Distributionally robust insurance under the Wasserstein distance

Tim J. Boonen[1] and Wenjun Jiang[2]

[1] *Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, People's Republic of China. E-mail: tjboonen@hku.hk*

[2] *Department of Mathematics and Statistics, University of Calgary, Calgary, AB, T2N 1N4, Canada. E-mail: wenjun.jiang@ucalgary.ca*

Abstract

This paper studies the optimal insurance contracting from the perspective of a decision maker (DM) who has an ambiguous understanding of the loss distribution. The ambiguity set of loss distributions is represented as a p-Wasserstein ball, with $p \in \mathbb{Z}^+$, centered around a specific benchmark distribution. The DM selects the indemnity function that minimizes the worst-case risk within the risk-minimization framework, considering the constraints of the Wasserstein ball. Assuming that the DM is endowed with a convex distortion risk measure and that insurance pricing follows the expected-value premium principle, we derive the explicit structures of both the indemnity function and the worst-case distribution using a novel survival-function-based representation of the Wasserstein distance. We examine a specific example where the DM employs the GlueVaR and provide numerical results to demonstrate the sensitivity of the worst-case distribution concerning the model parameters.

Keywords: Optimal insurance, robustness, distortion risk measure, Wasserstein distance, GlueVaR.

JEL code: C71, G22.

1 Introduction

Insurance is an effective risk-hedging tool which helps safeguard individuals against substantial financial losses. An insurance contract typically consists of two components: the indemnity function, which clarifies the insurer's promised indemnity for different values of the decision maker's (DM) loss, and the premium, which is the *ex ante* compensation received by the insurer. In practice, the premium is often a given function of the indemnity function, which is why the optimal insurance contracting problem boils down to the determination of the optimal indemnity function only. Substantial research on optimal insurance contract theory has emerged in the recent decades. We refer to Albrecher et al. (2017) or Cai and Chi (2020) for reviews of the most recent developments.

A fundamental concern in insurance contracting is the mechanism used to determine the shape of the insurance indemnities. The insurer may offer a menu of possible contracts, or the DM can freely select any feasible indemnity. Alternatively, the optimal contract may be designed from the perspective of both parties, which yield the so-called *Pareto optimal* contracts that have gained substantial popularity recently. Our focus is on the optimal insurance contracting that considers the DM's interest, who is able to purchase indemnity functions via a given premium principle. Regarding the preferences of the DM, there are two main streams in the literature: maximizing expected utility or minimizing a risk measure. This paper belongs to the second category, and we assume that the DM is endowed with a convex distortion risk measure. The distortion risk measure is now widely applied in optimal insurance contracting, and it includes the two prominent risk measures Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR) as special cases. Both risk measures have been extensively applied and studied in the banking and insurance industry. For the readers who are interested in the design of the optimal insurance under a distortion risk measure, we refer to Cheung (2010); Chi and Tan (2011); Assa (2015); Zhuang et al. (2016); Boonen and Jiang (2022) and the references therein.

A prevailing setting adopted by a vast literature on optimal insurance contracting is that the DM possesses complete knowledge about her loss distribution. This assumption has been challenged by many recent studies by considering an uncertainty set of loss distributions. However, a central question needs to be answered before adopting such a setting: how to choose the uncertainty set that can best depict the DM's ambiguity and knowledge towards her loss distribution? Within the realm of our topic, there are mainly three approaches for constructing such uncertainty sets. A simple choice is to consider a finite set of candidate distributions, where the specification of the candidate distributions is left to expert opinions. With such an uncertainty set, Asimit et al. (2017) and Jiang et al. (2020) study the optimal insurance contract using the minimax theorem. Alternatively, a second approach to generate the uncertainty set is via fixing some moments of the distribution. Representative papers for this approach are Liu and Mao (2022) and Xie et al. (2023), which investigated the optimal insurance under VaR, TVaR and expectile when the first two moments of the loss distribution are exactly known. The third approach is to consider distance metrics to determine an uncertainty set. The distance-based approach considers all distributions that are close enough to a given benchmark distribution, where the "closeness" is measured using a particular metric. Birghila and Pflug (2019); Birghila et al. (2023) and Boonen and Jiang (2024) study the related distributionally robust insurance problems with the 1-Wasserstein distance and the L^1 and L^2 distance metrics. Recently, Cai et al. (2023) re-examined the optimal stop-loss function under the distributional uncertainty, where the uncertainty set is specified by using both the moment-based and distance-based approaches. This paper is situated in the third and last stream, where we model the uncertainty set by using the p-Wasserstein distance, where $p \in \mathbb{Z}^+$. We also adopt the minimax framework, and the focus of this paper is to derive expressions of the worst-case distribution corresponding to the optimal insurance contract.

The Wasserstein distance is widely used in optimal transport and robustness for modeling distances between distributions, as it provides a meaningful and intuitive interpretation. It is a true metric, and is robust against small perturbations in the distributions (Villani, 2009). Moreover, the Wasserstein distance has desirable properties such as differentiability and strong convexity.

Compared to the existing literature, the contributions of this paper cover several aspects. First, we develop a new representation for the p-Wasserstein distance of any positive integer p by using the survival functions. Second, without assuming the parametric forms of the indemnity function, the explicit structures of both the optimal indemnity function and the worst-case distribution are derived using the minimax theorem and the Karush–Kuhn–Tucker (KKT) method. This generalizes the work done by Birghila and Pflug (2019), which derived the optimal indemnity function numerically under the 1-Wasserstein distance.

The rest of this paper is organized as follows. Section 2 reviews some basics of distortion risk measures and sets up the main problem of this paper. Section 3 presents the main results of this paper. Section 4 provides a detailed example with the GlueVaR. Section 5 concludes and provides directions for potential future research. All proofs are delegated to Appendix B.

2 Distortion risk measures and problem formulation

Throughout the paper, we use the notations $x \wedge y = \min\{x, y\}$, $x \vee y = \max\{x, y\}$ and $(x)_+ = \max\{x, 0\}$. Moreover, $\mathbb{1}_A(x)$ is the indicator function, which is equal to 1 if $x \in A$ and 0 otherwise.

2.1 Distortion risk measures

Let there be a one-period economy. We fix a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ such that Ω is atom-less and $\mathcal{B}$ is the Borel σ-algebra. A DM is facing an insurable, non-negative loss represented by a random variable X. Let $\mathcal{P}$ be the collection of probability measures of X with support being a subset of $[0, M]$, where $M > 0$. Let $F_P(x)$ be the cumulative distribution function (CDF) of X under probability measure $P \in \mathcal{P}$, for which the survival function is defined as $S_P(x) := 1 - F_P(x) = P(X > x)$.

The convex distortion risk measure of a random variable Z on the measurable space $(\Omega, \mathcal{B})$ is allowed to depend on a probability measure P, and is given by:

$$\rho_g^P(Z) = \int_0^\infty g(P(Z > z))dz + \int_{-\infty}^0 [g(P(Z > z)) - 1]dz, \tag{2.1}$$

where g is called the distortion function that is increasing[1] and concave over $[0, 1]$ and satisfies $g(0) = 0$ and $g(1) = 1$. This implies that the function g is differentiable almost everywhere. For a fixed probability measure P, the distortion risk measure ρ_g^P satisfies the following properties (Wang et al., 1997; Denuit et al., 2006):

- Comonotonic additivity: $\rho_g^P(Z + Y) = \rho_g^P(Z) + \rho_g^P(Y)$ for comonotonic random variables Z and Y[2].
- Sub-additivity: $\rho_g^P(Z + Y) \leq \rho_g^P(Z) + \rho_g^P(Y)$ for any two random variables Z and Y.

Note that Comonotonic additivity and the fact that $\rho_g^P(1) = 1$ imply Translation invariance, which is defined as $\rho_g^P(Z + c) = \rho_g^P(Z) + c$ for all $c \in \mathbb{R}$. Moreover, convex distortion risk measures are coherent in the sense of Artzner et al. (1999), and are averse to mean-preserving spreads (Yaari, 1987).

2.2 Problem formulation

Let an insurance contract be given by a pair $(I, \pi(I))$, where I is the indemnity function and $\pi(I)$ is the corresponding premium. If the DM purchases $(I, \pi(I))$, its end-of-period loss can be written as $X - I(X) + \pi(I)$.

The insurer uses its selected probability measure $\mathbb{Q} \in \mathcal{P}$ to price insurance, and we assume that the premium $\pi(I)$ is given by the expected value principle:

$$\pi(I) = (1 + \theta)\mathbb{E}^{\mathbb{Q}}[I(X)] = (1 + \theta)\int_0^M I(x)dF_{\mathbb{Q}}(x), \tag{2.2}$$

where $\theta \geq 0$ is called the safety loading factor.

For the indemnity function, we impose the so-called *incentive compatibility* condition. This condition, first proposed by Huberman et al. (1983), requires that the losses borne by the DM and insurer are both increasing. This reduces the DM's motivation of manipulating or under-reporting the losses and thus alleviates the *ex post* moral hazard issues. Under the incentive compatibility condition, the indemnity function must be in the following set:

$$\mathcal{I} = \left\{ I : [0, M] \mapsto [0, M] \,\middle|\, I(0) = 0,\ 0 \leq I(x_2) - I(x_1) \leq x_2 - x_1 \text{ for any } 0 \leq x_1 \leq x_2 \leq M \right\}.$$

Notably, if $I \in \mathcal{I}$, then $X - I(X)$ and $I(X)$ are comonotonic. The set $\mathcal{I}$ is quite large and includes many well known indemnity functions, such as the stop-loss, quota-share and truncated stop-loss

[1] We do not distinguish between "increasing" and "non-decreasing" in the paper.

[2] The random variables Z and Y are called comonotonic if $Z = K_1(T)$ and $Y = K_2(T)$ for some increasing functions K_1 and K_2, where T is a random variable.

functions[3]. Additionally, if $I \in \mathcal{I}$, then I is 1-Lipschitz continuous and admits the following integral representation

$$I(x) = \int_0^x \eta(t)dt, \quad x \in [0, M], \tag{2.3}$$

where η is called the marginal indemnity function (MIF) (Assa, 2015; Zhuang et al., 2016).

Throughout the paper, we assume that the DM aims to choose the insurance contract that can minimize its risk exposure, measured by a distortion risk measure. When distributional uncertainty is absent and the probability measure $P \in \mathcal{P}$ is known, the following problem is faced by the DM, which has been extensively studied in the literature:[4]

$$\min_{I \in \mathcal{I}} \rho_g^P(X - I(X) + \pi(I)). \tag{2.4}$$

The DM is assumed to be uncertain about the underlying distribution P of X due to its limited access to the market information and historical data. Then, it aims to minimize the worst-case risk measure, that is, the largest risk measure among a given set of possible distributions. The DM will consider a set of distributions around some benchmark distribution $F_{\mathbb{B}}$ with $\mathbb{B} \in \mathcal{P}$. We use the Wasserstein distance metric to measure the distance between the candidate distribution and the benchmark distribution. As explained in Pesenti and Jaimungal (2023), an important reason for using such metric in uncertainty modelling is that it allows comparison between distributions with differing supports. The following definition is for the Wasserstein distance between two cumulative distribution functions.

Definition 2.1. *The p-Wasserstein distance, where $p \in \mathbb{Z}^+$, between F_{P_1}, F_{P_2} with $P_1, P_2 \in \mathcal{P}$ is given by*

$$W(F_{P_1}, F_{P_2}) = \inf_{X \sim F_{P_1}, Y \sim F_{P_2}} \mathbb{E}\left[|X - Y|^p\right]^{\frac{1}{p}},$$

where $X \sim F_P$ means that the CDF of the random variable X is F_P.

