Check for updates

Received: 19 January 2021 | Accepted: 3 January 2023

DOI: 10.1111/mafi.12379

MATHEMATICAL FINANCE An International Journal of Mathematics, Statistics and Financial Economics WILEY

ORIGINAL ARTICLE

Preference robust distortion risk measure and its application

Wei Wang[1,2] | **Huifu Xu**[1]

[1]Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong

[2]School of Business, University of Southampton, Southampton, UK

Correspondence
Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong
Email: hfxu@se.cuhk.edu.hk

Funding information
CUHK direct grant

Abstract

Distortion risk measure (DRM) plays a crucial role in management science and finance particularly actuarial science. Various DRMs have been introduced but little is discussed on which DRM at hand should be chosen to address a decision maker's (DM's) risk preference. This paper aims to fill out the gap. Specifically, we consider a situation where the true distortion function is unknown either because it is difficult to identify/elicit and/or because the DM's risk preference is ambiguous. We introduce a preference robust distortion risk measure (PRDRM), which is based on the worst-case distortion function from an ambiguity set of distortion functions to mitigate the impact arising from the ambiguity. The ambiguity set is constructed under well-known general principles such as concavity and inverse S-shapedness of distortion functions (overweighting on events from impossible to possible or possible to certainty and underweighting on those from possible to more possible) as well as new user-specific information such as sensitivity to tail losses, confidence intervals to some lotteries, and preferences to certain lotteries over others. To calculate the proposed PRDRM, we use the convex and/or concave envelope of a set of points to characterize the curvature

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Mathematical Finance* published by Wiley Periodicals LLC.

of the distortion function and derive a tractable reformulation of the PRDRM when the underlying random loss is discretely distributed. Moreover, we show that the worst-case distortion function is a nondecreasing piecewise linear function and can be determined by solving a linear programming problem. Finally, we apply the proposed PRDRM to a risk capital allocation problem and carry out some numerical tests to examine the efficiency of the PRDRM model.

KEYWORDS
preference robust distortion risk measure, risk capital allocation, worst-case distortion function, Yaari's dual theory of choice

1 | INTRODUCTION

Distortion risk measure (DRM), a widely used approach for pricing insurance risks, is a risk-adjusted expected value where the underlying distribution of the risk is modified by a distortion function (or equivalently a probability weighting function) reflecting the decision maker's (DM's) risk preference.[1] Since its introduction by Denneberg (1990) on premium calculation, DRM has received extensive attention in that it enjoys some important properties such as monotonicity, positive homogeneity, translation invariance, comonotonic additivity, and as such it covers a wide range of monetary risk measures including VaR and CVaR. In actuarial science, a number of popular premium principles such as Gini's principle (Denneberg, 1990), Wang's proportional hazards principle (Wang, 1995), and the cumulative residual entropy principle (Sordo et al., 2016), are all proposed based on DRM. In the case when the probability weighting function is absolutely continuous, DRM can be reformulated as the weighted average of the quantile function of random loss (Dhaene et al., 2012). In particular, if the distortion function is concave, then DRM reduces to the well-known *spectral risk measure* (SRM) (Acerbi, 2002) where the derivative of the weighting function is called the *risk spectrum*, see, for example, Yaari (1987), Tsanakas and Desli (2003), and Henryk and Silvia (2006). Moreover, under some mild conditions, DRM/SRM can be used as a basis for representing any law invariant coherent risk measures, see, for example, Pichler and Shapiro (2015) and Wang and Xu (2020).

In the current research on behavioral economics and risk management, the distortion function is assumed to be known or can be elicited through a tolerable amount of introspective questioning (Bleichrodt & Pinto, 2000; Tversky & Wakker, 1995). There have been a lot of efforts on how to identify/elicit a probability weighting function for a DM including the design of surveys and experiments for assessing the decision weights and measuring beliefs under uncertainty, see, for example, Wu and Gonzalez (1996), Abdellaoui et al. (2021), and references therein. In practice, however, there may be some ambiguity about a DM's risk preference, which means that there is not a clear-cut distortion function to characterize his/her risk preference precisely. Such ambiguity may arise from a lack of accurate description of human behavior, cognitive difficulty, or incomplete information. The ambiguity may also arise from decision-making problems, which

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

involve several stakeholders who fail to reach a consensus. Under these circumstances, it might be sensible to construct a set of plausible distortion functions representing the DM's risk attitude and then base the optimal decision on the worst-case distortion function from the ambiguity set. This is the fundamental idea of the *preference robust optimization* (PRO) model.

Maccheroni (2002) seems to be the first to consider the worst-case utility evaluation among many available utilities when a conservative DM faces uncertain outcomes of lotteries. He derives necessary and sufficient conditions for the existence of this kind of robust decision-making framework. Armbruster and Delage (2015) give a comprehensive treatment of the problem from an optimization perspective by formally proposing a maximin PRO paradigm. Specifically, they consider a class of utility functions, which are concave or S-shaped and discuss how a DM's preference may be elicited through pairwise comparisons. Moreover, they demonstrate that solving the PRO is down to solving a linear programming problem under some mild conditions. Over the past few years, there has been increasing attention to the PRO models, see Hu and Mehrotra (2015), Haskell et al. (2016), Guo and Xu (2018), and Zhang et al. (2020) for the recent developments. Delage and Li (2017) first consider a PRO model for monetary risk measures where the DM's choice of risk measure is ambiguous. They construct the ambiguity set of convex risk measures by imposing some additional information such as scale invariance (also known as positive homogeneity), law invariance, and preference elicitation information from pairwise comparison, and then develop tractable reformulations using the acceptance set representation of convex risk measure established by Föllmer and Schied (2002). Delage et al. (2022) consider a PRO model for shortfall risk measures where the ambiguity set of convex loss functions is characterized using risk preference information such as certainty equivalent and sensitivity to tail events and give the tractable reformulations for the minimax optimization problem when the underlying random loss is discretely distributed. More recently, Wang and Xu (2020) consider a PRO model for decision-making problems where the DM's risk preference is ambiguous and can be characterized via a risk spectrum. They propose a *robust spectral risk model* (RSRM) where the ambiguity of the DM's risk preference is defined by a ball of risk spectra centered at a nominal risk spectrum. Guo and Xu (2021) complement the model by considering pairwise comparison approaches for constructing the ambiguity set and demonstrating how worst-case risk spectra converge to the truth as more information about the DM's preference is elicited.

In this paper, we follow up this stream of research by introducing a *preference robust distortion risk measure* (PRDRM) to address a situation where the DM's risk preference can be characterized by DRM and there is incomplete information about the true distortion function. Differing from the utility PRO models, which are fundamentally built on Von Neumann–Morgenstein's expected utility theory (see, e.g., Von Neumann et al. (2007)), our PRDRM model is based on Yaari's dual theory of choice (Yaari, 1987). We do so not only because DRMs are naturally based on the latter but also because there is a broad agreement among psychologists and economists that probabilities in many decision-making problems are not linear, see, for instance, Allais paradox (Allais, 1953) and Ellsberg paradox (Ellsberg, 1961). The main contributions of this paper can be summarized as follows.

First, we propose a PRDRM model where a DM's risk preference can be represented by a distortion function but information about such a function is incomplete. To hedge the risk arising from the ambiguity, we consider the worst-case distortion function from a set of plausible distortion functions for the calculation of distortion risk. In the case when the distortion functions are concave and differentiable, the new PRDRM model recovers the preference robust SRM models (Wang and Xu, 2020; Guo and Xu, 2021). Since distortion functions are not necessarily concave or differentiable, the PRDRM model covers a much larger class of problems where the DM's risk

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

preference can be represented by a DRM instead of a spectral risk measure. Second, we derive a tractable reformulation to calculate the proposed PRDRM by characterizing the concave or inverse S-shaped distortion function through the concave and/or convex envelope over a discrete set of points. In particular, we show that the worst-case distortion function is a nondecreasing piecewise linear function when the underlying random loss is discretely distributed and demonstrate through some simple examples how such a worst-case piecewise linear function can be identified. Third, we apply the proposed PRDRM to one-stage stochastic decision-making problems and derive tractable formulations for solving the latter when distortion functions are concave. We then apply the proposed robust model and computational scheme to a risk capital allocation problem and report some numerical results.

The rest of the paper is organized as follows. Section 2 describes the PRDRM model, Section 3 details the construction of the ambiguity set, and Section 4 develops tractable reformulations for the proposed PRDRMs. Section 5 applies the proposed PRDRM to optimal decision-making problems and proposes an alternating iterative algorithm for solving the latter. Section 6 reports the application of the PRDRM to a risk capital allocation problem and some numerical test results. Some technical details and supplementary results or statements are given in the appendices.

2 | PROBLEM SETUP

Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ be a (non-negative) random variable, where Ω is a sample space with sigma algebra $\mathcal{F}$ and $\mathbb{P}$ is a reference probability measure. Throughout the paper, X represents financial loss and the smaller value is preferred. For instance, X may model the claims of an insurance contract from the insurer's point of view. Let $\mathcal{L}^0$ denote the set of all random variables mapping from Ω to $\mathbb{R}$. For each $X \in \mathcal{L}^0$, let $F_X(x) := \mathbb{P}[X \le x]$ be its cumulative distribution function (CDF) and $S_X(x) := 1 - F_X(x)$ its survival function. In the literature of economics, X is also called "act" and F_X is called "lottery." For a given confidence level $\alpha \in (0, 1)$, the generalized inverse of the CDF for X is defined as $F_X^{\leftarrow}(\alpha) := \inf\{x \in \mathbb{R} : F_X(x) \ge \alpha\}$. $F_X^{\leftarrow}(\alpha)$ is also known as the value at risk (VaR) or as a quantile function mapping from (0,1) to $\mathbb{R}$ if α is viewed as a variable.

Let $X \in \mathcal{L}^0$ be a random loss and $g : [0, 1] \to [0, 1]$ be *a distortion function*, that is, g is a nondecreasing function with $g(0) = 0$ and $g(1) = 1$. Throughout the paper, we use $\mathcal{G}$ to denote the set of all distortion functions. The DRM induced by g is defined as

$$\rho_g(X) := \int_0^\infty g(S_X(x))dx + \int_{-\infty}^0 [g(S_X(x)) - 1]dx, \tag{1}$$

provided that at least one of the two integrals is finite. Equation (1) is a special Choquet integral if we regard $g \circ \mathbb{P}$ as a "measure" over measurable space $(\Omega, \mathcal{F})$ in which case g may be interpreted as a "distortion" function of $\mathbb{P}$. By the properties of Choquet integral, DRM is a law invariant, positively homogeneous, monotone and comonotonic additive risk measure. In addition, it is easy to verify that $\rho_g(X)$ is a coherent risk measure (satisfying monotonicity, subadditivity, positive homogeneity, and translation invariance (Artzner et al., 1999)) if and only if g is concave, see, for example, Wirch and Hardy (2001) and Acerbi (2002). Furthermore, when the distortion function g is left-continuous, Equation (1) can be alternatively represented as

$$\rho_g(X) = \int_{\mathbb{R}} x d\tilde{g}(F_X(x)) = \int_0^1 F_X^{\leftarrow}(t) d\tilde{g}(t) = \int_0^1 F_X^{\leftarrow}(1-t) dg(t), \tag{2}$$

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where $\tilde{g}(t) = 1 - g(1-t)$ and the integral is understood as the Lebesgue-Stieltjes integral, see, for example, Dhaene et al. (2012). Similar representations can be easily derived when the distortion function g is right-continuous (Dhaene et al., 2012). The first equality in Equation (2) shows that the probability of loss X is modified by $\tilde{g}$ and the second equality interprets $\rho_g(X)$ as the weighted average of the quantile losses, which is also known as the *spectral risk measure* when $\tilde{g}$ is differentiable (Acerbi, 2002). In this paper, our focus will be on the case that X is a non-negative random loss although all of the results developed in this paper can be applied to the random variables without this restriction. Note that when X is non-negative, DRM can be seen as a special loss-based risk measure introduced by Cont et al. (2013). In Appendix A.1, we list a few well-known DRMs in the literature.

2.1 | Preference robust distortion risk measure

In the current research on behavioral economics and risk management, the distortion function in DRM is assumed to be known or can be easily elicited by a tolerable number of questionnaires. In practice, however, identifying/eliciting a unique distortion function, which precisely characterizes a DM's risk attitude, may turn out to be difficult either because the decision-making problem involves several stakeholders who fail to reach a consensus or because the prospect space is too complex and/or too large to elicit such a distortion function. Under these circumstances, it might be sensible to use partially available information to construct a set of distortion functions to reflect such uncertainty of a DM's risk attitude and then consider the worst-case distortion function from the ambiguity set for the DRM to mitigate the risk arising from the ambiguity of true distortion function. Here we give a formal definition of preference robust DRM.

Definition 2.1 (Preference robust DRM). Let $\mathcal{G}'$ denote the ambiguity set of distortion functions constructed from the available preference information of a DM. The *preference robust distortion risk measure* of a random loss X based on $\mathcal{G}'$ is defined as

$$\text{(PRDRM)} \quad \rho_{\mathcal{G}'}(X) := \sup_{g \in \mathcal{G}'} \rho_g(X), \tag{3}$$

which is the worst-case DRM computed from the set of distortion functions $\mathcal{G}'$.

Let g^* denote the true distortion function. In the forthcoming discussions, we will construct the ambiguity set through queries for the DM and the DM's responses to the queries are consistent. We refer readers to Armbruster and Delage (2015) for PRO models with inconsistent response cases. The definition reflects the DM's conservatism on the risk in the absence of complete information about g^*. In the case when the distortion functions g in the set $\mathcal{G}'$ are all continuous and piecewise differentiable, the PRDRM coincides with the RSRM in Wang and Xu (2020). By definition, we know immediately that PRDRM is a law invariant risk measure. When g is a concave function, $\rho_g(X)$ is a coherent risk measure. In this case, PRDRM is also a coherent risk measure when all distortion functions in the ambiguity set are concave because the "sup" operation preserves the four axiomatic properties in the definition of coherent risk measure. Conversely, any law invariant coherent risk measure can be represented as a PRDRM. We refer readers to Wang and Xu (2020) for the related discussions.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2 | Optimization with PRDRM

As a motivation as well as an application, we will apply the proposed PRDRM to optimal decision-making problems where the optimal decision is determined by minimizing the PRDRM, that is,

$$\text{(PRDRM-Opt)} \quad \min_{z\in Z} \max_{g\in\mathcal{G}'} \rho_g(f(z,\xi)), \tag{4}$$

where $f : Z \times \mathbb{R}^k \to \mathbb{R}$ is a continuous loss function, z is a vector of decision variables, Z is a compact convex set in $\mathbb{R}^n$ and ξ is a vector of random variables. When $\mathcal{G}'$ is compact with respect to some weak topology, the inner maximum in (PRDRM-Opt) can be attained because $\rho_g(\cdot)$ is continuous, in this case, the outer minimum can also be attained because $\rho_{\mathcal{G}'}(f(z,\xi))$ is continuous in z and Z is compact. Moreover, if $f(z,\xi)$ is convex in z for every fixed ξ, then $\rho_g(f(z,\xi))$ is convex in z given the fact that $\rho_g(\cdot)$ is monotone increasing. In this case, the optimal value of the inner maximization problem of (PRDRM-Opt) is a convex function of z and hence the outer minimization problem of (PRDRM-Opt) is a convex program.

3 | AMBIGUITY SET OF THE DISTORTION FUNCTIONS

The nature of the proposed PRDRM is determined by the information structure of the ambiguity set. In general, the ambiguity set should comprise two types of information: (a) generic information such as being risk-averse, which is widely considered in the literature on behavioral economics and being inverse S-shaped, which is observed by many empirical tests, and (b) user-specific information such as preferring a lottery to another one, having "confidence" intervals to some lotteries and being sensitive to tail losses. The former can be identified through subjective judgment whereas the latter requires some careful elicitation process. In this section, we discuss how to construct the ambiguity set.

3.1 | Characterization of distortion functions using risk preference information

In this subsection, we discuss various approaches for eliciting a DM's risk preference and using them to construct an ambiguity set of distortion functions. In particular, we will investigate how to account for information about preferring certain lotteries over other lotteries, about having some "confidence" intervals for the risks of a list of random variables/prospects, about how sensitive the DM is regarding events that occur in the tail of prospects, about whether the DRM is coherent, and finally about the lower subadditivity and upper subadditivity, which are associated with the possibility effect and certainty effect.

1. Pairwise comparison. Let $\{G_m, B_m\}_{m=1}^M$ be a set of comparable prospects. The set $\mathcal{G}_{pair}$ denotes all of the distortion functions, which satisfy $\rho_g(G_m) \le \rho_g(B_m)$ for $m = 1, \dots, M$, that is,

$$\mathcal{G}_{pair} := \{g \in \mathcal{G} : \rho_g(G_m) \le \rho_g(B_m), \text{ for } m = 1, \dots, M\}. \tag{5}$$

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In this approach, we presume that the DM prefers G_m over B_m for $m = 1, \dots, M$ in order to ensure that the constructed ambiguity set of $\mathcal{G}_{pair}$ is not empty. Otherwise, the constructed pair of prospects should be deleted.

2. Certainty equivalent. Let $\{W_k\}_{k=1}^K$ be a list of random variables with an associated set of "confidence" intervals $[w_k^-, w_k^+] \subseteq [\text{ess inf } W_k, \text{ess sup} W_k]$ for the "certainty equivalent" of each W_k. The set $\mathcal{G}_{ce}$ denotes all of the distortion functions, which evaluate the risk of each W_k to be larger than w_k^- and lower than w_k^+, that is,

$$\mathcal{G}_{ce} := \left\{ g \in \mathcal{G} : w_k^- \le \rho_g(W_k) \le w_k^+, \text{ for } k = 1, \dots, K \right\}. \tag{6}$$

Note that a natural method that can be used to identify such two bounds for the riskiness of a random variable W_k would take the form of questions such as:

- Upper bound w_k^+: "What is the smallest amount of money that you would decline to pay instead of being exposed to the risk of W_k?"
- Lower bound w_k^-: "What is the largest amount of money that you would be willing to pay instead of being exposed to the risk of W_k?"

From the above questions, we can see that $w_k^- \le w_k^+$. Moreover, when the answers to both questions are such that $w_k^+ = w_k^- = \overline{w}_k$, it implies that we have identified the certainty equivalent of W_k, that is, $\rho_g(W_k) = \overline{w}_k$, yet in practice, it is more often the case that only an interval $[w_k^-, w_k^+]$ will be obtained for this value. The certainty equivalent approach is well-known in the literature of PRO, see Delage et al. (2022) and references therein.

To analyze how sensitive the distortion function is to a small probability, we consider the following class of distortion functions based on Delage et al. (2022).

3. Sensitivity to large losses with a small probability. Let $\varphi : (0,1) \to \mathbb{R}_+$ be a nonincreasing convex function and $\{Z_\varphi^\epsilon\}_{\epsilon<1}$ be a set of elementary lotteries with $\mathbb{P}(Z_\varphi^\epsilon = \varphi(\epsilon)) = \epsilon$ and $\mathbb{P}(Z_\varphi^\epsilon = 0) = 1 - \epsilon$. Denote by $\mathcal{G}_{bnd}(\varphi)$ the set of distortion functions where for each $g \in \mathcal{G}_{bnd}(\varphi)$, ρ_g assigns to each random variable in the set $\{Z_\varphi^\epsilon\}_{\epsilon<1}$ a value of the risk that is lower than the risk of a certain loss of one:

$$\mathcal{G}_{bnd}(\varphi) := \left\{ g \in \mathcal{G} : \rho_g(Z_\varphi^\epsilon) = \varphi(\epsilon) g(\epsilon) \le 1, \text{ for } \epsilon \le \epsilon_0 \right\}, \tag{7}$$

where ϵ_0 is a very small positive number. Typically, ϵ is small whereas $\varphi(\epsilon)$ is large. The distortion functions defined as such effectively provide an upper bound for the true distortion function with small probabilities, that is, $g(\epsilon) \le \frac{1}{\varphi(\epsilon)}$ for $\epsilon \le \epsilon_0$. For example, if we choose $\varphi(\epsilon) = -\frac{1}{\ln(1-\epsilon)}$, then we would have $g(\epsilon) \le -\ln(1-\epsilon)$ for $\epsilon \le \epsilon_0$. Practically speaking, identifying φ can be as difficult as identifying the distortion function g. A general guideline is to require $\varphi(t)t \le 1$ for $t \in (0,1)$ and $\lim_{t \to 0^+} \varphi(t) = \infty$. For instance, we may set $\varphi(t) = \frac{1}{t^\alpha}$ for $\alpha \in (0,1)$.

