Risk Measures of Stop-loss and Limited Loss Random Variables under Model Uncertainty with Applications in Insurance

by

Mingren Yin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2023

© Mingren Yin 2023

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this thesis, our focus is on the optimization of reinsurance design, accounting for the influence of model uncertainty. The following chapters outline our approach:

In Chapter 2, we identify the worst-case distributions for both insurers and reinsurers by assuming that insurers and reinsurers respectively have their own uncertainty sets. These distributions are structured to maximize their respective shares of the total loss, assessed by a distortion risk measure. We consider a reinsurance contract structured as a stop-loss treaty with a deductible. Our uncertainty sets adopt traditional two-moment characteristics, incorporated with distance constraints defined using Wasserstein distance. We provide numerical solutions for the worst-case distributions in a general scenario, along with analytical solutions for cases when uncertainty sets only have constraints on the first two moments of the underlying loss random variable. Based on that, we find the optimal stop-loss reinsurance policy from the perspective of the insurer taking model uncertainty into account.

In Chapter 3, we assume that uncertainty sets of insurers and reinsurers are defined only by Wasserstein distance. We consider the worst-case risk measures of limited stop-loss functions and determine the worst-case distributions for both insurers and reinsurers under limited stop-loss reinsurances. In addition, by conducting numerical experiments, we explore how the limits and deductibles of limited stop-loss reinsurances impact worst-case risk measures for both parties.

Moving into Chapter 4, we integrate the notion of distribution ambiguity into a negotiation framework, specifically Pareto optimality. Through numerical experiments based on results presented in Chapters 2 and 3, we investigate how the negotiation power between parties influences the equilibrium point.

Concluding our study, the final chapter outlines potential directions for future research and development, building upon the foundation laid out in this work.

Acknowledgements

I would like to express my deepest gratitude to all those who have contributed to the completion of this thesis. Without their unwavering support, guidance, and encouragement, this journey would not have been possible.

First and foremost, I extend my heartfelt appreciation to my supervisors, Professor Jun Cai and Professor Fangda Liu, for their invaluable guidance and mentorship throughout this research.

I am grateful to the members of my thesis committee, Professor Chengguo Weng, Professor David Landriault, Professor Thomas Parker and Professor Hailiang Yang for their insightful feedback and constructive criticism. Their diverse perspectives greatly enriched the quality of this research.

To my cherished best friends, you shared my laughter and tear. Your thought-provoking ideas and steadfast companionship have infused the past four years with vibrant colors and brilliance. I hereby sincerely wish all of you a journey of life that is as memorable and fruitful as the moments we have shared.

Last but certainly not least, I want to express my appreciation to my parents, Wenchang Yin and Yongmei Gao, for their unconditional love, sacrifices, and constant encouragement. Your belief in my potential has been my driving force.

Thank you all.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Optimal (re-)insurance problem

Reinsurance, a contractual arrangement offered by reinsurers to safeguard against an insurer's potential claims, works as a substantial instrument for risk distribution, gaining extensive attention in both academia and industry practice. Taking part in reinsurance contracts, insurers can effectively transfer a portion of their risk to reinsurers. This strategic move enables insurers to reduce their exposure and improve their capacity to withstand unforeseen shocks. Hence, a reinsurance contract can help the insurers stabilize their balance sheets and improve their liquidity and solvency, comprehensively contributing to the insurers' resistance against risks.

Initially, insurers operated under the assumption of risk independence, believing that a sufficient number of independent risks ensured help stabilize their cumulative claims. This perception persisted until the recognition of systemic and catastrophic risks, exemplified by events like financial crises, terrorism, and the merely past COVID-19 pandemic. In these scenarios, predicting aggregate reimbursements becomes challenging, exposing insurers to extraordinary losses that could jeopardize their financial stability.

Let us denote by X the underlying (aggregate) risk faced by the insurer. Conventionally, X is assumed to be a non-negative random variable. With the specification of a reinsurance contract, as the aggregate loss X occurs, the reinsurer agrees to pay indemnity $I(X)$ to the insurer and requires a premium P. The premium is usually decided either by a premium principle, see e.g. [Tan et al., 2009], or by the negotiation between the two counterparts, the insurer and the reinsurer, see [Jiang et al., 2019], in the literature. Thus, when a loss

$X = x$ occurs, $I(x)$ is the ceded loss to the reinsurer and the insurer will only need to cover the retained loss $x - I(x)$. The function $I(x)$ is commonly described as compensation function, indemnification function, or ceded loss function, while $R(x) \triangleq x - I(x)$ is known as retained loss function. The total loss faced by the insurer with a reinsurance contract now becomes $X - I(X) + P$. If we denote by W_0 the initial wealth of the insurer, after receiving indemnity from the reinsurer, the insurer's terminal wealth is $W_0 - X + I(X) - P$. In the literature, there are two most popular ways to quantify the effect brought by the reinsurance contract. The first way is based on behavioral economics that assigns a utility function $u(\cdot)$ to the terminal wealth. In this framework, the insurer would like to select a reinsurance contract that maximizes its expected utility of terminal wealth, hence the optimal reinsurance problem becomes:

$$\max_{I \in \mathcal{I}} \mathbb{E}\left[u\left(W_0 - X + I(X) - P\right)\right]. \tag{1.1}$$

Alternatively, we can assign a real number to each well defined retained loss, $X - I(X) + P$, of the insurer, with a risk functional ρ. Commonly used risk measures include Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR), which will be elaborated in Section 1.2. The optimal reinsurance problem then is formulated as:

$$\min_{I \in \mathcal{I}} \rho\left(X - I(X) + P\right). \tag{1.2}$$

In the both formulations of the optimal reinsurance problem, $\mathcal{I}$ is the collection of all feasible indemnity functions. Before giving more detailed explanations about the optimal reinsurance problem, we will first define a "feasible" reinsurance by properly specifying the feasible set $\mathcal{I}$.

Moral hazard is an important issue that needs to be sufficiently addressed when designing an reinsurance contract. An essential principle between the counterparts is that both parties face larger respective risks as the total loss enlarges. Otherwise, the effect of moral hazard will be brought into presence. In other word, if the ceded loss function $I(x)$ is not a non-decreasing function, there exists $0 \leqslant x < y$ such that $I(x) > I(y)$. Suppose, for instance, the aggregate loss reaches the level y, then the insurer may report the partial loss x to the reinsurer and get even more reimbursement than if she reports the entire loss as much as y. This may seriously harm the interest of the reinsurance company. Conversely, if the retained loss function $R(x) = x - I(x)$ is not a non-decreasing function, the insurer might exaggerate the reported loss and get "bonus" beyond the reimbursement. No matter which case it is, the incentive of the insurer to miss-report the loss on purpose could potentially impair the reinsurer's interest. As a result, a commonly accepted indemnity function usually satisfies the following two conditions:

1. $I : [0, +\infty) \to [0, +\infty)$ such that $I(0) = 0$ and I is non-decreasing;
2. $|I(y) - I(x)| \leqslant |y - x|$, for any non-negative x and y.

The second condition is also known as 1-Lipschitz continuous condition or "no-sabotage" condition. Obviously, all deductible contracts satisfy the above conditions and hence are "feasible" contracts according to our definition.

Note that it could be quite challenging to completely avoid moral hazard issues in a reinsurance contract, since there are many reasons that could trigger the occurrence of moral hazard. The above two conditions can only help eliminate those in particular related to intentional miss-reporting.

The "optimality" of a reinsurance contract can be either from the perspective of the insurer, or reflecting the "mutual interest" of the both parties. A "mutual interest"-oriented optimal reinsurance problem can be in the framework of game theory, which normally reflects the "negotiation" between the two parties. More details can be found in [Borch, 1960]. [Arrow, 1963] and [Arrow, 1996] provide the fundamental work on the optimal insurance design problem from the insurer's perspective. Arrow has shown that stop-loss insurance treaty is the optimal solution to the following maximization problem when utility function u is concave:

$$\max_{I \in \mathcal{I}} \mathbb{E}\left[u\left(W_0 - X + I(X) - \mathbb{E}[I(X)]\right)\right]. \tag{1.3}$$

This is a particular case of Problem (1.1) when the premium is calculated by the expected premium principle and fixed to p. The utility function $u(\cdot)$ is commonly assumed to be an increasing concave function. The concavity of u, on one hand, represents the marginal diminishing property of wealth and on the other hand, represents the risk-averse bearing of the insured who is seeking risk sharing. When the initial wealth is non-random, say $W_0 = w_0$ for some constant w_0, Problem (1.3) is equivalent to the following minimization problem

$$\min_{I \in \mathcal{I}} \mathbb{E}[u(X - I(X) + \mathbb{E}[I(X)])],$$

where $u(\cdot)$ is an increasing convex function.

Even though Arrow's insurance model is rooted in the center among the studies of optimal insurance, the work of [Allais, 1953] and [Ellsberg, 1961] put forward a challenge to the foundations of Expected-Utility-Theory (EUT). Those work have influenced decision theory study significantly by promoting preferences for non-EUT models that explains the Allais' paradox and captures other cognitive biases that are not illustrated by EUT. Substitutes to EUT include Rank-Dependent Expected-Utility (RDEU) by [Quiggin, 1982], Dual Theory by [Yaari, 1987] and Cumulative Prospect Theory (CPT) by [Kahneman and Tversky, 2013]. These modern decision making models have also made their way to the insurance studies. Among the recent literature, [Sung et al., 2011] showed that the optimal insurance policy is either an insurance layer or a stop-loss insurance, where an insured's decision-making behavior is modeled by Kahneman and Tversky's CPT with convex probability distortions. [Ghossoub, 2019] formulates an optimal insurance design problem under RDEU and characterized the optimal retention. [Cai and Weng, 2016] shows a reinsurance layer is optimal for the insurer whose risk exposure is quantified by the risk measure of expectile, which is a distributional quantity, as is first proposed by [Newey and Powell, 1987].

A vital component of research in optimal reinsurance is in the framework of utility maximization or risk measure minimization, the latter one being the emphasis of this thesis. [Cai and Tan, 2007] introduces the general risk measures VaR and Conditional Tail Expectation (CTE) into the reinsurance's model and seeks for the optimal stop-loss contracts and optimal quota-share contracts under various premium principle. [Cai et al., 2008] extends the previous work by investigating the optimal deductible and optimal quota-share coefficient corresponding to several kinds of premium principles other than expectation.

In the literature, most of the optimal reinsurance solutions are derived from an insurer's perspective. According to [Borch, 1969], however, an optimal reinsurance treaty for the insurer might not be optimal for its counter-part, or even unacceptable. As a result, it is quite substantial in the study of optimal reinsurance to design a policy that is fair enough and acceptable for the both parties. This framework that reflects such "fairness" is usually from the Pareto-optimal point of view. [Cai et al., 2017] characterizes all Pareto-optimal reinsurance contracts and gives explicit forms of them when both the insurer and the insured use TVaR to account for their losses. [Jiang, 2022] formulates and solves the Pareto-optimal insurance problem under the heterogeneous beliefs of the insurer and the insured.

In this thesis, we will also be committed to finding an optimal reinsurance treaty that is fair enough for the both parties. However, as varying from the traditional model, we will make our optimal policy conservative enough by taking into consideration the worst-case scenarios, with the uncertainty of the true distribution of the total loss. All risks will be quantified using risk measures, whose general preliminary and literature review will be found in the next section.

1.2 Risk measure

It has been extensively discussed how to quantify the risk of a financial position. Before the concept of general risk measures was introduced, researchers had started using the moments to measure a random risk X, with its distribution available, including variance. Variance, however, does not take into account the asymmetry of X, while it is always the right-tail that matters the most. To this end, VaR was introduced to describe the downside risk. Based upon this, other modern risk measures, such as TVaR, which is also called Expected Shortfall (ES), were introduced to better capture the scale of the tail risk beyond a quantile. These improvements also motivated the systematic investigation of the desirable axioms that are supposed to be satisfied by risk measures.

Let Ω be a sample space. In finance, A financial risk is represented by a random variable $X : \Omega \mapsto \mathbb{R}$, where $X(\omega)$ is the realized financial result. In the framework of insurance and risk sharing, in particular, X often represents the total risk born by the insurance players altogether. Since this thesis is in the framework of reinsurance and risk sharing, we always take X as a random variable of "loss". In other word, a large realized value of X represents an extreme loss. A profit is then associated with a negative X, in this sense. Denote $\mathcal{X}$ as the collection of all the risks X that we would like to quantify. When a probability space $(\Omega, \Sigma, \mathbb{P})$ is given, $\mathcal{X}$ is often chose as $L^p(\Omega, \Sigma, \mathbb{P}), p \geqslant 0$.

Definition 1.2.1 *A mapping $\rho : \mathcal{X} \mapsto \mathbb{R}$ is a monetary risk measure if it satisfies the following properties for $X, Y \in \mathcal{X}$:*

- *(C1) Monotonicity: If $X \leqslant Y$, then $\rho(X) \leqslant \rho(Y)$.*
- *(C2) Translation invariance: For $m \in \mathbb{R}$, $\rho(X + m) = \rho(X) + m$.*

These two properties are desirable especially in regulatory capital calculation. (C1) means when risk Y is larger than risk X in all scenarios, then it requires more capital to be put aside. (C2) implies that if a risk is to increase by m in all circumstances, then its capital requirement will increase by the same amount.

Without loss of generality, one would assume a monetary risk measure further satisfies the following property:

- (C3) Normalization: $\rho(0) = 0$.

Definition 1.2.2 *A monetary risk measure $\rho : \mathcal{X} \mapsto \mathbb{R}$ is called a convex risk measure if it further satisfies the following property:*

- *(C4) Convexity:* $\rho(\lambda X + (1-\lambda)Y) \leqslant \lambda\rho(X) + (1-\lambda)\rho(Y)$, *for* $\lambda \in [0,1]$ *and* $X, Y \in \mathcal{X}$.

The property (C4) is desirable because it illustrates the effect of diversification in reducing the portfolio risk. One allocating his/her resources between the financial positions X and Y can reduce the overall risk, compared with his/her holding onto only one type of them. If this property is satisfied, then it is implied that diversification will never increase the risk.

The notion of convex risk measures was introduced in the works by [Follmer and Schied, 2002] as generalization of coherent risk measures.

Definition 1.2.3 *A convex risk measure* $\rho : \mathcal{X} \to \mathbb{R}$ *is called a coherent risk measure if it further satisfies the following axioms:*

- *(C5) Positive homogeneity:* $\rho(\lambda X) = \lambda\rho(X)$, *for* $X \in \mathcal{X}$, $\lambda \geqslant 0$.
- *(C6) Subadditivity:* $\rho(X+Y) \leqslant \rho(X) + \rho(Y)$, *for* $X, Y \in \mathcal{X}$.

Note that with the assumption of positive homogeneity (C5), convexity (C4) is equivalent to subadditivity (C6). Positive homogeneity describes the phenomenon that when homogeneous financial positions are put into one portfolio, then typically the risk is not diversified. Subadditivity, on the other hand, says when two positions are put together into one portfolio, then the portfolio's risk is bounded above by the sum of the individual risks.

The notion of coherent risk measures was firstly introduced in [Artzner et al., 1997] and further developed in [Artzner et al., 1999]. The authors discuss about ways to measure market and non-market risks and propose a set of four desirable properties. Those properties were in response to the criticisms against VaR for not being subadditive and not taking into account the severity of an incurred damage event.

Definition 1.2.4 (Comonotonicity) *Two measurable functions* X *and* Y *on* (Ω, Σ) *are called comonotone if*

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0, \text{ for all } \omega, \omega' \in \Omega.$$

Besides the above mentioned properties, other desirable properties of risk measures are listed below:

- (C7) Comonotonic additivity: $\rho(X+Y) = \rho(X) + \rho(Y)$, whenever $X, Y \in \mathcal{X}$ are comonotone.
- (C0) Law-invariance: $\rho(X) = \rho(Y)$ if $X \overset{d}{=} Y$.

Law-invariance (C0) is a property assumed to be satisfied by all the risk measures that we are interested in, because we do not discriminate between random variables with identical distributions when quantifying their corresponding risks. Comonotonic additivity (C7) corresponds to the case that when two financial positions always move along the same direction, then including them in one portfolio is not a effective diversification in terms of reducing the risk.

Among all the risk measures, the most popular are VaR and TVaR are extensively used in regulatory capital calculation, decision making and risk management.

Definition 1.2.5 *The Value-at-Risk (*VaR*) of a random variable X at level $\alpha \in (0,1)$ is defined as the lower α-quantile of X*

$$\mathrm{VaR}_\alpha(X) = \inf\{t \in \mathbb{R} : \mathbb{P}(X \leqslant t) > \alpha\}.$$

As a valid alternative to VaR, TVaR was proposed by [Acerbi and Tasche, 2002], with the advantage of being a coherent risk measure. Its definition is as follows:

Definition 1.2.6 *The Tail Value-at-Risk of random variable X at level $\alpha \in (0,1)$ is defined as*

$$\mathrm{TVaR}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_p(X)\,\mathrm{d}p.$$

There has been great arguments between VaR and TVaR as the standard tool in quantifying financial risks. Among the extensive discussions, even though VaR is neither coherent, nor sensitive to the tail, being elicitable is one of its advantage against its counterpart, see [Gneiting, 2011]. Elicitability is another vital property for a risk measure since it provides a natural methodology to perform backtesting. TVaR replaces VaR and is taken as the standard risk measure for market risk in banking sector, serving as the most popular risk measure in financial regulation. [Wang and Zitikis, 2021] put forward an axiomatic foundation for TVaR.

Arguing that TVaR has a historical estimator that is non-robust, [Cont et al., 2010] initiates a new risk measure, as known as RVaR, by introducing another parameter that generalizes TVaR.

Definition 1.2.7 *For $0 < \alpha < \beta < 1$, the Range Value-at-Risk (*RVaR*) at levels (α, β) of random variable X is defined as*

$$\mathrm{RVaR}(X) = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} \mathrm{VaR}_p(X) \, \mathrm{d}p.$$

As can be easily seen, RVaR is more flexible than TVaR in the sense that it is the average of VaR levels across a certain range of probabilities, which is not necessarily the tail event.

Among the above three popular risk measures, only TVaR is coherent, since VaR and RVaR fail in satisfying (C4) Convexity and (C6) Sub-additivity. However, all of them satisfy (C7) Comonotonicitic additivity, since they all belong to a more general class of risk measures, distortion risk measure, as defined below:

Definition 1.2.8 (Distortion Risk Measure) *Let $g : [0, 1] \mapsto [0, 1]$ be non-decreasing such that $g(0) = 0$ and $g(1) = 1$. The distortion risk measure of a distribution G, with notation $\rho^g(G)$, is defined as*

$$\rho^g(G) = -\int_{-\infty}^{0} 1 - g(1 - G(x)) \, \mathrm{d}x + \int_{0}^{+\infty} g(1 - G(x)) \, \mathrm{d}x, \tag{1.4}$$

whenever at least one of the two integrals in (1.4) is finite. In this case, g is a distortion function.

Whenever the distortion function g is absolutely continuous, then the distortion risk measure ρ^g has the following representation

$$\rho^g(G) = \int_{0}^{1} \gamma(u) G^{-1}(u) \, \mathrm{d}u, \tag{1.5}$$

where the weight function $\gamma(u) = \partial_{-} g(x)|_{x=1-u}, 0 < u < 1$, satisfies $\int_0^1 \gamma(u) = 1$ and ∂_{-} denotes the derivative from the left.

In this thesis, We may also apply the same notation to a random variable, for example, $\rho^g(X) = \int_0^1 \gamma(u) \mathrm{VaR}_u(X) \, \mathrm{d}u$. Since distortion risk measure satisfies Law-invariance property, one can easily see $\rho^g(X) = \rho^g(G)$ whenever $X \overset{d}{\sim} G$.

1.3 Distributional uncertainty and worst-case distribution

In the classical literature in financial decision making , it is often assumed that the distribution of the crucial underlying random variable is perfectly known. For instance, among the literature attributed to solve optimal investment problems, [Linsmeier and Pearson, 1996] focuses on the probability of loss incurred by the investment portfolio. This chapter defines the portfolio VaR as the minimal level γ such that the probability that the portfolio loss is greater than γ is small enough. This requires complete information about the asset returns' distributions. In comparison, [Markowitz, 1968] considers the following familiar problem of minimizing the portfolio variance, subject to a minimal portfolio mean of return:

$$\min_{w \in \mathcal{W}} w^T \Gamma w, \text{ s.t., } \hat{x}^T w \geqslant \mu,$$

where w is a weight vector, $\hat{x}$ is the vector of mean return of the assets, Γ is the convariance matrix of the asset returns, and μ is a pre-specified lower bound for the portfolio return. In this model, only the two moments instead of the entire distribution function of the asset returns are required. Based on the results given in this chapter, we can derive a reliable investment portfolio as long as the two moments, instead of the entire distribution, are accurately estimated.

More examples in optimal insurance and reinsurance are based on the complete information of the distribution of the total loss X. Apart from the examples in Section (1.1), [Ghossoub, 2019] designs an optimal reinsurance problem under Rank-dependent Expected Utility (RDEU) and characterizes the optimal retention functions in both cases whether the insurer and the reinsurer's distortion functions are identical. [Sung et al., 2011] analyzes the optimal insurance policy from the perspective of an insurer, whose decision behavior is described by Cumulative Prospect Theory introduced by Kahneman and Tversky, see [Kahneman and Tversky, 2013], and shows the optimality of insurance layers as expected value premium principle is selected. As another vital direction, optimal insurance contracts that minimize certain risk measures are visited frequently. Among them, [Tan et al., 2009] investigates the optimal quota-share and stop-loss reinsurance contracts under 17 types of reinsurance premium principles, using both VaR and TVaR as the insurer's adopted risk measure.

In the above classical literature in financial decision making, no matter the risk, or the investment gain, is quantified using utility function or risk functionals, it has been commonly accepted that the underlying distribution of the total risk, or the asset return, is completely known to us. In reality, however, the true underlying distribution is intractable and its estimation is always susceptible to errors. An insightful discussion in [Black and Litterman, 1992] points out that inaccurate point estimates can substantially compromise the efficiency of portfolio investment optimization, by making the real VaR substantially worse than the theoretical value. The aforementioned modelling risk stimulated the development of worst-case risk analysis, where the worst-possible risk level is paid attention to over a set of candidate distributions that reflect the uncertainty or information incompleteness about the underlying distribution.

Distributional uncertainty, also called "distributional ambiguity", "model uncertainty" and so on, describes the case where the true distribution of a certain random variable is not completely known or not unique. Decision-making with the presence of distributional uncertainty has been taken as a question of great interest in fields, such as Insurance, Finance, Economics, Operation Management Science and Control System.

In the field of (re)insurance, the problem of evaluating a risk and making decisions with model uncertainty has been frequently revisited following the work in [Ghaoui et al., 2003]. In this work, the authors define the worst-case VaR and show that the problem can be converted into a semi-definite programming problem, where the distributions are partially known, such that only bounds on the mean and covariance matrix are available. As a cornerstone for the field of worst-case risk analysis, this chapter draws the following conclusion of worst-case VaR:

$$\sup_{F\in\mathcal{S}} \mathrm{VaR}_\alpha(X^F) = \mu + \sigma\sqrt{\frac{\alpha}{1-\alpha}}, \tag{1.6}$$

where $\mathcal{S} := \{F : \int x \,\mathrm{d}F(x) = \mu, \int x^2 \,\mathrm{d}F(x) = \mu^2 + \sigma^2\}, \mu \in \mathbb{R}, \sigma > 0$, represents the collection of all the distributions with fixed mean μ and fixed variance σ^2. In addition, the optimizer of Problem (1.6) is a two-point distribution. [Natarajan et al., 2010] revisits Problem (1.6), and replaces VaR by TVaR in the objective. The authors show that the worst-case TVaR is achieved by the same distribution as that maximizes VaR, and the worst-case VaR and worst-case TVaR are identical over such an uncertainty set.

Motivated by the conclusions about worst-case VaR and TVaR, [Li, 2018] investigates the worst-case convex distortion risk measures, also known as spectral risk measures, when only the first two moments of the distribution are known. As a main conclusion of the chapter,

$$\sup_{F\in\mathcal{S}} \rho^g(X^F) = \mu + \sigma\sqrt{||\gamma||_2^2 - 1}, \tag{1.7}$$

where $\gamma : (0,1) \mapsto \mathbb{R}$ satisfies $\gamma(u) = \partial_- g(x)|_{x=1-u}$ and $\mathcal{S}$, as previous, contains all the distributions with mean μ and variance σ^2. Note that this conclusion narrows down to the results in [Natarajan et al., 2010] when ρ^g is in particular a TVaR.

However, [Bernard et al., 2018] argues that the bounds obtained above are too loose to be relevant to practice and two-point distributions are not attractable enough to be applied in risk management applications. They hence add a Wasserstein distance constraint on the candidate distributions in [Bernard et al., 2020b], such that the uncertainty set excludes the distributions too far away from a reference distribution, even though their mean and variance meet the requirements. In addition, this work obtains the conclusions about all distortion risk measures, including the non-convex ones, such as RVaR. According to this work, the tighter the Wasserstein distance constraint is, the more worst-case distribution looks like the reference distribution, and the more practical it becomes.

Other than specifying the two moments, researchers have been investigating worst-case risk measures under probability distance constraints or other refined information. For instance, [Glasserman and Xu, 2014] considers the worst-case value of a risk measure defined through expectation, over all the distributions with a relative entropy small enough to a reference distribution. [Blanchet and Murthy, 2019] focuses on quantifying the impact brought by model misspecification by measuring a worst-case expected cost, over probability measures that are close enough to a given probability measure μ, i.e.,

$$\sup \int f \, \mathrm{d}\nu : d_c(\mu, \nu) \leqslant \delta, \tag{1.8}$$

for $d_c(\mu, \nu) := \inf\{\int c \, \mathrm{d}\pi : \pi \in \Pi(\mu, \nu)\}$, where $c(x, y)$ specifies the cost of transporting a unit mass from x to y. The work converts the above problem to a 2-dimensional optimization problem. Note that as a special case of Problem (1.8), the problem

$$\sup \mathbb{P}(A) : d_c(\mu, \mathbb{P}) \leqslant \delta$$

can also be converted into a 1-dimensional optimization problem, which provides us with the insights into computing worst-case probabilities and worst-case VaR, simultaneously. In addition, [Bernard et al., 2020a] derives the upper and lower bounds on RVaR of the portfolio loss with the knowledge of mean, variance and unimodality feature.

Other than quantifying the worst-case risk, researchers have been working on making financial decisions with the presence of model uncertainty. The paradigm of Distributionally robust optimization (DRO) is well accepted within this field:

$$\min_{\overrightarrow{w} \in \mathcal{W}} \sup_{F \in \mathcal{F}} \mathbb{E}[f(\overrightarrow{w}, \overrightarrow{X})], \tag{1.9}$$

where $\overrightarrow{w}$ denotes a feasible decision vector in $\mathcal{W}$, $\overrightarrow{X}$ represents a random vector with distribution $F \in \mathcal{F}$ and f is a cost function of $\overrightarrow{w}$ and $\overrightarrow{X}$. The literature then adopts the expected utility to account for the risk non-neutrality of the decision maker and the DRO problem becomes:

$$\min_{\overrightarrow{w} \in \mathcal{W}} \sup_{F \in \mathcal{F}} \mathbb{E}[u(f(\overrightarrow{w}, \overrightarrow{X}))]. \tag{1.10}$$

[Cai et al., 2020] then considers quantifying risks using distortion risk measures and introduces the formulation of Distributionally Robust Risk Optimization (DRRO):

$$\min_{\overrightarrow{w} \in \mathcal{W}} \sup_{F \in \mathcal{F}} \rho_h(f(\overrightarrow{w}, \overrightarrow{X})). \tag{1.11}$$

The authors characterize the worst-case distributions by discovering the hidden convexity in DRRO for a general class of distortion functions.

Other mounts of examples of decision-making with distributional uncertainty are in the field of optimal (re)insurance, which has seldom been visited until recently. In this field, the decision maker contributes to find a robust optimal (re)insurance contract that is insensitive to the underlying distribution of the risk. As instances, [Asimit et al., 2017] obtains the closed-form solution of the optimal robust contract with respect to VaR and provides Linear Programming formulations for TVaR when there are finitely many probability distribution candidates in the uncertainty set. [Birghila and Pflug, 2019], instead, focuses on identifying an optimal robust insurance contract that minimizes the distortion risk functional of the retained loss with budget constraint. The uncertainty set here is the convex hull of a finite set of distributions. The problem is formulated as below:

$$\begin{aligned}
&\inf_{I \in \mathcal{I}} \sup_{F \in \mathcal{C}} \rho^{g_1}(X^F - I(X^F) + \pi^{g,\theta}(I(X^{\hat{F}}))), \qquad (1.12)\\
&\text{s.t. } \pi^{g,\theta}(I(X^{\hat{F}})) \leqslant B.
\end{aligned}$$

[Liu and Mao, 2021] investigates the optimal reinsurance deductible problem using VaR and TVaR, respectively, to quantify the insurer's risk. This work uses two-moment uncertainty set and shows the worst-case VaR and TVaR are both achieved by three-point distributions and the both cases induce the same optimal robust deductible.

In this thesis, we work with a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$ be the set of random variables with finite p-th moment, $p \in [1, \infty]$, on that space. Denote by $\mathcal{M}^2 = \{G(x) = \mathbb{P}(X \leqslant x) | X \in L^2\}$ the space of distribution functions with finite second moments. A positive (resp. negative) realized value of random variable X represents a financial loss (resp. profit) in this chapter. For any random variable X, F_X represents the distribution function of X and $F_X^{-1}(p) = \inf\{x \in \mathbb{R} : F_X(x) > p\}$, $p \in (0, 1)$, is the right-continuous version of the quantile function of F_X. Let U be a uniform random variable on $(0, 1)$.

Chapter 2

The worst-case distributions for the insurer and the reinsurer with deductible insurance

In classic researches on reinsurance, the distribution of the risk X faced by the insurer is often assumed to be precisely known. In reality, however, it can be hard to find the true underlying distribution of the risk faced by the insurer. Typically, researchers use the empirical distribution as the underlying distribution but the estimation to an unknown distribution is prone to errors. To be conservative, the both counterparts should consider about the worst-possible risk levels over their perspective sets of candidate distributions. In this chapter, we seek the worst-case distributions from both the perspectives of the insurer and the reinsurer.

2.1 Notations and model settings

In this chapter, we consider an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let L^p be the set of all random variables having finite p-th moments. A risk measure ρ is a mapping from the set of random variables to $\mathbb{R}$. The class of well-known risk measures of our interest is the class of distortion risk measures defined below.

According to Definition 1.2.8, the distortion risk measure of a random variable Y induced by the function g is defined as

$$\rho^g(Y) = \int_0^{+\infty} g(1 - G_Y(x))\,\mathrm{d}x - \int_{-\infty}^0 (1 - g(1 - G_Y(x)))\,\mathrm{d}x,$$

where G_Y is the distribution of Y. In this case, g is called the distortion function of the distortion risk measure ρ^g.

As a restatement of (1.5), if a distortion function g is absolutely continuous, then the distortion risk measure $\rho^g(Y)$ has the following representation:

$$\rho^g(Y) = \int_0^1 \gamma_g(u)\,G_Y^{-1}(u)\,\mathrm{d}u = \int_0^1 \gamma(u)\,G^{-1}(u)\,\mathrm{d}u, \tag{2.1}$$

where the function $\gamma_g(u) = \gamma(u) = \frac{\partial^- g}{\partial x}\big|_{x=1-u}$, $0 < u < 1$, is called the weight function of the distortion risk measure ρ^g and satisfies $\int_0^1 \gamma_g(u)\,\mathrm{d}u = \int_0^1 \gamma(u)\,\mathrm{d}u = 1$, $\frac{\partial^- f}{\partial x}$ denotes the left derivative of a function f, and $G_Y^{-1}(\alpha) = G^{-1}(\alpha) = \mathrm{VaR}_\alpha(Y) \triangleq \inf\{y : \mathbb{P}(Y \leqslant y) \geqslant \alpha\}$ is the quantile function or the (left-continuous) value-at-risk (VaR) of the distribution G_Y. A distortion risk measure with expression (1.5) is also called a spectral risk measure. To simplify notations, in this chapter, we often omit the subscripts g and Y in γ_g and $G_Y^{-1}(u)$ and use the second expression in (1.5) to denote the distortion risk measure $\rho^g(Y)$ when g is absolutely continuous. In this chapter, we impose the following assumption.

Assumption 2.1.1 *Assume that the weight function γ in (1.5) satisfies $\int_0^1 |\gamma(u)|^2\,\mathrm{d}u < \infty$ and all the random variables considered in the chapter are contained in the L^2 space.*

In practice, if a decision maker has only partial information on the distribution of a loss random variable X, the decision maker may use the empirical distribution $\hat{F}$ of X as a base estimation of the distribution of X. However, due to observation errors or insufficient data, the decision maker would believe that the possible distributions of X lie within a 'range' of the empirical distribution. Such a range can be represented in different ways. One of the commonly used ways is to use Wasserstein distance with order 2.

Definition 2.1.1 (Wasserstein Distance) *Given two distributions F and G, Wasserstein distance between F and G with order 2 is defined as*

$$d_W(F, G) \triangleq \Big(\int_0^1 |F^{-1}(y) - G^{-1}(y)|^2\,\mathrm{d}y\Big)^{1/2} = \big(\mathbb{E}[(F^{-1}(U) - G^{-1}(U))^2]\big)^{1/2},$$

where and throughout this thesis, $H^{-1}(\alpha) = \inf\{x \in \mathbb{R} : H(x) \geqslant \alpha\}$ for $0 < \alpha < 1$ is the quantile function or the (left-continuous) value-at-risk (VaR) of a distribution H, and U represents a uniform random variable on $(0, 1)$.

In addition, in this thesis, when we write an integral, we assume this integral is finite. Note that the Wasserstein distance $d_W(F, G)$ is uniquely determined by the quantile functions of F and G. Hence, we often write $d_W(F^{-1}, G^{-1})$ instead of $d_W(F, G)$.

In the context of stop-loss reinsurances, $X \wedge d$ and $(X-d)_+$ are the loss covered by an insurer and a reinsurer, respectively, where X is the underlying insurance loss faced by an insurer and d is a (stop-loss) retention. The insurer and reinsurer may have different beliefs in the distribution of X and may also have different observed data on X. Therefore, with only partial information on X, both the insurer and reinsurer may use their own uncertainty sets to describe the possible distributions for X. In this chapter, we define an uncertainty set via

$$\mathcal{S}(\mu, \sigma, \hat{F}; \varepsilon) \triangleq \left\{F : \, d_W(F, \hat{F}) \leqslant \varepsilon, \; \mathbb{E}[X^F] = \mu, \, \mathrm{var}(X^F) = \sigma^2\right\}, \tag{2.2}$$

and assume that uncertainty sets of the insurer and reinsurer are

$$\mathcal{S}_1 \triangleq \mathcal{S}(\mu_1, \sigma_1, \hat{F}; \varepsilon_1) \quad \text{and} \quad \mathcal{S}_2 \triangleq \mathcal{S}(\mu_2, \sigma_2, \hat{G}; \varepsilon_2),$$

respectively. We denote the mean and variance of $\hat{F}$ $(\hat{G})$ in $\mathcal{S}_1$ $(\mathcal{S}_2)$ by $\hat{\mu}_1$ $(\hat{\mu}_2)$ and $\hat{\sigma}_1^2$ $(\hat{\sigma}_2^2)$. In addition, we assume that the insurer and reinsure use distortion functions g_1 and g_2, respectively, to measure their own risks.

