European Journal of Operational Research 316 (2024) 694–706

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/eor

Innovative applications of O.R.

Robust insurance design with distortion risk measures

Tim J. Boonen [a], Wenjun Jiang [b,*]

[a] Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China
[b] Department of Mathematics and Statistics, University of Calgary, Calgary, AB, T2N 1N4, Canada

ARTICLE INFO

Keywords:
Risk management
Optimal insurance
Distortion risk measure
L^2 distance
Tail Value-at-Risk

ABSTRACT

This paper studies the optimal insurance problem within the risk minimization framework and from a policyholder's perspective. We assume that the decision maker (DM) is uncertain about the underlying distribution of her loss and considers all the distributions that are close to a given (benchmark) distribution, where the "closeness" is measured by the L^2 or L^1 distance. Under the expected-value premium principle, the DM picks the indemnity function that minimizes her risk exposure under the worst-case loss distribution. By assuming that the DM's preferences are given by a convex distortion risk measure, we disentangle the structures of the optimal indemnity function and worst-case loss distribution in an analytical way, and provide the explicit forms for both of them under specific distortion risk measures. We also compare the results under the L^2 distance and the first-order Wasserstein (L^1) distance. Some numerical examples are presented at the end to illustrate the implications of our main results.

1. Introduction

Insurance can help individuals to be resilient to adverse events in their life. The study of optimal insurance contracting has a long history. The contract is characterized by an indemnity function, which maps the decision maker's (DM, also called policyholder) loss to indemnity, and a premium, which is paid upfront to the insurer. The premium is commonly modeled via a premium principle: a functional applied to the indemnity function. Then, the study of optimal insurance contracting leads to the problem of deriving the optimal indemnity function. Classical examples of the indemnity function include the quota-share and stop-loss functions, both of which are commonly used in practice. The popularity of these two functions is attributed not only to their simple forms but also to the theoretical foundations provided by Borch (1960) and Arrow (1974). We refer the interested readers to Albrecher et al. (2017) for a comprehensive review of some recent advances in insurance contracting.

Among the myriad literature, one main theme in optimal insurance contracting is the minimization of the DM's end-of-period risk exposure, which is measured by a risk measure. Risk measures such as Value-at-Risk (VaR), Tail Value-at-Risk (TVaR), or more general distortion risk measures possess a variety of mathematical properties and economic interpretations. The study of insurance contracting via risk minimization was conducted by Cai and Tan (2007), who derived the optimal retention point for a stop-loss function under the VaR and TVaR risk measures. Chi and Tan (2011) study the optimal insurance contracting under VaR and TVaR when restricting the indemnity function to different classes and show that the stop-loss function is always optimal if TVaR is used as the risk measure. By introducing the *marginal indemnity function* (MIF), Assa (2015) studies the optimal insurance contracting under the distortion risk measure and show that the optimal indemnity function is of the layered form. Cheung et al. (2019) extend the result of Assa (2015) to the case where the DM aims to minimize a generic law-invariant coherent risk measure of her net risk exposure and characterize the solution again by using the MIF. We refer interested readers to Cai and Chi (2020) for a review of some recent developments in optimal insurance contracting based on risk measures in static models.

An important assumption underlying many classical insurance models is that the DM has perfect knowledge of the underlying distribution of her losses. A relaxation of this assumption is usually obtained by seeking a robust solution to the classical insurance models by incorporating distributional uncertainty. For example, Balbás et al. (2015) and Asimit et al. (2017) are among the first ones to derive optimal insurance contracts under distributional uncertainty, and their focus is on robust notions of expectations, VaR and TVaR risk measures.[1] The literature on this topic can be categorized into several streams depending on the way of modeling the distributional uncertainty and uncertainty aversion. For instance, to model the distributional uncertainty, one

* Corresponding author.
E-mail addresses: tjboonen@hku.hk (T.J. Boonen), wenjun.jiang@ucalgary.ca (W. Jiang).

[1] Moreover, Balbás et al. (2011) study robust optimal insurance contracts in a setting where robustness is understood as stability of a solution with respect to several risk measures.

https://doi.org/10.1016/j.ejor.2024.02.002
Received 23 August 2023; Accepted 5 February 2024
Available online 8 February 2024

can either consider finitely many given candidate distributions (Asimit et al., 2017; Asimit & Boonen, 2018; Asimit et al., 2019) or infinitely many candidate distributions which are in the same parametric family or with the common statistical characteristics (Liu & Mao, 2022) or close enough to a given benchmark distribution under some distance metrics (Balbás et al., 2015; Birghila & Pflug, 2019; Pflug et al., 2017). To model the uncertainty aversion, a very popular model is the maxmin model (Birghila et al., 2023; Gilboa & Schmeidler, 1989) in which decisions are based on the worst-case distribution, while some alternative preferences such as the α-maxmin model (Ghirardato et al., 2004) and the smooth-ambiguity model (Jiang et al., 2020; Klibanoff et al., 2005) are also gaining popularity.

We assume that the DM considers loss distributions closely around a given benchmark distribution, where the "closeness" is measured by the L^2 or L^1 distance metric. The use of the L^p norm in modeling distributional uncertainty in finance and insurance can be found in, for example, Rachev et al. (2008), López-Díaz et al. (2012), Yang et al. (2014), and Bernard, De Vecchi, and Vanduffel (2022). It is well-known that the L^1 distance is equivalent to the first-order Wasserstein distance, which also has been gaining increasing popularity in finance and insurance. Interested readers are also referred to Pesenti and Jaimungal (2023) and Bernard, Pesenti, and Vanduffel (2023) for the further motivations of using such metrics in insurance. In this paper, we study the optimal insurance contract with a distortion risk measure subject to distributional uncertainty. The papers most closely related to our work are Birghila and Pflug (2019) and Liu and Mao (2022). Under the first-order Wasserstein distance, Birghila and Pflug (2019) study the optimal insurance contracting under Wang's premium principle and TVaR, and propose a numerical algorithm to approximate the worst-case distribution. Liu and Mao (2022) focus on the feasible set of indemnities that includes only the stop-loss functions. The question of identifying analytical forms of the optimal indemnity function and the worst-case distribution is still quite open. This paper fills this gap by identifying the analytical form of the optimal indemnity function without assuming its functional form under the maxmin model where the DM's preferences are modeled by some convex distortion risk measure. This is different from Liu and Mao (2022), in which the authors restrict the insurance indemnity functions to stop-loss functions only. In contrast to Birghila and Pflug (2019), we identify the worst-case distribution in an analytical way (see Section 5). Another related study is Bernard, Pesenti, and Vanduffel (2023), in which the authors only focus on the distribution that leads to the worst-case distortion risk measure. In contrast, our focus is on an optimal insurance problem in which the objective is to find both the optimal indemnity function and the worst-case loss distribution.

The remainder of this paper is structured as follows. Section 2 gives some preliminaries about the distortion risk measure and sets up the main problem. Section 3 solves the main problem by showing the structures of the optimal indemnity function and worst-case distribution. Section 4 presents some concrete examples illustrating the application and implication of our main results. Section 5 solves the main problem by changing the L^2 distance to L^1 distance. Section 6 presents some numerical examples. Section 7 concludes the paper, and all the proofs are delegated to the appendix.

2. Preliminaries and problem formulation

2.1. Distortion risk measure

Let there be a one-period economy. A DM is facing an insurable, non-negative loss represented by a random variable X. We fix the corresponding probability space $(\Omega, \mathcal{G}, \mathbb{P})$, where Ω is assumed to be atom-less and $\mathcal{G}$ is the σ-algebra generated by X. Let $\mathcal{P}([0, M])$ be the collection of probability measures on the measurable space $(\Omega, \mathcal{G})$ with support being a subset of $[0, M]$, where we assume $M < \infty$ is fixed.

The distortion risk measure stems from the dual utility theory of Yaari (1987), and is popular in decision theory and risk management. As an alternative to the expected utility theory in behavioral economics, dual utility theory describes people's behavior through a modification of the independence axiom. Distortion risk measures can also be interpreted as a risk-neutral evaluation under distorted beliefs, and in those distorted beliefs the more extreme events get a larger weight.

The distortion risk measure of random variable Z on the measurable space $(\Omega, \mathcal{G})$ 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[2] and concave over its domain $[0, 1]$ and satisfies $g(0) = 0$ and $g(1) = 1$. Eq. (2.1) shows that the distortion risk measure can be understood as the expectation of Z under a distorted probability measure (Balbás et al., 2009). For fixed probability measure P, the distortion risk measure ρ_g^P satisfies the following properties (Denuit et al., 2006; Wang et al., 1997):

- Comonotonic additivity: $\rho_g^P(Z + Y) = \rho_g^P(Z) + \rho_g^P(Y)$ for comonotonic random variables Z and Y.[3]
- Sub-additivity: $\rho_g^P(Z + Y) \le \rho_g^P(Z) + \rho_g^P(Y)$ for any two random variables Z and Y.

Note that Comonotonic additivity and the fact that $\rho_g^P(1) = 1$ imply Translation invariance: $\rho_g^P(Z + c) = \rho_g^P(Z) + c$ for all $c \in \mathbb{R}$. Moreover, distortion risk measures are coherent in the sense of Artzner et al. (1999), which can be written as spectral risk measures (Balbás et al., 2009), and are averse to mean-preserving spreads (Yaari, 1987).

The class of distortion risk measures is quite large and includes TVaR, which will be discussed further in Section 4.2.

2.2. Problem formulation

Throughout the paper, we use the notation $x \wedge y = \min\{x, y\}$, $x \vee y = \max\{x, y\}$, and $(x)_+ = \max\{x, 0\}$.

Suppose that the DM is interested in purchasing an insurance contract $(I, \pi(I))$ where I is the indemnity function and $\pi(I)$ is the corresponding premium. After buying the insurance contract, the DM's end-of-period loss becomes $X - I(X) + \pi(I)$.

Let $F^P(x)$ be the Cumulative Distribution Function (CDF) of X under probability measure $P \in \mathcal{P}([0, M])$. The survival function is defined as $S^P(x) := 1 - F^P(x) = P(X > x)$. 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. Moreover, the insurer uses a benchmark distribution $F^{\mathbb{Q}}$, where $\mathbb{Q} \in \mathcal{P}([0, M])$, to price insurance, and we assume that the premium $\pi(I)$ is given by the expected value principle:

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

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

For the indemnity function, we impose exogenously the so-called *incentive compatibility* condition on it. This condition is first proposed by Huberman et al. (1983) and requires that the losses borne by the DM and insurer are both increasing. This would reduce the DM's motivation of manipulating the losses and thus alleviate the *ex post* moral hazard issues. Under the incentive compatibility condition, the indemnity function must be in the following class:

$$\mathcal{I} = \left\{ I : \mathbb{R}^+ \mapsto \mathbb{R}^+ \;\middle|\; I(0) = 0,\ 0 \le I(x_2) - I(x_1) \le x_2 - x_1 \text{ for any } 0 \le x_1 \le x_2 \right\}.$$

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

[3] 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.

Notably, if $I \in \mathcal{I}$, then $X - I(X)$ and $I(X)$ are comonotonic. The class $\mathcal{I}$ is quite large and includes many well known indemnity functions, such as the stop-loss, quota-share and truncated stop-loss functions.[4] Additionally, if $I \in \mathcal{I}$, then I is 1-Lipschitz continuous and admits the following integral representation

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

where η is called the 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 is known, the following problem is faced by the DM, which has been extensively studied in the literature[5]:

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

Now given that the DM is uncertain about the distribution of X, it aims to minimize the worst-case risk measure, i.e. the largest one among the risks resulting from all the possible distributions. The DM will consider a set of distributions around the benchmark distribution $F^{\mathbb{Q}}$ with $\mathbb{Q} \in \mathcal{P}([0, M])$. In this paper, we use the L^p norm to measure the distance between the candidate distribution and the benchmark distribution. It has also been found that the L^1 distance is equivalent to the first-order Wasserstein distance (Panaretos & Zemel, 2019), and as explained in Pesenti and Jaimungal (2023), an important reason for using such metric in uncertainty modeling is that it allows comparison between distributions with differing supports. We will only apply the L^2 and L^1 distances in this paper, while mainly focusing on the L^2-distance-based problem as the L^1-distance-based problem can be solved in a similar way. The following definition is for a general L^p distance between two cumulative distribution functions.