We now discuss the structural properties of distortion functions. Wirch and Hardy (2001) show that a DRM (ρ_g) is subadditive if and only if g is a concave function, see Theorem 2.2 there. Since DRM is monotone, positively homogeneous and translation invariant, it means that ρ_g is coherent if and only if g is concave. Moreover, they relate the concavity of the distortion function g to the concept of second-order stochastic dominance (Hanoch & Levy, 1975), see Wirch and Hardy (2001, Theorem 3.1). The equivalence between the concavity of g and coherence of ρ_g motivates us to pay special attention to PRDRM with concave distortion functions.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4. Concave distortion functions. Let $\mathcal{G}_{coh}$ be the set of distortion functions such that the induced DRMs are coherent, that is,

$$\mathcal{G}_{coh} := \{g \in \mathcal{G} : g \text{ is concave on } [0,1]\}. \tag{8}$$

In the literature on economics, many empirical tests show that the distortion function has the so-called lower subadditivity and upper subadditivity, see, for example, Tversky and Wakker (1995), Prelec (1998), and Bleichrodt and Pinto (2000). The *lower subadditivity* of the distortion function means that a lower interval $[0, q]$ has more impact on a DM than an intermediate interval $[p, p+q]$, provided that $p+q$ is bounded away from one. Alternatively stated, lower subadditivity says that a change from impossible to possible has a stronger impact on an individual's decision than an equal change from possible to more possible. This effect is referred to as the *possibility effect*. The *upper subadditivity* of the distortion function means that an upper interval $[1-q, 1]$ has more impact than an intermediate interval $[p, p+q]$, provided that p is bounded away from zero. Hence, a change from possible to certain has more impact than an equal change from possible to more possible. This effect is referred to as the *certainty effect*. The effect of lower subadditivity and upper subadditivity is to produce an inverse S-shaped probability weighting function, overweighting small and high probabilities and underweighting intermediate probabilities (Bleichrodt and Pinto, 2000). This motivates us to consider PRDRM with S-shaped distortion functions.

5. S-shaped distortion functions. Let $\tilde{p} \in (0,1)$ be a fixed turning point and $\mathcal{G}_S$ be the set of all inverse S-shaped distortion functions with the fixed turning point (from concavity to convexity),

$$\mathcal{G}_S := \{g \in \mathcal{G} : g(t) \text{ is concave on } [0, \tilde{p}] \text{ and convex on } [\tilde{p}, 1]\}. \tag{9}$$

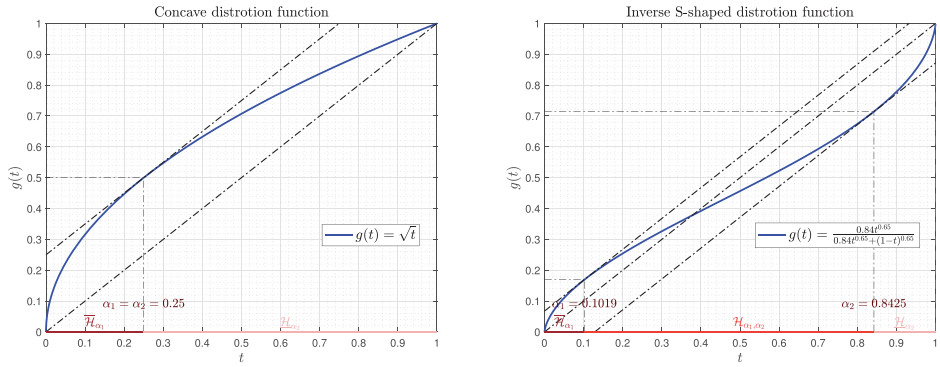
There are many classes of distortion functions that are inverse S-shaped for certain parameter values and we list the most common ones with specified values of the parameters in Table 1 based on Bleichrodt and Pinto (2000) and Baillon et al. (2020).

TABLE 1 Inverse S-shaped parametric distortion functions and their parametric estimates

Abbr	Functional form	References and parametric estimates
TK	$g(t) = \frac{t^\alpha}{[t^\alpha + (1-t)^\alpha]^{\frac{1}{\alpha}}}$, $\alpha \in [0.28, 1)$	Tversky and Kahneman (1992): $\alpha = 0.69$ (losses), $\alpha = 0.61$ (gains) Wu and Gonzalez (1996): $\alpha = 0.71$ (gains) Abdellaoui (2000): $\alpha = 0.70$ (losses), $\alpha = 0.60$ (gains)
GE	$g(t) = \frac{\beta t^\alpha}{\beta t^\alpha + (1-t)^\alpha}$, $\alpha, \beta > 0, \alpha < 1$	Goldstein and Einhorn (1987) Wu and Gonzalez (1996): $\alpha = 0.68, \beta = 0.84$ (gains) Gonzalez and Wu (1999): $\alpha = 0.44, \beta = 0.77$ (gains) Tversky and Wakker (1995): $\alpha = 0.69, \beta = 0.77$ (gains) Abdellaoui (2000): $\alpha = 0.60, \beta = 0.65$ (gains); $\alpha = 0.65, \beta = 0.84$ (losses)
Prelec	$g(t) = \exp(-\beta(-\ln p)^\alpha)$, $\alpha, \beta > 0$	Prelec (1998): $\alpha = 0.65, \beta = 1$ (losses) Wu and Gonzalez (1996): $\alpha = 0.74, \beta = 1$ (gains)

"Losses" means random loss, whereas "gains" means random gain.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) (b)

FIGURE 1 The graphical illustration of distortion function: (a) concave case and (b) inverse S-shaped case. [Color figure can be viewed at wileyonlinelibrary.com]

Example 3.1. To see how the shape of the distortion function is closely related to the DM's risk preference, we give a simple example to illustrate. Let $X \in \mathcal{L}^0$ be fixed and $\alpha_1, \alpha_2 \in (0,1)$ with $\alpha_1 \le \alpha_2$. Define the higher risk event, medium risk event, and lower risk event as the follows:

$$\overline{\mathcal{H}}_{\alpha_1} := \left\{\omega \in \Omega : X(\omega) > F_X^{\leftarrow}(1-\alpha_1)\right\},$$

$$\mathcal{H}_{\alpha_1,\alpha_2} := \left\{\omega \in \Omega : F_X^{\leftarrow}(1-\alpha_2) < X(\omega) \le F_X^{\leftarrow}(1-\alpha_1)\right\},$$

and

$$\underline{\mathcal{H}}_{\alpha_2} := \left\{\omega \in \Omega : X(\omega) \le F_X^{\leftarrow}(1-\alpha_2)\right\}.$$

Consider the case that $\alpha_1 = \alpha_2$. If the DM is risk-averse, then s/he would put more weight on the higher risk event $\overline{\mathcal{H}}_{\alpha_1}$ and less weight on the lower risk event $\underline{\mathcal{H}}_{\alpha_2}$. In this case, g is concave. Figure 1 a gives a graphical illustration of such a distortion function. From the figure, we can see that the interval [0,0.25] on the t-axis is the range of the survival function values over the higher risk event $\overline{\mathcal{H}}_{\alpha_1}$. Since $g'(t) \ge 1$ over the interval, the probability of X is scaled-up. In other words, the DM overweighs the event $\overline{\mathcal{H}}_{\alpha_1}$. Likewise, the distortion function scales down the probability of X over [0.25,1] (which corresponds to $\underline{\mathcal{H}}_{\alpha_2}$). In this case, the DM underweighs the risk of event $\underline{\mathcal{H}}_{\alpha_2}$.

Let us now consider the case when $\alpha_1 < \alpha_2$. If the DM is rational, then her/his distortion function would have the lower subadditivity and upper subadditivity as discussed earlier, that is, the DM puts more weight on the high risk event $\overline{\mathcal{H}}_{\alpha_1}$ and the lower risk event $\underline{\mathcal{H}}_{\alpha_2}$ and less weight on the medium risk event $\mathcal{H}_{\alpha_1,\alpha_2}$. This kind of risk preference leads to an inverse S-shaped distortion risk function. Figure 1b gives a graphical illustration of such a distortion function. From the figure, we can see that the distortion function is steep over the intervals [0,0.1019] and [0.8425, 1], which correspond to the ranges of the survival function over $\overline{\mathcal{H}}_{\alpha_1}$ and $\underline{\mathcal{H}}_{\alpha_2}$. Hence, the underlying probability of the random loss X over these ranges is scaled-up, which means that the DM overweighs the events $\overline{\mathcal{H}}_{\alpha_1}$ and $\underline{\mathcal{H}}_{\alpha_2}$. The distortion function is less steep over [0.1019, 0.8425], which corresponds to $\mathcal{H}_{\alpha_1,\alpha_2}$ and the underlying probability of a random loss X is scaled-down. Consequently, the DM underweighs the event.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The slope of the tangent line at a point $(t_0, g(t_0))$ reflects a DM's marginal distortion at loss x_0 ($S_X(x_0) = t_0$), which is a scaling of the probability of the loss. The steeper the slope is, the more significant will be the scaling. At point $t = 0$ (which corresponds to the loss ess sup X), the slope might be infinity, which means that a very large scaling is posed on a very small probability of large tail loss. If the slope of the graph of g at some points or over some intervals is zero, then the probability of loss is scaled to zero. For example, in the CVaR case (see Example A.1(ii)), the related distortion function (g_α) poses a constant scaling $\frac{1}{1-\alpha}$ over the underlying probability of higher risk event $\overline{\mathcal{H}}_\alpha$ and 0 over the rest.

3.2 | Ambiguity set of distortion functions

In this paper, we will consider the following two types of ambiguity sets of distortion functions for deriving the tractable reformulations of the proposed PRDRM.

$$\mathcal{G}_1 = \mathcal{G}_{coh} \cap \mathcal{G}_{pair} \cap \mathcal{G}_{ce} \cap \mathcal{G}_{bnd}; \quad \mathcal{G}_2 = \mathcal{G}_S \cap \mathcal{G}_{pair} \cap \mathcal{G}_{ce} \cap \mathcal{G}_{bnd}.$$

The reason why we specify the ambiguity set $\mathcal{G}_1$ is as follows: coherent risk measures are widely used in finance both for theoretical analysis and practical use, by choosing $\mathcal{G}_{coh}$, we restrict our attention to a specific class of DMs whose risk attitude can be characterized by concave distortion functions; the other three specifications are related to elicitation process: "interval confidence" of a list of lotteries can be easily estimated, pair-wise preference information can be easily obtained in practice, and sensitivity to the tail prospects is always necessary to describe a DM's risk attitude towards tail events. The reason why we choose $\mathcal{G}_2$ is that many empirical studies eliciting the probability weighting function indicate that the distortion function, which characterizes a DM's risk attitude, is always inverse S-shaped. Therefore, we believe that these two choices of ambiguity set of distortion functions would have a great impact both in theory and application.

4 | TRACTABLE REFORMULATION FOR PRDRM

In this section, we derive tractable reformulations for the proposed PRDRM based on the ambiguity sets $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively. To this end, we first consider how to calculate DRM for a discrete random variable and then derive the tractable reformulations.

4.1 | Computation of DRM

We discuss how to compute DRM for a non-negative discretely distributed random loss. For a continuously distributed random variable X, we can use sample average approximation to discretize the probability distribution and then calculate DRM approximately with the latter, see, for example, Acerbi (2002) and Guo and Xu (2021).

Let X be a non-negative finitely distributed random loss with $\mathbb{P}(X = x_i) = p_i$ for $i = 1, \dots, n$, where $0 \le x_1 < x_2 < \cdots < x_n$. For the simplicity of notation, let $x_0 = 0$, $x_{n+1} = \infty$, and $\pi_0 = 0$. The CDF of X is a step-like function with n breakpoints, that is, $F_X(x) = \pi_i$, when $x \in [x_i, x_{i+1})$, for $i = 0, 1, \dots, n$, where $\pi_i = \sum_{j \le i} p_j$ for $i = 1, 2, \dots, n$. Consequently we follow from

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Equation (1) to write the DRM of X as

$$\rho_g(X) = \int_0^\infty g(S_X(x))dx = \sum_{i=1}^n \int_{x_{i-1}}^{x_i} g(S_X(x))dx = \sum_{i=1}^n (x_i - x_{i-1})g(1-\pi_{i-1})$$
$$= \sum_{i=1}^n x_i[g(1-\pi_{i-1}) - g(1-\pi_i)] = \sum_{i=1}^n x_i\phi_i, \tag{10}$$

where $\phi_i = g(1-\pi_{i-1}) - g(1-\pi_i) \geq 0$ for $i = 1, \ldots, n$ and $\sum_{i=1}^n \phi_i = 1$. The last equation in Equation (10) also confirms the fact that $\rho_g(X)$ is a weighted expected value of X.

4.2 | Tractable reformulation of PRDRM under $\mathcal{G}_1$

In this subsection, we consider tractable reformulation of PRDRM when the ambiguity set of distortion functions is constructed as follows:

$$\mathcal{G}_1 = \mathcal{G}_{coh} \cap \mathcal{G}_{pair} \cap \mathcal{G}_{ce} \cap \mathcal{G}_{bnd} = \left\{ g \in \mathcal{G} : \begin{array}{c} w_k^- \leq \rho_g(W_k) \leq w_k^+; \text{ for } k = 1, \ldots, K; \\ \rho_g(G_m) \leq \rho_g(B_m); \text{ for } m = 1, \ldots, M; \\ g(\epsilon) \leq \dfrac{1}{\varphi(\epsilon)}, \epsilon \leq \epsilon_0, \text{and } g \text{ is concave.} \end{array} \right\}, \tag{11}$$

where ϵ_0 is a very small positive number.

To derive a tractable reformulation of PRDRM with $\mathcal{G}' = \mathcal{G}_1$, we need to establish an intermediate technical issue. Let $\Theta := \{\theta_1, \ldots, \theta_J\} \subset \mathbb{R}$ be a discrete set in $\mathbb{R}$ where $\theta_1 < \cdots < \theta_J$ and $\upsilon_1, \ldots, \upsilon_J \in \mathbb{R}$ be a set of numbers such that

$$l(\theta_j) = \upsilon_j, \text{ for } j = 1, \ldots, J \tag{12}$$

for some nondecreasing concave function $l(t)$ mapping from $\mathbb{R}$ to $\mathbb{R}$. Since $l(\cdot)$ is not unique, let $\mathcal{L}_\upsilon$ denote the set of all such functions, where we write υ for $(\upsilon_1, \ldots, \upsilon_J)$. Next, we consider points $t_i \in \mathbb{R}$, for $i = 1, \ldots, n$ (which can be any but must be fixed) and ask ourselves how we may find a function $l^* \in \mathcal{L}_\upsilon$ such that

$$\sum_{i=1}^n a_i l^*(t_i) = \sup_{l \in \mathcal{L}_\upsilon} \sum_{i=1}^n a_i l(t_i), \tag{13}$$

where a_i, for $i = 1, \ldots, n$ are fixed positive constants. The next lemma states that such l^* exists, has a piecewise linear structure and can be obtained from solving a linear program.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lemma 4.1. *Consider the following linear program:*

$$\sup_{s,\beta} \sum_{i=1}^{n} a_i s_i \tag{14a}$$

$$\text{s.t. } v_j + \beta_j(t_i - \theta_j) \geq s_i, \text{ for } j = 1, \dots, J; i = 1, \dots, n, \tag{14b}$$

$$v_j + \beta_j(\theta_{j+1} - \theta_j) \geq v_{j+1}, \text{ for } j = 1, \dots, J-1, \tag{14c}$$

$$v_j + \beta_{j+1}(\theta_{j+1} - \theta_j) \leq v_{j+1}, \text{ for } j = 1, \dots, J-1, \tag{14d}$$

$$\beta_j \geq 0, \text{ for } j = 1, \dots, J. \tag{14e}$$

Let ϑ^* denote its optimal value. Then there exists a piecewise linear concave function $l^* \in \mathcal{L}_v$ such that $\vartheta^* = \sum_{i=1}^{n} a_i l^*(t_i)$.

Proof. The thrust of the proof is to show problems (13) and (14) have the same optimal value. We proceed the proof by showing that any feasible solution of the former can be used to construct a feasible solution of the latter and conversely an optimal solution of the latter can be used to construct a feasible solution of the former. Note that $\mathcal{L}_v \neq \emptyset$ (which is our setup), β_j is lower bounded by 0 and upper bounded by constraints (14d). The latter ensures that s_i is upper bounded via inequalities (14b). Together with the fact that $a_i \geq 0$ for $i = 1, \dots, n$, this further guarantees that $\vartheta^* < \infty$ and the optimal value of the maximization problem (14) is attainable.

Let $l \in \mathcal{L}_v$. For $j = 1, \dots, J$, define the support functions $h_j(\theta) := v_j + \beta_j(\theta - \theta_j)$, which majorizes l at θ_j. By the concavity and nondecreasing property of l, the parameters (slopes) β_j, for $j = 1, \dots, J$, must satisfy inequalities (14c)–(14e). Moreover, the majorization property ensures that $h_j(t) \geq l(t)$ for any $t \in \mathbb{R}$. Let $s_i := l(t_i), i = 1, \dots, n$. Then we have

$$h_j(t_i) = v_j + \beta_j(t_i - \theta_j) \geq s_i, \text{ for } j = 1, \dots, J; i = 1, \dots, n$$

which give rise to constraint (14b). The discussions above imply that for $s := (s_1, \dots, s_n)$ and $\beta := (\beta_1, \dots, \beta_J)$ chosen as such, (β, s) forms a feasible solution to problem (14). This shows

$$\vartheta^* = \sup_{\beta,s} \sum_{i=1}^{n} a_i s_i \geq \sum_{i=1}^{n} a_i l(t_i) \text{ and hence } \vartheta^* \geq \sup_{l \in \mathcal{L}_v} \sum_{i=1}^{n} a_i l(t_i). \tag{15}$$

Conversely, let (β^*, s^*) be an optimal solution to problem (14). Then $\vartheta^* = \sum_{i=1}^{n} a_i s^*$. Let $h_j^*(\theta) = v_j + \beta_j^*(\theta - \theta_j)$, for $j = 1, \dots, J$ and $l^*(t) := \min_{j=1,\dots,J} h_j^*(t)$. Since $\beta_j^* \geq 0$, then l^* is a nondecreasing piecewise linear concave function. In what follows, we show

$$l^*(\theta_j) = v_j, \text{ for } j = 1, \dots, J, \text{ and } l^*(t_i) = s_i^*, \text{ for } i = 1, \dots, n. \tag{16}$$

Observe first that β_j^* satisfies

$$v_j + \beta_j^*(\theta_k - \theta_j) \geq v_k, \text{ for } k = 1, \dots, J;\ j = 1, \dots, J. \tag{17}$$

To see this, we first show Equation (17) for fixed j and $k = j+1, \dots, J$. We do so by induction. The inequality holds when $k = j+1$ because it coincides with constraint (14c). Assume now the

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

inequality holds for $k = k' \geq j+1$, thar is, $v_j + \beta_j^*(\theta_{k'} - \theta_j) \geq v_{k'}$. We show it also holds for $k = k'+1$. Since $\beta_{k'}^* \leq \beta_j^*$ (the monotonicity of β_j^* is implied by a combination of constraints (14c) and (14d) and θ_j is in an increasing order), then

$$\begin{aligned} v_j + \beta_j^*(\theta_{k'+1} - \theta_j) &= v_j + \beta_j^*(\theta_{k'+1} - \theta_{k'}) + \beta_j^*(\theta_{k'} - \theta_j) \\ &\geq \beta_{k'}^*(\theta_{k'+1} - \theta_{k'}) + v_j + \beta_j^*(\theta_{k'} - \theta_j) \\ &\geq v_{k'+1} - v_{k'} + v_j + \beta_j^*(\theta_{k'} - \theta_j) \quad \text{(by (14c))} \\ &\geq v_{k'+1}. \quad \text{(by assumption)} \end{aligned}$$

Likewise, we can also show Equation (17) holds for $k = 1, \ldots, j-1$ using constraint (14d) and monotonicity of β_j^*. From Equation (17), we have

$$h_j^*(\theta_{j'}) = v_j + \beta_j^*(\theta_{j'} - \theta_j) \geq v_{j'}, \text{ for } j' = 1, \ldots, J, \text{ and } h_{j'}^*(\theta_{j'}) = v_{j'},$$

which yield $l^*(\theta_{j'}) = \min_{j=1,\ldots,J} h_j^*(\theta_{j'}) = v_{j'}$ for $j' = 1, \ldots, J$ and hence $l^* \in \mathcal{L}_v$. On the other hand, since problem (14) is a linear convex program, the optimal solution must be located at a boundary of the feasible set. This means that there exist $j_i \in \{1, \ldots, J\}$, for $i = 1, \ldots, n$ such that

$$v_{j_i} + \beta_{j_i}^*(t_i - \theta_{j_i}) = s_i^*, \text{ for } i = 1, \ldots, n, \tag{18}$$

and hence $h_{j_i}^*(t_i) = s_i^*$, for $i = 1, \ldots, n$. Consequently, combining with constraint (14 b), we have

$$l^*(t_i) = \min_{j=1,\ldots,J} h_j^*(t_i) = s_i^*. \tag{19}$$

A combination of Equation (19) and the fact that $\vartheta^* = \sum_{i=1}^n a_i s_i^*$ gives rise to

$$\vartheta^* = \sum_{i=1}^n a_i s_i^* = \sum_{i=1}^n a_i l^*(t_i) \leq \sup_{l \in \mathcal{L}_v} \sum_{i=1}^n a_i l(t_i). \tag{20}$$

The conclusion follows by combining Equations (15) and (20). □

In the follow-up discussions, we will use Θ to denote the set of all breakpoints of the quantile functions of W_k, G_m, and B_m in certainty equivalent and pairwise comparison, respectively, for $k = 1, \ldots, K$ and $m = 1, \ldots, M$ and label them in the increasing order of the values, that is, we will use θ_j to denote the jth smallest element of set Θ and let $\theta_0 = 0$. For clarity of the exposition, scenarios in Ω will be indexed by i and elements in Θ will be indexed by j. Thus, the size of the optimization problem will be determined by the number of pairwise comparisons M, the number of certainty equivalent lotteries K, the size of scenarios $|\Omega| = n$, and the size of the total breakpoints $J = |\Theta|$.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The next theorem states that $\rho_{\mathcal{G}_1}(X)$ can be computed by solving a finite-dimensional linear program of reasonable size when X is discretely distributed with finite outcomes.

