A reverse ES (CVaR) optimization formula

Yuanying Guan[*]   Zhanyi Jiao[†]   Ruodu Wang[‡]

June 28, 2023

Abstract

The celebrated Expected Shortfall (ES, also known as TVaR and CVaR) optimization formula implies that ES at a fixed probability level is the minimum of a linear real function plus a scaled mean excess function. We establish a reverse ES optimization formula, which says that a mean excess function at any fixed threshold is the maximum of an ES curve minus a linear function. Despite being a simple result, this formula reveals elegant symmetries between the mean excess function and the ES curve, as well as their optimizers. The reverse ES optimization formula is closely related to the Fenchel-Legendre transforms, and our formulas are generalized from ES to optimized certainty equivalents, a popular class of convex risk measures. We analyze worst-case values of the mean excess function under two popular settings of model uncertainty to illustrate the usefulness of the reverse ES optimization formula, and this is further demonstrated with an application using insurance datasets.

Keywords: Tail Value-at-Risk, Conditional Value-at-Risk, mean excess loss, optimized certainty equivalents, Fenchel-Legendre transform

1 Introduction

The Value-at-Risk (VaR) and the Expected Shortfall (ES, also known as TVaR and CVaR) are the two most popular risk measures in banking and insurance, and they are widely employed in regulatory capital computation, decision making, performance analysis, and risk management. In particular, ES is the standard risk measure in the current Basel Accords as well as the Swiss Solvency Test, and VaR is standard in the insurance regulatory framework of Solvency II. For general treatments of VaR and ES in actuarial science and finance, see Denuit et al. (2005), Kaas et

---

[*]Department of Mathematical Sciences and Department of Finance & Real Estate, DePaul University, USA. ✉ yguan8@depaul.edu

[†]Department of Statistics and Actuarial Science, University of Waterloo, Canada. ✉ z27jiao@uwaterloo.ca

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

al. (2008) and McNeil et al. (2015). In this paper, we will stick to the term "ES" as used by BCBS (2019) for this risk measure, although the term "CVaR" is commonly found in the optimization literature.

An influential result on VaR and ES is an optimization formula obtained by Rockafellar and Uryasev (2000, 2002), which is the main motivation for this paper. We first give the definition of VaR and ES. Let $\mathcal{X}$ be the set of integrable random variables in an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. At a probability level $\alpha \in [0, 1]$, VaR has two versions as the left- and right-quantiles. For $X \in \mathcal{X}$, define

$$
\begin{aligned}
\mathrm{VaR}_\alpha^-(X) &= \inf\{t \in \mathbb{R} : \mathbb{P}(X \le t) \ge \alpha\}; \\
\mathrm{VaR}_\alpha^+(X) &= \inf\{t \in \mathbb{R} : \mathbb{P}(X \le t) > \alpha\}.
\end{aligned}
\tag{1}
$$

By definition, $\mathrm{VaR}_0^-(X) = -\infty$ and $\mathrm{VaR}_1^+(X) = \infty$. ES at probability level $\alpha \in [0, 1]$ is defined as

$$
\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_\beta^-(X) \, \mathrm{d}\beta, \quad X \in \mathcal{X}, \ \alpha \in [0, 1),
$$

and $\mathrm{ES}_1(X) = \mathrm{VaR}_1^-(X)$. It is well known that an ES is a coherent risk measure (Artzner et al. (1999)) and a convex risk measure (Föllmer and Schied (2016)), and it admits an axiomatization based on portfolio diversification (Wang and Zitikis (2021)). Below, we first present the celebrated formula of Rockafellar and Uryasev (2002).

Theorem 1 (ES optimization formula). *For $X \in \mathcal{X}$ and $\alpha \in (0, 1)$, it holds*

$$
\mathrm{ES}_\alpha(X) = \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{1-\alpha} \mathbb{E}[(X - t)_+] \right\},
\tag{2}
$$

and the set of minimizers for (2) *is* $[\mathrm{VaR}_\alpha^-(X), \mathrm{VaR}_\alpha^+(X)]$.

Theorem 1 has been a cornerstone of risk management and financial engineering since Rockafellar and Uryasev (2000, 2002) and Pflug (2000). This result has been tremendously useful in the optimization of ES (see Rockafellar and Uryasev (2013) for a review) and it has also been widely taught in actuarial science, see e.g., Denuit et al. (2005, Section 2.4.3) and Kaas et al. (2008, Section 5.6). Among other implications, this formula directly gives an elementary proof of subadditivity of ES; see Embrechts and Wang (2015) for a comparison with six other proofs.

In this paper, we establish a new optimization formula based on ES, which can be seen as a reverse formula to Theorem 1. This formula reveals nice symmetries between the ES curve and the mean excess function, as we discuss in Section 3. The mean excess loss, also known as the stop-loss premium, has a deep root in actuarial science (e.g., De Vylder and Goovaerts (1982)). In Section 4, we apply the new formula to two popular settings of model uncertainty, one induced

by information of mean and a higher moment and the other induced by a Wasserstein ball. In both settings, the worst-case ES admits an explicit formula in the recent literature (Pesenti et al. (2020); Liu et al. (2022)) whereas the worst-case mean excess function does not. Two insurance loss datasets are studied in Section 5 to illustrate the obtained results on the mean excess loss under model uncertainty induced by a Wasserstein ball. We present a few further technical results in Sections 6 and 7; more precisely, the reverse ES optimization formula are generalized to the class of optimized certainty equivalents introduced by Ben-Tal and Teboulle (2007) (Section 6), and two related formulas are obtained via Fenchel-Legendre transforms (Section 7). Section 8 concludes the paper.

2 A reverse ES optimization formula

We start from the observation from Theorem 1 that $\mathrm{ES}_\alpha(X)$ for a fixed $\alpha \in (0,1)$ can be obtained through taking the minimum of a function involving $\mathbb{E}[(X-t)_+]$ over $t \in \mathbb{R}$. Having a mathematical symmetry in mind, a natural question is whether $\mathbb{E}[(X-t)_+]$ for a fixed $t \in \mathbb{R}$ can be obtained through taking the maximum of a function involving $\mathrm{ES}_\alpha(X)$ over $\alpha \in [0,1]$. This leads to the reverse ES optimization formula, the main result of this paper. In what follows, we always use the convention $0 \times x = 0$ for $x \in [-\infty, \infty]$.

Theorem 2 (Reverse ES optimization formula)**.** *For $X \in \mathcal{X}$ and $t \in \mathbb{R}$, it holds*

$$\mathbb{E}[(X-t)_+] = \max_{\alpha \in [0,1]} \left\{ (1-\alpha)\left(\mathrm{ES}_\alpha(X) - t\right) \right\}, \tag{3}$$

and the set of maximizers for (3) *is* $[\mathbb{P}(X < t), \mathbb{P}(X \leq t)]$.

To prove Theorem 2, we first present a useful lemma which collects some standard properties of quantiles, which are known to specialists on quantiles. We provide a self-contained short proof since we could not find this precise formulation in the literature.

Lemma 1. *For $\alpha \in [0,1]$ and any random variable X, the following statements hold:*

(i) $\alpha > \mathbb{P}(X \leq t)$ if and only if $\mathrm{VaR}_\alpha^-(X) > t$;

(ii) $\alpha < \mathbb{P}(X < t)$ if and only if $\mathrm{VaR}_\alpha^+(X) < t$.

Remark 1. The statements in Lemma 1 can be equivalently stated in other forms, such as (i.a) $\mathbb{P}(X \leq t) \geq \alpha$ if and only if $\mathrm{VaR}_\alpha^-(X) \leq t$; (i.b) $\mathbb{P}(X > t) \leq 1-\alpha$ if and only if $\mathrm{VaR}_\alpha^-(X) \leq t$; (ii.a) $\mathbb{P}(X < t) \leq \alpha$ if and only if $\mathrm{VaR}_\alpha^+(X) \geq t$; (ii.b) $\mathbb{P}(X \geq t) \geq 1-\alpha$ if and only if $\mathrm{VaR}_\alpha^+(X) \geq t$.