The Wasserstein distance plays an important role in optimal transport (Villani, 2009), and it has a closed-form representation (Panaretos and Zemel, 2019) using the quantile functions:

$$W_p(F_{P_1}, F_{P_2}) = \left(\int_0^1 \left|F_{P_1}^{-1}(t) - F_{P_2}^{-1}(t)\right|^p dt\right)^{\frac{1}{p}}, \tag{2.5}$$

where the notation $K^{-1}(t)$ denotes the left-inverse of the function K at the point t:

$$K^{-1}(t) = \inf\left\{x \in \text{dom}(K) : K(x) \geq t\right\}.$$

With the definition of the p-Wasserstein distance, the uncertainty set that will be studied in this paper is a ball centering around a benchmark distribution $F_{\mathbb{B}}$:

$$\mathcal{P}_\epsilon := \left\{P \in \mathcal{P} : W_p(F_P, F_{\mathbb{B}}) \leq \epsilon^{\frac{1}{p}}\right\}. \tag{2.6}$$

[3]The stop-loss function is given by $I(x) = (x-d)_+$ for some $d \geq 0$. The quota-share function is given by $I(x) = cx$ for some fraction $c \in [0, 1]$. The truncated stop-loss function is given by $I(x) = (x - d_1)_+ \wedge d_2$ for some $d_1, d_2 \geq 0$.

[4]If $P = \mathbb{Q}$, we here refer to Assa (2015), Zhuang et al. (2016) and Lo (2017b), and for generic $P \in \mathcal{P}$ we refer to Boonen (2016).

If $p=2$, as shown in Bernard et al. (2023), the constraint in (2.6) implies

$$\epsilon \geq W(F_P, F_{\mathbb{B}})^2 \geq (\mu^P - \mu^{\mathbb{B}})^2 + (\sigma^P - \sigma^{\mathbb{B}})^2, \tag{2.7}$$

where $\mu^P, \mu^{\mathbb{B}}$ denote the expectations of X under the probability measures P and $\mathbb{B}$ respectively, and $\sigma^P, \sigma^{\mathbb{B}}$ denote the standard deviations of X under the probability measures P and $\mathbb{B}$ respectively. From this perspective, imposing the Wasserstein-type constraint also places some uncertainties on the moments of X. This point of view will be further demonstrated in the later analyses.

If $\epsilon = 0$, then the set $\mathcal{P}_0$ is a singleton, which yields the case without distributional uncertainty. This case is solved by Cui et al. (2013), Assa (2015) and Boonen (2016), where the first two focus on the homogeneous-belief-based problems (i.e., $\mathbb{B}=\mathbb{Q}$) and the last focuses on the heterogeneous-belief-based problem (i.e., $\mathbb{B}\neq\mathbb{Q}$). To avoid this case, we set $\epsilon > 0$ in (2.6) in the rest of this paper. Since $M<\infty$, it holds that $\mathbb{E}^{\mathbb{Q}}[X] < \infty$ and $\sup_{P\in\mathcal{P}_\epsilon} \rho_g^P(X) < \infty$.

The main problem that we aim to solve is presented below.

Problem 1. *For a given $\epsilon > 0$, solve*

$$\inf_{I\in\mathcal{I}} \sup_{P\in\mathcal{P}_\epsilon} \rho_g^P(X - I(X) + \pi(I)).$$

Next section is devoted to obtain the full analytical solution to Problem 1.

3 The optimal indemnity function and the worst-case distribution

3.1 The structure of indemnity function

The set $\mathcal{I}$ is convex and bounded. By using the distance metric $d(I^1, I^2) = \max_{t\in[0,M]} \mid I^1(t) - I^2(t) \mid$, for any $I^1, I^2 \in \mathcal{I}$, the set $\mathcal{I}$ is compact under this metric d by Arzelà-Ascoli's theorem. As the distortion risk measure is translation invariant and comonotonic additive, it holds that $\rho_g^P(X - I(X) + \pi(I))$ is linear in I, for $I\in\mathcal{I}$. Moreover, it is easy to verify that $\mathcal{P}_\epsilon$ is also convex, and $\rho_g^P(X - I(X) + \pi(I))$ is concave in F_P due to the concavity of the distortion function g. Hence, we can apply the minimax theorem (see Theorem A.1 in Appendix A for more details) to Problem 1, and we then obtain the following problem:

$$\sup_{P\in\mathcal{P}_\epsilon} \inf_{I\in\mathcal{I}} \rho_g^P(X - I(X) + \pi(I)). \tag{3.1}$$

The inner problem of (3.1) coincides with Problem (2.4), for which the solution is well-known in the literature (see, e.g., Boonen, 2016). We present its solution below.

Lemma 3.1. *For a fixed $P\in\mathcal{P}_\epsilon$, the optimal indemnity function to the inner problem of (3.1) is given by $I^*(x;P) = \int_0^x \eta^*(t;P)dt$ where*

$$\eta^*(t;P) = \mathbb{1}_{\{t:(1+\theta)S_{\mathbb{Q}}(t) < g(S_P(t))\}}(t) + \gamma(t)\cdot \mathbb{1}_{\{t:(1+\theta)S_{\mathbb{Q}}(t) = g(S_P(t))\}}(t),$$

where γ is a $[0,1]$-valued and Lebesgue-measurable function.

As shown in Lemma 3.1, the optimal indemnity function that solves the inner problem of (3.1) may not be unique due to the non-uniqueness of $\gamma(t)$. In this lemma, we can interpret the term $(1+\theta)S_{\mathbb{Q}}(x) - g(S_P(x))$ as the net price for purchasing the marginal coverage $I'(x)$ when the realized loss is x. Lemma 3.1 tells that the DM will purchase the maximum marginal coverage (i.e., $I'(x) = 1$) when this net price is negative and purchase zero marginal coverage (i.e., $I'(x) = 0$) when this net price is positive. Lemma 3.1 implies that if the worst-case survival function S_{P^*} (written as S^* in the sequel) is known, then $I^*(x; P^*)$ is the solution to Problem 1. In the next section, we derive the worst-case survival function S^* analytically.

3.2 The worst-case distribution

With the optimal indemnity function in Lemma 3.1, the objective function for Problem (3.1) can be written as

$$\begin{aligned}
\rho_g^P(X - I^*(X;P) + \pi(I)) &= \rho_g^P(X) - \rho_g^P(I^*(X;P)) + \pi(I^*(\cdot;P)) \\
&= \rho_g^P(X) + \int_0^M \big(-g(S_P(x)) + (1+\theta)S_{\mathbb{Q}}(x)\big)\, \mathbb{1}_{\{x: g(S_P(x)) > (1+\theta)S_{\mathbb{Q}}(x)\}}(x)dx \\
&= \rho_g^P(X) - \int_0^M \big(g(S_P(x) - (1+\theta)S_{\mathbb{Q}}(x))\big)_+ \, dx \\
&= \int_0^M \left\{ g(S_P(x)) - (g(S_P(x)) - (1+\theta)S_{\mathbb{Q}}(x))_+ \right\} dx \\
&= \int_0^M \big(g(S_P(x)) \wedge (1+\theta)S_{\mathbb{Q}}(x)\big)\, dx.
\end{aligned}$$

The focus now is to find the worst-case distribution that solves the outer problem of (3.1), or

$$\begin{cases}
\sup\limits_{P\in\mathcal{P}} \displaystyle\int_0^M \big(g(S_P(x)) \wedge (1+\theta)S_{\mathbb{Q}}(x)\big)\, dx, \\
\text{s.t.} \displaystyle\int_0^1 \left|F_P^{-1}(t) - F_{\mathbb{B}}^{-1}(t)\right|^p dt \le \epsilon.
\end{cases} \tag{3.2}$$

For convenience, let $\mathcal{S}$ be the class of all survival functions S_P of X for $P \in \mathcal{P}$. Since the problem in (3.2) only depends on P via S_P, it can be written as

$$\begin{cases}
\sup\limits_{S_P\in\mathcal{S}} \displaystyle\int_0^M \big(g(S_P(x)) \wedge (1+\theta)S_{\mathbb{Q}}(x)\big)\, dx, \\
\text{s.t.} \displaystyle\int_0^1 \left|F_P^{-1}(t) - F_{\mathbb{B}}^{-1}(t)\right|^p dt \le \epsilon.
\end{cases} \tag{3.3}$$

In the following, we denote by P^* the worst-case measure for Problem (3.3) and by S_{P^*} the associated survival function. The formulation of Problem (3.3) tells that: if $S_P(x) = S_{\mathbb{B}}(x)$ for all $x \in [0, M]$, then $F_P^{-1}(t) = F_{\mathbb{B}}^{-1}(t)$ for all $t \in [0,1]$. As the objective function of (3.3) is increasing in S_P, it must hold that $S_{P^*}(x) \ge S_{\mathbb{B}}(x)$ for all $x \in [0, M]$, or equivalently, $F_{P^*}^{-1}(t) \ge F_{\mathbb{B}}^{-1}(t)$ for all $t \in [0,1]$. For the ease of discussing the main results, we conclude this finding in the next lemma.

Lemma 3.2. *There exists a worst-case survival function that satisfies $S_{P^*}(x) \geq S_{\mathbb{B}}(x)$ for all $x \in [0, M]$, and the associated quantile function $F_{P^*}^{-1}(t) \geq F_{\mathbb{B}}^{-1}(t)$ for all $t \in [0, 1]$.*

With Lemma 3.2, to solve Problem (3.2) it suffices to restrict ourselves to the set

$$\tilde{\mathcal{P}}_\epsilon := \left\{ P \in \mathcal{P}_\epsilon : \; F_P^{-1}(t) \geq F_{\mathbb{B}}^{-1}(t) \; \forall \; t \in [0, 1] \right\}. \tag{3.4}$$

By focusing on the set $\tilde{\mathcal{P}}_\epsilon$, the left-hand-side of the constraint of (3.3) can be written as

$$\begin{aligned} \int_0^1 \left|F_P^{-1}(t) - F_{\mathbb{B}}^{-1}(t)\right|^p dt &= \int_0^1 \left(F_P^{-1}(t) - F_{\mathbb{B}}^{-1}(t) \right)^p dt \\ &= \sum_{k=0}^{p} \binom{p}{k} (-1)^{p-k} \int_0^1 (F_P^{-1}(t))^k (F_{\mathbb{B}}^{-1}(t))^{p-k} dt. \end{aligned} \tag{3.5}$$

The following lemma will play a key role in the subsequent discussions.

Lemma 3.3. *For any probability measures $P_1, P_2 \in \mathcal{P}$ and $k_1, k_2 \in \mathbb{Z}^+$, we have*

$$\int_0^1 (F_{P_1}^{-1}(t))^{k_1} (F_{P_2}^{-1}(t))^{k_2} dt = \int_0^M \int_0^M \left\{ k_1 x^{k_1 - 1} k_2 y^{k_2 - 1} S_{P_1}(x) \wedge S_{P_2}(y) \right\} dxdy. \tag{3.6}$$

Thanks to the survival-function-based representation of moments (Chakraborti et al., 2018):

$$\mathbb{E}^P[X^k] = \int_0^M x^k dF_P(x) = \int_0^1 \left(F_P^{-1}(t) \right)^k dt = \int_0^M k x^{k-1} S_P(x) dx, \tag{3.7}$$

and Lemma 3.3, by some algebraic manipulations, Eq. (3.5) can be further written as

$$\begin{aligned} &\sum_{k=0}^{p} \binom{p}{k} (-1)^{p-k} \int_0^1 (F_P^{-1}(t))^k (F_{\mathbb{B}}^{-1}(t))^{p-k} dt \\ =& \int_0^1 \left(F_P^{-1}(t) \right)^p dt + \sum_{k=1}^{p-1} \binom{p}{k} (-1)^{p-k} \int_0^1 (F_P^{-1}(t))^k (F_{\mathbb{B}}^{-1}(t))^{p-k} dt + (-1)^p \int_0^1 \left(F_{\mathbb{B}}^{-1}(t) \right)^p dt \\ =& \int_0^M p x^{p-1} S_P(x) dx - p(p-1) \int_0^M \int_0^M \left\{ (x-y)^{p-2} S_P(x) \wedge S_{\mathbb{B}}(y) \right\} dxdy + (-1)^p \int_0^M p x^{p-1} S_{\mathbb{B}}(x) dx. \end{aligned} \tag{3.8}$$

With (3.8) the constraint of the problem (3.3) can be re-written as

$$\int_0^M x^{p-1} S_P(x) dx - (p-1) \int_0^M \int_0^M \left\{ (x-y)^{p-2} S_P(x) \wedge S_{\mathbb{B}}(y) \right\} dxdy \leq \zeta, \tag{3.9}$$

where $\zeta = \frac{1}{p}\epsilon - (-1)^p \int_0^M x^{p-1} S_{\mathbb{B}}(x) dx$. The following lemma shows that the left-hand side of the above inequality is convex in S_P for $P \in \tilde{\mathcal{P}}_\epsilon$, which will be helpful for re-formulating the main problem.