Theorem 4.2. *Let X be finitely distributed with $\mathbb{P}(X = x_i) = p_i$, for $i = 1, \dots, n$, where $0 \le x_1 < x_2 < \dots < x_n$. Then $\rho_{\mathcal{G}_1}(X)$ is the optimal value of the following linear program:*

$$\sup_{\upsilon,\beta,s,\zeta} \sum_{i=1}^{n} (x_i - x_{i-1}) s_{i-1} \tag{21a}$$

$$\text{s.t. } w_k^- \le \sum_{j=1}^{J} \left[F^{\leftarrow}_{W_k}(\theta_j) - F^{\leftarrow}_{W_k}(\theta_{j-1}) \right] \upsilon_{j-1} \le w_k^+, \text{ for } k = 1, \dots, K, \tag{21b}$$

$$\sum_{j=1}^{J} \left[F^{\leftarrow}_{G_m}(\theta_j) - F^{\leftarrow}_{G_m}(\theta_{j-1}) \right] \upsilon_{j-1} \le \sum_{j=1}^{J} \left[F^{\leftarrow}_{B_m}(\theta_j) - F^{\leftarrow}_{B_m}(\theta_{j-1}) \right] \upsilon_{j-1}, \text{ for } m = 1, \dots, M, \tag{21c}$$

$$\upsilon_j \le \frac{1}{\varphi(1-\theta_j)}, \text{ for } j = 1, \dots, J \text{ such that } \theta_j \ge 1 - \epsilon_0, \tag{21d}$$

$$s_i \le \frac{1}{\varphi(1-\pi_i)}, \text{ for } i = 1, \dots, n \text{ such that } \pi_i \ge 1 - \epsilon_0, \tag{21e}$$

$$\upsilon_j + \beta_j(\theta_j - \pi_i) \ge s_i, \text{ for } i = 0, 1, \dots, n;\ j = 0, 1, \dots, J, \tag{21f}$$

$$\upsilon_j + \beta_j(\theta_j - \theta_{j+1}) \ge \upsilon_{j+1}, \text{ for } j = 0, 1, \dots, J-1, \tag{21g}$$

$$\upsilon_j + \beta_{j+1}(\theta_j - \theta_{j+1}) \le \upsilon_{j+1}, \text{ for } j = 0, 1, \dots, J-1, \tag{21h}$$

$$\beta_j \ge 0, \text{ for } j = 0, 1, \dots, J, \tag{21i}$$

$$s_i + \zeta_i(\pi_i - \theta_j) \ge \upsilon_j, \text{ for } i = 0, 1, \dots, n;\ j = 0, 1, \dots, J, \tag{21j}$$

$$s_i + \zeta_i(\pi_i - \pi_{i+1}) \ge s_{i+1}, \text{ for } i = 0, 1, \dots, n-1, \tag{21k}$$

$$s_i + \zeta_{i+1}(\pi_i - \pi_{i+1}) \le s_{i+1}, \text{ for } i = 0, 1, \dots, n-1, \tag{21l}$$

$$\zeta_i \ge 0, \text{ for } i = 0, 1, \dots, n, \tag{21m}$$

$$s_0 = 1, s_n = 0, \upsilon_0 = 1, \upsilon_J = 0, \tag{21n}$$

where $\pi_0 = 0$ and $\pi_i = \sum_{l \le i} p_l$ for $i = 1, \dots, n$.

Problem (21) is a linear program with $2(J + n)$ variables and $2nJ + 5n + M + 2K + 5J + 2$ constraints at most (not counting the non-negativity constraints). The objective function follows from Equation (10) and constraint (21b) corresponds to the certainty equivalents $w_k^- \le \rho_g(W_k) \le w_k^+$ for $k = 1, \dots, K$; constraint (21c) arises from pairwise comparisons $\rho_g(G_m) \le \rho_g(B_m)$ for $m = 1, \dots, M$; constraints (21d) and (21e) represent the boundedness condition from $g(\epsilon) \le \frac{1}{\varphi(\epsilon)}$, $\epsilon \le \epsilon_0$; finally constraints (21f)–(21m) characterize the monotonicity and concavity of g and constraint

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(21n) represents the normalization of g because g is a distortion function satisfying $g(0) = 0$ and $g(1) = 1$.

Proof. We use Lemma 4.1 to prove the result and do so in three steps.

Step 1. For each fixed $\upsilon = (\upsilon_0, \upsilon_1, \dots, \upsilon_J) \in \mathbb{R}^{J+1}$, we define the set

$$\mathcal{G}(\upsilon) := \{g \in \mathcal{G} : g(1-\theta_j) = \upsilon_j, \text{ for } j = 0, 1, \dots, J\}.$$

By the definition, $\mathcal{G}(\upsilon) \subset \mathcal{G}$. Note that $\mathcal{G}(\upsilon)$ may be empty. A necessary and sufficient condition for $\mathcal{G}(\upsilon) \neq \emptyset$ is that $\upsilon_0 = 1$, $\upsilon_1 = 0$, and υ_j for $j = 0, 1, \dots, J$ is in nonincreasing order because $\mathcal{G}$ is the set of all distortion functions. In the rest of discussions, we will only focus on the case that $\mathcal{G}(\upsilon) \neq \emptyset$. On the other hand, for any $g \in \mathcal{G}$, there exist $\tilde{\upsilon} \in \mathbb{R}^{J+1}$ such that $g \in \mathcal{G}(\tilde{\upsilon})$. Thus

$$\mathcal{G} = \bigcup_{\upsilon \in \mathbb{R}^{J+1}} \mathcal{G}(\upsilon) \quad \text{and} \quad \mathcal{G}_1 = \bigcup_{\upsilon \in \mathbb{R}^{J+1}} (\mathcal{G}_1 \cap \mathcal{G}(\upsilon))$$

and subsequently we can write $\rho_{\mathcal{G}_1}(X)$ as

$$\begin{aligned} \rho_{\mathcal{G}_1}(X) = \sup_{\upsilon} \ & \rho_{\mathcal{G}_1 \cap \mathcal{G}(\upsilon)}(X) \\ \text{s.t.} \ & \mathcal{G}(\upsilon) \cap \mathcal{G}_1 \neq \emptyset. \end{aligned}$$

Note that $\mathcal{G}(\upsilon)$ defines the set of all functions in $\mathcal{G}$ whose values on Θ are $\upsilon = (\upsilon_0, \upsilon_1, \dots, \upsilon_J)$, whereas $\mathcal{G}_{ce}$ is a set of specific functions in $\mathcal{G}$, which satisfy certainty equivalent conditions and their values are determined by a subset of Θ. Moreover, since $\mathcal{G}(\upsilon)$ is determined by υ, then for fixed υ, either υ satisfy the certainty equivalent conditions or not, which implies that either $\mathcal{G}(\upsilon)$ is a subset of $\mathcal{G}_{ce}$ or is disjoint from it. The same is true for the sets $\mathcal{G}_{pair}$ and $\mathcal{G}_{bnd}$. Since $\mathcal{G}_1 = \mathcal{G}_{coh} \cap \mathcal{G}_{pair} \cap \mathcal{G}_{ce} \cap \mathcal{G}_{bnd}$,

$$\rho_{\mathcal{G}_1}(X) = \sup_{\upsilon} \ \rho_{\mathcal{G}_1 \cap \mathcal{G}(\upsilon)}(X) \tag{23a}$$

$$\text{s.t.} \ \mathcal{G}(\upsilon) \cap \mathcal{G}_{coh} \neq \emptyset, \mathcal{G}(\upsilon) \subset \mathcal{G}_{pair}, \mathcal{G}(\upsilon) \subset \mathcal{G}_{ce}, \mathcal{G}(\upsilon) \subset \mathcal{G}_{bnd}. \tag{23b}$$

Step 2. Let Y be a discretely distributed non-negative random variable with $\mathbb{P}(Y = y_l) = p_l$ for $l = 1, \dots, L$. We want to represent $\rho_g(Y)$ via $\{(1-\theta_j, \upsilon_j)\}_{j=0}^{J}$ if the breakpoints of the quantile function of Y are contained in Θ. The quantile function of Y is a step-like function with

$$F_Y^{\leftarrow}(t) = y_{l+1}, \text{ for } t \in (\pi_l, \pi_{l+1}] \text{ and } l = 0, 1, \dots, L,$$

where $y_0 = 0$, $\pi_0 = 0$, $\pi_{L+1} = 1$, and $\pi_l = \sum_{j \leq l} p_j$ for $l = 1, \dots, L$. Moreover, since $\pi_l \in \Theta$ for $l = 1, \dots, L+1$ and θ_l is the lth smallest element in set Θ, then by Equation (10), we have

$$\rho_g(Y) = \sum_{l=1}^{L} (y_l - y_{l-1}) g(1-\pi_{l-1}) = \sum_{j=1}^{J} \left[F_Y^{\leftarrow}(\theta_j) - F_Y^{\leftarrow}(\theta_{j-1})\right] g(1-\theta_{j-1}). \tag{24}$$

Consequently, $\rho_g(W_k)$, $\rho_g(G_m)$, and $\rho_g(B_m)$ have the similar representations of Equation (24) for $k = 1, \dots, K$ and $m = 1, \dots, M$.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Step 3. We are now ready to reformulate problem (23). By Equations (3) and (10), the objective function of program (23) can be reformulated as

$$\rho_{\mathcal{G}_1\cap\mathcal{G}(\upsilon)}(X) = \sup_{g} \sum_{i=1}^{n}(x_i - x_{i-1})g(1-\pi_{i-1}) \tag{25a}$$

$$\text{s.t. } g \in \mathcal{G}_1 \cap \mathcal{G}(\upsilon). \tag{25b}$$

Note that υ in problem (25) is fixed and satisfies (23 b), thus $\mathcal{G}_{pair}$ and $\mathcal{G}_{ce}$ have no impact on problem (25). Consequently, constraint (25 b) can be represented as $g \in \mathcal{G}_{coh} \cap \mathcal{G}(\upsilon) \cap \mathcal{G}_{bnd}$. Problem (25) corresponds to the maximization of the objective function in Equation (21 a) w.r.t. β, s, ζ under the constraints (21 e)–(21 m). To see this, we consider two cases.

Case 1. $1-\pi_i \notin [0,\epsilon_0]$ for $i = 1,\dots,n$. In this case, constraint $\mathcal{G}_{bnd}$ in problem (25) is not invoked. Consequently, constraint (21e) will disappear and problem (25) reduces to Equation (13) except that here $\upsilon_0, \upsilon_1, \dots, \upsilon_J$ and $1-\theta_0, 1-\theta_1, \dots, 1-\theta_J$ are both in nonincreasing order. Let $\tilde{\theta}_j := 1-\theta_j$ for $j = 0,1,\dots,J$. Relabel the sequence $\tilde{\theta}_0, \tilde{\theta}_1, \dots, \tilde{\theta}_J$ in nondecreasing order and denote them by $\hat{\theta}_j$, for $j = 0,1,\dots,J$. Then $\hat{\theta}_j = \tilde{\theta}_{J-j}$, for $j = 0,1,\dots,J$. Let $\hat{\upsilon}_j := g(\hat{\theta}_j) = g(\hat{\theta}_{J-j}) = g(1-\theta_{J-j}) := \upsilon_{J-j}$. Consequently, we can apply Lemma 4.1 with $(\hat{\upsilon}_j, \hat{\theta}_j)$ for $j = 0,1,\dots,J$ and $t_i = 1-\pi_i$ for $i = 0,1,\dots,n$. To see how the application is spelt out, let us consider constraint (14c). In this context, it implies

$$\hat{\upsilon}_j + \hat{\beta}_j(\hat{\theta}_{j+1} - \hat{\theta}_j) \geq \hat{\upsilon}_{j+1}, \text{ for } j = 0,1,\dots,J-1. \tag{26}$$

Let $\beta_j = \hat{\beta}_{J-j}$ for $j = 0,1,\dots,J$. Then, Equation (26) is equivalent to

$$\upsilon_{J-j} + \beta_{J-j}(\theta_{J-j} - \theta_{J-j-1}) \geq \upsilon_{J-j-1}, \text{ for } j = 0,1,\dots,J-1. \tag{27}$$

Let $k := J-j-1$. Then, Equation (27) can be written as

$$\upsilon_{k+1} + \beta_{k+1}(\theta_{k+1} - \theta_k) \geq \upsilon_k, \text{ for } k = 0,1,\dots,J-1, \tag{28}$$

which coincides with constraint (21h). Likewise, we are able to use constraint (14d) to derive constraint (21g) and constraint (14b) to derive constraint (21f). Constraint (21i) corresponds to Equation (14e). Constraints (21j)–(21m) are redundant in this case. To explain this, we note that problem (25) corresponds to the maximization of the objective function in Equation (21 a) w.r.t. β, s, ζ under the constraints (21 e)–(21 m). Following a similar analysis to the converse part of the proof of Lemma 4.1, we are able to show the optimal solutions of the program without these constraints satisfy $s_i = g(1-\pi_i)$ for $i = 1,\dots,n$. Since g is concave, and $\upsilon_j = g(1-\theta_j)$ for $j = 0,1,\dots,J$, then these solutions satisfy (21j)–(21m), which means these constraints are redundant.

Case 2. There exists $i \in \{1,\dots,n\}$ such that $1-\pi_i \in [0,\epsilon_0]$. All arguments are the same as in case 1 except that in this case, constraint $\mathcal{G}_{bnd}$ is valid, see constraint (21e). Unfortunately, this constraint may destroy the nondecreasing concavity of g. To address the issue, we add constraints (21 j)–(21 m) to reinforce the nondecreasing and concave properties of the worst-case distortion function.

Now, we consider the constraints of program (23). Constraint $\mathcal{G}(\upsilon) \cap \mathcal{G}_{coh} \neq \emptyset$ means that there exist $\beta_j \geq 0$ for $j = 0,1,\dots,J$ such that constraints (21 g) and (21 h) hold. Constraint $\mathcal{G}(\upsilon) \subset \mathcal{G}_{pair}$ corresponds to Equation (21 c), $\mathcal{G}(\upsilon) \subset \mathcal{G}_{ce}$ corresponds to constraint (21 b), and $\mathcal{G}(\upsilon) \subset \mathcal{G}_{bnd}$ corresponds to constraint (21 d). Constraint (21n) is the normalization condition because g is a distortion function satisfying $g(0) = 0$ and $g(1) = 1$. □

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Remark 4.3. From the proof of Theorem 4.2, we know that if we drop the sensitivity condition $\mathcal{G}_{bnd}$ in $\mathcal{G}_1$, then we may not require constraints (21 d)–(21 e) and (21 j)–(21 m) in problem (21).

The following example illustrates the curvature of the worst-case concave distortion function.

Example 4.4. Consider the case $\mathcal{G}_1 = \mathcal{G}_{coh} \cap \mathcal{G}_{ce}$, where $\mathcal{G}_{ce} = \{g \in \mathcal{G} : w_1^- \le \rho_g(W_1) \le w_1^+\}$, Consider a random variable where $P(W_1 = 1) = \frac{1}{2}, P(W_1 = 0) = \frac{1}{2}$, and $w_1^- = w_1^+ = \frac{3}{4}$. We calculate the PRDRMs of the following random losses X_1 and X_2 with $P(X_1 = x) = \frac{1}{4}, P(X_1 = \lambda x) = \frac{1}{2}$, $P(X_1 = \mu x) = \frac{1}{4}, P(X_2 = x) = \frac{2}{5}, P(X_2 = \lambda x) = \frac{2}{5}, P(X_2 = \mu x) = \frac{1}{5}$, where $\mu \ge \lambda \ge 1$ and $x > 0$, on $\mathcal{G}_1$. By Equation (10), we have $\mathcal{G}_1 = \{g \in \mathcal{G}_{coh} : g(\frac{1}{2}) = \frac{3}{4}\}$.

In this setup, since there is no pairwise comparison condition and the sensitivity condition, then $M = 0, \varphi = 0$. There is only one certainty equivalent condition, thus $K = 1$ and $J = 2$ with $\theta_1 = \frac{1}{2}$ and $\theta_2 = 1$. For X_1, the quantile function has three breakpoints: $\pi_1 = \frac{1}{4}, \pi_2 = \frac{3}{4}$ and $\pi_3 = 1$. Let $\pi_0 = 0$ and $\theta_0 = 0$. By Theorem 4.2 and Remark 4.3, we can calculate $\rho_{\mathcal{G}_1}(X_1)$ by solving the following linear program

$$\sup_{s_1,s_2;\beta_0,\beta_1,\beta_2} \quad x[s_0 + (\lambda - 1)s_1 + (\mu - \lambda)s_2] \tag{29a}$$

$$\text{s.t.} \quad \upsilon_1 = \frac{3}{4}, \tag{29b}$$

$$0 \le \beta_0 \le \frac{1}{2} \le \beta_1 \le \frac{3}{2} \le \beta_2, \tag{29c}$$

$$s_2 \le \frac{1}{4}\beta_2, s_1 \le \frac{3}{4}\beta_2, \tag{29d}$$

$$s_2 \le \frac{3}{4} - \frac{1}{4}\beta_1, s_1 \le \frac{3}{4} + \frac{1}{4}\beta_1, \tag{29e}$$

$$s_2 \le 1 - \frac{3}{4}\beta_0, s_1 \le 1 - \frac{1}{4}\beta_0, \tag{29f}$$

$$\upsilon_0 = 1, \upsilon_2 = 0, s_0 = 1, s_3 = 0, \tag{29g}$$

To see how the LP works, we consider two cases for X_1. **Case 1:** $\lambda = 3, \mu = 4$. The optimal value of the LP is $\frac{7}{2}x$ and the optimal solution is $s_1 = 1, s_2 = \frac{1}{2}, \beta_0 = 0, \beta_1 = 1, \beta_2 = 2$. Thus $\rho_{\mathcal{G}_1}(X_1) = \frac{7}{2}x$. **Case 2:** $\lambda = 2, \mu = 12$. $\rho_{\mathcal{G}_1}(X_1) = \frac{65}{8}x$ with $s_1 = \frac{7}{8}, s_2 = \frac{5}{8}, \beta_0 = \frac{1}{2}, \beta_1 = \frac{1}{2}, \beta_2 = \frac{5}{2}$.

Figure 2a depicts the worst-case concave distortion function associated with X_1 with different λ and μ values. To explain how the worst-case concave distortion functions are determined, let us note first that in this example, $\Theta = \{0, \frac{1}{2}, 1\}$ and $\upsilon = (0, \frac{3}{4}, 1)$. Thus, for any $g \in \mathcal{G}(\upsilon)$, the graph of g must pass through points (0,0), $(\frac{1}{2}, \frac{3}{4})$, and (1,1). Moreover, by the concavity of g, the graph of the worst-case distortion function must fall within the two shaded triangle areas.

Let us now look into the blue curve. The t-axis of Figure 2a depicts the range of the survival function. For X_1 in case 1, when $t \in [0, 0.25)$, the corresponding loss is $4x$, whereas when $t \in [0.25, 0.75)$, the corresponding loss is $3x$ and when $t \in [0.75, 1)$, the corresponding loss is x. From Figure 2a, we can see that the worst-case distortion function over the interval [0.75,1] should be as largest as possible, and the large possible value in the shaded triangle area is $g(0.75) = 1$ because the difference of quantile losses at $t = 1$ and $t = 0.75$ is $2x$, which is larger than the difference of

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

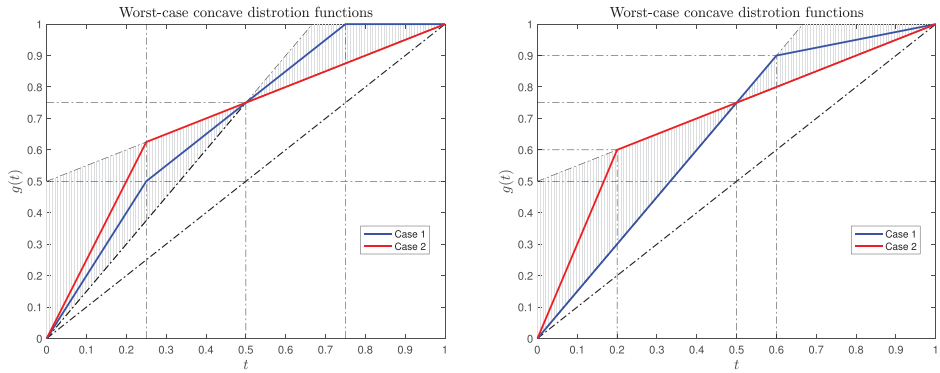


FIGURE 2 Worst-case concave distortion functions: (a) for X_1 and (b) for X_2. [Color figure can be viewed at wileyonlinelibrary.com]

quantile losses at $t = 0.75$ and $t = 0.25$ (is x). Since the worst-case distortion function is piecewise linear by Lemma 4.1, then its concavity determines that the function passing through points (0,0), $(\frac{1}{2}, \frac{3}{4})$, and (1,1) within the shaded triangle areas must be a straight line.