Definition 2.1. The L^p distance between $F^P, F^{\mathbb{Q}}$ with $\{P, \mathbb{Q}\} \subset \mathcal{P}([0, M])$ is given by

$$D_p(F^P, F^{\mathbb{Q}}) = \left(\int_{-\infty}^{\infty} (F^P(x) - F^{\mathbb{Q}}(x))^p dx \right)^{\frac{1}{p}}.$$

With the above definition, the first uncertainty set of the loss distribution is described as a ball centered on the benchmark distribution $F^{\mathbb{Q}}$ under the L^2 distance:

$$\mathcal{P}_\epsilon := \{P \in \mathcal{P}([0, M]) \mid D_2(F^P, F^{\mathbb{Q}}) \le \sqrt{\epsilon}\},$$

where $\epsilon \ge 0$. If $\epsilon = 0$, then the set $\mathcal{P}_0$ is a singleton, and given by $\mathcal{P}_0 = \{\mathbb{Q}\}$; this yields the case without distributional uncertainty. This case is addressed by Cui et al. (2013) and Assa (2015). Hence, we focus on the case where $\epsilon > 0$ in the rest of this paper. Since $M < \infty$, it holds that $\mathbb{E}^{\mathbb{Q}}[X] < \infty$ and $\sup_{P \in \mathcal{P}_1} \rho_g^P(X) < \infty$.

We next present the main problem that we study in this paper.

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

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

It is worth pointing out that our Problem 1 is different from Problem (P$_4$) of Birghila and Pflug (2019), where their uncertainty set is structured under the Wasserstein distance. Nevertheless, the existence of the solution to our Problem 1 can be proved analogously as their Proposition 4.2. In the next section, we provide the analytical solution to Problem 1.

Remark 2.1. Generally, the selection of ϵ is subjective and depends on the DM's information set or ambiguity-aversion level. Intuitively, a more ambiguity-averse DM will apply a smaller ϵ in her model. As will be shown in Section 3.2, ϵ is negatively related with the Lagrangian parameter β in (3.5), where β can be understood as a penalty parameter as in Uppal and Wang (2003). This penalty parameter penalizes distribution functions that are further away from the benchmark distribution $F^{\mathbb{Q}}$, and this penalty is high for a more ambiguity-averse DM.

Statistically, the selection of ϵ can depend on the collected data. Note that the L^2 distance between an empirical CDF F_n and a hypothesized CDF F bears much similarity with the quadratic empirical distribution function test statistic (Stephens, 2017):

$$Q = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \psi(x) dF(x),$$

where n is the size of dataset, and $\psi(x)$ is a function that assigns weights to the squared difference $(F_n(x) - F(x))^2$. If F is smooth enough, which is indeed the case for many parametric loss distributions introduced in Klugman et al. (2012), then taking $\psi(x) = \frac{1}{f(x)}$, where $f(x) = F'(x)$, yields

$$\int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dx = \frac{Q}{n}.$$

While generally the explicit distribution for Q is not available in the literature, one can still apply bootstrapping to get the confidence interval for Q, for which the bounds can be used to determine the value of ϵ. This is not further pursued in this paper.

3. The optimal indemnity function and the worst-case distribution

3.1. The structure of indemnity function

In this section, we derive the structure of the optimal indemnity function. We start by stating the well-known minimax theorem (Fan, 1953).

Theorem 3.1 (*Minimax Theorem*)**.** *Let Ξ_1 be a non-empty compact convex Hausdorff topological vector space*[6] *and let 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

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

The Minimax theorem states that under certain conditions, the infimum of the supremum of a real-valued function defined on a product of two sets is equal to the supremum of the infimum of the function. The class $\mathcal{I}$ is convex. Since $M < \infty$, applying Arzelà–Ascoli Theorem leads to the compactness of $\mathcal{I}$. 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. 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, the conditions for applying Theorem 3.1 are all met in our setting. By exchanging the "inf" and "sup" in Problem 1, we 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;

[4] The stop-loss function is given by $I(x) = (x - d)_+$ for some $d \ge 0$. 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 $d_1, d_2 \ge 0$.

[5] If $P = \mathbb{Q}$, we here refer to Assa (2015), Zhuang et al. (2016) and Lo (2017), and for generic $P \in \mathcal{P}([0, M])$ we refer to Boonen (2016).

[6] A Hausdorff topological vector space 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.

Cheung & Lo, 2017; Lo, 2017). We present its solution below. Here, $\mathbb{1}_A(x)$ is the indicator function, that is equal to 1 if $x \in A$ and 0 otherwise. Moreover, we recall that S^P and $S^{\mathbb{Q}}$ are the survival functions under the probability measures P and $\mathbb{Q}$, respectively.

Lemma 3.1. *For a fixed $P \in \mathcal{P}([0,M])$, optimal indemnity functions to the inner problem of (3.1) are given by $I^*(x;P) = \int_0^x \eta^*(t;P)dt$ with*

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

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

As shown in Lemma 3.1, the optimal indemnity function that solves the inner problem of (3.1) is not unique due to the non-uniqueness of $\gamma(t)$ for t such that $(1+\theta)S^{\mathbb{Q}}(t) = g\left(S^P(t)\right)$. 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, as it represents the difference between the cost of coverage under the insurer's probability measure and the DM's valuation of the coverage. Lemma 3.1 indicates that the DM will purchase the largest 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

Note that

$$D_2(F^P,F^{\mathbb{Q}}) = \left(\int_0^M (F^P(x)-F^{\mathbb{Q}}(x))^2dx\right)^{\frac{1}{2}} = \left(\int_0^M (S^P(x)-S^{\mathbb{Q}}(x))^2dx\right)^{\frac{1}{2}}.$$

With the indemnity function $I^*(x;P)$ given in Lemma 3.1, we can write Problem 1 as

$$\begin{cases} \sup\limits_{P\in\mathcal{P}([0,M])} \rho_g^P(X - I^*(X;P) + \pi(I^*)), \\ \quad\text{s.t.} \int_0^M (S^P(x)-S^{\mathbb{Q}}(x))^2dx \le \epsilon, \quad \epsilon>0. \end{cases} \tag{3.2}$$

The objective function of (3.2) could be further written as

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

Problem (3.2) can thus be written as

$$\begin{cases} \sup\limits_{P\in\mathcal{P}([0,M])} \int_0^M \left(g(S^P(x))\wedge(1+\theta)S^{\mathbb{Q}}(x)\right)dx, \\ \quad\text{s.t.} \int_0^M (S^P(x)-S^{\mathbb{Q}}(x))^2dx \le \epsilon, \quad \epsilon>0. \end{cases} \tag{3.3}$$

Since Ω is atom-less and the $\mathcal{G}$ is generated by X, we can transform the objective to find the worst-case survival distribution S^P of X that solves (3.3). That is, for the class of survival functions of X given by $\mathcal{S}_X := \{S^P : P\in\mathcal{P}([0,M])\}$, we write

$$\begin{cases} \sup\limits_{S^P\in\mathcal{S}_X} \int_0^M \left(g(S^P(x))\wedge(1+\theta)S^{\mathbb{Q}}(x)\right)dx, \\ \quad\text{s.t.} \int_0^M (S^P(x)-S^{\mathbb{Q}}(x))^2dx \le \epsilon, \quad \epsilon>0. \end{cases} \tag{3.4}$$

We next provide some observations for the worst-case survival function $S^*(x)$, where

$$\begin{aligned} S^* = arg\, sup_{S^P\in\mathcal{S}_X} &\int_0^M \left(g(S^P(x))\wedge(1+\theta)S^{\mathbb{Q}}(x)\right)dx, \\ \text{s.t.} &\int_0^M (S^P(x)-S^{\mathbb{Q}}(x))^2dx \le \epsilon, \quad \epsilon>0. \end{aligned}$$

- Note that $S^P(x) = S^{\mathbb{Q}}(x)$ is always feasible in Problem (3.4) and that $S^*(x)$ should be as large as possible in order to maximize the objective function. The constraint provides a bound on the L^2 distance between P and $\mathbb{Q}$, and it is straightforward to see that the worst-case survival function should satisfy $S^*(x)\ge S^{\mathbb{Q}}(x)$ for all $x\ge 0$. In other words, the worst-case risk faced by the DM is larger than the risk faced by the insurer in the sense of the first-order stochastic dominance.
- For fixed x, the objective function $g(S^P(x))\wedge(1+\theta)S^{\mathbb{Q}}(x)$ is concave in $S^P(x)$. Furthermore, the left-side of constraint in (3.4) is a convex function of S^P, and for $\epsilon>0$ there exists at least one $S^P\in\mathcal{S}_X$ that is strictly feasible to the problem (3.4): for instance, $S^P(x) = S^{\mathbb{Q}}(x)$ for $x\in[0,\infty)$. This is Slater's condition, and thus solving Problem (3.4) is equivalent to solving the following problem:

$$\sup_{S^P\in\mathcal{S}_X} \int_0^M \left(g(S^P(x))\wedge(1+\theta)S^{\mathbb{Q}}(x) - \beta(S^P(x)-S^{\mathbb{Q}}(x))^2\right)dx, \tag{3.5}$$

  for some $\beta\ge 0$ (i.e., strong duality holds).
- It is sufficient to solve Problem (3.5) for generic $S^P(x)$ such that $S^P(x)\in[S^{\mathbb{Q}}(x),1]$ for any fixed value of $x\ge 0$, and then check thereafter that S^P is indeed a survival function; that is, it is decreasing, right-continuous, and such that $S^P(0)=1$ and $\lim_{x\to\infty} S^P(x)=0$. This is the technique that we will use to solve Problem (3.5).

We treat the cases when the L^2 constraint is slack and when the constraint is binding separately. For the case when the Lagrangian parameter $\beta=0$ in Problem (3.5), we only need to consider the problem

$$\sup_{S^P\in\mathcal{S}_X} \int_0^M \left(g(S^P(x))\wedge(1+\theta)S^{\mathbb{Q}}(x)\right)dx. \tag{3.6}$$

For some distortion functions, the above problem might not give a unique solution. In that case, we need to find the solution that yields the threshold of ϵ such that $\beta=0$ if ϵ is greater than that threshold. In the next theorem, we derive a worst-case survival function and this threshold for the case when $\beta=0$.

Theorem 3.2. *Let*

$$x_0 := \sup\{x\in[0,M) : (1+\theta)S^{\mathbb{Q}}(x)\ge 1\},$$
$$x_1 := \sup\{x\in[x_0,M) : (1+\theta)S^{\mathbb{Q}}(x)\ge g(S^{\mathbb{Q}}(x))\},$$

and $t_0 := g^{-1}(1)$, where $g^{-1}(t) = \inf\{x\in[0,1] : g(x)\ge t\}$. If $\epsilon\ge\int_0^M(\tilde{S}^*(x)-S^{\mathbb{Q}}(x))^2dx$ with

$$\tilde{S}^*(x) = \left(t_0\vee S^{\mathbb{Q}}(x)\right)\mathbb{1}_{[0,x_0)}(x) + g^{-1}((1+\theta)S^{\mathbb{Q}}(x))\mathbb{1}_{[x_0,x_1)}(x) + S^{\mathbb{Q}}(x)\mathbb{1}_{[x_1,M]}(x) \tag{3.7}$$

for all $x\ge 0$, then a worst-case survival function of X that solves (3.6) is given by $S^*=\tilde{S}^*$.

If $\beta=0$, the worst-case survival function may not be unique. As shown in the proof of Theorem 3.2, the survival function in (3.7) is the one that, among the solutions to Problem (3.6), minimizes $\int_0^M(S^P(x)-S^{\mathbb{Q}}(x))^2dx$.

Next, we derive the worst-case survival function of X when the Lagrangian parameter $\beta>0$ in Problem (3.5) (and thus the L^2 distance constraint becomes binding). For the following theorem, we remark that g is not assumed to be differentiable. Here, since g is monotone, and it is therefore differentiable almost everywhere by the Lebesgue Theorem for the differentiability of monotone functions.