Proof. To show (i), denote by $A_\alpha = \{t \in \mathbb{R} : \mathbb{P}(X \leq t) \geq \alpha\}$. Note that A_α is closed in $\mathbb{R}$ since $t \mapsto \mathbb{P}(X \leq t)$ is upper semicontinuous. This gives $\mathrm{VaR}_\alpha^-(X) = \min A_\alpha$. Hence, $\alpha > \mathbb{P}(X \leq t) \Longleftrightarrow t \notin A_\alpha \Longleftrightarrow \mathrm{VaR}_\alpha^-(X) > t$. To show (ii), denote by $B_\alpha = \{t \in \mathbb{R} : \mathbb{P}(X < t) \leq \alpha\}$ which is closed in $\mathbb{R}$ since $t \mapsto \mathbb{P}(X < t)$ is lower semicontinuous. This gives $\mathrm{VaR}_\alpha^+(X) = \max B_\alpha$. It follows that $\alpha < \mathbb{P}(X < t) \Longleftrightarrow t \notin B_\alpha \Longleftrightarrow \mathrm{VaR}_\alpha^+(X) < t$. $\square$

Proof of Theorem 2. Let $g : [0,1] \to \mathbb{R}$, $\alpha \mapsto (1-\alpha)(\mathrm{ES}_\alpha(X) - t)$. Note that for any $\alpha, \alpha' \in [0,1]$,

$$g(\alpha) - g(\alpha') = \int_\alpha^{\alpha'} \left(\mathrm{VaR}_\beta^-(X) - t\right) \mathrm{d}\beta \tag{4}$$

$$= \int_\alpha^{\alpha'} \left(\mathrm{VaR}_\beta^+(X) - t\right) \mathrm{d}\beta. \tag{5}$$

Let $[c,d] = [\mathbb{P}(X < t), \mathbb{P}(X \leq t)]$. For $\alpha \leq d$, Lemma 1 (i) and (4) imply $g(\alpha) \leq g(d)$. For $\alpha < c$, Lemma 1 (ii) and (5) imply $g(\alpha) < g(c)$. For $\alpha \geq c$, Lemma 1 (ii) and (5) imply $g(\alpha) \leq g(c)$. For $\alpha > d$, Lemma 1 (i) and (4) imply $g(\alpha) < g(d)$. Summarizing the above inequalities, we obtain

$$g(\alpha_1) < g(c) = g(\alpha_2) = g(d) > g(\alpha_3) \quad \text{for any } \alpha_1 < c < \alpha_2 < d < \alpha_3.$$

Therefore, the set of maximizers for (3) is $[c,d]$. By using Lemma 1 (i) again,

$$g(d) = \int_{\mathbb{P}(X \leq t)}^1 \left(\mathrm{VaR}_\beta^-(X) - t\right) \mathrm{d}\beta = \int_0^1 \left(\mathrm{VaR}_\beta^-(X) - t\right)_+ \mathrm{d}\beta = \mathbb{E}[(X-t)_+],$$

thus showing (3). $\square$

From the reverse ES optimization formula, instead of directly calculating $\mathbb{E}[(X-t)_+]$ for fixed $t \in \mathbb{R}$, we can maximize a quantile-based function $\alpha \mapsto (1-\alpha)(\mathrm{ES}_\alpha(X) - t)$ over $\alpha \in [0,1]$. Some implications of this result are discussed in Section 3.

The next corollary on a formula for $\mathbb{E}[X \wedge x]$ can be obtained from Theorem 2. To state this result, we define the left-ES for $\alpha \in [0,1]$ as

$$\mathrm{ES}_\alpha^-(X) = \frac{1}{\alpha} \int_0^\alpha \mathrm{VaR}_\beta^-(X) \, \mathrm{d}\beta, \quad X \in \mathcal{X}, \ \alpha \in (0,1], \tag{6}$$

and $\mathrm{ES}_0^-(X) = \mathrm{VaR}_0^+(X)$.

Corollary 1. *For $t \in \mathbb{R}$ and $X \in \mathcal{X}$, it holds*

$$\mathbb{E}[X \wedge t] = \min_{\alpha \in [0,1]} \left\{\alpha \mathrm{ES}_\alpha^-(X) + (1-\alpha)t\right\}, \tag{7}$$

4

and the set of minimizers for (7) *is* $[\mathbb{P}(X < t), \mathbb{P}(X \leq t)]$.

Proof. The formula (3) directly leads to

$$
\begin{aligned}
\mathbb{E}[X \wedge t] &= \mathbb{E}[X] - \max_{\alpha \in [0,1]} \{(1-\alpha)(\mathrm{ES}_\alpha(X) - t)\} \\
&= \mathbb{E}[X] + \min_{\alpha \in [0,1]} \{(1-\alpha)(t - \mathrm{ES}_\alpha(X))\} \\
&= \min_{\alpha \in [0,1]} \{(1-\alpha)t + \mathbb{E}[X] - (1-\alpha)\mathrm{ES}_\alpha(X)\} \\
&= \min_{\alpha \in [0,1]} \{(1-\alpha)t + \alpha\mathrm{ES}_\alpha^-(X)\}.
\end{aligned}
$$

The corresponding statement on optimizers is the same as that in Theorem 2. □

Remark 2. A similar relation to Corollary 1 is found in Pflug and Römisch (2007, p.15) which is formulated for the integrated distribution function and the integrated quantile function.

3 Symmetries between Theorems 1 and 2

The function $t \mapsto \mathbb{E}[(X - t)_+]$ is called the *mean excess function* of X according to McNeil et al. (2015), and the function $\alpha \mapsto \mathrm{ES}_\alpha(X)$ will be called the *ES profile* of X according to Burzoni et al. (2022). The ES profile also relates to the Lorenz curve (see e.g., Gastwirth (1971)), which can be written as $\alpha \mapsto \alpha\mathrm{ES}_\alpha^-(X)/\mathrm{ES}_0(X)$ for a non-negative random variable X representing the wealth distribution of a population. For clarity, we distinguish between the terms "mean excess function" (as a function of t) and "mean excess loss" (as a function of X), and analogously between the terms "ES profile" and "ES".

To appreciate Theorem 2 and contrast it with Theorem 1, we need to understand the roles of the mean excess function and the ES profile. The reason why Theorem 2 has not been explored in the literature is perhaps due to the perception that the ES profile is harder to obtain or to optimize than the mean excess function. Based on this reasoning, it seems that using the mean excess function to compute ES is more natural than using the ES profile to compute the mean excess function. However, in theory, there is no such asymmetry: For a given random variable X, its mean excess function and its ES profile have perfectly symmetric roles, as we discuss below. Indeed, we shall see in Section 4 that in relevant applications, useful formulas for the mean excess function can be obtained from the ES profile via Theorems 2.

1. **Functional properties on $\mathcal{X}$.** Both the mean excess loss and ES have nice properties, symmetric to each other, as mappings on $\mathcal{X}$.

(a) For a fixed $t \in \mathbb{R}$, the mapping $X \mapsto \mathbb{E}[(X - t)_+]$ is linear in the distribution of X and convex in the quantile of X. Indeed, this mapping satisfies the independence axiom of von Neumann and Morgenstern (1947).

(b) For a fixed $\alpha \in (0, 1)$, the mapping $X \mapsto \mathrm{ES}_\alpha(X)$ is linear in the quantile of X and concave in the distribution of X (e.g., Wang et al. (2020)). Indeed, this mapping satisfies the dual independence axiom of Yaari (1987).

2. **Optimization problems.** As for the optimization problems (2) and (3), we have the following symmetry.

   (a) In the minimization (2) over $t \in \mathbb{R}$, the function $t \mapsto t + \frac{1}{1-\alpha}\mathbb{E}[(X - t)_+]$ is convex in t.

   (b) In the maximization (3) over $\alpha \in [0, 1]$, the function $\alpha \mapsto (1 - \alpha)(\mathrm{ES}_\alpha(X) - t)$ is concave in α.

3. **Solutions to the optimization problems.** The optimizers to the optimization problems (2) and (3) also admit nice symmetry, as one is the quantile interval, and the other one is the probability interval.

4. **Parametric forms.** For commonly used distributions in risk management and actuarial science, if one of the mean excess loss and ES has an explicit formula, then so is the other one (e.g., Pareto distributions; see Example 1 below). Moreover, each of the two curves determines the whole distribution of the random variable.

To summarize, writing one as a minimum or maximum of the other as in Theorems 1 and 2 leads to the following implications for optimization:

(a) Theorem 1 allows one to transform the non-linear (in distribution) objective $\mathrm{ES}_\alpha(X)$ as the minimum over t of linear (in distribution) functions convex in t.

(b) Theorem 2 allows one to transform the non-linear (in quantile) objective $\mathbb{E}[(X - t)_+]$ as the maximum over α of linear (in quantile) functions concave in α.

