambda)\rho(Y).
\end{aligned}$$

Hence (B2) holds. $\square$

As suggested by Proposition 2.7, if we replace (B2) by the weaker property (B2'), Theorem 2.2 and other results of Sections 2 and 3 hold in the case $\rho(1) - \rho(0) \neq 0$. Next, we show by the following example that (B2) and (B2') are not equivalent for the case $\rho(1) - \rho(0) = 0$.

Example 2.5. Let $\rho : \mathcal{L}^2 \to \mathbb{R}$ be given by $\rho(X) = \sigma(X)\mathbb{I}_{\{\sigma(X)\geq 1\}}$, where $\sigma(\cdot)$ is the standard deviation in Example 2.1. Clearly, ρ satisfies (B1) with $c = \rho(1) - \rho(0) = 0$. Note that σ is convex on $\mathcal{L}^2$, and hence it is quasi-convex on $\mathcal{L}^2$. We now show that ρ is quasi-convex, that is, for any $X, Y \in \mathcal{L}^2$ and $\lambda \in [0, 1]$,

$$\rho(\lambda X + (1-\lambda)Y) \leq \max\{\rho(X), \rho(Y)\}. \tag{12}$$

Note that

$$\max\{\rho(X), \rho(Y)\} = \max\{\sigma(X), \sigma(Y)\}\mathbb{I}_{\{\max\{\sigma(X),\sigma(Y)\}\geq 1\}}.$$

If $\max\{\sigma(X), \sigma(Y)\} < 1$, then by quasi-convexity of σ, we have $\sigma(\lambda X+(1-\lambda)Y) \leq \max\{\sigma(X), \sigma(Y)\} < 1$. Hence, $\rho(\lambda X + (1-\lambda)Y) = 0$ and (12) holds. If $\max\{\sigma(X), \sigma(Y)\} \geq 1$, then

$$\rho(\lambda X + (1-\lambda)Y) \leq \sigma(\lambda X + (1-\lambda)Y) \leq \max\{\sigma(X), \sigma(Y)\} = \max\{\rho(X), \rho(Y)\},$$

and thus (12) holds. Hence, ρ is quasi-convex.

On the other hand, if we take a standard normal random variable Z and let $X = 0.8Z$ and $Y = 1.2Z$, then $\rho(X) = 0$, $\rho(Y) = 1.2$, and thus

$$\rho\left(\frac{X}{2} + \frac{Y}{2}\right) = \rho(Z) = 1 > 0.6 = \frac{1}{2}\rho(X) + \frac{1}{2}\rho(Y).$$

Therefore, ρ is not convex. This shows that quasi-convexity does not imply convexity if $c = 0$. $\square$

In the next example, we show that quasi-convexity is not invariant by adding or subtracting the mean. Note that all properties (B1)-(B4) are preserved under such a transformation.

Example 2.6. Take ρ as in Example 2.5, and define another functional $\rho' : \mathcal{L}^2 \to \mathbb{R}$ by $\rho'(X) = \rho(X) + \mathbb{E}[X]$. The functional ρ' satisfies (B1) with $c = 1$. We shall see that ρ' is not quasi-convex although ρ is. For instance, we can take a standard normal random variable Z and let $X = 0.8Z+1.2$ and $Y = 1.2Z$. Then, $\rho'(X) = 1.2$, $\rho'(Y) = 1.2$, and thus

$$\rho'\left(\frac{X}{2} + \frac{Y}{2}\right) = \rho'(0.6 + Z) = 1.6 > 1.2 = \max\{\rho'(X), \rho'(Y)\}.$$

Hence, ρ' is not quasi-convex. Therefore, quasi-convexity (B2') is not invariant by adding or subtracting the mean functional. $\square$

In summary, quasi-convexity is equivalent to convexity if $\rho(1) - \rho(0) \neq 0$, which is the case of (not necessarily monotone) risk measures. In the case $\rho(1) - \rho(0) = 0$, which is the case of deviation measures, these two notions are not equivalent. For general results on the representation of quasi-convex risk functionals, we refer to Frittelli and Maggis (2011) and the references therein.

3 Worst-case values with moment information

In this section we consider convex risk functionals defined on $\mathcal{L}^2$. According to Theorem 2.2 and Remark 2.3, a convex risk functional $\rho : \mathcal{L}^2 \to (-\infty, \infty]$ can be written as

$$\rho(X) = \sup_{h \in \Phi_c^2} \left\{ \int_0^1 \mathrm{VaR}_\alpha(X) \mathrm{d}h(\alpha) - \beta(h) \right\}, \tag{13}$$

where $c \in \mathbb{R}$ and $\beta : \Phi_c^2 \to (-\infty, \infty]$ is convex and $\inf_{h \in \Phi_c^2} \beta(h) \in \mathbb{R}$. Without further specifying β, ρ in (13) may possibly take the value ∞ for some $X \in \mathcal{L}^p$. Certainly, (13) includes all law-invariant risk functionals in Definition 1.1, which are real-valued. A condition for the finiteness of ρ in (13) on $\mathcal{L}^p$, $p \in [1, \infty]$ is specified in Proposition 3.6 at the end of this section.

In the context of robust risk evaluation, one may only have partial information on a risk X to be evaluated. We consider the case in which one only knows the mean and the variance of X. This setup has wide applications in model uncertainty and portfolio optimization. We refer to Li et al. (2018), Li (2018) and Cornilly et al. (2019) and the references therein for more background on this problem. Later we generalize our results to the case of another moment instead of the variance.

Denote by $\mathcal{L}^2(m, v) = \{X \in \mathcal{L}^2 : \mathbb{E}[X] = m,\ \sigma^2(X) = v^2\}$. Define the worst-case value of ρ with mean $m \in \mathbb{R}$ and standard deviation $v > 0$ as

$$\bar{\rho}(m, v) = \sup\{\rho(X) : X \in \mathcal{L}^2(m, v)\}.$$

Since $h \in \Phi_c^2$ is concave and continuous, it one-to-one corresponds to its left derivative h'. Recall that $||h'||_2^2 = \int_0^1 (h'(t))^2 \mathrm{d}t$ and $||h' - c||_2^2 = \int_0^1 (h'(t) - c)^2 \mathrm{d}t = ||h'||_2^2 - c^2$. Clearly $||h' - c||_2^2 = 0$

if and only if $h'(t) = c$ a.e. Note that h' exits almost everywhere and it is a decreasing function. Recall that for $h \in \Phi_c^2$,

$$\rho_h(X) = \int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha) = \int_0^1 \mathrm{VaR}_\alpha(X)h'(\alpha)\mathrm{d}\alpha, \quad X \in \mathcal{L}^2, \tag{14}$$

and its corresponding worst-case value is

$$\bar{\rho}_h(m, v) = \sup\{\rho_h(X) : X \in \mathcal{L}^2(m, v)\}.$$

We aim to calculate the values $\bar{\rho}(m, v)$ and $\bar{\rho}_h(m, v)$ and, if possible, find the distributions of X attaining the worst-case values. The main results are summarized in the following theorem.

Theorem 3.1. *Suppose that $c \in \mathbb{R}$, $m \in \mathbb{R}$ and $v > 0$.*

(i) For $h \in \Phi_c^2$,

$$\bar{\rho}_h(m, v) = mc + v||h' - c||_2. \tag{15}$$

If $||h' - c||_2 > 0$, the above maximum value is attained by a random variable $X \in \mathcal{L}^2(m, v)$ with

$$\mathrm{VaR}_t(X) = m + v\frac{h'(t) - c}{||h' - c||_2}, \quad t \in (0, 1) \text{ a.e.}$$

If $||h' - c||_2 = 0$, the above maximum value is attained by any random variable $X \in \mathcal{L}^2(m, v)$.

(ii) For ρ given in (13), we have

$$\bar{\rho}(m, v) = mc + \sup_{h \in \Phi_c^2} \left\{v||h' - c||_2 - \beta(h)\right\}. \tag{16}$$

Proof. We first show (i). Hölder's inequality implies

$$\begin{aligned}
\bar{\rho}_h(0, 1) = \sup_{X \in \mathcal{L}^2(0,1)} \int_0^1 h'(t)\mathrm{VaR}_t(X)\mathrm{d}t &= \sup_{X \in \mathcal{L}^2(0,1)} \int_0^1 (h'(t) - c)\mathrm{VaR}_t(X)\mathrm{d}t \\
&\le \sup_{X \in \mathcal{L}^2(0,1)} ||h' - c||_2 \left(\int_0^1 (\mathrm{VaR}_t(X))^2\mathrm{d}t\right)^{1/2} \qquad (17) \\
&= ||h' - c||_2.
\end{aligned}$$

Assume $||h' - c||_2 > 0$. To attain the Hölder bound in (17), it is necessary and sufficient that $\mathrm{VaR}_t(X) = k(h'(t) - c)$ a.e. for some $k > 0$, leading to $k = 1/||h' - c||_2$, and namely,

$$\mathrm{VaR}_t(X) = \frac{h'(t) - c}{||h' - c||_2} \quad t \in (0, 1) \text{ a.e.} \tag{18}$$

Note that since h' is a decreasing function, such a random variable $X \in \mathcal{L}^2(0, 1)$ always exists, noting that for any decreasing function g in $(0, 1)$, the random variable $g(1-U)$ has $\mathrm{VaR}_t(g(1-U)) = g(t)$ a.e. for a uniform $[0, 1]$ random variable U.

If $||h'-c||_2=0$, it is trivial to see $\rho_h(X)=c\mathbb{E}[X]$, which is a constant for all random variables $X\in\mathcal{L}^2(0,1)$. Therefore, $\bar{\rho}_h(0,1)=||h'-c||_2$ for all $h\in\Phi_c^2$. Any random variable $Y\in\mathcal{L}^2(m,v)$ is one-to-one corresponding to a random variable $X\in\mathcal{L}^2(0,1)$ via $Y=vX+m$. Since ρ_h satisfies translation invariance (B1) and positive homogeneity (A2), we have $\bar{\rho}_h(m,v)=mc+v\bar{\rho}_h(0,1)$. Hence, we obtain (15), and the attaining random variable is obtained from (18).

The statement (ii) can be verified directly from (i) and $\rho(X)=\sup_{h\in\Phi_c^2}\{\rho_h(X)-\beta(h)\}$. $\square$

Remark 3.2. Results of a similar type as Theorem 3.1 are available in the literature. In particular, in the context of coherent or distortion risk measures, some special cases of Theorem 3.1 are obtained by Li (2018) and Zhu and Shao (2018). Theorem 3.1 covers these cases, and we provide a concise proof based on Hölder's inequality, different from the above literature. Cornilly et al. (2019) studied a related problem for convex distortion risk measures, where the optimization problem is taken over the set of random variables with a specified bounded range. Zhu and Shao (2018) also obtained results for non-convex distortion risk measures, which is not the focus of this paper.