Theorem 3.3. *Let x_0 and x_1 be as defined in Theorem 3.2. If $\beta > 0$ in Problem (3.5), the worst-case survival function of X is uniquely given by*

$$S^*(x;\beta) = \hat{S}(x;\beta)\mathbb{1}_{[0,x_0)}(x) + \left(\tilde{S}(x;\beta) \wedge g^{-1}((1+\theta)S^{\mathbb{Q}}(x))\right)\mathbb{1}_{[x_0,x_1)}(x) + S^{\mathbb{Q}}(x)\mathbb{1}_{[x_1,M]}(x), \tag{3.8}$$

where

$$\hat{S}(x;\beta) = \begin{cases} 1, & \text{if } g'(1^-) - 2\beta F^{\mathbb{Q}}(x) \geq 0, \\ \tilde{S}(x;\beta), & \text{if } g'(1^-) - 2\beta F^{\mathbb{Q}}(x) < 0, \end{cases} \tag{3.9}$$

and where $\tilde{S}(x;\beta)$ satisfies

$$g'(\tilde{S}(x;\beta)^-) - 2\beta(\tilde{S}(x;\beta) - S^{\mathbb{Q}}(x)) \geq 0, \quad g'(\tilde{S}(x;\beta)^+) - 2\beta(\tilde{S}(x;\beta) - S^{\mathbb{Q}}(x)) \leq 0. \tag{3.10}$$

Now, $S^*(x;\beta)$ solves Problem (3.4) where β is such that

$$\int_0^M (S^*(x;\beta) - S^{\mathbb{Q}}(x))^2 dx = \epsilon. \tag{3.11}$$

Furthermore, β is decreasing with respect to ϵ.

The result in Theorem 3.3 can be interpreted as follows. If the function $t \mapsto g(t) - (1+\theta)t$ has a positive root $\hat{t}$, then $g(t) \geq (1+\theta)t$ for $t \in [0, \hat{t}]$. As per the discussions at the beginning of this section, only the survival function $S^P(x) \geq S^{\mathbb{Q}}(x)$ for all $x \geq 0$ is of interest. Thus, for large x such that $S^{\mathbb{Q}}(x) \leq \hat{t}$, we have

$$g(S^P(x)) \geq g(S^{\mathbb{Q}}(x)) \geq (1+\theta)S^{\mathbb{Q}}(x).$$

Lemma 3.1 tells that the DM will purchase full coverage for that part of loss. Hence, the DM does not consider any worse distribution for the tail part. For x such that

$$(1+\theta)S^{\mathbb{Q}}(x) > g(S^P(x)) \geq g(S^{\mathbb{Q}}(x)),$$

the DM will retain these marginal losses. Thus, the DM would give more weight to the probabilities for the small- or medium-sized losses.

Appendix B studies the more general case where the DM has a benchmark distribution different from that under $\mathbb{Q}$. This can be interpreted as there being belief heterogeneity between the DM and insurer regarding the benchmark distribution (i.e., their respective baseline probability distributions for estimating risk). Belief heterogeneity in insurance without distributionally robust objectives has been studied in Boonen (2016), Chi (2019), Ghossoub (2017), Jiang et al. (2019), and we differ from these papers by considering a robust insurance problem formulation. By introducing belief heterogeneity, the worst-case survival functions are less straightforward to interpret than those in Theorems 3.2 and 3.3, and the resultant optimal insurance indemnities can have complex shapes. We provide Appendix B for completeness.

If we change the expectation premium principle in this paper to the distortion premium principle, i.e. $\pi(I) = (1+\theta)\rho^{\mathbb{Q}}_{\tilde{g}}(I(X))$, then

$$\begin{aligned} \rho^{\mathbb{Q}}_{\tilde{g}}(I(X)) &= \int_0^M \tilde{g}(S^{\mathbb{Q}}_{I(X)}(x))dx \\ &= \int_0^M I(x)d[1 - \tilde{g}(S^{\mathbb{Q}}_X(x))] \\ &= \mathbb{E}^{\hat{\mathbb{Q}}}[I(X)], \end{aligned}$$

where $F^{\hat{\mathbb{Q}}}(x) := 1 - \tilde{g}(S^{\mathbb{Q}}_X(x))$ does not depend on the function I (Boonen et al., 2021). Hence, adopting the distortion premium principle is equivalent to adopting another probability measure in the expectation premium principle, and this is a special case of our general setting in Appendix B.

Some specific examples illustrating the main results in this section will be presented in the next section.

4. Some concrete examples

4.1. A twice differentiable distortion function

To further compare the worst-case distribution with the benchmark distribution, we adopt the following additional assumption.

Assumption 1. The following two conditions are satisfied simultaneously:

- The distortion function g is twice differentiable, and
- there exist probability density functions for both the worst-case distribution and benchmark distribution.

We define the following two points:

$$x_0' := \inf\{x \in [0, x_0) : 2\beta F^{\mathbb{Q}}(x) > g'(1^-)\}, \tag{4.1}$$

$$x_1' := \inf\{x \in [x_0, x_1) : g'(g^{-1}((1+\theta)S^{\mathbb{Q}}(x))) \geq 2\beta(g^{-1}((1+\theta)S^{\mathbb{Q}}(x)) - S^{\mathbb{Q}}(x))\}, \tag{4.2}$$

with $\inf \emptyset$ being the left endpoint of the interval. The following proposition provides a description of the worst-case distribution under the above assumption.

Proposition 4.1. *Let Assumption 1 hold. For the worst-case survival function $S^*(\cdot;\beta)$, with β as in Theorem 3.3, it holds that*

(i). $S^*(x_0';\beta) = 1$, and
(ii). $f^{\mathbb{Q}}(x) \geq f^*(x;\beta)$ if $x \in [0, x_1')$, and $f^{\mathbb{Q}}(x) \leq f^*(x;\beta)$ if $x \in [x_1', x_1)$, where $f^*(x;\beta) := -\frac{\partial S^*(x;\beta)}{\partial x}$.

Proposition 4.1 shows that there may exist singularities between the worst-case distribution and the benchmark distribution, and that the DM would assign less weight to the probabilities for small-sized losses while assigning more weight to the probabilities for medium-sized losses, compared with the benchmark distribution.

Under Assumption 1, we can revisit and determine the optimal indemnity function for Problem 1. The following results are from Section 3.2 and Lemma 3.1.

- From Proposition 4.1, we get that the DM assigns zero probability to the event $\{\omega \in \Omega : X(\omega) \in [0, x_0')\}$. Hence the DM demands no insurance for the losses in $[0, x_0')$.
- When $x \in [x_0', x_1')$, by the proof of Proposition 4.1, we get that $(1+\theta)S^{\mathbb{Q}}(x) > g(S^*(x;\beta))$. Therefore the DM demands no insurance for the losses in $[x_0', x_1')$.
- When $x \in [x_1', x_1)$, as per Theorem 3.3 and Proposition 4.1, $S^*(x;\beta) = g^{-1}((1+\theta)S^{\mathbb{Q}}(x))$, or equivalently, $(1+\theta)S^{\mathbb{Q}}(x) = g(S^*(x;\beta))$. Hence, the DM is indifferent between purchasing insurance or no insurance for the losses in $[x_1', x_1)$.
- When $x \in [x_1, M)$, we know from Theorem 3.3 that $S^*(x;\beta) = S^{\mathbb{Q}}(x)$ and $(1+\theta)S^{\mathbb{Q}}(x) < g(S^{\mathbb{Q}}(x))$. Hence, the DM demands full insurance for the losses in $[x_1, M)$.

The above findings lead to the following theorem, that provides the explicit indemnity function that solves Problem 1.

Theorem 4.1. *Under Assumption 1, the optimal indemnity function that solves Problem 1 is given by*

$$I^*(x) = \int_0^x \left\{\gamma(t)\mathbb{1}_{[x_1',x_1)}(t) + \mathbb{1}_{[x_1,\infty)}(t)\right\} dt, \tag{4.3}$$

where $\gamma(t) \in [0,1]$, x_1' is defined in Eq. (4.2) and x_1 is defined in Theorem 3.3.

Theorem 4.1 shows that under Assumption 1, the stop-loss function $I_d(x) = (x-d)_+$ for some $d \in [x_1', x_1)$ is a solution to Problem 1.

In the classical setting where $\epsilon = 0$ (and thus $F^P = F^{\mathbb{Q}}$ for all $P \in \mathcal{P}_0$), as per Lemma 3.1, we have $I_c^*(x) = \int_0^x \eta^*(t;P)dt$ where

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

Based on the definition of x_1 in Theorem 3.2, it holds that if $(1+\theta)S^{\mathbb{Q}}(x) > g(S^{\mathbb{Q}}(x))$ on $[0,x_1)$, then $I_c^*(x) = (x-x_1)_+$. Note that when $\beta \to \infty$ (or equivalently $\epsilon = 0$), our setting reduces to the classical setting, and (4.2) tells that $x_1' = x_1$ (the right end-point when the set becomes empty). Hence, (4.3) becomes $I^*(x) = \int_0^x \mathbb{1}_{[x_1,\infty)}(t)dt = (x-x_1)_+$, which is exactly $I_c^*(x)$.

4.2. Tail value-at-risk

In this section, we study the case in which the DM is a TVaR minimizer.

Definition 4.1. The TVaR of a random variable Z at the confidence level $\alpha \in (0,1)$, under probability measure P, is defined as

$$\mathrm{TVaR}_\alpha^P(Z) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_t^P(Z)dt, \tag{4.4}$$

with

$$\mathrm{VaR}_\alpha^P(Z) = \inf\{z \in \mathbb{R} : P(Z \le z) \ge \alpha\}.$$

The distortion function for TVaR at the confidence level $\alpha \in (0,1)$ is given by $g(x) = \min\{1, \frac{x}{1-\alpha}\}$ (Dhaene et al., 2006), which is indeed increasing and concave. Note that this distortion function g is not (twice) differentiable, and thus Assumption 1 does not hold.

We focus on the following case:

$$1+\theta < \frac{1}{1-\alpha}, \tag{4.5}$$

which is the most common case in practice.[7]

We first look at the case when $\beta = 0$, and the L^2 constraint is slack. Then, it follows from Theorem 3.2 that $t_0 = g^{-1}(1) = 1-\alpha$. Since $S^{\mathbb{Q}}(x) > \frac{1}{1+\theta} > 1-\alpha = t_0$ for $x \in [0,x_0)$, it follows from Theorem 3.2 that $\tilde{S}^*(x) = S^{\mathbb{Q}}(x)$ for $x \in [0,x_0)$. It is easy to see that $(1+\theta)x < \min\{1, \frac{x}{1-\alpha}\}$ when $x < \frac{1}{1+\theta}$, and thus $x_1 = x_0$. Therefore, a worst-case survival function of X is given by

$$\tilde{S}^*(x) = S^{\mathbb{Q}}(x)\mathbb{1}_{[0,x_0)}(x) + S^{\mathbb{Q}}(x)\mathbb{1}_{[x_0,M]}(x) = S^{\mathbb{Q}}(x). \tag{4.6}$$

Thus, if $\epsilon \ge \int_0^M (\tilde{S}^*(x)-S^{\mathbb{Q}}(x))^2dx = 0$, a worst-case survival function of X is given by $S^{\mathbb{Q}}(x)$. From Lemma 3.1, we get that the corresponding optimal indemnity function is given by $I^*(x) = (x-x_0)_+$. Since $\epsilon \ge 0$ always holds, the case with $\beta > 0$ becomes irrelevant.

The intuition of these findings with TVaR is as follows. It will be optimal to insure risks larger than x_0 via a stop-loss indemnity, and thus the retained risk is capped by a deductible. The probability under $\mathbb{Q}$ that the retained risk is equal to the deductible (maximum loss) is large enough, according to (4.5) so that the TVaR is equal to this deductible. Under alternative distributions, the value of TVaR cannot increase strictly as the TVaR cannot exceed the maximum loss. Therefore, a worst-case probability measure is given by $S^{\mathbb{Q}}(x)$.

5. Under the L^1 distance

In this section, we revisit Problem 1 but with the L^1 distance. As shown by many works in the literature, e.g., Panaretos and Zemel (2019), Villani (2009), the L^1 distance coincides with the first-order Wasserstein distance[8]:

$$\begin{aligned} D_1(F^P,F^{\mathbb{Q}}) &:= \int_{-\infty}^{\infty}|F^P(x)-F^{\mathbb{Q}}(x)|dx = \int_0^1 |(F^P)^{-1}(t)-(F^{\mathbb{Q}})^{-1}(t)|dt \\ &= \inf_{X\sim F^P, Y\sim F^{\mathbb{Q}}} \mathbb{E}[|X-Y|], \end{aligned} \tag{5.1}$$

where F^P, $F^{\mathbb{Q}}$ are two CDFs, and $(F^P)^{-1}$ and $(F^{\mathbb{Q}})^{-1}$ are their corresponding quantile functions. The above equation also serves as a bridge to analytically compare the L^p distance and the pth-order Wasserstein distance. In particular, for $p \in \mathbb{Z}^+$, if $D_p(F^P,F^{\mathbb{Q}}) \ge 1$, then

$$\begin{aligned} D_p(F^P,F^{\mathbb{Q}}) &= \left(\int_{-\infty}^{\infty}|F^P(x)-F^{\mathbb{Q}}(x)|^p dx\right)^{\frac{1}{p}} \le \int_{-\infty}^{\infty}|F^P(x)-F^{\mathbb{Q}}(x)|^p dx \\ &\le \int_{-\infty}^{\infty}|F^P(x)-F^{\mathbb{Q}}(x)|dx = \int_0^1 |(F^P)^{-1}(t)-(F^{\mathbb{Q}})^{-1}(t)|dt \\ &\le \left(\int_0^1 |(F^P)^{-1}(t)-(F^{\mathbb{Q}})^{-1}(t)|^p dt\right)^{\frac{1}{p}}\left(\int_0^1 1^q dt\right)^{\frac{1}{q}} \\ &= \left(\int_0^1 |(F^P)^{-1}(t)-(F^{\mathbb{Q}})^{-1}(t)|^p dt\right)^{\frac{1}{p}}, \end{aligned}$$

where the last inequality is a consequence of the Hölder inequality with $\frac{1}{p}+\frac{1}{q} = 1$. As such, if the DM adopts $\epsilon > 1$ in her $\mathcal{P}_\epsilon$, applying the L^p distance would result in more choices of probability measures.