Due to the above discussed symmetries, Theorem 2 serves as a natural dual formula to Theorem 1. Indeed, Theorems 1 and 2 are closely related to Fenchel-Legendre transformations, which we will discuss in Section 7.

4 Worst-case risk under model uncertainty

As discussed in Section 3, one of the greatest advantages of the ES optimization formula in Theorem 1 is that it allows us to translate optimization problems of ES to those of the mean excess function. More precisely, for a set of actions A and a loss function $f : A \times \mathbb{R}^d \to \mathbb{R}$, Theorem 1 implies

$$\min_{y \in A} \mathrm{ES}_\alpha(f(y, \mathbf{X})) = \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{1-\alpha} \min_{y \in A} \mathbb{E}[(f(y, \mathbf{X}) - t)_+] \right\},$$

and thus, for the minimization of ES, it suffices to minimize the mean excess loss $\mathbb{E}[(f(y, \mathbf{X}) - t)_+]$ for each $t \in \mathbb{R}$, which is more convenient in many specific settings; see the review in Rockafellar and Uryasev (2013).

In contrast, Theorem 2 has a maximum operator in its formula (3), and it is useful in maximization problems. Moreover, even though risk often needs to be minimized, a maximum naturally appears in the presence of model uncertainty, which is often addressed via a worst-case approach. The worst-case risk evaluation under uncertainty appears in, e.g., Gilboa and Schmeidler (1989) and Maccheroni et al. (2006) in the context of decision making, Ghaoui et al. (2003) and Zhu and Fukushima (2009) in the context of optimization, and Embrechts et al. (2013) in the context of risk aggregation. More precisely, suppose that there is uncertainty about a random vector $\mathbf{X}$, assumed to be in a set $\mathcal{U}$, and $f : \mathbb{R}^d \to \mathbb{R}$ is a loss function. Theorem 2 implies that the worst-case mean excess loss can be computed by (recall that the convention is $0 \times \infty = 0$), via exchanging the order of two suprema,

$$\sup_{\mathbf{X} \in \mathcal{U}} \mathbb{E}[(f(\mathbf{X}) - t)_+] = \sup_{\alpha \in [0,1]} \left\{ (1-\alpha) \left(\sup_{\mathbf{X} \in \mathcal{U}} \mathrm{ES}_\alpha(f(\mathbf{X})) - t \right) \right\}, \tag{8}$$

which allows us to use rich existing results on worst-case ES. Moreover, the maximization over $\alpha \in [0,1]$ is attainable under a condition of uniform integrability, as we show below.

Proposition 1. *Let $\mathcal{Y}$ be a set of random variables and $t \in \mathbb{R}$. If $\mathcal{Y}$ is uniformly integrable, then*

$$\sup_{Y \in \mathcal{Y}} \mathbb{E}[(Y - t)_+] = \max_{\alpha \in [0,1]} \left\{ (1-\alpha) \left(\sup_{Y \in \mathcal{Y}} \mathrm{ES}_\alpha(Y) - t \right) \right\}. \tag{9}$$

In particular, if there exists $p > 1$ such that $\sup_{Y \in \mathcal{Y}} \mathbb{E}[|Y|^p] < \infty$, then $\mathcal{Y}$ is uniformly integrable and (9) holds.

Proof. We first show that uniform integrability of $\mathcal{Y}$ implies, for any $\alpha \in [0,1]$,

$$\lim_{\alpha' \to \alpha} \sup_{Y \in \mathcal{Y}} \int_{\alpha}^{\alpha'} |\mathrm{VaR}_{\beta}(Y)| \, \mathrm{d}\beta \to 0, \tag{10}$$

where the limit is one-sided if $\alpha = 0$ or $\alpha = 1$. Suppose that (10) does not hold, and without loss of generality we consider $\alpha' \downarrow \alpha$ (in this case, $\alpha < 1$). It follows that there exists $m > 0$ such that, for any $\epsilon \in (0, 1 - \alpha)$, there exists $Y_\epsilon \in \mathcal{Y}$ satisfying $\int_{\alpha}^{\alpha+\epsilon} |\mathrm{VaR}_{\beta}(Y_\epsilon)| \, \mathrm{d}\beta > m$. Since VaR_{α} is monotone in α, we have

$$m < \int_{\alpha}^{\alpha+\epsilon} |\mathrm{VaR}_{\beta}(Y_\epsilon)| \, \mathrm{d}\beta \leq \int_{0}^{\epsilon} |\mathrm{VaR}_{\beta}(Y_\epsilon)| \, \mathrm{d}\beta + \int_{1-\epsilon}^{1} |\mathrm{VaR}_{\beta}(Y_\epsilon)| \, \mathrm{d}\beta.$$

For any $K > 0$, let $\epsilon > 0$ be such that $4K\epsilon < m$. It follows that

$$\begin{aligned}
\mathbb{E}\left[|Y_\epsilon| \mathbb{1}_{\{|Y_\epsilon| > K\}}\right] &\geq \mathbb{E}[(|Y_\epsilon| - K)_+] \\
&= \int_{0}^{1} (|\mathrm{VaR}_{\beta}(Y_\epsilon)| - K)_+ \, \mathrm{d}\beta \\
&\geq \int_{0}^{\epsilon} (|\mathrm{VaR}_{\beta}(Y_\epsilon)| - K)_+ \, \mathrm{d}\beta + \int_{1-\epsilon}^{1} (|\mathrm{VaR}_{\beta}(Y_\epsilon)| - K)_+ \, \mathrm{d}\beta \\
&\geq \int_{0}^{\epsilon} |\mathrm{VaR}_{\beta}(Y_\epsilon)| \, \mathrm{d}\beta + \int_{1-\epsilon}^{1} |\mathrm{VaR}_{\beta}(Y_\epsilon)| \, \mathrm{d}\beta - 2\epsilon K > m/2.
\end{aligned}$$

Hence, $\sup_{Y \in \mathcal{Y}} \mathbb{E}[|Y| \mathbb{1}_{\{|Y| > K\}}] > m/2$, contradicting uniform integrability. Therefore, (10) holds.

Let $g : [0,1] \to \mathbb{R}$, $\alpha \mapsto \sup_{Y \in \mathcal{Y}} (1-\alpha)(\mathrm{ES}_{\alpha}(Y) - t)$. Note that for any $\alpha, \alpha' \in [0,1]$, using (10),

$$\begin{aligned}
|g(\alpha) - g(\alpha')| &= \left| \sup_{Y \in \mathcal{Y}} \int_{0}^{\alpha'} \left(\mathrm{VaR}_{\beta}^{-}(Y) - t \right) \mathrm{d}\beta - \sup_{Y \in \mathcal{Y}} \int_{0}^{\alpha} \left(\mathrm{VaR}_{\beta}^{-}(Y) - t \right) \mathrm{d}\beta \right| \\
&\leq \left| \sup_{Y \in \mathcal{Y}} \int_{\alpha}^{\alpha'} \left(\mathrm{VaR}_{\beta}^{-}(Y) - t \right) \mathrm{d}\beta \right| \leq \sup_{Y \in \mathcal{Y}} \int_{\alpha}^{\alpha'} |\mathrm{VaR}_{\beta}^{-}(Y)| \, \mathrm{d}\beta + |(\alpha' - \alpha)t|,
\end{aligned}$$

which converges to 0 as $\alpha' \to \alpha$. This shows that g is continuous on $[0,1]$, and hence the maximum in (9) is attained. The last statement that boundedness of $\mathbb{E}[|Y|^p]$ implies uniformly integrability can be found in Exercise 5.5.1 of Durrett (2010). □

In two settings of uncertainty based on moment information and the Wasserstein metric which we study below, explicit formulas for the worst-case ES are available in Pesenti et al. (2020) and Liu et al. (2022), whereas the worst-case mean excess loss does not have an explicit formula. In the popular case that f is a portfolio loss function (i.e., $f(\mathbf{x}) = \mathbf{w}^{\top}\mathbf{x}$ for some $\mathbf{w} \in \mathbb{R}^d$), the multi-dimensional uncertainty sets reduce to one-dimensional sets of the same type; see Mao et al. (2022,

Section 6). For this reason, we will focus on the one-dimensional uncertainty sets.