Next we present a corollary for the case of a positively homogeneous convex risk functional $\rho:\mathcal{L}^2\to\mathbb{R}$, given by (see Theorem 2.4)

$$\rho(X)=\sup_{h\in\Psi_c^2}\left\{\int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha)\right\},\qquad X\in\mathcal{L}^2. \tag{19}$$

Corollary 3.3. *Suppose that $c\in\mathbb{R}$, $m\in\mathbb{R}$, $v>0$ and ρ is given by (19). Then*

$$\bar{\rho}(m,v)=mc+v\sup_{h\in\Psi_c^2}||h'-c||_2.$$

Example 3.1. We list a few simple cases of Theorem 3.1 and Corollary 3.3. It is obvious that, if ρ is the standard deviation or the variance, the corresponding worst-case value $\bar{\rho}(m,v)$ should be v or v^2. This can be checked using Theorem 3.1, Corollary 3.3, (8) and (10). If $\rho(X)=\sigma(X)$, $X\in\mathcal{L}^2$, using Corollary 3.3 and (8), we arrive at the equality

$$\bar{\rho}(m,v)=v\,\sup\{||h'||_2: h\in\Phi_0,\ ||h'||_2^2\le 1\}=v.$$

If $\rho(X)=\sigma^2(X)$, $X\in\mathcal{L}^2$, using using Theorem 3.1 and (10), we arrive at the equality

$$\bar{\rho}(m,v)=\sup_{h\in\Phi_0^2}\left\{v||h'||_2-\frac{1}{4}||h'||_2^2\right\}=\sup_{x\ge0}\left\{vx-\frac{1}{4}x^2\right\}=v^2.$$

In the case of $\rho=\mathrm{ES}_\alpha$ for some $\alpha\in(0,1)$, we have $\rho(X)=\int_0^1\mathrm{VaR}_t(X)\mathrm{d}h(t)$ for $X\in\mathcal{L}^2$, where $h(t)=\min\{\frac{t}{\alpha},1\}$. Using Theorem 3.1, we have

$$\sup_{X\in\mathcal{L}^2(m,v)}\mathrm{ES}_\alpha(X)=m+v||h'-1||_2=m+v\sqrt{\int_0^1\left(\frac{1}{\alpha}\mathbb{I}_{[0,\alpha]}(t)-1\right)^2\mathrm{d}t}=m+v\sqrt{\frac{1-\alpha}{\alpha}}.$$

This is the well-known Cantelli-type formula for ES; see Li et al. (2018) for this formula and a collection of results on ES bounds with mean and variance constraints.

Finally, we illustrate that the results in Theorem 3.1 and Corollary 3.3 can be extended to the case of another central moment instead of the variance. For $p > 1$, $m \in \mathbb{R}$ and $v > 0$, denote by $\mathcal{L}^p(m, v) = \{X \in \mathcal{L}^p,\ \mathbb{E}[X] = m,\ \mathbb{E}[|X - m|^p] = v^p\}$. To establish the result in $\mathcal{L}^p$, we introduce the following quantities, for $h \in \Phi_c$ and $q \geq 1$,

$$[h]_q = \min_{x \in \mathbb{R}} ||h' - x||_q \quad \text{and} \quad c_{h,q} = \arg\min_{x \in \mathbb{R}} ||h' - x||_q.$$

For $q > 1$, to verify that $[h]_q$ and $c_{h,q}$ are well-defined, consider the mapping

$$\phi : \mathbb{R} \to \mathbb{R}, \quad x \mapsto \int_0^1 \left|h'(t) - x\right|^q \mathrm{d}t,$$

which is strictly convex, and thus continuous. Moreover, invoking Jensen's inequality, we end up with

$$\lim_{x \to -\infty} \phi(x) = \lim_{x \to \infty} \phi(x) = \infty.$$

From this we can conclude that ϕ has a unique minimum at $c_{h,q} \in \mathbb{R}$. Moreover, note that $[h]_q = ||h' - c_{h,q}||_q$, and for $q = 2$, $[h]_2 = ||h' - c||_2$ and $c_{h,2} = c$. The quantity $[h]_q$ may be interpreted as a q-central norm of the function h' and $c_{h,q}$ as a q-center.

Theorem 3.4. *For $c \in \mathbb{R}$, $p > 1$, $h \in \Phi_c^p$, $m \in \mathbb{R}$, and $v > 0$, we have*

$$\sup\{\rho_h(X) : X \in \mathcal{L}^p(m, v)\} = mc + v[h]_q,$$

where $q = (1 - 1/p)^{-1}$. If $[h]_q > 0$, the above maximum value is attained by a random variable $X \in \mathcal{L}^p(m, v)$ with

$$\mathrm{VaR}_t(X) = m + v\frac{|h'(t) - c_{h,q}|^q}{h'(t) - c_{h,q}}[h]_q^{1-q}, \quad t \in (0, 1)\ a.e.,$$

where by convention $0^q/0 = 0$. If $[h]_q = 0$, the above maximum value is attained by any random variable $X \in \mathcal{L}^p(m, v)$.

Proof. The proof is similar to that of Theorem 3.1. The case $[h]_q = 0$ is trivial and we assume $[h]_q > 0$ in the following. Again we use Hölder's inequality, and obtain

$$\begin{aligned} \sup_{X \in \mathcal{L}^p(0,1)} \int_0^1 h'(t)\mathrm{VaR}_t(X)\mathrm{d}t &= \sup_{X \in \mathcal{L}^p(0,1)} \int_0^1 (h'(t) - c_{h,q})\mathrm{VaR}_t(X)\mathrm{d}t \\ &\leq \sup_{X \in \mathcal{L}^p(0,1)} ||h' - c_{h,q}||_q \left(\int_0^1 |\mathrm{VaR}_t(X)|^p \mathrm{d}t\right)^{1/p} = [h]_q. \end{aligned}$$

To construct a random variable that attains the Hölder bound, we use the Hölder extremal equality, which states that, for any function f on $[0,1]$, it holds

$$||f||_q = \sup\left\{\left|\int_0^1 f(t)g(t)\mathrm{d}t\right| : ||g||_p \leq 1\right\},$$

and if $||f||_q < \infty$, the maximum is attained by $g(t) = ||f||_q^{1-q}|f(t)|^q/f(t)$, $t \in [0,1]$, which satisfies $||g||_p = 1$. Because h' is a decreasing function and $q > 1$, the function $|h' - c_{h,q}|^q/(h' - c_{h,q})$ is also a decreasing function. Let X be a random variable whose distribution is given by

$$\mathrm{VaR}_t(X) = \frac{|h'(t) - c_{h,q}|^q}{h'(t) - c_{h,q}}[h]_q^{1-q}, \quad t \in (0,1) \text{ a.e.}$$

Note that such a random variable X always exists. For instance, we can take

$$X = \frac{|h'(U) - c_{h,q}|^q}{h'(U) - c_{h,q}}[h]_q^{1-q}$$

where U is a uniform random variable on $[0,1]$. It is easy to see that X satisfies the above requirement noting that $|h' - c_{h,q}|^q/(h' - c_{h,q})$ is a decreasing function.

From the Hölder extremal equality, $\mathbb{E}[|X|^p] = 1$ and

$$\rho_h(X) = [h]_q^{1-q}\int_0^1 |h'(t) - c_{h,q}|^q \mathrm{d}t = [h]_q^{1-q}[h]_q^q = [h]_q.$$

It remains to verify $\mathbb{E}[X] = 0$, which boils down to

$$\begin{aligned}\int_0^1 \frac{|h'(t) - c_{h,q}|^q}{h'(t) - c_{h,q}}\mathrm{d}t &= \int_0^1 |h'(t) - c_{h,q}|^{q-1}\mathbb{I}_{\{h'(t)>c_{h,q}\}}\mathrm{d}t - \int_0^1 |h'(t) - c_{h,q}|^{q-1}\mathbb{I}_{\{h'(t)<c_{h,q}\}}\mathrm{d}t \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}x}\int_0^1 |h'(t) - x|^q \mathrm{d}t\right)\bigg|_{x=c_{h,q}} = \left(\frac{\mathrm{d}}{\mathrm{d}x}||h' - x||_q^q\right)\bigg|_{x=c_{h,q}} = 0,\end{aligned}$$

where the last equality is due to the fact that $c_{h,q}$ minimizes $||h' - x||_q$ over $x \in \mathbb{R}$. Therefore, $X \in \mathcal{L}^p(0,1)$ and $\rho_h(X) = [h]_q$. Similarly to the proof of Theorem 3.1, any random variable $Y \in \mathcal{L}^p(m,v)$ is one-to-one corresponding to a random variable $X \in \mathcal{L}^p(0,1)$ via $Y = vX + m$. Since ρ_h satisfies translation invariance (B1) and positive homogeneity (A2), we have

$$\sup\{\rho_h(Y) : Y \in \mathcal{L}^p(m,v)\} = v\sup\{\rho_h(X) : X \in \mathcal{L}^p(0,1)\} + mc = mc + v[h]_q.$$

The rest of the proof is straightforward. □

Remark 3.5. A similar setup with an absolute moment constraint instead of a central moment constraint is considered by Cornilly et al. (2019) for convex distortion risk measures and random variables with a specified bounded range. As far as we are aware of, formula of the type in Theorem 3.4 does not appear in the literature. The argument using the Hölder inequality can be applied to the problem of the absolute moment constraint in an analogous way. Note that if $p = 2$, having a mean-variance constraint is equivalent to having a mean-second-moment constraint.

Theorem 3.4 implies the following formula for the supremum over $\mathcal{L}^p(m,v)$ of ρ defined via (13) in the beginning of this section:

$$\sup\{\rho(X) : X \in \mathcal{L}^p(m,v)\} = mc + \sup_{h\in\Phi_c^p}\{v[h]_q - \beta(h)\}, \tag{20}$$

generalizing (16) in Theorem 3.1. Before ending this section, we present a proposition yielding the finiteness of ρ on $\mathcal{L}^p$.

Proposition 3.6. *For $p \in [1,\infty]$ and $c \in \mathbb{R}$, the functional ρ on $\mathcal{L}^p$ defined via*

$$\rho(X) = \sup_{h\in\Phi_c^p}\left\{\int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha) - \beta(h)\right\},$$

is finite if $\inf_{h\in\Phi_c^p}\beta(h) \in \mathbb{R}$ and $\sup_{h\in\Psi}||h'||_q < \infty$, where $\Psi = \{h \in \Phi_c^p : \beta(h) < \infty\}$ and $q = (1-1/p)^{-1} \in [1,\infty]$.