Lemma 3.4. *The integral*

$$\int_0^M x^{p-1} S_P(x) dx - (p-1) \int_0^M \int_0^M \left\{ (x-y)^{p-2} S_P(x) \wedge S_{\mathbb{B}}(y) \right\} dxdy$$

is convex in S_P for $P \in \tilde{\mathcal{P}}_\epsilon$.

With the aid of Lemma 3.4, since the objective function in (3.3) is concave in $S_P(x)$, and furthermore $S_P = S_{\mathbb{B}}$ is strictly feasible to Problem (3.3) due to $\epsilon > 0$, the Slater condition holds (Boyd and Vandenberghe, 2004). Therefore, strong duality holds, and solving Problem (3.3) is equivalent to solving

$$\sup_{S_P \in \tilde{\mathcal{S}}} \int_0^M \left\{ g(S_P(x)) \wedge (1+\theta)S_{\mathbb{Q}}(x) - \beta \left(x^{p-1} S_P(x) - \int_0^M (p-1)(x-y)^{p-2} S_P(x) \wedge S_{\mathbb{B}}(y)dy \right) \right\} dx \tag{3.10}$$

for some $\beta \geq 0$ such that the constraint of (3.9) is satisfied, where $\tilde{\mathcal{S}} = \left\{ S_P : P \in \tilde{\mathcal{P}}_\epsilon \right\}$.

To state the main results, we define

$$x_0 := \sup \left\{ x \in [0, M) : S_{\mathbb{Q}}(x) \geq \frac{1}{1+\theta} \right\}. \tag{3.11}$$

The following theorem summarizes the worst-case survival function for the case where $\beta = 0$, which indicates a slack constraint. Its proof is similar to that of Theorem B.1 of Boonen and Jiang (2024) and thus omitted.

Theorem 3.1. *Let x_0 be defined in (3.11), and let*

$$\begin{aligned} \mathcal{A} &:= \left\{ x \in [x_0, M] : (1+\theta)S_{\mathbb{Q}}(x) \geq g(S_{\mathbb{B}}(x)) \right\}, \\ \mathcal{B} &:= [x_0, M] \backslash \mathcal{A}. \end{aligned}$$

If

$$\int_0^M x^{p-1} \tilde{S}^*(x)dx - (p-1) \int_0^M \int_0^M \left\{ (x-y)^{p-2} \tilde{S}^*(x) \wedge S_{\mathbb{B}}(y) \right\} dxdy \leq \zeta,$$

where

$$\tilde{S}^*(x) = (t_0 \vee S_{\mathbb{B}}(x))\mathbb{1}_{[0,x_0)}(x) + g^{-1}((1+\theta)S_{\mathbb{Q}}(x))\mathbb{1}_{\mathcal{A}}(x) + S_{\mathbb{B}}(x)\mathbb{1}_{\mathcal{B}}(x), \tag{3.12}$$

and $t_0 := g^{-1}(1)$, then the worst-case survival function that solves the problem (3.3) is $S_{P^*} = \tilde{S}^*$.

To present the main result for the case where $\beta > 0$, which indicates a binding constraint, we define

$$K(t; x) = g(t) - \beta \left(x^{p-1} t - \int_0^M (p-1)(x-y)^{p-2} t \wedge S_{\mathbb{B}}(y)dy \right), \tag{3.13}$$

which is differentiable with respect to t almost everywhere over $[S_{\mathbb{B}}(x), 1]$.

Theorem 3.2. *Let x_0 be defined in (3.11) and $\mathcal{A}$ and $\mathcal{B}$ be defined in Theorem 3.1. If $\beta > 0$ in Problem (3.10), the worst-case survival function that solves Problem (3.10) is given by*

$$S_{P^*}(x; \beta) = \hat{S}(x; \beta)\mathbb{1}_{[0,x_0)}(x) + \left(\hat{S}(x; \beta) \wedge g^{-1}((1+\theta)S_{\mathbb{Q}}(x)) \right) \mathbb{1}_{\mathcal{A}}(x) + S_{\mathbb{B}}(x)\mathbb{1}_{\mathcal{B}}(x), \tag{3.14}$$

where

$$\hat{S}(x; \beta) = \inf \left\{ t \in [S_{\mathbb{B}}(x), 1] : K'(t; x) \leq 0 \right\}, \tag{3.15}$$

with $\inf \emptyset$ *being the right-end point of the interval. Here* β *is such that*

$$\int_0^M x^{p-1} S_{P^*}(x;\beta)dx - (p-1)\int_0^M \int_0^M \left\{(x-y)^{p-2} S_{P^*}(x;\beta) \wedge S_{\mathbb{B}}(y)\right\} dxdy = \zeta. \tag{3.16}$$

Similar results can be derived by replacing the p-Wasserstein distance with the L^p distance to calculate the distance between distribution functions. We present those results in Appendix C, those can be seen as an extension of the results in Boonen and Jiang (2024), who only study the L^1 and L^2 distance metrics.

It is found that $\tilde{S}^*(x) = S_{\mathbb{B}}(x)$ and $S_{P^*}(x) = S_{\mathbb{B}}(x)$ for $x \in \mathcal{B}$. Note that on $\mathcal{B}$, we have

$$(1+\theta)S_{\mathbb{Q}}(x) < g(S_{\mathbb{B}}(x)) \leq g(S_{P^*}(x)),$$

which leads to a negative net price for marginal coverage (see Section 3.1). As such, the DM has no incentive to consider worse situation for the losses on $\mathcal{B}$ as they are ceded out to the insurer. For $x \in \mathcal{A}$, if $\beta = 0$, or $\beta > 0$ and $\hat{S}(x;\beta) \geq g^{-1}((1+\theta)S_{\mathbb{Q}}(x))$, the worst-case survival function $S_{P^*}(x) = g^{-1}((1+\theta)S_{\mathbb{Q}}(x))$ (or $S_{P^*}(x;\beta) = g^{-1}((1+\theta)S_{\mathbb{Q}}(x))$), the net price for marginal coverage is equal to $(1+\theta)S_{\mathbb{Q}}(x) - g(S_{P^*}(x)) = 0$, indicating that the DM is indifferent between with and without insurance. If $\beta > 0$ and $\hat{S}(x;\beta) < g^{-1}((1+\theta)S_{\mathbb{Q}}(x))$, then the net price becomes $(1+\theta)S_{\mathbb{Q}}(x) - g(\hat{S}(x;\beta)) > 0$, indicating that the DM would retain those losses.

Among different p-Wasserstein distances, $p = 1$ is of particular interest (Birghila and Pflug, 2019). In that case, for any $P \in \tilde{\mathcal{P}}_\epsilon$, the constraint of the problem 3.3 becomes

$$\begin{aligned}
\int_0^1 \left|F_P^{-1}(t) - F_{\mathbb{B}}^{-1}(t)\right| dt =& \int_0^M \left|F_P(x) - F_{\mathbb{B}}(x)\right| dx \\
=& \int_0^1 \left(F_P^{-1}(t) - F_{\mathbb{B}}^{-1}(t)\right) dt \\
=& \mathbb{E}^P[X] - \mathbb{E}^{\mathbb{B}}[X] \leq \epsilon.
\end{aligned}$$

In other words, considering the 1-Wasserstein distance ball is equivalent to considering the L^1 ball or the uncertainty of the first moment.

For $p_1, p_2 \in \mathbb{Z}^+$, if $p_1 < p_2$, then by Holder's inequality we have

$$\begin{aligned}
\left(\int_0^1 \left|F_P^{-1}(t) - F_{\mathbb{B}}^{-1}(t)\right|^{p_1}\right)^{\frac{1}{p_1}} \leq& \left(\left(\int_0^1 \left|F_P^{-1}(t) - F_{\mathbb{B}}^{-1}(t)\right|^{p_2}\right)^{\frac{p_1}{p_2}} \left(\int_0^1 1^{\frac{p_2}{p_2-p_1}} dt\right)^{\frac{p_2-p_1}{p_2}}\right)^{\frac{1}{p_1}} \\
=& \left(\int_0^1 \left|F_P^{-1}(t) - F_{\mathbb{B}}^{-1}(t)\right|^{p_2}\right)^{\frac{1}{p_2}}.
\end{aligned}$$

As such, a p-Wasserstein distance with a higher p leads to a smaller uncertainty set, and therefore a less worse survival function.

Theorem 3.2 shows the worst-case survival function for the problem (3.1) (or Problem 1), where the optimal indemnity function for the inner problem relies on the survival function of the outer problem. In real-life applications, the inner problem of Problem 1, where the indemnity function is given, is also interesting. It is not difficult to show that the inner problem of Problem 1 can be written as

$$\begin{cases} \sup_{P\in\tilde{\mathcal{P}}_\epsilon} \displaystyle\int_0^M g(S_P(x))(1-I'(x))dx, \\ \text{s.t. } \displaystyle\int_0^M x^{p-1}S_P(x)dx-(p-1)\int_0^M\int_0^M \left\{(x-y)^{p-2}S_P(x)\wedge S_{\mathbb{B}}(y)\right\}dxdy\le\xi. \end{cases} \tag{3.17}$$

The Slater condition is again satisfied since the objective function of (3.17) is concave in S_P, leading to the equivalence between (3.17) and its dual problem

$$\sup_{S_P\in\tilde{\mathcal{S}}}\int_0^M\left\{g(S_P(x))(1-I'(x))-\beta\left(x^{p-1}S_P(x)-\int_0^M(p-1)(x-y)^{p-2}S_P(x)\wedge S_{\mathbb{B}}(y)\right)\right\}dx \tag{3.18}$$

for some $\beta\ge 0$. Now let

$$\widehat{K}(t;x)=g(t)(1-I'(x))-\beta\left(x^{p-1}t-\int_0^M(p-1)(x-y)^{p-2}t\wedge S_{\mathbb{B}}(y)dy\right).$$

The following corollary presents the solution to the problem (3.18) when I belongs to a subset of $\mathcal{I}$.

Corollary 3.1. *If $I\in\mathcal{I}$ is convex, the worst-case survival function that solves the problem (3.18) is given by*

$$S_{P^*}(x;\beta)=\inf\{t\in[S_{\mathbb{B}}(x),1]:\widehat{K}'(t;x)\le 0\},$$

where β is chosen to satisfy the Karush–Kuhn–Tucker (KKT) conditions:

$$\int_0^M x^{p-1}S_{P^*}(x)dx-(p-1)\int_0^M\int_0^M\left\{(x-y)^{p-2}S_{P^*}(x)\wedge S_{\mathbb{B}}(y)\right\}dxdy\le\xi$$

and

$$\beta\left(\int_0^M x^{p-1}S_{P^*}(x)dx-(p-1)\int_0^M\int_0^M\left\{(x-y)^{p-2}S_{P^*}(x)\wedge S_{\mathbb{B}}(y)\right\}dxdy-\xi\right)=0.$$

Corollary 3.1 is applicable for the cases where the indemnity function is a stop-loss function or a proportional-stop-loss function[5]. The convexity of the indemnity function can also guarantee that the well-known Vajda condition is satisfied (Vajda, 1962) such that the proportion of the loss borne by the insurer is increasing with respect to the total loss. We leave the study of the problem (3.18) for other kinds of indemnity functions for future research.