We point out that the parameters and reference distribution in $\mathcal{S}(\mu_1, \sigma_1, \hat{F}; \varepsilon_1)$ are not necessarily equal to the corresponding parameters and reference distribution in $\mathcal{S}(\mu_2, \sigma_2, \hat{G}; \varepsilon_2)$. For instance, even if the insurer and reinsure use the empirical distribution of X as the reference distributions $\hat{F}$ and $\hat{G}$, the two reference distributions may be still different since the available observed data on the underlying insurance loss X to the insurer and reinsurer may be different.

For any two distributions F and G with means μ_F and μ_G and variances σ_F^2 and σ_G^2, it is easy to see that

$$\begin{aligned}
(d_W(F, G))^2 &= \mathbb{E}[(F^{-1}(U) - G^{-1}(U))^2] \\
&= \mathbb{E}[(F^{-1}(U))^2] + \mathbb{E}[(G^{-1}(U))^2] - 2\,\mathbb{E}[F^{-1}(U)\, G^{-1}(U)] \\
&= (\mu_F - \mu_G)^2 + \sigma_F^2 + \sigma_G^2 - 2\,\sigma_F\,\sigma_G\,\mathrm{corr}(F^{-1}(U), G^{-1}(U)) \\
&\geqslant (\mu_F - \mu_G)^2 + (\sigma_F - \sigma_G)^2.
\end{aligned}$$

Thus, for $i = 1, 2$, to guarantee $\mathcal{S}_i$ to be non-empty, it must hold that $\sqrt{(\mu_i - \hat{\mu}_i)^2 + (\sigma_i - \hat{\sigma}_i)^2} \leqslant \varepsilon_i$. To avoid trivial cases, in this chapter, we impose the following assumption on the parameters in the sets $\mathcal{S}_1$ and $\mathcal{S}_2$.

Assumption 2.1.2 *Assume that the parameters in the sets $\mathcal{S}_1 = \mathcal{S}(\mu_1, \sigma_1, \hat{F}; \varepsilon_1)$ and $\mathcal{S}_2 = \mathcal{S}(\mu_2, \sigma_2, \hat{G}; \varepsilon_2)$ satisfy $(\mu_i - \hat{\mu}_i)^2 + (\sigma_i - \hat{\sigma}_i)^2 \leqslant \varepsilon_i^2$ for $i = 1, 2$.*

We remark that under Assumption 2.1.2, $\mathcal{S}(\mu_1, \sigma_1, \hat{F}; 0) = \{\hat{F}\}$ and $\mathcal{S}(\mu_2, \sigma_2, \hat{G}; 0) = \{\hat{G}\}$ since $(\mu_i - \hat{\mu}_i)^2 + (\sigma_i - \hat{\sigma}_i)^2 \leqslant 0$ implies that $\mu_i = \hat{\mu}_i$ and $\sigma_i = \hat{\sigma}_i$ for $i = 1, 2$. In the rest of the chapter, we first solve the following optimization problems in Sections 2.2 and 2.3:

$$\sup_{F \in \mathcal{S}_1} \rho^{g_1}\left(X^F \wedge d\right) \tag{2.3}$$

$$\sup_{G \in \mathcal{S}_2} \rho^{g_2}\left((X^G - d)_+\right), \tag{2.4}$$

and try to find distributions $F^* \in \mathcal{S}_1$ and distributions $G^* \in \mathcal{S}_2$ such that $\sup_{F \in \mathcal{S}_1} \rho^{g_1}\left(X^F \wedge d\right) = \rho^{g_1}\left(X^{F^*} \wedge d\right)$ and $\sup_{G \in \mathcal{S}_2} \rho^{g_2}\left((X^G - d)_+\right) = \rho^{g_2}\left((X^{G^*} - d)_+\right)$. We then consider the applications of these results in Section 2.4. To make solutions to these problems feasible, in this chapter, we assume that the distortion functions g_1 and g_2 are absolutely continuous. Thus, the expressions in (1.5) apply for ρ^{g_1} and ρ^{g_2}.

Other than that, selecting appropriate sizes of the uncertainty set, namely the values of ε_1 and ε_2, is another substantial topic in this field. Just as [Blanchet et al., 2022] point out, an uncertainty set that is too large can eliminate the significance of the reference distribution at its center, while an excessively small uncertainty set can lack robustness in terms of decision making. Also, the determination of $\varepsilon_i, i = 1, 2$, should be driven by data. In this sense, we would recommend a tentative ε_i selecting criteria that follows:

$$\varepsilon_1 \leqslant C_1 \cdot \tilde{\varepsilon}_1,$$
$$\varepsilon_2 \leqslant C_2 \cdot \tilde{\varepsilon}_2,$$

where $\tilde{\varepsilon}_1 = d_W(\hat{F}, \tilde{F})$, $\tilde{F}$ being the empirical distribution of the true underlying risk in the insurer's opinion and the notations are similar for the reinsurer. This criteria should be combined with Assumption 2.1.2 when specifying the parameters $\varepsilon_i, i = 1, 2$. We recommend taking $C_i \in [1, 2]$ for $i = 1, 2$, to accommodate for the principles given in the previous paragraph. In particular, with $C_i \geqslant 1$, the empirical distributions for each party will fall inside their individual uncertainty sets, in case the empirical distributions are in fact accurate estimations to the true distribution. A suitable selection of C_i should also tolerant errors brought by biased data, and that is why C_i should be larger than 1. But $C_i > 2$ can substantially exaggerate the model risk by allowing too many irrelevant distribution candidates into the decision makers' account.

Due to the incompleteness of the information regarding the decision makers' data accuracy, estimation process and other factors that can affect their assumptions, we only give such recommendations instead of specifications of the parameters ε_i. The decision makers should adjust the size of their uncertainty sets to meet their own needs. Generally speaking, the uncertainty sets should be smaller when the decision makers have more available data points and hence smoother empirical distributions, since the Wasserstein distance between the empirical distribution and the true distribution converges to 0 at a convergence rate of $O(n^{-1})$, according to [Blanchet et al., 2022].

Since the choices of ε_i depend on the empirical distributions, $\tilde{F}$ and $\tilde{G}$, which are the unbiased and consistent estimators of the distribution of the underlying loss random and are random, the choices of ε_i will be a "random process". Furthermore, with the empirical distributions evolving with the accumulation of information, we will obtain empirical distribution series $\{\tilde{F}_n\}$ and $\{\tilde{G}_n\}$. Hence the choices of ε_i' will be a stochastic process. That is another interesting topic that we will explore in the future research.

2.2 Worst-case risk measures of the limited loss random variable

In this section, we focus on problem (2.3) with limited loss random variable. Throughout this section, $\gamma_{g_1} = \gamma_1$ is the weight function in expression (1.5) for ρ^{g_1}. Define $\mathcal{Q}_1$ to be the set containing all the quantile functions whose corresponding distributions belong to $\mathcal{S}_1$. That is,

$$\mathcal{Q}_1 = \left\{F^{-1} : F \in \mathcal{S}_1\right\} = \left\{F^{-1} :\, d_W(F^{-1}, \hat{F}^{-1}) \leqslant \varepsilon_1,\, \mathbb{E}[X^F] = \mu_1,\, \mathrm{var}(X^F) = \sigma_1^2\right\}.$$

Meanwhile, for any $d \geqslant 0$, define $\mathcal{Q}_1^d$ to be the following set of quantile functions:

$$\mathcal{Q}_1^d = \left\{G^{-1} :\, d_W(G^{-1}, \hat{F}^{-1} - d) \leqslant \varepsilon_1,\, \mathbb{E}[X^G] = \mu_1 - d,\, \mathrm{var}(X^G) = \sigma_1^2\right\}.$$

Lemma 2.2.1 *For a given $d \geqslant 0$, it holds that*

$$\sup_{F \in \mathcal{S}_1} \rho^{g_1}(X^F \wedge d) = d + \sup_{G^{-1} \in \mathcal{Q}_1^d} \rho^{g_1}(G^{-1}(U) \wedge 0). \tag{2.5}$$

Proof. First we recall a well-know result that for any quantile function F^{-1}, the random variable $F^{-1}(U)$ has the distribution F. Therefore, by using the law-invariance property of distortion risk measures, we can write problem (2.3) as

$$\sup_{F \in \mathcal{S}_1} \rho^{g_1}\left(X^F \wedge d\right) = \sup_{F \in \mathcal{S}_1} \rho^{g_1}\left(F^{-1}(U) \wedge d\right) = \sup_{F^{-1} \in \mathcal{Q}_1} \rho^{g_1}\left(F^{-1}(U) \wedge d\right). \tag{2.6}$$

Now, for any $F^{-1} \in \mathcal{Q}_1$, we can define quantile function G^{-1} as $G^{-1}(u) = F^{-1}(u) - d$, $u \in (0,1)$. Note that for any distributions F and G and any constant a, $d_W(F^{-1}+a, G^{-1}) =$

$d_W(F^{-1}, G^{-1}-a)$. Thus, $d_W(G^{-1}, \hat{F}^{-1}-d) = d_W(G^{-1}+d,\ \hat{F}^{-1}) = d_W(F^{-1}, \hat{F}^{-1}) \leqslant \varepsilon_1$. Together with facts $\mathbb{E}[G^{-1}(U)] = \mathbb{E}[F^{-1}(U)] - d = \mu_1 - d$ and $\mathrm{var}(G^{-1}(U)) = \mathrm{var}(F^{-1}(U)) = \sigma_1^2$, we get $G^{-1} \in \mathcal{Q}_1^d$. Furthermore, note that for any $x, y \in \mathbb{R}$, $x \wedge y = (x-y) \wedge 0 + y$. Thus, by the cash-invariant property of distortion risk measures, we have

$$\rho^{g_1}\left(F^{-1}(U) \wedge d\right) = \rho^{g_1}((F^{-1}(U) - d) \wedge 0) + d = \rho^{g_1}\left(G^{-1}(U) \wedge 0\right) + d.$$

Conversely, for any $G^{-1} \in \mathcal{Q}_1^d$, we define quantile F^{-1} as $F^{-1}(u) = G^{-1}(u) + d$, $u \in (0,1)$. Hence, $d_W(F^{-1}, \hat{F}^{-1}) = d_W(G^{-1} + d,\ \hat{F}^{-1}) = d_W(G^{-1},\ \hat{F}^{-1} - d) \leqslant \varepsilon_1$, $\mathbb{E}[X^F] = \mathbb{E}[F^{-1}(U)] = \mathbb{E}[G^{-1}(U) + d] = \mu_1$, and $\mathrm{var}(X^F) = \mathrm{var}(G^{-1}(U) + d) = \sigma_1^2$, which mean that $F \in \mathcal{Q}_1$. Moreover, by using $x \wedge y = (x-y) \wedge 0 + y$ and the cash-invariant property of distortion risk measures again, we have $\rho^{g_1}(G^{-1}(U) \wedge 0) + d = \rho^{g_1}(F^{-1}(U) \wedge d)$. Therefore,

$$\sup_{F^{-1} \in \mathcal{Q}_1} \rho^{g_1}\left(F^{-1}(U) \wedge d\right) = \sup_{F^{-1} \in \mathcal{Q}_1^d} \left(\rho^{g_1}\left(G^{-1}(U) \wedge 0\right) + d\right) = d + \sup_{F^{-1} \in \mathcal{Q}_1^d} \rho^{g_1}\left(G^{-1}(U) \wedge 0\right),$$

which, together with (2.6), implies that (2.5) holds. ■

As a consequence of Lemma 2.2.1, the problem (2.3) is reduced to the following problem

$$\sup_{G^{-1} \in \mathcal{Q}_1^d} \rho^{g_1}\left(G^{-1}(U) \wedge 0\right), \tag{2.7}$$

which means that F^* is a maximizer to the problem (2.3) over $\mathcal{S}_1$ if and only if F^{*-1} is a maximizer to the problem (2.7) over $\mathcal{Q}_1$. To solve the problem (2.7) and simplify notations, for any given $\beta \in [0,1]$, we define

$$\gamma_{1,\beta}(u) \triangleq \gamma_1(u)\, \mathrm{I}_{[0,\beta]}(u), \qquad u \in (0,1), \tag{2.8}$$

where and throughout this chapter, I_A is an indicator function, which means that $\mathrm{I}_A(u) = 1$ if $u \in A$ and 0 otherwise. Furthermore, we define L_0 as a function of $\beta \in [0,1]$ and a quantile function F^{-1} with

$$L_0(\beta,\ F^{-1}) \triangleq \int_0^1 \gamma_{1,\beta}(u) F^{-1}(u)\, \mathrm{d}u. \tag{2.9}$$

In addition, for the weight function γ_1 of the distortion risk measure ρ^{g_1}, we define constant α_1 as follows: If $\gamma_1(u) > 0$ on an open interval $(0, \delta)$ for some $0 < \delta < 1$, define $\alpha_1 \triangleq 0$, otherwise, define

$$\alpha_1 \triangleq \sup\left\{0 < u < 1 : \int_0^u \gamma_1(t)\, \mathrm{d}t = 0\right\}. \tag{2.10}$$

Note that a distortion function is increasing. Thus, for an absolutely continuous distortion function g_1, we know that $0 \leqslant \alpha_1 < 1$, $\gamma_1(u) \geqslant 0$ on $(0,1)$, and $\gamma_1(u) = 0$ on $(0, \alpha_1]$ almost everywhere.

In addition, since for any distribution F, the distribution of the random variable $F^{-1}(U)$ is equal to F, and $x \wedge 0$ is an increasing function in x, by the invariance property of the VaR on an increasing transform, we have for any $u \in (0,1)$,

$$\mathrm{VaR}_u(F^{-1}(U) \wedge 0) = \mathrm{VaR}_u(F^{-1}(U)) \wedge 0 = F^{-1}(u) \wedge 0, \tag{2.11}$$

which means that the quantile function of the random variable $F^{-1}(U) \wedge 0$ is $F^{-1}(u) \wedge 0$ for $0 < u < 1$. By the monotonicity of a distortion risk measure, it holds that $\rho^{g_1}\left(G^{-1}(U) \wedge 0\right) \leqslant 0$ for any quantile function G^{-1}. Thus, for any absolutely continuous distortion function g_1, if a quantile function $G^{-1} \in \mathcal{Q}_1^d$ satisfies $G^{-1}(u) \geqslant 0$ for any $u \in (\alpha_1, 1)$, then by (1.5), (2.11), and the definition of α_1, we have $\rho^{g_1}(G^{-1}(U) \wedge 0) = \int_0^{\alpha_1} \gamma_1(u)\left(G^{-1}(u) \wedge 0\right) \mathrm{d}u + \int_{\alpha_1}^1 \gamma_1(u)\left(G^{-1}(u) \wedge 0\right) \mathrm{d}u = 0$. To avoid such trivial cases, we impose the following assumption in this section.

Assumption 2.2.1 *Assume that* $\sup\limits_{G^{-1} \in \mathcal{Q}_1^d} \rho^{g_1}\left(G^{-1}(U) \wedge 0\right) < 0$ *in* (2.7) *for a given* $d \geqslant 0$.

Lemma 2.2.2 *For any quantile function* F^{-1}, *it holds that*

$$\rho^{g_1}(F^{-1}(U) \wedge 0) = \min_{\beta \in [\alpha_1, 1]} L_0(\beta,\, F^{-1}) = L_0(F(0),\, F^{-1}). \tag{2.12}$$

In particular, if $F(0) \leqslant \alpha_1$, *then* $\rho^{g_1}(F^{-1}(U) \wedge 0) = 0$. *Moreover, for any* $d \geqslant 0$, *if Assumption* 2.2.1 *holds, then for all* $F^{-1} \in \mathcal{Q}_1^d$, *it must hold that* $F(0) > \alpha_1$.

Proof. Note that for a (left-continuous) quantile function F^{-1}, it holds that for any $0 < u < 1$, $u \leqslant F(x) \Longleftrightarrow F^{-1}(u) \leqslant x$, and $F(x) < u \Longleftrightarrow x < F^{-1}(u)$. Therefore, we have

$$\begin{cases} F^{-1}(u) \leqslant 0, & \text{for } 0 < u \leqslant F(0), \\ F^{-1}(u) > 0, & \text{for } F(0) < u < 1. \end{cases} \tag{2.13}$$

By (2.9) and (2.8), we have

$$L_0(\beta, F^{-1}) = \int_0^1 \gamma_{1,\beta}(u)\, F^{-1}(u)\, \mathrm{d}u = \int_0^\beta \gamma_1(u)\, F^{-1}(u)\, \mathrm{d}u. \tag{2.14}$$

Since $\gamma_1(u)$ is non-negative on $(0,1)$ and $L_0(\beta, F^{-1})$ is continuous in $\beta \in [0,1]$, we see by (2.13) and (2.14) that $L_0(\beta, F^{-1}) \leqslant 0$ and it decreases in $\beta \in [0, F(0)]$. Moreover, $L_0(\beta, F^{-1})$ increases in $\beta \in (F(0), 1]$. Hence, $\min_{0\leqslant\beta\leqslant 1} L_0(\beta, F^{-1})$ is attainable at $\beta = F(0)$, which means that

$$\min_{0\leqslant\beta\leqslant 1} L_0(\beta, F^{-1}) = L_0(F(0), F^{-1}) \leqslant 0. \tag{2.15}$$

By the definition of α_1, we have $\gamma_1(u) = 0$ on $(0, \alpha_1]$ almost everywhere. Thus, by (2.14), we have

$$L_0(\beta, F^{-1}) = L_0(\alpha_1, F^{-1}) = 0, \qquad \text{for any } 0 \leqslant \beta \leqslant \alpha_1. \tag{2.16}$$

It follows from (2.15) and (2.16) that

$$\min_{0\leqslant\beta\leqslant 1} L_0(\beta, F^{-1}) = \min_{\alpha_1\leqslant\beta\leqslant 1} L_0(\beta, F^{-1}) = L_0(F(0), F^{-1}) \leqslant 0,$$

which, together with (1.5), (2.11), and (2.13), implies that

$$\begin{aligned}\rho^{g_1}(F^{-1}(U) \wedge 0) &= \int_0^1 \gamma_1(u)\, \left(F^{-1}(u) \wedge 0\right)\, \mathrm{d}u = \int_0^{F(0)} \gamma_1(u)\, F^{-1}(u)\, \mathrm{d}u \\ &= L_0(F(0), F^{-1}) = \min_{\beta\in[\alpha_1,1]} L_0(\beta,\, F^{-1}) \leqslant 0. \end{aligned} \tag{2.17}$$

Hence, (2.12) holds. In particular, if $F(0) \leqslant \alpha_1$, by (2.17) and (2.16), we have $\rho^{g_1}(F^{-1}(U) \wedge 0) = 0$. Moreover, for any $d \geqslant 0$, if there exists a quantile function $F^{-1} \in \mathcal{Q}_1^d$ with $F(0) \leqslant \alpha_1$, then by what we just proved, we have $\rho^{g_1}(F^{-1}(U) \wedge 0) = 0$, which yields $\sup_{F^{-1}\in\mathcal{Q}_1^d} \rho^{g_1}(F^{-1}(U) \wedge 0) \geqslant 0$, a contradiction of Assumption 2.2.1. Thus, $F(0) > \alpha_1$. ■

By Lemma 2.2.2, for a given $d \geqslant 0$, we have

$$\sup_{F^{-1}\in\mathcal{Q}_1^d} \rho^{g_1}(F^{-1}(U) \wedge 0) = \sup_{F^{-1}\in\mathcal{Q}_1^d} \min_{\beta\in[\alpha_1,1]} L_0(\beta, F^{-1}) \leqslant \inf_{\beta\in[\alpha_1,1]} \sup_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta, F^{-1}), \tag{2.18}$$

where the inequality holds since $\sup_{F^{-1}\in\mathcal{Q}_1^d} \min_{\beta\in[\alpha_1,1]} L_0(\beta, F^{-1}) \leqslant \sup_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta, F^{-1})$ for any $\beta \in [\alpha_1, 1]$. Throughout this section, for a given $d \geqslant 0$, we denote the inf-sup value in (2.18) by $B = B(d)$, namely

$$B = B(d) = \inf_{\beta\in[\alpha_1,1]} \sup_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta, F^{-1}). \tag{2.19}$$

The value B is an upper bound for $\sup_{F^{-1}\in\mathcal{Q}_1^d}\rho^{g_1}(F^{-1}(U)\wedge 0)$ in (2.18). We first solve the inf-sup problem in (2.19) and obtain an expression for B and then show that the upper bound B is equal to $\sup_{F^{-1}\in\mathcal{Q}_1^d}\rho^{g_1}(F^{-1}(U)\wedge 0)$ in (2.18) under some conditions.

To solve the inf-sup problem in (2.19), we first consider the inner optimization problem in (2.19), namely, consider the problem $\sup_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta, F^{-1})$ for a fixed $\beta\in[\alpha_1, 1]$. Note that for $\beta>\alpha_1$ the L^1-norm of $\gamma_{1,\beta}$ is positive, i.e., $\|\gamma_{1,\beta}\|_1=\int_0^1\gamma_{1,\beta}(u)\,\mathrm{d}u>0$. Hence, the following function is well-defined:

$$\tilde{g}_{1,\beta}(t)=1-\int_0^{1-t}\frac{\gamma_{1,\beta}(u)}{\|\gamma_{1,\beta}\|_1}\,\mathrm{d}u,\qquad \text{for } t\in[0,1]. \tag{2.20}$$

It is easy to see that $\tilde{g}_{1,\beta}$ is absolutely continuous, non-decreasing, $\tilde{g}_{1,\beta}(0)=0$, and $\tilde{g}_{1,\beta}(1)=1$. Thus, $\tilde{g}_{1,\beta}$ is a well-defined distortion function and its corresponding distortion risk measure is denoted by $\rho^{\tilde{g}_{1,\beta}}$. By using (1.5) and noticing that $\frac{\partial^-\tilde{g}_{1,\beta}(t)}{\partial t}=\frac{\gamma_{1,\beta}(1-t)}{\|\gamma_{1,\beta}\|_1}$ almost everywhere on $t\in(0,1)$, we have

$$\rho^{\tilde{g}_{1,\beta}}(X^F)=\int_0^1\frac{\gamma_{1,\beta}(u)}{\|\gamma_{1,\beta}\|_1}F^{-1}(u)\,\mathrm{d}u=\frac{1}{\|\gamma_{1,\beta}\|_1}L_0(\beta,F^{-1}), \tag{2.21}$$

where $\|\gamma_{1,\beta}\|_1$ depends on β only. Therefore, for a fixed $\beta\in[\alpha_1,1]$, the problem $\sup_{F^{-1}\in\mathcal{Q}_1^d}L_0(\beta,F^{-1})$ is reduced to the problem $\sup_{F^{-1}\in\mathcal{Q}_1^d}\rho^{\tilde{g}_{1,\beta}}(X^F)$, and maximizers to $\sup_{F^{-1}\in\mathcal{Q}_1^d}L_0(\beta,F^{-1})$ are the same as those to $\sup_{F^{-1}\in\mathcal{Q}_1^d}\rho^{\tilde{g}_{1,\beta}}(X^F)$, which will be solved by applying Theorem 2 of [Bernard et al., 2020b]. For a convenient reference, Theorem 2 of [Bernard et al., 2020b] is restated as Lemma 2.6.1 in the appendix of our chapter. In doing so, we recall the concept of *isotonic projection*. Let

$$\mathcal{K}=\left\{k:(0,1)\mapsto\mathbb{R}\;\middle|\;\int_0^1 k(u)^2\,\mathrm{d}u<\infty,\ k(u)\ \text{ is a non-decreasing function on } (0,1)\right\} \tag{2.22}$$

be the space of square-integrable and non-decreasing functions on $(0,1)$. Denote the isotonic projection of a function $\gamma\in L^2(0,1)$ onto $\mathcal{K}$ as $k_\gamma^\uparrow=\arg\min_{k\in\mathcal{K}}||\gamma-k||_2$, where $||\cdot||_2$ denotes the L^2 norm. Note that $\mathcal{K}$ is a non-empty closed convex set and the isotonic projection $k_\gamma^\uparrow$ uniquely exists for any $k\in\mathcal{K}$. See, for example, [Németh, 2003] for detailed discussions and properties of isotonic projections.

Before moving on, we introduce new notation. For $\lambda\geqslant 0$ and $\beta\in[\alpha_1,1]$, let $\ell_{\beta,\lambda}^\uparrow$ be the isotonic projection of $\gamma_{1,\beta}+\lambda\hat{F}^{-1}$, i.e.,

$$\ell_{\beta,\lambda}^\uparrow=\arg\min_{\ell\in\mathcal{K}}||\ell-\gamma_{1,\beta}-\lambda\hat{F}^{-1}||_2. \tag{2.23}$$

Roughly speaking the isotonic projection $\ell^{\uparrow}_{\beta,\lambda}$ is the best approximation to the function $\gamma_{1,\beta} + \lambda\hat{F}^{-1}$ among all the square-integrable and non-decreasing functions on $(0, 1)$ under the L^2 norm, in the sense that it has the largest correlation with $\gamma_{1,\beta} + \lambda\hat{F}^{-1}$.

Assumption 2.2.2 *Assume that the isotonic projection $\ell^{\uparrow}_{\beta,\lambda}$ defined in (2.23) is not constant for any $\lambda > 0$ and any $\beta \in (\alpha_1, 1]$.*

Now, we are ready to present and prove the following Theorem 2.2.3, which is the first main result in this section. To do so, we define constants $c_{1,\beta}$ as

$$c_{1,\beta} \triangleq \operatorname{corr}(\hat{F}^{-1}(U),\, \ell^{\uparrow}_{\beta,0}(U)). \tag{2.24}$$

In addition, for any $\beta \in [\alpha_1, 1]$ and $\lambda \geqslant 0$, let $a_{\beta,\lambda} = \mathbb{E}\big[\ell^{\uparrow}_{\beta,\lambda}(U)\big]$ and $b_{\beta,\lambda} = \sqrt{\operatorname{var}(\ell^{\uparrow}_{\beta,\lambda}(U))}$, and let $F^{-1}_{\beta,\lambda}(u)$ for $0 < u < 1$ be a quantile function defined as

$$F^{-1}_{\beta,\lambda}(u) = \begin{cases} \mu_1 + \sigma_1\left(\frac{\ell^{\uparrow}_{\beta,\lambda}(u) - a_{\beta,\lambda}}{b_{\beta,\lambda}}\right), & \text{if } b_{\beta,\lambda} > 0, \\ \mu_1, & \text{if } b_{\beta,\lambda} = 0. \end{cases} \tag{2.25}$$

Theorem 2.2.3 *Suppose Assumptions 2.2.1 and 2.2.2 hold. Then, there exist $\beta_0 \in [\alpha_1, 1]$ and $\lambda_0 \geqslant 0$ satisfying*

$$\inf_{\beta\in[\alpha_1,1]} \sup_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta, F^{-1}) = L_0(\beta_0, F^{-1}_{\beta_0,\lambda_0} - d), \tag{2.26}$$

where L_0 is defined in (2.14) and $F^{-1}_{\beta_0,\lambda_0}$ is defined by (2.25). Furthermore,

$$\beta_0 = \operatorname*{arg\,min}_{\beta\in[\alpha_1,1]} \left\{ \|\gamma_{1,\beta}\|_1\, \rho^{\tilde{g}_{1,\beta}}(F^{-1}_{\beta,\lambda_\beta}(U) - d) \right\}, \tag{2.27}$$

and $\lambda_0 = \lambda_{\beta_0}$, where, for any $\beta \in [\alpha_1, 1]$, the following statements hold:

(i) If

$$(\hat{\mu}_1 - \mu_1)^2 + (\hat{\sigma}_1 - \sigma_1)^2 < \varepsilon_1^2 < (\hat{\mu}_1 - \mu_1)^2 + (\hat{\sigma}_1 - \sigma_1)^2 + 2\sigma_1\hat{\sigma}_1(1 - c_{1,\beta}), \tag{2.28}$$

then $\lambda_\beta > 0$ is the unique solution to the equation $d_W(\hat{F}^{-1}, F^{-1}_{\beta,\lambda}) = \varepsilon_1$ for $\lambda \in (0, \infty)$.

(ii) If

$$(\hat{\mu}_1 - \mu_1)^2 + (\hat{\sigma}_1 - \sigma_1)^2 + 2\sigma_1\hat{\sigma}_1(1 - c_{1,\beta}) \leqslant \varepsilon_1^2, \tag{2.29}$$

then $\lambda_\beta = 0$. Moreover, if $\ell^{\uparrow}_{\beta,0}$ is non-constant, then $F^{-1}_{\beta,0} - d \in \mathcal{Q}_1^d$; if $\ell^{\uparrow}_{\beta,0}$ is constant, then $F^{-1}_{\beta,0} - d = \mu_1 - d \notin \mathcal{Q}_1^d$.

Proof. We first show that the value B defined in (2.19) is achieved at some $\beta_0 \in [\alpha_1, 1]$. Write $\bar{L}_0(\beta) \triangleq \sup_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta, F^{-1})$. Then, for any $\beta_1 \in [\alpha_1, 1]$ and any $x < \bar{L}_0(\beta_1)$, there exists a quantile function $F_1^{-1} \in \mathcal{Q}_1^d$ such that $x < L_0(\beta_1, F_1^{-1})$. Since $L_0(\beta, F_1^{-1}) = \int_0^\beta \gamma_1(u)F_1^{-1}(u)\,\mathrm{d}u$ is continuous in $\beta \in [\alpha_1, 1]$, there is an open neighbourhood $I \subset [\alpha_1, 1]$ with $\beta_1 \in I$ such that $x < L_0(\beta, F_1^{-1})$ for all $\beta \in I$. Therefore $x < L_0(\beta, F_1^{-1}) \leqslant \sup_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta, F^{-1}) = \bar{L}_0(\beta)$ for all $\beta \in I$. Hence, $\bar{L}_0(\beta)$ is lower semicontinuous.[1]

By (2.19), $B = \inf_{\beta\in[\alpha_1,1]} \bar{L}_0(\beta)$. Thus, there is a sequence $\{\beta_n\} \subset [\alpha_1, 1]$ such that $\bar{L}_0(\beta_n) \downarrow B$ as $n \to \infty$. Without loss of generality, we assume $\beta_n \to \beta_0$ for some $\beta_0 \in [\alpha_1, 1]$ (indeed, there is a subsequence of $\{\beta_n\}$ such that the subsequence converges to some $\beta_0 \in [\alpha_1, 1]$ since $\{\beta_n\}$ is a bounded sequence). Hence, $B \leqslant \bar{L}_0(\beta_0) \leqslant \lim_{n\to\infty} \bar{L}_0(\beta_n) = B$, where the second inequality comes from the lower semicontinuity of $\bar{L}_0$. Thus, $B = \bar{L}_0(\beta_0)$, i.e., the value B is achieved at β_0. Therefore, we have

$$B = \bar{L}_0(\beta_0) = \sup_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta_0, F^{-1}) = \min_{\beta\in[\alpha_1,1]} \sup_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta, F^{-1}). \tag{2.30}$$

Note that the distortion function $\tilde{g}_{1,\beta}$ in (2.20) of the distortion risk measure $\rho^{\tilde{g}_{1,\beta}}$ is undefined when $\beta = \alpha_1$ since $\|\gamma_{1,\alpha_1}\|_1 = 0$. Now, in the following proof, we define $\rho^{\tilde{g}_{1,\alpha_1}}$ as a mapping such that $\rho^{\tilde{g}_{1,\alpha_1}}(X) = 0$ for any random variable X. Thus, (2.21) implies for any $F^{-1} \in \mathcal{Q}_1^d$,

$$L_0(\beta, F^{-1}) = \|\gamma_{1,\beta}\|_1\, \rho^{\tilde{g}_{1,\beta}}(X^F) \tag{2.31}$$

holds for any $\beta \in [\alpha_1, 1]$ since $L_0(\alpha_1, F^{-1}) = 0$ for $F^{-1} \in \mathcal{Q}_1^d$ by (2.16). Hence, by (2.30), we have

$$B = \min_{\beta\in[\alpha_1,1]} \Big\{ \|\gamma_{1,\beta}\|_1 \sup_{F^{-1}\in\mathcal{Q}_1^d} \rho^{\tilde{g}_{1,\beta}}(X^F) \Big\}. \tag{2.32}$$

For any $\beta \in (\alpha_1, 1]$ and $\lambda \geqslant 0$, denote

$$\tilde{\ell}^{\uparrow}_{\beta,\lambda,d} = \arg\min_{\ell\in\mathcal{K}} \Big\| \ell - \frac{\gamma_{1,\beta}}{\|\gamma_{1,\beta}\|_1} - \lambda(\hat{F}^{-1} - d) \Big\|_2, \tag{2.33}$$

[1]A function $f : [0, \infty] \mapsto [-\infty, +\infty]$ is lower semincontinuous at a point $d \in [0, \infty]$ if for every $y < f(d)$ there exists a neighborhood I of d such that $f(x) > y$ for all $x \in I$. Equivalently, f is lower semincontinuous at d if and only if $\liminf_{x\to d} f(x) \geqslant f(d)$.

with $\tilde{a}_{\beta,\lambda,d} = \mathbb{E}[\tilde{\ell}^{\uparrow}_{\beta,\lambda,d}(U)]$ and $\tilde{b}_{\beta,\lambda,d} = \sqrt{\mathrm{var}(\tilde{\ell}^{\uparrow}_{\beta,\lambda,d}(U))}$. For a fixed $\beta \in (\alpha_1, 1]$ and $d \geqslant 0$, it is easy to verify that for any $\lambda \geqslant 0$,

$$\tilde{\ell}^{\uparrow}_{\beta,\lambda,d} = \frac{1}{\|\gamma_{1,\beta}\|_1}\left(\arg\min_{\ell\in\mathcal{K}} \left\|\ell - \gamma_{1,\beta} - \lambda\|\gamma_{1,\beta}\|_1\, \hat{F}^{-1}\right\|_2\right) - \lambda d = \frac{1}{\|\gamma_{1,\beta}\|_1}\,\ell^{\uparrow}_{\beta,\lambda\|\gamma_{1,\beta}\|_1} - \lambda d. \tag{2.34}$$

For any random variables X_1 and X_2 and any constants $a_1 > 0$, b_1, $a_2 > 0$, b_2, we have $\mathrm{corr}(a_1X_1 + b_1, a_2X_2 + b_2) = \mathrm{corr}(X_1, X_2)$. Hence, $\mathrm{corr}(\hat{F}^{-1}(U),\ \ell^{\uparrow}_{\beta,\lambda\|\gamma_{1,\beta}\|_1}(U)) = \mathrm{corr}(\hat{F}^{-1}(U)-d,\ \tilde{\ell}^{\uparrow}_{\beta,\lambda,d})$ for any $\beta \in (\alpha_1, 1]$. Thus, by (2.24), $c_{1,\beta} = \mathrm{corr}(\hat{F}^{-1}(U),\ \ell^{\uparrow}_{\beta,0}(U)) = \mathrm{corr}(\hat{F}^{-1}(U) - d,\ \tilde{\ell}^{\uparrow}_{\beta,0,d}(U))$. Meanwhile, by (2.34), we see that $\tilde{\ell}^{\uparrow}_{\beta,\lambda,d}$ is a constant if and only if $\ell^{\uparrow}_{\beta,\lambda\|\gamma_{1,\beta}\|_1}$ is a constant. When $\tilde{b}_{\beta,\lambda,d} > 0$, we have $(\tilde{\ell}^{\uparrow}_{\beta,\lambda,d}(u) - \tilde{a}_{\beta,\lambda,d})/\tilde{b}_{\beta,\lambda,d} = (\ell^{\uparrow}_{\beta,\lambda\|\gamma_{1,\beta}\|_1}(u) - a_{\beta,\lambda\|\gamma_{1,\beta}\|_1})/b_{\beta,\lambda\|\gamma_{1,\beta}\|_1}$ for $0 < u < 1$. Therefore, under Assumption 2.2.2, we conclude the following results by applying Lemma 2.6.1 to the problem $\sup_{F^{-1}\in\mathcal{Q}_1^d} \rho^{\tilde{g}_{1,\beta}}(X^F)$ in (2.32) for a given $\beta \in (\alpha_1, 1]$:

(i) Assume that (2.28) holds. Then the maximizer $F^{*-1}_{\beta,\lambda_\beta} \in \mathcal{Q}_1^d$ to $\sup\limits_{F^{-1}\in\mathcal{Q}_1^d} \rho^{\tilde{g}_{1,\beta}}(X^F)$ is unique with $F^{*-1}_{\beta,\lambda_\beta} = F^{-1}_{\beta,\lambda_\beta} - d$, where $F^{-1}_{\beta,\lambda_\beta}$ is defined in (2.25) and $\lambda_\beta > 0$ is the unique positive solution to the equation $d_W(\hat{F}^{-1}, F^{-1}_{\beta,\lambda}) = \varepsilon_1$ for $\lambda \in (0,\infty)$.