Proof. We have $\rho(0) = -\inf_{h\in\Phi_\beta}\beta(h) \in \mathbb{R}$ and for any $h \in \Phi_c^p$, $\beta(h) + \rho(0) \geq 0$. For $X \in \mathcal{L}^p$,

$$\begin{aligned}
\rho(X) &= \sup_{h\in\Phi_c^p}\left\{\int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha) - \beta(h) - \rho(0)\right\} + \rho(0)\\
&\leq \sup\left\{\int_0^1 \mathrm{VaR}_\alpha(X)\mathrm{d}h(\alpha) : h \in \Phi_c^p,\ \beta(h) < \infty\right\} + \rho(0)\\
&\leq \sup\left\{\mathbb{E}[|X|^p]^{1/p}||h'||_q : h \in \Psi\right\} + \rho(0) < \infty,
\end{aligned}$$

where Hölder's inequality is applied to get the second inequality. ◻

4 Reinsurance design problem

4.1 Problem setup

In this section, we consider an application of the representation (2) in optimal reinsurance design problems. All proofs in this section are presented in the Appendix.

A reinsurance contract, bought by an insurer from a reinsurer to protect against the insurer's potential aggregate claim, is an important risk-sharing tool for an insurer. Denote by X the underlying (aggregate) risk faced by the insurer and assume that $X \in \mathcal{L}_+^\infty$, where $\mathcal{L}_+^\infty = \{X \in \mathcal{L}^\infty : X \geq 0\}$. We focus on the case that $X \in \mathcal{L}^\infty$ due to technical challenges. Consistently to the literature of reinsurance (Cai and Tan (2007), Cai et al. (2008) and Cheung (2010)) the survival function $S_X(x)$ of X is assumed to be continuous and strictly decreasing on $(0,\bar{X}]$ with a possible jump at 0, where $\bar{X}$ is the essential supremum of X. Under a reinsurance contract, the reinsurer agrees to cover a part of the risk $I(X)$ for the insurer and requires a premium. The function $I(x)$ is commonly described as the *ceded loss function*, while $R(x) \triangleq x - I(x)$ is known as the *retained loss function*. The total random loss faced by the insurer with a reinsurance contract becomes the retained loss $X - I(X)$ plus the premium. To avoid moral hazard, a "*feasible*" I needs to satisfy following two conditions, and we denote $\mathcal{I}$ to be the set of all feasible ceded loss functions satisfying

1. $I : [0, \bar{X}] \to [0, \bar{X}]$ is a non-decreasing function satisfies $I(0) = 0$,
2. $0 \le I(y) - I(x) \le y - x$, for any $0 \le x \le y \le \bar{X}$,

Note that any $I \in \mathcal{I}$ is 1-Lipschitz continuous on $[0, \bar{X}]$. Since a sequence $\{I_n\}_{n=1}^{\infty} \subset \mathcal{I}$ is uniformly bounded and equicontinuous, by the Arzelà-Ascoli theorem, there exists a subsequence that converges uniformly on $[0, \bar{X}]$ and the limit function also belongs to $\mathcal{I}$.

We assume that the insurer takes a law-invariant convex risk functional ρ on $\mathcal{L}^\infty$ with the representation (2) where $c \ge 0$, β is convex and lower semi-continuous with respect to 1-Lebesgue norm and $\lim_{n\to\infty} \beta(h_n) = \infty$ whenever $\lim_{n\to\infty} \|h_n\|_\infty = \infty$. In the context of reinsurance, all random variables represent risks in the sense that a positive value means a loss and a negative value means a gain. It is natural that adding a sure loss makes the position less favourable by the insurer, or equivalently, increases the value of the risk functional. Noting that (B1) implies $\rho(Y + 1) = \rho(Y) + c$ for all random variables Y, the assumption $c \ge 0$ is natural.

As the seller of a reinsurance contract, the reinsurer will assign a premium to a ceded loss $I(X)$. We assume that the reinsurer uses a *Wang's premium principle* (Wang et al. (1997)), which is a special case of the signed Choquet integrals in (1).

Definition 4.1 (Wang's premium principle)**.** Suppose that $g : [0, 1] \to [0, 1]$ is an increasing and concave function with $g(0) = 0$ and $g(1) = 1$. Then, g is called a *distortion function*, and the Choquet integral $\rho_{(1+\theta)g} : \mathcal{L}_+^\infty \to \mathbb{R}$, defined on the set of non-negative random variables,

$$\rho_{(1+\theta)g}(Y) = (1 + \theta) \int_0^\infty g \circ S_Y(t)\, \mathrm{d}t, \qquad Y \in \mathcal{L}_+^\infty,$$

for a constant $\theta \ge 0$ is called a *Wang's premium principle.*

Remark 4.1. In the insurance literature, Wang's premium principle is commonly defined with $\theta = 0$. Indeed, when the distortion function g is concave, the exceeding amount $\int_0^\infty g \circ S_Y(t)\, \mathrm{d}t - \mathbb{E}[Y]$ is non-negative and it is the risk loading added to the expected loss. In Definition 4.1, we impose an additional risk loading parameter θ in order to include the *expected value premium principle.* When $g(x) = x$ for $x \ge 0$, $\rho_{(1+\theta)g}(Y) = (1 + \theta)\mathbb{E}[Y]$ recovers the expected value premium principle.

In this section, we consider a general framework of the optimal reinsurance design problem from the perspective of the insurer. Without any constraint on premiums, the insurer is interested in the following *free premium problem*:

$$\min_{I \in \mathcal{I}} \rho \left(X - I(X) + \rho_{(1+\theta)g}(I(X)) \right). \tag{21}$$

The insurer may face a budget on his purchasing of reinsurance

$$\rho_{(1+\theta)g}(I(X)) \le p, \qquad \text{for some budget threshold } p > 0. \tag{22}$$

We call the minimization problem (21) subject to (22) as the *budget constraint problem*. Note that $\rho_{(1+\theta)g}(I(X)) \le \rho_{(1+\theta)g}(X) \le (1+\theta)\bar{X}$ for all $I \in \mathcal{I}$. Therefore, if $p \ge (1+\theta)\bar{X}$, then the constraint (22) is trivial. Hence, the budget constraint problem includes the free premium problem as a special case.

Due to the continuity of the mappings ρ and $\rho_{(1+\theta)g}$, the budget constraint problem (21)-(22) admits an optimal solution $I^* \in \mathcal{I}$. To be more precise, take a sequence $\{I_n\}_{n=1}^{\infty} \subset \mathcal{I}$ satisfying $\rho_{(1+\theta)g}(I_n(X)) \le p$ and

$$\lim_{n\to\infty} \rho\left(X - I_n(X) + \rho_{(1+\theta)g}(I_n(X))\right) = \inf_{I\in\mathcal{I},\ \rho_{(1+\theta)g}(I(X))\le p} \rho\left(X - I(X) + \rho_{(1+\theta)g}(I(X))\right).$$

Since there exists a subsequence $\{I_{n_k}\}_{k=1}^{\infty}$ that uniformly converges to $I^* \in \mathcal{I}$. we know, $I_{n_k}(X) \to I^*(X)$ in $\mathcal{L}^\infty$ as $k \to \infty$. Since both the mappings $Y \mapsto \rho\left(X - Y + \rho_{(1+\theta)g}(Y)\right)$ and $Y \mapsto \rho_{(1+\theta)g}(Y)$ are $\|\cdot\|_\infty$-continuous, we know that I^* is a minimizer for Problem (21)-(22).

The insurer's objective and the reinsurer's premium principle belongs to some families of functionals. In the one-reinsurer model, a lot of research has been done when one of the two functionals is given specifically while the other one is only given by a general expression. For example, Chi and Tan (2013) assigned the ES to the insurer; Cheung et al. (2014) chose the actuarial pricing principle for the reinsurer; Cheung and Lo (2017) used distorted risk measures for both the insurer and the reinsurer, which is generalized to coherent risk measures in Cheung et al. (2019). We provide a general formula for the optimal reinsurance contract, which can be applied to any law-invariant convex risk functionals and Wang's premium principle. The class of convex risk functionals we consider includes not only convex risk measures, but also deviation measures, and combinations of a convex risk measure and a deviation measure. Thus, our setting includes many classic settings with a great generality.

Before moving onto the next section, we quote the following well-known Sion's Minimax Theorem which will be helpful for solving Problem (21).

Theorem 4.2 (Sion's Minimax Theorem)**.** *Let Ξ_1 be a compact convex subset of a topological vector space, and Ξ_2 be a convex subset of a topological vector space. Let f be a real-valued function defined on $\Xi_1 \times \Xi_2$ such that*

1) $\xi_1 \mapsto f(\xi_1, \xi_2)$ is convex and lower-semicontinuous on Ξ_1 for each $\xi_2 \in \Xi_2$;

2) $\xi_2 \mapsto f(\xi_1, \xi_2)$ is concave and upper-semicontinuous on Ξ_2 for each $\xi_1 \in \Xi_1$.

Then

$$\inf_{\xi_1\in\Xi_1} \sup_{\xi_2\in\Xi_2} f(\xi_1, \xi_2) = \sup_{\xi_2\in\Xi_2} \inf_{\xi_1\in\Xi_1} f(\xi_1, \xi_2). \tag{23}$$

Remark 4.3. If the equation (23) holds, the value in (23) is called the *saddle-value* in the minimax problem. A pair $(\xi_1^*, \xi_2^*) \in \Xi_1 \times \Xi_2$ is called a *saddle-point* of f with respect to $\Xi_1 \times \Xi_2$, if it satisfies $\inf_{\xi_1\in\Xi_1} f(\xi_1, \xi_2^*) = \sup_{\xi_2\in\Xi_2} f(\xi_1^*, \xi_2)$. For an arbitrary real-value function f and space $\Xi_1 \times \Xi_2$, it is always true that $\inf_{\xi_1\in\Xi_1}\sup_{\xi_2\in\Xi_2} f(\xi_1,\xi_2) \geq \sup_{\xi_2\in\Xi_2}\inf_{\xi_1\in\Xi_1} f(\xi_1,\xi_2)$, but (23) may not hold. The existence of a saddle-point implies the existence of the saddle value because (23) is implied by the following observation

$$\inf_{\xi_1\in\Xi_1}\sup_{\xi_2\in\Xi_2} f(\xi_1,\xi_2) \leq \sup_{\xi_2\in\Xi_2} f(\xi_1^*,\xi_2) = \inf_{\xi_1\in\Xi_1} f(\xi_1,\xi_2^*) \leq \sup_{\xi_2\in\Xi_2}\inf_{\xi_1\in\Xi_1} f(\xi_1,\xi_2).$$

It should be pointed out that, the existence of a saddle-value is not a sufficient condition for the existence of a saddle-point.

4.2 The free premium problem

As we have seen above, the free premium problem (21) is a special case of the budget constraint problem (21)-(22). Nevertheless, we first analyze the free premium problem, because the arguments here are easier to follow, thus helping the reader to understand our main ideas.