4.1   Uncertainty sets induced by moment information

We first study the uncertainty set induced by mean and a higher moment. For $p > 1$, $m \in \mathbb{R}$ and $v \geq 0$, denote by

$$\mathcal{L}^p(m, v) = \{X \in \mathcal{X} : \mathbb{E}[X] = m, \ \mathbb{E}[|X - m|^p] \leq v^p\},$$

that is, the set of all random variables with given mean m and a p-th centralized absolute moment at most v^p. We are interested in the worst-case value of a functional over $\mathcal{L}^p(m, v)$. The special case of this problem when $p = 2$, i.e., the setting with mean and variance information, has been the most popular; see e.g., Ghaoui et al. (2003), Li (2018) and Liu et al. (2020) on various risk measures.

Let $\rho : \mathcal{L}^p \to \mathbb{R}$ be a mapping where $\mathcal{L}^p$ is the set of all random variables with finite p-th moment. Note that the problem of $\sup_{X \in \mathcal{L}^p(m,v)} \rho(X)$ is better suited for $\rho = \mathrm{ES}_\alpha$ or some other risk measures than for the mean excess loss $\rho : X \mapsto \mathbb{E}[(X - t)_+]$, because many risk measures, including VaR_α and ES_α, satisfy some simple properties which yield

$$\sup_{X \in \mathcal{L}^p(m,v)} \rho(X) = m + v \sup_{X \in \mathcal{L}^p(0,1)} \rho(X).$$

Therefore, we can convert the original problem to an optimization over $\mathcal{L}^p(0, 1)$. Such a relationship does not hold for the mean excess loss $\rho : X \mapsto \mathbb{E}[(X - t)_+]$.

The problem of the worst-case mean excess loss with moment conditions has a long history; see e.g., De Vylder and Goovaerts (1982) in actuarial science and Jagannathan (1977) in operations research. Pesenti et al. (2020, Corollary 1) obtained a closed-form formula for the worst-case ES_α over $\mathcal{L}^p(m, v)$, that is,

$$\sup_{X \in \mathcal{L}^p(m,v)} \mathrm{ES}_\alpha(X) = m + v\alpha(\alpha^p(1 - \alpha) + (1 - \alpha)^p\alpha)^{-1/p}. \tag{11}$$

In particular, in case $p = 2$, it becomes $\sup_{X \in \mathcal{L}^2(m,v)} \mathrm{ES}_\alpha(X) = m + v(\frac{\alpha}{1-\alpha})^{1/2}$. By exchanging the order of two suprema, the problem of worst-case mean excess loss can be obtained by combining (11) and Theorem 2. In what follows, we use the convention that $1/0 = \infty$ and $1/\infty = 0$.

Proposition 2. *For $p > 1$, $m, t \in \mathbb{R}$ and $v \geq 0$, we have*

$$\sup_{X \in \mathcal{L}^p(m,v)} \mathbb{E}[(X - t)_+] = \max_{\alpha \in [0,1]} \left\{ (1 - \alpha)(m - t) + v \left((1 - \alpha)^{1-p} + \alpha^{1-p} \right)^{-1/p} \right\}. \tag{12}$$

Proof. The proposition follows directly from putting together (9) and (11). The maximum is attainable because the set $\mathcal{L}^p(m, v)$ satisfies the condition in Proposition 1. □

In the most popular case $p = 2$, Proposition 2 gives

$$\sup_{X \in \mathcal{L}^2(m,v)} \mathbb{E}[(X - t)_+] = \max_{\alpha \in [0,1]} \left\{ (1 - \alpha)(m - t) + v\sqrt{\alpha(1 - \alpha)} \right\} = \frac{1}{2} \left(m - t + \sqrt{v^2 + (m - t)^2} \right),$$

which coincides with Jagannathan (1977, Corollary 1.1). The maximum value in (12) for $p \neq 2$ can be computed numerically. We provide a numerical example below by simply taking $m = 0$ and $v = 1$. Figure 1 shows value of worst-case mean excess loss with respect to different thresholds t under different moment conditions. We observe that a higher p leads to a lower value of the worst-case mean excess loss at any threshold level, and this is because the constraint $\mathbb{E}[|X - m|^p] \leq v^p$ is more stringent with larger p. With Proposition 2, we can easily identify the worst-case values given a fixed threshold without knowing the exact distribution of loss.
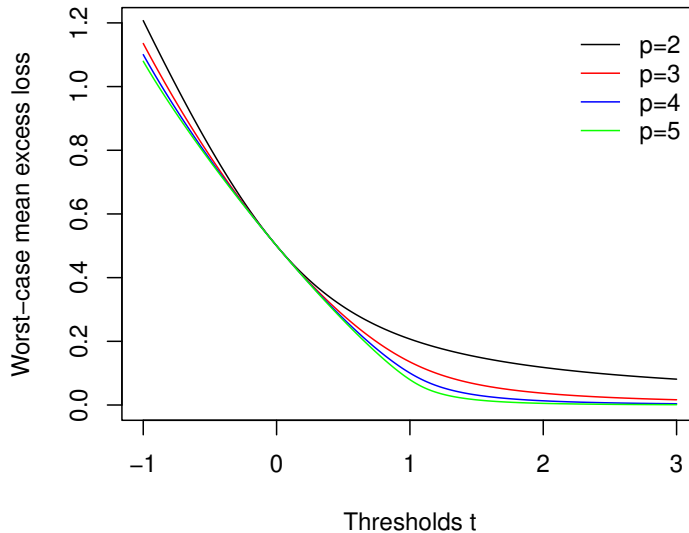

Figure 1: Worst-case mean excess loss with moment conditions in $\mathcal{L}^p(0, 1)$

10

4.2 Uncertainty sets induced by Wasserstein metrics

Next, we consider uncertainty sets induced by Wasserstein metrics (this setting of uncertainty will be called the Wasserstein uncertainty). Recall that the Wasserstein metric of order $p \geq 1$ between two distributions F and G on $\mathbb{R}$ is defined by

$$W_p(F, G) = \inf_{X \sim F, \, Y \sim G} (\mathbb{E}[|X - Y|^p])^{1/p} = \left(\int_0^1 |F^{-1}(x) - G^{-1}(x)|^p \, \mathrm{d}x \right)^{1/p},$$

where $X \sim F$ means that the distribution of X is F. For a benchmark loss $X \in \mathcal{X} = \mathcal{L}^1$ and an uncertainty level $\delta \geq 0$, the *Wasserstein ball* around X is $\{Y : W_p(F_X, F_Y) \leq \delta\}$, where F_X and F_Y are the distributions of X and Y, respectively. Note that $\delta = 0$ corresponds to the case of no model uncertainty. The worst-case value of a risk measure $\rho : \mathcal{X} \to \mathbb{R}$ under the above uncertainty setting around X is

$$\sup \{\rho(Y) : W_p(F_X, F_Y) \leq \delta\}.$$

The worst-case ES_α under Wasserstein uncertainty is obtained by Liu et al. (2022, Proposition 4) with the closed-form formula

$$\sup \{\mathrm{ES}_\alpha(Y) : W_p(F_X, F_Y) \leq \delta\} = \mathrm{ES}_\alpha(X) + \frac{\delta}{(1 - \alpha)^{1/p}}. \tag{13}$$

Based on (13) and Theorem 2, we can calculate the worst-case value of $\rho(X) = \mathbb{E}[(X - t)_+]$ for $t \in \mathbb{R}$, similarly to Proposition 2.

Proposition 3. *For $t \in \mathbb{R}$, $p \geq 1$, $\delta \geq 0$ and $X \in \mathcal{X}$, we have*

$$\sup \{\mathbb{E}[(Y - t)_+] : W_p(F_X, F_Y) \leq \delta\} = \max_{\alpha \in [0,1]} \left\{ (1 - \alpha)(\mathrm{ES}_\alpha(X) - t) + \delta(1 - \alpha)^{1 - 1/p} \right\}. \tag{14}$$

Proof. The proposition follows directly from putting together (8) and (13). The maximum is attainable because the function $\alpha \mapsto (1 - \alpha)(\mathrm{ES}_\alpha(X) - t) + \delta(1 - \alpha)^{1 - 1/p}$ is continuous. $\square$

Comparing (14) with (3) in Theorem 2, to compute the function $\mathbb{E}[(Y - t)_+]$ based on ES, there is an extra term of $\delta(1 - \alpha)^{1 - 1/p}$ in the maximization over $\alpha \in [0, 1]$ to compensate for model uncertainty. As far as we are aware of, both formulas (12) and (14) in this section are new.