(ii) Assume that (2.29) holds. Then $\lambda_\beta = 0$. Moreover, if $\tilde{\ell}^{\uparrow}_{\beta,0,d}$ is not a constant, i.e., $\ell^{\uparrow}_{\beta,0}$ is not a constant, then the maximizer $F^{*-1}_{\beta,0}$ to $\sup_{F^{-1}\in\mathcal{Q}_1^d} \rho^{\tilde{g}_{1,\beta}}(X^F)$ is unique with $F^{*-1}_{\beta,0} = F^{-1}_{\beta,0} - d \in \mathcal{Q}_1^d$, where $F^{-1}_{\beta,0}$ is defined in (2.25).

On the other hand, if $\tilde{\ell}^{\uparrow}_{\beta,0,d}$ is a constant, i.e., $\ell^{\uparrow}_{\beta,0}$ is a constant, then the maximizer $F^{*-1}_{\beta,0}$ to $\sup_{F^{-1}\in\mathcal{Q}_1^d} \rho^{\tilde{g}_{1,\beta}}(X^F)$ is $F^{*-1}_{\beta,0} = F^{-1}_{\beta,0} - d = \mu_1 - d \notin \mathcal{Q}_1^d$ and the supremum $\sup_{F^{-1}\in\mathcal{Q}_1^d} \rho^{\tilde{g}_{1,\beta}}(X^F)$ is not attained in $\mathcal{Q}_1^d$.

Consequently, by (2.30) and (2.32), we have

$$\begin{aligned} B = B(d) &= \min_{\beta\in[\alpha_1,1]} L_0(\beta, F^{-1}_{\beta,\lambda_\beta} - d) = L_0(\beta_0, F^{-1}_{\beta_0,\lambda_{\beta_0}} - d) = \|\gamma_{1,\beta_0}\|_1\, \rho^{\tilde{g}_{1,\beta_0}}(F^{-1}_{\beta_0,\lambda_{\beta_0}}(U) - d) \\ &= \min_{\beta\in[\alpha_1,1]} \left\{\|\gamma_{1,\beta}\|_1\, \rho^{\tilde{g}_{1,\beta}}(F^{-1}_{\beta,\lambda_\beta}(U) - d)\right\} \end{aligned}$$

and $L_0(\beta_0, F^{-1}_{\beta_0,\lambda_{\beta_0}} - d) = \sup\limits_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta_0, F^{-1})$. If (2.28) holds, then $\lambda_0 \triangleq \lambda_{\beta_0} > 0$ is the unique positive solution to the equation $d_W(\hat{F}^{-1}, F^{-1}_{\beta_0,\lambda}) = \varepsilon_1$ for $\lambda \in (0,\infty)$; if (2.29) holds, then $\lambda_0 \triangleq \lambda_{\beta_0} = 0$. Moreover, if (2.29) holds, $\lambda_\beta = 0$ for any $\beta \in [\alpha_1, 1]$ in (2.27) and

$$B = B(d) = \|\gamma_{1,\beta_0}\|_1 \; \rho^{\tilde{g}_{1,\beta_0}}(F^{-1}_{\beta_0,0}(U) - d) = \min_{\beta\in[\alpha_1,1]} \left\{ \|\gamma_{1,\beta}\|_1 \; \rho^{\tilde{g}_{1,\beta}}(F^{-1}_{\beta,0}(U) - d) \right\}. \quad (2.35)$$

Thus, we complete the proof. ■

Theorem 2.2.3 solves the inf-sup problem (2.19), gives the expression of the inf-sup value B that is the upper bound of the worst-case risk measure in (2.18), and characterizes the quantile function under which the upper bound can be attained. More importantly, with Theorem 2.2.3, one reduces the infinite-dimensional optimization problem (2.19) to the one-dimensional optimization problem (2.27). Especially, the parameters β_0 and λ_0 of the worst-case quantile function $F^{-1}_{\beta_0,\lambda_0}$ in Theorem 2.2.3 can be obtained by solving the one-dimensional optimization problem (2.27). Now, we show in the following Theorem 2.2.4 that the inf-sup value B in (2.19) is indeed equal to the worst-case risk measure in (2.18). The proof of Theorem 2.2.4 is given in Appendix 2.6.

Theorem 2.2.4 *Suppose Assumptions 2.2.1 and 2.2.2 hold, and $\|\gamma\|_\infty < \infty$. If $\varepsilon_1^2 < (\hat{\mu}_1 - \mu_1)^2 + (\hat{\sigma}_1 - \sigma)^2 + 2\sigma\hat{\sigma}_1(1 - c_{1,\beta})$ for all $\sigma \leqslant \sigma_1$ and $\beta \in [\alpha_1, 1]$, then the worst-case risk measure $\sup_{F^{-1}\in\mathcal{Q}_1^d} \rho^{g_1}(F^{-1}(U)\wedge 0)$ in (2.18) and the inf-sup value $\inf_{\beta\in[\alpha_1,1]} \sup_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta, F^{-1})$ in (2.18) are equal. Moreover, the distribution of the quantile function $F^{-1}_{\beta_0,\lambda_0}$ given in Theorem 2.2.3 is the maximizer to the problem (2.3).*

In the following, we use the distortion risk measure TVaR as an example to illustrate results of Theorem 2.2.4. Later, in Section 2.4 we provide more applications of these results by using TVaR.

Example 1 (Worst-case values of TVaR of the limited loss random variable) *In Theorems 2.2.4, suppose $\rho^{g_1} = \mathrm{TVaR}_{p_1}$ for some $0 < p_1 < 1$. That is, for a random variable X,*

$$\mathrm{TVaR}_{p_1}(X) = \frac{1}{1-p_1}\int_{p_1}^1 \mathrm{VaR}_q(X)\,\mathrm{d}q = \frac{1}{1-p_1}\int_{p_1}^1 F^{-1}(q)\,\mathrm{d}q.$$

The distortion function of the TVaR_{p_1} is $g_1(u) = \min\{\frac{u}{1-p_1}, 1\}$. Thus, $\gamma_1(u) = \frac{\partial^- g_1}{\partial x}|_{x=1-u} = \frac{1}{1-p_1}\,\mathrm{I}_{(p_1,1]}(u)$ and $\alpha_1 = p_1$, where α_1 is defined in (2.10). For any $\beta \in (\alpha_1, 1]$, we have $\gamma_{1,\beta}(u) = \frac{1}{1-\alpha_1}\,\mathrm{I}_{(\alpha_1,\beta]}(u) = \frac{\beta-\alpha_1}{1-\alpha_1}\,\tilde{\gamma}_{1,\beta}(u)$ for $0 \leqslant u \leqslant 1$, where $\tilde{\gamma}_{1,\beta} = \frac{1}{\beta-\alpha_1}\,\mathrm{I}_{(\alpha_1,\beta]}$. Note that $\frac{\beta-\alpha_1}{1-\alpha_1} = \|\gamma_{1,\beta}\|_1$ and

$$\tilde{g}_{1,\beta}(t) = 1 - \int_0^{1-t} \tilde{\gamma}_{1,\beta}(u)\,\mathrm{d}u = 1 - \int_0^{1-t} \frac{1}{\beta-\alpha_1}\,\mathbb{I}_{(\alpha_1,\beta]}(u)\,\mathrm{d}u$$

is the distortion function of Range Value-at-Risk (RVaR) *with levels* α_1 *and* β*, denoted by* $\mathrm{RVaR}_{\alpha_1,\beta}$*. In fact, for any random variable* X *with quantile function* F^{-1}*, if* $\beta > \alpha_1$*,* $\mathrm{RVaR}_{\alpha_1,\beta}(X) = \frac{1}{\beta-\alpha_1}\int_{\alpha_1}^{\beta} F^{-1}(q)\,\mathrm{d}q$*. From Proposition 1 of [Bernard et al., 2020b], for any* $\beta \in (\alpha_1, 1]$*,*

$$\sup_{F^{-1}\in\mathcal{Q}_1^d} \mathrm{RVaR}_{\alpha_1,\beta}(F^{-1}(U)) = \frac{1}{\beta-\alpha_1}\int_{\alpha_1}^{\beta} F^{-1}_{\beta,\lambda_\beta}(u)\,\mathrm{d}u, \tag{2.36}$$

where $F^{-1}_{\beta,\lambda_\beta}$ *has the expression given in* (2.25) *with*

$$\ell^{\uparrow}_{\beta,\lambda_\beta}(u) = \begin{cases} \lambda_\beta \hat{F}^{-1}(u), & 0 < u \leqslant \alpha_1, \\ \frac{1}{\beta-\alpha_1} + \lambda_\beta \hat{F}^{-1}(u), & \alpha < u \leqslant w_{\beta,0}, \\ c_\beta, & w_{\beta,0} < u \leqslant w_{\beta,1}, \\ \lambda_\beta \hat{F}^{-1}(u), & w_{\beta,1} < u < 1, \end{cases}$$

and $w_{\beta,0}$*,* $w_{\beta,1}$ *and* c_β *with* $\alpha_1 \leqslant w_{\beta,0} \leqslant \beta \leqslant w_{\beta,1}$*,* $c_\beta < \infty$*, satisfy*

$$\lambda_\beta \hat{F}^{-1}(w_0) = \begin{cases} c_\beta - \frac{1}{\beta-\alpha_1}, & \textit{if } \frac{1}{\beta-\alpha_1} \leqslant c_\beta - \lambda_\beta \hat{F}^{-1}(\alpha_1), \\ \lambda_\beta \hat{F}^{-1}(\alpha_1), & \textit{otherwise}, \end{cases}$$

$$\lambda_\beta \hat{F}^{-1}(w_1) = \begin{cases} \lambda_\beta \hat{F}^{-1}(1), & \textit{if } c_\beta \geqslant \lambda_\beta \hat{F}^{-1}(1), \\ c_\beta, & \textit{otherwise}, \end{cases}$$

$$c_\beta = \frac{1}{w_{\beta,1}-w_{\beta,0}}\Big(\frac{\beta-w_{\beta,0}}{\beta-\alpha_1}\Big) + \frac{\lambda_\beta}{w_{\beta,1}-w_{\beta,0}}\int_{w_{\beta,0}}^{w_{\beta,1}} \hat{F}^{-1}(u)\,\mathrm{d}u.$$

By (2.26)*,* (2.32)*, and* (2.36)*, we see that*

$$L_0(\beta_0, F^{-1}_{\beta_0,\lambda_0} - d) = \min_{\alpha_1\leqslant\beta\leqslant1}\Big\{\frac{1}{1-\alpha_1}\int_{\alpha_1}^{\beta} F^{-1}_{\beta,\lambda_\beta}(u)\,\mathrm{d}u\Big\}, \tag{2.37}$$

which means that β_0 *and* λ_{β_0} *can be obtained by solving the minimization problem in* (2.37)*. To illustrate the application of expression* (2.37) *in determining* β_0 *and* λ_0*, we assume that the sets* $\mathcal{Q}_1$ *and* $\mathcal{S}_1$ *have the following settings:* $\hat{F}(x) = 1 - (\frac{12}{x+12})^4$*,* $x \geqslant 0$*;* $\mu_1 = \hat{\mu}_1 = 4$*;* $\sigma_1 = \hat{\sigma}_1 = 4\sqrt{2}$*;* $\varepsilon_1 = 2$*, i.e., the reference distribution* $\hat{F}$ *is a Pareto distribution. In addition, take* TVaR_{p_1} *with* $p_1 = 0.9$*. In Figure 2.1, we plot the worst-case quantile function* $F^{-1}_{\beta_0,\lambda_0}$ *of the limited loss* $X \wedge d$ *under the settings. When* $d = 15$*, we obtained that* $\beta_0 = 0.9967$*,* $\lambda_0 = 0.7315$*,* $w_{\beta_0,0} = 0.9944$*,* $w_{\beta_0,1} = 0.9981$*, and* $\sup_{F^{-1}\in\mathcal{Q}_1} \mathrm{TVaR}_{0.9}(X^F \wedge 15) =$

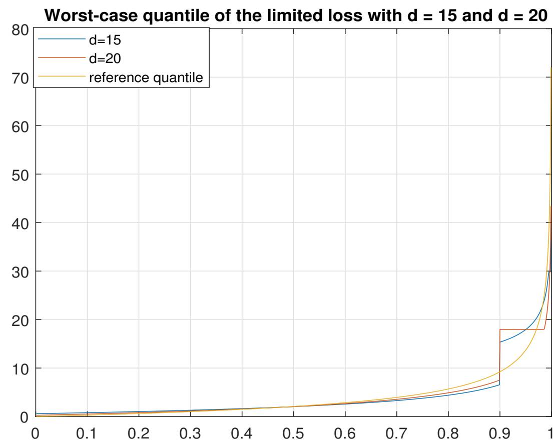


Figure 2.1: Worst-case quantile of the limited loss with $d = 15$ and $d = 20$

$\mathrm{TVaR}_{0.9}(X^{F_{\beta_0,\lambda_0}} \wedge 15) = 15$. *When* $d = 20$, *we obtained that* $\beta_0 = 0.9270$, $\lambda_0 = 0.3406$, $w_{\beta_0,0} = 0.9000$, $w_{\beta_0,1} = 0.9858$, *and* $\sup_{F^{-1}\in\mathcal{Q}_1} \mathrm{TVaR}_{0.9}(X^F \wedge 20) = \mathrm{TVaR}_{0.9}(X^{F_{\beta_0,\lambda_0}} \wedge 20) = 18.2269$. *In addition, for* $d = 15$, *the worst-case quantile function* $F^{-1}_{\beta_0,\lambda_0}(u)$ *of the limited loss* $X \wedge 15$ *jumps upward at* $p_1(= 0.9)$ *and across the value 15 that is the worst-case value of the* TVaR *in this case, which means that the worst-case distribution* F_{β_0,λ_0} *is a mixture distribution. For* $d = 20$, *the worst-case quantile function* $F^{-1}_{\beta_0,\lambda_0}(u)$ *of the limited loss* $X \wedge 20$ *jumps upward at* $p_1(= 0.9)$ *as well and across the value 18.2269 that is the worst-case value of the* TVaR *in this case, and* $F^{-1}_{\beta_0,\lambda_0}(u)$ *is flat from* $p_1(= 0.9)$ *to a value close to 1, which means that the worst-case distribution* F_{β_0,λ_0} *is a mixture distribution and is flat from* $p_1(= 0.9)$ *to a value close to 1. We point out that all the numerical results in this example and in Sections 2.3 and 2.4 are produced by using "fminbnd", a built-in function within Matlab.* □

2.3 Worst-case risk measures of the stop-loss random variable

In this section, we solve problem (2.4) and find the worst-case distribution G^* for the stop-loss random variable $(X - d)_+$ under the distortion risk measure ρ^{g_2}. For simplicity, write $\gamma_2 = \gamma_{g_2}$ which is the weight function in the expression (1.5) for ρ^{g_2}. Similarly to Section

2.2, we define:

$$\mathcal{Q}_2 = \left\{G^{-1} : G \in \mathcal{S}_2\right\} = \left\{G^{-1} :\, d_W(G, \hat{G}) \leqslant \varepsilon_2,\, \mathbb{E}[X^G] = \mu_2,\, \mathrm{var}(X^G) = \sigma_2^2\right\},$$
$$\mathcal{Q}_2^d = \left\{K^{-1} :\, d_W(K^{-1}, \hat{G}^{-1} - d) \leqslant \varepsilon_1,\, \mathbb{E}[K^{-1}(U)] = \mu_2 - d,\, \mathrm{var}(K^{-1}(U)) = \sigma_2^2\right\}, \quad \text{for any } d \geqslant 0.$$

In addition, for the weight function γ_2, define constant α_2 as follows: $\alpha_2 \triangleq 1$ if $\gamma_2(u) > 0$ in a neighbourhood of 1, i.e., $\gamma_2 > 0$ holds on some interval $(\delta, 1)$, $0 < \delta < 1$, and

$$\alpha_2 \triangleq \inf\left\{0 < u < 1 : \int_u^1 \gamma_2(t)\,\mathrm{d}t = 0\right\}, \tag{2.38}$$

otherwise. Note that $0 < \alpha_2 \leqslant 1$ for an absolutely continuous distortion function g_2. For any given $\beta \in [0, \alpha_2]$, define $\gamma_{2,\beta}(u) \triangleq \gamma_2(u)\,\mathrm{I}_{(\beta,1]}(u)$ for $u \in (0, 1)$. For any $\beta \in [0, 1]$ and quantile function G^{-1}, define

$$H_0(\beta,\, G^{-1}) = \int_0^1 \gamma_{2,\beta}(u) G^{-1}(u)\,\mathrm{d}u. \tag{2.39}$$

Lemma 2.3.1 *For a given $d \geqslant 0$, the problem (2.4) has the following equivalent expression:*

$$\sup_{G \in \mathcal{S}_2} \rho^{g_2}\left((X^G - d)_+\right) = \sup_{\beta \in [0, \alpha_2]} \sup_{K^{-1} \in \mathcal{Q}_2^d} H_0(\beta, K^{-1}). \tag{2.40}$$

Proof. By (1.5) and the invariance property of the VaR on the increasing function $(x-d)_+$, we have for any $G \in \mathcal{S}_2$,

$$\rho^{g_2}\left((X^G - d)_+\right) = \int_0^1 \gamma_2(u)(G^{-1}(u) - d)_+\,\mathrm{d}u = \int_{G(d)}^1 \gamma_2(u)(G^{-1}(u) - d)\,\mathrm{d}u. \tag{2.41}$$

Furthermore, take $K^{-1} = G^{-1} - d \in \mathcal{Q}_2^d$ and get $\rho^{g_2}\left((X^G - d)_+\right) = \int_0^1 \gamma_2(u)(K^{-1}(u))_+\,\mathrm{d}u = \rho^{g_2}\left((K^{-1}(U))_+\right)$. Reversely, for any $K^{-1} \in \mathcal{Q}_2^d$, define $G^{-1} = K^{-1} + d$. Clearly, $G \in \mathcal{S}_2$ satisfies the above equation. Hence,

$$\sup_{G \in \mathcal{S}_2} \rho^{g_2}\left((X^G - d)_+\right) = \sup_{K^{-1} \in \mathcal{Q}_2^d} \rho^{g_2}\left((K^{-1}(U))_+\right). \tag{2.42}$$

For any $K^{-1} \in \mathcal{Q}_2^d$, by (2.13), we have $K^{-1}(u) \leqslant 0$ for $0 < u \leqslant K(0)$, and $K^{-1}(u) \geqslant 0$ for $K(0) < u < 1$. Applying arguments similar to those for (2.17), we further have

$$\rho^{g_2}\left((K^{-1}(U))_+\right) = \int_0^1 \gamma_2(u)(K^{-1}(u))_+\,\mathrm{d}u = \max_{\beta \in [0, \alpha_2]} \int_\beta^1 \gamma_2(u) K^{-1}(u)\,\mathrm{d}u = \max_{\beta \in [0, \alpha_2]} H_0(\beta, K^{-1}).$$

By taking the supremum over the uncertainty set $\mathcal{Q}_2^d$ on both sides of the above equation, one can re-write problem (2.42) as

$$\sup_{K^{-1}\in\mathcal{Q}_2^d} \rho^{g_2}\left((K^{-1}(U))_+\right) = \sup_{K^{-1}\in\mathcal{Q}_2^d} \max_{\beta\in[0,\alpha_2]} H_0(\beta, K^{-1}). \tag{2.43}$$

The equation (2.40) follows consequently. ■

Lemma 2.3.1 transfers the problem (2.4) to the double supremum problem (2.40), which can be solved by using the same way as the inf-sup problem (2.19). In doing so, for $\lambda \geqslant 0$ and $\beta \in [0, \alpha_2]$, let $k^{\uparrow}_{\beta,\lambda}$ be the isotonic projection of $\gamma_{2,\beta} + \lambda\hat{G}^{-1}$ onto $\mathcal{K}$, i.e.,

$$k^{\uparrow}_{\beta,\lambda} = \arg\min_{\ell\in\mathcal{K}} ||k - \gamma_{2,\beta} - \lambda\hat{G}^{-1}||_2. \tag{2.44}$$

Assumption 2.3.1 *Assume that the isotonic projection $k^{\uparrow}_{\beta,\lambda}$ defined in (2.44) is not constant for any $\lambda > 0$ and any $\beta \in [0, \alpha_2)$.*

For any $\beta \in [0, \alpha_2]$, denote $c_{2,\beta}$ by $c_{2,\beta} \triangleq \mathrm{corr}(\hat{G}^{-1}(U), k^{\uparrow}_{\beta,0}(U))$. In addition, for any $\beta \in [0, \alpha_2]$ and $\lambda \geqslant 0$, let $m_{\beta,\lambda} = \mathbb{E}\big[k^{\uparrow}_{\beta,\lambda}(U)\big]$ and $v_{\beta,\lambda} = \sqrt{\mathrm{var}(k^{\uparrow}_{\beta,\lambda}(U))}$. We define the quantile function $G^{-1}_{\beta,\lambda}(u)$, $0 < u < 1$, as

$$G^{-1}_{\beta,\lambda}(u) = \begin{cases} \mu_2 + \sigma_2\Big(\frac{k^{\uparrow}_{\beta,\lambda}(u) - m_{\beta,\lambda}}{v_{\beta,\lambda}}\Big), & \text{if } v_{\beta,\lambda} > 0 \\ \mu_2, & \text{if } v_{\beta,\lambda} = 0. \end{cases} \tag{2.45}$$

Theorem 2.3.2 *Suppose Assumption 2.3.1 holds. Then*

$$\sup_{G\in\mathcal{S}_2} \rho^{g_2}\left((X^G - d)_+\right) = \sup_{\beta\in[0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d), \tag{2.46}$$

where $G^{-1}_{\beta,\lambda_\beta}$ is defined in (2.45). Furthermore, the following statements hold:

(i) If $(\hat{\mu}_2 - \mu_2)^2 + (\hat{\sigma}_2 - \sigma_2)^2 < \varepsilon_2^2 < (\hat{\mu}_2 - \mu_2)^2 + (\hat{\sigma}_2 - \sigma_2)^2 + 2\sigma_2\hat{\sigma}_2(1 - c_{2,\beta})$, then λ_β is the unique solution to the equation $d_W(\hat{G}^{-1}, G^{-1}_{\beta,\lambda}) = \varepsilon_2$ for $\lambda \in (0, \infty)$ and $G^{-1}_{\beta,\lambda_\beta} \in \mathcal{Q}_2$.

(ii) If $\varepsilon_2^2 \geqslant (\hat{\mu}_2 - \mu_2)^2 + (\hat{\sigma}_2 - \sigma_2)^2 + 2\sigma_2\hat{\sigma}_2(1 - c_{2,\beta})$, then $\lambda_\beta = 0$. If $k^{\uparrow}_{\beta,0}$ is non-constant, then $G^{-1}_{\beta,0} \in \mathcal{Q}_2$. If $k^{\uparrow}_{\beta,0}$ is a constant, then $G^{-1}_{\beta,0} = \mu_2 \notin \mathcal{Q}_2$.

Proof. Note that for any $K^{-1} \in \mathcal{Q}_2^d$, define $G^{-1} = K^{-1} + d$. Thus, $G^{-1} \in \mathcal{Q}_2$ and $H_0(\beta, K^{-1}) = \int_0^1 \gamma_{2,\beta}(u) K^{-1}(u)\,\mathrm{d}u = \int_0^1 \gamma_{2,\beta}(u)(G^{-1}(u) - d)\,\mathrm{d}u$ in (2.40) has the same structure as $L_0(\beta, F^{-1}) = \int_0^1 \gamma_{1,\beta}(u)(F^{-1}(u) - d)\,\mathrm{d}u$ in (2.26). By changing F^{-1} and $\gamma_{1,\beta}$ to G^{-1} and $\gamma_{2,\beta}$, respectively, in the proof of Theorem 2.3.2 for (2.31), we easily see that for any $K^{-1} \in \mathcal{Q}_2^d$,

$$H_0(\beta, K^{-1}) = \|\gamma_{2,\beta}\|_1 \; \rho^{\tilde{g}_{2,\beta}}(X^K), \tag{2.47}$$

where

$$\tilde{g}_{2,\beta}(t) = 1 - \int_0^{1-t} \frac{\gamma_{2,\beta}(u)}{\|\gamma_{2,\beta}\|_1}\,\mathrm{d}u, \qquad \text{for } t \in [0,1]. \tag{2.48}$$

is a distortion function and $\|\gamma_{2,\beta}\|_1 = \int_0^1 \gamma_{2,\beta}(u)\,\mathrm{d}u > 0$ for any $\beta \in [0, \alpha_2)$ and $\tilde{g}_{2,\alpha_2}(t) = 0$ for any $t \in [0,1]$. Therefore, by (2.40) and (2.47), we have

$$\sup_{G \in \mathcal{S}_2} \rho^{g_2}\left((X^G - d)_+\right) = \sup_{\beta \in [0,\alpha_2]} \sup_{K^{-1} \in \mathcal{Q}_2^d} H_0(\beta, K^{-1}) = \sup_{\beta \in [0,\alpha_2]} \left\{ \|\gamma_{2,\beta}\|_1 \sup_{K^{-1} \in \mathcal{Q}_2^d} \rho^{\tilde{g}_{2,\beta}}(X^K) \right\}. \tag{2.49}$$

Then, by applying Lemma 2.6.1 to the problem $\sup_{K^{-1} \in \mathcal{Q}_2^d} \rho^{\tilde{g}_{2,\beta}}(X^K)$ and using the arguments similar to those after (2.32) in Theorem 2.2.3, we see that $\|\gamma_{2,\beta}\|_1 \sup_{K^{-1} \in \mathcal{Q}_2^d} \rho^{\tilde{g}_{2,\beta}}(X^K) = H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d)$ for any $\beta \in [0, \alpha_2]$ and statements (i) and (ii) in Theorem 2.3.2 hold. It completes the proof of Theorem 2.3.2. ■

Remark 2.3.1 *From the proof for the lower-semicontinuity of $L_0(\beta, F^{-1}_{\beta,\lambda_\beta} - d)$ in the proof of Theorem 2.2.3, we see that $H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d)$ is lower-semicontinuous in $\beta \in [0, \alpha_2]$. However, the lower-semicontinuity of $H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d)$ is not sufficient to guarantee the existence of a maximizer to $\sup_{\beta \in [0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d)$ in (2.46). In fact, the optimization problem in (2.26) for Theorem 2.2.3 is an inf-sup problem, while the optimization problem (2.49) for Theorem 2.3.2 is a sup-sup problem. However, if $H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d)$ is continuous in $\beta \in [0, \alpha_2]$, then there exists $\beta^* \in [0, \alpha_2]$ such that $H_0(\beta^*, G^{-1}_{\beta^*,\lambda_{\beta^*}} - d) = \sup_{G \in \mathcal{S}_2} \rho^{g_2}\left((X^G - d)_+\right)$. In addition, like Theorems 2.2.3, with expression (2.46) in Theorem 2.3.2, one reduces the infinite-dimensional optimization problem (2.4) to the feasible one-dimensional optimization $\sup_{\beta \in [0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d)$.* □

Proposition 2.3.3 *Suppose Assumption 2.3.1 holds. Then, for a given $d \geqslant 0$, the following two statements hold:*

(i) If the problem (2.4) has a maximizer $G^* \in \mathcal{S}_2$, then $\beta^* \wedge \alpha_2$ is a maximizer to $\sup\limits_{\beta\in[0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d)$ in (2.46), and $G^{-1}_{\beta^*,\lambda_{\beta^*}} = G^{*-1}$ a.e., where $\beta^* \triangleq G^*(d)$ and λ_{β^*} solves $d_W(\hat{G}^{-1}, G^{-1}_{\beta^*,\lambda_{\beta^*}}) = \varepsilon_2$.

(ii) If the problem $\sup_{\beta\in[0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d)$ in (2.46) has a maximizer $\beta^* = \beta^*(d) \in [0, \alpha_2]$ and $G_{\beta^*,\lambda_{\beta^*}}(d-) \leqslant \beta^* \leqslant G_{\beta^*,\lambda_{\beta^*}}(d)$, then $G_{\beta^*,\lambda_{\beta^*}}$ is a maximizer to the problem (2.4).

The proof of Proposition 2.3.3 is given in Appendix 2.6. This proposition shows that a quantile function with form (2.45) is necessary for the corresponding distribution of the quantile function to be a maximizer of problem (2.4). We use the following example to illustrate the applications of Theorem 2.3.2 and Proposition 2.3.3.

Example 2 (Worst case values of TVaR of the stop-loss random variable) *To illustrate the applications of Theorem 2.3.2 and Proposition 2.3.3, we assume that the reference distribution $\hat{G}$ in $\mathcal{Q}_2^d$ and $\mathcal{S}_2$ is a Pareto distribution and the sets $\mathcal{Q}_2^d$ and $\mathcal{S}_2$ have the following settings: $\hat{G}(x) = 1 - \left(\frac{8}{x+8}\right)^3$, $x \geqslant 0$; $\mu_2 = \hat{\mu}_2 = 4$; $\sigma_2 = \hat{\sigma}_2 = 4\sqrt{3}$; $\varepsilon_2 = 2$. In addition, suppose $\rho^{g_2} = \mathrm{TVaR}_{p_2}$ in Theorem 2.3.2 for some $0 < p_2 < 1$. Then, $\gamma_2(u) = \frac{1}{1-p_2}\,\mathrm{I}_{(p_2,1]}(u)$ with $\alpha_2 = 1$ and $\gamma_{2,\beta}(u) = \frac{1}{1-p_2}\,\mathrm{I}_{(p_2\vee\beta,1]}(u)$ for $0 \leqslant u \leqslant 1$ and any $\beta \in [0,1]$. Under the settings, we obtained β^* and λ_{β^*} satisfying $\sup_{\beta\in[0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d) = H_0(\beta^*, G^{-1}_{\beta^*,\lambda_{\beta^*}} - d)$ by solving the one-dimensional optimization problem $\sup_{\beta\in[0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d)$ for $d = 10$ and $d = 20$.*

In Figure 2.2, we plot the quantile function $G^{-1}_{\beta^*,\lambda_{\beta^*}}$ under the settings. When $d = 10$, we obtained that $\beta^* = 0.9$, $\lambda_{\beta^*} = 0.7942$, and $\sup_{G\in\mathcal{S}_2} \mathrm{TVaR}_{0.9}\big((X^G - 10)_+\big) = 1.1702$. When $d = 20$, we obtained that $\beta^* = 0.9669$, $\lambda_{\beta^*} = 0.4621$, and $\sup_{G\in\mathcal{S}_2} \mathrm{TVaR}_{0.9}\big((X^G - 20)_+\big) = 0.5357$. We observed from the numerical results that in both the cases $d = 10$ and $d = 20$, $\sup_{G\in\mathcal{S}_2} \mathrm{TVaR}_{0.9}\big((X^G - d)_+\big) = \mathrm{TVaR}_{0.9}\big((X^{G_{\beta^*,\lambda_{\beta^*}}} - d)_+\big)$, which means that the supremum in (2.46) is attained at the distribution $G_{\beta^*,\lambda_{\beta^*}}$. In addition, the worst-case quantile functions $G^{-1}_{\beta^*,\lambda_{\beta^*}}(u)$ of the stop-loss $(X-10)_+$ and $(X-20)_+$ jump upward at $u = 0.9$ and $u = 0.97$, respectively, and across the their own stop-loss retentions $d = 10$ and $d = 20$, which means that the worst-case distributions $G_{\beta^*,\lambda_{\beta^*}}$ are mixture distributions in both the cases. Also, we found from the numerical results that $G_{\beta^*,\lambda_{\beta^*}}(d) = \beta^*$ in both the cases. These findings are consistent with the statements of Proposition 2.3.3. $\square$

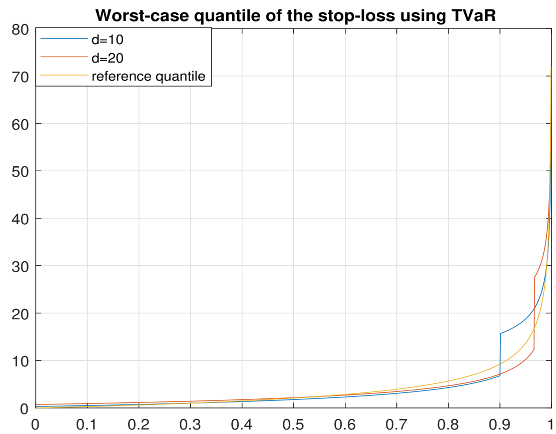


Figure 2.2: Worst-case quantile of the stop-loss using TVaR

2.4 Worst-case TVaRs of the stop-loss and limited loss random variables with applications in robust stop-loss reinsurances

In this section, we illustrate the applications of Theorems 2.2.3 and 2.3.2 by using the distortion risk measure of TVaR and special cases of the uncertainty sets $\mathcal{S}_1 = \mathcal{S}(\mu_1, \sigma_1, \hat{F}; \varepsilon_1)$ and $\mathcal{S}_2 = \mathcal{S}(\mu_2, \sigma_2, \hat{G}; \varepsilon_2)$. We adopted simplified notations for $\mathcal{S}_1 = \mathcal{S}(\mu_1, \sigma_1, \hat{F}; \varepsilon_1)$ and $\mathcal{S}_2 = \mathcal{S}(\mu_2, \sigma_2, \hat{G}; \varepsilon_2)$ for special values of ε_1 and ε_2 as follows:

$$\mathcal{S}_1^\infty = \mathcal{S}(\mu_1, \sigma_1, \hat{F}; \infty), \quad \mathcal{S}_2^\infty = \mathcal{S}(\mu_2, \sigma_2, \hat{G}; \infty), \quad \mathcal{S}_1^0 = \mathcal{S}(\hat{\mu_1}, \hat{\sigma}_1, \hat{F}; 0), \quad \mathcal{S}_2^0 = \mathcal{S}(\hat{\mu}_2, \hat{\sigma}_2, \hat{G}; 0).$$

In fact, $\mathcal{S}_i^\infty$ is the set of all the distributions with mean μ_i and variance σ_i for $i = 1, 2$. In addition, $\mathcal{S}_i^0$ is reduced to a set of a single distribution that is the reference distribution $\hat{F}$ for $i = 1$ or $\hat{G}$ for $i = 2$, namely, $\mathcal{S}_1^0 = \{\hat{F}\}$ and $\mathcal{S}_2^0 = \{\hat{G}\}$. In this cases, both of the insurer and reinsurer use deterministic distribution functions to evaluate the underlying insurance loss X. However, $\hat{F}$ and $\hat{G}$ may be different. Such a model setting represents that the insurer and reinsurer may have different beliefs in the distribution of X.