The existence of optimal solutions to Problem (21) is guaranteed by the continuity of the risk functionals. To obtain insight of the form of optimal reinsurance ceded functions, we need to first rely on the main representation result in Theorem 2.2 to transfer the minimization problem (21) to a minimax problem, and then to characterize the optimal solution by applying Theorem 4.2. For a given $I \in \mathcal{I}$, using Lemma 2.1, we can write

$$\begin{aligned}\rho\left(X - I(X) + \rho_{(1+\theta)g}(I(X))\right) &= \sup_{h\in\Phi_c}\left(\int_0^1 \mathrm{VaR}_\alpha(R(X))\mathrm{d}h(\alpha) - \beta(h)\right) + c\rho_{(1+\theta)g}(I(X))\\ &= \sup_{h\in\Phi_c} f(I,h),\end{aligned}$$

where $f : \mathcal{I}\times\Phi_c \to \mathbb{R}$ is defined via

$$f(I,h) \triangleq \int_0^{\bar{X}} h\circ S_{R(X)}(x)\mathrm{d}x + c(1+\theta)\int_0^{\bar{X}} g\circ S_{I(X)}(t)\,\mathrm{d}t - \beta(h). \tag{24}$$

Therefore, Problem (21) has the following minimax expression (25).

$$\min_{I\in\mathcal{I}}\sup_{h\in\Phi_c} f(I,h), \tag{25}$$

which can be solved by changing the order of the minimum and the supremum with the help of Theorem 4.2. Such technique is crucial in solving minmax problems when general risk measures are used; see also Cheung et al. (2014, 2019).

For each $h \in \Phi_c$, write:

1. $\phi_h(x) \triangleq \min\{h(x), c(1+\theta)g(x)\} = h(x)\wedge(c(1+\theta)g(x))$, for any $x\in[0,1]$;
2. $G_h \triangleq \{0\leq t\leq \bar{X} : c(1+\theta)g\circ S_X(t) < h\circ S_X(t)\}$, and $G_h^c \triangleq [0,\bar{X}]\setminus G_h$;
3. $E_h \triangleq \{0\leq t\leq \bar{X} : c(1+\theta)g\circ S_X(t) = h\circ S_X(t)\}$, and $(G_h\cup E_h)^c \triangleq [0,\bar{X}]\setminus(G_h\cup E_h)$;

and define $I_h(t)$ to be the reinsurance policy satisfying the following conditions:

$$I_h(0) = 0, \text{ and } I_h'(t) = \mathbb{I}_{G_h}(t) \text{ a.e., for } t \geq 0, \tag{26}$$

where $\mathbb{I}_{G_h}(t) = 1$ for $t \in G_h$ and $\mathbb{I}_{G_h}(t) = 0$ for $t \notin G_h$ is the indicator function associated with G_h.

Lemma 4.4. *For a given $X \in \mathcal{L}_+^\infty$, there exists $h_0 \in \Phi_c$ such that*

$$S \triangleq \sup_{h \in \Phi_c} \left\{ \int_0^{\bar{X}} \phi_h \circ S_X(t)\mathrm{d}t - \beta(h) \right\} = \int_0^{\bar{X}} \phi_{h_0} \circ S_X(t)\mathrm{d}t - \beta(h_0) \tag{27}$$

is the saddle-value of the minimax problem (25). Moreover, denoted by I_0 an optimal solution to Problem (21), we have $\rho\left(X - I_0(X) + \rho_{(1+\theta)g}(I_0(X))\right) = S$.

The next proposition provides a necessary condition for the expression of the minimizer of Problem (21).

Proposition 4.5. *Each optimal solution I_0 for Problem (21) has the following form:*

$$I_0(x) = \int_0^x \left(\mathbb{I}_{G_{h_0}}(t) + \alpha(t)\mathbb{I}_{E_{h_0}}(t)\right) \mathrm{d}t, \qquad \text{for } x \geq 0, \tag{28}$$

where $\alpha : \mathbb{R}_+ \to [0,1]$ is any measurable function, and h_0 is defined in Lemma 4.4.

Remark 4.6. The feasibility constraints $I \in \mathcal{I}$ on the ceded loss function are mathematically unnecessary for the solution of our main problem (21) in this paper. That is, by allowing all possible choices of measurable function I in (21), one would still obtain an optimal solution I which belongs to $\mathcal{I}$. This conclusion is based on the celebrated result of comonotone improvement; see, for instance, Jouini et al. (2008) in the context of optimal risk sharing between two agents. To be consistent with the reinsurance literature and its practice, we still impose the feasibility conditions on the ceded loss function I.

In addition to the assumptions of Proposition 4.5, if the set E_{h_0} has Lebesgue measure zero, then the optimal reinsurance contract is simplified to $I_{h_0}(x) \triangleq \int_0^x \mathbb{I}_{G_{h_0}}(t)\,\mathrm{d}t$ for $x \geq 0$. Then, the necessary condition for optimality of reinsurance contract given by the expression (28) becomes a sufficient condition. It implies that (I_{h_0}, h_0) is a saddle point of the function $f(I,h)$ on $\mathcal{I} \times \Phi_c$, i.e. $\sup_{h \in \Phi_c} f(I_{h_0}, h) = f(I_{h_0}, h_0) = \min_{I \in \mathcal{I}} f(I, h_0)$.

If ρ is a law-invariant and comonotonic additive coherent risk measure. There exists $h \in \Phi_1$ such that $\rho(X) = \int_0^1 \mathrm{VaR}_q(X)\mathrm{d}h(q)$ for all $X \in \mathcal{L}^\infty$ (Kusuoka (2001)). Thus, any reinsurance contract with the form of

$$I_0(x) = \int_0^x \left(\mathbb{I}_{G_h}(t) + \alpha(t)\mathbb{I}_{E_h}(t)\right) \mathrm{d}t, \qquad \text{for } x \geq 0, \tag{29}$$

where $\alpha : \mathbb{R}_+ \to [0,1]$ is any measurable function, will be the optimal reinsurance ceded loss function for such a choice of ρ. This result is mathematically equivalent to Proposition 3.1 of Jouini et al. (2008) in the context of risk sharing using convex monetary utility functions.

Remark 4.7. The Minimax Theorem is applied to obtain the result in Proposition 4.5. In order to make the Minimax Theorem applicable to very general risk functionals with representation (2), we need to work on $\mathcal{L}^\infty$ space because the set of all feasible ceded loss function $\mathcal{I}$ is a compact set. On the other hand, if we impose certain restrictions on risk functionals, then the boundness assumption on X can be relaxed. To see this, we take $X \geq 0$ and $X \in \mathcal{L}^p$, $1 \leq p < \infty$, and assume

$$\rho(X) = \sup_{h \in \Psi} \left\{ \int_0^1 \mathrm{VaR}_q(X) \mathrm{d}h(q) - \beta(h) \right\}, \tag{30}$$

where $\Psi \subseteq \Phi_c$ is a convex and compact subset and β is continuous with respect to the supremum norm on Ψ. Note that $\inf_{I \in \mathcal{I}} \sup_{h \in \Psi} f(I, h) = -\sup_{I \in \mathcal{I}} \inf_{h \in \Psi} -f(I, h)$. We can check that all conditions in the Minimax Theorem are satisfied.

1. Ψ is convex and compact.
2. $\mathcal{I}$ is convex.
3. $-f(\cdot, h)$ is linear and thus concave on $\mathcal{I}$.
4. $-f(I, \cdot)$ is convex and lower-semicontinuous on Ψ.

Therefore, we can apply the Minimax Theorem to interchange the supremum and infimum

$$\sup_{I \in \mathcal{I}} \inf_{h \in \Psi} -f(I, h) = \inf_{h \in \Psi} \sup_{I \in \mathcal{I}} -f(I, h),$$

and conduct similar argument in proof of Proposition 4.5 to solve the optimization problem (21). A simple example of (30) is $\Psi = \{h\}$ and in this case the Minimax Theorem is always applicable. Then we can get the optimal result (29) for unbounded random variable $X \in \mathcal{L}^p$.

A direct consequence of Proposition 4.5 gives a mathematical support to the optimality of an insurance with deductible. If $I(x) = 0$ for $0 \leq x \leq d$, then we call d the deductible of I. Insurance with deductible is commonly observed in practice. The most popular examples are deductible insurance defined as $I(x) = (x - d)^+$ and excess-of-loss insurance defined as $I(x) = \max\{(x - d)^+, M\}$. Many studies in the insurance literature discuss the optimality of either stop-loss or excess-of-loss insurance under different model assumptions, e.g. dependent insurable risks in Cai and Wei (2012), the reinsurer's default risk in Cai et al. (2014), the presence of exclusion clauses in Chi and Liu (2017), and the Pareto-optimal arrangement in Cai et al. (2017). The next corollary gives a general result on the presence of a deductible part for a monotone law-invariant convex risk functional.

Corollary 4.8. *Assume $\theta > 0$ and either $c = 0$ or ρ satisfies monotonicity (A1). There exists an insurance policy with deductible as an optimal solution to Problem (21). In particular, if the expected premium principle $\rho_{(1+\theta)g}(I(X)) = (1+\theta)\mathbb{E}[I(X)]$ is used, then a stop-loss insurance policy is an optimal solution.*

Example 4.1. Assume that the insurer uses ES at level $\alpha \in (0,1)$. Define a convex function $\beta : \Phi_1 \to \mathbb{R} \cup \{+\infty\}$ via $\beta(h) = 0$ if $h(t) = h_\alpha(t) \triangleq \frac{1}{\alpha} t\, \mathbb{I}_{[0,\alpha]}(t) + \mathbb{I}_{(\alpha,1]}(t)$ otherwise $\beta(h) = +\infty$. Then, ES_α can be induced by substituting this particular β into the expression (11) for the convex risk measure in Theorem 2.2. Since β only takes finite value at h_α, the function $\int_0^{\bar{X}} \phi_h \circ S_X(t)\mathrm{d}t - \beta(h)$ achieves its maximal value at h_α.

In the first case that $(1+\theta)g(\alpha) \geq 1$, since g is concave and h_α is linear on both $[0,\alpha]$ and $[\alpha, 1]$, we know that $(1+\theta)g \geq h_\alpha$ on $[0,1]$. Thus, (29) implies that buying no reinsurance is the optimal solution for the insurer. In the second case, we assume that $(1+\theta)g(\alpha) < 1$. Note that, functions h_α and $(1+\theta)g$ will cross at most once on $[0,\alpha)$. When $(1+\theta)g'(0) > \frac{1}{\alpha}$, they do cross, and denote by d_2^* the root of equation

$$\frac{1}{(1+\theta)\alpha} = \frac{g \circ S_X(d_2^*)}{S_X(d_2^*)},$$

then $d_2^* > a$. When $(1+\theta)g'(0) \leq \frac{1}{\alpha}$, g is always smaller or equal to h_α on $[0,1]$ and we use $d_2^* = \bar{X}$ in this case. It can be easily checked that $\mathrm{VaR}_\alpha(X) \leq d_2^*$ and $G_{h_\alpha} = [d_2^*, \bar{X}]$. Meanwhile, h_{δ_α} and $(1+\theta)g$ cross at most once $(\alpha, 1]$ at the point d_1^* such that $(1+\theta)g \circ S_X(d_1^*) = 1$. Thus, $[0, d_1^*] \subset (G_{h_\alpha} \cup E_{h_\alpha})^c$. Therefore, by using the expression (29), the optimal solution to Problem (21) when $\rho = \mathrm{ES}_\alpha$ is $I^*(x) = (x - d_1^*)^+ - (x - d_2^*)^+$, and the corresponding minimal value is

$$\begin{aligned}
\min_{I \in \mathcal{I}} \mathrm{ES}_\alpha \left(X - I(X) + \rho_{(1+\theta)g}(I(X))\right) &= \int_0^{\bar{X}} \phi_{h_\alpha} \circ S_X(t)\mathrm{d}t - \beta(h_\alpha) \\
&= d_1^* + \int_{d_2^*}^{\bar{X}} S_X(t)\mathrm{d}t + (1+\theta)\int_{d_1^*}^{d_2^*} g \circ S_X(t)\mathrm{d}t.
\end{aligned}$$

This result is consistent with the known result in Chi and Tan (2013).