Next, we look into the red curve. For X_1 in case 2, at $t = 0.25$, the corresponding difference of quantile loss is $12x - 2x = 10x$, whereas at $t = 0.75$, the corresponding difference of quantile loss is $2x - x = x$. The difference of quantile loss at 0.25 is much more significant than that at 0.75. From Figure 2a, we can see that the worst-case distortion function over the interval $[0, \frac{1}{4}]$ should be as large as possible, and the largest possible value in the shaded triangle area is $g(\frac{1}{4}) = \frac{5}{8}$. Since the worst-case distortion function is piecewise linear by Lemma 4.1, then its concavity determines that the functions passing through points (0,0), $(\frac{1}{2}, \frac{3}{4})$, and (1,1) within the shaded triangle areas must be a straight line.

For X_2, we have $K = 1$, $M = 0$, $\varphi = 0$, $J = 2$, and $n = 3$; $\pi_0 = 0$, $\pi_1 = \frac{2}{5}$, $\pi_2 = \frac{4}{5}$, and $\pi_3 = 1$; $\theta_0 = 0, \theta_1 = \frac{1}{2}$, and $\theta_2 = 1$. Similar to X_1, we have the following:

- **Case 1:** $\lambda = 3, \mu = 4$. In this case, $\rho_{\mathcal{G}_1}(X_2) = \frac{31}{10}x$ with $s_1 = \frac{9}{10}, s_2 = \frac{3}{10}, \beta_0 = \frac{1}{4}, \beta_1 = \frac{3}{2}, \beta_2 = \frac{3}{2}$.
- **Case 2:** $\lambda = 2, \mu = 12$. In this case, $\rho_{\mathcal{G}_1}(X_2) = \frac{39}{5}x$ with $s_1 = \frac{4}{5}, s_2 = \frac{3}{5}, \beta_0 = \frac{1}{2}, \beta_1 = \frac{1}{2}, \beta_2 = 3$.

Figure 2b depicts the worst-case concave distortion function associated with X_2 with different λ and μ values. Similar arguments apply to X_2.

From Figure 2a, we can see that the worst-case distortion function in both cases can be piecewise linear. Note that they are not unique because we can revise their values, for example, over interval $(0, \frac{1}{4})$ so long as the concavity and monotonicity are preserved. We consider the piecewise linear function only for simplicity. Moreover, the difference between the two curves indicates that they depend on realizations of X_1. Similar comments apply to Figure 2b. Furthermore, by comparing the graph of case 1 in Figure 2a and the graph of case 1 in Figure 2b, we can see that the worst-case distortion function also depends on the probabilities of X_1 and X_2 even though have the same realizations.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.3 | Tractable reformulation for PRDRM on $\mathcal{G}_2$

In this subsection, we consider the tractable formulation for the proposed PRDRM where the ambiguity set of distortion functions is $\mathcal{G}_2$. From our definition, we have

$$\mathcal{G}_2 = \mathcal{G}_S \cap \mathcal{G}_{pair} \cap \mathcal{G}_{ce} \cap \mathcal{G}_{bnd} = \left\{ g \in \mathcal{G} : \begin{array}{l} w_k^- \le \rho_g(W_k) \le w_k^+ \text{ for } k = 1, \dots, K; \\ \rho_g(G_m) \le \rho_g(B_m) \text{ for } m = 1, \dots, M; \\ g(\epsilon) \le \dfrac{1}{\varphi(\epsilon)}\epsilon \le \epsilon_0; \text{ and } g \text{ is inverse S-shaped} \end{array} \right\},$$

where ϵ_0 is a small positive number.

To derive tractable reformulation of PRDRM with ambiguity set $\mathcal{G}_2$, we need to consider how to determine the worst-case convex function through a set of points $\{(\theta_j, \upsilon_j)\}_{j=1}^J$. As in Delage et al. (2022), we will use the support function approach to reformulate the robust constraint, which is convex over $[\tilde{p}, 1]$ as a linear programming problem, and then derive the tractable reformulation of $\rho_{\mathcal{G}_2}(\mathrm{X})$ for a discretely distributed non-negative random loss X.

In the follow-up discussions, we refer to Θ as the set of all breakpoints of quantile functions of W_k, G_m, and B_m in certainty equivalent and pairwise comparison, respectively, for $k = 1, \dots, K$ and $m = 1, \dots, M$, and label them in the increasing order of the values, that is, we use θ_j to denote the jth smallest entry of Θ and $\theta_0 = 0$. To ease the exposition, we will include $1 - \tilde{p}$ in Θ and define the sets $\mathcal{J} = \{1, \dots, J\}$, $\mathcal{J}^- = \{j \in \mathcal{J} : 1 - \theta_j \le \tilde{p}\}$, and $\mathcal{J}^+ = \{j \in \mathcal{J} : 1 - \theta_j \ge \tilde{p}\}$, where $J = |\Theta|$. The next theorem states that $\rho_{\mathcal{G}_2}(X)$ can be computed by solving a finite-dimensional linear program of reasonable size when X is finitely distributed, that is, X has n different realizations. The linear program has $2(J + n)$ variables and $2nJ + 5n + M + 2K + 5J + 2$ constraints at most (not counting the non-negativity constraints). Since the proof is similar to Theorem 4.2, we defer it to Appendix A.2.

Theorem 4.5. *Let X be finitely distributed with $\mathbb{P}(X = x_i) = p_i$, for $i = 1, \dots, n$ with $0 < x_1 < x_2 < \dots < x_n$. Then $\rho_{\mathcal{G}_2}(X)$ is the optimal value of the following linear program:*

$$\sup_{\upsilon, \beta, s, \zeta, \alpha, b} \sum_{i=1}^{n} (x_i - x_{i-1})[s_{i-1}\mathbf{1}\{1 - \pi_{i-1} < \tilde{p}\} + (\alpha_{i-1}(1 - \pi_{i-1}) + b_{i-1})\mathbf{1}\{1 - \pi_{i-1} \ge \tilde{p}\}] \tag{30a}$$

$$\text{s.t. } w_k^- \le \sum_{j=1}^{J} \left[F^{\leftarrow}_{W_k}(\theta_j) - F^{\leftarrow}_{W_k}(\theta_{j-1})\right] \upsilon_{j-1} \le w_k^+, \text{ for } k = 1, \dots, K, \tag{30b}$$

$$\sum_{j=1}^{J} \left[F^{\leftarrow}_{G_m}(\theta_j) - F^{\leftarrow}_{G_m}(\theta_{j-1})\right] \upsilon_{j-1} \le \sum_{j=1}^{J} \left[F^{\leftarrow}_{B_m}(\theta_j) - F^{\leftarrow}_{B_m}(\theta_{j-1})\right] \upsilon_{j-1}, \text{ for } m = 1, \dots, M, \tag{30c}$$

$$\upsilon_j \le \frac{1}{\varphi(1 - \theta_j)}, \text{ for } j \in \mathcal{J}, \text{ such that } \theta_j \ge 1 - \epsilon_0, \tag{30d}$$

$$s_i \le \frac{1}{\varphi(1 - \pi_i)}, \text{ for } i = 1, 2, \dots, n, \text{ such that } \pi_i \ge 1 - \epsilon_0, \tag{30e}$$

$$\upsilon_j + \beta_j(\theta_j - \pi_i) \ge s_i, \text{ for } i = 0, 1, \dots, n; j \in \mathcal{J}^-, \tag{30f}$$

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$v_j + \beta_j(\theta_j - \theta_{j+1}) \geq v_{j+1}, \text{ for } j \in \mathcal{J}^- \setminus \{J\}, \tag{30g}$$

$$v_j + \beta_{j+1}(\theta_j - \theta_{j+1}) \leq v_{j+1}, \text{ for } j \in \mathcal{J}^- \setminus \{J\}, \tag{30h}$$

$$\alpha_i(1 - \theta_j) + b_i \leq v_j, \text{ for } i = 0, 1, \dots, n; j \in \mathcal{J}^+, \tag{30i}$$

$$\alpha_i \geq 0, \text{ for } i = 0, 1, \dots, n, \tag{30j}$$

$$v_j + \beta_j(\theta_j - \theta_{j-1}) \leq v_{j-1}, \text{ for } j \in \mathcal{J}^+, \tag{30k}$$

$$v_j + \beta_{j-1}(\theta_j - \theta_{j-1}) \geq v_{j-1}, \text{ for } j \in \mathcal{J}^+, \tag{30l}$$

$$\beta_j \geq 0, \text{ for } j = 0, 1, \dots, J, \tag{30m}$$

$$s_i + \zeta_i(\pi_i - \theta_j) \geq v_j, \text{ for } j \in \mathcal{J}^-; i = 0, 1, \dots, n-1, \text{such that } \pi_{i+1} \geq 1 - \tilde{p}, \tag{30n}$$

$$s_i + \zeta_i(\pi_i - \pi_{i+1}) \geq s_{i+1}, \text{ for } i = 0, 1, \dots, n-1, \text{such that } \pi_{i+1} \geq 1 - \tilde{p}, \tag{30o}$$

$$s_i + \zeta_{i+1}(\pi_i - \pi_{i+1}) \leq s_{i+1}, \text{ for } i = 0, 1, \dots, n-1, \text{such that } \pi_{i+1} \geq 1 - \tilde{p}, \tag{30p}$$

$$\zeta_i \geq 0, \text{ for } i = 0, 1, \dots, n, \text{such that } \pi_{i+1} \geq 1 - \tilde{p}, \tag{30q}$$

$$v_0 = 1, v_J = 0, s_0 = 1, s_n = 0, \tag{30r}$$

where $\pi_0 = 0$ *and* $\pi_i = \sum_{l \leq i} p_l$ *for* $i = 1, \dots, n$.

Remark 4.6. Similar to Remark 4.3, we know that if we drop the sensitivity condition $\mathcal{G}_{bnd}$ in $\mathcal{G}_2$, then we may not require constraints (30 d)–(30 e) and (21 n)–(21 q) in problem (30).

The next example illustrates the change of concavity and convexity in the worst-case nondecreasing inverse S-shaped distortion functions.

Example 4.7. Consider the case $\mathcal{G}_2 = \mathcal{G}_S \cap \mathcal{G}_{ce}$, where $\mathcal{G}_{ce} = \{g \in \mathcal{G} : w_1^- \leq \rho_g(W_1) \leq w_1^+, w_2^- \leq \rho_g(W_2) \leq w_2^+\}, P(W_1 = 1) = \frac{1}{6}, P(W_1 = 0) = \frac{5}{6}, P(W_2 = 1) = \frac{5}{6}, P(W_2 = 0) = \frac{1}{6}$, and $w_1^- = w_1^+ = \frac{2}{5}, w_2^- = w_2^+ = \frac{3}{5}$. We calculate the PRDRMs of the following random losses Y_1 and Y_2 with $P(Y_1 = x) = \frac{1}{8}, P(Y_1 = \lambda_1 x) = \frac{1}{8}, P(Y_1 = \lambda_2 x) = \frac{1}{2}, P(Y_1 = \lambda_3 x) = \frac{1}{8}, P(Y_1 = \lambda_4 x) = \frac{1}{8}, P(Y_2 = x) = \frac{1}{12}$, $P(Y_2 = \lambda_1 x) = \frac{1}{6}$, $P(Y_2 = \lambda_2 x) = \frac{1}{2}$, $P(Y_2 = \lambda_3 x) = \frac{1}{6}$, $P(Y_2 = \lambda_4 x) = \frac{1}{12}$, where $\lambda_4 \geq \lambda_3 \geq \lambda_2 \geq \lambda_1 \geq 1$ and $x > 0$. By Equation (10), we have

$$\mathcal{G}_2 = \left\{ g \in \mathcal{G} : g\left(\frac{1}{6}\right) = \frac{2}{5}, g\left(\frac{5}{6}\right) = \frac{3}{5}, g \text{ is concave on } \left[0, \frac{1}{3}\right] \text{ and convex on } \left[\frac{1}{3}, 1\right] \right\}.$$

In this setup, since there is no pairwise comparison condition and the sensitivity condition, then $M = 0$, $\varphi = 0$. There are only two certainty equivalent conditions, thus $K = 2$ and $J = 4$ with $\theta_1 = \frac{1}{6}, \theta_2 = \frac{2}{3}, \theta_3 = \frac{5}{6}$, and $\theta_4 = 1$. In this case, $\mathcal{J}^- = \{2, 3, 4\}$ and $\mathcal{J}^+ = \{1, 2\}$. For Y_1, the quantile function has five breakpoints: $\pi_1 = \frac{1}{8}, \pi_2 = \frac{1}{4}, \pi_3 = \frac{3}{4}, \pi_4 = \frac{7}{8}$, and $\pi_5 = 1$. Let $\theta_0 = 0$ and $\pi_0 = 0$. By Theorem 4.5 and Remark 4.6, we can calculate $\rho_{\mathcal{G}_2}(Y_1)$ by solving the following

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

linear program:

$$\sup_{\upsilon,\beta,s,\alpha,b} \; x\left[(\lambda_4-\lambda_3)s_4+(\lambda_3-\lambda_2)s_3+(\lambda_2-\lambda_1)\left(\frac{3}{4}\alpha_2+b_2\right)+(\lambda_1-1)\left(\frac{7}{8}\alpha_1+b_1\right)+\alpha_0+b_0\right]$$

$$\begin{aligned}
\text{s.t. } & \upsilon_1=\frac{3}{5},\upsilon_3=\frac{2}{5},\\
& 0\le\beta_1\le\frac{12}{5}\le\beta_0,0\le\beta_3\le\frac{12}{5}\le\beta_4,\beta_2\ge 0,\\
& \upsilon_2-\frac{1}{12}\beta_2\ge s_3,\upsilon_2-\frac{5}{24}\beta_2\ge s_4,\upsilon_2-\frac{1}{3}\beta_2\ge 0,\frac{1}{4}\beta_4\ge s_3,\upsilon_2\ge 0,\\
& \frac{2}{5}+\frac{1}{12}\beta_3\ge s_3,\frac{2}{5}-\frac{1}{24}\beta_3\ge s_4,\frac{1}{8}\beta_4\ge s_4,\\
& \upsilon_2-\frac{1}{6}\beta_2\ge\frac{2}{5},\upsilon_2-\frac{1}{6}\beta_3\le\frac{2}{5},\upsilon_2+\frac{1}{2}\beta_2\le\frac{3}{5},\upsilon_2+\frac{1}{2}\beta_1\ge\frac{3}{5},\\
& \alpha_0+b_0\le 1,\frac{5}{6}\alpha_0+b_0\le\frac{3}{5},\frac{1}{3}\alpha_0+b_0\le\upsilon_2,\\
& \alpha_1+b_1\le 1,\frac{5}{6}\alpha_1+b_1\le\frac{3}{5},\frac{1}{3}\alpha_1+b_1\le\upsilon_2,\\
& \alpha_2+b_2\le 1,\frac{5}{6}\alpha_2+b_2\le\frac{3}{5},\frac{1}{3}\alpha_2+b_2\le\upsilon_2,\\
& \upsilon_0=1,\upsilon_4=0,s_0=1,s_5=0.
\end{aligned}$$

To see how the LP works, we consider three cases for Y_1.

- **Case 1:** $\lambda_1=2,\lambda_2=3,\lambda_3=4,\lambda_4=5$. In this case, $\rho_{\mathcal{G}_2}(Y_1)=\frac{16}{5}x$ with $s_4=\frac{3}{10},s_3=\frac{3}{5},\upsilon_2=\frac{3}{5},\frac{3}{4}\alpha_2+b_2=\frac{3}{5},\frac{7}{8}\alpha_1+b_1=\frac{7}{10},\alpha_0+b_0=1$.
- **Case 2:** $\lambda_1=2,\lambda_2=3,\lambda_3=4,\lambda_4=10$. In this case, $\rho_{\mathcal{G}_2}(Y_1)=\frac{76}{15}x$ with $s_4=\frac{2}{5},s_3=\frac{2}{5},\upsilon_2=\frac{17}{30},\frac{3}{4}\alpha_2+b_2=\frac{3}{5},\frac{7}{8}\alpha_1+b_1=\frac{7}{10},\alpha_0+b_0=1$.
- **Case 3:** $\lambda_1=2,\lambda_2=5,\lambda_3=6,\lambda_4=10$. In this case, $\rho_{\mathcal{G}_2}(Y_1)=\frac{27}{5}x$ with $s_4=\frac{419}{1154},s_3=\frac{1195}{2522},\upsilon_2=\frac{316}{577},\frac{3}{4}\alpha_2+b_2=\frac{366}{619},\frac{7}{8}\alpha_1+b_1=\frac{7}{10},\alpha_0+b_0=1$.

Figure 3a depicts the worst-case inverse S-shaped distortion functions associated with Y_1 with different $\lambda_1,\lambda_2,\lambda_3$, and λ_4 values. To see how the worst-case inverse S-shaped distortion functions are determined, let us note first that in this example, $\Theta=\{0,\frac{1}{6},\frac{1}{3},\frac{5}{6},1\}$ and $\upsilon=(0,\frac{2}{5},\upsilon_2,\frac{3}{5},1)$. Thus, for any $g\in\mathcal{G}(\upsilon)$, the graph of g must pass through points (0,0), $(\frac{1}{6},\frac{2}{5})$, $(\frac{5}{6},\frac{3}{5})$, and (1,1). Moreover, by the increasing concavity of g over the interval $[0,\frac{1}{3}]$ and increasing convexity of g over the interval $[\frac{1}{3},1]$, the graph of the worst inverse S-shaped distortion function must fall within the shaded area.

Let us now look into the blue curve. The t-axis of Figure 3a depicts the range of survival functions. Since from Equation (10), the adjacent differences of the losses of Y_1 (sorted in increasing order) are all equal to x, then the worst-case inverse S-shaped distortion function should be as large as possible on the interval $[\frac{1}{3},1]$ and the largest value of g at point $\frac{1}{3}$ is $\frac{3}{5}$. Moreover, since

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

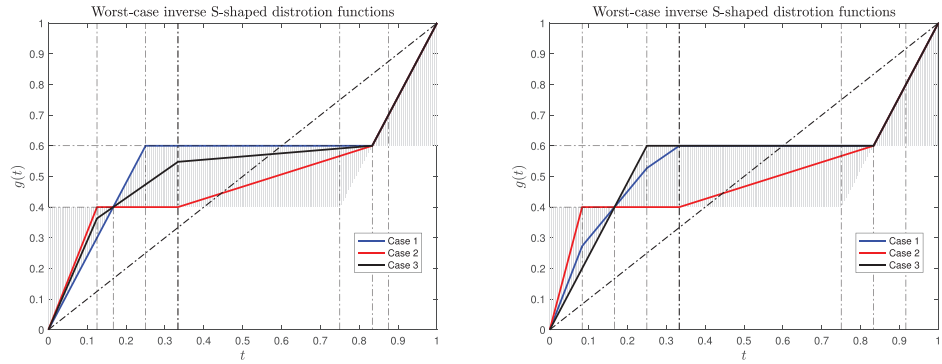


FIGURE 3 Worst-case inverse S-shaped distortion functions: (a) for Y_1 and (b) for Y_2. [Color figure can be viewed at wileyonlinelibrary.com]

the distortion function on the interval $[0, \frac{1}{3}]$ is concave and should pass through the points (0,0), $(\frac{1}{6}, \frac{2}{5})$, and $(\frac{1}{3}, \frac{3}{5})$, then the worst-case nondecreasing concave function over the interval may be the one such that the value of g at point $\frac{1}{4}$ is lifted to the largest possible value (which is $\frac{3}{5}$), and this in turn forces the part over the interval $[0, \frac{1}{6}]$ to take a linear form. Similar arguments can be applied to cases 2 and 3 for Y_1.