We first give the worst-case TVaRs of the limited loss and stop-loss random variables and then find optimal retentions that minimize the worst-case TVaR of the insurer's risk exposure in a stop-loss reinsurance when both of the insurer and reinsurer face an uncertain distribution on the underlying insurance loss X. Mathematically, we discuss the following three optimization problems:

$$\sup_{F\in\mathcal{S}_1} \mathrm{TVaR}_{p_1}(X^F \wedge d), \tag{2.50}$$

$$\sup_{G\in\mathcal{S}_2} \mathrm{TVaR}_{p_2}\left((X^G - d)_+\right), \tag{2.51}$$

$$\min_{d\geqslant 0} \sup_{F\in\mathcal{S}_1} \mathrm{TVaR}_{p_1}\left(X^F \wedge d + \pi(d)\right), \tag{2.52}$$

where $0 < p_1 < 1$, $0 < p_2 < 1$, and $\pi(d)$ is a deterministic reinsurance premium under a given retention $d \geqslant 0$. In problem (2.52), we assume that the random variable X is the underlying loss faced by the insurer. A positive value (resp. a negative value) of X represents a loss (resp. a profit). The idea of model (2.52) is to minimize the worst-case TVaR of the insurer's risk exposure in a stop-loss reinsurance and to find optimal retentions d^* under distribution uncertainty.

In a stop-loss reinsurance agreement, for a given retention $d \geqslant 0$, the reinsurer needs to offer the insurer a deterministic premium $\pi(d)$ even if the distribution of X is uncertain. Normally, the calculation of $\pi(d)$ depends on the distribution of X and d. For example, $\pi(d) = (1+\theta)\mathbb{E}(X-d)_+$ is the expected value principle, $\pi(d) = (1+\theta)\rho^g(X-d)_+$ is the distortion principle, where $\theta \geqslant 0$ is a loading factor, ρ^g is a distortion risk measure, and $\mathbb{E}(X-d)_+$ is called the stop-loss premium. However, when the distribution of the underlying insurance loss X is uncertain, these principles will not yield a deterministic premium. In this section, we propose two methods for the reinsurer to determine the reinsurance premium $\pi(d)$ under distribution uncertainty. One method is to assume

$$\pi(d) = \sup_{G\in\mathcal{S}_2} (1+\theta)\,\mathbb{E}\left[(X^G - d)_+\right], \tag{2.53}$$

which is the highest premium charged by the reinsurer or the worst-case premium for the insurer under distribution uncertainty and the expected value principle. However, in practice, it could be too conservative for the reinsurer to price the premium by using the worst-case premium for the insurer. This may result in an unacceptably high premium for the insurer and consequently make the reinsurance contract less competitive in reinsurance market. As a result, the reinsurer may determine the premium based on the reference/pricing distribution $\hat{G}$ by using the expected value principle and calculate the premium by

$$\pi(d) = (1+\theta)\mathbb{E}\big[(X^{\hat{G}} - d)_+\big] = \sup_{G\in\mathcal{S}_2^0} (1+\theta)\mathbb{E}\big[(X^G - d)_+\big]. \tag{2.54}$$

In this premium, the reinsurer uses the reference distribution $\hat{G}$ to evaluate the underlying insurance loss X.

Stop-loss reinsurances are not only optimal forms among many optimal reinsurance design problems, but also one of the most popular forms of insurance/reinsurance used in practice. In addition, that the settings of the optimal stop-loss reinsurance problem (2.52) are different from those used in [Hu et al., 2015] and [Liu and Mao, 2021]. In fact, the uncertainty sets used in problem (2.52) are different from those used in [Hu et al., 2015] and [Liu and Mao, 2021]. More importantly, the reinsurance premium $\pi(d)$ in problem (2.52) is deterministic while the reinsurance premium in [Liu and Mao, 2021] is $(1+\theta)\mathbb{E}\left[(X-d)_+\right]$, which involves model uncertainty of X and affects the process of looking for the overall worst-case distribution.

2.4.1 Explicit and closed-form solutions under the uncertainty sets $\mathcal{S}_1^\infty$ and $\mathcal{S}_2^\infty$

In this subsection, we assume that the insurer and reinsurer use uncertainty sets $\mathcal{S}_1^\infty$ and $\mathcal{S}_2^\infty$, respectively, to represent the sets of possible distributions for the underlying insurance loss X. Under the uncertainty sets $\mathcal{S}_1^\infty$ and $\mathcal{S}_2^\infty$, we obtain the explicit and closed-form expressions for the solutions to problems (2.50)-(2.52) in the following Theorems 2.4.1 and 2.4.2. In this section, for any $p_1, p_2 \in (0,1)$, $i=1,2$, we define

$$d_i = \mu_i + \sigma_i \sqrt{\frac{p_i}{1-p_i}} \quad \text{and} \quad d_3 = \mu_2 - \sigma_2 \frac{1-2p_2}{2\sqrt{p_2(1-p_2)}}. \tag{2.55}$$

Theorem 2.4.1 *For any $d \geqslant 0$ and $p_1, p_2 \in (0,1)$, one has*

$$\sup_{F\in\mathcal{S}_1^\infty} \mathrm{TVaR}_{p_1}(X^F \wedge d) = d_1 \wedge d, \tag{2.56}$$

and

$$\sup_{G\in\mathcal{S}_2^\infty} \mathrm{TVaR}_{p_2}\left((X^G - d)_+\right) = \begin{cases} d_2 - d, & d \leqslant d_3, \\ \frac{1}{2(1-p_2)}\left(\mu_2 - d + \sqrt{(\mu_2-d)^2+\sigma_2^2}\right), & d > d_3. \end{cases} \tag{2.57}$$

Proof. We first prove (2.56). To do so, let $\varepsilon_1 \to \infty$ or $\varepsilon_1 = \infty$ in Theorem 2.2.3. Under the notations used in Theorem 2.2.3 and its proof, in this case $\varepsilon_1 = \infty$, we see that condition

(2.29) holds and thus $\lambda_0 = 0$ according to Theorem 2.2.3 (ii). In addition, in this case, it is easy to verify that for any $\beta \in [\alpha_1, 1] = [p_1, 1]$, where α_1 is defined in (2.10), we have

$$F_{\beta,0}^{-1}(u) = \begin{cases} \mu_1 - \sigma_1\sqrt{\frac{1-p_1}{p_1}}, & u \leqslant p_1 \\ \mu_1 + \sigma_1\sqrt{\frac{p_1}{1-p_1}}, & u > p_1, \end{cases} \tag{2.58}$$

and

$$\|\gamma_{1,\beta}\|_1 \ \rho^{\tilde{g}_{1,\beta}}(F_{\beta,0}^{-1}(U) - d) = \frac{\beta - p_1}{1 - p_1}\Big(\mu_1 + \sigma_1\sqrt{\frac{p_1}{1-p_1}} - d\Big) = \frac{\beta - p_1}{1 - p_1}\,(d_1 - d) \triangleq f(\beta). \tag{2.59}$$

Thus, by (2.5), (2.35), and Theorem 2.2.4, we have $\sup_{F\in\mathcal{S}_1^\infty} \mathrm{TVaR}_{p_1}(X^F \wedge d) = d + \min_{\beta\in[p_1,1]} f(\beta)$. It is easy to see that $\beta_0 = \arg\min_{\beta\in[p_1,1]} f(\beta) = p_1$ if $d_1 - d > 0$ and 1 if $d_1 - d \leqslant 0$. Hence, (2.56) holds.

Next, we use Theorem 2.3.2 to prove (2.57). Note that for TVaR_{p_2}, $p_2 = 1$, where p_2 is defined in (2.38). In this case $\varepsilon_2 = \infty$, the condition in Theorem 2.3.2 (ii) holds, and it is easy to verify that

$$G_{\beta,0}^{-1}(u) = \begin{cases} \mu_2 - \sigma_2\sqrt{\frac{1-(p_2\vee\beta)}{p_2\vee\beta}}, & u \leqslant p_2 \vee \beta, \\ \mu_2 + \sigma_2\sqrt{\frac{p_2\vee\beta}{1-(p_2\vee\beta)}}, & u > p_2 \vee \beta, \end{cases} \tag{2.60}$$

and

$$H_0(\beta, G_{\beta,0}^{-1} - d) = \int_\beta^1 \frac{1}{1-p_2}\,\mathrm{I}_{[p_2,1]}(u)(G_{\beta,0}^{-1}(u) - d)\,\mathrm{d}u \triangleq h(\beta),$$

where

$$h(\beta) = \begin{cases} \mu_2 + \sigma_2\sqrt{\frac{p_2}{1-p_2}} - d = d_2 - d, & \beta \leqslant p_2, \\ \frac{1}{1-p_2}\big(\sigma_2\sqrt{\beta(1-\beta)} + (\mu_2 - d)(1-\beta)\big), & \beta > p_2. \end{cases} \tag{2.61}$$

By Theorem 2.3.2, $\sup_{G\in\mathcal{S}_2^\infty} \mathrm{TVaR}_{p_2}((X^G - d)_+) = \sup_{0\leqslant\beta\leqslant1} h(\beta)$. It is easy to see that $h(\beta)$ is continuous on $[0,1]$. Hence, there exists $\beta^* \in [0,1]$ such that $\sup_{0\leqslant\beta\leqslant1} h(\beta) = h(\beta^*)$. If fact, it is easy to verify that if $d \leqslant d_3$, then for any $\beta^* \in [0, p_2]$, $\sup_{0\leqslant\beta\leqslant1} h(\beta) = h(\beta^*) = d_2 - d$; and if $d > d_3$, then $\beta^* = \frac{1}{2}(1 - \frac{\mu_2 - d}{\sqrt{(\mu_2-d)^2+\sigma_2^2}}) \in [0,1]$ and $\sup_{0\leqslant\beta\leqslant1} h(\beta) = h(\beta^*) = \frac{1}{2(1-p_2)}\big(\mu_2 - d + \sqrt{(\mu_2 - d)^2 + \sigma_2^2}\big)$. Therefore, (2.57) holds. ∎

Remark 2.4.1 *It is easy to versify that $\beta^*(= \beta^*(d))$ in the proof of Theorem 2.4.1 satisfies $G_{\beta^*,0}(d^-) < \beta^* \leqslant G_{\beta^*,0}(d)$, which means that the conditions in Proposition 2.3.3 (ii) holds.*

Hence, by Proposition 2.3.3(ii), the corresponding distribution $G_{\beta^*,0}$ of the quantile function $G_{\beta,0}^{-1}$ given in (2.60) is the worst-case distribution to the problem $\sup_{G\in\mathcal{S}_2^\infty} \mathrm{TVaR}_{p_2}\big((X^G-d)_+\big)$. In addition, by Theorem 2.2.4, the corresponding distribution $F_{\beta_0,0}$ of the quantile function $F_{\beta,0}^{-1}$ given in given in (2.58) is the worst-case distribution to the problem $\sup_{F\in\mathcal{S}_1^\infty} \mathrm{TVaR}_{p_1}(X^F\wedge d)$.

Moreover, note that $\mathrm{TVaR}_0=\mathbb{E}$ and $\lim_{p_2\downarrow 0} d_3=-\infty$. Thus, with $p_2=0$ in (2.57), the second case in (2.57) always applies because $d\geqslant 0$. Therefore, by (2.57), we obtain for any $d\geqslant 0$,

$$\sup_{G\in\mathcal{S}_2^\infty} \mathbb{E}[(X^G-d)_+]=\frac{1}{2}\Big(\mu_2-d+\sqrt{(\mu_2-d)^2+\sigma_2^2}\,\Big), \tag{2.62}$$

which recovers Corollary 1.1 of [Jagannathan, 1977]. Thus, under the uncertainty set $\mathcal{S}_2^\infty$, the premium given in (2.53) is equal to

$$\pi(d)=\frac{1+\theta}{2}\Big(\mu_2-d+\sqrt{(\mu_2-d)^2+\sigma_2^2}\,\Big). \tag{2.63}$$

□

Now, by using Theorem 2.4.1, we obtain the explicit and closed-form solutions to problem (2.52) in the following theorem.

Theorem 2.4.2 *Assume that the reinsurance premium in problem (2.52) is calculated by (2.53). Let d^* be the optimal solution to problem (2.52) with $\mathcal{S}_1=\mathcal{S}_1^\infty$ and $\mathcal{S}_2=\mathcal{S}_2^\infty$.*

(i) Assume $0<\mu_2<d_1$. Then

$$d^*=\begin{cases}0, & \text{if } 0\leqslant\theta<1,\ \mu_2+\frac{\theta-1}{2\sqrt{\theta}}\sigma_2\leqslant 0,\ \frac{1+\theta}{2}\big(\mu_2+\sqrt{\mu_2^2+\sigma_2^2}\,\big)<d_1,\\ \infty, & \text{if } 0\leqslant\theta<1,\ \mu_2+\frac{\theta-1}{2\sqrt{\theta}}\sigma_2\leqslant 0,\ \frac{1+\theta}{2}\big(\mu_2+\sqrt{\mu_2^2+\sigma_2^2}\,\big)\geqslant d_1,\\ \mu_2+\frac{\theta-1}{2\sqrt{\theta}}\sigma_2, & \text{if } \theta>0,\ 0<\mu_2+\frac{\theta-1}{2\sqrt{\theta}}\sigma_2<d_1,\ \mu_2+\sqrt{\theta}\,\sigma_2<d_1,\\ \infty, & \text{if } \theta>0,\ 0<\mu_2+\frac{\theta-1}{2\sqrt{\theta}}\sigma_2<d_1,\ \mu_2+\sqrt{\theta}\,\sigma_2\geqslant d_1.\\ \infty, & \text{if if } \theta>1,\ \mu_2+\frac{\theta-1}{2\sqrt{\theta}}\sigma_2\geqslant d_1,\end{cases} \tag{2.64}$$

(ii) Assume $\mu_2\geqslant d_1$. Then, $d^*=\infty$.

Proof. By (2.63) and (2.56), we see that problem (2.52) is reduced to problem $\min_{d\geqslant 0} f(d)$, where

$$f(d) = d_1 \wedge d + \frac{(1+\theta)}{2}\Big(\mu_2 - d + \sqrt{(d-\mu_2)^2 + \sigma_2^2}\Big)$$

Obviously, f is positive and continuous on $[0,\infty)$ and differentiable on $(0,d_1)\cup(d_1,\infty)$ with $f(\infty) \triangleq \lim_{d\to\infty} f(d) = d_1$ and

$$f'(d) = \begin{cases} 1 + \frac{1+\theta}{2}\Big(\frac{d-\mu_2}{\sqrt{(d-\mu_2)^2+\sigma_2^2}} - 1\Big), & 0 < d < d_1, \\ \frac{1+\theta}{2}\Big(\frac{d-\mu_2}{\sqrt{(d-\mu_2)^2+\sigma_2^2}} - 1\Big), & d > d_1. \end{cases}$$

Note that $f'(d) < 0$ on (d_1,∞). Hence, f strictly decreases on $(d_1,\,\infty)$ and $\min_{d\in(d_1,\infty]} f(d) = \lim_{d\to\infty} f(d) = f(\infty) = d_1$. In addition, since f is continuous on the closed interval $[0,d_1]$, there exists $d_1^* \in [0,d_1]$ such that $\min_{d\in[0,d_1]} f(d) = f(d_1^*)$. Therefore, there exists $d^* \in [0,\infty]$ satisfying $\min_{d\in[0,\infty]} f(d) = f(d^*) = \min\{f(d_1^*), f(\infty)\}$. Hence, $d^* = d_1^*$ if $f(d_1^*) < f(\infty) = d_1$ and $d^* = \infty$ if $f(d_1^*) \geqslant f(\infty) = d_1$. Next, we obtain the explicit expression for d_1^* by discussing the following cases:

(i) Assume $0 < \mu_2 < d_1$. Then for $d \in (0,\mu_2)$, we see that $d - \mu_2 < 0$ and that $f'(d) \leqslant 0$ on $(0,\mu_2)$ if and only if

$$\Big(1 + \frac{\sigma_2^2}{(d-\mu_2)^2}\Big)^{-\frac{1}{2}} \geqslant \frac{1-\theta}{1+\theta}. \tag{2.65}$$

On the other hand, if $0 < \mu_2 < d_1$, for $d \in (\mu_2, d_1)$, we have $d - \mu_2 > 0$ and that $f'(d) \leqslant 0$ on (μ_2, d_1) if and only if

$$\Big(1 + \frac{\sigma_2^2}{(d-\mu_2)^2}\Big)^{-\frac{1}{2}} \leqslant \frac{\theta-1}{1+\theta}. \tag{2.66}$$

Thus, by checking conditions (2.65) and (2.66), we have the following conclusions:

(a) If $\theta = 0$, then $f'(d) > 0$ on $(0,\mu_2)$ and $f'(d) > 0$ on (μ_2, d_1). In this case, $d_1^* = 0$.

(b) If $0 < \theta < 1$, then $f'(d) > 0$ on $(\mu_2,\, d_1)$. On the other hand, if $0 < \theta < 1$, by (2.65), it is easy to verify that $f'(d) \leqslant 0$ on $(0,\,\mu_2)$ if and only if $d \leqslant d_0$, where, for any $\theta \geqslant 0$, $d_0 = d_0(\theta)$ is given by

$$d_0 = d_0(\theta) = \mu_2 + \frac{\theta - 1}{2\sqrt{\theta}}\,\sigma_2, \tag{2.67}$$

where, by convention, if $\theta = 0$, $d_0 = d_0(0) = -\infty$.

(b1) If $d_0 \leqslant 0$, then $f'(d) > 0$ on $(0,\, \mu_2)$. In this case, $d_1^* = 0$.

(b2) If $d_0 > 0$, which is equivalent to $0 < d_0 < \mu_2$ as $d_0 < \mu_2$ if $0 < \theta < 1$, then $f'(d) \leqslant 0$ on $(0,\, d_0)$ and $f'(d) > 0$ on (d_0, μ_2). In this case, $d_1^* = d_0 = d_0(\theta)$ and

$$f(d_0) = \mu_2 + \sqrt{\theta}\,\sigma_2. \tag{2.68}$$

(c) If $\theta = 1$, then $f'(d) \leqslant 0$ on $(0,\, \mu_2)$ and $f'(d) > 0$ on (μ_2, d_1). In this case, $d_1^* = \mu_2 = d_0(1)$ and $f(\mu_2) = \mu_2 + \sigma_2 = f(d_0(1))$.

(d) If $\theta > 1$, then $f'(d) \leqslant 0$ on $(0, \mu_2)$. On the other hand, if $\theta > 1$, by (2.66), we see that $f'(d) \leqslant 0$ on (μ_2, d_1) if and only if $d \leqslant d_0$.

(d1) If $d_0 \in (\mu_2, d_1)$, then $f'(d) \leqslant 0$ on $(\mu_2,\, d_0)$ and $f'(d) > 0$ on (d_0, d_1). In this case, $d_1^* = d_0$ and $f(d_0)$ is given in (2.68).

(d2) If $d_0 \geqslant d_1$, then $f'(d) \leqslant 0$ on $(\mu_2,\, d_1)$. In this case, $d_1^* = d_1$.

Note that $f(0) = \frac{1+\theta}{2}\big(\mu_2 + \sqrt{\mu_2^2 + \sigma_2^2}\,\big)$, $f(\infty) = d_1$, and $f(d_0)$ is given in (2.68) and that if $d_1^* = d_1$, then $d^* = \infty$. Thus, by noticing $d_0(0) \leqslant 0$ and combining cases (a) and (b1), we obtain the first two cases in (2.64). By combining cases (b2), (c), and (d1), we obtain the third and fourth cases in (2.64). The last case (the fifth case) in (2.64) corresponds to (d2).

(ii) Assume $\mu_2 \geqslant d_1$. Then for $d \in (0, d_1)$, we see that $d - \mu_2 < 0$ and that $f'(d) \leqslant 0$ on $(0, d_1)$ if and only if condition (2.65) holds. Thus, by checking condition (2.65), we have the following conclusions:

(a) If $\theta = 0$, then $f'(d) > 0$ on $(0,\, d_1)$. In this case, $d_1^* = 0$. However, in this case, $f(0) = \frac{1}{2}\big(\mu_2 + \sqrt{\mu_2^2 + \sigma_2^2}\,\big) > \mu_2 \geqslant f(\infty) = d_1$ as $\mu_2 \geqslant d_1$, and thus $d^* = \infty$.

(b) If $0 < \theta < 1$, by (2.65), it is easy to verify that $f'(d) \leqslant 0$ on $(0, d_1)$ if and only if $d \leqslant d_0$.

(b1) If $d_0 \leqslant 0$, then $f'(d) > 0$ on $(0,\, d_1)$. In this case, $d_1^* = 0$ and thus $d^* = \infty$ as $f(0) = \frac{1+\theta}{2}\big(\mu_2 + \sqrt{\mu_2^2 + \sigma_2^2}\,\big) > \mu_2 \geqslant f(\infty) = d_1$.

(b2) If $0 < d_0 < d_1$, then $f'(d) \leqslant 0$ on $(0, d_0)$ and $f'(d) > 0$ on (d_0, d_1). In this case, $d_1^* = d_0$ and $f(d_0)$ is given in (2.68). However, in this case, $f(d_1^*) = f(d_0) = \mu_2 + \sqrt{\theta}\,\sigma_2 > f(\infty) = d_1$ as $\mu_2 \geqslant d_1$, and thus $d^* = \infty$. In addition, if $d_0 \geqslant d_1$, then $f'(d) \leqslant 0$ on $(0,\, d_1)$. In this case, $d_1^* = d_1$ and $d^* = \infty$.

(c) If $\theta \geqslant 1$, then $f'(d) \leqslant 0$ on $(0,\, d_1)$. In this case, $d_1^* = d_1$.

Thus, by combining cases (a)-(c), we see that $d^* = \infty$ if $d_1 \leqslant \mu_2$. ∎

Remark 2.4.2 *In Theorem 2.4.2(i), note that* $0 < \mu_2 < d$ *and* $\mu_2 + \frac{\theta-1}{2\sqrt{\theta}}\sigma_2$ *is strictly increases in* $\theta \in (0,\infty)$. *Thus, conditions* $0 \leqslant \theta < 1$ *and* $\mu_2 + \frac{\theta-1}{2\sqrt{\theta}}\sigma_2 \leqslant 0$ *imply that* θ *must be in the interval* $[0,\theta_0]$ *for a unique* $\theta_0 \in (0,1)$. *That means a larger loading factor* $\theta \in (\theta_0,\theta_1)$. *Furthermore, conditions* $\theta > 0$ *and* $0 < \mu_2 + \frac{\theta-1}{2\sqrt{\theta}}\sigma_2 < d_1$ *imply that* θ *must be in the interval* $(\theta_0,\, \theta_1)$ *for a unique* $\theta_1 > 1$. *That means a larger loading factor* $\theta \in (\theta_0,\theta_1)$.

Hence, from (2.64) *of Theorem 2.4.2, we see that if the worst-case value* d_1 *of the insurer's TVaR is not larger than the threshold value* $\max\{f(0), f(d_0)\} = \max\big\{\frac{1+\theta}{2}\big(\mu_2 + \sqrt{\mu_2^2+\sigma_2^2}\big),\ \mu_2 + \sqrt{\theta}\,\sigma_2\big\}$, *the insurer will not worry about its loss and will not buy a reinsurance* ($d^* = \infty$). *If the worst-case value* d_1 *exceeds the threshold value* $\min\{f(0), f(d_0)\} = \min\big\{\frac{1+\theta}{2}\big(\mu_2 + \sqrt{\mu_2^2+\sigma_2^2}\big),\ \mu_2 + \sqrt{\theta}\,\sigma_2\big\}$, *the insurers would like to buy a full reinsurance with* $d^* = 0$ *if* $d_1 > f(0)$ *and the premium is cheaper or a partial reinsurance with* $d^* = d_0$ *if* $d_1 > f(d_0)$ *and the premium is more expensive. In addition, the results in Theorem 2.4.2 (ii) are also reasonable. If fact, if the worst-case value* d_1 *of the insurer's TVaR is bounded by* μ_2*, which is the expected underlying insurance loss evaluated by the reinsurer, the insurers will not worry about its loss and would not buy a reinsurance.* □

Theorem 2.4.3 *Assume that the reinsurance premium in problem* (2.52) *is calculated by* (2.53). *Let* d^* *be the optimal solution to problem* (2.52) *with* $\mathcal{S}_1 = \mathcal{S}_1^\infty$ *and* $\mathcal{S}_2 = \mathcal{S}_2^0$. *Then*

$$d^* = \begin{cases} \hat{G}^{-1}(\frac{\theta}{1+\theta}), & \text{if } d_1 > \hat{G}^{-1}(\frac{\theta}{1+\theta}) + (1+\theta)\int_{\hat{G}^{-1}(\frac{\theta}{1+\theta})}^{\infty}(1-\hat{G}(x))\,\mathrm{d}x, \\ \infty, & \text{if } d_1 \leqslant \hat{G}^{-1}(\frac{\theta}{1+\theta}) + (1+\theta)\int_{\hat{G}^{-1}(\frac{\theta}{1+\theta})}^{\infty}(1-\hat{G}(x))\,\mathrm{d}x. \end{cases} \tag{2.69}$$

where d_1 *is defined in* (2.55).

Proof. According to Theorem 2.4.1 and (2.62), in this case, problem (2.52) is equivalent to $\min_{d\geqslant 0} g(d)$, where

$$g(d) = d_1 \wedge d + (1+\theta)\int_d^\infty (x-d)\,\mathrm{d}\hat{G}(x) = d_1 \wedge d + (1+\theta)\int_d^\infty (1-\hat{G}(x))\,\mathrm{d}x.$$

Note that $g(d) = d + (1+\theta)\int_d^\infty (1-\hat{G}(x))dx$ on $[0,d_1]$ and thus $g(d)$ is a convex function on $[0,d_1]$ by Lemma 2.2(i) of [Wang and Zitikis, 2021]. Therefore, $g(d)$ has a minimum value on $[0,d_1]$ and $g(d)$ attains the minimum value on any $d_1^* \in (0,d_1)$ if and only in $d_1^* \in \{d \in (0,d_1) : 0 \in \partial g\}$, where ∂g is the set of subgradients of g at d and $0 \in \partial g$ if and only if $0 \in [\frac{\partial^- g}{\partial d}, \frac{\partial^+ g}{\partial d}]$, which, together with the left derivative $\frac{\partial^- g}{\partial d} = 1 + (1+\theta)(\hat{G}(d-) - 1)$ and the right derivative $\frac{\partial^+ g}{\partial d} = 1 + (1+\theta)(\hat{G}(d) - 1)$, and Lemma 4.1(ii) of [Cai and Wang, 2021], implies $d_1^* \in \left[\hat{G}^{-1}(\frac{\theta}{1+\theta}),\, \hat{G}^{-1+}(\frac{\theta}{1+\theta})\right]$ and

$$\min_{0 \leqslant d \leqslant d_1} g(d) = g(d_1^*) = g\Big(\hat{G}^{-1}\big(\frac{\theta}{1+\theta}\big)\Big),$$

where, $F^{-1+}(p) = \sup\{t : F(t) \leqslant p\}$ is the right-continuous quantile function of a distribution function F for $p \in (0,1)$. Moreover, $g(d) = d_1 + (1+\theta)\int_d^\infty (1 - \hat{G}(x))dx$ on $[d_1, \infty)$ and $g(d)$ is decreasing on $[d_1, \infty)$ with $g(\infty) = \lim_{d\to\infty} g(d) = d_1$. Thus, $d^* = d_1^*$ if $g(d_1^*) < g(\infty)$ and $d^* = \infty$ if $g(d_1^*) \geqslant g(\infty)$. If fact, if $g(d_1^*) < g(\infty)$, d^* can be any point in the interval $\left[\hat{G}^{-1}(\frac{\theta}{1+\theta}),\, \hat{G}^{-1+}(\frac{\theta}{1+\theta})\right]$. Without loss generality, we choose d^* to be $\hat{G}^{-1}(\frac{\theta}{1+\theta})$. Hence, expression (2.69) holds. Note that if $\hat{G}$ is continuous at $\frac{\theta}{1+\theta}$, then $\hat{G}^{-1}(\frac{\theta}{1+\theta}) = \hat{G}^{-1+}(\frac{\theta}{1+\theta})$. ■

Remark 2.4.3 *When $\mathcal{S}_1 = \mathcal{S}_1^0$ and $\mathcal{S}_2 = \mathcal{S}_2^0$ with $\hat{F} = \hat{G}$, or when the distribution of the underlying insurance loss X is known to both of the insurer and reinsurer, problem (2.52) is solved in Theorem 3.1 of [Cai and Tan, 2007]. It is interesting to see that the non-trivial optimal retention $\hat{G}^{-1}(\frac{\theta}{1+\theta})$ in Theorem 2.4.3 is the same as that in Theorem 3.1 of [Cai and Tan, 2007] and is determined by the pricing distribution $\hat{G}$ used by the reinsurer. However, the optimal retentions in Theorem 2.4.3 depend on $d_1 = \mu_1 + \sigma_1\sqrt{\frac{p_1}{1-p_1}}$ that is the worst case* TVaR_{p_1} *of the underlying insurance loss X over the insurer's uncertainty set $\mathcal{S}_1^\infty$. Theorem 2.4.3 shows that for a larger worst-case TVaR, the insurer would like to buy a non-trivial stop-loss reinsurance, while for a smaller worst-case TVaR, the insurer is willing to undertake all the underlying insurance loss X.* □

2.4.2 Existence of solution and numerical solutions under the general uncertainty sets $\mathcal{S}_1$ and $\mathcal{S}_2$

For the general uncertainty sets $\mathcal{S}_1 = \mathcal{S}(\mu_1, \sigma_1, \hat{F}; \varepsilon_1)$ and $\mathcal{S}_2 = \mathcal{S}(\mu_2, \sigma_2, \hat{G}; \varepsilon_2)$ with $0 < \varepsilon_1,\, \varepsilon_2 < \infty$, the explicit and closed-form solutions to problems (2.50)-(2.52) are not available. In this subsection, we show that optimal solution to Problem (2.52) exists under certain conditions and then use examples to illustrate solutions to these problems. We can consider a generalization of Problem (2.52) as

$$\min_{d \geqslant 0} \sup_{F \in \mathcal{S}_1} \rho^{g_1}\left(X^F \wedge d + \pi(d)\right), \tag{2.70}$$

where $\pi(d)$ is a deterministic reinsurance premium under a given retention $d \geqslant 0$ or a deterministic function of $d \geqslant 0$, and $\mathcal{S}_1$ is an uncertainty set for X. We show in the following theorem that optimal solutions to Problem (2.70) exist under certain conditions.

Theorem 2.4.4 *Assume that the uncertainty set $\mathcal{S}_1$ in problem (2.70) is non-empty and g_1 is an absolutely continuous distortion function. If $\pi(d)$ is lower-semicontinuous and $\pi(d) < \infty$ for any $d \in [0, \infty]$, then an optimal solution $d^* \in [0, \infty]$ to problem (2.70) exists.*

Proof. For simplicity, we write $H(d, F) = \rho^{g_1}\left(X^F \wedge d + \pi(d)\right)$ for $F \in \mathcal{S}_1$ and $d \geqslant 0$. Due to the cash invariance property and monotonicity of a distortion risk measure, we have for all $d \in [0, \infty)$,

$$H(d, F) = \rho^{g_1}\left(X^F \wedge d\right) + \pi(d) \leqslant d + \pi(d) < \infty.$$

Hence, $H(d, F)$ is well-defined for all $d \in [0, \infty)$. Furthermore, from (1.5), (2.11), and (2.13), it is easy to see that $H(d, F) = \pi(d) + h_F(d)$, where, $h_F(d) = d + \int_0^{F(d)} \gamma_1(u)(F^{-1}(u) - d)\,\mathrm{d}u$. We show that for any distribution F, $h_F(d)$ is continuous in $d \in [0, \infty)$. In fact, for any $\delta > 0$, it holds that

$$\begin{aligned}
|h_F(d+\delta) - h_F(d)| &= \left|\delta - \delta \int_0^{F(d+\delta)} \gamma_1(u)\,\mathrm{d}u + \int_{F(d)}^{F(d+\delta)} \gamma_1(u)(F^{-1}(u) - d)\,\mathrm{d}u\right| \\
&\leqslant \delta - \delta \int_0^{F(d+\delta)} \gamma_1(u)\,\mathrm{d}u + \delta \int_{F(d)}^{F(d+\delta)} \gamma_1(u)\,\mathrm{d}u = \delta\, g_1(1 - F(d)) \leqslant \delta,
\end{aligned}$$

where the first inequality is because $1 - \int_0^{F(d+\delta)} \gamma_1(u)\,\mathrm{d}u = 1 - g_1(F(d+\delta)) \geqslant 0$ and $0 \leqslant F^{-1}(u) - d \leqslant \delta$. Similarly, we have $|h_F(d) - h_F(d-\delta)| \leqslant \delta$. As a result, $h_F(d)$ is continuous on $[0, \infty)$. Together with the lower semicontinuity of $\pi(d)$, we know that the function $H(d, F)$ is lower semicontinuous at $d \in [0, \infty)$ for any given distribution F. Write $\bar{H}(d) \triangleq \sup_{F \in \mathcal{S}_1} H(d, F)$, and note that problem (2.70) can also be expressed as $\inf_{d \geqslant 0} \bar{H}(d)$. By the similar argument in the proof of Theorem 2.2.3, we can conclude that $\bar{H}(d) \triangleq \sup_{F \in \mathcal{S}_1} H(d, F)$ is lower semicontinuous in $d \in [0, \infty)$, and there exists $d^* \in [0, \infty]$ such that $\bar{H}$ achieves its minimual value at d^*, i.e., d^* can serve as an optimal solution. ■

Note that $\sup_{G \in \mathcal{S}_2}(1 + \theta)\,\mathbb{E}\left[(X^G - d)_+\right]$ is a lower-semicontinuous function of d. Thus, by Theorem 2.4.4, we see that an optimal solution $d^* \in [0, \infty]$ to problem (2.52) always exists if the premium is calculated by (2.53). To numerically illustrate the solutions obtained in Sections 3-5, in this rest of this section, we assume that $p_1 = 0.90$ in problem (2.52) and $\pi(d)$ is calculated by (2.53) with $\theta = 2$. Further, for $\mathcal{S}_1 = \mathcal{S}(\mu_1, \sigma_1, \hat{F}; \varepsilon_1)$, $\hat{F}$ is the Pareto distribution in Example 1 with $\varepsilon_1 = 0.05$, 1, 2.5; and for $\mathcal{S}_2 = \mathcal{S}(\mu_2, \sigma_2, \hat{G}; \varepsilon_2)$, $\hat{G}$ is the Pareto distribution in Example 2 with $\varepsilon_2 = 0.5$. Note that Assumption 2.1.2 holds under these settings.