We close this section by remarking the applicability of our methodology in solving other robust optimization problems in the literature, for example, the one in Bernard et al. (2023). They worked out the worst-case distortion risk measure by considering the uncertainty for the loss distribution,

[5]A proportional-stop-loss function is of the form $I(x)=c(x-d)_+$ for some $c\in[0,1]$ and $d\ge 0$.

where the uncertainty set is depicted by a 2-Wasserstein ball and the fixed first- and second-order moments of the loss variable. Their problem can be formulated as

$$\begin{cases} \sup_{P \in \mathcal{P}_\epsilon} \ \rho_g^P(X), \\ \quad \text{s.t. } \mathbb{E}^P[X] = \mu^P, \quad \mathbb{E}^P[X^2] = (\sigma^P)^2, \end{cases} \tag{3.19}$$

where $\mu^P, (\sigma^P)^2$ are fixed quantities that satisfy (2.7). If assuming a convex distortion risk measure ρ_g, we can extend the results of Bernard et al. (2023) to the case with the p-Wasserstein ball, where $p \in \mathbb{Z}^+$. By adopting the survival-function-based representation, the above problem can be written as

$$\begin{cases} \sup_{S_P \in \mathcal{S}} \ \int_0^M g\left(S_P(x)\right) dx, \\ \quad \text{s.t. } \int_0^M S_P(x) dx = \mu^p, \quad \int_0^M 2x S_P(x) dx = (\sigma^P)^2, \\ \qquad \int_0^M x^{p-1} S_P(x) dx - (p-1) \int_0^M \int_0^M \left\{(x-y)^{p-2} S_P(x) \wedge S_{\mathbb{B}}(y)\right\} dx dy \le \zeta, \end{cases} \tag{3.20}$$

where ζ is defined right after (3.9). Through the application of the KKT theorem, one can easily solve the dual problem of (3.20) to get the worst-case survival function. Note that Bernard et al. (2023) apply a quantile formulation for the problem (3.19) and characterizes the worst-case quantile function via the isotonic projection. Our method provides a perspective alternative to theirs.

4 A concrete example when ρ_g^P is GlueVaR

In this section, we study a concrete case where the DM applies GlueVaR as the risk measure. The GlueVaR is first proposed by Belles-Sampera et al. (2014) as a generalization of the commonly used risk measures – VaR and TVaR, all of which belong to the family of distortion risk measures. The GlueVaR has four parameters, and its distortion function is given by

$$g_{\alpha,\gamma}^{r_1,r_2}(x) = \begin{cases} \frac{r_1}{1-\gamma} x, & \text{if } 0 \le x < 1-\gamma, \\ r_1 + \frac{r_2 - r_1}{\gamma - \alpha}\left[x - (1-\gamma)\right], & \text{if } 1-\gamma \le x < 1-\alpha, \\ 1, & \text{if } 1-\alpha \le x \le 1, \end{cases} \tag{4.1}$$

where $0 \le \alpha \le \gamma \le 1$, $r_1 \in [0,1]$ and $r_2 \in [r_1, 1]$. It is not difficult to verify that GlueVaR reduces to VaR if $g_{\alpha,\gamma}^{r_1,r_2} = g_{\alpha,\alpha}^{0,0}$, to TVaR if $g_{\alpha,\gamma}^{r_1,r_2} = g_{\alpha,\alpha}^{1,1}$, and to Range VaR if $g_{\alpha,\gamma}^{r_1,r_2} = g_{\alpha,\gamma}^{0,1}$. We refer to Lv and Wei (2023) for another study of a distributionally robust reinsurance problem under GlueVaR, in which the uncertainty set is generated by the set of all probability measures with given first two moments.

To adapt to the setting of our model, we will only focus on the case where $r_2 = 1$ and $\frac{1-r_1}{\gamma-\alpha} < \frac{r_1}{1-\gamma}$, which results in $g_{\alpha,\gamma}^{r_1,r_2}$ being a concave function on $[0,1]$ (see Figure 1 for an illustration). For simplicity in the discussion, we assume that the DM applies the probability measure $\mathbb{Q}$ of the insurer as the benchmark probability measure, i.e., $\mathbb{B} = \mathbb{Q}$. We also assume that $F_{\mathbb{Q}}$ is continuous and strictly increasing for the ease of later calculations. We note that if $r_1 = 1$, then ρ_g^P reduces to TVaR, which has been studied in Boonen and Jiang (2024) for the L^2 and L^1 ball. Hence, we assume that $r_1 < 1$. For p-Wasserstein distance, we choose $p \geq 2$.

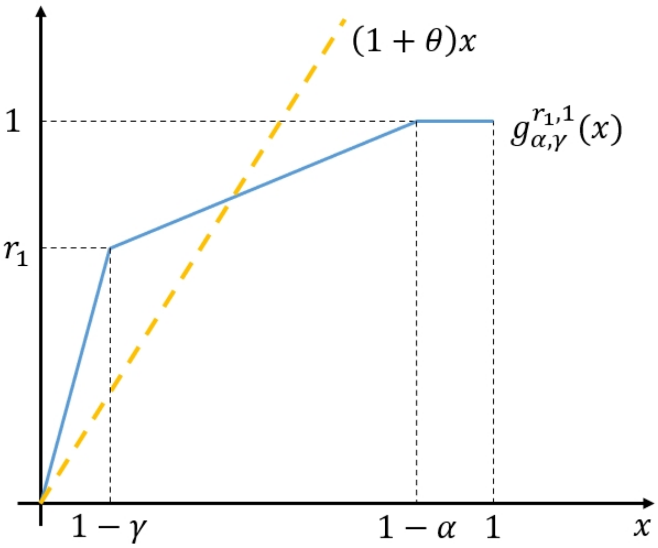

Figure 1: The illustration of the distortion function $g_{\alpha,\gamma}^{r_1,1}$ that is used in Section 4.1.

When $\mathbb{B} = \mathbb{Q}$, we define

$$x_1 = \sup\{x \in [x_0, M) : (1+\theta)S_{\mathbb{Q}}(x) \geq g(S_{\mathbb{Q}}(x))\},$$

and it can be easily shown that $\mathcal{A} = [x_0, x_1)$ and $\mathcal{B} = [x_1, M]$. Note that the net price for marginal coverage is always positive for $x \in [0, x_0)$ and negative for $x \in [x_1, M]$. According to Lemma 3.1, the optimal indemnity function satisfies $I'(x) = 0$ for $x \in [0, x_0)$ and $I'(x) = 1$ for $x \in [x_1, M]$. When $x \in [x_0, x_1)$, the net price is non-negative. If the DM does not bother to pay the *ex ante* premiums for the losses for which she is indifferent between with and without insurance, then $I'(x) = 0$ for $x \in [x_0, x_1)$. In such a case, the optimal indemnity function is given by $I^*(x) = (x - x_1)_+$, which is a classical stop-loss function. In the remainder of the section, our focus will be on the worst-case survival function, first using an analytical approach and then a numerical one.

4.1 Analytical results

Following Theorems 3.1 and 3.2, we derive the worst-case survival functions for the cases where $\beta = 0$ and $\beta > 0$.

Case 1. $\beta = 0$

In this case, using Eq. (3.11) we obtain $x_0 = F_{\mathbb{Q}}^{-1}\left(\frac{\theta}{1+\theta}\right)$.

- If $1-\alpha \le \frac{1}{1+\theta}$, then it is easy to get that $\mathcal{A}$ is either $\emptyset$ or a singleton set $\{x_0\}$. In this case, we can get that: for any $x \in [0, x_0)$

$$S_{\mathbb{Q}}(x) \ge S_{\mathbb{Q}}(x_0) = \frac{1}{1+\theta} \ge 1-\alpha = g^{-1}(1) = t_0.$$

Thus, the worst-case survival function is given by

$$S_{P^*}(x) = \big(t_0 \vee S_{\mathbb{Q}}(x)\big)\,\mathbb{1}_{[0,x_0)}(x) + S_{\mathbb{Q}}(x)\mathbb{1}_{[x_0,M]}(x) = S_{\mathbb{Q}}(x).$$

- If $\frac{1-\gamma}{r_1} < \frac{1}{1+\theta} < 1-\alpha$, then we get

$$\mathcal{A} = \big\{x \in [x_0, M] : (1+\theta)S_{\mathbb{Q}}(x) \ge g(S_{\mathbb{Q}}(x))\big\} = [x_0, x_1]\,,$$

where $x_1 = F_{\mathbb{Q}}^{-1}\left(\frac{\alpha r_1 - \alpha + \theta\gamma - \theta\alpha}{(\gamma-\alpha)(1+\theta)-(1-r_1)}\right)$. In this case, the worst-case survival function is given by

$$S_{P^*}(x) = \big((1-\alpha) \vee S_{\mathbb{Q}}(x)\big)\,\mathbb{1}_{[0,x_0)}(x) + \left(\frac{\gamma-\alpha}{1-r_1}\big[(1+\theta)S_{\mathbb{Q}}(x) - r_1\big] + (1-\gamma)\right)\mathbb{1}_{[x_0,x_1)}(x)$$
$$+ S_{\mathbb{Q}}(x)\mathbb{1}_{[x_1,M]}(x).$$

- If $\frac{1}{1+\theta} \le \frac{1-\gamma}{r_1}$, then it is easy to get $\mathcal{A} = [x_0, M]$. Let $x_2 = F_{\mathbb{Q}}^{-1}(\gamma)$. If furthermore $\frac{1}{1+\theta} > 1-\gamma$, or equivalently $x_0 < x_2$, the worst-case survival function is then given by

$$S_{P^*}(x) = \big((1-\alpha) \vee S_{\mathbb{Q}}(x)\big)\,\mathbb{1}_{[0,x_0)}(x) + \left(\frac{\gamma-\alpha}{1-r_1}\big[(1+\theta)S_{\mathbb{Q}}(x) - r_1\big] + (1-\gamma)\right)\mathbb{1}_{[x_0,x_2)}(x)$$
$$+ \frac{(1-\gamma)(1+\theta)}{r_1}S_{\mathbb{Q}}(x)\mathbb{1}_{[x_2,M]}(x).$$

Otherwise, the worst-case survival function is given by

$$S_{P^*}(x) = \big((1-\alpha) \vee S_{\mathbb{Q}}(x)\big)\,\mathbb{1}_{[0,x_0)}(x) + \frac{(1-\gamma)(1+\theta)}{r_1}S_{\mathbb{Q}}(x)\mathbb{1}_{[x_0,M]}(x).$$

Case 2. $\beta > 0$

For this case, we first calculate $\hat{S}(x;\beta)$ by using Eq. (3.15). Note that if $t \in (1-\alpha, 1]$, then $\mathcal{H}'(t;x) = 0 - \beta\left(x - F_{\mathbb{Q}}^{-1}(1-t)\right)^{p-1}$, which results in $\hat{S}(x;\beta) = S_{\mathbb{Q}}(x)$. This further implies that $\hat{S}(x;\beta) = S_{\mathbb{Q}}(x)$ when $x \in [0, F_{\mathbb{Q}}^{-1}(\alpha))$. By combining the results for different values of t, we obtain

$$\hat{S}(x;\beta) = \begin{cases} S_{\mathbb{Q}}(x), & x \in [0, \tilde{x}_1), \\ 1-\alpha, & x \in [\tilde{x}_1, \tilde{x}_2), \\ S_{\mathbb{Q}}\left(x - \left(\frac{1-r_1}{\beta(\gamma-\alpha)}\right)^{\frac{1}{p-1}}\right), & x \in [\tilde{x}_2, \tilde{x}_3), \\ 1-\gamma, & x \in [\tilde{x}_3, \tilde{x}_4), \\ S_{\mathbb{Q}}\left(x - \left(\frac{r_1}{\beta(1-\gamma)}\right)^{\frac{1}{p-1}}\right), & x \in [\tilde{x}_4, M], \end{cases} \tag{4.2}$$

where

$$\tilde{x}_1 = F_{\mathbb{Q}}^{-1}(\alpha), \quad \tilde{x}_2 = \left(F_{\mathbb{Q}}^{-1}(\alpha) + \left(\frac{1-r_1}{\beta(\gamma-\alpha)} \right)^{\frac{1}{p-1}} \right) \wedge M,$$

$$\tilde{x}_3 = \left(F_{\mathbb{Q}}^{-1}(\gamma) + \left(\frac{1-r_1}{\beta(\gamma-\alpha)} \right)^{\frac{1}{p-1}} \right) \wedge M, \quad \tilde{x}_4 = \left(F_{\mathbb{Q}}^{-1}(\gamma) + \left(\frac{r_1}{\beta(1-\gamma)} \right)^{\frac{1}{p-1}} \right) \wedge M.$$

- If $1-\alpha \leq \frac{1}{1+\theta}$, then $x_0 = F_{\mathbb{Q}}^{-1}\left(\frac{\theta}{1+\theta}\right) \leq F_{\mathbb{Q}}^{-1}(\alpha) = \tilde{x}_1$. Similar to the first bullet point of Case 1, the set $\mathcal{A}$ is either empty or a singleton set. Thus, the worst-case survival function is given by

$$S_{P^*}(x;\beta) = \hat{S}(x;\beta)\mathbb{1}_{[0,x_0)}(x) + S_{\mathbb{Q}}(x)\mathbb{1}_{[x_0,M]}(x) = S_{\mathbb{Q}}(x).$$

  However, it should be noted that the worst-case survival function is equal to the benchmark survival function, indicating that this case can never happen if the constraint is to be binding for some $\epsilon > 0$.