4.3 The budget constraint problem

Next, we consider the budget constraint problem (21)-(22). In Proposition 4.5, an optimal solution I_0 to the free premium problem (21) has premium larger than or equal to $(1+\theta)\int_{G_{h_0}} g \circ S_X(x)\mathrm{d}x$. Thanks to (A1) monotonicity which is satisfied by Wang's premium principle, the budget constraint (22) is not binding if $p > (1+\theta)\int_{G_{h_0}} g \circ S_X(x)\mathrm{d}x$. Therefore, to avoid redundant argument, we assume

$$p \leq (1+\theta)\int_{G_{h_0}} g \circ S_X(x)\mathrm{d}x,$$

that is p is no larger than the minimal premium for optimal solutions to the free premium problem (21).

For each $h \in \Phi_c$ and $\lambda \geq 0$, write

$$\begin{aligned}
G_{h,\lambda} &= \{0 \leq x \leq \bar{X} : (\lambda + c)(1+\theta)g \circ S_X(x) < h \circ S_X(x)\}, \\
E_{h,\lambda} &= \{0 \leq x < \bar{X} : (\lambda + c)(1+\theta)g \circ S_X(x) = h \circ S_X(x)\}.
\end{aligned}$$

Theorem 4.9. *Assume $p \le (1+\theta)\int_{G_{h_0}} g \circ S_X(x)\mathrm{d}x$. There exists $h^* \in \Phi_c$ and $\lambda^* \ge 0$ such that*

$$I^*(x) = \int_0^x \left(\mathbb{I}_{G_{h^*,\lambda^*}}(t) + \alpha(t)\mathbb{I}_{E_{h^*,\lambda^*}}(t) \right) \mathrm{d}t, \tag{31}$$

is an optimal solution to Problem (21)-(22), where $\alpha : \mathbb{R}_+ \to [0,1]$ is any measurable function such that $\rho_{(1+\theta)g}(I^*(X)) = p$.

Example 4.2 (Mean-variance)**.** In the budget constraint problem, suppose the insurer's objective is given by the mean-variance measure in (9) and the reinsurer uses an expected premium principle with $\theta > 0$. According to Example 2.2, for any retained loss function R, the supremum in (9) is attained at h such that $h'(q) = 2(\mathrm{VaR}_q(R(X)) - \mathbb{E}[R(X)])$. For each $\lambda > 0$, since $(\lambda(1+\theta)+c)g(\cdot)$ is linear and $h(\cdot)$ is concave, there exists $x_{h,\lambda} \ge 0$ such that $G_{h,\lambda} = (x_{h,\lambda}, \bar{X}]$. Taking $\alpha(x) = 0$ in (31) implies that I^* is a stop-loss function, a result that is well known in the literature since the seminal work of Arrow (1963). Under our assumption that $S_X(\cdot)$ is continuous and strictly decreasing, there exists a deductible d^* such that the budget constraint is achieved at the boundary, i.e. $(1+\theta)\mathbb{E}[(X-d^*)^+] = p$.

Example 4.3 (Signed Choquet integrals)**.** Suppose the insurer's objective is given by a signed Choquet integral with a concave distortion function $h \in \Phi_c$ and the reinsurer uses Wang's premium principle with an increasing concave distortion function g.

The optimal solution I_0 to the free premium problem (21) has form given in (28) with $h_0 = h$. In particular, write $I_h(x) = \int_0^x \mathbb{I}_{G_h}(t)\mathrm{d}t$ for $x \ge 0$.

For the budget constraint problem (21)-(22), if $p \ge \rho_{(1+\theta)g}(I_h(X))$, then there exists measurable function $\alpha : \mathbb{R}_+ \to [0,1]$ such that I_0 in (28) satisfies the budget constraint.

Next, we assume $p < \rho_{(1+\theta)g}(I_h(X)) = (1+\theta)\int_{G_h} g \circ S_X(x)\mathrm{d}x$. Applying the Lagrangian multiplier method in the proof of Theorem 4.9, there exists $\lambda_0 > 0$ such that

$$(1+\theta)\int_{G_{h,\lambda_0}} g \circ S_X(x)\mathrm{d}x \le p \le (1+\theta)\int_{G_{h,\lambda_0} \cup E_{h,\lambda_0}} g \circ S_X(x)\mathrm{d}x.$$

Thus, we can conclude that an optimal solution I^* to Problem (21)-(22) satisfies

$$I^*(x) = \int_0^x (\mathbb{I}_{G_{h,\lambda_0}}(t) + \alpha(t)\mathbb{I}_{E_{h,\lambda_0}}(t))\mathrm{d}t$$

where $\alpha : \mathbb{R}_+ \to [0,1]$ is a measurable function and $\rho_{(1+\theta)g}(I^*(X)) = p$. In particular, we take $g(x) = x$ for $x \in [0,1]$. Since h is concave and continuous on $(0,1)$, there exits $t_0 \in [0,1]$ such that $(c+\lambda_0)(1+\theta)t > h(t)$ for $t \in (t_0, 1)$ if the interval is non-empty, and $(c+\lambda_0)(1+\theta)t$ is either strictly smaller than or equal to $h(t)$ on the whole interval $(0, t_0)$. A direct consequence is that $I^*(x) = 0$ for $x \in [0, \mathrm{VaR}_{t_0}(X)]$, i.e., I^* is has a positive deductible if $t_0 < S_X(0)$. On the interval $(0, t_0)$, if $h(t) > (c+\lambda_0)(1+\theta)t$ for all $t \in (0, t_0)$, then $G_{h,\lambda_0} = (\mathrm{VaR}_{t_0}(X), \infty)$ and E_{h,λ_0} has measure zero; if $h(t) = (c+\lambda_0)(1+\theta)t$ for all $t \in (0, t_0)$, then $G_{h,\lambda_0} = \emptyset$ and $G_{h,\lambda_0} \cup E_{h,\lambda_0} = [\mathrm{VaR}_{t_0}(X), \infty)$. In both case, we can write an optimal solution in the stop-loss form $I^*(x) = (x-d)^+$ where $d \ge \mathrm{VaR}_{t_0}(X)$

such that $\rho_{(1+\theta)g}(I^*(X)) = p$. Note that in this example, the insurer's objective is not necessarily monotone, and the stop-loss contract is still optimal.

5 Conclusion

In this paper, we introduce the family of law-invariant convex risk functionals and obtain its representation. This family covers a broad range of existing risk measures such as convex risk measures and deviation measures. Two applications of our main representation result in optimization problems are studied. First, we derive the expression of the worst-case value of a convex risk functional when only partial information (the mean and the variance or a higher moment) of the risk is available. Second, we obtain forms and properties of optimal reinsurance polices when the insurer adopts a law-invariant convex risk functional while the reinsurer uses Wang's premium principle. In general, we show the existence of optimal solutions and the optimality of multi-layer ceded loss functions. These applications illustrate that the new family of risk functionals is powerful and flexible, and many classic results in the literature can be generalized without assuming monotonicity of the underlying objectives.

Remark 5.1. Our framework is built for static risk functionals, that is, risks are modelled as one-period random losses. In the literature of risk measures, risk functionals are also studied in a dynamic setting; see, for instance, Weber (2006) and Kupper and Schachermayer (2009) for dynamic consistency of risk measures and its relation to loss functions. In view of relevant actuarial applications, such as the robust risk assessment and the optimal reinsurance designin Sections 3 and 4, we leave dynamic aspects of convex risk functionals out of the scope of this paper, and they are certainly interesting for future research.

Acknowledgements. The authors thank Niushan Gao, Ludger Rüschendorf, Steven Vanduffel, two reviewers, an Associate Editor, and an Editor for helpful discussions and comments on various technical issues. Liu acknowledges financial support from the 111 Project (B17050) and the National Natural Science Foundation of China (NNSFC-11601540). Cai, Lemieux and Wang acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-03975, RGPIN-2015-238959 and RGPIN-2018-03823/RGPAS-2018-522590, respectively). Wang also acknowledges support from the Society of Actuaries CAE Grant. A preliminary version of Section 4 of this paper was titled "Optimal reinsurance in a market of multiple reinsurers under law-invariant convex risk measures" (SSRN: 3062767).

A Appendix: Proofs of results in Section 4

First, we present an auxiliary result[2] which will be useful in the proofs of Lemma 4.4 and Theorem 4.9.

Lemma A.1. *For $X \geq 0$ with continuous survival function $S_X(x)$ on $(0, \bar{X}]$ where $\bar{X} < \infty$ is the essential supremum of X, let $J : [0, \bar{X}] \to [0, \infty)$ be a nondecreasing 1-Lipschitz continuous mapping with $J(0) = 0$ and derivative J'. Then for any concave function $h : [0, 1] \to \mathbb{R}$ satisfying $h(0) = 0$, the associated signed*

[2] We thank a referee for suggesting this lemma.