Example 3 *(***Numerical solutions under** $\mathcal{S}_1$ **and** $\mathcal{S}_2$*) In this example, we consider the effects of* ε_1 *on the worst-case* $\mathrm{TVaR}_{0.90}$ *of the insurer's risk exposure*

$$V_1(d) \triangleq \sup_{F\in\mathcal{S}_1} \mathrm{TVaR}_{0.90}(X^F \wedge d + \pi_1(d))$$

and optimal retentions d^**, where*

$$\pi_1(d) = \sup_{G\in\mathcal{S}_2} (1+\theta)\mathbb{E}[(X^G - d)_+].$$

We calculate and plot $V_1(d)$ *for* $d \in [0, 30]$ *under three values* $\varepsilon_1 = 0.05, 1, 2.5$ *in Figure 2.3,* $\varepsilon_2 = 0.05$ *and obtain the optimal retention* $d^* = 3.5522$ *that is the same for all the three values of* ε_1*. From the numerical results, we see that the larger the uncertainty set* $\mathcal{S}_1$*, the larger the worst-case* $\mathrm{TVaR}_{0.90}$ *of the insurer's risk exposure. This finding is reasonable since more uncertainties will produce more risks. In addition, from Figure 2.3, we see that for three values of* ε_1*, the worst-case values* $V_1(d)$ *are identical on* $[0, 10]$*, in which the worst-case values* $V_1(d)$ *attain the same minimum value at the same minimizer* $d^* = 3.5522$*; and* ε_1 *affects the worst-case value* $V_1(d)$ *only when* $d > 10$*. We only calculate and plot the the worst-case value* $V_1(d)$ *on [0, 30]. However, we claim that* $d^* = 3.5522$ *is the global minimizer of* $V_1(d)$ *on* $[0, \infty]$*. To see that, for any* $d > 30$*, we have*

$$\begin{aligned}
&\sup_{F\in\mathcal{S}_1} \mathrm{TVaR}_{0.90}(X^F \wedge d + \pi(d)) \geqslant \sup_{F\in\mathcal{S}_1} \mathrm{TVaR}_{0.90}(X^F \wedge d) \geqslant \mathrm{TVaR}_{0.90}(X^{\hat{F}} \wedge d) \\
\geqslant\ &\mathrm{TVaR}_{0.90}(X^{\hat{F}} \wedge 30) = 15.52 \geqslant 9.37 = \sup_{F\in\mathcal{S}_1} \mathrm{TVaR}_{0.90}(X^F \wedge d^* + \pi(d^*)),
\end{aligned}$$

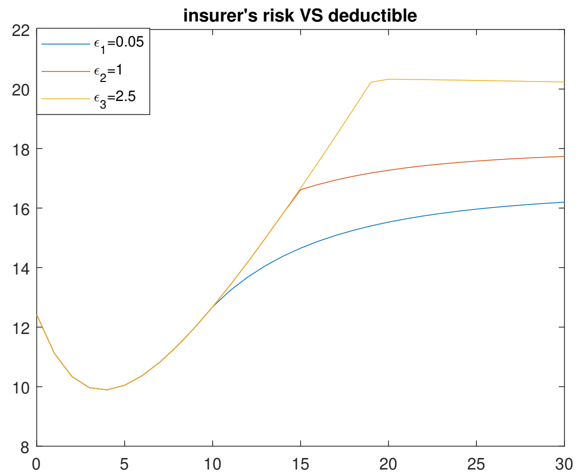
where the third inequality comes from the monotonicity of TVaR and the fact that $X^{\hat{F}} \wedge d \geqslant X^{\hat{F}} \wedge 30$ *for* $d > 30$*. Therefore,* $d^* = 3.5522$ *is the global minimizer of* $V_1(d)$ *on* $[0, \infty]$*.* $\square$

Example 4 *(***Numerical solutions under** $\mathcal{S}_1$ **and** $\mathcal{S}_2^0$*) Under the uncertainty sets* $\mathcal{S}_1$ *and* $\mathcal{S}_2^0$*, we calculate and plot*

$$V_2(d) \triangleq \sup_{F\in\mathcal{S}_1} \mathrm{TVaR}_{0.90}(X^F \wedge d + \pi_2(d))$$

for $d \in [0, 30]$ *under three different values of* $\varepsilon_1 = 0.05, 1, 2.5$ *in Figure 2.4. We obtain the optimal retention* $d^* = 3.5380$ *that is the same for all the three values of* ε_1 *in this example, where*

$$\pi_2(d) = \sup_{G\in\mathcal{S}_2^0} (1+\theta)\mathbb{E}[(X^G - d)_+] = (1+\theta)\mathbb{E}[(X^{\hat{G}} - d)_+].$$



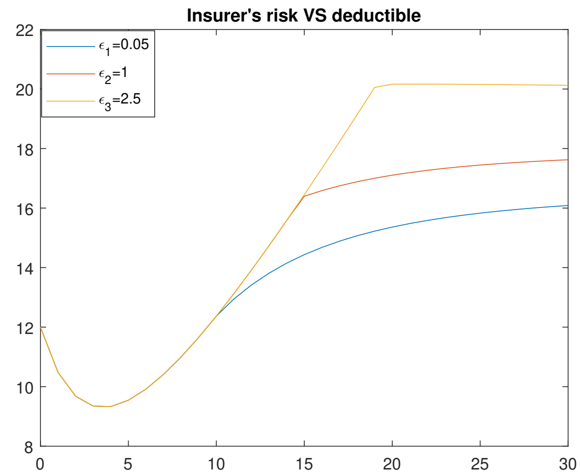
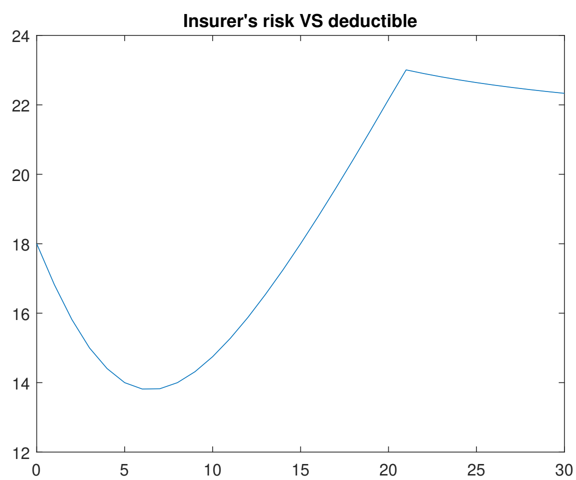


Figure 2.3: Numerical Illustration for Example 3

Figure 2.4: Numerical Illustration for Example 4



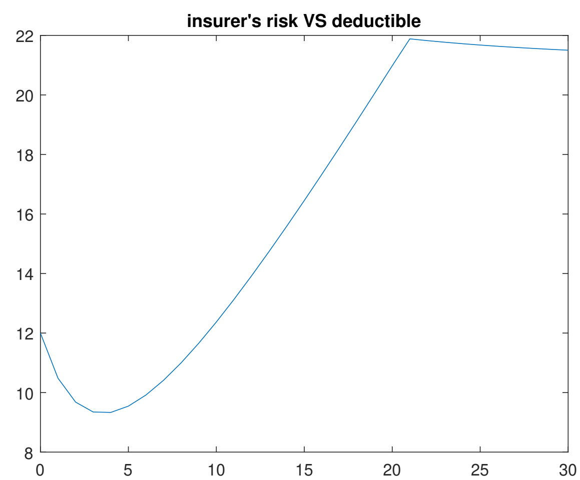


Figure 2.5: Numerical Illustration for Example 5

Figure 2.6: Numerical Illustration for Example 6

The behaviour of the worst-case value $V_2(d)$ is similar to the worst-case value $V_1(d)$ discussed in Example 3. Also, we claim by the same arguments used in Example 3 that $d^* = 3.5380$ is the global minimizer of $V_2(d)$ on $[0, \infty]$ for all the three values of ε_1. In addition, we find the optimal retention $d^* = 3.5380$ in this example is smaller than the optimal retention $d^* = 3.5522$ in Example 3. Note that the premium $\pi_2(d)$ in this example is lower than the premium $\pi_1(d)$ in Example 3 while the worst-case value $V_2(d)$ in this example is smaller than the worst-case value $V_2(d)$ in Example 3. These numerical results suggest that with a lower premium and a smaller worst-case value, the optimal reinsurance for the insurer is to retain less risks. □

At the end of this section, we also want to compare the numerical results obtained in Examples 3 and 4 with those calculated by using Theorem 2.4.2 and Theorem 2.4.3.

Example 5 *(**Numerical solutions under $\mathcal{S}_1^\infty$ and $\mathcal{S}_2^\infty$**) With the uncertainty sets $\mathcal{S}_1^\infty$ and $\mathcal{S}_2^\infty$, the worst-case* $\mathrm{TVaR}_{0.90}$ *of the insurer's risk exposure*

$$V_3(d) \triangleq \sup_{F \in \mathcal{S}_1^\infty} \mathrm{TVaR}_{0.90}(X^F \wedge d + \pi_3(d)),$$

for $d \in [0, 30]$ and under three different values of $\varepsilon_1 = 0.05, 1, 2.5$, is plotted in Figure 2.5, where

$$\pi_3(d) = \sup_{G \in \mathcal{S}_2^\infty} (1+\theta)\mathbb{E}[(X^G - d)_+] = \frac{1+\theta}{2}\left(\mu_2 - d + \sqrt{(\mu_2 - d)^2 + \sigma_2^2}\right)$$

by (2.63). In this example, by Theorem 2.4.2 (i), we find the optimal deductible is $d^* = 6.4495$. Comparing with Examples 3 and 4, we notice that when the uncertainty set is getting larger in the sense of Wasserstein distances ($\varepsilon_i = \infty$, $i = 1, 2$ in this example) or when the insurer and reinsurer are facing more uncertainties, the premium $\pi_3(d)$ in this example is higher than the premiums $\pi_1(d)$ and $\pi_2(d)$ in Examples 3 and 4; the worst-case value $V_3(d)$ is larger than the worst-case values $V_2(d)$ and $V_3(d)$; and the optimal retention $d^* = 6.4495$ in this example is larger than those in Examples 3 and 4. This is also a reasonable finding as with more uncertainties, the reinsurers need to charge a higher premium and the insurer has to retain more risks. □

Example 6 *(**Numerical solutions under $\mathcal{S}_1^\infty$ and $\mathcal{S}_2^0$**) With the uncertainty sets $\mathcal{S}_1^\infty$ and $\mathcal{S}_2^0$, the worst-case* $\mathrm{TVaR}_{0.90}$ *of the insurer's risk exposure*

$$V_4(d) \triangleq \sup_{F \in \mathcal{S}_1^\infty} \mathrm{TVaR}_{0.90}(X^F \wedge d + \pi_4(d)),$$

for $d \in [0, 30]$ and under three different values of $\varepsilon_1 = 0.05, 1, 2.5$, is plotted in Figure 2.6, where

$$\pi_4(d) = \sup_{G \in \mathcal{S}_2^0} (1+\theta)\mathbb{E}[(X^G - d)_+] = (1+\theta)\mathbb{E}[(X^{\hat{G}} - d)_+].$$

The premium in this example is the same as in Example 4 while the insurer is facing more uncertainties as $\varepsilon_1 = \infty$. Therefore, the worst-case value $V_4(d)$ is not less than the worst-case value $V_3(d)$ for all $d \geqslant 0$ and all the three values of ε_1 in Example 4. However, the worst-case value $V_4(d)$ is equal to the worst-case value $V_3(d)$ on $[0, 5]$, in which the worst-case values $V_4(d)$ and $V_3(d)$ both attain the same minimum value at the same minimizer $d^* = 3.5380$. In this example, the minimizer $d^* = 3.5380$ is obtained by using Theorem 2.4.3. □

2.5 Concluding remarks

In this chapter, we investigate the worst-case values of the distortion risk measures of the stop-loss $(X - d)_+$ and limited loss $X \wedge d$ when the distribution of an underlying loss random variable X is uncertain. We use the uncertainty set $\mathcal{S}(\mu, \sigma, \hat{F}; \varepsilon)$, defined in (2.2) and introduced in [Bernard et al., 2020b], to represent all the possible distributions of X. The uncertainty set $\mathcal{S}(\mu, \sigma, \hat{F}; \varepsilon)$ contains the information on the mean, variance, and empirical/reference distribution of X available to a decision maker. The uncertainty set $\mathcal{S}(\mu, \sigma, \hat{F}; \varepsilon)$ is reasonable and practical and it expresses the beliefs of a decision maker in the possible distributions of X. We derive the expressions for the worst-case values of the distortion risk measures of $(X-d)_+$ and $X \wedge d$ when distortion functions are absolutely continuous. We also find the distributions in (2.2) under which the worst-case values are attainable. These results are important in robust risk management for insurance, finance, operations research, and many other fields. Our results can recover the classical results of [Jagannathan, 1977] and [Lo, 1987] on the worst-case values of the expectations of $(X-d)_+$ and $X \wedge d$ when only the information of the mean and variance of X is available.

To illustrate the applications of the worst-case values of the distortion risk measures of $(X - d)_+$ and $X \wedge d$, in this chapter, we discuss the optimal stop-loss reinsurance that minimizes the worst-case value of the TVaR of the insurer's risk exposure in a stop-loss reinsurance. We propose two methods for the reinsurer to determine reinsurance premiums when the distribution of an underlying loss random variable X is uncertain. The model settings and reinsurance premiums used in our chapter are different from [Hu et al., 2015] and [Liu and Mao, 2021]. Especially, the reinsurance premiums used in [Liu and Mao, 2021] are uncertain while the reinsurance premiums in our model are deterministic. A deterministic premium charged by the reinsurer or offered to the insurer is a realistic requirement in a reinsurance agreement even if the distribution of an underlying loss random variable X is uncertain. In addition, another important difference between our model settings and those used in [Hu et al., 2015] and [Liu and Mao, 2021] is the beliefs of the insurer and reinsurer in the distribution of the underlying insurance loss random variable X. [Hu et al., 2015] and [Liu and Mao, 2021] assume that the insurer and reinsurer have the same beliefs in the distribution of the underlying insurance loss random variable X, while our model assumes the beliefs of the insurer and reinsurer in the distribution may be different. Such heterogeneous beliefs in the distribution of the underlying insurance loss random variable X are more realistic assumptions, which have attracted a lot of interest in the recent studies of insurance. See [Jiang et al., 2019] and the references therein for more examples of heterogeneous beliefs. We will explore more applications of the results obtained under the uncertainty set $\mathcal{S}(\mu, \sigma, \hat{F}; \varepsilon)$ to robust risk management problems in the future research.

2.6 Appendix

Lemma 2.6.1 (Theorem 2, [Bernard et al., 2020b]) *Consider a distortion risk measure ρ^g with distortion function g and the corresponding weight function γ, and an uncertainty set $\mathcal{S}(\mu, \sigma, \hat{G}; \varepsilon)$ defined in (2.2), where the distribution $\hat{G}$ has mean $\hat{\mu}$ and variance $\hat{\sigma}^2$ and For $\lambda \geqslant 0$, define $\ell_\lambda^\uparrow = \arg\min_{\ell \in \mathcal{K}} ||\ell - \gamma - \lambda \hat{G}^{-1}||_2$, where $\mathcal{K}$ is defined in (2.22). Denote $c_0 = \operatorname{corr}(\hat{G}^{-1}(U), l_0^\uparrow(U))$. Assume $\ell_\lambda^\uparrow$ is not constant for any $\lambda > 0$. Then, solutions to the problem $\sup_{G \in \mathcal{S}(\mu, \sigma, \hat{G}; \varepsilon)} \rho^g(X^G)$ are given as below:*

(i) If $(\hat{\mu} - \mu)^2 + (\hat{\sigma} - \sigma)^2 < \varepsilon^2 < (\hat{\mu} - \mu)^2 + (\hat{\sigma} - \sigma)^2 + 2\sigma\hat{\sigma}(1 - c_0)$, then the maximizer G_λ^* to $\sup_{G \in \mathcal{S}(\mu, \sigma, \hat{G}; \varepsilon)} \rho^g(X^G)$ is unique and the quantile function of G_λ^* is given by $G_\lambda^{*-1}(u) = \mu + \sigma\left(\frac{l_\lambda^\uparrow(u) - a_\lambda}{b_\lambda}\right)$, where $a_\lambda = \mathbb{E}[l_\lambda^\uparrow(U)]$, $b_\lambda = \sqrt{\operatorname{var}(l_\lambda^\uparrow(U))}$, and $\lambda > 0$ is the unique positive solution to the equation $d_W(\hat{G}^{-1}, G_\lambda^{*-1}) = \varepsilon$.

(ii) Let $(\hat{\mu} - \mu)^2 + (\hat{\sigma} - \sigma)^2 + 2\sigma\hat{\sigma}(1 - c_0) \leqslant \varepsilon^2$. If $\ell_0^\uparrow$ is not constant, then the maximizer G_0^* to $\sup_{G \in \mathcal{S}(\mu, \sigma, \hat{G}; \varepsilon)} \rho^g(X^G)$ is unique and the quantile function of G_0^* is given by $G_0^{*-1}(u) = \mu + \sigma\left(\frac{l_0^\uparrow(u) - a_0}{b_0}\right)$. If $\ell_0^\uparrow$ is constant, then the quantile function of the maximizer G_0^* to $\sup_{G \in \mathcal{S}(\mu, \sigma, \hat{G}; \varepsilon)} \rho^g(X^G)$ is the constant μ and the supremum $\sup_{G \in \mathcal{S}(\mu, \sigma, \hat{G}; \varepsilon)} \rho^g(X^G)$ is not attained on $\mathcal{S}(\mu, \sigma, \hat{G}; \varepsilon)$.

Lemma 2.6.2 (Proposition 3, [Bernard et al., 2020b]) *Let* $K = \{(\mu, \sigma) \,|\, \underline{\mu} \leqslant \mu \leqslant \bar{\mu},\, |\underline{\mu}| \leqslant |\bar{\mu}|,\, \underline{\sigma} \leqslant \sigma \leqslant \bar{\sigma}\}$ *and* $\mathcal{S}_\varepsilon(K) = \{G \in \mathcal{S}(\mu, \sigma, \hat{G}; \varepsilon) \,|\, (\mu, \sigma) \in K\}$. *If for all* $(\mu, \sigma) \in K$, *it holds that* $(\hat{\mu} - \mu)^2 + (\hat{\sigma} - \sigma)^2 < \varepsilon^2 < (\hat{\mu} - \mu)^2 + (\hat{\sigma} - \sigma)^2 + 2\sigma\hat{\sigma}(1 - c_0)$, *then*

$$\sup_{G \in \mathcal{S}_\varepsilon(K)} \rho^g(X^G) \;=\; \sup_{G \in \mathcal{S}(\mu^K_{max}, \sigma^K_{max}, \hat{G}; \varepsilon)} \rho^g(X^G),$$

where $(\mu^K_{max}, \sigma^K_{max}) = (\underline{\mu}, \bar{\sigma})$ *if* $\hat{\mu} < -1/\lambda$ *and* $(\mu^K_{max}, \sigma^K_{max}) = (\bar{\mu}, \bar{\sigma})$, *otherwise.*

Lemma 2.6.3 *The set* $\mathcal{S}'_1 \triangleq \left\{F : d_W(F, \hat{F}) \leqslant \varepsilon_1,\, \mathbb{E}[X^F] = \mu_1,\, \mathrm{var}(X^F) \leqslant \sigma_1^2\right\}$ *is weakly compact.*

Proof. First note that $\mathcal{Q}'^d_1 \subset L^2(0,1)$, where the space $L^2(0,1)$ with the metric induced by the L^2-norm is a complete metric space. For any $F^{-1} \in \mathcal{Q}'^d_1$, the L^2-norm of F^{-1} is bounded by the constant $\|\hat{F}^{-1} - d\|_2 + \varepsilon < \infty$ because

$$\|F^{-1}\|_2 \leqslant \|\hat{F}^{-1} - d\|_2 + \|\hat{F}^{-1} - d - F^{-1}\|_2 = \|\hat{F}^{-1} - d\|_2 + d_W(\hat{F}^{-1} - d, F^{-1}) \leqslant \|\hat{F}^{-1} - d\|_2 + \varepsilon_1.$$

We first show that $\mathcal{S}'_1$ is closed. Suppose F is a limit of $\mathcal{S}'_1$ in the sense that there exists a sequence $\{F_n, n = 1, 2, ...\} \subset \mathcal{S}'_1$ such that $F_n^{-1}(U)$ converges to $F^{-1}(U)$ in distribution, i.e., $\lim_{n\to\infty} F_n(x) = F(x)$ for every continuous point x of F. Let $X_n = F_n^{-1}(U)$ for $n = 1, 2, ...$, and $X_\infty = F^{-1}(U)$. Since $\mathbb{E}[X_n^2] = \mathrm{var}(X_n) + \mathbb{E}[X_n]^2 \leqslant \mu_1^2 + \sigma_1^2$ for all n, by Theorem 4.6.2 of [Durrett, 2019], $\{X_n, n = 1, 2, ...\}$ is uniformly integrable. Note that $X_n = F_n^{-1}(U)$ also converges to $X_\infty = F^{-1}(U)$ in probability. Also we have $\mathbb{E}[|X_n|] < \infty$ for all n, then Theorem 4.6.3 of [Durrett, 2019] says that X_n also converges to X_∞ in L^1, which means that $\mathbb{E}[|X_n - X_\infty|] = \int_0^1 |F_n^{-1}(u) - F^{-1}(u)|\, du \to 0$ as $n \to \infty$, and $\mathbb{E}[X_\infty] = \lim_{n\to\infty} \mathbb{E}[X_n] = \mu_1$. Furthermore, since the power function $h(x) = x^2$ is continuous and bounded below, X_n converges to X_∞ in distribution implies $\mathbb{E}[X_\infty^2] \leqslant \liminf \mathbb{E}[X_n^2] = \mu_1^2 + \sigma_1^2$. Together with $\mathbb{E}[X_\infty] = \mu_1$, we have $\mathrm{var}(X_\infty) \leqslant \sigma_1^2$. Let $Y_n = F_n^{-1}(U) - \hat{F}^{-1}(U)$ and $Y_\infty = F^{-1}(U) - \hat{F}^{-1}(U)$. Then $\mathbb{E}[Y_n] = \mu_1 - \hat{\mu}_1$ and $\mathbb{E}[Y_n^2] = d_W(F, \hat{F})^2 \leqslant \varepsilon_1^2$ for all $n = 1, 2, ...$ Using the similar argument for X_n, we can also verify that $\mathbb{E}[Y_\infty^2] \leqslant \varepsilon_1^2$. Therefore, $F \in \mathcal{S}'_1$, i.e., $\mathcal{S}'_1$ is closed.

Second, we show $\mathcal{S}'_1$ is tight. For any $\delta > 0$, define $K_\delta = (\mu_1 - d)/\delta < \infty$. By Markov inequality, we then have that for all $F \in \mathcal{S}'_1$, the associated distribution F satisfies

$$F(K_\delta) \geqslant 1 - \frac{\mathbb{E}[F^{-1}(U)]}{K_\delta} = 1 - \frac{\mu_1 - d}{K_\delta} \geqslant 1 - \delta$$

which implies (uniform) tightness of $\mathcal{S}_1'$.

Since $\mathcal{S}_1'$ is closed and tight, by Theorem 6.2 of [Birghila and Pflug, 2019], we conclude that $\mathcal{S}_1'$ is weakly compact. ■

Proof of Theorem 2.2.4. To prove this theorem, we first expand sets $\mathcal{Q}_1$ and $\mathcal{Q}_1^d$ to larger sets $\mathcal{Q}_1'$ and $\mathcal{Q}_1'^d$, respectively, and consider the problem $\sup_{F^{-1}\in\mathcal{Q}_1'} \rho^{g_1}(X^F \wedge d)$, which has the following expression by using the arguments similar to those for Lemma 2.2.1: For any $d \geqslant 0$,

$$\sup_{F^{-1}\in\mathcal{Q}_1'} \rho^{g_1}(X^F \wedge d) = d + \sup_{F^{-1}\in\mathcal{Q}_1'^d} \rho^{g_1}(F^{-1}(U) \wedge 0), \tag{2.71}$$

where $\mathcal{Q}_1' \triangleq \left\{F^{-1} : d_W(F, \hat{F}) \leqslant \varepsilon_1,\ \mathbb{E}[X^F] = \mu_1,\ \mathrm{var}(X^F) \leqslant \sigma_1^2\right\}$ and

$$\mathcal{Q}_1'^d \triangleq \left\{F^{-1} : F^{-1} + d \in \mathcal{Q}_1'\right\}. \tag{2.72}$$

It is easy to see that $\mathcal{Q}_1 \subset \mathcal{Q}_1'$ and $\mathcal{Q}_1^d \subset \mathcal{Q}_1'^d$.

We first show that the supremum $\sup_{F^{-1}\in\mathcal{Q}_1'^d} \rho^{g_1}(F^{-1}(U) \wedge 0)$ can be attained at some quantile function in $\mathcal{Q}_1'^d$. In doing so, note that equation (2.12) in Lemma 2.2.2 implies

$$\sup_{F^{-1}\in\mathcal{Q}_1'^d} \rho^{g_1}(F^{-1}(U) \wedge 0) = \sup_{F^{-1}\in\mathcal{Q}_1'^d} \min_{\beta\in[p_1,1]} L_0(\beta, F^{-1}). \tag{2.73}$$

There exists a sequence $\{G_n^{-1}, n = 1, 2, ...\} \subset \mathcal{Q}_1'^d$ s.t. $\sup\limits_{F^{-1}\in\mathcal{Q}_1'^d} \rho^{g_1}(F^{-1}(U)\wedge 0) = \lim\limits_{n\to\infty} \rho^{g_1}(G_n^{-1}(U)\wedge$ $0)$.

Write $\mathcal{Q}_1'^d = \{F^{-1} - d : F \in \mathcal{S}_1'\}$ where $\mathcal{S}_1'$ defined in Lemma 2.6.3 is weakly compact. Thus, the set $\{G_n^{-1}, n = 1, 2, ...\}$ has a limit in $\mathcal{Q}_1'^d$, denoted by $G_\infty^{-1} \in \mathcal{Q}_1'^d$. Using the similar argument in the proof of Lemma 2.6.3, $G_n^{-1}(U)$ converges to $G_\infty^{-1}(U)$ in L^1. By Proposition 4 of [Wang and Wang, 2020], if $\|\gamma\|_\infty < \infty$, then the risk measure ρ^{g_1} is continuous with respect to the L^1-convergences of random variables. Thus, $\rho^{g_1}(G_\infty^{-1}(U) \wedge 0) = \lim\limits_{n\to\infty} \rho^{g_1}(G_n^{-1}(U) \wedge 0) = \max\limits_{F^{-1}\in\mathcal{Q}_1'^d} \rho^{g_1}(F^{-1}(U) \wedge 0)$.

In the rest of the proof, we denote $\beta_\infty = G_\infty(0)$. By equations (2.12) and (2.73), we see that

$$\begin{aligned} L_0(\beta_\infty, G_\infty^{-1}) = \rho^{g_1}(G_\infty^{-1}(U) \wedge 0) = \max_{F^{-1}\in\mathcal{Q}_1'^d} \rho^{g_1}(F^{-1}(U) \wedge 0) &= \max_{F^{-1}\in\mathcal{Q}_1'^d} \min_{\beta\in[\alpha_1,1]} L_0(\beta, F^{-1}) \\ &= \min_{\beta\in[\alpha_1,1]} \max_{F^{-1}\in\mathcal{Q}_1'^d} L_0(\beta, F^{-1}), \end{aligned} \tag{2.74}$$

where the last equality comes from the min-max theorem. The detailed proof of equality (2.74) is given after this proof. Therefore, $L_0(\beta_\infty, G_\infty^{-1})$ is a saddle value for the above min-max problem.

Next, we show that $L_0(\beta_0, F_{\beta_0,\lambda_0}^{-1} - d)$ is also a saddle value for the same min-max problem in (2.74), where $F_{\beta_0,\lambda_0}^{-1}$ is given in Theorem 2.2.3. We can write $\mathcal{Q}_1'^d = \{G^{-1} \in \mathcal{S}(\mu, \sigma, \hat{F}^{-1} - d; \varepsilon_1) \,|\, (\mu, \sigma) \in K)\}$, where $K = \{(\mu, \sigma) \,|\, \mu = \mu_1 - d, \;\; 0 \leqslant \sigma \leqslant \sigma_1\}$. Under the assumption that $\varepsilon_1^2 < (\hat{\mu}_1 - \mu_1)^2 + (\hat{\sigma}_1 - \sigma)^2 + 2\sigma\hat{\sigma}_1(1 - c_{1,\beta_0})$ for all $\sigma \leqslant \sigma_1$, where $\beta_0 \in [\alpha_1, 1]$ is given in Theorem 2.2.3, we have

$$\begin{aligned}\max_{F^{-1}\in\mathcal{Q}_1'^d} L_0(\beta_0, F^{-1}) &= \|\gamma_{1,\beta_0}\|_1 \max_{F^{-1}\in\mathcal{Q}_1'^d} \rho^{\tilde{g}_{1,\beta_0}}(X^F) \\ &= \|\gamma_{1,\beta_0}\|_1 \max_{F^{-1}\in\mathcal{Q}_1^d} \rho^{\tilde{g}_{1,\beta_0}}(X^F) = \max_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta_0, F^{-1}),\end{aligned} \tag{2.75}$$

where the first and third equalities come from (2.31) and the second equality comes from Lemma 2.6.2. Since $\mathcal{Q}_1^d \subset \mathcal{Q}_1'^d$, we always have $\max\limits_{F^{-1}\in\mathcal{Q}_1'^d} L_0(\beta, F^{-1}) \geqslant \sup\limits_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta, F^{-1})$ for any $\beta \in [\alpha_1, 1]$. Hence

$$L_0(\beta_\infty, G_\infty^{-1}) = \min_{\beta\in[\alpha_1,1]} \max_{F^{-1}\in\mathcal{Q}_1'^d} L_0(\beta, F^{-1}) \geqslant \min_{\beta\in[\alpha_1,1]} \max_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta, F^{-1}) = L_0(\beta_0, F_{\beta_0,\lambda_0}^{-1} - d), \tag{2.76}$$

where the first equality is given by (2.74) and second equality is given by (2.26). In the rest of the proof, we write $G_0^{-1} = F_{\beta_0,\lambda_0}^{-1} - d$ for simplicity. Meanwhile, (2.12) and (2.75) implies

$$L_0(\beta_\infty, G_\infty^{-1}) \leqslant L_0(\beta_0, G_\infty^{-1}) \leqslant \max_{F^{-1}\in\mathcal{Q}_1'^d} L_0(\beta_0, F^{-1}) = \max_{F^{-1}\in\mathcal{Q}_1^d} L_0(\beta_0, F^{-1}) = L_0(\beta_0, G_0^{-1}). \tag{2.77}$$

Equations (2.76) and (2.77) together imply

$$L_0(\beta_\infty, G_\infty^{-1}) = L_0(\beta_0, G_\infty^{-1}) = L_0(\beta_0, G_0^{-1}), \tag{2.78}$$

i.e., $L_0(\beta_0, G_0^{-1})$ is also a saddle value for the same min-max problem in (2.74)

From the equation (2.76) and the proof of Theorem 2.2.3, both G_0^{-1} and G_∞^{-1} are maximizers of the problem $\max_{F^{-1}\in\mathcal{Q}_1'^d} L_0(\beta_0, F^{-1})$, which has unique solution from the proof of Proposition 3 of [Bernard et al., 2020b]. Therefore, it holds true that $G_0^{-1} = G_\infty^{-1}$ a.s. Since $G_\infty^{-1} \in \mathcal{Q}_1^d$ and $\mathcal{Q}_1^d \subset \mathcal{Q}_1'^d$, we finally conclude that

$$L_0(\beta_\infty, G_\infty^{-1}) \leqslant \sup_{F^{-1}\in\mathcal{Q}_1^d} \rho^{g_1}(F^{-1}(U) \wedge 0) \leqslant \max_{F^{-1}\in\mathcal{Q}_1'^d} \rho^{g_1}(F^{-1}(U) \wedge 0) = L_0(\beta_0, G_0^{-1}) = L_0(\beta_\infty, G_\infty^{-1}).$$

Therefore, $G_\infty^{-1} + d = G_0^{-1} + d = F_{\beta_0,\lambda_0}^{-1} \in \mathcal{Q}_1$ is the solution to the problem (2.3), and the worst case risk measure and the inf-sup value in (2.18) are equal. ■

Definition 2.6.1 (Quasi-convexity and quasi-concavity) *(See [Di Guglielmo, 1977]) For a function $f : \mathcal{T} \mapsto \mathbb{R}$ is a real-valued function defined on a convex subset $\mathcal{T}$ of a real vector space, if for all $x, y \in \mathcal{T}$ and $\lambda \in [0,1]$, we have*

$$f(\lambda x + (1-\lambda)y) \leqslant \max\{f(x), f(y)\},$$

then f is quasi-convex. If for all $x, y \in \mathcal{T}$ and $\lambda \in [0,1]$, we have

$$f(\lambda x + (1-\lambda)y) \geqslant \min\{f(x), f(y)\},$$

we say f is quasi-concave.

With the help of the definition of quasi-convexity and quasi-concavity, we can step into the following founding theorem:

Lemma 2.6.4 (Min-Max Theorem, [Sion, 1958]) *Let $\mathbb{X}$ be a compact convex subset of a linear topological space and $\mathbb{Y}$ a compact convex subset of a linear topological space. If f is a real-valued function on $\mathbb{X} \times \mathbb{Y}$ with $f(x, \cdot)$ upper semi-continuous and quasi-concave on $\mathbb{Y}$, $\forall x \in \mathbb{X}$, and $f(\cdot, y)$ lower semi-continuous and quasi-convex on $\mathbb{X}, \forall y \in \mathbb{Y}$, then* $\max_{y\in\mathbb{Y}} \min_{x\in\mathbb{X}} f(x,y) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} f(x,y)$.

Proof of Equation (2.74). To apply Lemma 2.6.4, we need to verify that all the conditions in Lemma 2.6.4 are fulfilled.

(i) It is easy to see that $\mathcal{Q}_1'^d$ is convex and $\mathcal{Q}_1'^d \subset \mathcal{K}$, where $\mathcal{K}$ is defined in (2.22) and is a linear topological space. Also, Lemma 2.6.3 shows that $\mathcal{Q}_1'^d$ is compact.

(ii) For any $\beta \in [\alpha_1, 1]$, $L_0(\beta, F^{-1}) = \int_0^\beta \gamma_1(u) F^{-1}(u)\,\mathrm{d}u$ is an affine transformation from $\mathcal{Q}_1'^d$ to $\mathbb{R}$. Hence, $L_0(\beta, \cdot)$ is quasi-concave on $\mathcal{Q}_1'^d$ for any given $\beta \in [\alpha_1, 1]$.

(iii) We show that for any $\beta \in [\alpha_1, 1]$, $L_0(\beta, \cdot)$ is continuous in $\mathcal{Q}_1'^d$. For any $F_0^{-1} \in \mathcal{Q}_1'^d$ and every sequence of quantile functions $\{F_n^{-1}\} \subset \mathcal{Q}_1'$ such that $F_n^{-1} \xrightarrow{L^2} F_0^{-1}$, one has

$$\begin{aligned}\left|L_0(\beta, F_n^{-1}) - L_0(\beta, F_0^{-1})\right| &= \left|\int_0^1 \gamma_{1,\beta}(u)\left(F_n^{-1}(u) - F_0^{-1}(u)\right)\mathrm{d}u\right| \\ &\leqslant \left\|\gamma_{1,\beta}\right\|_2 \left\|F_n^{-1} - F_0^{-1}\right\|_2 \;\to\; 0, \quad \text{as } n \to \infty.\end{aligned}$$

Therefore, $L_0(\beta, \cdot)$ is continuous in $\mathcal{Q}_1'^d$.