- If $\frac{1-\gamma}{r_1} < \frac{1}{1+\theta} < 1-\alpha$, then $\mathcal{A} = [x_0, x_1]$, where x_1 has been defined in the second bullet point of Case 1. It is straightforward that $x_0 > \tilde{x}_1$ and $x_1 \leq F_{\mathbb{Q}}^{-1}(\gamma) < \tilde{x}_3$. Note that $\tilde{x}_2$ approaches to $\tilde{x}_1$ if $\beta \to \infty$ and to M if $\beta \to 0$. Therefore, the order of x_0, x_1, and $\tilde{x}_2$ must be considered to fully identify the worst-case survival function.

  - If $x_0 < x_1 \leq \tilde{x}_2$, the worst-case survival function is given by

$$\begin{aligned} S_{P^*}(x;\beta) =& S_{\mathbb{Q}}(x)\mathbb{1}_{[0,\tilde{x}_1)}(x) + (1-\alpha)\mathbb{1}_{[\tilde{x}_1,x_0)}(x) \\ &+ \left(\frac{\gamma-\alpha}{1-r_1}\left[(1+\theta)S_{\mathbb{Q}}(x) - r_1\right] + (1-\gamma) \right)\mathbb{1}_{[x_0,x_1)}(x) + S_{\mathbb{Q}}(x)\mathbb{1}_{[x_1,M]}(x). \end{aligned}$$

    As in the first bullet point, the worst-case survival function is equal to that when $\beta = 0$, suggesting that this case cannot occur if the constraint is to be binding for some $\epsilon > 0$.

  - If $x_0 \leq \tilde{x}_2 < x_1$, the worst-case survival function is given by

$$\begin{aligned} S_{P^*}(x;\beta) =& S_{\mathbb{Q}}(x)\mathbb{1}_{[0,\tilde{x}_1)}(x) + (1-\alpha)\mathbb{1}_{[\tilde{x}_1,x_0)}(x) \\ &+ \left(\frac{\gamma-\alpha}{1-r_1}\left[(1+\theta)S_{\mathbb{Q}}(x) - r_1\right] + (1-\gamma) \right)\mathbb{1}_{[x_0,\tilde{x}_2)}(x) \\ &+ \left[S_{\mathbb{Q}}\left(x - \left(\frac{1-r_1}{\beta(\gamma-\alpha)} \right)^{\frac{1}{p-1}} \right) \wedge \left(\frac{\gamma-\alpha}{1-r_1}\left[(1+\theta)S_{\mathbb{Q}}(x) - r_1\right] + (1-\gamma) \right) \right]\mathbb{1}_{[\tilde{x}_2,x_1)}(x) \\ &+ S_{\mathbb{Q}}(x)\mathbb{1}_{[x_1,M]}(x). \end{aligned}$$

- If $\tilde{x}_2 < x_0 < x_1$, the worst-case survival function is given by

$$
\begin{aligned}
S_{P^*}(x;\beta) =& S_{\mathbb{Q}}(x)\mathbb{1}_{[0,\tilde{x}_1)}(x) + (1-\alpha)\mathbb{1}_{[\tilde{x}_1,\tilde{x}_2)}(x) + S_{\mathbb{Q}}\left(x - \left(\frac{1-r_1}{\beta(\gamma-\alpha)}\right)^{\frac{1}{p-1}}\right)\mathbb{1}_{[\tilde{x}_2,x_0)}(x) \\
&+ \left[S_{\mathbb{Q}}\left(x - \left(\frac{1-r_1}{\beta(\gamma-\alpha)}\right)^{\frac{1}{p-1}}\right) \wedge \left(\frac{\gamma-\alpha}{1-r_1}\left[(1+\theta)S_{\mathbb{Q}}(x) - r_1\right] + (1-\gamma)\right)\right]\mathbb{1}_{[x_0,x_1)}(x) \\
&+ S_{\mathbb{Q}}(x)\mathbb{1}_{[x_1,M]}(x).
\end{aligned}
$$

- If $\frac{1}{1+\theta} \leq \frac{1-\gamma}{r_1}$, then $\mathcal{A} = [x_0, M]$. Let x_2 be defined in the third bullet point of Case 1, we easily get $x_2 < \tilde{x}_3$. In order to identify the worst-case survival function, the order of x_0, x_2 and $\tilde{x}_2$ needs to be discussed. Since a sufficiently large θ is required in this case, which is less realistic in practice, we will only examine one sub-case without further discussion.

  - If $x_0 < x_2 \leq \tilde{x}_2$, the worst-case survival function is given by

$$
\begin{aligned}
S_{P^*}(x;\beta) =& S_{\mathbb{Q}}(x)\mathbb{1}_{[0,\tilde{x}_1)}(x) + (1-\alpha)\mathbb{1}_{[\tilde{x}_1,x_0)}(x) \\
&+ \left(\frac{\gamma-\alpha}{1-r_1}\left[(1+\theta)S_{\mathbb{Q}}(x) - r_1\right] + (1-\gamma)\right)\mathbb{1}_{[x_0,x_2)}(x) \\
&+ \frac{(1-\gamma)(1+\theta)}{r_1}S_{\mathbb{Q}}(x)\mathbb{1}_{[x_2,\tilde{x}_2)}(x) \\
&+ \left(\frac{(1-\gamma)(1+\theta)}{r_1}S_{\mathbb{Q}}(x)\right) \wedge \hat{S}(x;\beta)\mathbb{1}_{[\tilde{x}_2,M]}(x),
\end{aligned}
$$

  where we use $\hat{S}(x;\beta)$ for $x \in [\tilde{x}_2, M]$ to shorten the lengthy expression.

Although the solution is complex, analytical forms of the worst-case survival function can be derived if the DM has a convex GlueVaR preference. These results can be easily extended to other distortion-risk-measure-based preferences, provided that the distortion function is concave and piece-wise linear. This perspective enables the approximation of the worst-case distribution for any convex distortion risk measure. A sensitivity analysis will be given in the next subsection.

4.2 Numerical results

The effect of the p of the p-Wasserstein distance on the worst-case survival function has been analytically studied in Section 3.2. In this section, we mainly focus on the effect of the safety loading factor θ, as well as the risk-aversion level of the DM, on the derived worst-case survival function and the resulting net price, i.e., $(1+\theta)S_{\mathbb{Q}}(x) - g(S_{P^*}(x))$.

For the effect of the safety loading factor, given the GlueVaR preference of the DM, the following settings are adopted.

- The loss variable X follows an exponential-type distribution, whose CDF is

$$
F_X(x) = \begin{cases} 1 - e^{-\frac{x}{100}}, & x \in [0, 5000], \\ 1, & x \in (5000, \infty). \end{cases}
$$

- The parameters of the GlueVaR preference are $r_1 = 0.6$, $r_2 = 1$, $\alpha = 0.05$ and $\gamma = 0.7$.

- We adopt a 2-Wasserstein distance throughout this section.

- To conduct the sensitivity test, we will test $\theta = 0.1, 0.5$ and 0.9.

For the ease of presentation, we also denote by S_1^*, S_2^* and S_3^* the worst-case survival function for the three cases of θ, and by X_1, X_2 and X_3 the random variables whose survival functions are S_1^*, S_2^* and S_3^*.

We choose $\epsilon^{\frac{1}{2}} = 10$ such that the case $\theta = 0.1$ has a slack constraint, while the cases $\theta = 0.5$ and $\theta = 0.9$ both have binding constraints. The left plot of Figure 2 shows the resulting worst-case survival functions for the three cases. The underlying intuition is straightforward: given a larger safety loading factor, the DM tends to retain more losses, which leaves itself a larger room for considering the worse distribution. Thus, the worst-case survival function under $\theta = 0.5$ or 0.9 dominates that under $\theta = 0.1$. Note that the constraint is binding for $\theta = 0.5$ or 0.9, hence S_3^* cannot dominate S_2^*. However, by computing the expectations of X_2 and X_3 we get

$$\int_0^{5000} S_2^*(x)dx = 106.59 < 107.78 = \int_0^{5000} S_3^*(x)dx.$$

As per Definition 2.2 and Theorem 2.3 of Cheung et al. (2015), we have

$$X_1 \leq_{st} X_2 \leq_{icx} X_3,$$

where $\leq_{st}$ and $\leq_{icx}$ denote respectively the first-order stochastic dominance and increasing convex order.

The right plot of Figure 2 shows the net price of the marginal coverage, which is defined in Section 3.1. The figure tells that a larger safety loading factor leads to higher net price, though the DM's deemed worst-case survival function becomes riskier, which therefore would reduce the DM's demand for insurance.

Next, we investigate the effect of the DM's risk-aversion level on the worst-case survival function and the net price. For that purpose, we fix the safety loading factor $\theta = 0.5$, and try $r_1 = 0.6, 0.7$ and 0.8. As shown by Figure 1, a larger r_1 leads to a more concave distortion function, which further leads to a more risk-averse DM. Under the same ϵ, the constraints for the three cases are all binding. Figure 3 shows the worst-case survival functions, as well as the net price for marginal coverage, under the different values of r_1. Due to the same reasoning, none of the survival functions dominate any of the others. By computing the expectations of X_1, X_2 and X_3, we find

$$\int_0^{5000} S_1^*(x)dx = 106.59 > \int_0^{5000} S_2^*(x)dx = 105.99 > \int_0^{5000} S_3^*(x)dx = 105.53,$$

which, as per Definition 2.2 and Theorem 2.3 of Cheung et al. (2015), indicates that $X_3 \leq_{icx} X_2 \leq_{icx} X_1$. Understandably, a more risk-averse DM, who applies a larger r_1, would transfer more tail risk to the insurer. This is also implied by the right plot of Figure 3, where the net price for marginal

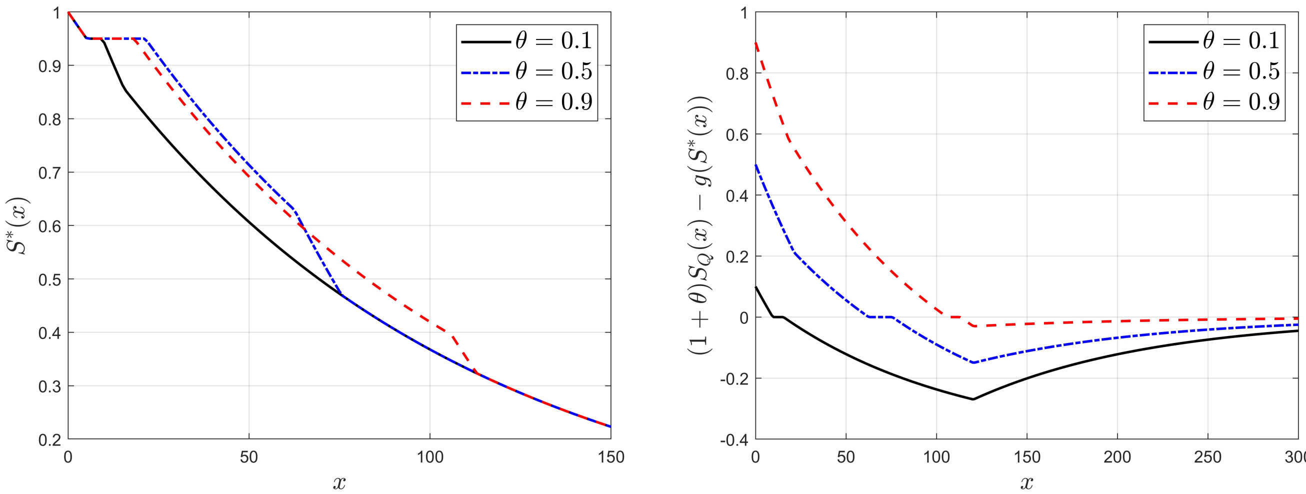

Figure 2: Under the 2-Wasserstein distance: (left) the worst-case survival functions under different θ; (right) the net price under different θ.

coverage is lower for a more risk-averse DM, which leads to a higher demand for insurance. Hence, there is no need for the DM to distort the distribution for the losses which have been ceded out. This explains the seemingly counter-intuitive observation where a more risk-averse DM over-weighs less the tail risk.