For Y_2, we have $K = 2, M = 0, \varphi = 0, J = 4$, and $n = 5$; $\pi_0 = 0, \pi_1 = \frac{1}{12}, \pi_2 = \frac{1}{4}, \pi_3 = \frac{3}{4}, \pi_4 = \frac{11}{12}$, and $\pi_5 = 1$; $\theta_0 = 0, \theta_1 = \frac{1}{6}, \theta_2 = \frac{2}{3}, \theta_3 = \frac{5}{6}$, and $\theta_4 = 1$; $\mathcal{J}^- = \{2, 3, 4\}$ and $\mathcal{J}^+ = \{1, 2\}$. Then similar to Y_1, we have the following:

- **Case 1:** $\lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 4, \lambda_4 = 5$, then $\rho_{\mathcal{G}_2}(Y_1) = \frac{16}{5}x$ with $s_4 = \frac{855}{3134}, s_3 = \frac{1212}{2299}, \upsilon_2 = \frac{3}{5}, \frac{3}{4}\alpha_2 + b_2 = \frac{3}{5}, \frac{11}{12}\alpha_1 + b_1 = \frac{4}{5}, \alpha_0 + b_0 = 1$;
- **Case 2:** $\lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 4, \lambda_4 = 10$, then $\rho_{\mathcal{G}_2}(Y_1) = \frac{31}{6}x$ with $s_4 = \frac{2}{5}, s_3 = \frac{2}{5}, \upsilon_2 = \frac{17}{30}, \frac{3}{4}\alpha_2 + b_2 = \frac{3}{5}, \frac{11}{12}\alpha_1 + b_1 = \frac{4}{5}, \alpha_0 + b_0 = 1$;
- **Case 3:** $\lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 9, \lambda_4 = 10$, then $\rho_{\mathcal{G}_2}(Y_1) = \frac{31}{5}x$ with $s_4 = \frac{1}{5}, s_3 = \frac{3}{5}, \upsilon_2 = \frac{3}{5}, \frac{3}{4}\alpha_2 + b_2 = \frac{3}{5}, \frac{7}{8}\alpha_1 + b_1 = \frac{4}{5}, \alpha_0 + b_0 = 1$;

Figure 3b depicts the worst-case inverse S-shaped distortion functions associated with Y_2 with different $\lambda_1, \lambda_2, \lambda_3$, and λ_4 values.

Similar to the comments after Example 4.4, from Example 4.7, especially from Figure 3a,b, we can see that the worst-case nondecreasing inverse S-shaped distortion functions for Y_1 and Y_2 in these three cases depend on both the probability distribution and the realization of the random losses. Therefore, we conclude that the worst-case piecewise linear nondecreasing inverse S-shaped distortion function for $\rho_{\mathcal{G}_2}(Y)$ depends on Y.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5 | PREFERENCE ROBUST DISTORTION RISK OPTIMIZATION

We now move on to discuss how to solve the optimal decision-making problem (4) with ambiguity set $\mathcal{G}_1$. To do so, we assume that the underlying uncertainty ξ is discretely distributed with $P(\xi = \xi_i) = p_i$ for $i = 1, \dots, n$. In the case when ξ is continuously distributed, we may discrete it with standard methods in stochastic programming such as sample average approximation, see, for example, Shapiro et al. (2014) and Guo and Xu (2021).

The main challenge that we have to tackle is the ordered statistics of $\{f(z, \xi_i), \text{for } i = 1, \dots, n\}$, which depends on the decision variable z. Wang and Xu (2020) propose the alternating iterative algorithm for solving a minimax spectral risk optimization problem. Here we continue to use the approach but under a significantly different setup. We begin with the case that $p_i = \frac{1}{n}$ for $i = 1, \dots, n$ before moving to the general case not only because it is relatively easy to handle but also because in the data-driven problems, empirical data are often assumed to be uniformly distributed.

5.1 | Discrete uniform distribution case

Let $p_i = \frac{1}{n}$ for $i = 1, \dots, n$ and $\phi_0 = 0$. For each fixed $z \in Z$ and $g \in \mathcal{G}$, it follows from Equation (10) that

$$\begin{aligned}\rho_g(f(z,\xi)) &= \sum_{i=1}^{n} f(z,\xi_i)\phi_i = \sum_{i=1}^{n}(\phi_i - \phi_{i-1})\left(\sum_{j=i}^{n} f(z,\xi_j)\right)\\ &= \sum_{i=1}^{n}(\phi_i - \phi_{i-1})(n-i+1)\left(\frac{1}{n-i+1}\sum_{j=i}^{n} f(z,\xi_j)\right)\\ &= \sum_{i=1}^{n}\gamma_i \mathrm{CVaR}_{\beta_i}(f(z,\xi)),\end{aligned} \tag{32}$$

where $\gamma_i = (\phi_i - \phi_{i-1})(n-i+1)$, $\beta_i = \frac{i-1}{n}$, and $\phi_i = g(\frac{n-i+1}{n}) - g(\frac{n-i}{n})$ for $i = 1, \dots, n$. The simple derivation shows that $\rho_g(f(z,\xi))$ can be represented as a linear combination of $\mathrm{CVaR}_{\beta_i}(f(z,\xi))$. The derivation requires ordered statistics $f(z,\xi_1) < f(z,\xi_2) < \cdots < f(z,\xi_n)$ but the representation does not depend on the ordered statistics. Recall that CVaR can be reformulated as a convex minimization problem

$$\mathrm{CVaR}_\alpha(X) = \inf_{\eta \in \mathbb{R}} \left\{\eta + \frac{1}{1-\alpha}\mathbb{E}[(X-\eta)_+]\right\}, \tag{33}$$

where $(a)_+ = \max\{a, 0\}$ and the infimum is attained, see, for example, Pflug (2000) and Rockafellar and Uryasev (2002). In the rest of this section, we will discuss how to solve the PRO model (4) with the ambiguity set $\mathcal{G}_1$.

For $g \in \mathcal{G}_1$, since g is nondecreasing and concave, then $\gamma_i \geq 0$ for $i = 1, \dots, n$. Consequently, we combine Equations (32) and (33) and obtain

$$\rho_g(f(z,\xi)) = \inf_{\eta_1, \dots, \eta_n} \sum_{i=1}^{n} \gamma_i \left\{\eta_i + \frac{1}{1-\beta_i}\mathbb{E}[(f(z,\xi) - \eta_i)_+]\right\}. \tag{34}$$

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

For fixed z, this is a separable convex program with respect to variables η_i, $i = 1, \dots, n$. By introducing auxiliary variables $\tau_i^k := (f(z, \xi_k) - \eta_i)_+$ and μ_i for $k = 1, \dots, n$ and $i = 1, \dots, n$, we can reformulate problem (34) as

$$\inf_{\mu,\eta,\tau} \sum_{i=1}^{n} \gamma_i \mu_i \tag{35a}$$

$$\text{s.t. } \mu_i \geq \eta_i + \frac{1}{1-\beta_i} \sum_{k=1}^{n} \frac{1}{n} \tau_i^k, \text{ for } i = 1, \dots, n, \tag{35b}$$

$$\tau_i^k \geq f(z, \xi_k) - \eta_i, \text{ for } i = 1, \dots, n; k = 1, \dots, n, \tag{35c}$$

$$\tau_i^k \geq 0, \text{ for } i = 1, \dots, n; k = 1, \dots, n. \tag{35d}$$

The equivalence of the two formulations is evident by the fact that an optimal solution to one program can be used to construct a feasible solution to the other program. For example, let $(\eta_1^*, \dots, \eta_n^*)$ be an optimal solution to the problem (34). Then we can define $\mu_i^* := \eta_i^* + \frac{1}{1-\beta_i} \sum_{k=1}^{n} \frac{1}{n} (\tau_i^k)^*$, $(\tau_i^k)^* := (f(z, \xi_k) - \eta_i^*)_+$ and easily verify that (μ^*, η^*, τ^*) is a feasible solution to the problem (35). Based on Equation (35), we can effectively recast $\min_{z \in Z} \rho_g(f(z, \xi))$ as

$$\inf_{z,\mu,\eta,\tau} \sum_{i=1}^{n} \gamma_i \mu_i \tag{36a}$$

$$\text{s.t. } \mu_i \geq \eta_i + \frac{1}{1-\beta_i} \sum_{k=1}^{n} \frac{1}{n} \tau_i^k, \text{ for } i = 1, \dots, n, \tag{36b}$$

$$\tau_i^k \geq f(z, \xi_k) - \eta_i, i = 1, \dots, n; \text{ for } k = 1, \dots, n, \tag{36c}$$

$$\tau_i^k \geq 0, \text{ for } i = 1, \dots, n; k = 1, \dots, n, \tag{36d}$$

$$z \in Z, \tag{36e}$$

which is a convex program and can be easily solved by using, for example, `fmincon` or `CVX` in MATLAB.

Since for each fixed $z \in Z$, the ordered statistics of $\{f(z, \xi_i), \text{for } i = 1, \dots, n\}$ can be sorted out, then from Theorem 4.2, the worst-case concave distortion function (which is determined by the reference points and test points $\frac{i}{n}$ for $i = 1, \dots, n$) for $\rho_{\mathcal{G}_1}(f(z, \xi))$ can be solved by a linear program (21). Moreover, once the worst-case distortion function is determined, then γ_i is fixed for $i = 1, \dots, n$, and the optimal $z \in Z$ for minimizing the distortion risk can be obtained by solving a convex program (36). Therefore, the resulting decision-making problem (4) on $\mathcal{G}_1$ is essentially a saddle point problem motivates us to consider the following procedures based on the alternating iterative algorithm for solving this saddle point problem.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Algorithm 5.1 (Computational procedures for solving Problem (4) with ambiguity set $\mathcal{G}_1$)

Step 0. Choose some initial $z^0 \in Z$. For $j = 1, 2, ...$, do the following steps.

Step 1. Sort the function values $\{f(z^{j-1}, \xi_i) : \text{for } i = 1, \dots, n\}$ in nondecreasing order and denote the sorted values by $f(z^{j-1}, \xi_{[1]}) \le f(z^{j-1}, \xi_{[2]}) \le \cdots \le f(z^{j-1}, \xi_{[n]})$.

Step 2. Based on the sorted order of function values, solve the following problem:

$$\max_{\upsilon,\beta,s,\zeta} \quad \sum_{i=1}^{n} \left[f(z^{j-1}, \xi_{[i]}) - f(z^{j-1}, \xi_{[i-1]}) \right] s_{i-1}$$

$$\text{s.t.} \quad (21b) - (21n),$$

where $\pi_0 = 0$ and $\pi_i = \frac{i}{n}$ for $i = 1, \dots, n$ in constraints (21e)–(21f) and (21j)–(21l). Denote the optimal solution of the component s by s^j.

Step 3. Let $\phi_i = s^j_{i-1} - s^j_i$ for $i = 1, \dots, n$ (consequently, $\gamma_i = (\phi_i - \phi_{i-1})(n - i + 1)$ is fixed). Solve problem (36) and let z^j denote the z-component of the optimal solution.

Step 4. Stop when $z^j = z^{j-1}$ and $s^j = s^{j-1}$

The next proposition states convergence of Algorithm 5.1.

Proposition 5.1. *Consider Algorithm 5.1 for solving problem (4) with $\mathcal{G}' = \mathcal{G}_1$. Suppose that for each fixed ξ, f is convex in z over Z. Then the algorithm either terminates in a finite number of steps with a solution to the problem or generates a sequence $\{(z^j, s^j)\}$ whose cluster points are optimal solutions to the problem.*

The proof is similar to that of Wang and Xu (2020, Proposition 3.1). For the completeness of this paper, we give a proof in Appendix A.3. For the discussions of the inverse S-shaped preference robust distortion risk optimization, we refer readers to an online version of this paper (Wang & Xu, 2021).

5.2 | Discrete nonuniform distribution case

In the case that ξ is not uniformly distributed, we assume without loss of generality that p_i, for $i = 1, \dots, n$, are rational numbers given the fact that rational numbers are dense on [0,1]. Consequently, we can find a least common multiple, denoted by M, to the denominators for all of the p_is such that they have the same denominator M. We then add some auxiliary scenarios to $\{\xi_1, \dots, \xi_n\}$ such that there are M scenarios in total denoted by $\{\hat{\xi}_1, \dots, \hat{\xi}_M\}$ and each scenario has an equal probability, that is, $\mathbb{P}(\hat{\xi} = \hat{\xi}_j) = \frac{1}{M}$ for $j = 1, \dots, M$. With this treatment, we may duplicate $f(z, \xi_i)$ for $p_i M$ times for $i = 1, \dots, n$ to generate a new random variable $f(z, \hat{\xi})$ with realizations as

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

follows:

$$f(z,\hat{\xi}_j)=\begin{cases} f(z,\xi_1), & \text{for } j=1,\dots,p_1M,\\ f(z,\xi_2), & \text{for } j=p_1M+1,\dots,p_2M,\\ \quad\vdots & \quad\vdots\\ f(z,\xi_n), & \text{for } j=(1-p_n)M+1,\dots,M. \end{cases} \tag{38}$$

For example, if $n=4$, $\mathbb{P}(\xi=\xi_1)=p_1=\frac{1}{5}$, $\mathbb{P}(\xi=\xi_2)=p_2=\frac{1}{2}$, $\mathbb{P}(\xi=\xi_3)=p_3=\frac{1}{4}$ and $\mathbb{P}(\xi=\xi_4)=p_4=\frac{1}{20}$, then we can find the least common multiple to the denominators of all these four rational numbers, which is $M=20$. Consequently, we can construct a new random variable $\hat{\xi}$ which has 20 scenarios with $\mathbb{P}(\hat{\xi}=\hat{\xi}_i)=\frac{1}{20}$ for $i=1,\dots,20$. Since $p_1=\frac{4}{20}$, $p_2=\frac{10}{20}$, $p_3=\frac{5}{20}$, and $p_4=\frac{1}{20}$, then we may duplicate $f(z,\xi_1)$ such that $f(z,\hat{\xi}_1)=\cdots=f(z,\hat{\xi}_4)=f(z,\xi_1)$, $f(z,\xi_2)$ such that $f(z,\hat{\xi}_5)=\cdots=f(z,\hat{\xi}_{14})=f(z,\xi_2)$, and so on. In this way, we can use the artificially new random variable $\hat{\xi}$, which is uniformly distributed for solving problem (4) with the computational procedures proposed in Section 5.1.

It is important to note that introducing an auxiliary random variable $f(z,\hat{\xi})$ to replace $f(z,\xi)$ will introduce a large number of additional variables to the problem (36). However, since the worst-case distortion function is piecewise linear, most of the γ_i values are zero, and consequently, the scale of the problem (36) is not large when the number of the scenarios for ξ is small. The next proposition states this.

Proposition 5.2. *Let $z\in Z$ and consider maximization problem*

$$\sup_{g\in\mathcal{G}}\rho_g(f(z,\xi)) \tag{39}$$

where $\mathcal{G}$ is $\mathcal{G}_1$ or $\mathcal{G}_2$, let g^* be an optimal solution to the problem. Let $f(z,\hat{\xi}_j)$ be defined as in Equation (38) for $j=1,\dots,M$. Then,

$$\rho_{g^*}(f(z,\xi))=\sum_{j=1}^{M}\gamma_j\mathrm{CVaR}_{\beta_j}(f(z,\hat{\xi})) \tag{40}$$

where $\gamma_j=(\phi_j-\phi_{j-1})(M-j+1),\beta_j=\frac{j-1}{M}$, and $\phi_j=g^*(\frac{M-j+1}{M})-g^*(\frac{M-j}{M})$ for $j=1,\dots,M$. Moreover, $\gamma_j\neq 0$ if and only if $1-\beta_j$ is a breakpoint of g^* and hence the number of nonzero γ_j coincides with the number of breakpoints of g^*.

Proof. From Equation (32), we see that for $g=g^*$,

$$\rho_{g^*}(f(z,\xi))=\rho_{g^*}(f(z,\hat{\xi}))=\sum_{j=1}^{M}\gamma_j\mathrm{CVaR}_{\beta_j}(f(z,\hat{\xi}))=\sum_{j=1}^{M}\gamma_j\mathrm{CVaR}_{\beta_j}(f(z,\xi)),$$

where $f(z,\hat{\xi}_j)$ is defined via Equation (38), $\gamma_j:=(\phi_j-\phi_{j-1})(M-j+1)$, $\phi_j:=g^*(\frac{M-j+1}{M})-g^*(\frac{M-j}{M})$ for $j=1,\dots,M$ with $\phi_0=0$ and $\beta_j:=\frac{j-1}{M}$. This gives rise to Equation (40). Note that the first and the third equalities are due to the law invariant property of DRM and the second one

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

is from Equation (32). Consequently, we can use Equation (5.2) to calculate $\rho_g(f(z,\xi))$ for discrete nonuniform distribution cases.

Next, we show the rest of the claims. For fixed z, it follows by Theorems 4.2 and 4.5 that a worst-case distortion function can be piecewise linear for both concave case and inverse S-shaped case. The breakpoints of $g^*(t)$ lie in the set $\{1-\pi_1, 1-\pi_2, \dots, 1-\pi_n; 1-\theta_1, 1-\theta_2, \dots, 1-\theta_J\}$, where $\pi_i = \sum_{l\le i} p_l$ for $i = 1, \dots, n$, θ_j is the jth smallest of the set consisting of all breakpoints of the quantile functions of W_k, G_m, and B_m for $k = 1, \dots, K$ and $m = 1, \dots, M$. To ease the exposition, let us consider the case that $g^*(t)$ has a linear piece over interval $[1-\pi_1, 1-\pi_2]$. If the consecutive three points $\frac{M-j-1}{M}, \frac{M-j}{M}, \frac{M-j+1}{M}$ fall in the interval $[1-\pi_1, 1-\pi_2]$, then $\phi_j = g^*(\frac{M-j+1}{M}) - g^*(\frac{M-j}{M}) = g^*(\frac{M-j}{M}) - g^*(\frac{M-j-1}{M}) = \phi_{j-1}$ and subsequently $\gamma_j = 0$. This means that $\gamma_j \neq 0$ if and only if when the middle point $\frac{M-j}{M}$ lies in set $\{1-\pi_1, 1-\pi_2, \dots, 1-\pi_n; 1-\theta_1, 1-\theta_2, \dots, 1-\theta_J\}$. This implies that the number of nonzero γ_j coincides with the number of breakpoints of g^*. □

6 | NUMERICAL RESULTS

In this section, we apply the proposed PRDRM model to the capital allocation problem of a life insurance company, which holds portfolios of life annuities and death benefit insurance. This application is based on the case study investigated by Van Gulick et al. (2012) and Wang et al. (2023). We use the same example to examine the performance of our proposed model and computational procedures.

6.1 | Setup

Let us start by describing the capital allocation problem in a life insurance company and introducing some notation. Analogous to Van Gulick et al. (2012) and Wang et al. (2023), we consider the case that there are two portfolios of life annuities and one death benefit insurance in an insurance company. To make the background of the case study clearer, we extract some basic definitions of the life insurance portfolios from Van Gulick et al. (2012):

- the *(deferred) single life annuity* that yields a yearly payment in every year that the insured is alive and older than 65;
- the *survivor annuity* that yields a yearly payment every year that the spouse outlives the insured if the insured dies before age 65;
- the *death benefit insurance* that yields a single payment in the year the insured die, if the insured dies before age 65.

Let X_i, $i \in \mathcal{I} = \{\text{sl}, \text{surv}, \text{db}\}$ denote the present value of the payments to all insureds in portfolio i. For the completeness of the paper, we provide details about how to get the insurance liability data set for portfolios considered in Appendix B. From there, we can see that there are at most 78 scenarios for each insurance liability and consequently throughout this section, we have $\Omega = \{1, 2, \dots, 78\}$.

Suppose that we have an amount of risk capital, denoted by $\bar{z}$, to be allocated among $X_{\text{sl}}, X_{\text{surv}}$, and X_{db}. In practice, the total amount of risk capital is always determined by a risk measure,

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

denoted by ρ, see, for example, Bauer and Zanjani (2016). Here we use $\text{CVaR}_{99\%}$ to estimate $\overline{z}$ because it is widely accepted to calculate the amount of risk capital in the literature, see, for example Van Gulick et al. (2012). Let $z_i, i \in \mathcal{I}$, denote the amount of risk capital allocated to portfolio i and $z = (z_1, z_2, z_3)^{\mathrm{T}} \in Z \subset \mathbb{R}^3$, where Z represents the set of all feasible risk capital allocations:

$$Z := \left\{ z \in \mathbb{R}^3 : \sum_{i\in\mathcal{I}} z_i = \overline{z},\ z_i \geq 0,\ \text{for } i \in \mathcal{I} \right\}. \tag{41}$$

Our first experiment is to determine the risk capital held by each portfolio if it is isolated from the insurance company and the risk capital held for the aggregated portfolio $X = \sum_{i\in\mathcal{I}} X_i$ under the proposed PRDRM. Meanwhile, we will determine the curvature of the worst-case distortion functions for X_i and X on the ambiguity sets $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively.