(iv) We show that for any given $F^{-1} \in \mathcal{Q}_1'^d$, $L_0(\cdot, F^{-1})$ is quasi-convex on $[0,1]$. It is easy to see that $L_0(\cdot, F^{-1})$ is non-increasing on $[0, F(d))$, and is non-decreasing on $(F(d), 1]$. Thus, with a given $F^{-1} \in \mathcal{Q}_1'^d$, for any $\alpha_1 \leqslant \beta_1 < \beta_2 \leqslant 1$, it is easy to check that, for any $\delta \in [0,1]$,

$$L_0(\delta\beta_1 + (1-\delta)\beta_2, F^{-1}) \leqslant \begin{cases} L_0(\beta_1, F^{-1}), & \text{if } \beta_1 < \delta\beta_1 + (1-\delta)\beta_2 \leqslant F(d), \\ L_0(\beta_2, F^{-1}), & \text{if } F(d) \leqslant \delta\beta_1 + (1-\delta)\beta_2 < \beta_2, \end{cases}$$
$$\leqslant \max\left\{L_0(\beta_1, F^{-1}), L_0(\beta_2, F^{-1})\right\},$$

which means that $L_0(\cdot, F^{-1})$ is quasi-convex on $[0,1]$.

(v) The continuity of $L_0(\cdot, F^{-1})$ on $[\alpha_1, 1]$ is obvious since $F^{-1}, \gamma_1 \in L^2([0,1])$.

Hence, by (i)-(v), we obtain equation (2.74) by using Lemma 2.6.4. ■

Proof of Proposition 2.3.3.

(i) Assume that $\sup_{G \in \mathcal{S}_2} \rho^{g_2}\left((X^G - d)_+\right) = \rho^{g_2}\left((X^{G^*} - d)_+\right)$, i.e., G^* is a maximizer to problem (2.4). By equation (2.41) with $\beta^* = G^*(d)$, we have

$$\rho^{g_2}\left((G^{*-1}(U) - d)_+\right) = \int_{\beta^*}^1 \gamma_2(u)\left(G^{*-1}(u) - d\right)\,\mathrm{d}u = H_0(\beta^*, G^{*-1} - d).$$

If $\beta^* = G^*(d) > \alpha_2$, then by the definition of α_2, we have $\rho^{g_2}\left((G^{*-1}(U) - d)_+\right) = 0$. Together with the observation $(G^{-1}(U) - d)_+ \geqslant 0$ for all $G \in \mathcal{S}_2$, we see that $\rho^{g_2}\left((X^G - d)_+\right) = 0$, i.e., $G(d) \geqslant \alpha_2$, for any $G \in \mathcal{S}_2$. Therefore, for any $\beta \in [0, \alpha_2]$, $\sup_{\beta \in [0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d) = 0$ by (2.41). In particular, we can take α_2 as the maximizer. In the following, we consider the case $\beta^* = G^*(d) \leqslant \alpha_2$. Like the problem $\max_{F^{-1} \in \mathcal{Q}_1'^d} L_0(\beta_0, F^{-1})$ discussed in the proof of Theorem 2.2.4, by the proof of Proposition 3 of [Bernard et al., 2020b], we see that the following optimization problem

$$\sup_{K^{-1} \in \mathcal{Q}_2^d} H_0(\beta^*, K^{-1}) \tag{2.79}$$

has a unique solution $G^{-1}_{\beta^*,\lambda_{\beta^*}} - d \in \mathcal{Q}_2^d$, where λ_{β^*} solves $d_W(\hat{G}^{-1}, G^{-1}_{\beta^*,\lambda_{\beta^*}}) = \varepsilon_2$. For simplicity, we write $\tilde{G}^{-1} = G^{-1}_{\beta^*,\lambda_{\beta^*}} - d$.

First, if $\tilde{G}^{-1}(\beta^*) < 0$, then $\tilde{G}(0) > \beta^*$, and $\tilde{G}^{-1}(u) \leqslant 0$ for all $\beta^* \leqslant u < \tilde{G}(0)$ by (2.13). Since $G^{*-1} - d \in \mathcal{Q}_2^d$, we have

$$\begin{aligned}
H_0(\beta^*, G^{*-1} - d) \leqslant H_0(\beta^*, \tilde{G}^{-1}) &= \int_{\tilde{G}(0)}^{1} \gamma_2(u)\tilde{G}^{-1}(u)\,\mathrm{d}u + \int_{\tilde{G}(0)}^{\beta^*} \gamma_2(u)\tilde{G}^{-1}(u)\,\mathrm{d}u \\
&\leqslant \int_{\tilde{G}(0)}^{1} \gamma_2(u)\tilde{G}^{-1}(u)\,\mathrm{d}u = H_0(\tilde{G}(0), \tilde{G}^{-1}) \\
&\leqslant \sup_{\beta\in[0,1]} \sup_{K^{-1}\in\mathcal{Q}_2^d} H_0(\beta, K^{-1}) = H_0(\beta^*, G^{*-1} - d).
\end{aligned}$$

Therefore, all the inequalities above are equalities.

Second, if $\tilde{G}^{-1}(\beta^*) \geqslant 0$. Then $\tilde{G}(0) \geqslant \beta^*$, and $\tilde{G}^{-1}(u) \geqslant 0$ for all $\tilde{G}(0) \leqslant u \leqslant \beta^*$ by (2.13). It follows that

$$\begin{aligned}
H_0(\beta^*, G^{*-1} - d) &= \sup_{\beta\in[0,1]} \sup_{K^{-1}\in\mathcal{Q}_2^d} H_0(\beta, K^{-1}) \\
&\geqslant H_0(\tilde{G}(0), \tilde{G}^{-1}) = \int_{\tilde{G}(0)}^{1} \gamma_2(u)\tilde{G}^{-1}(u)\,\mathrm{d}u \\
&= \int_{\tilde{G}(0)}^{\beta^*} \gamma_2(u)\tilde{G}^{-1}(u)\,\mathrm{d}u + \int_{\beta^*}^{1} \gamma_2(u)\tilde{G}^{-1}(u)\,\mathrm{d}u \\
&\geqslant \int_{\beta^*}^{1} \gamma_2(u)\tilde{G}^{-1}(u)\,\mathrm{d}u = H_0(\beta^*, \tilde{G}^{-1}) \\
&\geqslant H_0(\beta^*, G^{*-1} - d).
\end{aligned}$$

Therefore, all inequalities above are equalities.

In both the cases $\beta^* > \alpha_2$ and $\beta^* \leqslant \alpha_2$, we have $H_0(\beta^*, G^{*-1} - d) = H_0(\beta^*, \tilde{G}^{-1})$. Since $\tilde{G}^{-1}$ is the unique solution to problem (2.79), it must hold that $G^{*-1} - d = \tilde{G}^{-1}$ a.s., and equivalently, $G^{*-1} = G^{-1}_{\beta^*,\lambda_{\beta^*}}$. Furthermore,

$$\sup_{\beta\in[0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d) = \rho^{g_2}\left((G^{*-1}(U) - d)_+\right) = H_0(\beta, G^{-1}_{\beta^*,\lambda_{\beta^*}} - d),$$

i.e., β^* is a maximizer to the problem $\sup_{\beta\in[0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d)$ in (2.46).

(ii) Assume that $\sup_{\beta\in[0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d)$ in (2.46) has a maximizer β^*. Then, by (2.46), we have

$$\sup_{G\in\mathcal{S}_2} \rho^{g_2}\left((X^G - d)_+\right) = \max_{\beta\in[0,\alpha_2]} H_0(\beta, G^{-1}_{\beta,\lambda_\beta} - d) = H_0(\beta, G^{-1}_{\beta^*,\lambda_{\beta^*}} - d).$$

Thus, by (2.40), we have

$$\begin{aligned}\sup_{G\in\mathcal{S}_2}\rho^{g_2}\left((X^G-d)_+\right) &= \sup_{\beta\in[0,1]}\sup_{K^{-1}\in\mathcal{Q}_2^d} H_0(\beta,K^{-1})\\ &= \max_{\beta\in[0,\alpha_2]} H_0(\beta,G^{-1}_{\beta,\lambda_\beta}-d)\\ &= H_0(\beta,G^{-1}_{\beta^*,\lambda_{\beta^*}}-d).\end{aligned}$$

It is easy to check that

$$\rho^{g_2}((G^{-1}_{\beta^*,\lambda_{\beta^*}}(U)-d)_+) = \int_{\beta^*}^{1}\gamma(u)(G^{-1}_{\beta^*,\lambda_{\beta^*}}(u)-d)\,\mathrm{d}u + \int_{G_{\beta^*,\lambda_{\beta^*}}(0)}^{\beta^*}\gamma(u)(G^{-1}_{\beta^*,\lambda_{\beta^*}}(u)-d)\,\mathrm{d}u.$$

Under the condition $G_{\beta^*,\lambda_{\beta^*}}(d-)\leqslant\beta^*\leqslant G_{\beta^*,\lambda_{\beta^*}}(d)$, we know that $G^{-1}_{\beta^*,\lambda_{\beta^*}}(u)-d=0$ for any u between β^* and $G_{\beta^*,\lambda_{\beta^*}}(d)$. Therefore, the second term in the above equation becomes 0. It follows that

$$\begin{aligned}\rho^{g_2}((G^{-1}_{\beta^*,\lambda_{\beta^*}}(U)-d)_+) &= H_0(\beta^*,G^{-1}_{\beta^*,\lambda_{\beta^*}}-d)\\ &= \sup_{\beta\in[0,1]}\sup_{K^{-1}\in\mathcal{Q}_2^d} H_0(\beta,K^{-1})\\ &= \sup_{G\in\mathcal{S}_2}\rho^{g_2}\left((X^G-d)_+\right).\end{aligned}$$

Thus, $G_{\beta^*,\lambda_{\beta^*}}$ is a maximizer to the problem $\sup_{G\in\mathcal{S}_2}\rho^{g_2}\left((X^G-d)_+\right)$.

■

Chapter 3

The worst-case distributions for limited stop-loss functions with model uncertainty

In Chapter 2, we explored the worst-case distributions with respect to both a stop-loss function and a limited loss function. However, a loss function that possesses the features of both types of loss functions, i.e., a general limited stop-loss function, is of great interest as well. Indeed, it corresponds to a reinsurance layer in the industry. As a result, we will try to find the worst-case distribution with respect to a limited stop-loss function. Based on that, we will look at how the change of limit and deductible will influence the worst-case risk measure for both sides. To increase the simplicity and the tractability of the problem, we will adjust the uncertainty set from Chapter 2 by removing the moment constraints on the candidate distributions. Another major change in the assumption is that we consider a convex distortion risk measure in Chapter 3.

3.1 Problem Formulation

Following the assumption in Chapter 1, we assume g is absolutely continuous and concave. By (1.5), the distortion risk measure ρ^g has the following representation

$$\rho^g(Y^G) = \int_0^1 \gamma(u) G^{-1}(u)\, \mathrm{d}u, \tag{3.1}$$

where the weight function $\gamma(u) = \partial_- g(x)|_{x=1-u}$, $0 < u < 1$, satisfies $\int_0^1 \gamma(u) = 1$ and ∂_- denotes the derivative from the left.

In an optimal strategy design problem, the agent is interested in finding the best strategy which optimizes her personal objective function. Mathematically, when the agent's objective is given by a risk measure, the optimization problem can be formulated as $\min_\ell \rho(\ell(X))$, where X is the random position faced by the agent, ℓ is the loss function chosen by the agent to against X, and ρ is the risk measure used to quantify the risk exposure. To solve such optimization problem, it is crucial to know the distribution of X to calculate the value of the risk measure after the transformation of loss function ℓ. Nevertheless, in practice, one is not able to know the distribution of the loss X precisely due to the limitation of knowledge. The agent may propose a reference distribution for X according to her available information, while she is aware of uncertainty thanks to the estimation errors. As a consequence, the agent can use an uncertainty set $\mathcal{S}$ to cover all candidate distributions of X. From the agent's concern about the risk management, her has special interests in the worst-case when the risk measure value achieve the supremum in the uncertainty set $\mathcal{S}$. Then, the agent's optimization problem with model uncertainty can be formulated as $\min_\ell \sup_{F\in\mathcal{S}} \rho\left(\ell(X^F)\right)$. The inner maximization problem represents the agent's conservative concern about the distribution of X in the worst-case scenario, while the outer minimization problem is used to determine the best strategy for the agent. Therefore, the optimal strategy $\ell^* := \arg\min_\ell \sup_{F\in\mathcal{S}} \rho(\ell(X^F))$, if exists, can minimize the agent's risk exposure in the worst-case.

We should note that the worst-case distribution solved from the inner problem $\sup_{F\in\mathcal{S}} \rho(\ell(X^F))$ depends on the loss transform ℓ. Indeed, the worst-case distributions associated with different loss transforms are not necessary to be the same. Therefore, it is crucial to characterize the worst-case distribution associated with a given loss function, i.e., solve the worst-case problem

$$\sup_{F\in\mathcal{S}} \rho(\ell(X^F)) \tag{3.2}$$

for a given ℓ. Motivated by insurance markets, we consider the family of limited stop-loss functions in the presenting chapter. Mathematically, define the set of all strategies by

$$\mathcal{L} := \left\{\ell : \mathbb{R} \to \mathbb{R} \text{ satisfying } \ell(x) = \min\left\{d + (x-d)_+, M\right\},\ \text{where} -\infty \leqslant d < M \leqslant \infty\right\}. \tag{3.3}$$

When $d > -\infty$, and $M = \infty$, the function $\ell(x) = d + (x-d)_+$ is determined by a *stop-loss function* $(x-d)_+$, in which the agent's risk position is the loss above the deductible d only, i.e., the risk is truncated from below. On the other hand, when $M < \infty$,

$$\ell(x) = \begin{cases} d, & x \leqslant d \\ x, & d < x < M \\ M, & x \geqslant M \end{cases} \longrightarrow x \wedge M = \begin{cases} x, & -\infty < x < M \\ M, & x \geqslant M, \end{cases} \qquad \text{as } d \to -\infty.$$

In the limiting case of $d = -\infty$, we define $\ell(x) = x \wedge M$ which is called *limited-loss function.* It means that the agent has limited responsibility up to the maximal amount M, i.e., the risk is truncated from above. The stop-loss and limited loss functions are not only two especial transforms but also building blocks of the set $\mathcal{L}$. In section 3.2, we first consider stop-loss and limited loss functions, and then extend results to an arbitrary transform in $\mathcal{L}$.

To characterize the worst-case distribution in the problem (3.2), it is also crucial to identify the uncertainty set. In practice, the agent can use her information about the risk to propose a reference distribution. A candidate distribution should be close to the reference distribution in the sense of certain distance metric. In this chapter, we adopt Wasserstein distance and its definition is given below.

Definition 3.1.1 (Wasserstein distance with order k) *Let $k \geqslant 1$. For two distributions F and G, the Wasserstein metric of order k is given by*

$$W_k(F, G) = W_k(F^{-1}, G^{-1}) = \left(\int_0^1 \left| F^{-1}(x) - G^{-1}(x) \right|^k \mathrm{d}x \right)^{1/k},$$

where F^{-1} and G^{-1} are the quantile functions of F and G respectively.

The above definition generalizes that of Wasserstein distance in Chapter 2 by allowing a larger range of the order k. In the presenting chapter, we sometimes abuse the notation of Wasserstein distance by applying it to quantile functions instead of distribution functions. There is no difference in the meaning between the two presentations. We assume that the agent has a reference distribution $\hat{F}$, and define the agent's the uncertainty set of distributions as

$$\mathcal{S} = \left\{ F : W_k(F, \hat{F}) \leqslant \varepsilon \right\} \qquad \text{for } k \geqslant 1 \ \text{ and } \ \varepsilon \geqslant 0. \tag{3.4}$$

Symmetrically, we define the uncertainty set of quantile functions as

$$\mathcal{Q} = \{F^{-1} : F \in \mathcal{S}\} = \left\{ F^{-1} : W_k(F^{-1}, \hat{F}^{-1}) \leqslant \varepsilon \right\} \qquad \text{for } k \geqslant 1 \ \text{ and } \ \varepsilon \geqslant 0. \tag{3.5}$$

Assumption 3.1.1 *Put $\bar{k} = (1 - 1/k)^{-1}$ for $k > 1$ and $\bar{k} = \infty$ for $k = 1$. Note that $\bar{k} = 1$ when $k = \infty$. Assume $\|\gamma\|_k$ and $\|\gamma\|_{\bar{k}}$ are well-defined whenever appear.*

Now, we are ready to formulate the agent's worst-case problem considered in this chapter. We assume that the agent adopts a convex distortion risk measure ρ with a weight function γ defined in (1.5), the uncertainty set $\mathcal{S}$ defined in (3.4), and a limited stop-loss function $\ell \in \mathcal{L}$. Then the agent's risk measure value in the worst-case scenario is

$$[\rho]_{\varepsilon}^{k}(X^{\hat{F}}; \ell) := \sup_{F \in \mathcal{S}} \rho\left(\ell(X^F)\right) = \sup_{F \in \mathcal{S}} \left\{ \rho\left(\min\left\{d + (X^F - d)_+, M\right\}\right) : W_k(\hat{F}, F) \leqslant \varepsilon \right\}. \tag{3.6}$$

If $\ell(x) = x$, in which case we can set $d = -\infty$ and $M = \infty$, then (3.6) is the robust version of a coherent distortion risk measure ρ via the Wasserstein metric, which is solved in [Liu et al., 2022]. For $\ell \in \mathcal{L}$ with $-\infty < d$ and/or $M < \infty$, it is a non-linear transformation, and moreover, it might not be neither concave nor convex. Therefore, results in [Liu et al., 2022] cannot be applied directly.

3.2 Worst-case distribution

3.2.1 Stop-loss function

In this section, we first consider the worst-case distribution in the problem (3.6) with respect to a transform $d + (x - d)_+$ with $d > \text{ess-inf} X^{\hat{F}} \geqslant -\infty$ given. Since ρ is cash-invariant, solving the problem (3.6) is equivalent to solving the following problem

$$[\rho]_{\varepsilon}^{k}(X^{\hat{F}}; \ell_1) = \sup \left\{ \rho\left((X^F - d)_+\right) : W_k(\hat{F}, F) \leqslant \varepsilon \right\}, \qquad d > \text{ess-inf} X^{\hat{F}}, \tag{3.7}$$

where $\ell_1(x) \triangleq (x - d)_+$ is a stop-loss function.

For any given $\beta \in [0, 1]$, define function

$$\gamma_{1,\beta}(u) := \gamma(u) \cdot \mathbb{I}_{[\beta,1]}(u), \qquad 0 < u < 1.$$

Since the distortion function g is concave, its weight function γ is non-negative and non-decreasing on $(0, 1)$ with $\|\gamma\|_1 = 1$. Therefore, there exists $\delta \in (0, 1)$ such that $\gamma > 0$ holds on the interval $(\delta, 1)$. Consequently, $\|\gamma_{1,\beta}\|_1 \geqslant \int_{\beta\vee\delta}^1 \gamma(u)\,\mathrm{d}u > 0$ for all $\beta < 1$. By applying the same argument in Lemma 2.3.1 in Chapter 2, we can write

$$\begin{aligned} [\rho]_\varepsilon^k(X^{\hat{F}}; \ell_1) = \sup_{G\in\mathcal{S}} \rho\left((X^G - d)_+\right) &= \sup_{G\in\mathcal{S}} \max_{\beta\in[0,1]} \int_\beta^1 \gamma(u)\left(G^{-1}(u) - d\right)\,\mathrm{d}u \\ &= \sup_{\beta\in[0,1]} \sup_{G\in\mathcal{S}} \int_\beta^1 \gamma(u)\left(G^{-1}(u) - d\right)\,\mathrm{d}u \qquad (3.8) \\ &= \sup_{\beta\in[0,1]} \sup_{G^{-1}\in\mathcal{Q}} \int_0^1 \gamma_{1,\beta}(u)\left(G^{-1}(u) - d\right)\,\mathrm{d}u. \end{aligned}$$

Intuitively, in the above expression, we can first fix a $\beta \in [0, 1]$ and solve the inner maximization problem, in which the integral $\int_0^1 \gamma_{1,\beta}(u)\,(G^{-1}(u) - d)\,\mathrm{d}u$ can viewed as a new distortion risk measure with a concave distortion function. Then Proposition 3 of [Liu et al., 2022] can be applied to find a solution to the inner problem.

Theorem 3.2.1 *Suppose Assumption 3.1.1 holds and* $\ell_1(x) = (x-d)_+$ *with* $d > \text{ess-inf}X^{\hat{F}}$. *Then we have*

$$[\rho]_\varepsilon^k(X^{\hat{F}}; \ell_1) = \max_{\beta\in[0,1]} \left\{\int_0^1 \gamma_{1,\beta}(u)\hat{F}^{-1}(u)\,\mathrm{d}u + \varepsilon\|\gamma_{1,\beta}\|_{\bar{k}} - d\|\gamma_{1,\beta}\|_1\right\}. \qquad (3.9)$$

Proof. To express the right hand side of (3.9) in a more simply way, we write

$$H(\beta) \triangleq \int_0^1 \gamma_{1,\beta}(u)\hat{F}^{-1}(u)\,\mathrm{d}u + \varepsilon\|\gamma_{1,\beta}\|_{\bar{k}} - d\|\gamma_{1,\beta}\|_1, \quad 0 \leqslant \beta \leqslant 1. \qquad (3.10)$$

First note that shifting two quantile functions by a same constant does not change their Wasserstein distance. Therefore, define $\mathcal{Q}_d \triangleq \mathcal{Q} - d$ and write (3.8) as

$$[\rho]_\varepsilon^k(X^{\hat{F}}; \ell_1) = \sup_{\beta\in[0,1]} \sup_{G^{-1}\in\mathcal{Q}_d} \int_0^1 \gamma_{1,\beta}(u)G^{-1}(u)\,\mathrm{d}u. \qquad (3.11)$$

If $\beta = 1$, then $\gamma_{1,\beta} = 0$ and $\sup_{G^{-1}\in\mathcal{Q}_d} \int_0^1 \gamma_{1,\beta}(u)G^{-1}(u)\,\mathrm{d}u = 0 = H(1)$ for all G^{-1}. If $\beta \in [0, 1)$, $\gamma_{1,\beta}$ is not the constant zero, and $\gamma_{1,\beta} \leqslant \gamma$ implies $\|\gamma_{1,\beta}\|_1$, $\|\gamma_{1,\beta}\|_{\bar{k}}$ and $\|\gamma_{1,\beta}\|_k$ are well-defined, where $\bar{k} = (1 - 1/k)^{-1}$. Note that, the function $\frac{\gamma_{1,\beta}(u)}{\|\gamma_{1,\beta}\|_1} \geqslant 0$ is non-decreasing in $u \in (0, 1)$ and $\int_0^1 \frac{\gamma_{1,\beta}(u)}{\|\gamma_{1,\beta}\|_1}\,\mathrm{d}u = 1$. Therefore, function defined via

$$g_{1,\beta}(q) := 1 - \int_0^{1-q} \frac{\gamma_{1,\beta}(u)}{\|\gamma_{1,\beta}\|_1}\,\mathrm{d}u, \quad q \in (0, 1), \qquad (3.12)$$

is non-decreasing and concave with $g_{1,\beta}(0) = 0$ and $g_{1,\beta}(1) = 1$. That is, $g_{1,\beta}$ is a concave distortion function, whose weight function is $\frac{\gamma_{1,\beta}}{\|\gamma_{1,\beta}\|_1}$. Denote $\rho_{1,\beta}$ to be the convex distortion risk measure induced by $g_{1,\beta}$. Then, for any $\beta \in [0,1)$ and G^{-1}, we have

$$\int_0^1 \gamma_{1,\beta}(u) G^{-1}(u)\,\mathrm{d}u = \|\gamma_{1,\beta}\|_1 \int_0^1 \frac{\gamma_{1,\beta}(u)}{\|\gamma_{1,\beta}\|_1} G^{-1}(u)\,\mathrm{d}u = \|\gamma_{1,\beta}\|_1 \cdot \rho_{1,\beta}(X^G).$$

For a fixed $\beta \in [0,1)$, we apply Proposition 4 in [Liu et al., 2022] to conclude that

$$\sup_{G^{-1}\in\mathcal{Q}_d} \rho_{1,\beta}(X^G) = [\rho_{1,\beta}]^k_\varepsilon(X^{\hat{F}} - d) = \rho_{1,\beta}(X^{\hat{F}}) - d + \varepsilon \left\| \frac{\gamma_{1,\beta}}{\|\gamma_{1,\beta}\|_1} \right\|_{\bar{k}}.$$

Furthermore we have

$$\begin{aligned}
\sup_{G^{-1}\in\mathcal{Q}_d} \int_0^1 \gamma_{1,\beta}(u) G^{-1}(u)\,\mathrm{d}u &= \|\gamma_{1,\beta}\|_1 \cdot \sup_{G^{-1}\in\mathcal{Q}_d} \rho_{1,\beta}(X^G) \\
&= \|\gamma_{1,\beta}\|_1 \left(\rho_{1,\beta}(X^{\hat{F}}) - d + \varepsilon \cdot \frac{\|\gamma_{1,\beta}\|_{\bar{k}}}{\|\gamma_{1,\beta}\|_1} \right) \\
&= \|\gamma_{1,\beta}\|_1 \left(\int_0^1 \frac{\gamma_{1,\beta}(u)}{\|\gamma_{1,\beta}\|_1} \hat{F}^{-1}(u)\,\mathrm{d}u - d + \varepsilon \cdot \frac{\|\gamma_{1,\beta}\|_{\bar{k}}}{\|\gamma_{1,\beta}\|_1} \right) = H(\beta).
\end{aligned}$$

In short, we verify that $H(\beta) = \sup_{G^{-1}\in\mathcal{Q}_d} \int_0^1 \gamma_{1,\beta}(u) G^{-1}(u)\,\mathrm{d}u$ for all $0 \leqslant \beta \leqslant 1$. Together with (3.11), we have $[\rho]^k_\varepsilon(X^{\hat{F}}; \ell_1) = \sup_{\beta\in[0,1]} H(\beta)$. Since $|\gamma(u)| < \infty$ and $|\hat{F}(u)| < \infty$ for all $u \in (0,1)$, all integrals $\int_0^1 \gamma_{1,\beta}(u)\hat{F}^{-1}(u)\,\mathrm{d}u = \int_\beta^1 \gamma(u)\hat{F}^{-1}(u)\,\mathrm{d}u$, $\|\gamma_{1,\beta}\|_{\bar{k}} = \left(\int_\beta^1 \gamma(u)^{\bar{k}}\,\mathrm{d}u\right)^{1/\bar{k}}$ and $\|\gamma_{1,\beta}\|_1 = \int_\beta^1 \gamma(u)\,\mathrm{d}u$ are continuous in β. Therefore, $H(\beta)$ is a continuous function. Hence, the supremum of $H(\beta)$ can be achieved on the compact set $[0,1]$. The expression (3.9) holds as a consequence. ■

Given a distortion risk measure ρ with distortion function g and its derivative γ, we can easily calculate all quantities in (3.9), and then determines the worst-case risk measure value.

Remark 3.2.1 *As $d \downarrow \text{ess-inf} X^{\hat{F}}$, the function $d + (x-d)^+$ converges to x pointwisely for all $x > \text{ess-inf} X^{\hat{F}}$. We consider a simpler case in detail when $\text{ess-inf} X^{\hat{F}} > -\infty$. Take $d = \text{ess-inf} X^{\hat{F}}$, $\ell_1(x) = (x - \text{ess-inf} X^{\hat{F}})_+$ and $\ell(x) = x$. By Proposition 3 of [Liu et al., 2022], we know $[\rho]^k_\varepsilon(X^{\hat{F}}; \ell) = \rho(X^{\hat{F}}) + \varepsilon \cdot \|\gamma\|_{\bar{k}}$.*

Meanwhile, we note that $\text{ess-inf}X^{\hat{F}} + \ell_1(X) = X$*. To determine* $[\rho]_\varepsilon^k(X^{\hat{F}}; \ell_1)$*, we obtain from* (3.10) *that*

$$H'(\beta) = \gamma(\beta)\left(\text{ess-inf}X^{\hat{F}} - \hat{F}^{-1}(\beta)\right) - \bar{k}\|\gamma_{1,\beta}\|_{\bar{k}}^{(\bar{k}-1)/\bar{k}}\gamma(\beta) \leqslant 0$$

for all $\beta > 0$*. It says that* $H(\beta)$ *is decreasing in* β*, and therefore, the maximal value of* $H(\beta)$ *is achieved at* $\beta = 0$*. Consequently,* $[\rho]_\varepsilon^k(X^{\hat{F}}; \ell_1) = \int_0^1 \gamma(u)\hat{F}^{-1}(u)\,\mathrm{d}u + \varepsilon \cdot \|\gamma\|_{\bar{k}} - \text{ess-inf}X^{\hat{F}} \cdot \|\gamma\|_1$ *since* $\|\gamma_{1,0}\|_1 = \|\gamma\|_1 = 1$*. Therefore,*

$$[\rho]_\varepsilon^k(X^{\hat{F}}; \text{ess-inf}X^{\hat{F}} + \ell_1) = \rho(X^{\hat{F}}) + \varepsilon \cdot \|\gamma\|_{\bar{k}} = [\rho]_\varepsilon^k(X^{\hat{F}}; \ell),$$

i.e., Theorem 3.2.1 and Proposition 3 of [Liu et al., 2022] give the same result. For the case when $\text{ess-inf}X^{\hat{F}} = -\infty$*, convergent argument can be applied to verify from* (3.9) *that* $[\rho]_\varepsilon^k(X^{\hat{F}}; d + (x-d)^+) \to \rho(X^{\hat{F}}) + \varepsilon \cdot \|\gamma_{1,\beta}\|_{\bar{k}}$ *as* $d \to -\infty$*. The detail argument is omitted here.*

Example 7 (Worst-case Tail Value-at-Risk) *In this example, we take* $\rho = \mathrm{TVaR}_\alpha$ *with* $0 < \alpha < 1$*, i.e.,*

$$\rho(X^F) = \mathrm{TVaR}_\alpha(X^F) = \frac{1}{1-\alpha}\int_\alpha^1 F^{-1}(u)\,\mathrm{d}u.$$

The problem (3.7) *becomes*

$$[\rho]_\varepsilon^k(X^{\hat{F}}; \ell_1) := \sup_{F \in \mathcal{S}} \mathrm{TVaR}_\alpha\left((X^F - d)_+\right) = \sup\left\{\mathrm{TVaR}_\alpha\left((X^F - d)_+\right) : W_k(\hat{F}, F) \leqslant \varepsilon\right\}. \tag{3.13}$$

It is well know that TVaR *has distortion function* $g(t) = \min\{\frac{1}{1-\alpha}t, 1\}$*,* $0 < t < 1$*, and its weight function is* $\gamma(t) = \frac{1}{1-\alpha}\mathbb{I}_{[\alpha,1]}(t)$*,* $0 < t < 1$*. Consequently, for any* $0 < \beta < 1$*, we have*

$$\gamma_{1,\beta}(t) = \frac{1}{1-\alpha}\mathbb{I}_{[\alpha\vee\beta,1]}(t),\ \ 0 < t < 1, \quad \textit{with } \|\gamma_{1,\beta}\|_1 = \frac{1-\alpha\vee\beta}{1-\alpha} \textit{ and } \|\gamma_{1,\beta}\|_{\bar{k}} = \frac{(1-\alpha\vee\beta)^{1/\bar{k}}}{1-\alpha}.$$

Then (3.9) *can be reduced to*

$$
\begin{aligned}
[\mathrm{TVaR}_\alpha]_\varepsilon^k(X^{\hat F};\ell_1) &= \max_{\beta\in[0,1]}\left\{\frac{1}{1-\alpha}\int_{\alpha\vee\beta}^1 \hat F^{-1}(u)\,\mathrm{d}u + \varepsilon\cdot\frac{(1-\alpha\vee\beta)^{1/\bar k}}{1-\alpha} - d\cdot\frac{1-\alpha\vee\beta}{1-\alpha}\right\}\\
&= \max_{\beta\in[0,1]}\left\{\frac{1}{1-\alpha}\int_{\alpha\vee\beta}^1 \left(\hat F^{-1}(u)-d\right)\,\mathrm{d}u + \varepsilon\cdot\frac{(1-\alpha\vee\beta)^{1/\bar k}}{1-\alpha}\right\}\\
&= \max_{\beta\in[0,1]}\left\{\frac{1-\alpha\vee\beta}{1-\alpha}\frac{1}{1-\alpha\vee\beta}\int_{\alpha\vee\beta}^1 \left(\hat F^{-1}(u)-d\right)\,\mathrm{d}u + \varepsilon\cdot\frac{(1-\alpha\vee\beta)^{1/\bar k}}{1-\alpha}\right\}\\
&= \frac{1}{1-\alpha}\max_{\beta\in[\alpha,1]}\left\{(1-\beta)\left(\mathrm{TVaR}_\beta(X^{\hat F})-d\right)+\varepsilon(1-\beta)^{1/\bar k}\right\}. \qquad (3.14)
\end{aligned}
$$

In particular, if we take $\alpha=0$*, then* $\mathrm{TVaR}_0=\mathbb{E}$*. The expression* (3.14) *implies that*

$$
[\mathbb{E}]_\varepsilon^k(X^{\hat F};\ell_1) = \max_{\beta\in[0,1]}\left\{(1-\beta)\left(\mathrm{TVaR}_\beta(X^{\hat F})-d\right)+\varepsilon(1-\beta)^{1-1/k}\right\}, \qquad (3.15)
$$

which covers the result of Proposition 2 in [Guan et al., 2022].