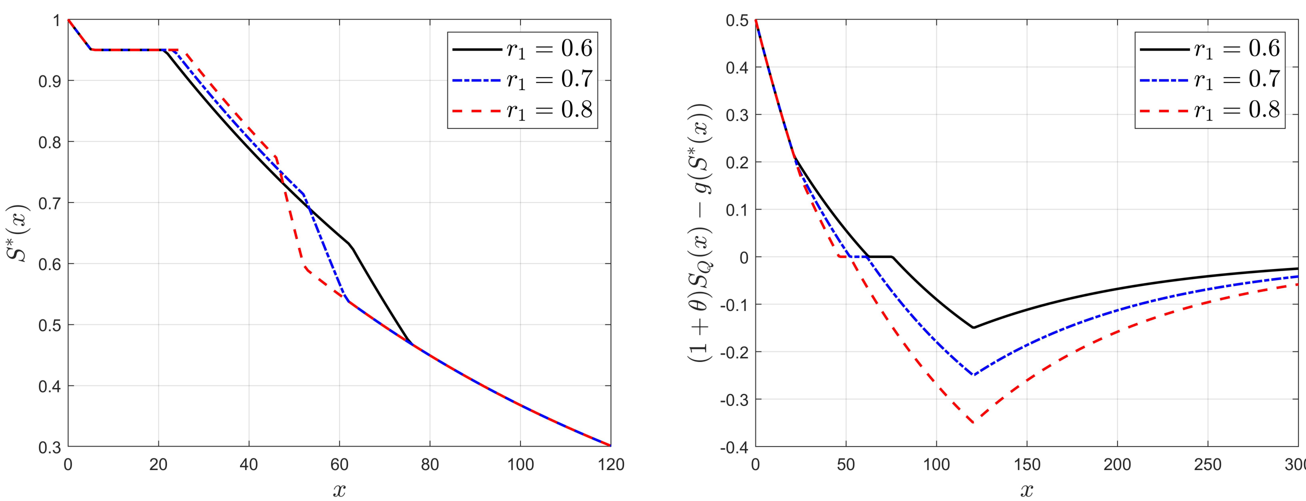

Figure 3: Under the 2-Wasserstein distance: (left) the worst-case survival functions under different r_1; (right) the net price under different r_1.

4.3 Comparison with the L^2 distance

In this section, we compare the 2-Wasserstein distance with the L^2 distance, which was studied in Boonen and Jiang (2024) (see also Appendix C), and show the worst-case survival functions and net prices when using the L^2 distance. We readily compute the following:

$$\max_{\theta\in\{0.1,0.5,0.9\},r_1\in\{0.6,0.7,0.8\}} \left(\int_0^M \left(\tilde{S}^*(x) - S_{\mathbb{Q}}(x)\right)^2 dx\right)^{\frac{1}{2}} = 2.5779 < 10,$$

where $\tilde{S}^*(x)$ is as shown in (3.12). Thus, for all the cases considered in this section the worst-case survival function under the L^2 distance is given by (3.12). As shown by Boonen and Jiang (2024) (see the derivation below Equation (5.1) therein), the L^p distance between two distribution functions is less than the p-Wasserstein distance, leaving the DM with more choices of the loss distribution if the same distance ϵ is used. The worst-case survival functions and net price under the L^2 distance are presented in Figures 4 (for various risk loadings θ) and 5 (for various parameters r_1). Comparing Figure 4 with Figure 2 (or Figure 5 with Figure 3), the DM chooses the riskier survival functions under the L^2 distance as compared with under the 2-Wasserstein distance. Straightforwardly, the net price for the marginal full coverage is lowered when the L^2 distance is applied. Nevertheless, as per the discussions at the beginning of Section 4, the optimal indemnity function remains the same when switching from the 2-Wasserstein distance to the L^2 distance if the DM buys zero insurance when she is indifferent between with and without insurance.

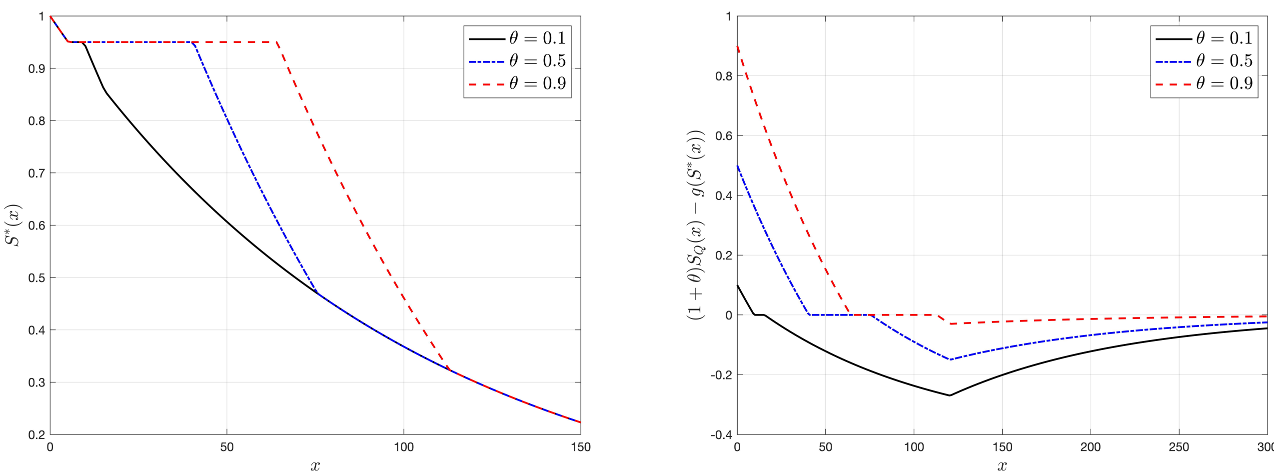

Figure 4: Under the L^2 distance: (left) the worst-case survival functions under different θ; (right) the net price under different θ.

5 Concluding remarks and future research

This study conducts an in-depth examination of optimal insurance contracting from a decision-maker's (DM) perspective, specifically focusing on instances where the DM is endowed with a convex distortion risk measure. This paper considers the case in which the DM has an ambiguous understanding about her loss distribution. We model the set of possible loss distributions utilizing the Wasserstein ball, which encompasses all distributions that are close to a given benchmark

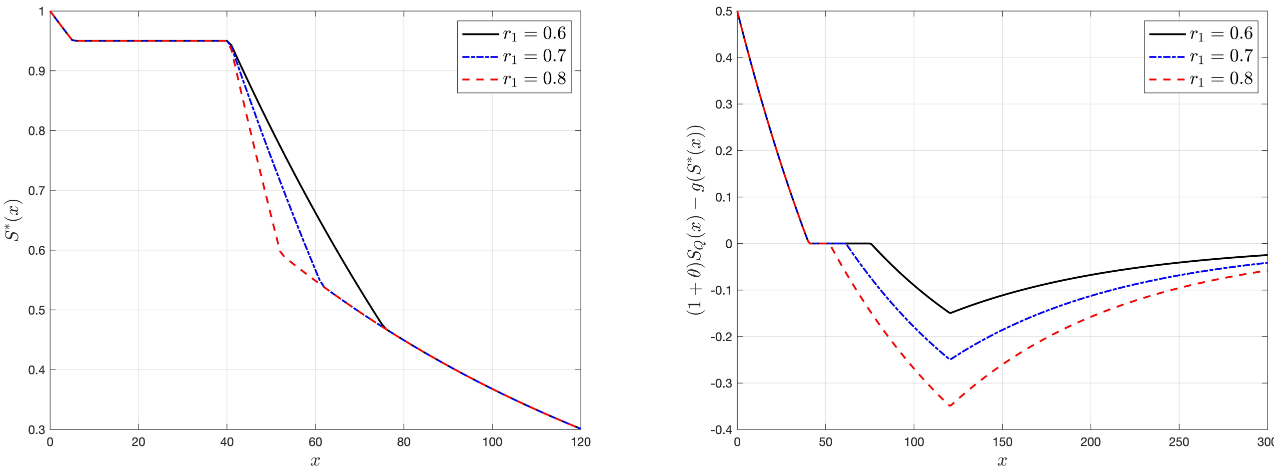

Figure 5: Under the L^2 distance: (left) the worst-case survival functions under different r_1; (right) the net price under different r_1.

distribution. To measure the distance, we study the p-Wasserstein distance. The optimal indemnity function is selected by considering the worst-case distribution from such a Wasserstein ball, which leads to the so-called distributionally robust insurance contract. For the pricing of the insurance contract, we assume that the insurer adopts the expected-value premium principle. By applying the well-known minimax theorem and the marginal indemnity function approach, the explicit structure of the optimal indemnity function can be determined. With the help of the newly developed survival-function-based representation of the Wasserstein distance, the explicit structure of the worst-case distribution is also worked out. It is further found that the DM would apply the benchmark distribution for the losses which have negative net price for marginal coverage under the benchmark distribution. To showcase the applicability of our main results, a concrete case is studied in detail, where the DM applies GlueVaR as her preference. Some numerical examples are also presented to demonstrate the effects of safety loading factor and level of risk aversion on the worst-case distribution.

Our research paves the way for numerous opportunities for further investigation. For example, within the same framework, novel techniques are needed to derive the optimal indemnity function and the worst-case distribution when the DM holds a non-convex distortion-risk-measure-based preference. Another potential extension is to incorporate the optimal effort (Robert and Therond, 2014) into our framework. If the DM can lower her ambiguity level through investing more resources in, for example, data collection, then another trade-off between the reduced ambiguity and the increased *ex ante* monetary loss should be considered within our framework. Such topics are left for future research.

Acknowledgements

The authors are indebted to two anonymous reviewers for their constructive comments, which helped improving this paper substantially. Wenjun Jiang acknowledges the financial support received from the Natural Sciences and Engineering Research Council of Canada (grant RGPIN-2020-04204), Alberta Innovates and the University of Calgary.

References

Albrecher, H., Teugels, J. L., and Beirlant, J. (2017). *Reinsurance: Actuarial and Statistical Aspects*. John Wiley & Sons.

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

Asimit, A. V., Bignozzi, V., Cheung, K. C., Hu, J., and Kim, E.-S. (2017). Robust and Pareto optimality of insurance contracts. *European Journal of Operational Research*, 262(2):720–732.

Assa, H. (2015). On optimal reinsurance policy with distortion risk measures and premiums. *Insurance: Mathematics and Economics*, 61:70–75.