Example 1. Let the benchmark loss X follow a Pareto distribution with tail parameter $\theta > 1$, that is, $\mathbb{P}(X > x) = x^{-\theta}$ for $x \geq 1$. For simplicity we take $\theta = 2$ and consider the Wasserstein metric

W_2. By straightforward calculation, $\text{ES}_\alpha(X) = 2(1-\alpha)^{-1/2}$ for $\alpha \in [0,1)$. Using (14), we get

$$\sup\left\{\mathbb{E}[(Y-t)_+] : W_2(F_X, F_Y) \leq \delta\right\} = \max_{\alpha \in [0,1]}\left\{(2+\delta)(1-\alpha)^{1/2} - (1-\alpha)t\right\}$$
$$= \frac{(1+\delta/2)^2}{t}\mathbb{1}_{\{t>1+\delta/2\}} + (2+\delta-t)\mathbb{1}_{\{t \leq 1+\delta/2\}}. \qquad (15)$$

Example 1 also illustrates how the level of model uncertainty affects the evaluation of the worst-case mean excess loss. Note that for the benchmark loss X,

$$\mathbb{E}[(X-t)_+] = \int_t^\infty \mathbb{P}(X>x)\,\mathrm{d}x = \int_{t\vee 1}^\infty x^{-2}\,\mathrm{d}x + (1-t)_+ = \frac{1}{t}\mathbb{1}_{\{t>1\}} + (2-t)\mathbb{1}_{\{t \leq 1\}}. \qquad (16)$$

If $\delta = 0$, then there is no model uncertainty, and (15) and (16) coincide. If $\delta > 0$, then for $t > 1+\delta/2$, the worst-case value (15) of the mean excess loss increases from the benchmark value (16) by a factor of $(1+\delta/2)^2 > 1$; for $t \leq 1$, the worst-case value (15) increases from the benchmark value (16) by a constant $\delta > 0$. We observe from (13) that for ES_α with a fixed $\alpha \in [0,1)$, the level of model uncertainty δ always affects the worst-case risk evaluation linearly; this also holds for any coherent distortion risk measures as shown by Liu et al. (2022). In contrast, for the mean excess loss, the effect of δ is no longer linear in the interesting domain where t is large.

Example 2. Let the benchmark loss X follow a normal distribution with mean μ and standard deviation σ. We can calculate $\text{ES}_\alpha(X) = \mu + \sigma\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}$ for $\alpha \in [0,1)$, where where ϕ and Φ^{-1} represent the density function and quantile function of the standard normal distribution, respectively. Using (14), we get

$$\sup\left\{\mathbb{E}[(Y-t)_+] : W_p(F_X, F_Y) \leq \delta\right\} = \max_{\alpha \in [0,1]}\left\{\sigma\phi(\Phi^{-1}(\alpha)) + (\mu-t)(1-\alpha) + \delta(1-\alpha)^{1-1/p}\right\}.$$

Although the above expression is not explicit, it can be easily computed numerically.

For a better understanding of Proposition 3, we provide a numerical example. In Figure 2 (a), by choosing the parameter $p = 2$ and the uncertainty level $\delta = 0.1$, we show how the worst-case values of the mean excess loss vary with the threshold t under different distributions, including Pareto, exponential, normal and t distributions. The obtained curves are similar to those in Figure 1. In Figure 2 (b), by taking $p = 2$ and $t = 2$, we report the worst-case values of the mean excess loss increases with the uncertainty level δ. As we can see, the effect of δ on the worst-case value of mean excess loss is non-linear, as we discussed in Example 1 for a Pareto distribution.
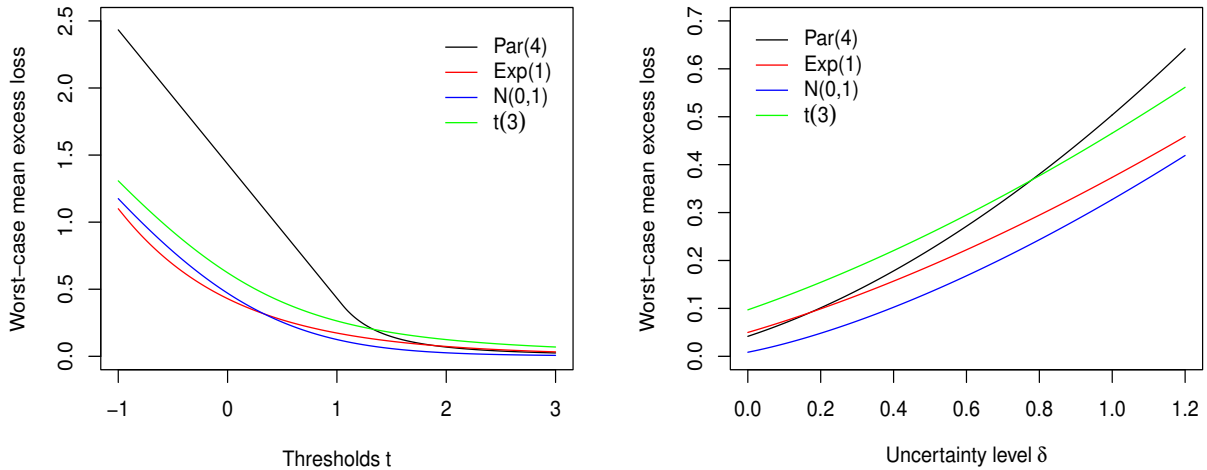
(a) Worst-case value changes with t (fixed $\delta = 0.1$)   (b) Worst-case value changes with δ (fixed $t = 2$)

Figure 2: Worst-case mean excess loss with Wasserstein uncertainty

5 Empirical analysis for insurance data

In this section, we use insurance data to calculate the worst-case mean excess loss under uncertainty governed by the Wasserstein metric with $p = 2$. In addition, we check how the uncertainty level δ and the threshold level t may influence the value of worst-case mean excess loss compared to the one without uncertainty, and see their different performances in different datasets.

We choose two univariate datasets from the R package CASdatasets: normalized hurricane damages (ushurricane, 1900-2005) and normalized French commercial fire losses (frecomfire, 1982-1996) pooled by each month. Both datasets have around 180 observations and the distributions are highly right-skewed. We shall use R to fit the data with lognormal, Gamma and Weibull distributions as our benchmark distributions.

In the first part of the empirical analysis, we fix the threshold level t and let the uncertainty level δ vary to visualize how the worst-case mean excess loss varies. Since the two datasets are quite different, it is important to calibrate t and δ to make two datasets to be comparable. In particular, δ should be chosen in a statistically relevant range; see e.g., Blanchet et al. (2022) for a discussion on this point. Generally, if the uncertainty level δ is too large, then the data become less relevant; if δ is too small, we are not protected against model uncertainty, thus losing the desired robustness. For a meaningful comparison, we make the following heuristic choices.

	Lognormal	Weibull	Gamma
Hurricane loss	43.83	37.99	47.45
Fire loss	186.69	248.99	244.24

Table 1: Values of δ_0 for the lognormal, Weibull and Gamma distributions and for the hurricane loss and fire loss datasets. The level δ_0 is the Wasserstein metric with $p = 2$ between the empirical and the fitted distributions

For each benchmark distribution, we let δ vary in $[\delta_0, 2\delta_0]$, where δ_0 is the Wasserstein distance (metric) between the fitted distribution and the empirical distribution. This choice ensures that the empirical distribution is inside the Wasserstein ball around the fitted distribution; intuitively, a poorly fitted distribution is associated with a larger δ_0, thus requiring a higher uncertainty level to be considered as robust. Moreover, δ_0 is of the order $n^{-1/2}$ if the estimation is $n^{-1/2}$-efficient, where n is the sample size. Table 1 shows the values of δ_0.

We are interested in the ratio $r(t, \delta)$ of the worst-case mean excess loss to that of the benchmark distribution, defined by

$$r(\delta, t) = \frac{\sup\{\mathbb{E}[(Y - t)_+] : W_2(F_X, F_Y) \leq \delta\}}{\mathbb{E}[(X - t)_+]},$$

where X follows one of the benchmark distributions (3 choices for each dataset). We first fix the threshold level t as the first quartile (25% quantile) t_0 of its corresponding benchmark distribution and let δ vary (Figures 3 and 4), and then we fix $\delta = \delta_0$ and let t vary (Figures 5 and 6).