Next, we proceed to define the risk capital allocation problem. For a particular allocation strategy $z \in Z$, we follow Van Gulick et al. (2012) to define the cost of mis-allocation to the portfolio i by

$$C_i(z) = (X_i - z_i)_+, \text{ for } i \in \mathcal{I},$$

where $(a)_+ = \max\{a, 0\}$ for $a \in \mathbb{R}$. The quantity $C_i(z)$ can be viewed as the shortfall to portfolio i for $i \in \mathcal{I}$. To simplify the discussion, we adopt the main decision criterion in Van Gulick et al. (2012) to determine the capital allocation by minimizing the aggregate shortfall. In this numerical study, we consider three models. The first one is allocation strategy based on CVaR:

$$\min_{z\in Z} \text{CVaR}_\alpha\left(\sum_{i\in\mathcal{I}} C_i(z)\right), \tag{42}$$

where $\alpha \in (0, 1)$ is the level of confidence and in applications, one typically sets $\alpha = 0.95$ or 0.99, see, for example, Rockafellar and Uryasev (2002). The second one is the DRM model where the true distortion function g_1^* is given:

$$\min_{z\in Z} \rho_{g_1^*}\left(\sum_{i\in\mathcal{I}} C_i(z)\right). \tag{43}$$

The last one is our PRDRM model on $\mathcal{G}_1$:

$$\min_{z\in Z} \rho_{\mathcal{G}_1}\left(\sum_{i\in\mathcal{I}} C_i(z)\right). \tag{44}$$

6.2 | Construction of the ambiguity sets

In this subsection, we consider how to construct ambiguity sets $\mathcal{G}_l$, for $l = 1, 2$, of the distortion functions using the pairwise comparison approach, certainty equivalent approach, and sensitivity conditions that we discussed in Section 4. Let,

$$g_1^*(t) = \sqrt{t}, \text{ for } t \in [0, 1] \quad \text{and} \quad g_2^*(t) = \frac{0.84t^{0.65}}{0.84t^{0.65} + (1-t)^{0.65}}, \text{ for } t \in [0, 1].$$

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Since we do not have a real DM in this numerical study, we assume that g_l^* is the DM's true distortion function for the construction of $\mathcal{G}_l$, that is, we use g_l^* to determine which prospect is preferred in pairwise comparison and the confidence interval in certainty equivalent. For example, we design a pair of prospects: Y_1 and Y_2 and then examine whether $\rho_{g_l^*}(Y_2) \leq \rho_{g_l^*}(Y_1)$. If the inequality holds, then we say the DM prefers Y_2 over Y_1.

For the simplification of notation, we here use elementary lotteries to elicit the DM's attitude and reduce the ambiguity. An *elementary lottery* is defined as

$$Y = \begin{cases} x, & \text{w.p. } p, \\ 0, & \text{w.p. } 1-p. \end{cases} \tag{45}$$

Here $x > 0$ is randomly generated from a given distribution and $p \in (0,1)$ is also randomly generated from the uniform distribution on (0,1), in other words, for each pair of (x, p), we determine a lottery. For such a lottery, the DRM of X induced by g is $\rho_g(X) = xg(p)$. In order to make the elicited data comparable with the random loss of each insurance liability and their aggregated liability, we generate the loss x from the uniform distribution on

$$\left[r_1 \times \min_{i \in \mathcal{I}; \omega \in \Omega} X_i(\omega), r_2 \times \max_{i \in \mathcal{I}, \omega \in \Omega} X_i(\omega) \right]$$

where $r_1 < 1 < r_2$ and X_i is ith portfolio in the insurance company for $i \in \mathcal{I}$. In our experiment, we set $r_1 = \frac{1}{2}, r_2 = 1.2$ and then the range of x to be [0,370610]. The outline of questionnaires to be asked and elicited data set example are given in Appendix C.

6.3 | Experiments

We divide the experiments into five sets: (i) determining the worst-case distortion functions for both concave and inverse S-shaped cases; (ii) calculating risk capital allocation under different strategies; (iii) analyzing the impact of elicitation information on the risk capital allocation policy; (iv) investigating the scalability of the PRDRM model; and finally (v) checking the sensitivity of our model to data perturbation. All of the experiments are carried out on MATLAB 2018b installed on a Macbook Pro (i5-5257 CPU, 2.90GHz dual core processor, 8GB memory).

6.3.1 | Worst-case distortion functions

In this experiment, we examine the impact of the number of the certainty equivalent constraints and the pairwise comparison constraints on the worst-case distortion function in the proposed PRDRM with ambiguity set $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively for the three liabilities and their aggregated liability in an insurance company. Specifically, we may choose the number of pairwise comparisons $M \in \{2, 50, 200\}$ and the number of certainty equivalent pairs $K \in \{2, 10, 20\}$ to conduct the experiments.

Figure 4a depicts the change of the worst-case concave distortion function w.r.t. increase of the number of certainty equivalent pairs for X_{sl} when the number of pairwise comparisons is fixed by $M = 2$. Figure 4b depicts the change of the worst-case concave distortion functions w.r.t. increase of the number of pairwise comparisons for X_{sl} when the number of certainty equivalent is fixed by

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

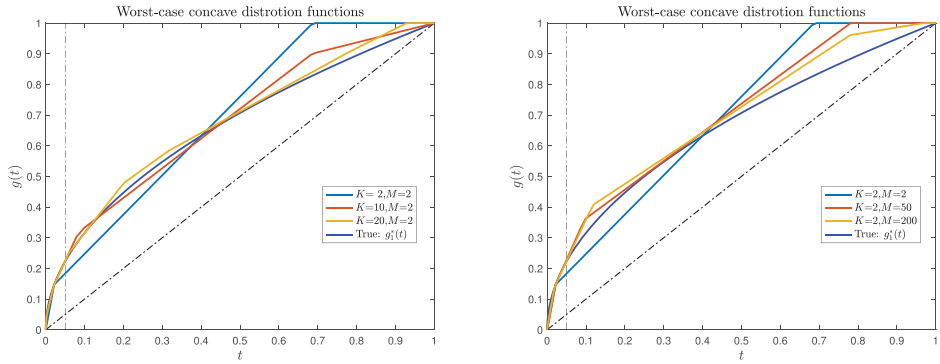


FIGURE 4 Worst-case concave distortion functions for X_{sl}: (a) w.r.t. the number of certainty equivalent pairs and (b) w.r.t. the number of pair-wised comparisons. [Color figure can be viewed at wileyonlinelibrary.com]

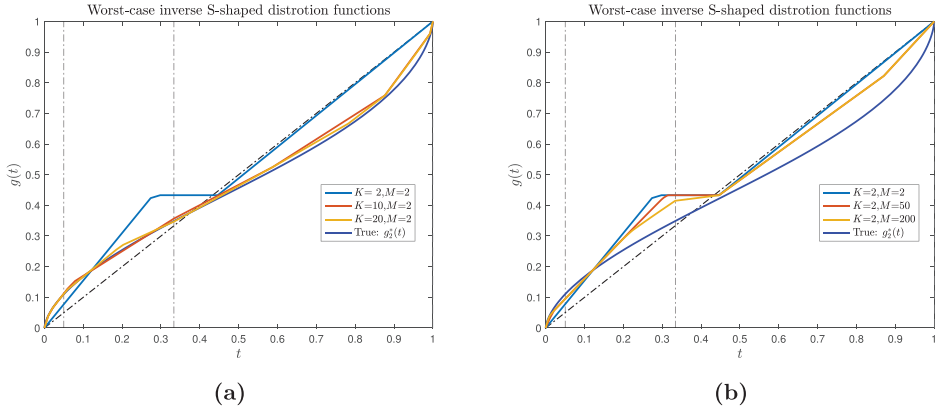


FIGURE 5 Worst-case inverse S-shaped distortion functions for X_{sl}: (a) w.r.t. the number of certainty equivalent pairs and (b) w.r.t. the number of pair-wised comparisons. [Color figure can be viewed at wileyonlinelibrary.com]

$K = 2$. From Figure 4a, we can see that for fixed $M = 2$, the worst-case concave distortion function moves closer and closer to the true one as the number of certainty equivalent pairs increases. This is because the ambiguity set $\mathcal{G}_1$ becomes smaller as the number of certainty equivalent pairs increases. Likewise, from Figure 4b, for fixed $K = 2$, the worst-case concave distortion function moves closer to the true one as the number of pair-wised comparisons increases. In both cases, the worst-case distortion function is bounded by $\sqrt{t}$ over the interval [0,0.05], this is because the sensitivity constraint (7) is valid over the interval, see the left -hand side of the vertical dashed line (at $t = 0.05$).

Figure 5a depicts the worst-case inverse S-shaped distortion functions w.r.t. the number of certainty equivalent pairs for X_{sl} when the number of pairwise comparisons is fixed by $M = 2$, and Figure 5b depicts the worst-case inverse S-shaped distortion functions w.r.t. the number of pairwise comparisons for X_{sl} when the number of certainty equivalent is fixed by $K = 10$. From Figure 5a, we can see that for fixed $M = 2$, the worst-case inverse S-shaped distortion function moves closer and closer to the true one as the number of certainty equivalent pairs increases.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This is because the ambiguity set $\mathcal{G}_2$ becomes smaller as the number of certainty equivalent pairs increases and hence a more precise distortion function may be obtained. Similarly, from Figure 5b, for fixed $K = 10$, the worst-case inverse S-shaped distortion function also moves closer to the true one as the number of pair-wised comparisons increases. In both cases, the worst-case distortion function is bounded by $g_2^*(t)$ over the interval [0,0.05], this is because the sensitivity constraint (7) is valid over the interval, see the left-hand side of the vertical dashed line (at $t = 0.05$). The figures of the worst-case concave distortion functions and the worst-case inverse S-shaped distortion functions for $X_{\text{surv}}, X_{\text{db}}$, and X can be similarly obtained, here we omit them.

6.3.2 | Risk capital allocation strategy

In this experiment, we carry out comparative studies on risk capital allocations among three liabilities (single life annuities, survivor annuity, and death benefit insurance) in an insurance company based on the CVaR model (42), DRM model (43), and PRDRM model (44). To this end, we first calculate the risk capital held by each liability if it is isolated from the insurance company and the risk capital held for the aggregated liability under different risk measures: CVaR, DRM, and PRDRM. Specifically, the risk capital based on CVaR is calculated via Equation (33) with $\alpha = 0.95$ and 0.99, respectively, the risk capital based on DRM is calculated via Equation (10) with $g(t) = g_1^*(t)$ for the concave case and with $g(t) = g_2^*(t)$ for inverse S-shaped case, and finally, the risk capital based on PRDRM is calculated via Equation (21) for the concave case and via Equation (30) for the inverse S-shaped case with $M = 50, K = 10$. Table 2 summarizes the statistical information of three liabilities and their aggregated liability. From Table 2, we see that

$$\rho_{\mathcal{G}_1}(X_i) > \rho_{g_1^*}(X_i) \text{ and } \rho_{\mathcal{G}_2}(X_i) > \rho_{g_2^*}(X_i), \text{ for } i \in \mathcal{I}.$$

This is because $g_l^* \in \mathcal{G}_l$ and PRDRM is defined as the maximum of DRM over set $\mathcal{G}_l$ for $l = 1, 2$.

Note that

$$\rho_{\mathcal{G}_2}(X_i) < \rho_{\mathcal{G}_1}(X_i) < \text{CVaR}_{95\%}(X_i) < \text{CVaR}_{99\%}(X_i), \text{ for } i \in \mathcal{I}.$$

Intuitively speaking, this may be because the distortion function $g_\alpha(t)$ of $\text{CVaR}_\alpha(\cdot)$ defined in Example A.1(ii) dominates both $g_1^*(t)$ and $g_2^*(t)$ for a large α in a point-wise sense. Moreover, by

TABLE 2 The information on three liabilities and their aggregated liability

Description	Single life ann. (X_{sl})	Surv.ann. (X_{surv})	Death benefit (X_{db})	Total (X)
$\mathbb{E}(X_i)$	207,460	11,020	9,620	228,094
$\sigma(X_i)/\mathbb{E}(X_i)$*	0.43	3.10	2.11	0.32
$\text{Corr}(X_i, X)$	0.80	0.05	−0.02	1
$\text{CVaR}_{95\%}(X_i)$	305,630	144,130	79,350	312,890
$\text{CVaR}_{99\%}(X_i)$	308,840	217,570	101,000	324,390
$\rho_{g_1^a}(X_i)$	231,620	35,227	23,524	245,510
$\rho_{g_2^a}(X_i)$	179,790	19,224	13,487	201,330
$\rho_{\mathcal{G}_1}(X_i)$	255,122	48,880	30,961	269,305
$\rho_{\mathcal{G}_2}(X_i)$	190,461	23,487	15,574	212,601

*$\sigma(X_i)$ is the standard deviation for X_i.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Amount of risk capital allocated to the three liabilities under three models

Allocation strategy	Single life ann. (X_{sl})	Surv.ann. (X_{surv})	Death benefit (X_{db})
(42) with $\alpha = 75\%$	293,740	28,720	1,930
(42) with $\alpha = 90\%$	250,530	58,250	15,600
(42) with $\alpha = 95\%$	208,580	63,610	52,200
(42) with $\alpha = 99\%$	154,710	98,170	71,510
(43) with $l = 1$	260,740	37,100	26,550
(44) with $l = 1$	263,760	36,240	24,390

simple calculations, we can see that the quantities satisfy the inequality:

$$\rho(X_{sl}) + \rho(X_{surv}) + \rho(X_{db}) > \rho(X) \tag{46}$$

for $\rho = \rho_{\mathcal{G}_1}, \rho_{\mathcal{G}_2}, \rho_{g_1^*}, \rho_{g_2^*}, \text{CVaR}_{95\%}, \text{CVaR}_{99\%}$. This is guaranteed in theory by subadditivity of coherent risk measure for $\rho_{\mathcal{G}_1}, \rho_{g_1^*}, \text{CVaR}_{95\%}$, and $\text{CVaR}_{99\%}$. However, it is not theoretically guaranteed for $\rho_{\mathcal{G}_2}$ and $\rho_{g_2^*}$. The inequality (46) happens to hold perhaps because of interdependence among X_{sl}, X_{surv}, and X_{db}.

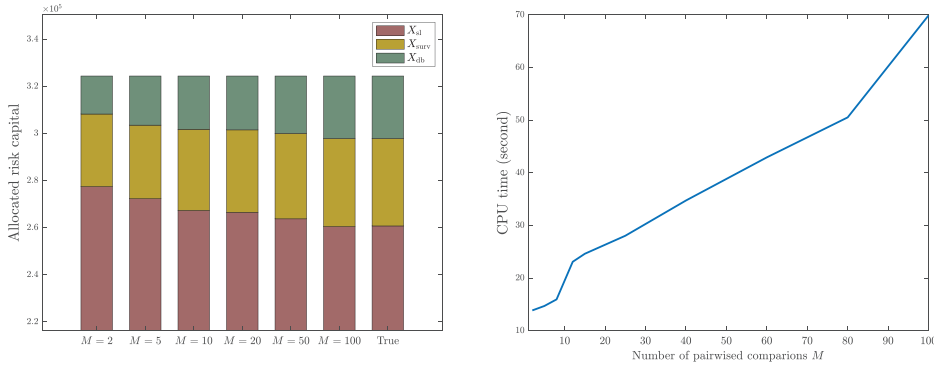
Next, we calculate the risk capitals to be allocated to the three liabilities: X_{sl}, X_{surv}, and X_{db}. In the current insurance regulation, the total amount of risk capital is usually calculated by CVaR_α with $\alpha = 0.99$. Therefore, we set the total amount of risk capital to be $\bar{z} = 324,390$. Table 3 displays the amount of risk capital allocated to three liabilities (single life annuity, survival annuity, and death benefit) under three models: CVaR model (42) with different parameters, DRM model (43), and PRDRM model (44). From Table 3, we can see that most risk capital is allocated to the single life annuity under three models because it has the largest risk compared with the survival annuity and death benefit (see Table 2). As we can see, the amount of risk capital allocated to the single life annuity reduces from 293,740 to 154,710 under CVaR_α when α increases from 0.75 to 0.99. This is because CVaR_α captures larger average tail risks and X_{sl} dominates the tail losses of X_{surv} and X_{db} (see Figure B.2a). The change in allocations under DRM model to the proposed PRDRM model is less drastic. This is because in this experiment, we use the relative larger elicitation information (i.e., the smaller ambiguity set of $\mathcal{G}_1$). Moreover, the allocations under DRM model and the proposed PRDRM model seem to be the weighted average of the allocations of these CVaRs. This is because DRM and PRDRM are both the weighted average of CVaR theoretically.

6.3.3 | Effect of the number of pairwise comparisons on risk capital allocation for concave case

In this experiment, we consider the impact of the number of pairs for eliciting the ambiguity set $\mathcal{G}_1$ on the amount of risk capital allocated to the three liabilities: single life annuity, survival annuity, and death benefit. Specifically, we fix the number of certainty equivalents with $K = 10$ and increase the number of pairwise comparisons M from 2 to 5, 10, 20, 50, and 100. We solve the corresponding minimax optimization problem (44) to examine the impact of the increase of M on the allocated risk capital among X_{sl}, X_{surv}, and X_{db}. The results are depicted in Figure 6a.

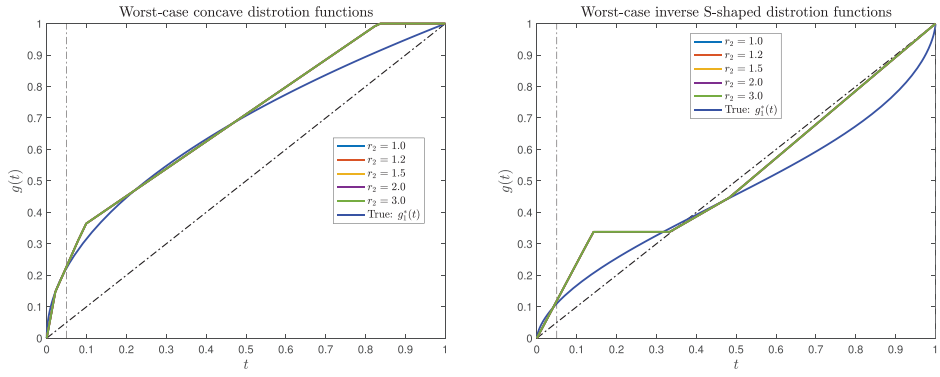
From Figure 6a, we can see that the optimal solutions (allocations) of PRDRM (44) converge to the true optimal solution based on the DRM (see the last column in the figure) where the true distortion function is used. Moreover, it seems that an increase in the number of pairwise

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) (b)

FIGURE 6 (a) Effect of increase of M on the optimal allocated risk capital for concave case. (b) CPU time w.r.t. the number of pairwise comparisons M for the concave case. [Color figure can be viewed at wileyonlinelibrary.com]



(a) (b)

FIGURE 7 (a) Worst-case concave distortion functions for X_{sl} w.r.t. change of r_2 for $M = 2, K = 2$. (b) Worst-case inverse S-shaped distortion functions for X_{sl} w.r.t. change of r_2 for $M = 2, K = 2$. [Color figure can be viewed at wileyonlinelibrary.com]

comparisons M does not have a significant impact on the optimal allocated risk capital for each liability, which means the optimal solution of PRDRM model is not very sensitive to the increase of preference information. Similar results for the effect of the number of certainty equivalent on risk capital allocation can be obtained and so we skip them.

6.3.4 | CPU time w.r.t. the number of pairs for concave case

In this experiment, we look into the computational time for solving the PRDRM model (44) for concave case in terms of CPU time as the number of pairwise comparisons M increases from 2 to 5, 8, 10, 12, 25, 40, 60, 80, and 100 for a fixed number of certainty equivalent $K = 10$. For fixed M (e.g., $M = 2$), we generate two pairs of pairwise comparison questionnaires and then solve the PRDRM model (44) with `linprog` in MATLAB to solve the LPs and record the CPU time. Since

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

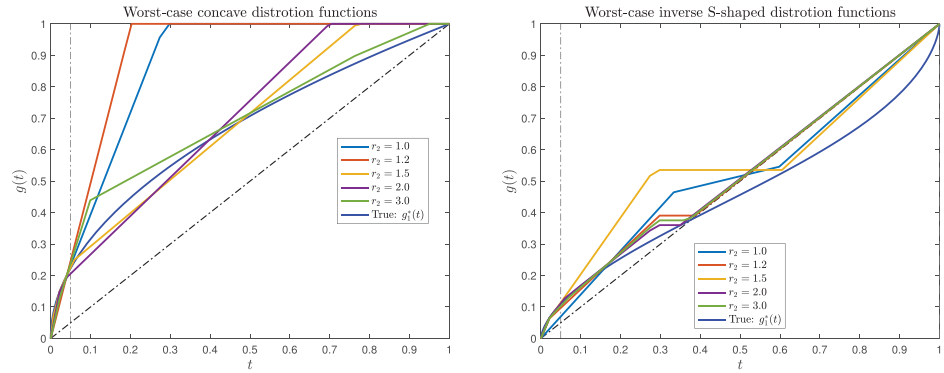


FIGURE 8 (a) Worst-case concave distortion functions for X_{sl} w.r.t. the change of r_2 for $M = 2, K = 2$. (b) Worst-case inverse S-shaped distortion functions for X_{sl} w.r.t. the change of r_2 for $M = 2, K = 2$. [Color figure can be viewed at wileyonlinelibrary.com]

the questionnaires are generated randomly, the resulting PRDRM model varies from simulation to simulation and so does the CPU time. To offset the fluctuations, we calculate the average CPU time over 100 simulations and depict the changes in the CPU time as M increases in Figure 6b. As we can see from Figure 6b, the CPU time changes from 13 s (with $M = 2$) to 70 s (with $M = 100$), which means the M increases by 49-folds whereas the CPU time increases by about 4.4-folds. The results show that the PRDRM model is fairly scalable w.r.t. the change of M.