Choquet integral ρ_h satisfies

$$\rho_h(J(X)) = \int_0^1 \mathrm{VaR}_\alpha(J(X))\mathrm{d}h(\alpha) = \int_0^{\bar{X}} J'(x)h\left(\mathbb{P}(X > x)\right)\mathrm{d}x.$$

Proof. Since J is nondecreasing and continuous, $\mathrm{VaR}_\alpha(J(X)) = J(\mathrm{VaR}_\alpha(X))$ for all $\alpha \in [0,1]$. For $X \geq 0$ with continuous survival function $S_X(x)$ on $(0, \bar{X}]$, we know $\mathrm{VaR}_0(X) = \bar{X}$, $\mathrm{VaR}_\alpha(X) = 0$ for $\alpha \in [S_X(0), 1]$ and $S_X(\mathrm{VaR}_\alpha(X)) = \alpha$ for $\alpha \in [0, S_X(0)]$. For a given $h \in \Phi$, by Lemma 2.1 (i), a change of variable and integration by parts, we have

$$\begin{aligned}
\rho_h(J(X)) &= \int_0^1 \mathrm{VaR}_\alpha(J(X))\mathrm{d}h(\alpha) \\
&= \int_0^1 J(\mathrm{VaR}_\alpha(X))\mathrm{d}h(\alpha) \\
&= \int_0^1 \mathbb{I}_{(0,1)}(\alpha)J(\mathrm{VaR}_\alpha(X))\mathrm{d}h(\alpha) + h(0+)J(\mathrm{VaR}_0(X)) + \left(h(1-) - h(1)\right)J(\mathrm{VaR}_1(X)) \\
&= -\int_0^{\bar{X}} \mathbb{I}_{(0,\bar{X})}(x)J(x)\mathrm{d}h(S_X(x)) + h(0+)J(\bar{X}) \\
&= \int_0^{\bar{X}} \mathbb{I}_{(0,\bar{X})}(x)h(S_X(x))J'(x)\mathrm{d}x = \int_0^{\bar{X}} J'(x)h(\mathbb{P}(X > x))\mathrm{d}x.
\end{aligned}$$

□

Proof of Lemma 4.4. To apply Theorem 4.2 to interchange the minimum sign and the supremum sign in the Problem (25), the following conditions should be checked carefully:

1) $\mathcal{I}$ is a compact set under the usual supremum norm $\|\cdot\|_\infty$ and $\mathcal{I}$ is convex.
2) Φ_c is convex.
3) For each fixed $h \in \Phi_c$, $f(\cdot, h)$ is continuous in I under the supremum norm. Note that VaR is comonotonic additive. Given $I_1, I_2 \in \mathcal{I}$ and $\lambda \in [0,1]$, $f(\lambda I_1 + (1-\lambda)I_2, h) = \lambda f(I_1, h) + (1-\lambda)f(I_2, h)$ for all $h \in \Phi_c$. Thus, $f(\cdot, h)$ is convex on $\mathcal{I}$.
4) Given $h_1, h_2 \in \Phi_c$ and $\lambda \in [0,1]$, the function $f(I, \lambda h_1 + (1-\lambda)h_2) \leq \lambda f(I, h_1) + (1-\lambda)f(I, h_2)$ for all $I \in \mathcal{I}$ due to the convexity of the function β. Thus, $f(I, \cdot)$ is concave on Φ_c. Also note that $f(I, \cdot)$ is upper-semicontinuous since β is lower-semicontinuous.

Therefore, by applying Theorem 4.2, we get

$$\min_{I \in \mathcal{I}} \sup_{h \in \Phi_c} f(I, h) = \sup_{h \in \Phi_c} \min_{I \in \mathcal{I}} f(I, h).$$

For each $h \in \Phi_c$, we first solve $\min_{I \in \mathcal{I}} f(I, h)$ and then express the minimizer I_h in term of h.

To this end, we are going to find the minimizer of function $f(\cdot, h)$ among $\mathcal{I}$. Since functions g, h, I' and $1 - I'$ are all non-negative, and by Lemma A.1, one obtains

$$\begin{aligned}
f(I, h) &= c(1+\theta)\int_0^{\bar{X}} g \circ S_X(t)I'(t)\,\mathrm{d}t + \int_0^{\bar{X}} h \circ S_X(t)\left(1 - I'(t)\right)\mathrm{d}t - \beta(h) \\
&\geq \int_0^{\bar{X}} \min\left\{c(1+\theta)g \circ S_X(t), h \circ S_X(t)\right\}\left[I'(t) + 1 - I'(t)\right]\mathrm{d}t - \beta(h) \\
&= \int_0^{\bar{X}} \phi_h \circ S_X(t)\,\mathrm{d}t - \beta(h).
\end{aligned}$$

Conversely, we can check that the function $I_h(t)$ defined in (26) satisfies $I_h \in \mathcal{I}$, and

$$f(I_h, h) = c(1+\theta)\int_{G_h} g \circ S_X(t)\mathrm{d}t + \int_{G_h^c} h \circ S_X(t)\mathrm{d}t - \beta(h) = \int_0^{\bar{X}} \phi_h \circ S_X(t)\mathrm{d}t - \beta(h).$$

Thus, the function I_h is a minimizer of $\min_{I\in\mathcal{I}} f(I,h)$, and moreover, $S \triangleq \sup_{h\in\Phi_c}\left\{\int_0^{\bar{X}} \phi_h \circ S_X(t)\mathrm{d}t - \beta(h)\right\}$ is the saddle value of the minimax problem (25).

Now, take a sequence $h_n \in \Phi_c$ such that $S = \lim_{n\to\infty} f(I_{h_n}, h_n)$. First assume that $\{h_n : n \in \mathbb{N}\}$ is uniformly bounded in the usual supremum norm on $[0,1]$, that is $\{\|h_n\|_\infty : n \in \mathbb{N}\}$ is a bounded sequence. Since h_n', $n \in \mathbb{N}$ is a decreasing function, $\|h_n'\|_1 \le 2\|h_n\|_\infty - c$, that is $\{h_n : n \in N\}$ has uniformly bounded total variation on $[0,1]$. By the Helly selection theorem, there exists a subsequence $\{h_{n_k} : k \in \mathbb{N}\}$, which pointwise converges to some function, denoted by h_0. For simplicity, we say that $\{h_n : n \in \mathbb{N}\}$ pointwise converges to h_0. For each $\lambda \in (0,1)$ and $x, y \in [0,1]$,

$$h_0(\lambda x + (1-\lambda)y) = \lim_{n\to\infty} h_n(\lambda x + (1-\lambda)y) \ge \lim_{n\to\infty} (\lambda h_n(x) + (1-\lambda)h_n(y)) = \lambda h_0(x) + (1-\lambda)h_0(y).$$

Meanwhile $h_0(0) = \lim_{n\to\infty} h_n(0) = 0$ and $h_0(1) = \lim_{n\to\infty} h_n(1) = c$. Thus, $h_0 \in \Phi_c$. Second, we assume that the sequence $\{\|h_n\|_\infty : n \in \mathbb{N}\}$ is unbounded. Then $\lim_{n\to\infty}\beta(h_n) = \infty$ and $S = -\infty$ contradicting the definition of S. In short, we can alway find $h_0 \in \Phi_c$ such that h_n is uniformly bounded and pointwise converges to h_0 by choosing a subsequence. It follows that ϕ_{h_n} converges to ϕ_{h_0} pointwise. Since $0 \le \phi_{h_n}(S_X(x)) \le g(S_X(x))$ for all $x \in [0, \bar{X}]$ and all $n \in \mathbb{N}$, and $\int_0^{\bar{X}} g(S_X(x))\mathrm{d}x < \infty$ by assumption, the Dominated Convergence Theorem implies that

$$\lim_{n\to\infty}\int_0^{\bar{X}} \phi_{h_n}(S_X(x))\mathrm{d}x = \int_0^{\bar{X}} \phi_{h_0}(S_X(x))\mathrm{d}x.$$

Furthermore, the fact that β is non-negative and lower-semi continuos implies that

$$\limsup_{n\to\infty} -\beta(h_n) = -\liminf_{n\to\infty}\beta(h_n) \le -\beta(h_0).$$

Thus,

$$\begin{aligned} S = \lim_{n\to\infty} f(I_{h_n}, h_n) &\le \limsup_{n\to\infty}\int_0^{\bar{X}} \phi_{h_n} \circ S_X(t)\mathrm{d}t + \limsup_{n\to\infty} -\beta(h_n) \\ &\le \int_0^{\bar{X}} [h_0 \circ S_X(t)] \wedge [g \circ S_X(t)]\,\mathrm{d}t - \beta(h_0) = f(I_{h_0}, h_0) \le S. \end{aligned}$$

As a consequence, the minimax problem (25) has the saddle-value $S = \int_0^{\bar{X}} \phi_{h_0} \circ S_X(t)\mathrm{d}t - \beta(h_0)$, and an optimal solution I_0 to Problem (21) satisfies

$$\rho\left(X - I_0(X) + \rho_{(1+\theta)g}(I_0(X)\right) = \min_{I\in\mathcal{I}}\sup_{h\in\Phi_c} f(I,h) = \sup_{h\in\Phi_c}\min_{I\in\mathcal{I}} f(I,h) = S. \quad \square$$

Proof of Proposition 4.5. Suppose that I_0 is an optimal solution of Problem (21). By Theorem 4.4, we have $\rho\left(X - I_0(X) + \rho_{(1+\theta)g}(I_0(X))\right) = \sup\limits_{h\in\Phi_c} f(I_0, h) = S$. It follows that

$$S = f(I_{h_0}, h_0) = \min_{I\in\mathcal{I}} f(I, h_0) \le f(I_0, h_0) \le \sup_{h\in\Phi_c} f(I_0, h) = S,$$

and thus $f(I_{h_0}, h_0) = f(I_0, h_0)$. By conditions in (26), we have

$$f\left(I_{\mu_0}, \mu_0\right) = \int_{G_{\mu_0}} g \circ S_X(t)\mathrm{d}t + \int_{G_{\mu_0}^c} h_{\mu_0} \circ S_X(t)\mathrm{d}t.$$

A direct calculation gives us

$$\begin{aligned}
0 = f(I_0, h_0) - f(I_{h_0}, h_0) &= \int_0^{\bar{X}} c(1+\theta) g \circ S_X(t) I_0'(t)\mathrm{d}t + \int_0^{\bar{X}} h_0 \circ S_X(t)[1 - I_0'(t)]\mathrm{d}t \\
&\quad - \int_{G_{h_0}} c(1+\theta) g \circ S_X(t)\mathrm{d}t - \int_{G_{h_0}^c} h_{\mu_0} \circ S_X(t)\mathrm{d}t \\
&= \int_{G_{h_0}} \left[h_0 \circ S_X(t) - c(1+\theta) g \circ S_X(t)\right][1 - I_0'(t)]\mathrm{d}t \qquad (32) \\
&\quad + \int_{(G_{h_0} \cup E_{h_0})^c} \left[c(1+\theta) g \circ S_X(t) - h_0 \circ S_X(t)\right] I_0'(t)\mathrm{d}t \\
&\quad + \int_{E_{h_0}} \left[c(1+\theta) g \circ S_X(t) I_0'(t) + h_0 \circ S_X(t)(1 - I_0'(t)) - h_0 \circ S_X(t)\right]\mathrm{d}t.
\end{aligned}$$

The first two integrands on the right hand of the equality (32) are both non-negative and the third term is zero. Therefore, we have $I_0'(t) = 0$ for $t \in (G_{h_0} \cup E_{h_0})^c$ and $I_0'(t) = 1$ for $t \in G_{h_0}$. Since I satisfies the 1-Lipschitz continuity property, its right derivative I' on the set E_{h_0} is a measurable function α taking values in $[0, 1]$. $\square$

Proof of Corollary 4.8. If $c = 0$, Proposition 4.5 implies that a full insurance, which is a special case of deductible insurance, is the optimal solution to Problem (21).