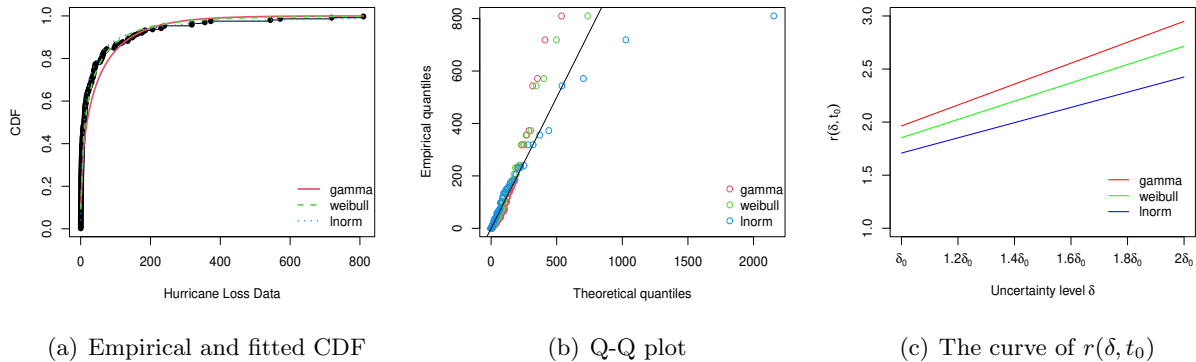(a) Empirical and fitted CDF   (b) Q-Q plot   (c) The curve of $r(\delta, t_0)$

Figure 3: Empirical results on the hurricane loss data

Figure 3 (a) and (b) present goodness-of-fit plots for the fitted lognormal, Weibull and Gamma distributions to the hurricane data. We observe that the lognormal and Weibull distributions fit
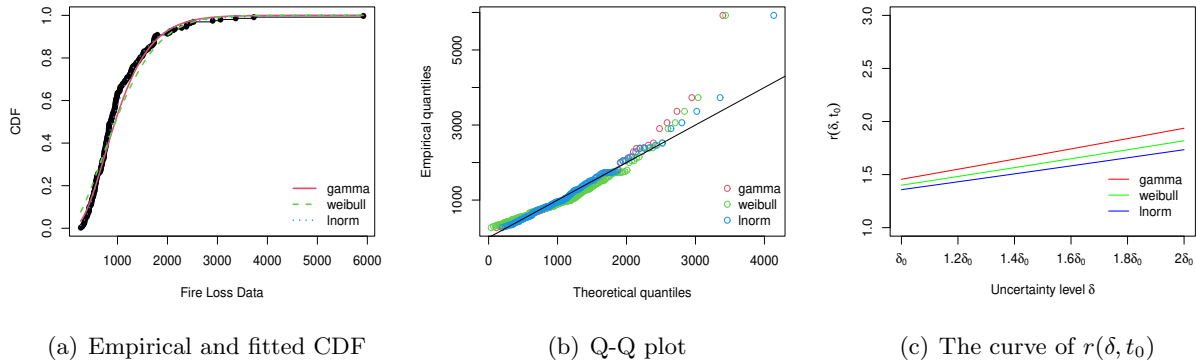
| | | (a) Empirical and fitted CDF | | (b) Q-Q plot | | (c) The curve of $r(\delta, t_0)$ |

Figure 4: Empirical results on the fire loss data

		δ_0	$1.2\delta_0$	$1.4\delta_0$	$1.6\delta_0$	$1.8\delta_0$	$2\delta_0$
	Lognormal	1.708	1.839	1.985	2.132	2.279	2.425
Hurricane loss	Weibull	1.853	2.012	2.193	2.352	2.534	2.715
	Gamma	1.964	2.149	2.334	2.539	2.724	2.950
	Lognormal	1.358	1.431	1.505	1.582	1.657	1.735
Fire loss	Weibull	1.400	1.481	1.564	1.649	1.733	1.819
	Gamma	1.456	1.548	1.644	1.740	1.837	1.937

Table 2: Values of $r(\delta, t_0)$ for the hurricane loss and the fire loss datasets. The threshold level t_0 is the first quartile of the benchmark distribution and the parameter δ_0 is the Wasserstein metric between the empirical and the fitted distributions

better to this dataset than the Gamma distribution. In Figure 3 (c), the Gamma model is penalized more for model uncertainty. We note that the curves $\delta \mapsto r(\delta, t_0)$ are almost linear in δ. The numerical values of $r(\delta, t_0)$ in Table 2 show that $r(\delta, t_0)$ is actually convex in δ, implying that the worst-case mean excess loss becomes more sensitive to δ for large values of δ, consistent with the numerical analysis in Section 4. Figure 4 based on the fire loss data exhibits similar patterns to the hurricane loss data. The lognormal distribution fits better to the fire loss than other two distributions so that the mean excess loss will be less affected by the Wasserstein uncertainty.

Comparing the curves $r(\delta, t_0)$ for two datasets, we can notice that the values of $r(\delta, t_0)$ for the hurricane data are much higher than the ones for the fire data, which means the hurricane loss is more severely affected by model uncertainty than fire loss. It may be explained by the fact that the hurricane losses are more catastrophic and right-skewed than fire losses so that more penalties should be added to hurricane case if model uncertainty is a concern.

In an insurance pricing context, the mean excess loss can be used to price the stop-loss pre-

mium where the threshold t can be seen as a deductible, and our method can be used to analyze the sensitivity of the stop-loss premium to the Wasserstein uncertainty. Taking the lognormal distribution as an example in Table 2, if we use $\delta = \delta_0$ and $t = t_0$ for pricing a hurricane insurance, the stop-loss premium will increase 70.8% compared to the one without considering model uncertainty. For the same choice $\delta = \delta_0$ and $t = t_0$ (although both t_0 and δ_0 depend on the dataset), the stop-loss premium will only increase 35.8% when pricing the fire insurance. It intuitively means that the hurricane insurance pricing is more sensitive to the Wasserstein uncertainty than the fire insurance pricing.

Next, we investigate how different threshold levels t may influence the mean excess loss with and without the Wasserstein uncertainty. The uncertainty level δ is fixed as δ_0 in this experiment and we look at $r(\delta_0, t)$ as t varies. Figures 5 and 6 report the ratio $r(\delta_0, t)$ in these settings, as well as the ratio

$$\hat{r}(\delta_0, t) = \frac{\sup\{\mathbb{E}[(Y - t)_+] : W_2(F_X, F_Y) \leq \delta_0\}}{\mathbb{E}[(\hat{X} - t)_+]},$$

where $\hat{X}$ follows the empirical distribution of the data. Note that $\hat{r}(\delta_0, t) \geq 1$ since δ_0 is chosen such that the empirical distribution is inside the Wasserstein ball. For both datasets, we let the threshold level t vary between the first quartile and the third quartile of the loss data. We observe from Figure 5 that the ratio $r(\delta_0, t)$ for the hurricane loss data is relatively stable with respect to the threshold level t, whereas Figure 6 shows that the ratio $r(\delta_0, t)$ for the fire loss data increases with the threshold level t in all selected benchmark distributions. Hence, compared to hurricane loss, the mean excess loss of fire loss data is more sensitive to model uncertainty with larger threshold levels. This observation is less pronounced for the better fitted lognormal distribution in Figure 6 (a). The other ratio $\hat{r}(\delta_0, t)$ is quite stable for the fire loss data and it shows a decreasing trend in t for the hurricane loss data.