In view of the equivalence between the L^1 distance and the first-order Wasserstein distance, the results obtained in this section also partially generalize those in Birghila and Pflug (2019).

For a given worst-case survival function, the results in Section 3.1 still hold. With the L^1 distance, we then focus on the following problem:

$$\begin{cases} \sup\limits_{S^P\in\mathcal{S}_X} \displaystyle\int_0^M \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x)\right)dx, \\ \text{s.t.} \displaystyle\int_0^M |S^P(x)-S^{\mathbb{Q}}(x)|dx \le \epsilon, \quad \epsilon > 0. \end{cases} \tag{5.2}$$

Similar to the discussions in Section 3.2, we know that the worst-case survival function $S^{**}(x)$, given by

$$\begin{aligned} S^{**} = arg\ sup_{S^P\in\mathcal{S}_X} &\int_0^M \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x)\right)dx, \\ \text{s.t.} &\int_0^M |S^P(x)-S^{\mathbb{Q}}(x)|dx \le \epsilon, \quad \epsilon > 0. \end{aligned}$$

should satisfy $S^{**}(x) \ge S^{\mathbb{Q}}(x)$ for all $x \ge 0$. Since $S^{\mathbb{Q}}$ is strictly feasible to Problem (5.2), as per Slater's condition, solving (5.2) is equivalent to solving its dual:

$$\sup_{S^P\in\mathcal{S}_X} \int_0^M \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x) - \beta|S^P(x)-S^{\mathbb{Q}}(x)|\right)dx, \tag{5.3}$$

for some $\beta \ge 0$.

Again, we treat the cases when $\beta = 0$ and $\beta > 0$ in different ways. The following theorem summarizes the worst-case survival function for the case when $\beta = 0$.

Theorem 5.1. *If $\epsilon \ge \int_0^M |\tilde{S}^*(x)-S^{\mathbb{Q}}(x)|dx$ where*

$$\tilde{S}^*(x) = \left(t_0 \vee S^{\mathbb{Q}}(x)\right)\mathbb{1}_{[0,x_0)}(x) + g^{-1}((1+\theta)S^{\mathbb{Q}}(x))\mathbb{1}_{[x_0,x_1)}(x) + S^{\mathbb{Q}}(x)\mathbb{1}_{[x_1,M]}(x),$$

where t_0, x_0 and x_1 are defined in Theorem 3.2, then a worst-case survival function of X is given by $S^{**} = \tilde{S}^*$.

[7] In insurance regulation, α is usually chosen to be close to 1, e.g. 0.99 or 0.975 based on the Swiss Solvency Test (SST) or BCBS (2013). Then, it follows $1+\theta < \frac{1}{1-\alpha}$ for reasonable values of the risk loading θ.

[8] Here, the notation $X \sim F^P$ means that the CDF of the random variable X is F^P.

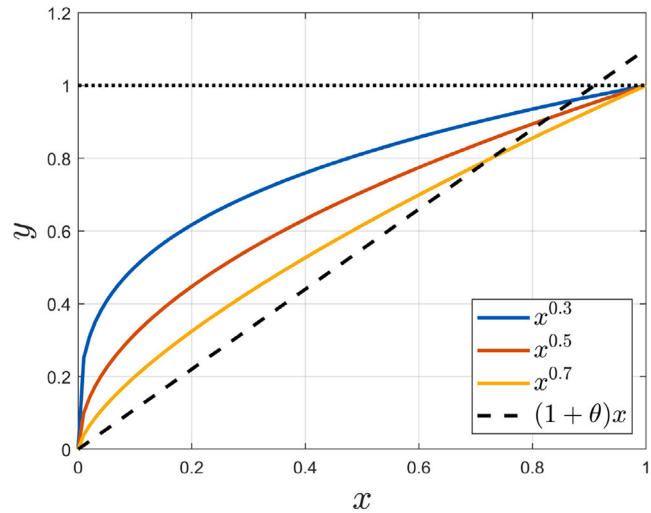

Fig. 1. The distortion functions for different parameters.

In the case where $\beta > 0$ in (5.3), we have the following theorem.

Theorem 5.2. *Let x_0 and x_1 be defined in Theorem 3.2. For Problem (5.3), when $\beta > 0$, the worst-case survival function of X is uniquely given by*

$$S^{**}(x;\beta) = \hat{S}(x;\beta)\mathbb{1}_{[0,x_0)}(x) + \left(\hat{S}(x;\beta) \wedge g^{-1}((1+\theta)S^{\mathbb{Q}}(x))\right)\mathbb{1}_{[x_0,x_1)}(x) + S^{\mathbb{Q}}(x)\mathbb{1}_{[x_1,M]}(x), \tag{5.4}$$

where

$$\hat{S}(x;\beta) = \begin{cases} 1, & \text{if } g'(1^-) \geq \beta, \\ t_1, & \text{if } g'(1^-) < \beta < g'(S^{\mathbb{Q}}(x)^+), \\ S^{\mathbb{Q}}(x), & \text{if } g'(S^{\mathbb{Q}}(x)^+) \leq \beta. \end{cases} \tag{5.5}$$

where t_1 satisfies $g'(t_1^-) \geq \beta$ and $g'(t_1^+) \leq \beta$. Now, $S^{**}(x;\beta)$ solves Problem (5.2) where β is such that

$$\int_0^M |S^{**}(x;\beta) - S^{\mathbb{Q}}(x)|dx = \epsilon. \tag{5.6}$$

Furthermore, β is decreasing with respect to ϵ.

Clearly, it is true that

$$|S^P(x) - S^{\mathbb{Q}}(x)| \geq (S^P(x) - S^{\mathbb{Q}}(x))^2 \quad \text{for} \quad x \geq 0,$$

and so it follows that any distribution function that satisfies the constraint of Problem (5.2) also satisfies the constraint of Problem (3.4) under the same ϵ.[9] Thus, applying the L^1 distance results in a smaller uncertainty set for the distributions, which is a subset of the uncertainty set under the L^2 distance. If β in (5.5) is larger (which represents a stronger effect of the constraint in (5.2)), then Theorem 5.2 tells that $S^{**}(x;\beta) = S^{\mathbb{Q}}(x)$ for small x, which implies that the DM would only assign more weight to the probabilities for medium-sized losses. This is partially attributable to the smaller distributional uncertainty set.

6. Numerical examples

In this section, we present a numerical example to analyze the effect of the distortion function on the worst-case survival function as well as the optimal indemnity function. We also present another example to study the effect of the order of the L^p distance on the worst-case survival function and the optimal indemnity function.

6.1. The effect of the distortion function on the worst-case survival function and indemnity function

We provide a numerical analysis under the L^2 distance under the following setup:

- The concave distortion function of the DM is of the power type, also known as proportional hazards (PH) transform (Wang, 1995), i.e.,

  $$g(x) = x^p$$

  where we focus on $p = 0.3, 0.5$ or 0.7. A lower value of p corresponds to a more concave distortion function and, thus, a higher aversion to mean-preserving spreads (Yaari, 1987).
- The benchmark loss distribution is the exponential distribution with mean 1000, i.e.,

  $$S^{\mathbb{Q}}(x) = e^{-\frac{x}{1000}}, \quad x \in [0, 10^6].$$

- The safety loading factor is equal to $\theta = 0.1$.

Fig. 1 illustrates the distortion functions under the different power parameters. A larger value of p corresponds with a less concave distortion function. For $p = 0.3, 0.5$ and 0.7, the smallest values of ϵ for the L^2-distance-based constraints to be slack are given by $0.377, 0.514$ and 0.807 respectively. To investigate the effect of p on the worst-case survival function, we look into two cases — when the L^2 constraint is always slack and when the L^2 constraint is always binding. For the second case, we select $\epsilon = 0.2$. By applying Theorems 3.2 and 3.3, the worst-case survival functions are shown in Fig. 2. Two interesting observations can be made:

- For the case when the L^2 constraint is slack, we have $S_1^*(x) \leq S_2^*(x) \leq S_3^*(x)$. In other words, if denoting by X_1, X_2 and X_3 the random variables whose survival functions are S_1^*, S_2^* and S_3^*, we have $X_1 \leq_{st} X_2 \leq_{st} X_3$, where $Z \leq_{st} Y$ means that Z is smaller than Y in terms of the first order stochastic dominance.
- For the case when the L^2 constraint is binding, we have

  $$\int_0^\infty S_1^*(x)dx = 1004.8 < \int_0^\infty S_2^*(x)dx = 1005.6 < \int_0^\infty S_3^*(x)dx = 1007.4.$$

  Note that any two of the three worst-case survival functions cross each other only once. Based on Definition 2.2 and Theorem 2.3 of Cheung et al. (2015), we have $X_1 \leq_{icx} X_2 \leq_{icx} X_3$, where X_1, X_2, X_3 denote the random variables whose survival functions are S_1^*, S_2^*, S_3^*, and $Z \leq_{icx} Y$ means that Z is smaller than Y in terms of the increasing convex order.

It may seem counter-intuitive that a DM who is less risk averse (larger p) would assign more weight to the probabilities of large losses. We attribute this observation to the effect of insurance. Fig. 3 displays the net price, i.e. $(1+\theta)S^{\mathbb{Q}}(x) - g(S^*(x))$, for the marginal coverage $I'(x)$ for a loss x. Note that the DM would purchase insurance only when the net price is negative. In either left or right panel of Fig. 3, the DM who is more risk averse (small p) purchases more insurance. Under this situation, the DM would not assign more weight to the part of the risk that has been transferred to the insurer.

[9] The problems being compared here are Problems (5.2) and (3.4), where we study the effect of changing the order of L^p distance only. For the constraints of original problems, we have $D_2(F^P, F^{\mathbb{Q}}) = (\int_0^\infty (S^P(x) - S^{\mathbb{Q}}(x))^2 dx)^{\frac{1}{2}} \leq \sqrt{\epsilon}$ under the L^2 distance and $D_1(F^P, F^{\mathbb{Q}}) = \int_0^\infty (S^P(x) - S^{\mathbb{Q}}(x))dx \leq \epsilon$ under the L^1 distance.

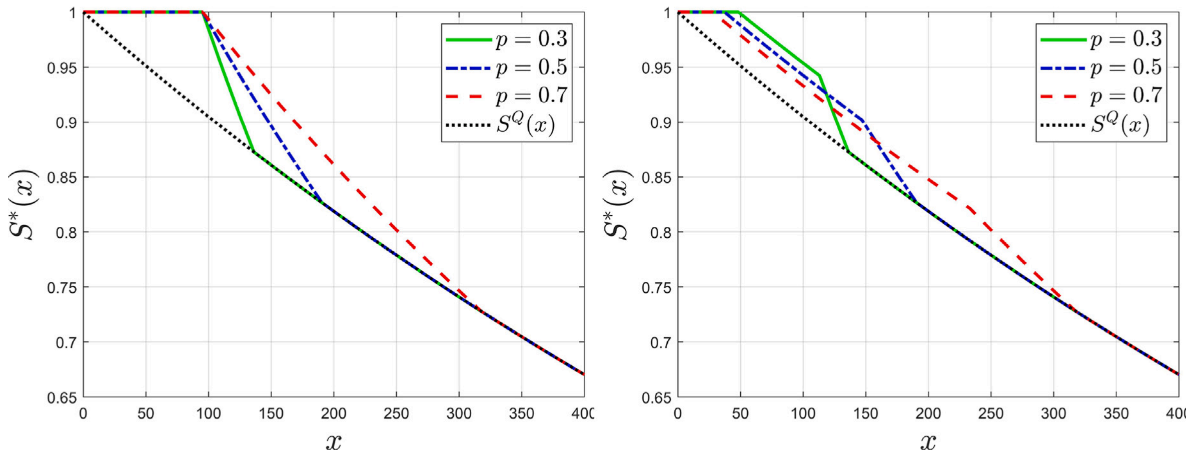

Fig. 2. (Left) the worst-case survival functions for different parameters p when the L^2 constraints are slack; (right) the worst-case survival functions for different parameters p when the L^2 constraints are binding.