Belles-Sampera, J., Guillén, M., and Santolino, M. (2014). Beyond value-at-risk: GlueVaR distortion risk measures. *Risk Analysis*, 34(1):121–134.

Bernard, C., Pesenti, S. M., and Vanduffel, S. (2023). Robust distortion risk measures. *Mathematical Finance*. Forthcoming.

Birghila, C., Boonen, T. J., and Ghossoub, M. (2023). Optimal insurance under maxmin expected utility. *Finance and Stochastics*, 27:467–501.

Birghila, C. and Pflug, G. C. (2019). Optimal xl-insurance under Wasserstein-type ambiguity. *Insurance: Mathematics and Economics*, 88:30–43.

Boonen, T. and Jiang, W. (2024). Robust insurance design with distortion risk measures. *European Journal of Operational Research*. Forthcoming.

Boonen, T. J. (2016). Optimal reinsurance with heterogeneous reference probabilities. *Risks*, 4(3):26.

Boonen, T. J. and Jiang, W. (2022). A marginal indemnity function approach to optimal reinsurance under the Vajda condition. *European Journal of Operational Research*, 303(2):928–944.

Boyd, S. P. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

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

Cai, J., Liu, F., and Yin, M. (2023). Worst-case risk measures of stop-loss and limited loss random variables under distribution uncertainty with applications to robust reinsurance. *Available at SSRN 4424759.*

Chakraborti, S., Jardim, F., and Epprecht, E. (2018). Higher-order moments using the survival function: The alternative expectation formula. *The American Statistician.*

Cheung, K. C. (2010). Optimal reinsurance revisited–a geometric approach. *ASTIN Bulletin*, 40(01):221–239.

Cheung, K. C., Chong, W. F., and Yam, S. C. P. (2015). Convex ordering for insurance preferences. *Insurance: Mathematics and Economics*, 64:409–416.

Chi, Y. and Tan, K. S. (2011). Optimal reinsurance under VaR and CVaR risk measures: a simplified approach. *ASTIN Bulletin: Journal of the IAA*, 41(2):487–509.

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

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2006). *Actuarial theory for dependent risks: measures, orders and models.* John Wiley & Sons.

Fan, K. (1953). Minimax theorems. *Proceedings of the National Academy of Sciences*, 39(1):42–47.

Huberman, G., Mayers, D., and Smith Jr, C. W. (1983). Optimal insurance policy indemnity schedules. *Bell Journal of Economics*, 14(2):415–426.

Jiang, W., Escobar-Anel, M., and Ren, J. (2020). Optimal insurance contracts under distortion risk measures with ambiguity aversion. *ASTIN Bulletin: The Journal of the IAA*, 50(2):619–646.

Liu, H. and Mao, T. (2022). Distributionally robust reinsurance with Value-at-Risk and Conditional Value-at-Risk. *Insurance: Mathematics and Economics*, 107:393–417.

Lo, A. (2017a). Functional generalizations of Hoeffding's covariance lemma and a formula for Kendall's tau. *Statistics & Probability Letters*, 122:218–226.

Lo, A. (2017b). A Neyman-Pearson perspective on optimal reinsurance with constraints. *ASTIN Bulletin: The Journal of the IAA*, 47(2):467–499.

Lv, W. and Wei, L. (2023). Distributionally robust reinsurance with Glue Value-at-Risk and expected value premium. *Mathematics*, 11(18):3923.

Panaretos, V. M. and Zemel, Y. (2019). Statistical aspects of wasserstein distances. *Annual Review of Statistics and its Application*, 6:405–431.

Pesenti, S. and Jaimungal, S. (2023). Portfolio optimisation within a Wasserstein ball. *SIAM Journal of Financial Mathematics*, 14(4):1175–1214.

Robert, C. Y. and Therond, P.-E. (2014). Distortion risk measures, ambiguity aversion and optimal effort. *ASTIN Bulletin: The Journal of the IAA*, 44(2):277–302.

Vajda, S. (1962). Minimum variance reinsurance. *ASTIN Bulletin*, 2(02):257–260.

Villani, C. (2009). *Optimal transport: Old and new*, volume 338. Springer.

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

Xie, X., Liu, H., Mao, T., and Zhu, X. B. (2023). Distributionally robust reinsurance with expectile. *ASTIN Bulletin: The Journal of the IAA*, 53(1):129–148.

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

Zhuang, S. C., Weng, C., Tan, K. S., and Assa, H. (2016). Marginal indemnification function formulation for optimal reinsurance. *Insurance: Mathematics and Economics*, 67:65–76.

A Minimax theorem

In this appendix, we state the minimax theorem. To do so, we first define a Hausdorff topological vector space. This is a topological vector space with the separation property, i.e., any two distinct points in the space can be separated by disjoint open sets.

Theorem A.1 (Minimax theorem, Fan (1953))**.** *Let Ξ_1 be a non-empty compact convex Hausdorff topological vector space and Ξ_2 be a convex set. If $\mathcal{H}$ is a real-valued function defined on $\Xi_1 \times \Xi_2$ such that*

- *$\xi_1 \mapsto \mathcal{H}(\xi_1, \xi_2)$ is convex and lower semi-continuous on Ξ_1 for each $\xi_2 \in \Xi_2$;*
- *$\xi_2 \mapsto \mathcal{H}(\xi_1, \xi_2)$ is concave on Ξ_2 for each $\xi_1 \in \Xi_1$,*

then

$$\inf_{\xi_1 \in \Xi_1} \sup_{\xi_2 \in \Xi_2} \mathcal{H}(\xi_1, \xi_2) = \sup_{\xi_2 \in \Xi_2} \inf_{\xi_1 \in \Xi_1} \mathcal{H}(\xi_1, \xi_2).$$

B Proofs of the main results

Proof of Lemma 3.3

Let U denote a uniform random variable on $(0,1)$, then $F_P^{-1}(U)$ is a random variable whose CDF is F_P. Recall the Hoeffding's formula on the covariance of two non-negative random variables X and

Y which are defined on the probability space $(\Omega, \mathcal{B}, \mathbb{P})$ (Lo, 2017a):

$$\begin{aligned}
\mathbf{C}ov[X,Y] &= \int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \left\{\mathbb{P}(X>x,Y>y) - \mathbb{P}(X>x)\mathbb{P}(Y>y)\right\} dxdy \\
&= \int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \mathbb{P}(X>x,Y>y)dxdy - \left(\int_{\mathbb{R}^+} \mathbb{P}(X>x)dx\right)\left(\int_{\mathbb{R}^+} \mathbb{P}(Y>y)dy\right) \\
&= \int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \mathbb{P}(X>x,Y>y)dxdy - \mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}$$

We therefore have

$$\mathbb{E}[XY] = \int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \mathbb{P}(X>x,Y>y)dxdy.$$

This leads to

$$\begin{aligned}
\int_0^1 (F_{P_1}^{-1}(t))^{k_1}(F_{P_2}^{-1}(t))^{k_2}dt =& \mathbb{E}[(F_{P_1}^{-1}(U))^{k_1}(F_{P_2}^{-1}(U))^{k_2}] \\
&= \int_0^{M^{k_2}}\int_0^{M^{k_1}} \mathbb{P}((F_{P_1}^{-1}(U))^{k_1} > x, (F_{P_2}^{-1}(U))^{k_2} > y)dxdy \\
&= \int_0^{M^{k_2}}\int_0^{M^{k_1}} \mathbb{P}(F_{P_1}^{-1}(U) > x^{\frac{1}{k_1}}, F_{P_2}^{-1}(U) > y^{\frac{1}{k_2}})dxdy \\
&= \int_0^{M}\int_0^{M} \mathbb{P}(F_{P_1}^{-1}(U) > s, F_{P_2}^{-1}(U) > z)ds^{k_1}dz^{k_2} \\
&= \int_0^{M}\int_0^{M} \left\{k_1 s^{k_1-1}k_2 z^{k_2-1}\mathbb{P}(F_{P_1}^{-1}(U) > s, F_{P_2}^{-1}(U) > z)\right\} dsdz \\
&= \int_0^{M}\int_0^{M} \left\{k_1 s^{k_1-1}k_2 z^{k_2-1}\mathbb{P}(U > F_{P_1}(s), U > F_{P_2}(z))\right\} dsdz \\
&= \int_0^{M}\int_0^{M} \left\{k_1 s^{k_1-1}k_2 z^{k_2-1}\mathbb{P}(U > F_{P_1}(s) \vee F_{P_2}(z)\right\} dsdz \\
&= \int_0^{M}\int_0^{M} \left\{k_1 s^{k_1-1}k_2 z^{k_2-1}\left(1 - F_{P_1}(s) \vee F_{P_2}(z)\right)\right\} dsdz \\
&= \int_0^{M}\int_0^{M} \left\{k_1 s^{k_1-1}k_2 z^{k_2-1}(1 - F_{P_1}(s)) \wedge (1 - F_{P_2}(z))\right\} dsdz \\
&= \int_0^{M}\int_0^{M} \left\{k_1 s^{k_1-1}k_2 z^{k_2-1}S_{P_1}(s) \wedge S_{P_2}(z)\right\} dsdz.
\end{aligned}$$

The proof is complete. □

Proof of Lemma 3.4

It suffices to prove the concavity of

$$\int_0^M\int_0^M \left\{(x-y)^{p-2}S_P(x) \wedge S_{\mathbb{B}}(y)\right\} dxdy$$

in S_P. This is obvious if $p \in \mathbb{Z}^+$ is an even number. Our focus will be on the case where $p \in \mathbb{Z}^+$ is an odd number.

Note that if $x \le y$, then for any $P \in \tilde{\mathcal{P}}_\epsilon$, we have $S_P(x) \ge S_{\mathbb{B}}(x) \ge S_{\mathbb{B}}(y)$. Thus, when $x \le y$, given any $\lambda \in (0,1)$ and $P_1, P_2 \in \tilde{\mathcal{P}}_\epsilon$, we have

$$\big(\lambda S_{P_1}(x) + (1-\lambda) S_{P_2}(x)\big) \ge S_{\mathbb{B}}(y),$$

which leads to

$$\big(\lambda S_{P_1}(x) + (1-\lambda) S_{P_2}(x)\big) \wedge S_{\mathbb{B}}(y) = S_{\mathbb{B}}(y) = \lambda S_{P_1}(x) \wedge S_{\mathbb{B}}(y) + (1-\lambda) S_{P_2}(x) \wedge S_{\mathbb{B}}(y).$$

Hence, for any $P_1, P_2 \in \tilde{\mathcal{P}}_\epsilon$ and $\lambda \in (0,1)$, we have

$$\begin{aligned}
&\int_0^M \int_0^M \left\{ (x-y)^{p-2} \big(\lambda S_{P_1}(x) + (1-\lambda) S_{P_2}(x)\big) \wedge S_{\mathbb{B}}(y) \right\} dxdy \\
= &\int_0^M \left\{ \int_0^y (x-y)^{p-2} \big(\lambda S_{P_1}(x) + (1-\lambda) S_{P_2}(x)\big) \wedge S_{\mathbb{B}}(y) dx \right\} dy \\
&+ \int_0^M \left\{ \int_y^M (x-y)^{p-2} \big(\lambda S_{P_1}(x) + (1-\lambda) S_{P_2}(x)\big) \wedge S_{\mathbb{B}}(y) dx \right\} dy \\
\ge &\int_0^M \left\{ \int_0^y (x-y)^{p-2} \big(\lambda S_{P_1}(x) \wedge S_{\mathbb{B}}(y) + (1-\lambda) S_{P_2}(x) \wedge S_{\mathbb{B}}(y)\big) dx \right\} dy \\
&+ \int_0^M \left\{ \int_y^M (x-y)^{p-2} \big(\lambda S_{P_1}(x) \wedge S_{\mathbb{B}}(y) + (1-\lambda) S_{P_2}(x) \wedge S_{\mathbb{B}}(y)\big) dx \right\} dy \\
= &\lambda \int_0^M \int_0^M \left\{ (x-y)^{p-2} S_{P_1}(x) \wedge S_{\mathbb{B}}(y) \right\} dxdy \\
&+ (1-\lambda) \int_0^M \int_0^M \left\{ (x-y)^{p-2} S_{P_2}(x) \wedge S_{\mathbb{B}}(y) \right\} dxdy.
\end{aligned}$$