6 Optimized certainty equivalents

We proceed to offer some more theoretical results and discussions on the reverse ES optimization formula. It would be interesting to see whether Theorem 2 can be generalized to other risk measures than the class ES_α. Note that ES_α belongs to the class of optimized certainty equivalents (OCE) of Ben-Tal and Teboulle (2007). The class of OCE includes ES and the entropic risk measures (Föllmer and Schied (2016)) as special cases. In this section, we work with the set $\mathcal{X}_B$ of essentially bounded random variables to avoid integrability issues. Let V be the set of increasing and convex functions $v : \mathbb{R} \to \mathbb{R}$ satisfying $v(0) = 0$, $\bar{v} \geq 1$ and $\lim_{t \to \infty} v'_+(-t) = 0$ where
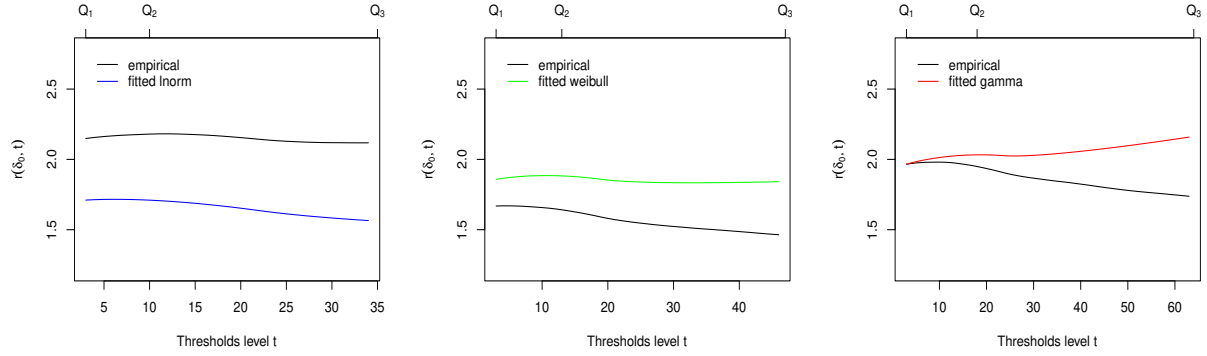
(a) Fitted to a lognormal distribution (b) Fitted to a Weibull distribution (c) Fitted to a Gamma distribution

Figure 5: Values of the ratios $r(\delta_0, t)$ and $\hat{r}(\delta_0, t)$ for the hurricane loss data, where Q_1, Q_2 and Q_3 represent the 1st, 2nd and 3rd quartiles of the data
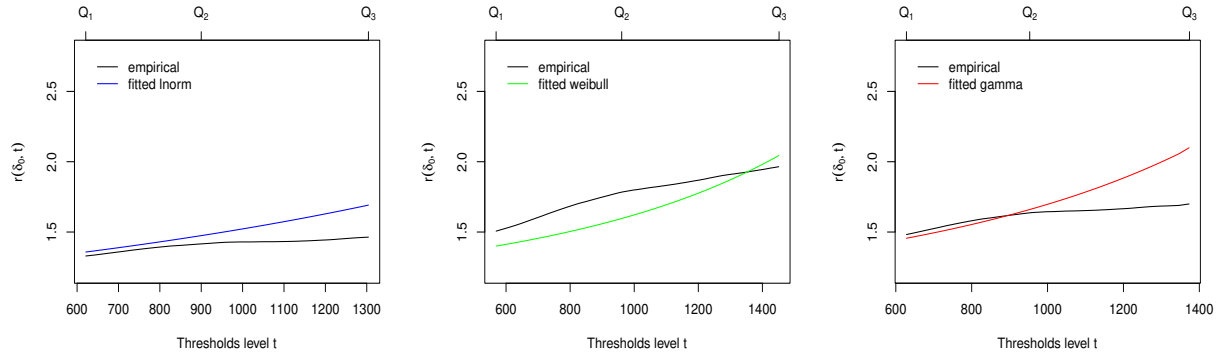
(a) Fitted to a lognormal distribution (b) Fitted to a Weibull distribution (c) Fitted to a Gamma distribution

Figure 6: Values of the ratios $r(\delta_0, t)$ and $\hat{r}(\delta_0, t)$ for the fire loss data, where Q_1, Q_2 and Q_3 represent the 1st, 2nd and 3rd quartiles of the data

$\bar{v} = \sup_{x \in \mathbb{R}} v'_+(x)$ and v'_+ is the right derivative of v. An OCE is a risk measure R defined by

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

The finiteness of R is guaranteed if $v'_+(x) \le 1 \le v'_+(y)$ for some $x, y \in \mathbb{R}$ which is satisfied by $v \in V$ if $\bar{v} > 1$. If R is finite, then it is a convex risk measure in the sense of Föllmer and Schied (2016). In particular, if $v(x) = x_+/(1 - \alpha)$ for $\alpha \in [0, 1)$, then R is ES_α as in Theorem 1. Moreover, under a continuity condition, ES_α is the only class of coherent risk measures in the class of OCE (Theorem 3.1 of Embrechts et al. (2021)).

Inspired by Theorem 2, we define a parametric family of OCE risk measures. For $v \in V$ and $\beta \in (0, \bar{v}]$, let

$$R_\beta^v(X) = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{\beta} \mathbb{E}[v(X - t)] \right\}, \quad X \in \mathcal{X}_B.$$

Here, the convention is $1/\infty = 0$. If $v(x) = x_+$, then $\bar{v} = 1$ and $R_\beta^v = \mathrm{ES}_{1-\beta}$ for $\beta \in (0, 1]$. The next result gives a reverse optimization formula for OCE, which includes the formula (3) as a special case. This result is related to the Fenchel-Legendre transformation as we discuss in Section 7.

Theorem 3 (Reverse OCE optimization formula). *For $X \in \mathcal{X}_B$, $t \in \mathbb{R}$ and $v \in V$, it holds*

$$\mathbb{E}[v(X - t)] = \sup_{\beta \in (0, \bar{v}]} \left\{ \beta(R_\beta^v(X) - t) \right\}. \tag{17}$$

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(t) = \mathbb{E}[v(X + t)]$, which is an increasing convex function on $\mathbb{R}$. As a convex function on $\mathbb{R}$, f is automatically continuous. Its Fenchel-Legendre transform f^*, called the conjugate function of f, is given by

$$f^*(\beta) = \sup_{t \in \mathbb{R}} \left\{ t\beta - f(t) \right\}, \quad \beta \in \mathbb{R}, \tag{18}$$

which is not necessarily finite.

If $\beta < 0$, then letting $t \to -\infty$ gives $\sup_{t \in \mathbb{R}} \{ t\beta - f(t) \} = \infty$ since f is increasing. On the other hand, if $\beta > \bar{v}$, then, since $v'_+(x) \le \bar{v}$ for $x \in \mathbb{R}$, letting $t \to \infty$ gives $\sup_{t \in \mathbb{R}} \{ t\beta - f(t) \} = \infty$.

Let $s \in \mathbb{R}$ be such that $f'_+(s) > \bar{v}/2$; such s exists since $\lim_{t \to \infty} f'_+(t) = \bar{v}$. For $\beta \in (0, \bar{v}/2]$, we have

$$f^*(\beta) = \sup_{t \in (-\infty, s]} \left\{ t\beta - f(t) \right\} \le s\beta + \sup_{t \in \mathbb{R}} \{ -f(t) \} = s\beta + f^*(0).$$

Hence,

$$\limsup_{\beta \downarrow 0} f^*(\beta) \le f^*(0). \tag{19}$$

Summarizing the above observations, for a fixed $t \in \mathbb{R}$,

$$\sup_{\beta \in \mathbb{R}} \left\{ -t\beta - f^*(\beta) \right\} = \sup_{\beta \in [0, \bar{v}]} \left\{ -t\beta - f^*(\beta) \right\} = \sup_{\beta \in (0, \bar{v}]} \left\{ -t\beta - f^*(\beta) \right\}, \tag{20}$$

where the last equality is due to (19). For $\beta \in (0, \bar{v}]$,

$$\frac{-f^*(\beta)}{\beta} = \inf_{t \in \mathbb{R}} \left\{ -t + \frac{1}{\beta} f(t) \right\} = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{\beta} f(-t) \right\} = R_\beta^v(X).$$

Thus, $f^*(\beta) = -\beta R_\beta^v(X)$. The Fenchel-Legendre theorem in the form of Proposition A.9 of Föllmer

and Schied (2016) gives $f^{**} = f$. Therefore, using (20),

$$f(-t) = \sup_{\beta \in (0, \bar{v}]} \left\{ -t\beta - f^*(\beta) \right\} = \sup_{\beta \in (0, \bar{v}]} \left\{ -t\beta + \beta R_\beta^v(X) \right\} = \sup_{\beta \in (0, \bar{v}]} \left\{ \beta (R_\beta^v(X) - t) \right\},$$

and hence (17) holds. □

As we can see from Theorem 3, the symmetry between the ES optimization formula (2) and the reverse formula (3) can be seen as a consequence of the Fenchel-Legendre transform mechanism.

7   Related Fenchel-Legendre transforms

As mentioned above, the reverse ES optimization formula and reverse OCE optimization formula is closely related to the Fenchel-Legendre transformation (e.g., Definition A.8 of Föllmer and Schied (2016)). In this section, we give two pairs of conjugate functions related to Theorem 2.