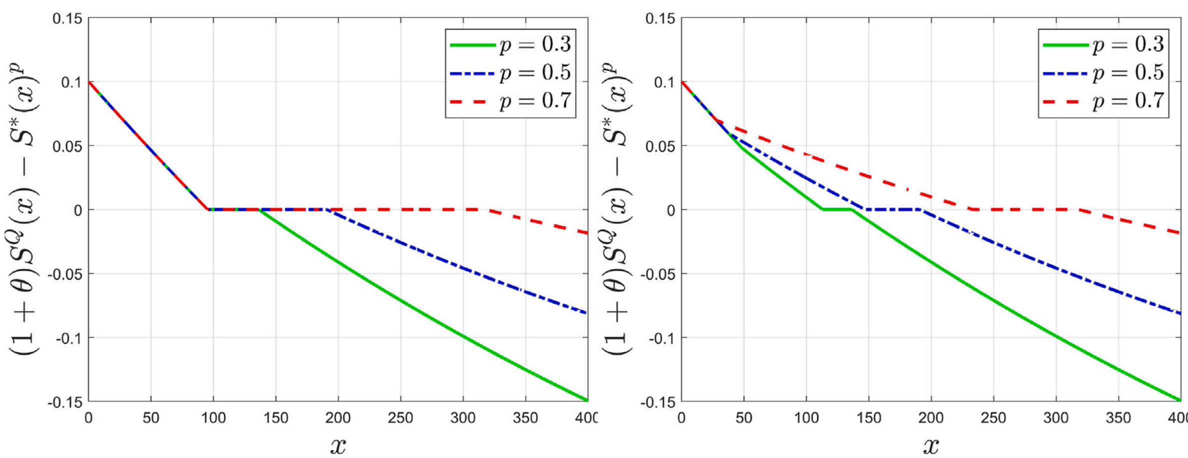

Fig. 3. (Left) the net price for purchasing $I'(x)$ when the L^2 constraints are all slack; (right) the net price for purchasing $I'(x)$ when the L^2 constraints are all binding. The net price is given by $(1+\theta)S^{\mathbb{Q}}(x) - (S^*(x))^p$.

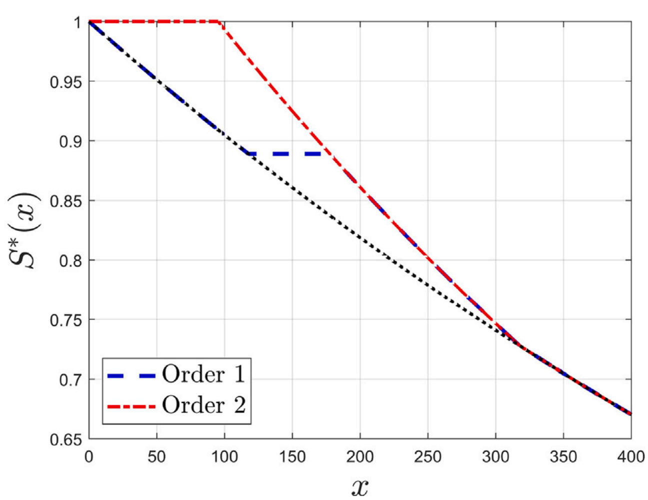

Fig. 4. The worst-case survival functions under the L^1 and L^2 distance metrics.

6.2. The effect of the order of L^p distance on the worst-case survival function and indemnity function

In this section, we investigate the effect of the order of L^p distance on the worst-case survival function and the optimal indemnity function. We consider the DM whose distortion function is $g(x) = x^{0.7}$. As shown in Section 5, the constraint under the L^1 distance is more restrictive than that under the L^2 distance for the same value of ϵ. Under the setting of our example, the smallest values of ϵ for the L^1 and L^2 constraints to be slack are given by 13.66 and 0.807 respectively. In what follows, we select $\epsilon = 5$ such that the constraint under the L^1 distance becomes binding while the constraint under the L^2 distance is slack. Fig. 4 shows the worst-case survival functions under the two different L^p distance metrics. As expected, the worst-case survival function under the L^1 distance gets closer to the benchmark survival function. Furthermore, under the L^1 distance, the DM would assign more weight to the probabilities for the medium-sized losses.

Fig. 5 exhibits the net prices for purchasing $I'(x)$ under the worst-case survival functions as shown in Fig. 4. Since the uncertainty set for distributions under the L^1 distance is a subset of that under the L^2 distance, it follows that the net price under the L^1 distance is higher than that under the L^2 distance. Also, the DM who applies the L^2 distance would retain less risk to herself.

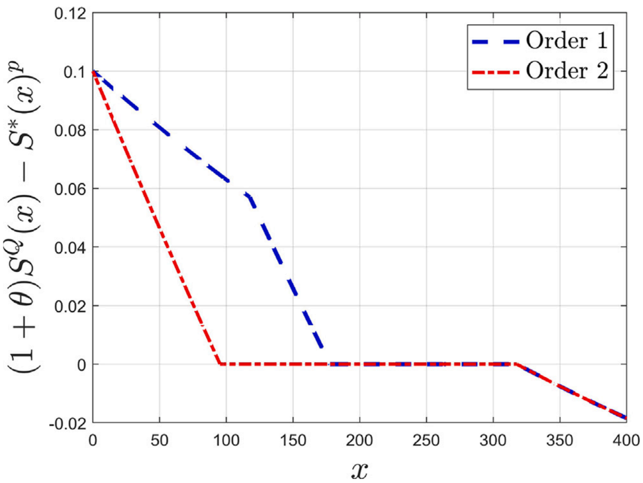

Fig. 5. The net prices for purchasing $I'(x)$ under the worst-case survival functions in Fig. 4, given by $(1+\theta)S^{\mathbb{Q}}(x) - (S^*(x))^p$.

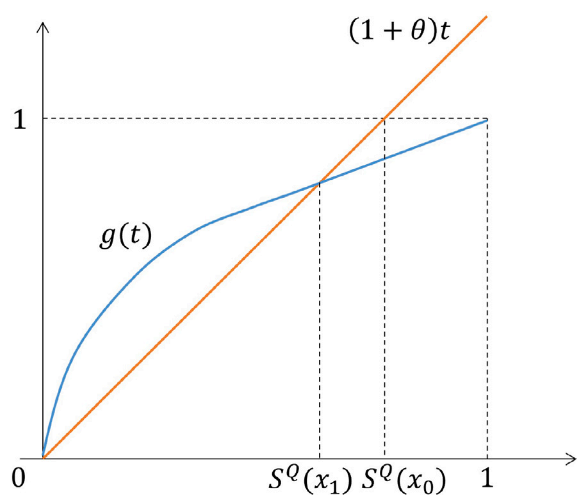

Fig. 6. The illustration of x_0 and x_1.

7. Concluding remarks and future research

This paper investigates a distributionally robust insurance problem from the viewpoint of a decision-maker (DM). The DM is assumed to consider all distributions close to the benchmark distribution, with "closeness" measured by either the L^2 or L^1 distance. Subsequently, the DM minimizes the distortion risk measure of the terminal loss under the worst-case distribution. We explicitly derive both the worst-case distribution and the optimal indemnity. The optimal insurance indemnity is typically of a layer-type form. If the benchmark distribution of the DM is the one used for insurance pricing, then the worst-case distribution has three pieces, and its tail will follow that of the benchmark distribution. In a specific example where the DM uses TVaR, we find that the worst-case distribution exactly matches the benchmark distribution in the most common situation. Additionally, we compare the worst-case distributions using the L^1 and L^2 distance metrics. We present some numerical examples to demonstrate the effects of risk aversion level and the order of L^p distance on the worst-case distribution and indemnity function.

We also extend our results to cases where the DM's benchmark distribution differs from the distribution used for pricing. Then, there exists belief heterogeneity between the DM and insurer regarding the benchmark distribution, and the effect of such belief heterogeneity is studied for the robust insurance problem. However, this introduces significantly more complexity to specific problems, even in cases where the DM uses TVaR. We leave the study of specific cases in which the solution is tractable for future research. While this paper mainly focuses on the case where the expected-value premium principle is applied, other premium principles, such as the mean–variance or mean-standard-deviation premium principles, also warrant investigation, which we leave for future research.

Acknowledgments

W.J. Jiang acknowledges the financial support received from the Natural Sciences and Engineering Research Council (RGPIN-2020-04204, DGECR-2020-00332) of Canada. W.J. Jiang is also grateful to the start-up grant received from the University of Calgary.

Appendix A. Proofs

Proof of Lemma 3.1

This result is well-known in the literature, but we provide a proof here for completeness.

Note that by using the Comonotonic additivity and Translation invariance of the distortion risk measure, as well as the integral representation of $\rho_g^P(I(X))$ (Cheung & Lo, 2017):

$$\rho_g^P(I(X)) = \int_0^M g\left(S^P(x)\right) dI(x),$$

the objective of Problem 1 can be written as

$$\begin{aligned}\rho_g^P(X - I(X) + \pi(I)) &= \rho_g^P(X) - \rho_g^P(I(X)) + \pi(I)\\ &= \rho_g^P(X) + \int_0^M ((1+\theta)S^{\mathbb{Q}}(x) - g(S^P(x)))dI(x).\end{aligned} \tag{A.1}$$

Based on the above equation, we can re-write the inner problem as

$$\begin{aligned}&\inf_{I\in\mathcal{I}} \int_0^M \left((1+\theta)S^{\mathbb{Q}}(x) - g(S^P(x))\right) dI(x)\\ =&\inf_{\eta\in\tilde{\mathcal{I}}} \int_0^M \left((1+\theta)S^{\mathbb{Q}}(x) - g(S^P(x))\right) I'(x)(x)dx\\ =&\inf_{\eta\in\tilde{\mathcal{I}}} \int_0^M \left((1+\theta)S^{\mathbb{Q}}(x) - g(S^P(x))\right) \eta(x)dx,\end{aligned} \tag{A.2}$$

where the equation holds due to (2.3), and

$$\tilde{\mathcal{I}} := \left\{\eta : [0,M] \mapsto [0,1] \;\middle|\; \eta \text{ is Lebesgue measurable}\right\}.$$

Then, to minimize the integral of (A.2), we only need to minimize its integrand function at each $x \in [0,\infty)$, which directly yields

$$\eta^*(x) = \begin{cases}1, & \text{if } (1+\theta)S^{\mathbb{Q}}(x) < g(S^P(x)),\\ \gamma(x), & \text{if } (1+\theta)S^{\mathbb{Q}}(x) = g(S^P(x)),\\ 0, & \text{if } (1+\theta)S^{\mathbb{Q}}(x) > g(S^P(x)),\end{cases}$$

for any γ that is a Lebesgue measurable and $[0,1]$-valued function. This concludes the proof. □

Proof of Theorem 3.2

Since g is a concave function and satisfies $g(0) = 0$ and $g(1) = 1$, the mapping $t \mapsto g(t) - (1+\theta)t$ can have 0, 1, or infinitely many roots on the interval $(0,1)$.

With the definitions of x_0 and x_1 (illustrated in Fig. 6), we first examine the following problem:

$$\sup_{S^P\in\mathcal{S}_X} \int_0^M \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x)\right) dx$$

$$= \sup_{S^P\in\mathcal{S}_X} \int_0^{x_0} \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x)\right) dx + \int_{x_0}^{x_1} \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x)\right) dx + \int_{x_1}^{M} \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x)\right) dx. \tag{A.3}$$

The survival function in the solutions to the above problem may not be unique, and we denote the set of solutions as $\mathcal{S}_X^s$. We find the solution in $\mathcal{S}_X^s$ that solves the next problem

$$\min_{S^P\in\mathcal{S}_X^s} \int_0^{M} \left(S^P(x) - S^{\mathbb{Q}}(x)\right)^2 dx = \min_{S^P\in\mathcal{S}_X^s} \int_0^{x_0} \left(S^P(x) - S^{\mathbb{Q}}(x)\right)^2 dx + \int_{x_0}^{x_1} \left(S^P(x) - S^{\mathbb{Q}}(x)\right)^2 dx + \int_{x_1}^{M} \left(S^P(x) - S^{\mathbb{Q}}(x)\right)^2 dx. \tag{A.4}$$

Given the definition of x_0, if $x < x_0$, then $(1+\theta)S^{\mathbb{Q}}(x) \ge 1$ holds true, resulting in the equation $g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x) = g(S^P(x))$. Since g is increasing and concave, we define $t_0 = g^{-1}(1)$, then any $t \in [t_0, 1]$ can maximize $g(t)$ over $[0,1]$. To minimize $\int_0^{x_0} \left(S^P(x) - S^{\mathbb{Q}}(x)\right)^2 dx$ over all $S^P \in \mathcal{S}_X^s$, we take $\tilde{S}^*(x) = t_0 \vee S^{\mathbb{Q}}(x)$.