Assume $c > 0$ and ρ is monotone. In this case, the representation (2) can be taken such that $\beta(h) = \infty$ for all h that is not increasing (Theorem 4.59 of Föllmer and Schied (2011)). For each increasing $h \in \Phi_c$, $h(1) = c$. Since g is continuous on $(0, 1]$, h is increasing on $(0, 1)$ and $c(1+\theta)g(1) = c(1+\theta)$, there exists $t_0 \in [0, 1)$ such that $c(1+\theta)g(t) > c \geq h(t)$ for all $t \in (t_0, 1)$. Thus, $(0, \mathrm{VaR}_{t_0}(X)) \subset (G_h \cup E_h)^c$ and $I_h(x) = 0$ for $x \in [0, \mathrm{VaR}_{t_0}(X)]$.

If g is linear, i.e., $g(t) = t$, since h is a concave function, there exists $t_0 \in [0, 1)$ such that $h(t) \geq c(1+\theta)g(t)$ for $t \in [0, t_0]$ and $h(t) < c(1+\theta)g(t)$ for $t \in (t_0, 1]$. Applying (28) with $\alpha(t) = 1$, we obtain an optimal solution of the stop-loss type. $\square$

Proof of Theorem 4.9. Recall that $c \geq 0$, and β, θ, and g are fixed. Without the budget constraint, Proposition 4.5 shows that the free premium problem (21) has solutions in the form (28) for some $h_0 \in \Phi_c$, and h_0 depends on c. Moreover, premium amounts of optimal solutions in (28) are all larger than or equal to $\rho_{(1+\theta)g}(I_{h_0}(X)) = (1+\theta)\int_{G_{h_0}} g \circ S_X(t)\mathrm{d}t$, where $I_{h_0}(x) = \int_0^x \mathbb{I}_{G_{h_0}}(t)\mathrm{d}t$ for $x \geq 0$. If $p > \rho_{(1+\theta)g}(I_{h_0}(X))$, then I_{h_0} is an optimal solution to the constrained problem (21)-(22), and the constraint (22) is not binding.

For $p \leq \rho_{(1+\theta)g}(I_{h_0}(X))$, we claim that the constraint (22) is binding. Suppose Problem (21)-(22) admits a solution $\tilde{I}$ and $\rho_{(1+\theta)g}(\tilde{I}(X)) < p$. First note that

$$\rho\left(X - I_{h_0}(X) + \rho_{(1+\theta)g}(I_{h_0}(X))\right) < \rho\left(X - \tilde{I}(X) + \rho_{(1+\theta)g}(\tilde{I}(X))\right).$$

Since the Choquet integral $\rho_{(1+\theta)g}$ is additive for comonotonic random variables, there exists $\lambda \in [0, 1)$ such that

$$\rho_{(1+\theta)g}(I(X)) = \lambda\rho_{(1+\theta)g}(\tilde{I}(X)) + (1-\lambda)\rho_{(1+\theta)g}(I_{h_0}(X)) = p$$

where $I = \lambda\tilde{I} + (1-\lambda)I_{h_0}$, i.e., I satisfies (22). Due to the convexity of ρ, we have

$$\begin{aligned}\rho(X - I(X) + \rho_{(1+\theta)g}(I(X))) &\le \lambda\rho\left(X - \tilde{I}(X) + \rho_{(1+\theta)g}(\tilde{I}(X))\right) + (1-\lambda)\rho\left(X - I_{h_0}(X) + \rho_{(1+\theta)g}(I_{h_0}(X))\right)\\ &< \rho\left(X - \tilde{I}(X) + \rho_{(1+\theta)g}(\tilde{I}(X))\right),\end{aligned}$$

which contradicts the optimality of $\tilde{I}$. The claim is proved.

In particular, if $p = \rho_{(1+\theta)g}(I_{h_0}(X))$, then I_{h_0} is an optimal solution to the constraint problem (21)-(22) and the constraint (22) is binding.

In the rest of the proof, we assume $p < \rho_{(1+\theta)g}(I_{h_0}(X)) = (1+\theta)\int_{G_{h_0}} g \circ S_X(t)\mathrm{d}t$. We translate the constrained minimization problem (21)-(22) to a non-constrained problem by using the Lagrangian multiplier method. Consider the following minimization problem

$$\max_{\lambda\ge 0}\min_{I\in\mathcal{I}}\left\{\rho\left(R(X) + g_{(1+\theta)g}(I(X))\right) + \lambda g_{(1+\theta)g}(I(X)) - \lambda p\right\} = \max_{\lambda\ge 0}\min_{I\in\mathcal{I}}\sup_{h\in\Phi_c}\left\{f_\lambda(I,h) - \lambda p\right\}, \tag{33}$$

where (again using Lemma A.1)

$$f_\lambda(I,h) = \int_0^{\bar{X}} h\circ S_X(x)R'(x)\mathrm{d}x + (1+\theta)(\lambda + c)\int_0^{\bar{X}} g\circ S_X(x)I'(x)\mathrm{d}x - \beta(h). \tag{34}$$

Similar as the free premium problem, we define

$$\begin{aligned}G_{h,\lambda} &= \{x \ge 0 : (\lambda + c)(1+\theta)g\circ S_X(x) < h\circ S_X(x)\},\\ E_{h,\lambda} &= \{0 \le x < \bar{X} : (\lambda + c)(1+\theta)g\circ S_X(x) = h\circ S_X(x)\},\end{aligned}$$

and $I_{h,\lambda}(x) = \int_0^x \mathbb{I}_{G_{h,\lambda}}(t)\mathrm{d}t$ for $x \ge 0$. Thus, for each $\lambda \ge 0$ and $h \in \Phi_c$, we know $f_\lambda(I,h) \ge f_\lambda(I_{h,\lambda},h)$. For a fixed $\lambda \ge 0$, because λp is a constant and does not affect the min-sup problem, we have

$$\min_{I\in\mathcal{I}}\sup_{h\in\Phi_c}\left\{f_\lambda(I,h) - \lambda p\right\} = \sup_{h\in\Phi_c}\min_{I\in\mathcal{I}}\left\{f_\lambda(I,h) - \lambda p\right\} = \sup_{h\in\Phi_c}\left\{f_\lambda(I_{h,\lambda},h) - \lambda p\right\}.$$

Thus, (33) becomes

$$\max_{\lambda\ge 0}\min_{I\in\mathcal{I}}\left\{\rho\left(R(X) + g_{(1+\theta)g}(I(X))\right) + \lambda g_{(1+\theta)g}(I(X)) - \lambda p\right\} = \sup_{h\in\Phi_c}\max_{\lambda\ge 0}\left\{f_\lambda(I_{h,\lambda},h) - \lambda p\right\}.$$

Since $f_\lambda(I_{h,\lambda},h) = (\lambda + c)(1+\theta)\int_{G_{h,\lambda}} g\circ S_X(x)\mathrm{d}x + \int_{G^c_{h,\lambda}} h\circ S_X(x)\mathrm{d}x - \beta(h)$, the right and left derivatives of $f_\lambda(I_{h,\lambda},h) - \lambda p$ with respect to λ are

$$\frac{\mathrm{d}}{\mathrm{d}\lambda^+}f_\lambda(I_{h,\lambda},h) = (1+\theta)\int_{G_{h,\lambda}} g\circ S_X(x)\mathrm{d}x, \quad \frac{\mathrm{d}}{\mathrm{d}\lambda^-}f_\lambda(I_{h,\lambda},h) = (1+\theta)\int_{G_{h,\lambda}\cup E_{h,\lambda}} g\circ S_X(x)\mathrm{d}x.$$

Define

$$\lambda(h) = \sup\left\{\lambda \ge 0 : (1+\theta)\int_{G_{h,\lambda}} g\circ S_X(x)\mathrm{d}x > p\right\}, \tag{35}$$

where $\lambda(h) = 0$ if the set on the right hand side is an empty set.

In the rest of argument, for simplicity, we assume $\theta = 0$ without loss of generality. Note that for a given

$h \in \Phi_c$, if $\lambda_1 \leq \lambda_2$ then $G_{h,\lambda_2} \subseteq G_{h,\lambda_1}$; if $\lambda_1 < \lambda_2$ then

$$G_{h,\lambda_2} \cup E_{h,\lambda_2} = \{0 \leq x < \bar{X} : \lambda_2 + c \leq \frac{h \circ S_X(x)}{g \circ S_X(x)}\} \subseteq \{0 \leq x < \bar{X} : \lambda_1 + c < \frac{h \circ S_X(x)}{g \circ S_X(x)}\} = G_{h,\lambda_1}.$$

For each $x \notin G_{h,\lambda(h)} \cup E_{h,\lambda(h)}$, if any, there exists λ such that $\lambda(h) > \lambda > \frac{h \circ S_X(x)}{g \circ S_X(x)} - c$, i.e. $x \notin G_{h,\lambda}$. Thus, $\mathbb{I}_{G_{h,\lambda}} \to \mathbb{I}_{G_{h,\lambda(h)} \cup E_{h,\lambda(h)}}$ pointwise for $\lambda \uparrow \lambda(h)$. It follows that

$$p \leq \lim_{\lambda \uparrow \lambda(h)} \int \mathbb{I}_{G_{h,\lambda}}(x)\, g \circ S_X(x) \mathrm{d}x = \int \mathbb{I}_{G_{h,\lambda(h)} \cup E_{h,\lambda(h)}}(x)\, g \circ S_X(x) \mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}\lambda^-} f_{\lambda(h)}(I_{h,\lambda(h)}, h).$$

Similarly, $\mathbb{I}_{G_{h,\lambda} \cup E_{h,\lambda}} \to \mathbb{I}_{G_{h,\lambda(h)}}$ pointwise as $\lambda \downarrow \lambda(h)$, and

$$\lim_{\lambda \downarrow \lambda(h)} \int \mathbb{I}_{G_{h,\lambda} \cup E_{h,\lambda}}(x)\, g \circ S_X(x) \mathrm{d}x = \int \mathbb{I}_{G_{h,\lambda(h)}}(x)\, g \circ S_X(x) \mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}\lambda^+} f_{\lambda(h)}(I_{h,\lambda(h)}, h).$$

Suppose there exists $\lambda_3 > \lambda(h)$ such that $\frac{\mathrm{d}}{\mathrm{d}\lambda^-} f_{\lambda_3}(I_{h,\lambda_3}, h) \geq p$. Take $\tilde{\lambda}_3 \in (\lambda(h), \lambda_3)$. Then we get the contradiction

$$p \leq \frac{\mathrm{d}}{\mathrm{d}\lambda^-} f_{\lambda_3}(I_{h,\lambda_3}, h) < \frac{\mathrm{d}}{\mathrm{d}\lambda^+} f_{\tilde{\lambda}_3}(I_{h,\tilde{\lambda}_3}, h) \leq p$$

where the last inequality is induced from the definition of $\lambda(h)$. Thus,

$$p \geq \lim_{\lambda \downarrow \lambda(h)} \int \mathbb{I}_{G_{h,\lambda} \cup E_{h,\lambda}}(x)\, g \circ S_X(x) \mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}\lambda^+} f_{\lambda(h)}(I_{h,\lambda(h)}, h).$$