The Fenchel-Legendre transformation converts convex functions to their conjugate. For a convex function $f : \mathbb{R} \to \mathbb{R}$, its Legendre-Fenchel transform is the function f^* on $\mathbb{R}$ defined by

$$f^*(\beta) = \sup_{t \in \mathbb{R}} \left\{ t\beta - f(t) \right\}, \qquad \beta \in \mathbb{R},$$

where β may be constrained to a subset of $\mathbb{R}$ such that f^* is real.

As we have seen in Theorem 3, Fenchel-Legendre transforms are closely related to our reverse ES optimization formula, as Theorems 1 and 2 can be expressed from each other via a Fenchel-Legendre transform. Below, we identify two other pairs of conjugate functions, one being quantile-based and one being expectation-based, analogously to the case of ES and the mean excess function.

Proposition 4. *Fix $X \in \mathcal{X}$.*

(i) The Fenchel-Legendre transform of the convex quantile-based function $f_1 : [0, 1] \to \mathbb{R}$,

$$f_1(\alpha) = -(1 - \alpha) \mathrm{ES}_\alpha(X),$$

is given by

$$f_1^*(t) = \max_{\alpha \in [0, 1]} \left\{ \alpha t - f_1(\alpha) \right\} = \mathbb{E}[X \vee t].$$

(ii) The Fenchel-Legendre transform of the convex quantile-based function $f_2 : [0, 1] \to \mathbb{R}$,

$$f_2(\alpha) = \alpha \mathrm{ES}_\alpha^-(X),$$

19

is given by

$$f_2^*(t) = \max_{\alpha \in [0,1]} \{\alpha t - f_2(\alpha)\} = \mathbb{E}[(t - X)_+].$$

Moreover, the set of maximizers for both maximization problems is $[\mathbb{P}(X < t), \mathbb{P}(X \le t)]$.

Proof. For the first statement, it is straightforward to identify that the quantile-based function $f_1 : \alpha \mapsto -(1 - \alpha)\mathrm{ES}_\alpha(X)$ is convex by taking a derivative with respect to α. By definition of the Legendre-Fenchel transformation, we have

$$\begin{aligned}
f_1^*(t) &= \sup_{\alpha \in [0,1]} \{\alpha t + (1 - \alpha)\mathrm{ES}_\alpha(X)\} \\
&= \sup_{\alpha \in [0,1]} \{(\alpha - 1)t + (1 - \alpha)\mathrm{ES}_\alpha(X)\} + t \\
&= \sup_{\alpha \in [0,1]} \{(1 - \alpha)(\mathrm{ES}_\alpha(X) - t)\} + t.
\end{aligned}$$

By the reverse ES optimization formula in Theorem 2, we know that $\alpha \in [\mathbb{P}(X < t), \mathbb{P}(X \le t)]$ is a maximizer of function $\alpha \mapsto (1 - \alpha(\mathrm{ES}_\alpha(X) - t))$, and hence the supremum above is attainable. Thus, we can conclude that

$$f_1^*(t) = \mathbb{E}[(X - t)_+] + t = \mathbb{E}[X \vee t].$$

The proof of the second Fenchel-Legendre transform follows the same routine. We apply the Fenchel-Legendre transform to the convex function $f_2(\alpha) = \alpha \mathrm{ES}_\alpha^-(X)$. Then we have

$$\begin{aligned}
f_2^*(t) &= \sup_{\alpha \in [0,1]} \left\{\alpha t - \alpha \mathrm{ES}_\alpha^-(X)\right\} \\
&= \sup_{\alpha \in [0,1]} \left\{\alpha t - \mathbb{E}(X) + \int_0^1 \mathrm{VaR}_\beta^-(X)\,\mathrm{d}\beta - \int_0^\alpha \mathrm{VaR}_\beta^-(X)\,\mathrm{d}\beta\right\} \\
&= \sup_{\alpha \in [0,1]} \{(1 - \alpha)(\mathrm{ES}_\alpha(X) - t)\} + t - \mathbb{E}[X].
\end{aligned}$$

By Theorem 2, we conclude that

$$f_2^*(t) = \mathbb{E}[(X - t)_+] + t - \mathbb{E}[X] = \mathbb{E}[(t - X)_+].$$

We can check that both $f_1^* : t \mapsto \mathbb{E}[X \vee t]$ and $f_2^* : t \mapsto \mathbb{E}[(t - X)_+]$ are convex functions. $\qquad \square$

By applying Legendre-Fenchel transform mechanism to the convex functions $f_1^* : t \mapsto \mathbb{E}[X \vee t]$ and $f_2^* : t \mapsto \mathbb{E}[(t - X)_+]$, one obtains the corresponding quantile-based functions f_1 and f_2 in

Proposition 4 as their conjugate functions.

8   Conclusion

The reverse ES optimization formula obtained in Theorem 2 serves as a dual formula to the celebrated ES optimization formula of Rockafellar and Uryasev (2000, 2002), and they are connected via the Fenchel-Legendre transforms. This new formula reveals profound symmetries between these two formulas regarding to their functional properties, parametric forms, optimization problems and the solutions to the optimization problems, and it can be generalized for the class of OCE of Ben-Tal and Teboulle (2007). The reverse ES optimization formula is particularly useful when directly calculating the mean excess loss is cumbersome, and this is illustrated by two settings of model uncertainty. The new formulas are applied to settings of model uncertainty and two insurance datasets.

The new formula may appear simple to risk experts, although we could not find it in the literature. The reason why such a natural formula has not been studied could partially be explained by the fact that the need for utilizing existing ES results to compute the mean excess loss mainly arises in the recent years, when model uncertainty is actively studied in quantitative risk management, as we discuss in Sections 4 and 5.

The main purpose of this paper is to introduce the new formula and discuss its direct implications. Given the importance of both the mean excess function and ES in actuarial science and risk management, we are optimistic about other potential applications of the formula, which will need future exploration.

Acknowledgements

References

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

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

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

Blanchet, J., Chen, L., and Zhou, X. Y. (2022). Distributionally robust mean-variance portfolio selection with Wasserstein distances. *Management Science*, **68**(9), 6382–6410.

Burzoni, M., Munari, C. and Wang, R. (2022). Adjusted Expected Shortfall. *Journal of Banking and Finance*, **134**, 106297.

Denuit, M., Dhaene, J., Goovaerts, M.J., Kaas, R. (2005). *Actuarial Theory for Dependent Risks*. Wiley: Chichester, UK.

De Vylder, F. and Goovaerts, M. J. (1982). Analytical best upper bounds on stop-loss premiums. *Insurance: Mathematics and Economics*, **1**(3), 163–175.

Durrett, R. (2010). *Probability: Theory and Examples*. Cambridge University Press. Fourth Edition.

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

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

Embrechts, P. and Wang, R. (2015). Seven proofs for the subadditivity of Expected Shortfall. *Dependence Modeling*, **3**, 126–140.

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

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica*, **39**(6), 1037–1039.

Ghaoui, L. M. E., Oks, M. and Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization. *Operations Research*, **51**, 543–556.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, **18**, 141–153.

Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R.* Springer Science & Business Media.

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.

Liu, F., Cai, J., Lemieux, C. and Wang, R. (2020). Convex risk functionals: Representation and applications. *Insurance: Mathematics and Economics*, **90**, 66–79.

Liu, F., Mao, T., Wang, R. and Wei, L. (2022). Inf-convolution, optimal allocations, and model uncertainty for tail risk measures. *Mathematics of Operations Research*, **47**(3), 2494–2519.

Maccheroni, F., Marinacci, M. and Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, **74**(6), 1447–1498.

Mao, T., Wang, R. and Wu, Q. (2022). Model aggregation for risk evaluation and robust optimization. *arXiv*: 2201.06370.

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

Jagannathan, R. (1977). Technical note–Minimax procedure for a class of linear programs under uncertainty.

Operations Research. **25**(1), 173-177.

Pesenti, S., Wang, Q. and Wang, R. (2020). Optimizing distortion riskmetrics with distributional uncertainty. *arXiv*: 2011.04889.

Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic Constrained Optimization* (Ed. Uryasev, S.), pp. 272–281, Springer, Dordrecht.

Pflug, G. C. and Römisch, W. (2007). *Modeling, Measuring and Managing Risk.*. World Scientific, Singapore.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, **2**, 21–42.

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

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

von Neumann, J. and Morgenstern, O. (1947). *Theory of Games and Economic Behavior.* Princeton University Press, Second Edition.

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

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

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

Zhu, S., and Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, **57**(5), 1155–1168.