When $x \in [x_0, x_1)$, we note that $g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x) \le (1+\theta)S^{\mathbb{Q}}(x)$. To maximize $g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x)$ over $S^P \in \mathcal{S}_X$, it is clear that solutions $S^P \in \mathcal{S}_X^s$ satisfy $S^P(x) \ge g^{-1}((1+\theta)S^{\mathbb{Q}}(x))$. To minimize $\int_{x_0}^{x_1} \left(S^P(x) - S^{\mathbb{Q}}(x)\right)^2 dx$ over all $S^P \in \mathcal{S}_X^s$, we take $\tilde{S}^*(x) = g^{-1}((1+\theta)S^{\mathbb{Q}}(x))$ for $x \in [x_0, x_1)$.

When $x \in [x_1, M]$, based on the definition of x_1 we have

$$\left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x)\right) \le \left(g(S^P(x)) \wedge g(S^{\mathbb{Q}}(x))\right) \le g(S^{\mathbb{Q}}(x)).$$

Therefore, we take $\tilde{S}^*(x) = S^{\mathbb{Q}}(x)$ for $x \in [x_1, M]$, and this automatically minimizes $\int_{x_1}^{M} \left(S^P(x) - S^{\mathbb{Q}}(x)\right)^2 dx$ over $S^P \in \mathcal{S}_X^s$.

Now, let

$$\tilde{S}^*(x) = \left(t_0 \vee S^{\mathbb{Q}}(x)\right) \mathbb{1}_{[0,x_0)}(x) + g^{-1}\left((1+\theta)S^{\mathbb{Q}}(x)\right) \mathbb{1}_{[x_0,x_1)}(x) + S^{\mathbb{Q}}(x)\mathbb{1}_{[x_1,\infty)}(x),$$

we show that this is indeed a survival function. First, $\tilde{S}^*$ is decreasing on $[0,x_0)$, $[x_0,x_1)$ and $[x_1,M]$, and is right-continuous (since $S^{\mathbb{Q}}$ is right-continuous). Now take arbitrary $a_0 \in [0,x_0)$, $a_1 \in [x_0,x_1)$ and $a_2 \in [x_1,M]$. Since $(1+\theta)S^{\mathbb{Q}}(x) \le 1$ on $[x_0,x_1)$, we have $g^{-1}((1+\theta)S^{\mathbb{Q}}(a_1)) \le g^{-1}(1) = t_0 \le t_0 \vee S^{\mathbb{Q}}(a_0)$. Since $(1+\theta)S^{\mathbb{Q}}(x) \ge g(S^{\mathbb{Q}}(x))$ on $[x_0,x_1)$, we have $g^{-1}((1+\theta)S^{\mathbb{Q}}(a_1)) \ge S^{\mathbb{Q}}(a_1) \ge S^{\mathbb{Q}}(a_2)$. Thus, $\tilde{S}^*(x)$ is decreasing on $[0,x_0) \cup [x_0,x_1) \cup [x_1,M]$. This confirms that $\tilde{S}^*$ is a survival function.

It is apparent that $\tilde{S}^*$ is the survival function S^{P^*} that solves

$$\max_{S^P\in\mathcal{S}_X} \int_0^{M} \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x)\right) dx.$$

Furthermore, $\tilde{S}^*$ minimizes $\int_0^M \left(S^P(x) - S^{\mathbb{Q}}(x)\right)^2 dx$ among the survival functions that maximize $\int_0^M \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x)\right) dx$. Thus, if $\epsilon \ge \int_0^M \left(\tilde{S}^*(x) - S^{\mathbb{Q}}(x)\right)^2 dx$, then the worst-case survival function is given by $S^*(x) = \tilde{S}^*(x)$. □

Proof of Theorem 3.3

For a given $\beta > 0$, we can rewrite Problem (3.5) as follows:

$$\sup_{S^P\in\mathcal{S}_X} \int_0^{x_0} \left(g(S^P(x)) - \beta(S^P(x) - S^{\mathbb{Q}}(x))^2\right) dx + \int_{x_0}^{M} \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x) - \beta(S^P(x) - S^{\mathbb{Q}}(x))^2\right) dx. \tag{A.5}$$

To facilitate the subsequent discussions, we define the following functions:

$$K_1(t) = g(t) - \beta(t - S^{\mathbb{Q}}(x))^2,$$
$$K_2(t) = (1+\theta)S^{\mathbb{Q}}(x) - \beta(t - S^{\mathbb{Q}}(x))^2,$$
$$K_3(t) = g(t) \wedge (1+\theta)S^{\mathbb{Q}}(x) - \beta(t - S^{\mathbb{Q}}(x))^2$$

for $t \in [0,1]$. To solve Problem (A.5), we will consider the following cases:

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

In this case, our goal is to maximize $g(S^P(x)) - \beta(S^P(x) - S^{\mathbb{Q}}(x))^2$ for each $x \in [0,x_0)$, with $S^P \in \mathcal{S}_X$. Since $g(t)$ and $-\beta(t - S^{\mathbb{Q}}(x))^2$ are both concave functions of t, the function $K_1(t)$ is concave. Note that $K_1'(t) = g'(t) - 2\beta(t - S^{\mathbb{Q}}(x))$, which is strictly decreasing due to $\beta > 0$. We further note that $g'(0^+) - 2\beta(0 - S^{\mathbb{Q}}(x)) > 0$, and thus the optimal t that maximizes $K_1(t)$ over $[0,1]$ is positive. Next, we will discuss two sub-cases:

Sub-case 1: If $g'(1^-) - 2\beta F^{\mathbb{Q}}(x) < 0$, then the optimal t that maximizes $K_1(t)$ is an element of $(0,1)$. To accommodate the points at which g is not differentiable, the solution that maximizes $K_1(t)$ over $[0,1]$, which is denoted by $\tilde{S}(x)$, satisfies

$$g'(\tilde{S}(x)^-) - 2\beta(\tilde{S}(x) - S^{\mathbb{Q}}(x)) \ge 0, \quad g'(\tilde{S}(x)^+) - 2\beta(\tilde{S}(x) - S^{\mathbb{Q}}(x)) \le 0 \tag{A.6}$$

where $g'(t^-) := \lim_{x\to t^-} g'(x)$ and $g'(t^+) = \lim_{x\to t^+} g'(x)$. Note that if g is differentiable at $\tilde{S}(x)$, then (A.6) becomes

$$g'(\tilde{S}(x)) - 2\beta(\tilde{S}(x) - S^{\mathbb{Q}}(x)) = 0,$$

which is the traditional first-order condition for maximizing the concave function $K_1(t)$. If $x_1 > x_2$, then $S^{\mathbb{Q}}(x_1) \le S^{\mathbb{Q}}(x_2)$, and

$$g'(\tilde{S}(x_1)^-) - 2\beta\tilde{S}(x_1) \ge -2\beta S^{\mathbb{Q}}(x_1) \ge -2\beta S^{\mathbb{Q}}(x_2) \ge g'(\tilde{S}(x_2)^+) - 2\beta\tilde{S}(x_2).$$

Hence $\tilde{S}(x_1) \le \tilde{S}(x_2)$. This implies that $\tilde{S}(x)$ is decreasing. Note that $g'(S^{\mathbb{Q}}(x)^+) - 2\beta(S^{\mathbb{Q}}(x) - S^{\mathbb{Q}}(x)) \ge 0$, and thus $\tilde{S}(x) \ge S^{\mathbb{Q}}(x)$.

Sub-case 2: If $g'(1^-) - 2\beta F^{\mathbb{Q}}(x) \ge 0$, then the optimal t that maximizes $K_1(t)$ over $[0,1]$ is equal to 1.

Now define

$$\hat{S}(x) = \begin{cases} 1, & \text{if } g'(1^-) - 2\beta F^{\mathbb{Q}}(x) \ge 0, \\ \tilde{S}(x), & \text{if } g'(1^-) - 2\beta F^{\mathbb{Q}}(x) < 0, \end{cases}$$

Note that $g'(1^-) - 2\beta F^{\mathbb{Q}}(x)$ is decreasing in x, so $\hat{S}(x)$ is decreasing. Hence, the optimal survival function on $[0,x_0]$ is then given by $\hat{S}(x)$.

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

In this case, let $t^*(x) = g^{-1}((1+\theta)S^{\mathbb{Q}}(x)) \in [0,1]$. If $t \le t^*(x)$, or equivalently, $g(t) \le (1+\theta)S^{\mathbb{Q}}(x)$, then $K_3(t) = K_1(t)$, which is concave on $[0,t^*(x)]$. If $t \ge t^*(x)$, or equivalently, $g(t) \ge (1+\theta)S^{\mathbb{Q}}(x)$, then $K_3(t) = K_2(t)$, which is concave on $[t^*(x),1]$. Note that

$$K_3'(t^*(x)^-) = K_1'(t^*(x)^-) = g'\left(t^*(x)^-\right) - 2\beta\left(t^*(x) - S^{\mathbb{Q}}(x)\right)$$

and

$$K_3'(t^*(x)^+) = K_2'(t^*(x)^+) = -2\beta\left(t^*(x) - S^{\mathbb{Q}}(x)\right).$$

We have $K_3'(t^*(x)^-) \ge K_3'(t^*(x)^+)$, which implies that $K_3(t)$ is concave on $[0,1]$.

Next, we will discuss two sub-cases:

Sub-case 1: If $x < x_1$, then $(1+\theta)S^{\mathbb{Q}}(x) \ge g(S^{\mathbb{Q}}(x))$, or equivalently, $K_3'(t^*(x)^+) \le 0$. In this case, the maximum of $K_3(t)$ on $[0,1]$ is attained within $[0,t^*(x)]$. Note that $K_3(t) = K_1(t)$ when $t \in [0,t^*(x)]$. Then, similar to the discussion of **Case 1**, the optimal survival function in this case is represented by $S^*(x) = \tilde{S}(x) \wedge g^{-1}((1+\theta)S^{\mathbb{Q}}(x))$, where $\tilde{S}(x)$ satisfies (A.6).

Sub-case 2: If $x \ge x_1$, then $(1+\theta)S^{\mathbb{Q}}(x) \le g(S^{\mathbb{Q}}(x))$, or equivalently, $K_3'(t^*(x)^+) \ge 0$. In this case, the maximum of $K_3(t)$ on $[0,1]$ is attained within $[t^*(x),1]$. Note that $K_3(t) = K_2(t)$ when $t \in [t^*(x),1]$, where $K_2(t)$ is apparently a quadratic function. Therefore, in this case the optimal survival function is given by $S^*(x) = S^{\mathbb{Q}}(x)$.

Now let

$S^*(x) = \hat{S}(x)\mathbb{1}_{[0,x_0)}(x) + \left(\tilde{S}(x) \wedge g^{-1}((1+\theta)S^{\mathbb{Q}}(x))\right)\mathbb{1}_{[x_0,x_1)}(x) + S^{\mathbb{Q}}(x)\mathbb{1}_{[x_1,M]}(x).$

We have shown that S^* element-wisely maximizes the integrand function of the problem (A.5). Furthermore, S^* is right continuous and decreasing on $[0, x_0)$, $[x_0, x_1)$ and $[x_1, M]$. Finally, we will prove that S^* is a decreasing function on the interval $[0, M]$. Take arbitrary $a_1 \in [0, x_0)$, $a_2 \in [x_0, x_1)$ and $a_3 \in [x_1, M]$, we have

$$\tilde{S}(a_2) \wedge g^{-1}((1+\theta)S^{\mathbb{Q}}(a_2)) \le \tilde{S}(a_2) \le \tilde{S}(a_1) \le \hat{S}(a_1),$$

and

$$S^{\mathbb{Q}}(a_3) \le S^{\mathbb{Q}}(a_2) \le \tilde{S}(a_2) \wedge g^{-1}((1+\theta)S^{\mathbb{Q}}(a_2)), \tag{A.7}$$

where the second inequality of (A.7) is due to $S^{\mathbb{Q}}(x) \le \tilde{S}(x)$ and $g(S^{\mathbb{Q}}(x)) \le (1+\theta)S^{\mathbb{Q}}(x)$ on $[x_0, x_1)$. This shows that S^* is decreasing on its domain $[0, M]$.