This ends the proof. □

Proof of Theorem 3.2

Based on the definition of x_0, we have $g(S_P(x)) \le (1+\theta) S_{\mathbb{Q}}(x)$ on $[0, x_0)$. Thus, the objective function of (3.10) can be written as

$$\begin{aligned}
\sup_{S_P \in \mathcal{S}} &\int_0^{x_0} \left\{ g(S_P(x)) - \beta \left(x^{p-1} S_P(x) - \int_0^M (p-1)(x-y)^{p-2} S_P(x) \wedge S_{\mathbb{B}}(y) dy \right) \right\} dx \\
&+ \int_{x_0}^M \left\{ g(S_P(x)) \wedge (1+\theta) S_{\mathbb{Q}}(x) - \beta \left(x^{p-1} S_P(x) - \int_0^M (p-1)(x-y)^{p-2} S_P(x) \wedge S_{\mathbb{B}}(y) dy \right) \right\} dx.
\end{aligned} \tag{A.1}$$

The following functions are defined to facilitate the later discussions:

$$K_1(t;x) = g(t) - \beta\left(x^{p-1}t - \int_0^M (p-1)(x-y)^{p-2} t \wedge S_{\mathbb{B}}(y)dy\right),$$

$$K_2(t;x) = (1+\theta)S_{\mathbb{Q}}(x) - \beta\left(x^{p-1}t - \int_0^M (p-1)(x-y)^{p-2} t \wedge S_{\mathbb{B}}(y)dy\right),$$

$$K_3(t;x) = g(t) \wedge (1+\theta)S_{\mathbb{Q}}(x) - \beta\left(x^{p-1}t - \int_0^M (p-1)(x-y)^{p-2} t \wedge S_{\mathbb{B}}(y)dy\right).$$

We adopt the element-wise maximization approach. More specifically, we look at the following cases.

Case 1. $x \in [0, x_0)$

In this case, as per the element-wise maximization, we have

$$S_{P^*}(x) = \underset{t\in[0,1]}{\arg\max}\ K_1(t;x).$$

Since g is concave, it is straightforward that $K_1(t;x)$ is concave. Note that $S_{\mathbb{B}}(y) \le t \iff y \ge F_{\mathbb{B}}^{-1}(1-t)$. Thus,

$$K_1(t;x) = g(t) - \beta x^{p-1}t + \beta(p-1)\left(\int_{F_{\mathbb{B}}^{-1}(1-t)}^M (x-y)^{p-2} S_{\mathbb{B}}(y)dy + \int_0^{F_{\mathbb{B}}^{-1}(1-t)} (x-y)^{p-2} t dy\right),$$

of which the derivative exists almost everywhere and is decreasing over $[S_{\mathbb{B}}(x), 1]$. As such, $S_{P^*}(x)$ for $x \in [0, x_0)$ is given by (3.15). Furthermore, note that $F_{\mathbb{B}}(F_{\mathbb{B}}^{-1}(1-t)) \ge 1-t$ for any $t \in [0,1]$, and $(F_{\mathbb{B}}^{-1})'(1-t) = 0$ when $F_{\mathbb{B}}(F_{\mathbb{B}}^{-1}(1-t)) > 1-t$. It can easily be verified that

$$K_1'(t_0^+;x) = \lim_{t\to t_0^+} K_1'(t;x) = g'(t_0^+) - \beta(x - F_{\mathbb{B}}^{-1}(1-t_0^+))^{p-1}$$

for any $t_0 \in (0,1)$. If $0 \le x_1 \le x_2 < x_0$, we have $K_1'(t^+;x_1) \ge K_1'(t^+;x_2)$ for any $t \in [S_{\mathbb{B}}(x_1), 1]$. Thus, $S_{P^*}(x_1) \ge S_{P^*}(x_2)$.

Case 2. $x \in [x_0, M]$

In this case, as per the element-wise maximization, we have

$$S_{P^*}(x) = \underset{t\in[0,1]}{\arg\max}\ K_3(t;x).$$

It is easy to check that $K_3(t;x)$ is concave in t. Denote $t^*(x) = g^{-1}((1+\theta)S_{\mathbb{Q}}(x))$, which is between 0 and 1 for $x \in [x_0, M]$, then

$$K_3(t;x) = \begin{cases} K_1(t;x), & t \in [0, t^*(x)], \\ K_2(t;x), & t \in (t^*(x), 1]. \end{cases}$$

Then, we have

$$K_3'(t^*(x)^+;x) = g'(t^*(x)^+) - \beta(x - F_{\mathbb{B}}^{-1}(1-t^*(x)^+))^{p-1},$$
$$K_3'(t^*(x)^+;x) = -\beta(x - F_{\mathbb{B}}^{-1}(1-t^*(x)^+))^{p-1}.$$

On $\mathcal{A}$, we have

$$\begin{aligned} t^*(x) \geq S_{\mathbb{B}}(x) \Longrightarrow\ & F_{\mathbb{B}}(x) \geq 1 - t^*(x) \\ \Longrightarrow\ & x \geq F_{\mathbb{B}}^{-1}(F_{\mathbb{B}}(x)) \geq F_{\mathbb{B}}^{-1}(1-t^*(x)). \end{aligned}$$

Thus,

$$K_3'(t^*(x)^+;x) = -\beta(x - F_{\mathbb{B}}^{-1}(1-t^*(x)^+))^{p-1} \leq 0$$

Hence, on $\mathcal{A}$, the maximum of $K_3(t;x)$ over $[S_{\mathbb{B}}(x),1]$ is attained within $[S_{\mathbb{B}}(x),t^*(x)]$. Note that the maximum of $K_1(t;x)$ over $[S_{\mathbb{B}}(x),1]$ is attained when $t=\hat{S}(x;\beta)$, the worst-case survival function $S_{P^*}(x)=\hat{S}(x;\beta)\wedge t^*(x)$ for $x\in\mathcal{A}$.

Similarly, on $\mathcal{B}$, we have $t^*(x) < S_{\mathbb{B}}(x)$, which leads to $K_3(t;X)=K_2(t;x)$ for $t\in[S_{\mathbb{B}}(x),1]$. Since

$$K_2'(t^+;x) = -\beta(x - F_{\mathbb{B}}^{-1}(1-t^+)^{p-1} \leq 0$$

for any $t\in[S_{\mathbb{B}}(x),1]$, the maximum of $K_3(t;x)$ is attained at $S_{\mathbb{B}}(x)$.

The decreasing property of $S_{P^*}(x)$ over $[0,x_0)\cup\mathcal{A}\cup\mathcal{B}$ can be proved similarly as that in the proof of Theorem B.2 of Boonen and Jiang (2024). The existence of $\beta>0$ for (3.16) to hold is guaranteed by the Karush–Kuhn–Tucker conditions (see Chapter 5 in Boyd and Vandenberghe (2004)). Alternatively, one can refer to Boonen and Jiang (2024) for a proof by using the Lebesgue Dominated Theorem. □

C Results under the L^p distance

The distributionally robust optimal insurance problem with the L^1 or L^2 distance metric is studied by Boonen and Jiang (2024). In this appendix, we show that those results can be extended to general $p\in\mathbb{Z}^+$ using the techniques used to prove Theorems 3.1 and 3.2

We first recall the definition of the L^p distance, where $p\in\mathbb{Z}^+$, between two CDFs F_{P_1} and F_{P_2} for any $P_1,P_2\in\mathcal{P}$:

$$D_p(F_{P_1},F_{P_2}) = \left(\int_{\mathbb{R}} \left|F_{P_1}(x)-F_{P_2}(x)\right|^p dx\right)^{\frac{1}{p}} = \left(\int_{\mathbb{R}} \left|S_{P_1}(x)-S_{P_2}(x)\right|^p dx\right)^{\frac{1}{p}}. \tag{A.1}$$

By replacing the Wasserstein distance with $D_p(\cdot,\cdot)$ in $\mathcal{P}_\epsilon$, the problem (3.3) can be formulated as

$$\begin{cases} \sup\limits_{S_P\in\mathcal{S}} \displaystyle\int_0^M \left(g(S_P(x))\wedge(1+\theta)S_{\mathbb{Q}}(x)\right)dx, \\ \quad \text{s.t.} \displaystyle\int_0^M \left|S_P(x)-S_{\mathbb{B}}(x)\right|^p dx \leq \epsilon. \end{cases} \tag{A.2}$$

Due to the same reasoning, we note that Lemma 3.2 still holds for the problem (A.2). As such, to solve the problem (A.2), it suffices to restrict our attention to the set $\tilde{\mathcal{P}}_\epsilon$, which is defined in (3.4). In other words, the constraint of (A.2) can now be written as

$$\int_0^M \left(S_P(x) - S_{\mathbb{B}}(x)\right)^p dx \leq \epsilon \quad \text{and} \quad S_P(x) \geq S_{\mathbb{B}}(x) \; \forall \; x \in [0, M].$$

It is easy to verify that the integral $\int_0^M \left(S_P(x) - S_{\mathbb{B}}(x)\right)^p dx$ is convex in S_P for $P \in \tilde{\mathcal{P}}_\epsilon$. As such, the Slater condition holds for (A.2), which allows us to solve its dual problem to get the solution: for some $\beta \geq 0$,

$$\sup_{S_P \in \mathcal{S}} \int_0^M \left\{ g(S_P(x)) \wedge (1+\theta) S_{\mathbb{Q}}(x) - \beta \left(S_P(x) - S_{\mathbb{B}}(x)\right)^p \right\} dx. \tag{A.3}$$

The following two theorems are quite similar to Theorems 3.1 and 3.2, with the only difference that the distance metric is replaced by $D_p(\cdot, \cdot)$. Thus, we omit their proofs here.

Theorem C.1. *Let x_0 be defined in (3.11), and $\mathcal{A}$ and $\mathcal{B}$ be defined in Theorem 3.1. If*

$$\int_0^M \left(\tilde{S}^*(x) - S_{\mathbb{B}}(x)\right)^p dx \leq \epsilon,$$

where

$$\tilde{S}^*(x) = (t_0 \vee S_{\mathbb{B}}(x)) \mathbb{1}_{[0,x_0)}(x) + g^{-1}((1+\theta) S_{\mathbb{Q}}(x)) \mathbb{1}_{\mathcal{A}}(x) + S_{\mathbb{B}}(x) \mathbb{1}_{\mathcal{B}}(x),$$

where $t_0 = g^{-1}(1)$ is defined in Theorem 3.1, then the worst-case survival function that solves the problem (A.3) with $\beta = 0$ is $S_{P^*} = \tilde{S}^*$.

Theorem C.2. *Let x_0 be defined by (3.11) and $\mathcal{A}$ and $\mathcal{B}$ be defined by Theorem 3.1. If $\beta > 0$ in Problem (A.3), the worst-case survival function that solves Problem (A.3) is given by*

$$S_{P^*}(x;\beta) = \hat{S}(x;\beta) \mathbb{1}_{[0,x_0)}(x) + \left(\hat{S}(x;\beta) \wedge g^{-1}((1+\theta) S_{\mathbb{Q}}(x))\right) \mathbb{1}_{\mathcal{A}}(x) + S_{\mathbb{B}}(x) \mathbb{1}_{\mathcal{B}}(x), \tag{A.4}$$

where

$$\hat{S}(x;\beta) = \inf \left\{ t \in [S_{\mathbb{B}}(x), 1] : g'(t) - \beta p (t - S_{\mathbb{B}}(x))^{p-1} \leq 0 \right\}, \tag{A.5}$$

with $\inf \emptyset$ being the right-end point of the interval. Here β is such that

$$\int_0^M \left(S_{P^*}(x;\beta) - S_{\mathbb{B}}(x)\right)^p dx = \epsilon. \tag{A.6}$$