In short, we now have

$$\sup_{h \in \Phi_c} \max_{\lambda \geq 0} \{f_\lambda(I_{h,\lambda}, h) - \lambda p\} = \sup_{h \in \Phi_c} \{f_{\lambda(h)}(I_{h,\lambda(h)}, h) - \lambda(h) p\}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\lambda^+} f_{\lambda(h)}(I_{h,\lambda(h)}, h) \leq p \leq \frac{\mathrm{d}}{\mathrm{d}\lambda^-} f_{\lambda(h)}(I_{h,\lambda(h)}, h).$$

Take a sequence $\{h_n\}_{n=1}^\infty \subseteq \Phi_c$ such that $f_{\lambda(h_n)}(I_{h_n,\lambda(h_n)}, h) - \lambda(h_n) p \to \sup_{h \in \Phi_c} f_{\lambda(h)}(I_{h,\lambda(h)}, h) - \lambda(h) p$. By the argument in the proof of Lemma 4.4, there exists a subsequence of $\{h_n\}$ converging to some $h^* \in \Phi_c$. Correspondingly, we obtain a sequence of non-negative real number (extended to positive infinity) $\{\lambda(h_n)\}$, which has a subsequence converging to some $\lambda^* \in [0, \infty]$. Without loss of generality, we take $\{h_n\}$ converging to h_0 pointwise and $\{\lambda(h_n)\}$ converging to λ^*. For each $n \in \mathbb{N}$ and $x \geq 0$, we have $\frac{h_n \circ S_X(x)}{\lambda(h_n)} \leq \frac{\max_{m \geq n}\{h_m \circ S_X(x)\}}{\min_{m \geq n}\{\lambda(h_m)\}}$. Denote by $\hat{h}_n = \max_{m \geq n}\{h_m\}$ and $\hat{\lambda}_n = \min_{m \geq n}\{\lambda(h_m)\}$. Then $G_{h_n,\lambda(h_n)} \subseteq G_{\hat{h}_n,\hat{\lambda}_n}$, $G_{h_n,\lambda(h_n)} \cup E_{h_n,\lambda(h_n)} \subseteq G_{\hat{h}_n,\hat{\lambda}_n} \cup E_{\hat{h}_n,\hat{\lambda}_n}$ and

$$p \leq \int \mathbb{I}_{G_{h_n,\lambda(h_n)} \cup E_{h_n,\lambda(h_n)}}(x)\, g \circ S_X(x) \mathrm{d}x \leq \int \mathbb{I}_{G_{\hat{h}_n,\hat{\lambda}_n} \cup E_{\hat{h}_n,\hat{\lambda}_n}}(x)\, g \circ S_X(x) \mathrm{d}x.$$

Since $\hat{\lambda}_n \leq \hat{\lambda}_{n+1}$ and $\hat{h}_n(x) \geq \hat{h}_{n+1}(x)$ for $x \geq 0$, we have $(G_{\hat{h}_{n+1},\hat{\lambda}_{n+1}} \cup E_{\hat{h}_{n+1},\hat{\lambda}_{n+1}}) \subseteq (G_{\hat{h}_n,\hat{\lambda}_n} \cup E_{\hat{h}_n,\hat{\lambda}_n})$ and $\cap_{n=1}^\infty (G_{\hat{h}_n,\hat{\lambda}_n} \cup E_{\hat{h}_n,\hat{\lambda}_n}) = G_{h^*,\lambda^*} \cup E_{h^*,\lambda^*}$. Thus, we have

$$p \leq \int \mathbb{I}_{G_{h^*,\lambda^*} \cup E_{h^*,\lambda^*}}(x)\, g \circ S_X(x) \mathrm{d}x.$$

On the contrary, since $\frac{h_n\circ S_X(x)}{\lambda(h_n)} \geq \frac{\inf_{m\geq n}\{h_m\circ S_X(x)\}}{\sup_{m\geq n}\{\lambda(h_m)\}}$, we have

$$p \geq \int_0^{\bar{X}} \mathbb{I}_{G_{h_n,\lambda(h_n)}}(x)\, g\circ S_X(x)\mathrm{d}x \geq \int_0^{\bar{X}} \mathbb{I}_{G_{\underline{h}_n,\bar{\lambda}_n}}(x)\, g\circ S_X(x)\mathrm{d}x,$$

where $\underline{h}_n = \inf_{m\geq n}\{h_m\circ S_X(x)\}$ and $\bar{\lambda}_n = \sup_{m\geq n}\{\lambda(h_m)\}$. Taking the limit leads to

$$p \geq \int_0^{\bar{X}} \mathbb{I}_{G_{h^*,\lambda^*}}(x)\, g\circ S_X(x)\mathrm{d}x.$$

Therefore, $\lambda^* = \lambda(h^*)$. It further implies that

$$\begin{aligned}
\lim_{n\to\infty} f_{\lambda(h_n)}(I_{h_n,\lambda(h_n)},h_n) &\leq \int_{G_{h^*,\lambda(h^*)}} (\lambda(h^*)+c)g\circ S_X(x)\mathrm{d}x + \int_{G^c_{h^*,\lambda(h^*)}} h^*\circ S_X(x)\mathrm{d}x + \limsup_{n\to\infty} -\beta(h_n) \\
&\leq \int_{G_{h_0,\lambda(h^*)}} (\lambda(h^*)+c)g\circ S_X(x)\mathrm{d}x + \int_{G^c_{h^*,\lambda(h^*)}} h^*\circ S_X(x)\mathrm{d}x - \beta(h^*).
\end{aligned}$$

Thus, (33) is solved that

$$\max_{\lambda\geq 0}\min_{I\in\mathcal{I}} \left\{\rho(R(X)+\rho_{(1+\theta)g}(I(X))) + \lambda\rho_{(1+\theta)g}(I(X)) - \lambda p\right\} = f_{\lambda(h^*)}(I_{h^*,\lambda(h^*)},h^*) - \lambda(h^*)p.$$

The optimal reinsurance policy I^* satisfying $I^*(x) = \int_0^x (\mathbb{I}_{G_{h^*,\lambda^*}}(t) + \alpha(t)\mathbb{I}_{E_{h^*,\lambda^*}}(t))\mathrm{d}t$, where $\alpha : \mathbb{R}_+ \to [0,1]$ is any measurable function such that $\rho_{(1+\theta)g}(I^*(X)) = p$. $\square$

References

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & 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., F. Delbaen, J. M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance 9(3)*, 203–228.

Cai, J., C. Lemieux, and F. Liu (2014). Optimal reinsurance with regulatory initial capital and default risk. *Insurance: Mathematics & Economics 57*, 13–24.

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

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

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

Cai, J. and W. Wei (2012). Optimal reinsurance with positively dependent risks. *Insurance: Mathematics & Economics 50(1)*, 57–63.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2012). Signed integral representations of comonotonic additive functionals. *Journal of Mathematical Analysis and Applications 385*(2), 895–912.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2015). Choquet integration on Riesz spaces and dual comonotonicity. *Transactions of the American Mathematical Society 367*(12), 8521–8542.

Cheung, K. C. (2010). Optimal reinsurance revisited–A geometric approach. *ASTIN Bulletin 40(1)*, 221–239.

Cheung, K. C., W. F. Chong, and A. Lo (2019). Budget-constrained optimal reinsurance design under coherent risk measures. *Scandinavian Actuarial Journal 2019*(9), 729–751.

Cheung, K. C. and A. Lo (2017). Characterizations of optimal reinsurance treaties: a cost-benefit approach. *Scandinavian Actuarial Journal 2017*(1), 1–28.

Cheung, K. C., K. C. J. Sung, S. C. P. Yam, and S. P. Yung (2014). Optimal reinsurance under general law-invariant risk measures. *Scandinavian Actuarial Journal 2014*(1), 72–91.

Chi, Y. and F. Liu (2017). Optimal insurance design in the presence of exclusion clauses. *Insurance: Mathematics & Economics 76*, 185–195.

Chi, Y. and K. S. Tan (2013). Optimal reinsurance with general premium principles. *Insurance: Mathematics and Economics 52(2)*, 180–189.

Cornilly, D., L. Rüschendorf, and S. Vanduffel (2019). Upper bounds for concave distortion risk measures on moment spaces. *Insurance: Mathematics & Economics 82*, 141–151.

Delbaen, F. (2012). *Monetary Utility Functions*. Osaka University CSFI lecture notes series. Osaka University Press. ISBN: 978-4-87259-278-8.

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

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

Föllmer, H. and A. Schied (2011). *Stochastic Finance: An Introduction in Discrete Time, 3rd Edition*. Berlin: Walter de Gruyter GmbH & Co. KG.

Frittelli, M. and M. Maggis (2011). Dual representation of quasi-convex conditional maps. *SIAM Journal on Financial Mathematics 2*(1), 357–382.

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

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

Gerber, H. U., E. S. Shiu, and H. Yang (2019). A constraint-free approach to optimal reinsurance. *Scandinavian Actuarial Journal 2019*(1), 62–79.

Grechuk, B., A. Molyboha, and M. Zabarankin (2009). Maximum entropy principle with general deviation measures. *Mathematics of Operations Research 34*(2), 445–467.

Jouini, E., W. Schachermayer, and N. Touzi (2008). Optimal risk sharing for law invariant monetary utility functions. *Mathematical Finance 18*(2), 269–292.

Kosmol, P. and D. Müller-Wichards (2011). *Optimization in function spaces: with stability considerations in Orlicz spaces*, Volume 13. Walter de Gruyter.

Kupper, M. and W. Schachermayer (2009). Representation results for law invariant time consistent functions. *Mathematics and Financial Economics 2*(3), 189–210.

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

Li, L., H. Shao, R. Wang, and J. Yang (2018). Worst-case range value-at-risk with partial information. *SIAM Journal on Financial Mathematics 9*(1), 190–218.

Li, Y.-M. J. (2018). Technical note–closed-form solutions for worst-case law invariant risk measures with application to robust portfolio optimization. *Operations Research 66*(6), 1533–1541.

McNeil, A. J., R. Frey, and P. Embrechts (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton university press.

Rockafellar, R. (1974). *Conjugate duality and optimization*. Philadelphia: Regional Conference Series in Applied Mathematics, 16, SIAM.

Rockafellar, R. T., S. Uryasev, and M. Zabarankin (2006). Generalized deviations in risk analysis. *Finance and Stochastics 10*(1), 51–74.

Wang, R., Y. Wei, and G. Willmot (2019). Characterization, robustness and aggregation of signed choquet integrals. *Mathematics of Operations Research*. forthcoming.

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

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

Zhu, W. and H. Shao (2018). Closed-form solutions for extreme-case distortion risk measures and applications to robust portfolio management. Available at SSRN: https://ssrn.com/abstract=3103458.