Next, we show the existence of β such that $\int_0^M (S^*(x;\beta) - S^{\mathbb{Q}}(x))^2 dx = \epsilon$. We only discuss the case where the support of X under $\mathbb{Q}$ is $[0, M]$ as the proofs for other cases are similar. Note that if $\beta \to 0$, then as per (3.9) $\hat{S}(x;\beta) \to 1$, which leads to $S^*(x;\beta) \ge \tilde{S}^*(x)$ for all $x \in [0, M]$, where $\tilde{S}^*(x)$ is given by (3.7). As such,

$$\int_0^M (S^*(x;\beta) - S^{\mathbb{Q}}(x))^2 dx \ge \int_0^M (\tilde{S}^*(x) - S^{\mathbb{Q}}(x))^2 dx > \epsilon.$$

If $\beta \to \infty$, then based on (3.9) and (3.10) $\hat{S}(x;\beta) \to S^{\mathbb{Q}}(x)$, which results in

$$\int_0^M (S^*(x;\beta) - S^{\mathbb{Q}}(x))^2 dx = 0 < \epsilon.$$

Now, consider an arbitrary sequence $\{\beta_n\}_{n=1,2,\ldots} \to \beta_0$, where $\{\beta_i\}_{i=0,1,\ldots} \in (0,\infty)$. Since $g(\cdot)$ is almost everywhere differentiable on $[0, 1]$, we have that $S^*(x;\beta_n)$ converges pointwisely to $S^*(x;\beta_0)$ almost everywhere. Since

$$\int_0^M (S^*(x;\beta_n) - S^{\mathbb{Q}}(x))^2 dx \le \int_0^M (1 - S^{\mathbb{Q}}(x))^2 dx < \infty,$$

by using Lebesgue's Dominated Convergence Theorem, we have

$$\int_0^M (S^*(x;\beta_n) - S^{\mathbb{Q}}(x))^2 dx \to \int_0^M (S^*(x;\beta_0) - S^{\mathbb{Q}}(x))^2 dx.$$

As such, $\beta \to \int_0^M (S^*(x;\beta) - S^{\mathbb{Q}}(x))^2 dx$ is a continuous mapping. Therefore, for any $\epsilon \in (0, \int_0^M (\tilde{S}^*(x) - S^{\mathbb{Q}}(x))^2 dx)$, where $\tilde{S}^*(x)$ is given by (3.7), it follows from the intermediate value theorem that there exists a $\beta > 0$ such that $\int_0^M (S^*(x;\beta) - S^{\mathbb{Q}}(x))^2 dx = \epsilon$.

When $\beta > 0$, then the integrand function of (3.5) is strictly concave in $S^P(x)$, which naturally leads to the uniqueness of $S^*(x;\beta)$ that solves (3.5).

At last, we show that β is decreasing with respect to ϵ. Given $\beta_2 > \beta_1 > 0$, if $g'(1^-) - 2\beta_2 F^{\mathbb{Q}}(x) < 0$ while $g'(1^-) - 2\beta_1 F^{\mathbb{Q}}(x) \ge 0$, then $\hat{S}(x;\beta_1) = 1 > \hat{S}(x;\beta_2) = \tilde{S}(x;\beta_2)$. If $g'(1^-) - 2\beta_i F^{\mathbb{Q}}(x) < 0$ for both $i = 1, 2$ and we assume that $\tilde{S}(x;\beta_2) > \tilde{S}(x;\beta_1)$, then we have

$$\begin{aligned}
0 &\ge g'(\tilde{S}(x;\beta_1)^+) - 2\beta_1(\tilde{S}(x;\beta_1) - S^{\mathbb{Q}}(x)) \\
&\ge g'(\tilde{S}(x;\beta_1)^+) - 2\beta_2(\tilde{S}(x;\beta_1) - S^{\mathbb{Q}}(x)) \\
&> g'(\tilde{S}(x;\beta_2)^-) - 2\beta_2(\tilde{S}(x;\beta_2) - S^{\mathbb{Q}}(x)),
\end{aligned}$$

which contradicts with

$$g'(\tilde{S}(x;\beta_2)^-) - 2\beta_2(\tilde{S}(x;\beta_2) - S^{\mathbb{Q}}(x)) \ge 0.$$

As such $\tilde{S}(x;\beta_2) \le \tilde{S}(x;\beta_1)$. In other words, $\hat{S}(x;\beta)$ is decreasing in β. Then, it is straightforward that β is decreasing with respect to ϵ. This concludes the proof. □

Proof of Proposition 4.1

Here, (i) follows directly from the fact that when $x \in [0, x_0')$, $g'(1^-) - 2\beta F^{\mathbb{Q}}(x) \ge 0$, and thus $S^*(x;\beta) = \hat{S}(x;\beta) = 1$. In other words, the DM assigns zero probability to the event $\{\omega \in \Omega : X(\omega) \in [0, x_0')\}$.

To prove (ii), it should be noted that

$$\frac{d\left(g^{-1}((1+\theta)t) - t\right)}{dt} = \frac{1+\theta}{g'\left(g^{-1}((1+\theta)t)\right)} - 1 \ge \frac{1+\theta}{g'(t)} - 1 \ge 0,$$

if $g^{-1}((1+\theta)t) \ge t$ and $1+\theta \ge g'(t)$. If $x \in [x_0, x_1)$, we have $g^{-1}((1+\theta)S^{\mathbb{Q}}(x)) \ge S^{\mathbb{Q}}(x)$ and $1+\theta \ge g'(S^{\mathbb{Q}}(x))$ (as illustrated by Fig. 6). Therefore, $g^{-1}((1+\theta)S^{\mathbb{Q}}(x)) - S^{\mathbb{Q}}(x)$ is decreasing in x. Since $g'(g^{-1}((1+\theta)S^{\mathbb{Q}}(x)))$ is increasing in x, then

$$g'(g^{-1}((1+\theta)S^{\mathbb{Q}}(x))) - 2\beta\left(g^{-1}((1+\theta)S^{\mathbb{Q}}(x)) - S^{\mathbb{Q}}(x)\right)$$

is increasing in x. Therefore, when $x \in [x_0', x_1')$, we have

$$g'(g^{-1}((1+\theta)S^{\mathbb{Q}}(x))) - 2\beta\left(g^{-1}((1+\theta)S^{\mathbb{Q}}(x)) - S^{\mathbb{Q}}(x)\right) < 0.$$

Note that $\tilde{S}(x;\beta)$ satisfies

$$g'(\tilde{S}(x;\beta)) - 2\beta(\tilde{S}(x;\beta) - S^{\mathbb{Q}}(x)) = 0.$$

Thus, $g^{-1}((1+\theta)S^{\mathbb{Q}}(x)) > \tilde{S}(x;\beta)$, which results in $S^*(x;\beta) = \tilde{S}(x;\beta)$ for $x \in [x_0', x_1')$. Now, consider the following two cases.

- For $x \in [0, x_0')$, $S^*(x;\beta) = 1$, then $f^*(x;\beta) = 0 \le f^{\mathbb{Q}}(x)$.
- For $x \in [x_0', x_1')$, $S^*(x;\beta) = \tilde{S}(x;\beta)$, then it holds that

  $$g'(\tilde{S}(x;\beta)) = 2\beta\left(\tilde{S}(x;\beta) - S^{\mathbb{Q}}(x)\right),$$

  which, after differentiating both sides with respect to x, becomes

  $$g''(\tilde{S}(x;\beta))(-\tilde{f}(x;\beta)) = 2\beta(f^{\mathbb{Q}}(x) - \tilde{f}(x;\beta)) \ge 0$$

  where $\tilde{f}(x;\beta) = -\frac{\partial \tilde{S}(x;\beta)}{\partial x}$. Hence, $f^{\mathbb{Q}}(x) \ge \tilde{f}(x;\beta)$.

In words, compared with the benchmark distribution, the DM puts less weight on $[0, x_1')$ under the worst-case distribution. Similarly, it can be proven that the DM puts more weight on $[x_1', x_1)$ under the worst-case distribution. □

Proof of Theorem 5.2

With the definition of x_0, Problem (5.3) can be written as

$$\begin{aligned}
\sup_{S^P \in S_X} &\int_0^{x_0} \left(g(S^P(x)) - \beta|S^P(x) - S^{\mathbb{Q}}(x)|\right) dx \\
&+ \int_{x_0}^M \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x) - \beta|S^P(x) - S^{\mathbb{Q}}(x)|\right) dx.
\end{aligned} \tag{A.8}$$

Now, we solve Problem (A.8) by considering the following cases.

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

Based on the discussion following Problem (5.2), we are only interested in the survival functions which are greater than $S^{\mathbb{Q}}(x)$ for all $x \ge 0$. Therefore, we propose to solve the following problem:

$$\max_{t \in [S^{\mathbb{Q}}(x),1]} g(t) - \beta(t - S^{\mathbb{Q}}(x)), \tag{A.9}$$

which is concave in t.

If $g'(S^{\mathbb{Q}}(x)^+) \le \beta$, then

$$S^{\mathbb{Q}}(x) = \arg\max_{t \in [S^{\mathbb{Q}}(x),1]} g(t) - \beta(t - S^{\mathbb{Q}}(x)).$$

If $g'(1^-) \ge \beta$, then

$$1 = \arg\max_{t \in [S^{\mathbb{Q}}(x),1]} g(t) - \beta(t - S^{\mathbb{Q}}(x)).$$

If $g'(S^{\mathbb{Q}}(x)^+) > \beta > g'(1^-)$, then

$$t_1 = \arg\max_{t \in [S^{\mathbb{Q}}(x),1]} g(t) - \beta(t - S^{\mathbb{Q}}(x)),$$

where t_1 satisfies $g'(t_1^-) \geq \beta$ and $g'(t_1^+) \leq \beta$. Thus, the worst-case survival function at $x \in [0, x_0)$ is given by

$$\hat{S}(x) = \begin{cases} 1, & \text{if } g'(1^-) \geq \beta, \\ t_1, & \text{if } g'(1^-) < \beta < g'(S^{\mathbb{Q}}(x)^+), \\ S^{\mathbb{Q}}(x), & \text{if } g'(S^{\mathbb{Q}}(x)^+) \leq \beta. \end{cases}$$

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

Since we are only interested in the survival functions that are greater than $S^{\mathbb{Q}}(x)$, we will solve the following problem:

$$\max_{t \in [S^{\mathbb{Q}}(x), 1]} K(t) := g(t) \wedge (1+\theta)S^{\mathbb{Q}}(x) - \beta(t - S^{\mathbb{Q}}(x)). \tag{A.10}$$

Define $t^*(x) = g^{-1}((1+\theta)S^{\mathbb{Q}}(x))$. When $t > t^*(x)$, or equivalently, $g(t) > (1+\theta)S^{\mathbb{Q}}(x)$, we have $K(t) = (1+\theta)S^{\mathbb{Q}}(x) - \beta(t - S^{\mathbb{Q}}(x))$. When $t \leq t^*(x)$, or equivalently, $g(t) \leq (1+\theta)S^{\mathbb{Q}}(x)$, we have $K(t) = g(t) - \beta(t - S^{\mathbb{Q}}(x))$. Let x_1 be defined in Theorem 3.2, we have the following two sub-cases.

Sub-case 1: When $x < x_1$, or equivalently, $t^*(x) \geq S^{\mathbb{Q}}(x)$, we can calculate that

$$\lim_{t \to t^*(x)^-} K_3'(t) = K_3'(t^*(x)^-) = g'(t^*(x)^-) - \beta$$

and

$$\lim_{t \to t^*(x)^+} K_3'(t) = K_3'(t^*(x)^+) = -\beta.$$

Thus, the maximum of $K_3(t)$ on $[0, 1]$ can only be attained within $[S^{\mathbb{Q}}(x), t^*(x)]$. Note that on this interval, $K_3(t) = g(t) - \beta(t - S^{\mathbb{Q}}(x))$, thus the worst-case survival function for this case is given by $\hat{S}(x) \wedge g^{-1}((1+\theta)S^{\mathbb{Q}}(x))$.