6.3.5 | Results for other specifications: sensitivity tests

The questionnaires for eliciting a DM's risk attitude are designed by using elementary lotteries (45). By a simple calculation, we know that

$$\min_{i\in I;\omega\in\Omega} X_i(\omega) = 0 \text{ and } \max_{i\in I,\omega\in\Omega} X_i(\omega) = 308,840.$$

This means that only the value of r_2 affects the range of the outcomes of the elementary lotteries. We generate the elementary lotteries as described in Section 6.2 for each r_2 where $r_2 \in \{1.0, 1.2, 1.5, 2.0, 3.0\}$ and set the number of pairwise comparisons M from 2 to 10,50, the number of certainty equivalent K from 2 to 10,20 and solve the problem (21) and (30), respectively. We divide the experiments into two parts: (i) The corresponding probabilities for the elementary lotteries are fixed. (ii) The corresponding probabilities for the elementary lotteries are also randomly generated.

Figure 7a,b depicts the sensitivity results of the worst-case distortion function ($M = 2, K = 2$) w.r.t. change of r_2 for X_{sl} in concave case and inverse S-shaped case, respectively, in the part (i) of the tests. It seems that the change of r_2 does not have an observable effect on the learning of the true distortion function (convergence of the worst-case distortions to the true) as the number of questionnaires increases. The similar results are also obtained for other choices of $M \in \{2, 10, 50\}$ and $K \in \{2, 10, 20\}$ and here we skip them.

In the case that p is also randomly generated, the results are depicted in Figure 8a,b. We can see that the change of the value r_2 has a more significant effect. Moreover, as M and K increase,

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

we find that the effect diminishes under both (i) and (ii), see Appendix D. Comparing Figures 7 and 8, we find that it is the probability of the elementary lottery rather than the value of the lottery has a more significant effect on the worst-case distortion function for both concave and inverse S-shaped cases.

7 | CONCLUDING REMARKS

In this paper, we explore a preference robust distortion risk model for decision-making problems where the information of the true distortion function is incomplete. We propose various ways how to construct an ambiguity set of distortion functions with elicited information about the DM's risk preferences. To compute the proposed PRDRM, we derive tractable formulations where the PRDRM can be calculated via solving a linear programming problem and show through some simple examples how the worst-case distortions may be identified. Finally, we apply the proposed robust DRM to the risk capital allocation problem in insurance.

The robust distortion model complements the current distortion model by allowing ambiguity. Obviously, the smaller the ambiguity set, the less conservative the PRDRM model will impose. It would, therefore, be interesting to link the degree of ambiguity to the level of conservatism of the model. We leave this for future research.

ACKNOWLEDGMENTS

We would like to thank the two anonymous referees for insightful comments and constructive suggestions, which help us strengthen the paper significantly. We would also like to thank the Associate Editor for the effective handling of the review. This work is supported by a CUHK direct grant.

DATA AVAILABILITY STATEMENT

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

ORCID

Wei Wang https://orcid.org/0000-0002-4213-7799
Huifu Xu https://orcid.org/0000-0001-8307-2920

ENDNOTES

[1] Throughout this paper, we use terminologies "risk preference" and "risk attitude" interchangeably as they are both used in the literature of the behavioral economics.

[2] https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables.

REFERENCES

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, *46*(11), 1497–1512.

Abdellaoui, M., Bleichrodt, H., Kemel, E., & l'Haridon, O. (2021). Measuring beliefs under ambiguity. *Operations Research*, *69*(2), 599–612.

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

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, *21*, 503–546.

Armbruster, B., & Delage, E. (2015). Decision making under uncertainty when preference information is incomplete. *Management Science*, *61*(1), 111–128.

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

Baillon, A., Bleichrodt, H., Emirmahmutoglu, A., Jaspersen, J., & Peter, R. (2020). When risk perception gets in the way: Probability weighting and underprevention. *Operations Research*, *70*, 1293–1952.

Bauer, D., & Zanjani, G. (2016). The marginal cost of risk, risk measures, and capital allocation. *Management Science*, *62*(5), 1431–1457.

Bleichrodt, H., & Pinto, J. L. (2000). A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science*, *46*(11), 1485–1496.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Cont, R., Deguest, R., & He, X. D. (2013). Loss-based risk measures. *Statistics & Risk Modeling*, *30*(2), 133–167.

Delage, E., Guo, S., & Xu, H. (2022). Shortfall risk models when information of loss function is incomplete, *Operations Research*, *70*(6), 3511–3518.

Delage, E., & Li, J. Y.-M. (2017). Minimizing risk exposure when the choice of a risk measure is ambiguous. *Management Science*, *64*(1), 327–344.

Denneberg, D. (1990). Premium calculation: Why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin: The Journal of the IAA*, *20*(2), 181–190.

Dhaene, J., Kukush, A., Linders, D., & Tang, Q. (2012). Remarks on quantiles and distortion risk measures. *European Actuarial Journal*, *2*(2), 319–328.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, *75*, 643–669.

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

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

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

Gerber, H. U. (2013). *Life insurance mathematics*. Springer Science & Business Media.

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*, *94*(2), 236–254.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, *38*(1), 129–166.

Guo, S., & Xu, H. (2018). Distributionally robust shortfall risk optimization model and its approximation. *Mathematical Programming*, *174*, 1–26.

Guo, S., & Xu, H. (2021). Robust spectral risk optimization when the subjective risk aversion is ambiguous: A moment-type approach. *Mathematical Programming*, *194*, 1–36.

Hanoch, G., & Levy, H. (1975). The efficiency analysis of choices involving risk. In *Stochastic optimization models in finance* (pp. 89–100). Elsevier.

Haskell, W. B., Fu, L., & Dessouky, M. (2016). Ambiguity in risk preferences in robust stochastic optimization. *European Journal of Operational Research*, *254*(1), 214–225.

Henryk, G., & Silvia, M. (2006). On a relationship between distorted and spectral risk measures. *Online paper*.

Hu, J., & Mehrotra, S. (2015). Robust decision making over a set of random targets or risk-averse utilities with an application to portfolio optimization. *IIE Transactions*, *47*(4), 358–372.

Karlin, S. (1959). *Mathematical methods and theory in games, programming and economics*. Addison-Wesley.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American Statistical Association*, *87*(419), 659–671.

Maccheroni, F. (2002). Maxmin under risk. *Economic Theory*, *19*(4), 823–831.

Merkle, M., Marinescu, D., Merkle, M. M., Monea, M., & Stroe, M. (2014). Lebesgue–Stieltjes integral and young's inequality. *Applicable Analysis and Discrete Mathematics*, *8*, 60–72.

Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic constrained optimization* (pp. 272–281). Springer.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Pichler, A., & Shapiro, A. (2015). Minimal representation of insurance prices. *Insurance: Mathematics and Economics*, *62*, 184–193.
Prelec, D. (1998). The probability weighting function. *Econometrica*, *66*, 497–527.
Rockafellar, R. T., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, *26*(7), 1443–1471.
Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2014). *Lectures on stochastic programming: Modeling and theory*. SIAM.
Sordo, M. A., Castaño-Martínez, A., & Pigueiras, G. (2016). A family of premium principles based on mixtures of TVaRs. *Insurance: Mathematics and Economics*, *70*, 397–405.
Tasche, D. (2002). Expected shortfall and beyond. *Journal of Banking & Finance*, *26*(7), 1519–1533.
Tsanakas, A., & Desli, E. (2003). Risk measures and theories of choice. *British Actuarial Journal*, *9*(4), 959–991.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.
Tversky, A., & Wakker, P. (1995). Risk attitudes and decision weights. *Econometrica: Journal of the Econometric Society*, *63*, 1255–1280.
Van Gulick, G., De Waegenaere, A., & Norde, H. (2012). Excess based allocation of risk capital. *Insurance: Mathematics and Economics*, *50*(1), 26–42.
Von Neumann, J., Morgenstern, O., & Kuhn, H. W. (2007). *Theory of games and economic behavior (commemorative edition)*. Princeton University Press.
Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, *17*(1), 43–54.
Wang, W., & Xu, H. (2020). Robust spectral risk optimization when information on risk spectrum is incomplete. *SIAM Journal on Optimization*, *30*(4), 3198–3229.
Wang, W., & Xu, H. (2021). Preference robust distortion risk measure and its application. Available at SSRN 3763632.
Wang, W., Xu, H., & Ma, T. (2023). Optimal scenario-dependent multivariate shortfall risk measure and its application in capital allocation. *European Journal of Operational Research*, *306*(1), 322–347.
Wirch, J. L., & Hardy, M. R. (2001). Distortion risk measures: Coherence and stochastic dominance. In *International Congress on Insurance: Mathematics and Economics*, (pp. 15–17).
Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, *42*(12), 1676–1690.
Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society*, *55*, 95–115.
Yitzhaki, S. (1982). Stochastic dominance, mean variance, and Gini's mean difference. *The American Economic Review*, *72*(1), 178–185.
Zhang, Y., Xu, H., & Wang, W. (2020). Preference robust models in multivariate utility-based shortfall risk minimization. *Optimization Methods and Software*, *37*, 1–41.

How to cite this article: Wang, W., & Xu, H. (2023). Preference robust distortion risk measure and its application. *Mathematical Finance*, *33,* 389–434. https://doi.org/10.1111/mafi.12379

APPENDIX A: SUPPLEMENTARY MATERIAL

A.1 | Some known examples of distortion risk measures

Example A.1. In this example, we assume that X is a non-negative random variable.

(i) Value at Risk (VaR). Let $g_\nu(t) = \mathbf{1}_{(1-\nu,1]}(t)$, where $\mathbf{1}_C(t)$ is the indicator function over C and $\nu \in (0,1)$. Observe that $F_X(x) \in [0,\nu)$ iff $S_X(x) \in (1-\nu,1]$, that is, $x \in [0, F_X^{\leftarrow}(\nu))$ iff

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$g_\nu(S_X(x)) = 1$. Consequently, it follows from (1) that

$$\rho_{g_\nu}(X) = \int_0^{F_X^{-1}(\nu)} dx = F_X^{-1}(\nu) =: \mathrm{VaR}_\nu(X).$$

(ii) Conditional Value at Risk (CVaR) (also called Expected Shortfall (Tasche, 2002)). Let $g_\alpha(t) = \min\{\frac{t}{1-\alpha}, 1\}$ with $\alpha \in (0,1)$. From (1), using integration by parts of Lebesgue-Stieltjes integral, see, e.g., Merkle et al. (2014), we have

$$\begin{aligned}
\rho_{g_\alpha}(X) &= \int_0^{F_X^{\leftarrow}(\alpha)} 1dx + \int_{F_X^{\leftarrow}(\alpha)}^{\infty} \frac{S_X(x)}{1-\alpha} dx \\
&= F_X^{\leftarrow}(\alpha) + \frac{1}{1-\alpha}\left[S_X(x)x\big|_{F_X^{\leftarrow}(\alpha)_-}^{\infty} - \int_{F_X^{\leftarrow}(\alpha)}^{\infty} x dS_X(x) \right] \\
&= \frac{1}{1-\alpha}\left\{ \mathbb{E}[X\mathbf{1}_{X \geq F_X^{\leftarrow}(\alpha)}] - F_X^{\leftarrow}(\alpha)[P(X < F_X^{\leftarrow}(\alpha)) - \alpha] \right\} \\
&=: \mathrm{ES}_\alpha(X) = \mathrm{CVaR}_\alpha(X),
\end{aligned} \tag{A.1}$$

where $(\cdot)_- = \lim_{\delta \downarrow 0}(\cdot - \delta)$ and the third equality is due to the fact that $\lim_{x\to\infty} x g(S_X(x)) = 0$. The relation can also be derived from (2), where we have

$$\rho_{g_\alpha}(X) = \int_0^1 F_X^{\leftarrow}(t) d\tilde{g}_\alpha(t) = \frac{1}{1-\alpha}\int_\alpha^1 F_X^{\leftarrow}(t)dt =: \mathrm{CVaR}_\alpha(X), \tag{A.2}$$

which is shown in Acerbi (2002).

From an insurance premium point of view, it might be more advantageous to have a non-flat tail distortion function because the whole loss distribution will be utilized, see, e.g., Wang (1995).

(iii) Proportional hazards transform risk measure (Wang, 1995). Let $g_\gamma(t) = t^{\frac{1}{\gamma}}$ for $\gamma > 1$. Then

$$\rho_{g_\gamma}(X) = \int_0^\infty S_X(x)^{\frac{1}{\gamma}} dx =: \rho_{\mathrm{PH}}(X).$$

(iv) Gini's risk measure (Denneberg, 1990). Let $g_s(t) = t - s(t^2 - t)$ for $s \in (0,1)$. Then from (2),

$$\begin{aligned}
\rho_{g_s}(X) &= \int_0^1 F_X^{\leftarrow}(1-t) dg_s(t) = \int_0^1 F_X^{\leftarrow}(1-t)[1 - s(2t-1)]dt \\
&= \int_0^1 F_X^{\leftarrow}(1-t)dt - s\int_0^1 F_X^{\leftarrow}(1-t)(2t-1)dt \\
&= \mathbb{E}[X] + s\int_0^1 F_X^{\leftarrow}(t)(2t-1)dt = \mathbb{E}[X] + \frac{1}{2}s\mathbb{E}[|X - X'|] =: \mathrm{Gini}_s(X),
\end{aligned}$$

where X' is an independent copy of X. The second last equality is due to the fact that $\mathbb{E}[|X - X'|] = 2\int_0^1 (2t-1)F_X^{\leftarrow}(t)dt$ (can be calculated easily with Fubini's Theorem, or see, e.g., Furman et al. (2017)), which is called the Gini's mean difference and is closely related to stochastic dominance, see, e.g., Yitzhaki (1982).

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

A.2 | Proof of Theorem 4.2

The proof requires some theoretical results on majorization of convex function. To make the paper self-contained, we begin by recalling the results. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. A function g is said to be *majorized* by f if

$$g(x) \leq f(x), \quad \forall x \in \text{dom } f.$$

g is a *support function* of f at $x \in \mathbb{R}^n$ if g is majorized by f and $g(x) = f(x)$. A vector $s \in \mathbb{R}^n$ is called a subgradient of f at $x \in \mathbb{R}^n$ if

$$f(y) \geq f(x) + \langle s, y - x \rangle, \quad \forall y \in \text{dom } f,$$

where $\langle a, b \rangle$ denotes the scalar product of two vectors a and b. Let $\partial f(x)$ denote the set of subgradients of f at x, which is known as subdifferential of f at x. When f is convex and subdifferentiable at x, the linear function

$$l(y) = f(x) + \langle a, y - x \rangle$$

is a support function of f at x for any $a \in \partial f(x)$. The next theorem states some properties of support functions and their relationship to convex functions.

Lemma A.2 (Convex majorization). *Let $f : \mathbb{R}^n \to \mathbb{R}$. The following assertions hold:*

(i) f is a convex function if and only if there exists an index set $\mathcal{J}$ such that

$$f(x) = \sup_{j \in \mathcal{J}} l_j(x), \quad \forall x \in dom f,$$

where $\mathcal{J}$ is possibly infinite and $l_j(x) = \langle a_j, x \rangle + b_j$ for all $j \in \mathcal{J}$.

(ii) For any finite set $\Theta \subset \mathbb{R}^n$ and values $\{v_\theta\}_{\theta \in \Theta} \subset \mathbb{R}$, $\hat{f} : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\hat{f}(x) = \max_{a,b} \{\langle a, x \rangle + b : \langle a, \theta \rangle + b \leq v_\theta, \ \forall \theta \in \Theta\} \tag{A.3}$$

is convex and $\hat{f}$ majorizes any convex functions passing $\{(\theta, v_\theta)\}_{\theta \in \Theta}$.

These results are well known, see e.g., Boyd and Vandenberghe (2004). Part (i) states that a convex function can be recovered by taking the supremum of its support functions, and Part (ii) gives conditions for constructing the largest convex function majorized by $\{v_\theta\}_{\theta \in \Theta}$.

Proof of Theorem 4.2. Let $v = (v_0, v_1, \ldots, v_J)$. We start by defining $\mathcal{G}(v) = \{g \in \mathcal{G} : g(1 - \theta_j) = v_j \text{ for } j = 0, 1, \ldots, J\}$. Then we can write $\rho_{\mathcal{G}_2}(X)$ as

$$\rho_{\mathcal{G}_2}(X) = \sup_v \ \rho_{\mathcal{G}_2 \cap \mathcal{G}(v)}(X)$$
$$\text{s.t.} \ \ \mathcal{G}(v) \cap \mathcal{G}_2 \neq \emptyset.$$

Since $\mathcal{G}(v)$ is determined by v, then for each fixed v, either $\mathcal{G}(v)$ is a subset of $\mathcal{G}_{ce}$ or is disjoint from it. The same is true for the sets $\mathcal{G}_{pair}$ and $\mathcal{G}_{bnd}$. Since $\mathcal{G}_2 = \mathcal{G}_S \cap \mathcal{G}_{pair} \cap \mathcal{G}_{ce} \cap \mathcal{G}_{bnd}$, it follows

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that

$$\rho_{\mathcal{G}_2}(X) = \sup_{\upsilon} \rho_{\mathcal{G}_S \cap \mathcal{G}(\upsilon)}(X) \tag{A.5a}$$

$$\text{s.t. } \mathcal{G}(\upsilon) \cap \mathcal{G}_S \neq \emptyset, \mathcal{G}(\upsilon) \subset \mathcal{G}_{pair}, \mathcal{G}(\upsilon) \subset \mathcal{G}_{ce}, \mathcal{G}(\upsilon) \subset \mathcal{G}_{bnd}. \tag{A.5b}$$

Following a similar analysis to the proof of Theorem 4.2 and Lemma A.2, we can reformulate the objective function of problem (A.5) as

$$\sup_{\beta,s,\zeta,\alpha,b} \sum_{i=1}^{n} (x_i - x_{i-1})[s_{i-1}\mathbf{1}\{1-\pi_{i-1} < \tilde{p}\} + (\alpha_{i-1}(1-\pi_{i-1}) + b_{i-1})\mathbf{1}\{1-\pi_{i-1} \geq \tilde{p}\}] \tag{A.6a}$$

$$\text{s.t. } (30e) - (30j) \text{ and } (30m) - (30q). \tag{A.6b}$$

Now, we consider the constraints of problem (A.5). Constraint $\mathcal{G}(\upsilon) \cap \mathcal{G}_S \neq \emptyset$ means that there exist $\beta_j \geq 0$ for $j = 0, 1, \ldots, J$ such that (30 g)-(30 h) and (30 k)-(30 l) hold. Constraint $\mathcal{G}(\upsilon) \subset \mathcal{G}_{pair}$ corresponds to (30 c), constraint $\mathcal{G}(\upsilon) \subset \mathcal{G}_{ce}$ to (30 b), and constraint $\mathcal{G}(\upsilon) \subset \mathcal{G}_{bnd}$ to (30 d). The last constraint (30 r) is a normalization condition.

A.3 | Proof of Proposition 5.1

Proof. Let $g \in \mathcal{G}_1$ be fixed and $s_i = g(1-\pi_i) = g(1-\frac{i}{n})$ for $i = 0, 1, \ldots, n$. To ease the notation, we set $s_{-1} := s_0$. Based on (32), we define

$$\upsilon(z, s) := \sum_{i=1}^{n} \gamma_i \text{CVaR}_{\beta_i}(f(z, \xi)),$$

where $\beta_i = \frac{i-1}{n}, \gamma_i = (\phi_i - \phi_{i-1})(n-i+1), \phi_i = s_{i-1} - s_i$ for $i = 1, \ldots, n$. In other words, we have

$$\gamma_i = (2s_{i-1} - s_i - s_{i-2})(n-i+1), \text{ for } i = 1, \ldots, n.$$

Since $f(z, \xi)$ is a convex function in z for every fixed ξ and the operator $\text{CVaR}_{\beta_i}(\cdot)$ is non-decreasing and convex, then $\text{CVaR}_{\beta_i}(f(z, \xi))$ is a convex function of z. Moreover, since $g \in \mathcal{G}_1$, then $\gamma_i \geq 0$ for $i = 1, \ldots, n$. Thus, $\upsilon(z, s)$ is a convex function in z for every fixed s (depending on g). On the other hand, for fixed $z \in Z$, $\upsilon(z, s)$ is linear in s. Since $s_i \in [0, 1]$ for any $g \in \mathcal{G}_1$, then we can write problem (4) in this context as

$$\min_{z \in Z} \max_{s \in S} \upsilon(z, s), \tag{A.7}$$

where Z is a convex and compact set of $\mathbb{R}^k$ and $S \subset [0, 1]^n$ is a convex and compact set of $\mathbb{R}^n$. By Fan (1953, Theorem 1 (ii)),

$$\min_{z \in Z} \max_{s \in S} \upsilon(z, s) = \max_{s \in S} \min_{z \in Z} \upsilon(z, s), \tag{A.8}$$

which, by Karlin (1959, Corollary 1.3.1), is sufficient and necessary for the existence of a saddle point. Let (z^*, s^*) denote the saddle point. Then

$$\upsilon(z^*, s) \leq \upsilon(z^*, s^*) \leq \upsilon(z, s^*). \tag{A.9}$$

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

We are now ready to show the convergence of the sequence $\{(z^j, s^j)\}$ generated by Algorithm 5.1. For $j = 1, 2, ...$, it follows from Step 2 of Algorithm 5.1,

$$v(z^{j-1}, s^j) \geq v(z^{j-1}, s), \text{ for all } s \in S. \tag{A.10}$$

Likewise, it follows from Step 3 of the algorithm

$$v(z^j, s^j) \leq v(z, s^j), \text{ for all } z \in Z. \tag{A.11}$$

In the case when the algorithm terminates in finite steps, we have $z^{j-1} = z^j$ and $s^{j-1} = s^j$ for some j and consequently (z^j, s^j) satisfies (A.9).