Example 8 (Wang's premium in the worst-case scenario) *In this example, we let* $\hat F(x)=1-(\frac{12}{x+12})^4, x\geqslant 0$*,* $\varepsilon=2$*,* $k=2$ *and look for the worst-case distribution with respect to a stop-loss function* $\ell_1(x)=(x-d)_+$*. We adopt* ρ^g*, a Wang's risk measure, to quantify* $\ell_1(X)$*, with* $g(u)=\Phi(\Phi^{-1}(u)+0.5), 0\leqslant u\leqslant 1$*. The definition of* ρ^g *follows Definition 1.2.8. The worst-case quantiles with different deductible* d *are plotted in Figure 3.1:*

3.2.2 Limited loss function

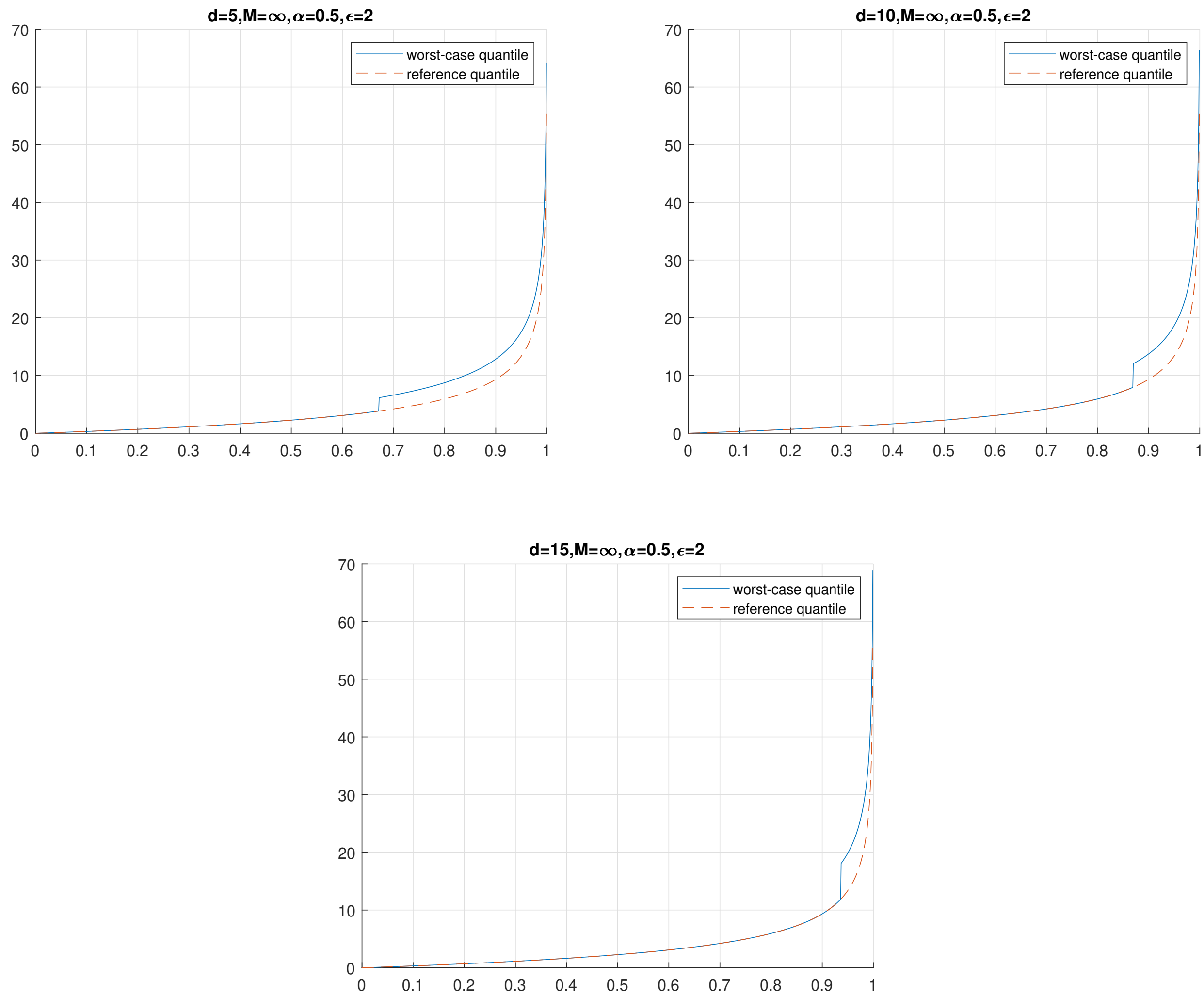
In this section, we consider the worst-case distribution in the problem (3.6) with a limited-loss function $\ell_2(x)=x\wedge M$ where $M<\text{ess-sup}X^{\hat F}$. We re-write the problem (3.6) as

$$
[\rho]_\varepsilon^k(X^{\hat F};\ell_2) = \sup\left\{\rho\left(X^F\wedge M\right) : W_k(\hat F,F)\leqslant\varepsilon\right\}. \qquad (3.16)
$$

To proceed, we first denote $q_1:=\hat F(M)$ and

$$
\begin{aligned}
q_0^k &:= \inf\left\{q\geqslant 0 : \int_q^{q_1}\left|M-\hat F^{-1}(u)\right|^k\,\mathrm{d}u\leqslant\varepsilon^k\right\} \qquad (3.17)\\
&= \inf\left\{q\geqslant 0 : \int_q^{1}\left|M-\hat F^{-1}(u)\wedge M\right|^k\,\mathrm{d}u\leqslant\varepsilon^k\right\}.
\end{aligned}
$$

Figure 3.1: Worst-case distributions with stop-loss transformation.



It worth pointing out that, if $q_0^k = 0$, then we have

$$\left(W_k(M \vee \hat{F}^{-1}, \hat{F}^{-1}\right)^k = \int_0^1 \left|M \vee \hat{F}^{-1}(u) - \hat{F}^{-1}(u)\right|^k \mathrm{d}u = \int_0^{q_1} \left|M - \hat{F}^{-1}(u)\right|^k \mathrm{d}u \leqslant \varepsilon^k,$$

i.e., $M \vee \hat{F}^{-1} \in \mathcal{Q}$. It is easy to see that $\rho\left((M \vee \hat{F}^{-1}) \wedge M\right) = \rho(M) = M$. Meanwhile, $F^{-1} \wedge M \leqslant M$ for any quantile function F^{-1}. By the monotonicity of ρ, we know that $\rho(X^F \wedge M) \leqslant M$ for any $F \in \mathcal{S}$. Therefore, $q_0^k = 0$ implies a trivial case in which the worst-case risk measure value is M and is achieved at a worst-case distribution $M \vee \hat{F}^{-1} \in \mathcal{S}$. If $q_0^k > 0$, then $M \vee \hat{F}^{-1} \notin \mathcal{S}$. In this case, since $\hat{F}^{-1}(u)$ is finite for any $0 < u < 1$, the integral $\int_q^1 \left|M - \hat{F}^{-1}(u) \wedge M\right|^k \mathrm{d}u$ is continuous in q. Therefore, q_0^k satisfies the equation

$$\int_q^{q_1} \left|M - \hat{F}^{-1}(u)\right|^k \mathrm{d}u = \int_q^1 \left|M - \hat{F}^{-1}(u) \wedge M\right|^k \mathrm{d}u = \varepsilon^k. \tag{3.18}$$

The non-trivial case of the problem (3.16) is more challenging compared to the problem (3.7) because $\ell_2(x) = x \wedge M$ is a concave function while the distortion risk measure ρ is convex. To see this mathematically, we apply Lemma 2.2.1 and Lemma 2.2.2 in Chapter 2 to obtain

$$\sup_{F \in \mathcal{S}} \rho(X^F \wedge M) = M + \sup_{F \in \mathcal{S}} \int_0^{F(d)} \gamma(u) \left(F^{-1}(u) - M\right) \mathrm{d}u = M + \sup_{F \in \mathcal{S}} \min_{\beta \in [0,1]} L(\beta, F^{-1}), \tag{3.19}$$

where

$$L(\beta, F^{-1}) \triangleq \int_0^1 \gamma_{2,\beta}(u) \left(F^{-1}(u) - M\right) \mathrm{d}u$$

and

$$\gamma_{2,\beta} \triangleq \gamma \cdot \mathbb{I}_{[0,\beta]}.$$

In (3.19), the problem (3.16) is expressed by a "sup-inf" problem, which is not necessarily equivalent to its "inf-sup" problem. In other words, the step of exchanging the inner and outer optimization problems in (3.8) may not hold true to (3.19). Furthermore, we should note that the weight function $\gamma_{2,\beta}$ in (3.19) is not a non-decreasing function. Therefore, the distortion risk measure induced by $\gamma_{2,\beta}$ is not coherent, and the argument for solving the problem (3.7) cannot be applied to the problem (3.16).

In this section, we solve the problem (3.16) for two cases when $k = 1$ and $k = 2$. Then we provide a partial characterization of the worst-case distribution for other values of k. To proceed, we introduce sets

$$\mathcal{A}_1 := \left\{ F^{-1} : W_k(F^{-1}, \hat{F}^{-1}) \leqslant \varepsilon, \quad \hat{F}^{-1} \wedge M \leqslant F^{-1} \right\}$$
$$\mathcal{A}_2 := \left\{ F^{-1} : W_k(F^{-1}, \hat{F}^{-1} \wedge M) \leqslant \varepsilon, \quad \hat{F}^{-1} \wedge M \leqslant F^{-1} \leqslant M \right\}.$$

Lemma 3.2.2 *Let U be a uniform random variable on $(0,1)$. The following equations hold:*

$$[\rho]_{\varepsilon}^{k}(X^{\hat{F}}; \ell_2) = \sup_{F^{-1} \in \mathcal{A}_i} \rho\left(F^{-1}(U) \wedge M\right), \quad i = 1, 2. \tag{3.20}$$

Proof. For any quantile function F^{-1} and its distribution function F, we have $X^F \stackrel{\mathrm{d}}{=} F^{-1}(U)$, $X^F \wedge M \stackrel{\mathrm{d}}{=} F^{-1}(U) \wedge M$, and we can write

$$\rho(X^F \wedge M) = \int_0^1 \gamma(u) \left(F^{-1}(u) \wedge M\right) \mathrm{d}u. \tag{3.21}$$

From the proof of Proposition 3 of [Liu et al., 2022], we have

$$[\rho]_{\varepsilon}^{k}(X^{\hat{F}}; \ell_2) = \sup \left\{ \rho(X^F \wedge M) : W_k(F, \hat{F}) \leqslant \varepsilon \text{ and } \hat{F}^{-1} \leqslant F^{-1} \right\}. \tag{3.22}$$

For any $F \in \mathcal{S}$ satisfying $\hat{F}^{-1} \leqslant F^{-1}$, we can check that $\hat{F}^{-1} \wedge M \leqslant F^{-1} \wedge M \leqslant M$, and

$$\begin{aligned} \left| F^{-1}(u) \wedge M - \hat{F}^{-1}(u) \wedge M \right| &= \begin{cases} \left| F^{-1}(u) - \hat{F}^{-1}(u) \right|; & \text{if } \hat{F}^{-1}(u) \leqslant F^{-1}(u) \leqslant M \\ \left| F^{-1}(u) - M \right|; & \text{if } \hat{F}^{-1}(u) \leqslant M \leqslant F^{-1}(u) \\ 0, & o/w. \end{cases} \\ &\leqslant \left| F^{-1}(u) - \hat{F}^{-1}(u) \right|, \qquad u \in (0,1). \end{aligned}$$

It follows that $W_k(F^{-1} \wedge M, \hat{F}^{-1} \wedge M) \leqslant W_k(F^{-1}, \hat{F}^{-1}) \leqslant \varepsilon$. Thus, $F^{-1} \wedge M \in \mathcal{A}_2$. Together with (3.21) and (3.22), we have

$$[\rho]_{\varepsilon}^{k}(X^{\hat{F}}; \ell_2) \leqslant \sup \left\{ \int_0^1 \gamma(u) F^{-1}(u) \, \mathrm{d}u : F \in \mathcal{A}_2 \right\}. \tag{3.23}$$

For any $F^{-1} \in \mathcal{A}_2$, we can define

$$\tilde{F}^{-1}(u) = \max\{F^{-1}(u), \hat{F}^{-1}(u)\} = \begin{cases} F^{-1}(u), & \text{for } u \leqslant q_1, \text{ i.e., } \hat{F}^{-1}(u) \leqslant M \\ \hat{F}^{-1}(u), & \text{for } u > q_1, \text{ i.e., } M < \hat{F}^{-1}(u). \end{cases}$$

It is easy to see that $\tilde{F}^{-1} \geqslant \hat{F}^{-1} \geqslant \hat{F}^{-1} \wedge M$, $\tilde{F}^{-1} \wedge M = F^{-1} \wedge M = F^{-1}$, and

$$\begin{aligned} \left(W_k(\tilde{F}^{-1}, \hat{F}^{-1})\right)^k = \int_0^{q_1} \left|\tilde{F}^{-1}(u) - \hat{F}^{-1}(u)\right|^k \mathrm{d}u &= \int_0^{q_1} \left|F^{-1}(u) - \hat{F}^{-1}(u) \wedge M\right|^k \mathrm{d}u \\ &\leqslant \int_0^{1} \left|F^{-1}(u) - \hat{F}^{-1}(u) \wedge M\right|^k \mathrm{d}u \leqslant \varepsilon^k, \end{aligned}$$

where the last inequality is because $F^{-1} \in \mathcal{A}_2$. Therefore $\tilde{F}^{-1} \in \mathcal{A}_1$ and $\rho(X^{\tilde{F}} \wedge M) = \rho(X^F)$. It implies that

$$\sup\left\{\int_0^1 \gamma(u) F^{-1}(u)\,\mathrm{d}u : F \in \mathcal{A}_2\right\} \leqslant \sup\left\{\int_0^1 \gamma(u)\left(\tilde{F}^{-1}(u) \wedge M\right)\mathrm{d}u : \tilde{F} \in \mathcal{A}_1\right\}. \tag{3.24}$$

Obviously, $\mathcal{A}_1 \subset \{F^{-1} : F \in \mathcal{S},\, \hat{F}^{-1} \leqslant F^{-1}\}$. Together with (3.22), we have

$$[\rho]_{\varepsilon}^{k}(X^{\hat{F}}; \ell_2) \geqslant \sup\left\{\int_0^1 \gamma(u)\left(F^{-1}(u) \wedge M\right)\mathrm{d}u : F^{-1} \in \mathcal{A}_1\right\}.$$

The above inequality, together with (3.23) and (3.24), imply that (3.20) is achieved as desired. ■

We first apply Lemma 3.2.2 to characterize the worst-case distribution when $k = 1$. Later, Lemma 3.2.2 will help us to partially determine the worst-case distribution for general cases.

Theorem 3.2.3 *Assume $k = 1$ and $\ell_2(x) = x \wedge M$ with $M <$ ess-sup$X^{\hat{F}}$. Then the worst-case distribution to the problem (3.16) is*

$$(F^*)^{-1}(u) = \begin{cases} \hat{F}^{-1}(u), & 0 \leqslant u \leqslant q_0^1; \\ M, & q_0^1 < u \leqslant q_1; \\ \hat{F}^{-1}(u), & q_1 < u \leqslant 1, \end{cases} \tag{3.25}$$

where $q_1 = \hat{F}(M)$ *and* q_0^1 *is defined in* (3.17)*. Furthermore, the worst-case risk measure value is*

$$[\rho]_\varepsilon^1(X^{\hat{F}}; \ell_2) = \min\left\{\rho(X^{\hat{F}}) + \varepsilon, M\right\} = \begin{cases} \rho(X^{\hat{F}}) + \varepsilon, & \textit{if } q_0^1 > 0 \\ M, & \textit{if } q_0^1 = 0. \end{cases} \tag{3.26}$$

Proof. For $k = 1$, the Wasserstein metric becomes $W_1(F, \hat{F}) = \int_0^1 \left|F^{-1}(u) - \hat{F}^{-1}(u)\right| \mathrm{d}u$. Moreover, for any $F \in \mathcal{S}$ with $\hat{F}^{-1} \leqslant F^{-1}$, we have $W_1(F, \hat{F}) = \int_0^1 F^{-1}(u) - \hat{F}^{-1}(u) \, \mathrm{d}u = \mathbb{E}[X^F] - \mathbb{E}[X^{\hat{F}}]$.

Arbitrarily take $F \in \mathcal{A}_2$, i.e., $\hat{F}^{-1} \wedge M \leqslant F^{-1} \leqslant M$ and

$$\mathbb{E}[F^{-1}(U)] - \mathbb{E}[\hat{F}^{-1}(U) \wedge M] = W_1(F^{-1}, \hat{F}^{-1} \wedge M) \leqslant \varepsilon. \tag{3.27}$$

Define $h(p) \triangleq \int_0^p \hat{F}^{-1}(u) \, \mathrm{d}u + M(1-p)$ as a function of $p \in [0,1]$. Since $\hat{F}^{-1}(u)$ is bounded for all $u \in (0,1)$, the function $h(p)$ is continuous in p. With $q_1 = \hat{F}(M)$, we have

$$h(q_1) = \int_0^{q_1} \hat{F}^{-1}(u) \, \mathrm{d}u + M(1-q_1) = \mathbb{E}[\hat{F}^{-1} \wedge M] \leqslant \mathbb{E}[F^{-1}(U)], \tag{3.28}$$

where the second equality is because, for any $0 < u < 1$ and quantile function G^{-1}, $u \leqslant G(x)$ if and only if $G^{-1}(u) \leqslant x$. On the other hand, if q_0^1 defined in (3.17) is zero, then $h(0) = M \geqslant \mathbb{E}[F^{-1}(U)]$. If $q_0^1 > 0$, then we have $\hat{F}^{-1}(u) \leqslant M$ for $0 < u \leqslant q_0^1 \leqslant q_1$, and

$$\begin{aligned} h(q_0^1) &= \int_0^{q_0^1} \hat{F}^{-1}(u) \, \mathrm{d}u + M(1 - q_0^1) \\ &= \int_{q_0^1}^1 \left(M - \hat{F}^{-1}(u) \wedge M\right) \mathrm{d}u + \int_0^{q_0^1} \hat{F}^{-1}(u) \, \mathrm{d}u + \int_{q_0^1}^1 \hat{F}^{-1}(u) \wedge M \, \mathrm{d}u \\ &= \int_{q_0^1}^1 \left(M - \hat{F}^{-1}(u) \wedge M\right) \mathrm{d}u + \int_0^1 \hat{F}^{-1}(u) \wedge M \, \mathrm{d}u \\ &= \varepsilon + \mathbb{E}[\hat{F}^{-1}(U) \wedge M], \end{aligned}$$

where the third equality comes from (3.18) with $k = 1$. From (3.27), we have $h(q_0^1) \geqslant \mathbb{E}[F^{-1}(U)]$. Together with (3.28) and the continuity of $h(p)$, there exists $p_0 \in [q_0^1, q_1]$ such that $\mathbb{E}[F^{-1}(U)] = h(p_0) = \int_0^{p_0} \hat{F}^{-1}(u) \, \mathrm{d}u + M(1 - p_0)$. Then define

$$G^{-1}(u) = \begin{cases} \hat{F}^{-1}(u), & 0 \leqslant u \leqslant p_0; \\ M, & p_0 < u \leqslant 1 \end{cases} \tag{3.29}$$

satisfying $\mathbb{E}[G^{-1}(U)] = h(p_0) = \mathbb{E}[F^{-1}(U)]$. Since $p_0 \leqslant q_1$, we have $\hat{F}^{-1}(u) \leqslant M$ for $0 \leqslant u \leqslant p_0$, and futhermore, $\hat{F}^{-1}(u) = \hat{F}^{-1}(u) \wedge M \leqslant F^{-1}(u)$ for all $0 < u < p_0$. Therefore, $G^{-1}(u) = \hat{F}^{-1}(u) \leqslant F^{-1}(u)$ for $0 \leqslant u \leqslant p_0$, and $G^{-1}(u) = M \geqslant F^{-1}(u)$ for $p_0 < u \leqslant 1$. That is, the function G^{-1} up-crosses the function F^{-1}. Together with $\mathbb{E}[G^{-1}(U)] = \mathbb{E}[F^{-1}(U)]$, by Lemma 3 of [Ohlin, 1969], $G^{-1}(U)$ is larger than $F^{-1}(U)$ in the sense of convex order. Since ρ is a coherent distortion risk measure which preserves the convex order, we have $\rho(X^F) = \rho(F^{-1}(U)) \leqslant \rho(G^{-1}(U)) = \rho(X^G)$. Also note that, by the definition of q_0^1 in (3.17), we have $W_1(G^{-1}, \hat{F}^{-1} \wedge M) = \int_{p_0}^1 \left| M - \hat{F}^{-1}(u) \wedge M \right| \mathrm{d}u \leqslant \varepsilon$, i.e., $G \in \mathcal{A}_2$. Since F^{-1} is arbitrarily taken from $\mathcal{A}_2$, we can conclude that

$$\sup \left\{ \rho(X^F) : F \in \mathcal{A}_2 \right\} = \sup \left\{ \rho(X^G) : G \in \mathcal{A}_3 \right\},$$

where

$$\mathcal{A}_3 = \{ G : G^{-1} = \hat{F}^{-1}\mathbb{I}_{[0,p]} + M\mathbb{I}_{(p,1]} \text{ for some } p \in [q_0^1, q_1] \}.$$

Lemma 3.2.2 further implies

$$[\rho]_\varepsilon^1(X^{\hat{F}}; \ell_2) = \sup \left\{ \rho(X^G) : G \in \mathcal{A}_3 \right\}.$$

The set $\mathcal{A}_3$ can be viewed as a set indexed by a single parameter $p \in [q_0^1, q_1]$. As p increases, the associated quantile function $G = \hat{F}^{-1}\mathbb{I}_{[0,p]} + M\mathbb{I}_{(p,1]}$ decreases in the sense of the first stochastic dominance order (FOD). Since ρ preserves FOD, $\rho(X^G)$ decreases as p increases. As a consequence, we can conclude that

$$(G^*)^{-1}(u) = \begin{cases} \hat{F}^{-1}(u), & 0 \leqslant u \leqslant q_0^1; \\ M, & q_0^1 < u \leqslant 1, \end{cases}$$

which is the largest quantile in $\mathcal{A}_3$ in the sense of FOD, maximizes the integral $\int_0^1 \gamma(u) G^{-1}(u)\, \mathrm{d}u$ on $\mathcal{A}_3$. It is easy to see that F^* defined in (3.25) satisfies $(G^*)^{-1}(u) = (F^*)^{-1}(u) \wedge M$ for $0 < u < 1$ and $F^* \in \mathcal{S}$. Therefore $\rho(X^{F^*} \wedge M) = [\rho]_\varepsilon^1(X^{\hat{F}}; \ell_2)$, i.e., F^* is the worst-case distribution to the problem (3.16) with $k = 1$. The equation (3.26) can be verified directly by calculating $\rho(X^{F^*} \wedge M)$ with (3.16). ■

It is interesting to point out two observations of Theorem 3.2.3. First the worst-case distribution (3.25) only depends on the upper limit M and the uncertainty set $\mathcal{S}$, but does not depend on the choice of the coherent distortion risk measure ρ. For the cases when $k \neq 1$, we normally expect that the weight function γ of ρ also plays an important role in characterizing the worst-case distribution, see for example [Bernard et al., 2020b], [Liu et al., 2022] and Chapter 2 of this thesis. When $k = 1$, the Wasserstein distance between two distributions ordered in FOD is simplified to the difference of their means. As a consequence, if a law-invariant risk measure ρ preserves FOD and the convex order, then similar argument in the proof of Theorem 3.2.3 can be applied to obtain the worst-case distribution in (3.25). In this sense, the result in Theorem 3.2.3 can be generalized to non-distortion coherent risk measures.

Second, from (3.26), we have $[\rho]_\varepsilon^1(X^{\hat{F}};\ell_2) - \rho(X^{\hat{F}}) = \varepsilon$ in the non-trivial case, i.e., $q_0^1 > 0$. From Proposition 3 of [Liu et al., 2022], we also have $\sup\{\rho(X^F) : W_1(F,\hat{F}) \leqslant \varepsilon\} = \rho(X^{\hat{F}})+\varepsilon$. Therefore, regardless the choice of ρ, when $M > \rho(X^{\hat{F}})+\varepsilon$, the worst-case distribution should fully utilize the distance tolerance ε between the reference distribution and a candidate distribution in the left tail (before hitting M) such that the difference between the worst-case risk measure value and $\rho(X^{\hat{F}})$ remains to be ε. Mathematically, $[\rho]_\varepsilon^1(X^{\hat{F}};\ell_2) = \rho(X^{\hat{F}}) + \varepsilon$ is a constant for all $M \in [\rho(X^{\hat{F}}) + \varepsilon, \text{ess-sup}X^{\hat{F}}]$.

Next we consider the case when $k = 2$. Before moving on, we need first to introduction the concept of *isotonic projection*. Same as in Chapter 2, let

$$\mathcal{K} = \left\{ k : (0,1) \mapsto \mathbb{R} \,\middle|\, \int_0^1 k(u)^2 \,\mathrm{d}u < \infty,\, k \text{ non-decreasing} \right\}$$

be the space of square-integrable non-decreasing functions on $(0,1)$. Denote the metric projection of a function $f \in L^2(0,1)$ onto the space $\mathcal{K}$ as

$$f^\uparrow = \underset{k\in\mathcal{K}}{\arg\min}\, ||f-k||_2.$$

Theorem 3.2.4 *Let $k = 2$ and Assumption 3.1.1 hold. Assume $q_0^2 > 0$. There exists a worst-case distribution $F^* \in \mathcal{S}$ such that $[\rho]_\varepsilon^2(X^{\hat{F}};\ell_2) = \rho\left(X^{F^*} \wedge M\right)$, and*

$$(F^*)^{-1}(u) = \begin{cases} \hat{F}^{-1}(u) + \lambda^*\gamma(u), & \textit{for } 0 < u \leqslant \theta^*; \\ M, & \textit{for } \theta^* < u \leqslant q_1; \\ \hat{F}^{-1}(u), & \textit{for } q_1 < u < 1, \end{cases} \tag{3.30}$$

where $\lambda^* \geqslant 0$ and $\theta^* \in (0, q_1)$ satisfies $W_2(F^*, \hat{F}) = \varepsilon$. Furthermore, the worst-case risk measure value is

$$[\rho]_\varepsilon^2(X^{\hat{F}};\ell_2) = \rho(X^{\hat{F}}) + \lambda^* \int_0^{\theta^*} \gamma(u)^2 \,\mathrm{d}u + \int_{\theta^*}^{q_1} \left(M - \hat{F}^{-1}(u)\right) \gamma(u) \,\mathrm{d}u. \tag{3.31}$$

Proof. From Lemma 2.2.1 and Lemma 2.2.2 in Chapter 2, we have

$$\rho(X^F \wedge M) = M + \int_0^{F(M)} \gamma(u) \left(F^{-1}(u) - M\right) \mathrm{d}u = M + \min_{\beta \in [0,1]} L(\beta, F^{-1}), \tag{3.32}$$

where $L(\beta, F^{-1}) = \int_0^1 \gamma_{2,\beta}(u) \left(F^{-1}(u) - M\right) \mathrm{d}u$ and $\gamma_{2,\beta} = \gamma \cdot \mathbb{I}_{[0,\beta]}$. Then we have

$$\sup_{F \in \mathcal{S}} \rho(X^F \wedge M) = M + \sup_{F \in \mathcal{S}} \min_{\beta \in [0,1]} L(\beta, F^{-1}) = M + \sup_{F^{-1} \in \mathcal{Q}} \min_{\beta \in [0,1]} L(\beta, F^{-1}). \tag{3.33}$$

We first introduce auxiliary notation used in the rest of proof. Define

$$\beta_0 = \begin{cases} 0, \text{ if } \gamma > 0 \text{ holds on some interval } (0, \delta), 0 < \delta < 1, \\ \sup\{0 < u < 1 : \int_0^u \gamma(t) \, \mathrm{d}t = 0\}, \text{ otherwise.} \end{cases} \tag{3.34}$$

Since γ is non-negative, it is easy to see $\gamma(u) = 0$ for $0 \leqslant u \leqslant \beta_0$. For any $\beta \leqslant \beta_0$, we have $\gamma_{1,\beta} = 0$ and $L(\beta, F^{-1}) = 0$ for all F^{-1}. For any $\beta_0 < \beta < 1$, $\gamma_{2,\beta}$ is not the constant zero and $\|\gamma_{2,\beta}\|_1 = \int_0^\beta \gamma(u) \, \mathrm{d}u > 0$ is well-defined. Furthermore, we can define the following function for any given $\beta_0 < \beta < 1$

$$g_{2,\beta}(x) = 1 - \int_0^{1-x} \frac{\gamma_{2,\beta}(u)}{\|\gamma_{2,\beta}\|_1} \, \mathrm{d}u, \qquad x \in [0,1].$$

Note that $g_{2,\beta}$ is a non-decreasing function with $g_{2,\beta}(0) = 0$ and $g_{2,\beta}(1) = 1$. We can use $g_{2,\beta}$ as a distortion function to induce a distortion risk measure $\rho_{2,\beta}$. It should be pointed out that $\rho_{2,\beta}$ is not coherent since $g_{2,\beta}$ is not a concave distortion function. For $\beta \in (\beta_0, 1)$, the function $L(\beta, F^{-1})$ can be expressed as

$$L(\beta, F^{-1}) = \|\gamma_{2,\beta}\|_1 \int_0^1 \frac{\gamma_{2,\beta}(u)}{\|\gamma_{2,\beta}\|_1} \left(F^{-1}(u) - M\right) \mathrm{d}u = \|\gamma_{2,\beta}\|_1 \cdot \rho_{2,\beta}(F^{-1}(U) - M).$$

With the help of Theorem 2.6.4 and similarly argument in the proof of Theorem 2.2.3 in Chapter 2, we can re-write the sup-min problem in (3.33) as

$$\sup_{F^{-1} \in \mathcal{Q}} \min_{\beta \in [0,1]} L(\beta, F^{-1}) = \min_{\beta \in [0,1]} \max_{F^{-1} \in \mathcal{Q}} L(\beta, F^{-1}). \tag{3.35}$$

Therefore, the worst-case risk measure value can be expressed as

$$\sup_{F \in \mathcal{S}} \rho(X^F \wedge M) = M + \min_{\beta \in [0,1]} \max_{F^{-1} \in \mathcal{Q}} L(\beta, F^{-1}).$$

For any $\beta \leqslant \beta_0$, $L(\beta, F^{-1}) = 0$ for all F^{-1} implies $\max_{F^{-1}\in\mathcal{Q}} L(\beta, F^{-1}) = 0$. For any $\beta_0 < \beta < 1$, we have

$$\max_{F^{-1}\in\mathcal{Q}} L(\beta, F^{-1}) = \|\gamma_{2,\beta}\|_1 \cdot \max_{F\in\mathcal{Q}} \rho_{2,\beta}(F^{-1}(U) - M) = \|\gamma_{2,\beta}\|_1 \left(-M + \max_{F\in\mathcal{Q}} \rho_{2,\beta}(F^{-1}(U)) \right).$$

For simplicity, write $M_\beta := \max_{F\in\mathcal{Q}} \rho_{2,\beta}(F^{-1}(U))$ for a given $\beta_0 < \beta < 1$. Let $\mathcal{M}$ be the set of quantile functions with finite second moments. From Theorem 1 of [Pesenti, 2022], the problem

$$\min_{F\in\mathcal{M}} W_2(F^{-1}, \hat{F}^{-1}) \qquad \text{s.t. } \rho_{2,\beta}(X^F) = M_\beta$$

has a unique solution given by $(F^*_\beta)^{-1} = \left(\hat{F}^{-1} + \lambda_\beta \gamma_{2,\beta}\right)^{\uparrow}$ with $\lambda_\beta \geqslant 0$ such that $\rho_{2,\beta}((F^*_\beta)^{-1}(U)) = M_\beta$. Next, we are going to show $W_2((F^*_\beta)^{-1}, \hat{F}^{-1}) = \varepsilon$.

(i) Suppose $W_2((F^*_\beta)^{-1}, \hat{F}^{-1}) > \varepsilon$. Then there exists δ such that $0 < \delta < W_2((F^*_\beta)^{-1}, \hat{F}^{-1}) - \varepsilon$. Take a sequence $\{F_n^{-1}, n = 1, 2, ...\} \subset \mathcal{Q}$ such that $M_\beta = \lim_{n\to\infty} \rho_{2,\beta}(F_n^{-1}(U))$. For any $n = 1, 2, ...$, $W_2(F_n^{-1}+\delta, \hat{F}^{-1}) \leqslant \delta + W_2(F_n^{-1}, \hat{F}^{-1}) \leqslant \delta + \varepsilon < W_2((F^*_\beta)^{-1}, \hat{F}^{-1})$. Meanwhile, $\rho_{2,\beta}(F_n^{-1}(U) + \delta) = \rho_{2,\beta}(F_n^{-1}(U)) + \delta \to M_\beta + \delta$ as $n \to \infty$. There exists N such that $\rho_{2,\beta}(F_n^{-1}(U) + \delta) > M_\beta$ for all $n \geqslant N$. Meanwhile, we know $W_2(F_n^{-1} + \delta, \hat{F}^{-1}) < W_2((F^*_\beta)^{-1}, \hat{F}^{-1})$ for all $n \geqslant N$. In particular, we take N and the problem

$$\min_{F\in\mathcal{M}} W_2(F^{-1}, \hat{F}^{-1}) \qquad \text{s.t. } \rho_{2,\beta}(X^F) = \rho_{2,\beta}(F_N^{-1}(U) + \delta)$$

again has a unique solution given by $\tilde{F}^{-1} = \left(\hat{F}^{-1} + \tilde{\lambda}\gamma_{2,\beta}\right)^{\uparrow}$ with $\tilde{\lambda} \geqslant 0$ such that $\rho_{2,\beta}(\tilde{F}^{-1}(U)) = \rho_{2,\beta}(F_N^{-1}(U) + \delta)$. Since $\rho_{2,\beta}(\tilde{F}^{-1}(U)) = \rho_{2,\beta}(F_N^{-1}(U) + \delta) > M_\beta = \rho_{2,\beta}((F^*_\beta)^{-1}(U))$, we have $\lambda_\beta < \tilde{\lambda}$. It implies $W_2((F^*_\beta)^{-1}, \hat{F}^{-1}) < W_2(\tilde{F}^{-1}, \hat{F}^{-1}) \leqslant W_2(F_N^{-1}+\delta, \hat{F}^{-1})$, which contradicts with the fact $W_2(F_n^{-1}+\delta, \hat{F}^{-1}) < W_2((F^*_\beta)^{-1}, \hat{F}^{-1})$ for all $n \geqslant N$.

(ii) If $W_2((F^*_\beta)^{-1}, \hat{F}^{-1}) < \varepsilon$, then take δ such that $0 < \delta < \varepsilon - W_2((F^*_\beta)^{-1}, \hat{F}^{-1})$. It is easy to see that

$$W_2((F^*_\beta)^{-1} + \delta, \hat{F}^{-1}) \leqslant W_2((F^*_\beta)^{-1} + \delta, (F^*_\beta)^{-1}) + W_2((F^*_\beta)^{-1}, \hat{F}^{-1}) < \varepsilon,$$

i.e., $(F^*_\beta)^{-1} + \delta \in \mathcal{Q}$. Meanwhile, $\rho_{2,\beta}((F^*_\beta)^{-1}(U) + \delta) = \rho_{2,\beta}((F^*_\beta)^{-1}(U) + \delta > \rho_{2,\beta}((F^*_\beta)^{-1}(U)) = M_\beta$. It contradicts with the definition that $M_\beta = \max_{F^{-1}\in\mathcal{Q}} \rho_{2,\beta}(F^{-1}(U))$.

Therefore, $(F_\beta^*)^{-1}$ satisfies $W_2((F_\beta^*)^{-1}, \hat{F}^{-1}) = \varepsilon$ and $\rho_\beta(X^{F_\beta^*}) = M_\beta$. Consequently, for any $\beta \in (\beta_0, 1)$,

$$\max_{F^{-1}\in\mathcal{Q}} L(\beta, F^{-1}) = \|\gamma_{2,\beta}\|_1 \cdot \rho_{2,\beta}(X^{F_\beta^*} - M),$$

where

$$(F_\beta^*)^{-1} = \left(\hat{F}^{-1} + \lambda_\beta\gamma_{2,\beta}\right)^{\uparrow} \tag{3.36}$$

with $\lambda_\beta \geqslant 0$ satisfying $W_2((F_\beta^*)^{-1}, \hat{F}^{-1}) = \varepsilon$. Furthermore, we have

$$\begin{aligned}
\sup_{F\in\mathcal{S}} \rho(X^F \wedge M) &= M + \min_{\beta\in[0,1]} \max_{F^{-1}\in\mathcal{Q}} L(\beta, F^{-1}) \\
&= M + \min_{\beta\in[0,1]} \left\{\|\gamma_{2,\beta}\|_1 \cdot \rho_{2,\beta}(X^{F_\beta^*} - M)\right\} \\
&= M + \|\gamma_{2,\beta^*}\|_1 \cdot \rho_{2,\beta^*}(X^{F_{\beta^*}^*} - M),
\end{aligned}$$

where $\beta^* \in \arg\min_{\beta\in[0,1]} \left\{\|\gamma_{2,\beta}\|_1 \cdot \rho_{2,\beta}(X^{F_\beta^*} - M)\right\}$.