Sub-case 2: When $x \geq x_1$, or equivalently, $t^*(x) \leq S^{\mathbb{Q}}(x)$, Problem (A.10) becomes

$$\max_{t \in [S^{\mathbb{Q}}(x), 1]} (1+\theta)S^{\mathbb{Q}}(x) - \beta(t - S^{\mathbb{Q}}(x)),$$

which is decreasing in t. Thus, the worst-case survival function for this case is given by $S^{\mathbb{Q}}(x)$.

Appendix B. The case with a different benchmark distribution

In the main text, we have discussed the worst-case distribution for the DM when her benchmark distribution is given by $\mathbb{Q}$. In this section we consider the case when the DM uses another benchmark distribution. For convenience, we denote by $\mathbb{B} \in \mathcal{P}([0, M])$ the DM's benchmark distribution. Under the L^2 distance, the main problem is now given by the following:

$$\begin{cases} \sup_{S^P \in S_X} \int_0^M \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x) \right) dx, \\ \text{s.t.} \int_0^M (S^P(x) - S^{\mathbb{B}}(x))^2 dx \leq \epsilon, \ \epsilon > 0. \end{cases} \tag{B.1}$$

Under the same reasoning as for (3.5), solving Problem (B.1) is equivalent to solving

$$\sup_{S^P \in S_X} \int_0^M \left(g(S^P(x)) \wedge (1+\theta)S^{\mathbb{Q}}(x) - \beta(S^P(x) - S^{\mathbb{B}}(x))^2 \right) dx \tag{B.2}$$

for some $\beta \geq 0$.

As in Section 3.2, we first present the result for the case when $\beta = 0$.

Theorem B.1. *Let x_0 be defined in Theorem 3.3. Define*

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

If $\epsilon \geq \int_0^M (\tilde{S}^*(x) - S^{\mathbb{B}}(x))^2 dx$, 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 is defined in Theorem 3.2, then the worst-case survival function that solves (B.2) is given by $S^* = \tilde{S}^*$.

Proof. The proof is similar to the proof of Theorem 3.2, and the only thing that needs to be shown is that $\tilde{S}^*$ is decreasing on $[0, x_0) \cup \mathcal{A} \cup \mathcal{B}$. To do so, we take $a_1 \in [0, x_0)$, $a_2 \in \mathcal{A}$ and $a_3 \in \mathcal{B}$, and analyze the following cases.

- If $a_1 < a_2$, then
$$S^*(a_2) = g^{-1}((1+\theta)S^{\mathbb{Q}}(a_2)) \leq g^{-1}(1) \leq t_0 \vee S^{\mathbb{Q}}(a_1) = S^*(a_1).$$
- If $a_1 < a_3$, then
$$S^*(a_3) = S^{\mathbb{B}}(a_3) \leq t_0 \vee S^{\mathbb{B}}(a_1) = S^*(a_1).$$
- If $a_2 < a_3$, then
$$S^*(a_3) = S^{\mathbb{B}}(a_3) \leq S^{\mathbb{B}}(a_2) \leq g^{-1}((1+\theta)S^{\mathbb{Q}}(a_2)) = S^*(a_2).$$
- If $a_3 < a_2$, then
$$S^*(a_2) = g^{-1}((1+\theta)S^{\mathbb{Q}}(a_2)) \leq g^{-1}((1+\theta)S^{\mathbb{Q}}(a_3)) \leq S^{\mathbb{B}}(a_3) = S^*(a_3).$$

Summarizing the above results indicates that S^* is indeed decreasing on $[0, x_0) \cup \mathcal{A} \cup \mathcal{B}$. □

When the constraint of Problem (B.1) is binding (i.e., $\beta > 0$), as in Theorem 3.3, we have the following result for the worst-case survival function.

Theorem B.2. *Let x_0 be as defined in Theorem 3.2 and $\mathcal{A}$ and $\mathcal{B}$ be as defined in Theorem B.1. If $\beta > 0$ in Problem (B.2), the worst-case survival function of X is given by*

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

where

$$\hat{S}(x;\beta) = \begin{cases} 1, & \text{if } g'(1^-) - 2\beta F^{\mathbb{B}}(x) \geq 0, \\ \tilde{S}(x;\beta), & \text{if } g'(1^-) - 2\beta F^{\mathbb{B}}(x) < 0, \end{cases}$$

where $\tilde{S}(x;\beta)$ satisfies

$$g'(\tilde{S}(x;\beta)^-) - 2\beta(\tilde{S}(x;\beta) - S^{\mathbb{B}}(x)) \geq 0, \quad g'(\tilde{S}(x;\beta)^+) - 2\beta(\tilde{S}(x;\beta) - S^{\mathbb{B}}(x)) \leq 0.$$

Here, β is such that

$$\int_0^M (S^*(x;\beta) - S^{\mathbb{B}}(x))^2 dx = \epsilon.$$

Proof. The proof is similar to the proof of Theorem 3.3, except for the proof that S^* is decreasing on $[0, x_0) \cup \mathcal{A} \cup \mathcal{B}$. This can be shown in the same way as in the proof of Theorem B.1. □

References

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

Arrow, K. J. (1974). Optimal insurance and generalized deductibles. *Scandinavian Actuarial Journal*, *1974*(1), 1–42.

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

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

Asimit, V., & Boonen, T. J. (2018). Insurance with multiple insurers: A game-theoretic approach. *European Journal of Operational Research*, *267*(2), 778–790.

Asimit, A. V., Hu, J., & Xie, Y. (2019). Optimal robust insurance with a finite uncertainty set. *Insurance: Mathematics & Economics*, *87*, 67–81.

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

Balbás, A., Balbás, B., Balbás, R., & Heras, A. (2015). Optimal reinsurance under risk and uncertainty. *Insurance: Mathematics & Economics*, *60*, 61–74.

Balbás, A., Balbás, B., & Heras, A. (2011). Stable solutions for optimal reinsurance problems involving risk measures. *European Journal of Operational Research*, *214*(3), 796–804.

Balbás, A., Garrido, J., & Mayoral, S. (2009). Properties of distortion risk measures. *Methodology and Computing in Applied Probability*, *11*(3), 385–399.
BCBS (2013). Fundamental review of the trading book: A revised market risk framework.
Bernard, C., De Vecchi, C., & Vanduffel, S. (2022). Robust assessment of life insurance products. Available at SSRN: https://ssrn.com/abstract=4087173 or http://dx.doi.org/10.2139/ssrn.4087173.
Bernard, C., Pesenti, S. M., & Vanduffel, S. (2023). Robust distortion risk measures. *Mathematical Finance*, (forthcoming).
Birghila, C., Boonen, T. J., & Ghossoub, M. (2023). Optimal insurance under maxmin expected utility. *Finance and Stochastics*, *27*, 467–501.
Birghila, C., & Pflug, G. C. (2019). Optimal XL-insurance under Wasserstein-type ambiguity. *Insurance: Mathematics & Economics*, *88*, 30–43.
Boonen, T. J. (2016). Optimal reinsurance with heterogeneous reference probabilities. *Risks*, *4*(3), 26.
Boonen, T. J., Liu, F., & Wang, R. (2021). Competitive equilibria in a comonotone market. *Economic Theory*, *72*(4), 1217–1255.
Borch, K. (1960). An attempt to determine the optimum amount of stop loss reinsurance. *Transactions of the XVI International Congress of Actuaries*, *1*, 597–610.
Cai, J., & Chi, Y. (2020). Optimal reinsurance designs based on risk measures: A review. *Statistical Theory and Related Fields*, *4*(1), 1–13.
Cai, J., & Tan, K. S. (2007). Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measures. *ASTIN Bulletin: Journal of the IAA*, *37*(1), 93–112.
Cheung, K. C., Chong, W. F., & Lo, A. (2019). Budget-constrained optimal reinsurance design under coherent risk measures. *Scandinavian Actuarial Journal*, *2019*(9), 729–751.
Cheung, K. C., Chong, W. F., & Yam, S. C. P. (2015). Convex ordering for insurance preferences. *Insurance: Mathematics & Economics*, *64*, 409–416.
Cheung, K. C., & Lo, A. (2017). Characterizations of optimal reinsurance treaties: A cost-benefit approach. *Scandinavian Actuarial Journal*, *2017*(1), 1–28.
Chi, Y. (2019). On the optimality of a straight deductible under belief heterogeneity. *ASTIN Bulletin: The Journal of the IAA*, *49*(1), 243–262.
Chi, Y., & 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., & Wu, L. (2013). Optimal reinsurance minimizing the distortion risk measure under general reinsurance premium principles. *Insurance: Mathematics & Economics*, *53*(1), 74–85.
Denuit, M., Dhaene, J., Goovaerts, M., & Kaas, R. (2006). *Actuarial theory for dependent risks: Measures, orders and models*. John Wiley & Sons.
Dhaene, J., Vanduffel, S., Goovaerts, M. J., Kaas, R., Tang, Q., & Vyncke, D. (2006). Risk measures and comonotonicity: A review. *Stochastic Models*, *22*(4), 573–606.
Fan, K. (1953). Minimax theorems. *Proceedings of the National Academy of Sciences*, *39*(1), 42–47.
Ghirardato, P., Maccheroni, F., & Marinacci, M. (2004). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, *118*(2), 133–173.
Ghossoub, M. (2017). Arrow's theorem of the deductible with heterogeneous beliefs. *North American Actuarial Journal*, *21*(1), 15–35.
Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, *18*(2), 141–153.
Huberman, G., Mayers, D., & Smith Jr, C. W. (1983). Optimal insurance policy indemnity schedules. *Bell Journal of Economics*, *14*(2), 415–426.
Jiang, W., Escobar-Anel, M., & Ren, J. (2020). Optimal insurance contracts under distortion risk measures with ambiguity aversion. *ASTIN Bulletin: The Journal of the IAA*, *50*(2), 619–646.
Jiang, W., Ren, J., Yang, C., & Hong, H. (2019). On optimal reinsurance treaties in cooperative game under heterogeneous beliefs. *Insurance: Mathematics & Economics*, *85*, 173–184.
Klibanoff, P., Marinacci, M., & Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, *73*(6), 1849–1892.
Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss models: From data to decisions*: *vol. 715*, John Wiley & Sons.
Liu, H., & Mao, T. (2022). Distributionally robust reinsurance with value-at-risk and conditional value-at-risk. *Insurance: Mathematics & Economics*, *107*, 393–417.
Lo, A. (2017). A Neyman-Pearson perspective on optimal reinsurance with constraints. *ASTIN Bulletin: The Journal of the IAA*, *47*(2), 467–499.
López-Díaz, M., Sordo, M. A., & Suárez-Llorens, A. (2012). On the L_p-metric between a probability distribution and its distortion. *Insurance: Mathematics & Economics*, *51*(2), 257–264.
Panaretos, V. M., & Zemel, Y. (2019). Statistical aspects of Wasserstein distances. *Annual Review of Statistics and its Application*, *6*, 405–431.
Pesenti, S., & Jaimungal, S. (2023). Portfolio optimisation within a Wasserstein ball. *SIAM Journal of Financial Mathematics*, *14*(4), 1175–1214.
Pflug, G. C., Timonina-Farkas, A., & Hochrainer-Stigler, S. (2017). Incorporating model uncertainty into optimal insurance contract design. *Insurance: Mathematics & Economics*, *73*, 68–74.
Rachev, S. T., Stoyanov, S. V., & Fabozzi, F. J. (2008). *Advanced stochastic models, risk assessment, and portfolio optimization: The ideal risk, uncertainty, and performance measures*. Wiley.
Stephens, M. A. (2017). Tests based on EDF statistics. In *Goodness-of-fit techniques* (pp. 97–194). Routledge.
Uppal, R., & Wang, T. (2003). Model misspecification and underdiversification. *The Journal of Finance*, *58*(6), 2465–2486.
Villani, C. (2009). *Optimal transport: Old and new*: *vol. 338*, Springer.
Wang, S. S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics & Economics*, *17*(1), 43–54.
Wang, S. S., Young, V. R., & Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics & Economics*, *21*(2), 173–183.
Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, *55*(1), 95–115.
Yang, J., Zhuang, W., & Hu, T. (2014). L_p-metric under the location-independent risk ordering of random variables. *Insurance: Mathematics & Economics*, *59*, 321–324.
Zhuang, S. C., Weng, C., Tan, K. S., & Assa, H. (2016). Marginal indemnification function formulation for optimal reinsurance. *Insurance: Mathematics & Economics*, *67*, 65–76.