Next, we consider the case that the algorithm generates an infinite sequence $\{(z^j, s^j)\}$. Let $(\hat{z}, \hat{s})$ be a cluster point of $\{(z^j, s^j)\}$. By taking a subsequence if necessary, we assume for the simplicity of notation that $(z^j, s^j) \to (\hat{z}, \hat{s})$ as $j \to \infty$. Assume for the sake of a contradiction that $(\hat{z}, \hat{s})$ is not a solution to the problem (A.7). Then $(\hat{z}, \hat{s})$ would violate one of the inequalities in (A.9). Consider the case that the second inequality of (A.9) is violated. Then there exists $z_0 \in Z$ such that

$$v(\hat{z}, \hat{s}) > v(z_0, \hat{s}).$$

Since v is continuous, the inequality means that for j sufficiently large,

$$v(z^j, s^j) > v(z_0, s^j),$$

which contradicts (A.11). Likewise, we can show that $(\hat{z}, \hat{s})$ satisfies the first inequality in (A.9). This completes the proof. □

APPENDIX B: INSURANCE LIABILITY DATA SET

Similar to the settings of the numerical illustration in Van Gulick et al. (2012), we assume that the annual benefit level of the single life annuity is normalized to 1, the annual benefit level of the survivor annuity is set to 0.75, and the (lump-sum) benefit level of the death benefit insurance is set to 7.5. For simplicity, we also assume that for each of the three portfolios, each insured has the same insured rights. Therefore, each of the life insurance portfolios yields an uncertain stream of payments at future dates and the present date is Dec 31, 2018. Let X_i, $i \in \{\text{sl}, \text{surv}, \text{db}\}$ denote the present value of the payments to all insureds in portfolio i, i.e.,

$$X_{\text{sl}} = \sum_{j=1}^{N_{\text{sl}}} \sum_{\tau=\max\{65-x_{\text{sl},j},0\}}^{100-x_{\text{sl},j}} \frac{\mathbf{1}_{(T_{\text{sl},j} \geq \tau)}}{(1+r)^\tau}, \tag{B.1}$$

$$X_{\text{surv}} = 0.75 \cdot \sum_{j=1}^{N_{\text{surv}}} \sum_{\tau=0}^{100-x'_{\text{surv},j}} \frac{\mathbf{1}_{(T_{\text{surv},j} < \min\{\tau, 65-x_{\text{surv},j}\})} \mathbf{1}_{(T'_{\text{surv},j} \geq \tau)}}{(1+r)^\tau}, \tag{B.2}$$

$$X_{\text{db}} = 7.5 \cdot \sum_{j=1}^{N_{\text{db}}} \sum_{\tau=1}^{\max\{65-x_{\text{db},j}\}} \frac{\mathbf{1}_{(\tau-1 \leq T_{\text{db},j} < \tau)}}{(1+r)^\tau}, \tag{B.3}$$

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

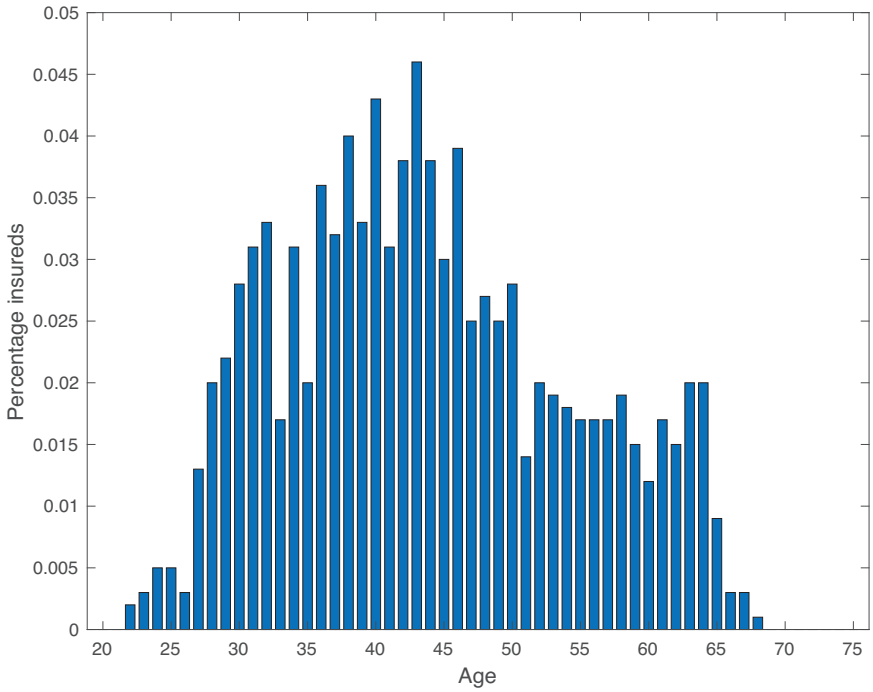


FIGURE B.1 Age composition: Percentage insureds as a function of age. For the portfolio of survivor annuities, the partner of the (male) insured is assumed to be three years younger than the insured. [Color figure can be viewed at wileyonlinelibrary.com]

where $N_i, i \in \{\text{sl, surv, db}\}$, denotes the number of the insureds in portfolio i, $x_{i,j}(x'_{i,j})$ denotes the age of insured j (the partner of insured j) in portfolio i, $T_{i,j}(T'_{i,j})$ denotes the remaining lifetime of insured j (the partner of insured j) in portfolio i, and r denotes the discount rate (see, e.g., Gerber (2013)). We present numerical results for a single life annuity portfolio with 45 000 male insureds ($N_{\text{sl}} = 45000$), a survivor annuity portfolio with 15 000 male insureds ($N_{\text{surv}} = 15000$), and a death benefit insurance portfolio with 15 000 male insureds ($N_{\text{db}} = 15000$). The age composition of the three products is identical, and given in Figure B.1. In addition, we assume that the maximum life span for all insureds is 100. The discount rate is equal to $r = 3\%$.

From Figure B.1, we can see that the age of the insureds ranges from 22 to 68, so the remaining lifetime of the insureds varies from 1 to 78=100-22, which means that the values of $T_{i,j}$ lie in the set $\{1, 2, \dots, 78\}$ for insured j and $i \in \{\text{sl, surv, db}\}$.

Scenarios for future survival probabilities of all insureds are generated by a standard Lee and Carter (1992) model, estimated on the historical data of the age-specific mortality rates in the UK from 1980 to 2017 in the national life tables from the National Statistics[2]. Since the remaining lifetime for an insured is mainly determined by her/his age (statistically speaking), then the random variable $T_{\text{sl},j}$ ($T_{\text{surv},j}$ or $T_{\text{db},j}$) is identical if the age of insured j is the same. Thus, there are totally 47=68-22+1 different random variables conditional on the age of the insured and for each such random variable, there are 78 different values at most. Since the maximum life span for all insureds is 100 and the age of all insureds is from 22 to 68, then there are precisely $2585 = \sum_{i=22}^{68}(100 - i)$ scenarios for the remaining lifetime conditional on the age of the insured. From the above discussions and (B.1), we see that the present value of the payment to insured j, i.e., the random

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

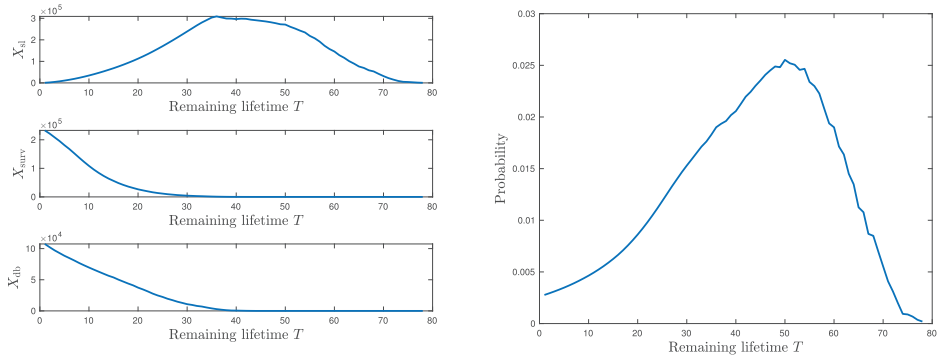


FIGURE B.2 **(a)** The value of the random variables X_{sl}, X_{surv} and X_{db} as a function of remaining lifetime T. **(b)** The probability that remaining lifetime T. [Color figure can be viewed at wileyonlinelibrary.com]

variable

$$\sum_{\tau=\max\{65-x_{sl,j},0\}}^{100-x_{sl,j}} \frac{\mathbf{1}_{(T_{sl,j}\geq\tau)}}{(1+r)^{\tau}}$$

has 78 different realizations at most depending on the realizations of T_{sl}. Note that the realizations of the present value of the payment for different insured in single life annuity are in the same set, i.e.,

$$\left\{\frac{1}{1+r}, \frac{1}{1+r}+\frac{1}{(1+r)^2}, \dots, \sum_{\tau=1}^{78}\frac{1}{(1+r)^{\tau}}\right\}.$$

Therefore, we may divide 2585 scenarios (for all insureds buying single life annuity) into 78 classes based on the realizations of the present value of the payment to each insured. With a slight abuse of notation, we use T to denote these classes. Note that the range of T is from 1 to 78 and we regard the values of T as our considered scenarios: "remaining lifetime". Since there are N_{sl} insureds buying single life annuity, then the sum of present value of the payments to all insureds in single life annuity is a compound random variable. The realization of X_{sl} for each scenario is calculated via (B.1) based on the age composition, which is a weighted sum of the realizations of the present values of the payments for insureds aging from 22 to 78 for each scenario. Similar explanations apply to X_{surv} and X_{db}.

Next, we consider how to obtain the probability of T at each scenario. Since T is determined by merging 2585 scenarios (scenarios of remaining lifetime conditional on the age of the insured) into 78 classes based on the realizations of the present value of the payment, then the probability of scenario $T = j$ is the weighted average of the (age conditional) probabilities of insureds whose remaining lifetime is j, where the weighting is the percentage age composition of insureds. Since the age composition of the three products is identical, then the probability of scenario j for each product is the same. Figure B.2 shows the value and the probability of the random variables X_{sl}, X_{surv} and X_{db} as a function of remaining lifetime T taking on a value in $1, 2, \dots, 78$.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

APPENDIX C: QUESTIONNAIRES AND ELICITED DATA SET EXAMPLE

The questionnaires to be asked are designed as follows.

Q1. Is the DM risk-averse?
If the answer is yes, then we use the ambiguity set $\mathcal{G}_1$, otherwise, we use $\mathcal{G}_2$.

Q2. Which risky lottery from each of the m-th pairs of lotteries does the DM prefer for $m = 1, \dots, M$?
The preferred lottery is denoted by G_m and the other is denoted by B_m. Note that when generating such lotteries, we require that the difference between the DRMs of each pair of lotteries is greater than the specified threshold, denoted by κ, to ensure that the compared lotteries have a significant difference and the DM can easily choose which one s/he prefers to. In the numerical tests, we set $\kappa = 1$. The decision is made based on the values of DRMs for each pair of the lotteries induced by the true distortion function g_1^*.

Q3. What is the smallest amount of money, denoted by w_k^+, that the DM would decline to pay instead of being exposed to the risk of W_k and what is the largest amount of money, denoted by w_k^-, that the DM would be willing to pay instead of being exposed to the risk of W_k for $k = 1, \dots, K$?
In this kind of questionnaire, we assume that w_k^+ and w_k^- are determined based on the values of DRMs for each lottery induced by the true distortion function g_1^* as follows:

$$w_k^+ = (1 + r) w_k g_1^*(p_k), \quad w_k^- = (1 - r) w_k g_1^*(p_k),$$

where w_k and p_k determine the elementary lottery W_k through (45) and r is randomly generated from the uniform distribution on $[0, 5\%]$.

Q4. How sensitive is the DM to large losses with a small probability? Here we may treat the probability which is less than 5% as a small probability.
For simplicity, we assume that the sensitivity function is defined as $\varphi(\epsilon) = \frac{1}{g_1^*(\epsilon)}$ with $\epsilon \leq 5\%$.

We give an example of the generated data set in Table C.1.

TABLE C.1 Generated data set example.

	$\mathcal{G}_1$						$\mathcal{G}_2$					
	G_m		B_m		W_k		G_m		B_m		W_k	
No.	x	p	x	p	x	p	x	p	x	p	x	p
1	269,184	0.25	339,867	0.36	332,487	0.39	17,452	0.49	73,233	0.59	351,747	0.12
2	268,199	0.13	164,189	0.46	119,308	0.46	3,193	0.69	276,701	0.31	261,657	0.44
3	66,982	0.21	55,857	0.39	73,477	0.81	18,087	0.22	226,438	0.85	198,852	0.58
4	109,711	0.95	256,891	0.40	278,295	0.05	71,882	0.75	144,167	0.56	6,571	0.50
5	208,140	0.12	226,130	0.18	113,867	0.05	247,103	0.62	281,560	0.97	87,763	0.81

APPENDIX D: RESULTS FOR SENSITIVITY TEST

In this section, we report the results for sensitivity test w.r.t. the change of r_2 only for the case (ii), i.e., the probability is also randomly generated because the results for the case (i) have no observable difference with Figure 7. Similar results are observed for X_{surv}, X_{db}, and X, here we omit them due to limitation of space.

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

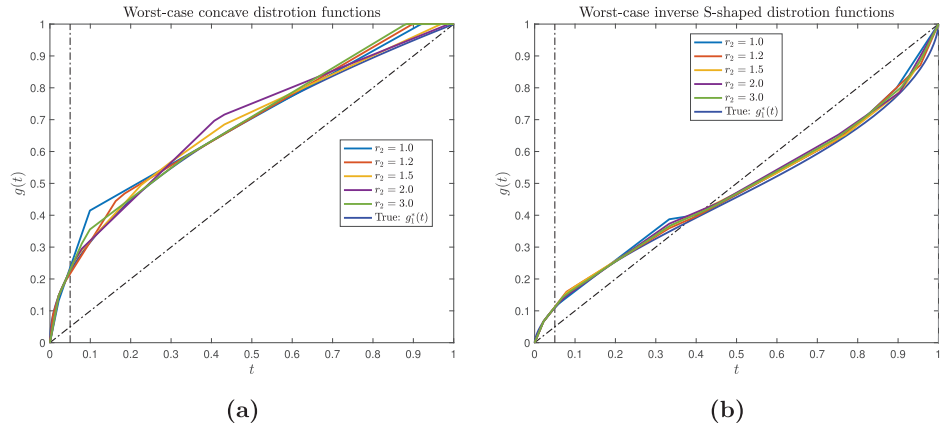


FIGURE D.1 **(a)** Worst-case concave distortion functions for X_{sl} w.r.t. the value of r_2 for $M = 2, K = 10$. **(b)** Worst-case inverse S-shaped distortion functions for X_{sl} w.r.t. the value of r_2 for $M = 2$ and $K = 10$. [Color figure can be viewed at wileyonlinelibrary.com]

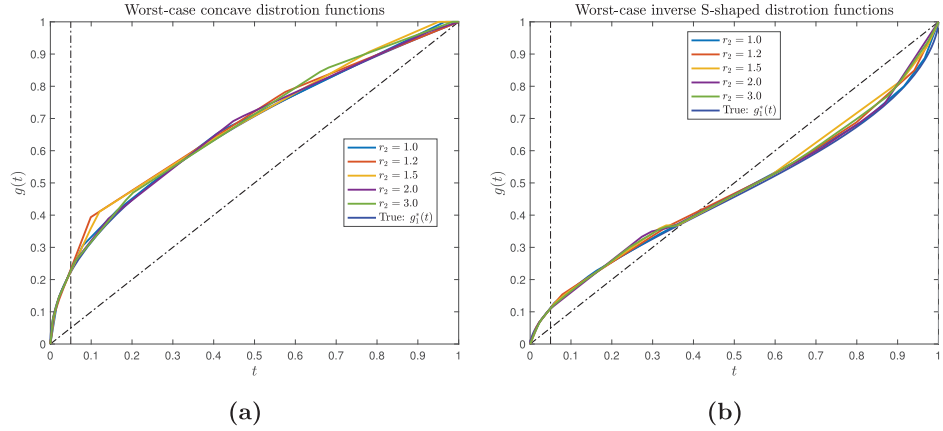


FIGURE D.2 **(a)** Worst-case concave distortion functions for X_{sl} w.r.t. the value of r_2 for $M = 2, K = 20$. **(b)** Worst-case inverse S-shaped distortion functions for X_{sl} w.r.t. the value of r_2 for $M = 2$ and $K = 20$. [Color figure can be viewed at wileyonlinelibrary.com]

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

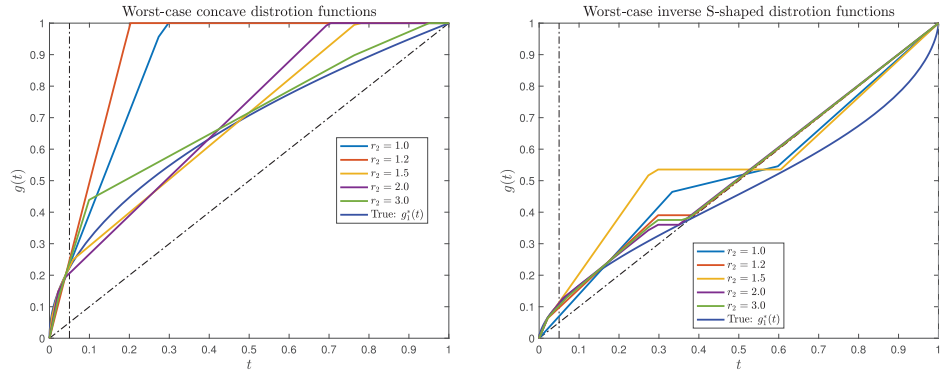


FIGURE D.3 **(a)** Worst-case concave distortion functions for X_{sl} w.r.t. the value of r_2 for $M = 10, K = 2$. **(b)** Worst-case inverse S-shaped distortion functions for X_{sl} w.r.t. the value of r_2 for $M = 10$ and $K = 2$. [Color figure can be viewed at wileyonlinelibrary.com]

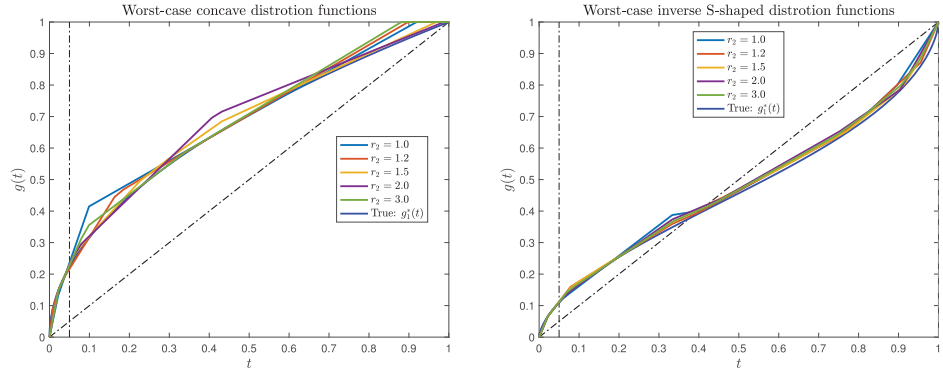


FIGURE D.4 **(a)** Worst-case concave distortion functions for X_{sl} w.r.t. the value of r_2 for $M = 10, K = 10$. **(b)** Worst-case inverse S-shaped distortion functions for X_{sl} w.r.t. the value of r_2 for $M = 10$ and $K = 10$. [Color figure can be viewed at wileyonlinelibrary.com]

14679965, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mafi.12379 by University Of Southampton, Wiley Online Library on [12/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License