From its projection isotonic representation given in (3.36) and similar argument used in the proof of Proposition 1 of [Bernard et al., 2020b], the quantile function $F_{\beta^*}^*$ can be expressed as

$$(F_{\beta^*}^*)^{-1}(u) = \begin{cases} \hat{F}^{-1}(u) + \lambda_{\beta^*}\gamma_{2,\beta^*}(u), & 0 < u \leqslant \theta_{\beta^*} \\ c, & \theta_{\beta^*} < u \leqslant p, \\ \hat{F}^{-1}(u), & u > p \end{cases} \tag{3.37}$$

for some $\lambda_{\beta^*} \geqslant 0$, $\theta_{\beta^*} \leqslant \beta^* \leqslant p \leqslant 1$ and constant c. Since $\theta_{\beta^*} \leqslant \beta^*$, $\gamma_{2,\beta^*}(u) = \gamma(u)$ for $0 < u \leqslant \theta_{\beta^*}$. We can also write $(F_{\beta^*}^*)^{-1}(u) = \hat{F}^{-1}(u) + \lambda_{\beta^*}\gamma(u)$ for $0 < u \leqslant \theta_{\beta^*}$.

We first claim that $\hat{F}(c-) \leqslant p \leqslant \hat{F}(p)$. Suppose $\hat{F}(c-) > p$, then there exists small $\delta > 0$ such that $(F_{\beta^*}^*)^{-1}(u) = \hat{F}^{-1}(u) < c = (F_{\beta^*}^*)^{-1}(p)$ for $u \in (p, p+\delta)$. It implies that $(F_{\beta^*}^*)^{-1}$ is not non-decreasing, which contradicts with the fact that $(F_{\beta^*}^*)^{-1}$ is a quantile function. Suppose $p > \hat{F}(p)$, then there exists small $\delta > 0$ such that $(F_{\beta^*}^*)^{-1}(u) = c < \hat{F}^{-1}(u)$ for $u \in (p-\delta, p)$. Since $\hat{F}^{-1}(u) + \lambda_{\beta^*}\gamma_{2,\beta^*}(u) \geqslant \hat{F}^{-1}(u)$ for al $u \in (0,1)$, we can strictly decrease the Wasserstein distance between $\hat{F}^{-1} + \lambda_{\beta^*}\gamma_{2,\beta^*}$ and $(F_{\beta^*}^*)^{-1}$ in (3.37) by taking $(F_{\beta^*}^*)^{-1}(u) = \hat{F}^{-1}(u)$ for $u \in (p-\delta, p)$. This is a contradiction with $(F_{\beta^*}^*)^{-1}$ in (3.37) is the isotonic projection of $\hat{F}^{-1} + \lambda_{\beta^*}\gamma_{2,\beta^*}$. Therefore, $\hat{F}(c-) \leqslant p \leqslant \hat{F}(c)$. In particular, we can take $p = \hat{F}(c)$.

Second we verify that $c = M$. Indeed, if $c > M$, we can take

$$G^{-1} = \min\{(F^*_{\beta^*})^{-1}, \max\{M, \hat{F}^{-1}\}\}.$$

Then $G^{-1} \wedge M = (F^*_{\beta^*})^{-1} \wedge M$ and thus $\rho(G^{-1}(U) \wedge M) = \rho((F^*_{\beta^*})^{-1}(U) \wedge M)$. Meanwhile, $\hat{F}^{-1}(u) \leqslant G^{-1}(u) \leqslant (F^*_{\beta^*})^{-1}(u)$ for all $0 < u < 1$ with $G^{-1}(u) < (F^*_{\beta^*})^{-1}(u)$ for some $u \in (\theta, p]$. Therefore, $G^{-1}(U)$ is strictly smaller than $(F^*_{\beta^*})^{-1}(U)$ in the sense of FOD, and $W_2(G^{-1}, \hat{F}^{-1}) < W_2(F^*_{\beta^*})^{-1}, \hat{F}^{-1}) \leqslant \varepsilon$. Take $0 < \delta < W_2(F^*_{\beta^*})^{-1}, \hat{F}^{-1}) - W_2(G^{-1}, \hat{F}^{-1})$, and construct $\tilde{G}^{-1} = G^{-1} + \delta$. Then $W_2(\tilde{G}^{-1}, \hat{F}^{-1}) \leqslant W_2(\tilde{G}^{-1}, G^{-1}) + W_2(G^{-1}, \hat{F}^{-1}) \leqslant \varepsilon$, i.e., $\tilde{G}^{-1} \in \mathcal{Q}$. Since it is assumed $q_0^2 > 0$, $(F^*_{\beta^*})^{-1}(u) < M$ for some u, and so does G^{-1}. Then $\rho(\tilde{G}^{-1}(U) \wedge M) > \rho(G^{-1}(U) \wedge M) = \rho((F^*_{\beta^*})^{-1}(U) \wedge M)$, which contradicts with the optimality of $F^*_{\beta^*}$. On the other hand, if $c < M$, then we have $\beta^* \leqslant p < q_1 := \hat{F}(M)$ because $(F^*_{\beta^*})^{-1} \geqslant \hat{F}^{-1}$. Since $(F^*_{\beta^*})^{-1}(u) = \hat{F}^{-1}(u)$ for $u > p$, we also have $q_1 = \hat{F}(M) = F^*_{\beta^*}(M)$. Then $\rho((F^*_{\beta^*})^{-1} \wedge M) = M + L(q_1, (F^*_{\beta^*})^{-1}) > M + L(\beta^*, (F^*_{\beta^*})^{-1}) = \sup_{F \in \mathcal{S}} \rho(X^F \wedge M)$ which is a contradiction. Therefore, we proof that $c = M$, and furthermore, we can take $p = \hat{F}(M) = p_1$. In conclusion, we characterize the optimal quantile function as given in (3.30) by taking $\lambda^* = \lambda_{\beta^*}$ and $\theta^* = \theta_{\beta^*}$. ■

Obviously, the worst-case quantile in (3.30) when $k = 2$ depends on not only the uncertainty set but also the risk measure ρ with weight function γ, which is a sharp difference with the worst-case quantile in (3.25) when $k = 1$. Nevertheless, the worst-case quantile in two cases present a common feature in the right tail part: there exists $0 < p \leqslant q_1$ such that $(F^*)^{-1}(u) = M$ for $p < u \leqslant q_1$ and $(F^*)^{-1}(u) = \hat{F}^{-1}(u)$ for $q_1 < u \leqslant 1$. Indeed, with the help of Lemma 3.2.2, we can show that the same feature in the right tail part of a worst-case distribution holds true for all $k \geqslant 1$.

Proposition 3.2.5 *Suppose $k \geqslant 1$ and Assumption 3.1.1 holds. If there exists*

$$(\tilde{F}^*)^{-1} \in \arg\max \left\{\rho(F^{-1}(U)) : F^{-1} \in \mathcal{A}_2\right\},$$

then

$$(F^*)^{-1}(u) = \max\{(\tilde{F}^*)^{-1}(u), \hat{F}^{-1}(u)\} = \begin{cases} (\tilde{F}^*)^{-1}(u), & \textit{if } 0 \leqslant u \leqslant \tilde{F}^*(M-) \\ M, & \textit{if } \tilde{F}^*(M-) < u \leqslant q_1 \\ \hat{F}^{-1}(u); & \textit{if } q_1 < u \leqslant 1 \end{cases} \tag{3.38}$$

is the worst-case quantile to the problem (3.16).

Proof. Define $(F^*)^{-1}(u) = \max\{(\tilde{F}^*)^{-1}(u), \hat{F}^{-1}(u)\}$. Note that $\hat{F}^{-1}(u) \leqslant (\tilde{F}^*)^{-1}(u) < M$ for $u < \tilde{F}^*(M-)$ and $(\tilde{F}^*)^{-1}(u) = M$ for $u > \tilde{F}^*(M-)$. Then we can further write $(F^*)^{-1}$ in (3.38). It is easy to check that $W_k((F^*)^{-1}, \hat{F}^{-1}) = W_k((\tilde{F}^*)^{-1}, \hat{F}^{-1} \wedge M) \leqslant \varepsilon$ and

$$\begin{aligned}\rho(X^{F^*} \wedge d) &= \int_0^1 \gamma(u) \left((F^*)^{-1}(u) \wedge M\right) \mathrm{d}u = \int_0^1 \gamma(u)(\tilde{F}^*)^{-1}(u) \, \mathrm{d}u \\ &= \sup\left\{\int_0^1 \gamma(u)F^{-1}(u) \, \mathrm{d}u : F \in \mathcal{A}_2\right\} = \sup\left\{\rho(F^{-1}(U)) : F \in \mathcal{A}_2\right\} \\ &= [\rho]_\varepsilon^k(X^{\hat{F}}; \ell_2).\end{aligned}$$

By Lemma 3.2.2, F^* is a worst-case scenario to the problem (3.16) with the given $\ell_2(x) = x \wedge M$. ∎

Example 9 *In this example, we let $\hat{F}(x) = 1-(\frac{12}{x+12})^4, x \geqslant 0$, $\varepsilon = 2$, $k = 2$ and look for the worst-case distribution with respect to a limited loss function ℓ_2 such that $\ell_2(x) = x \wedge M$. We still adopt ρ^g, a Wang's risk measure, to quantify $\ell_2(X)$, with $g(u) = \Phi(\Phi^{-1}(u) + 0.5), 0 \leqslant u \leqslant 1$. The worst-case quantile functions with different limits M are plotted in Figure 3.2:*

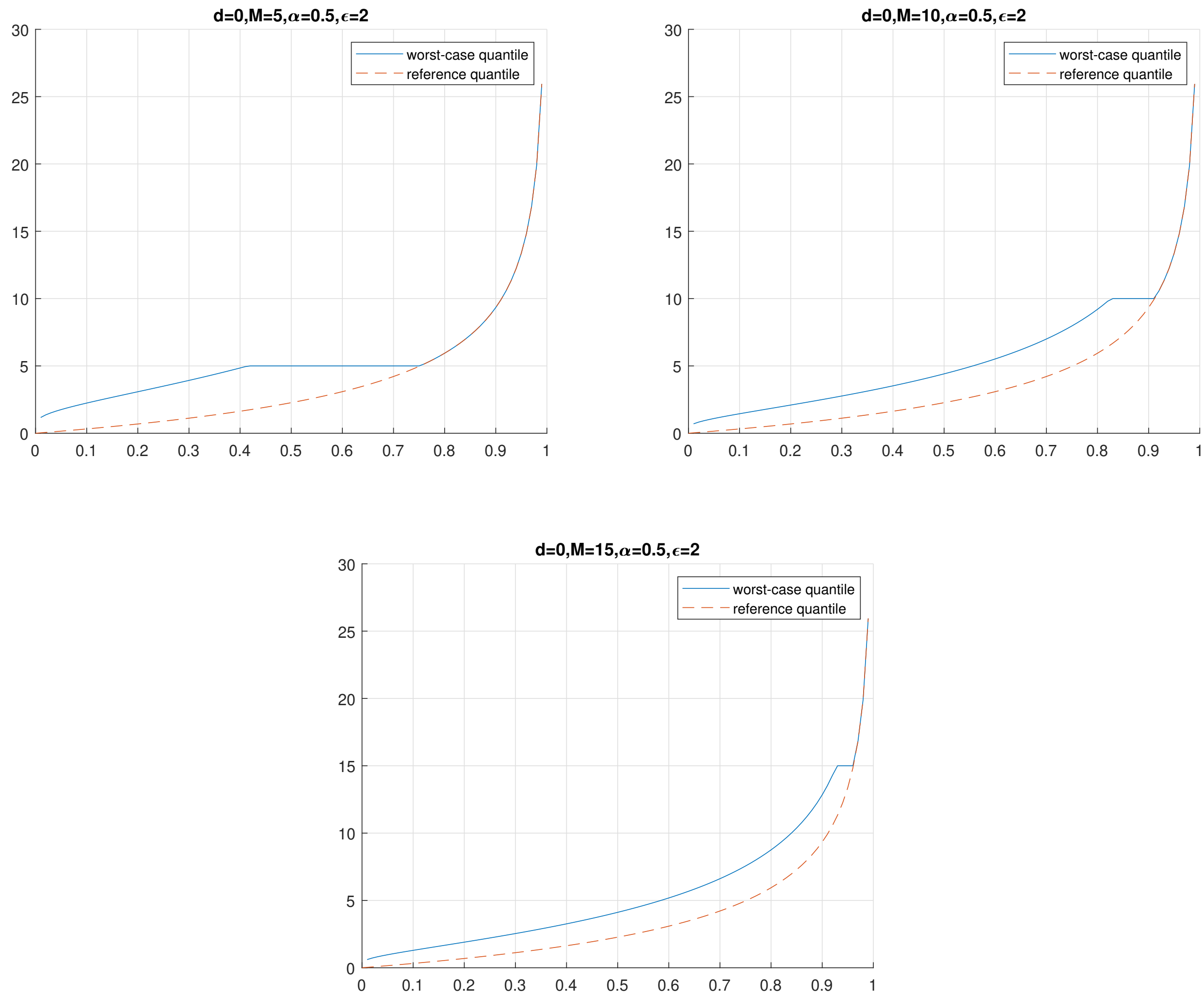
3.2.3 General limited stop-loss function

In a general case, we arbitrarily take a transform $\min\{d+(x-d)^+, M\} \in \mathcal{L}$ with $\text{ess-inf}(X^{\hat{F}}) < d < M < \text{ess-sup}(X^{\hat{F}})$. Again, since ρ is cash-invariant, we take $\ell = \min\{(x-d)^+, m\}$ where $m = M - d > 0$, and consider $[\rho]_\varepsilon^k(X^{\hat{F}}; \ell) = \sup_{F \in \mathcal{S}} \rho\left(\ell(X^F)\right)$. Heuristically, $\ell(x)$ is a two-side truncated transform. We can apply results from sections 3.2.1 and 3.2.2 to handle the lower and upper side truncation respectively. Because the limitation we have for the values of k in Section 3.2.2, in the following theorem, we consider cases when $k = 1$ and $k = 2$. Recall in Section 3.2.1 we define $\mathcal{Q}_d = \mathcal{Q} - d$ and ρ_{1,β_1} is a coherent distortion risk measure induced by distortion function g_{1,β_1} defined in (3.12).

Proposition 3.2.6 *Let $k = 1, 2$ and Assumption 3.1.1 hold. For $\ell = \min\{(x-d)_+, m\}$ with $m > 0$ and $\text{ess-inf}(X^{\hat{F}}) < d < d + m < \text{ess-sup}(X^{\hat{F}})$, we have*

$$[\rho]_\varepsilon^k(X^{\hat{F}}; \ell) = \sup_{0 \leqslant \beta \leqslant 1}\left\{\|\gamma_{1,\beta}\|_1 \cdot \max_{F^{-1} \in \mathcal{Q}_d}\left\{\rho_{1,\beta}(X^F \wedge m)\right\}\right\}. \tag{3.39}$$

Figure 3.2: Worst-case distributions with limited loss transformation.



Proof. For any $F \in \mathcal{S}$ and $\ell(x) = \min\{(x-d)_+, m\} \in \mathcal{L}$, we have

$$\begin{aligned}
\rho(\ell(X^F)) &= \int_{F(d)}^{F(d+m)} \gamma(u)\left(F^{-1}(u) - d\right)\mathrm{d}u + m\int_{F(d+m)}^{1} \gamma(u)\,\mathrm{d}u \\
&= \max_{0 \leqslant \beta \leqslant F(d+m)} \left\{ \int_{\beta}^{F(d+m)} \gamma(u)\left(F^{-1}(u) - d\right)\mathrm{d}u \right\} + m\left(1 - \int_0^{F(d+m)} \gamma(u)\,\mathrm{d}u\right) \\
&= \max_{0 \leqslant \beta \leqslant F(d+m)} \left\{ \int_{\beta}^{F(d+m)} \gamma(u)\left(F^{-1}(u) - d - m\right)\mathrm{d}u + m - m\int_0^{\beta} \gamma(u)\,\mathrm{d}u \right\} \\
&= \max_{0 \leqslant \beta \leqslant F(d+m)} \left\{ \int_{0}^{F(d+m)} \gamma_{1,\beta}(u)\left(F^{-1}(u) - d - m\right)\mathrm{d}u + m - m\int_0^{\beta} \gamma(u)\,\mathrm{d}u \right\},
\end{aligned}$$

where $\gamma_{1,\beta} = \gamma \cdot \mathbb{I}_{[\beta,1]}$ with $\|\gamma_{1,\beta}\|_1 = \int_\beta^1 \gamma(u)\,\mathrm{d}u = 1 - \int_0^\beta \gamma(u)\,\mathrm{d}u$. Using the same argument in Section 3.2.1, for any $\beta < 1$, we can define a coherent distortion risk measure $\rho_{1,\beta}$ induced by distortion function $g_{1,\beta}$ defined in (3.12). Write $F_{-d}^{-1} = F^{-1} - d$, and then $F(d+M) = F_{-d}(M)$. It is easy to check that

$$\begin{aligned}
\rho(\ell(X^F)) &= \max_{0 \leqslant \beta \leqslant F_{-d}(m)} \left\{ \int_0^{F_{-d}(m)} \gamma_{1,\beta}(u)\left(F_{-d}^{-1}(u) - m\right)\mathrm{d}u + m - m\int_0^\beta \gamma(u)\,\mathrm{d}u \right\} \\
&= \max_{0 \leqslant \beta \leqslant F_{-d}(m)} \left\{ \|\gamma_{1,\beta}\|_1 \int_0^{F_{-d}(m)} \frac{\gamma_{1,\beta}(u)}{\|\gamma_{1,\beta}\|_1}\left(F_{-d}^{-1}(u) - m\right)\mathrm{d}u + m\|\gamma_{1,\beta}\|_1 \right\} \\
&= \max_{0 \leqslant \beta \leqslant F_{-d}(m)} \left\{ \|\gamma_{1,\beta}\|_1 \cdot \rho_{1,\beta}(X^{F_{-d}} \wedge m) \right\}, \qquad (3.40)
\end{aligned}$$

where the first equality is from (3.32). If $\beta \geqslant F_{-d}(m)$, then $\gamma_{1,\beta}(u) = \gamma(u) \cdot \mathbb{I}_{[\beta,1]}(u) = 0$ for all $u < F_{-d}(m)$ and the integral $\int_0^{F_{-d}(m)} \gamma_{1,\beta}(u)\left(F_{-d}^{-1}(u) - m\right)\mathrm{d}u = 0$. Therefore, for any $\beta \geqslant F_{-d}(m)$, we have

$$\begin{aligned}
\|\gamma_{1,\beta}\|_1 \cdot \rho_{1,\beta}(X^{F_{-d}} \wedge m) &= m\int_\beta^1 \gamma(u)\,\mathrm{d}u \\
&\leqslant m\int_{F(d+m)}^1 \gamma(u)\,\mathrm{d}u = \|\gamma_{1,F_{-d}(m)}\|_1 \cdot \rho_{1,F_{-d}(m)}(X^{F_{-d}} \wedge m).
\end{aligned}$$

It says that $F_{-d}(m)$ is sub-optimal to all larger probability levels for the maximization problem (3.40). As a consequence, we have

$$
\begin{aligned}
[\rho]_{\varepsilon}^{k}(X^{\hat{F}};l) = \sup_{F\in\mathcal{S}} \rho\left(l(X^F)\right) &= \sup_{F\in\mathcal{S}} \left\{ \max_{0\leqslant\beta\leqslant F_{-d}(M)} \|\gamma_{1,\beta}\|_1 \cdot \rho_{1,\beta}(X^{F_{-d}} \wedge m) \right\} \\
&= \sup_{F\in\mathcal{S}} \left\{ \max_{0\leqslant\beta\leqslant 1} \|\gamma_{1,\beta}\|_1 \cdot \rho_{1,\beta}(X^{F_{-d}} \wedge m) \right\} \\
&= \sup_{F^{-1}\in\mathcal{Q}_d} \left\{ \max_{0\leqslant\beta\leqslant 1} \|\gamma_{1,\beta}\|_1 \cdot \rho_{1,\beta}(X^{F} \wedge m) \right\} \\
&= \sup_{0\leqslant\beta\leqslant 1} \left\{ \|\gamma_{1,\beta}\|_1 \sup_{F^{-1}\in\mathcal{Q}_d} \rho_{1,\beta}(X^{F} \wedge m) \right\}.
\end{aligned}
$$

For any $\beta \in [0,1]$, the function $\gamma_{1,\beta} = \mathbb{I}_{\{[\beta,1]\}}$ is increasing and thus the risk measure $\rho_{1,\beta}$ is coherent. The problem $\sup_{F^{-1}\in\mathcal{Q}_d} \rho_{1,\beta}(X^F \wedge m)$ can be solved by Theorem 3.2.3 and Theorem 3.2.4 when $k=1$ and $k=2$, respectively. Therefore, the expression (3.39) is obtained. ■

When $k=1$, we can further simplify (3.39) in Proposition 3.2.6. We know from (3.25) that for any β the maximum $\max_{F^{-1}\in\mathcal{Q}_d} \left\{\rho_{1,\beta}(X^F \wedge m)\right\}$ is achieved at

$$
(F^*)^{-1}(u) = \begin{cases} \hat{F}^{-1}(u) - d, & 0 \leqslant u \leqslant \tilde{q}_0^1; \\ m - d, & \tilde{q}_0^1 \leqslant u \leqslant \tilde{q}_1; \\ \hat{F}^{-1}(u) - d, & \tilde{q}_1 < u \leqslant 1, \end{cases}
$$

where $\tilde{q}_1 = \hat{F}(m+d)$ and $\tilde{q}_0^1 = \inf\left\{q \geqslant 0 : \int_q^{q_1} \left| m - \hat{F}^{-1}(u) + d \right| \,\mathrm{d}u \leqslant \varepsilon \right\}$. Therefore,

$$
[\rho]_{\varepsilon}^{1}(X^{\hat{F}};\ell) = \max_{0\leqslant\beta\leqslant 1} \left\{ \|\gamma_{1,\beta}\|_1 \cdot \rho_{1,\beta}(X^{F^*} \wedge m) \right\}.
$$

When $k=2$, (3.30) implies that $\max_{F^{-1}\in\mathcal{Q}_d} \left\{\rho_{1,\beta}(X^F \wedge m)\right\}$ is achieved at

$$
(F_\beta^*)^{-1}(u) = \begin{cases} \hat{F}^{-1}(u) + \lambda^*\gamma_{1,\beta}(u) - d, & \text{for } 0 < u < \theta^*, \\ m - d, & \text{for } \theta^* < u < \tilde{q}_1, \\ \hat{F}^{-1}(u) - d, & \text{for } \tilde{q}_1 < u < 1 \end{cases}
$$

where $\lambda^* > 0$ and $\theta^* \in (0,1)$ satisfies $W_2(F_\beta^*, \hat{F} - d) = \varepsilon$. Therefore

$$
[\rho]_{\varepsilon}^{2}(X^{\hat{F}};\ell) = \max_{0\leqslant\beta\leqslant 1} \left\{ \|\gamma_{1,\beta}\|_1 \cdot \rho_{1,\beta}(X^{F_\beta^*} \wedge m) \right\}.
$$

Example 10 *In this example, we let* $\hat{F}(x) = 1 - (\frac{12}{x+12})^4, x \geqslant 0$, $\varepsilon = 2$, $k = 2$ *and look for the worst-case distribution with respect to a stop-loss function* l *such that* $\ell(x) = (x-d)_+ \wedge M$. *We adopt* ρ^g, *a Wang's risk measure, to quantify* $\ell(X)$, *with* $g(u) = \Phi(\Phi^{-1}(u) + 0.5), 0 \leqslant u \leqslant 1$. *The definition of* ρ^g *follows Definition 1.2.8. The worst-case quantiles with different deductible* d *and* M *are plotted in Figure 3.2:*

3.3 Application in reinsurance premium

3.3.1 Wang's premium in the worst-case scenario

In practice, a limited stop-loss is commonly used in a reinsurance treaty. A reinsurer can adopt a certain premium principle to calculate the reinsurance premium. Among popular principles, Wang's premium principle is introduced by [Wang et al., 1997]. Indeed, Wang's premium principle uses a coherent distortion risk measure to quantify the loss ceded to the reinsurer. To apply Wang's premium principle, the reinsurer may need to assume the distribution of the underlying risk. Due to the limited information, the true distribution can vary from the reference one used by the reinsurer. Therefore, the reinsurer can be interested in the reinsurance premium in the worst case. Mathematically, for a limited stop-loss reinsurance $l \in \mathcal{L}$, Wang's reinsurance premium induced by a distortion risk measure ρ_w is defined as

$$\pi(\ell(X)) = (1+\theta)\rho_w(\ell(X)) \tag{3.41}$$

where $\theta \geqslant 0$ is a given risk loading and ρ_w is the distortion risk measure induced by the distribution function

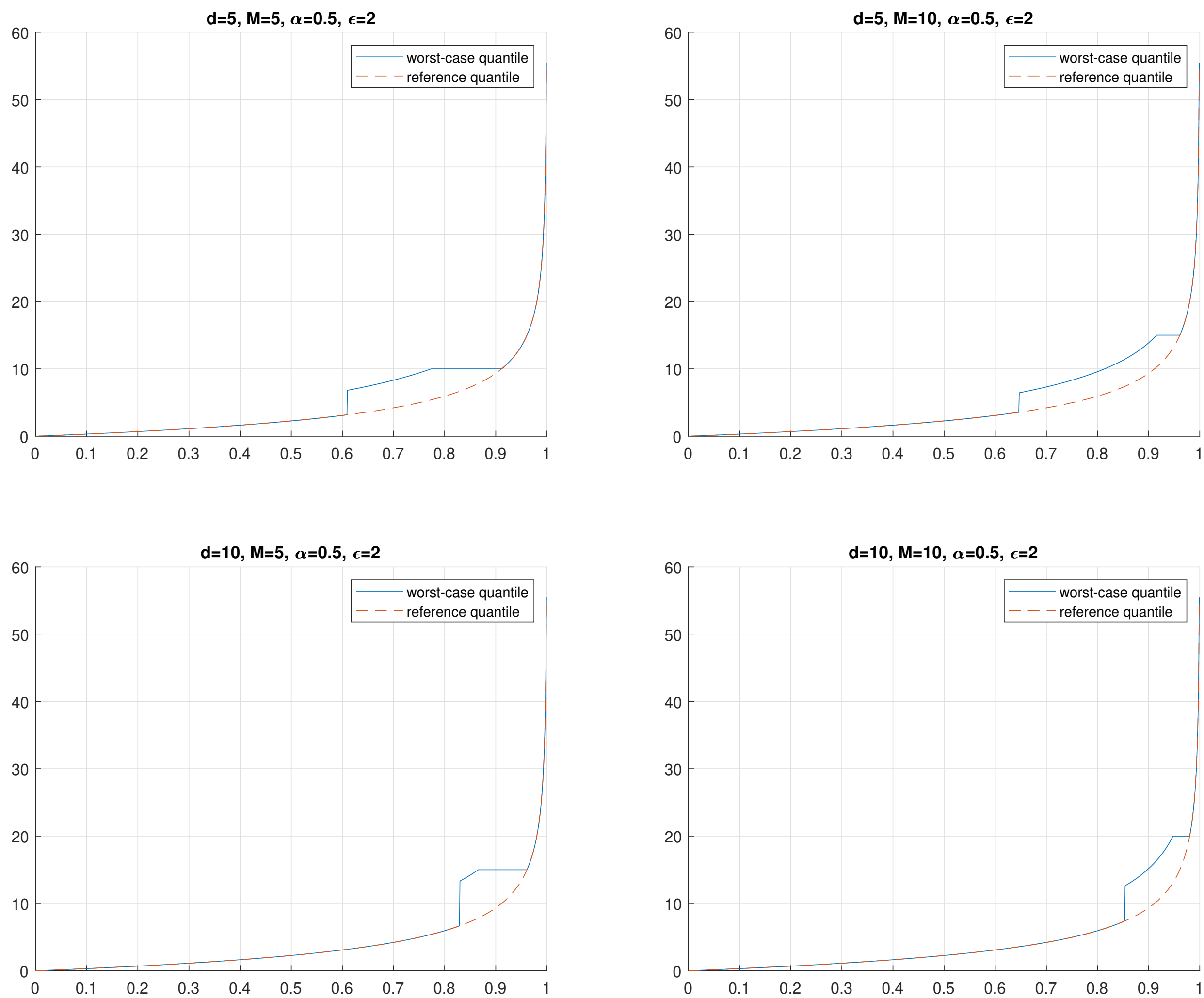
$$g_w(u) = \Phi(\Phi^{-1}(u) + \alpha), \qquad 0 \leqslant u \leqslant 1 \text{ for a given } \alpha \in (0,1).$$

Suppose the reinsurer has a reference distribution $\hat{F}$, and defines the uncertainty set $\mathcal{S}$ as the one given in (3.4). The worst-case Wang's premium can be formulated as

$$\pi^{\uparrow}(\ell(X)) := \sup_{F \in \mathcal{S}} (1+\theta)\rho_w(\ell(X^F)) = (1+\theta) \cdot [\rho_w]_{\varepsilon}^{k}(X^{\hat{F}}; \ell), \tag{3.42}$$

which can be solved by the results obtained in previous sections.

Figure 3.3: Worst-case distributions with limited stop-loss transformation.



Example 11 *We first consider the changes of the worst-case Wang's premium with respect to ε and retention levels of the reinsurance policies. Mathematically, we fix $k = 2$, $\theta = 0$ and two reference distributions*

$$\hat{F}_1(x) = 1 - e^{-x/4}, x \geqslant 0, \qquad \hat{F}_2 = 1 - \left(\frac{12}{x+12}\right)^4, x \geqslant 0.$$

Note that we set the means of the two reference distributions to be equal for better comparison. Their variances, however, cannot be matched on that condition, with the Pareto distribution's variance always larger.

In the context of insurance, exponential distribution family and Pareto distribution family are commonly used to fit a light-tail risk and a heavy-tail risk, respectively. Both exponential distributions and Pareto distributions are well-defined on the whole non-negative part of the real line. We consider a general limited stop-loss reinsurance policy $\ell(x) = (x-d)^+ \wedge M$ with $0 < d, M < \infty$, and take $\varepsilon > 0$. By default, we set $\varepsilon = 2$, $M = 5$ and $d = 5$. To show the effect of various parameters on the worst-case risk measure of our interest, we change one parameter at a time and illustrate its effect on the objective worst-case risk measure. The other two parameters maintain as the default setting at the mean time.

Figure 3.4, Figure 3.5 and Figure 3.6 respectively show the worst-case Wang's premium when ε changes in the range $(0.1, 1.9)$, limit M changes in the range $(4, 13)$, and deductible d changes in the range $(0.5, 9.5)$. One can refer to Table 3.1 to track the corresponding numerical results.

In Figure 3.4, Wang's premium of the Pareto reference distribution is measured as 1.4748, while that of the exponential distribution is 1.5355, with ε changing.

3.3.2 Risk measure based loss ratio

In the P&C insurance market, the concept of Loss Elimination Ratio (LER) is commonly used to calculate the portion of loss removed from the insurance seller's payment liability in the sense of the mean. Given the underlying risk $X^{\hat{F}}$ of the insurer and the reinsurance policy ℓ, the LER is defined as

$$\text{LER} = 1 - \frac{\mathbb{E}[\ell(X^{\hat{F}})]}{\mathbb{E}[X^{\hat{F}}]}.$$

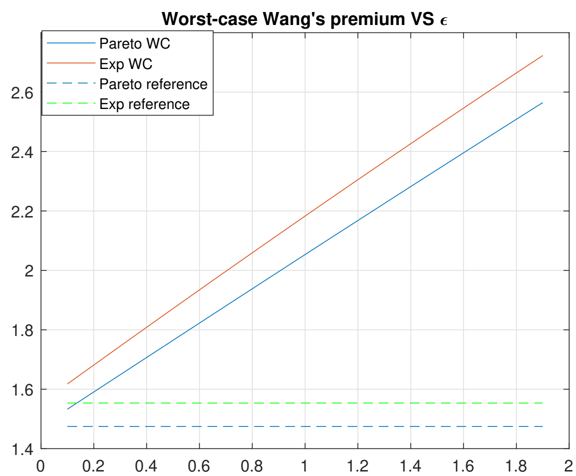


Figure 3.4: $\pi^{\uparrow}(\ell(X))$ vs ε

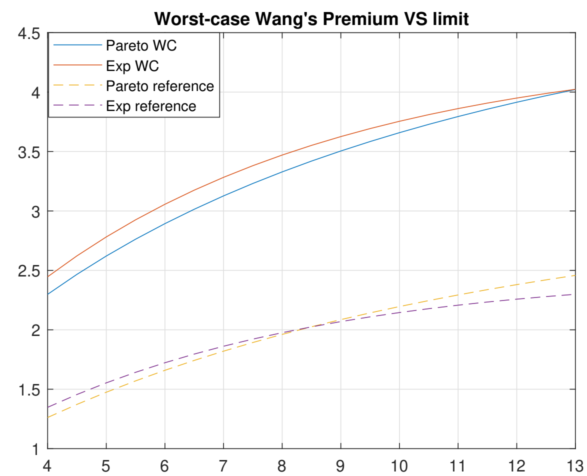


Figure 3.5: $\pi^{\uparrow}(\ell(X))$ vs M

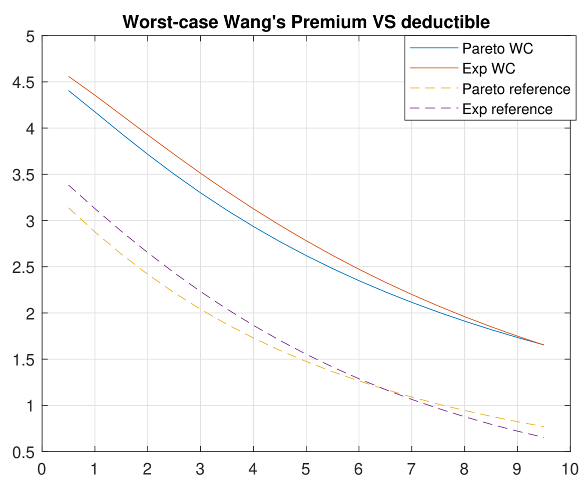


Figure 3.6: $\pi^{\uparrow}(\ell(X))$ vs d

Furthermore, if the reinsurer uses the expected-value premium principle, LER also represent the portion of the premium amount removed from the full insurance premium due to the policy adjustments d and M, while $1 - \text{LER}$ is the portion of the premium earned by the reinsurer. It worth pointing out that a reinsurance policy can be considered as a risk sharing tool used between two participants, and the mean is one particular risk measure chosen to quantify the risk exposure level for two participants. As a consequence, LER and 1-LER represent the portion of liability of assigned to the buyer and sellers under a given ℓ, respectively.

In general, the reinsurer and the insurer may choose other risk measures to quantify risk exposure levels. Inspired by the classical definition of LER, we propose the following general definition for the loss ratio based on a distortion risk measure.

Definition 3.3.1 (ρ-based Loss Ratio) *Given an underlying risk $X^{\hat{F}}$, a distortion risk measure ρ^g, and an indemnity function ℓ in a reinsurance policy, the ρ-based loss ratios for the reinsurer and the insurer are defined as*

$$\mathrm{LR}_R^{\rho} = \frac{\rho(\ell(X^{\hat{F}}))}{\rho(X^{\hat{F}})} \quad \textit{and} \quad \mathrm{LR}_I^{\rho} = \frac{\rho(X^{\hat{F}} - \ell(X^{\hat{F}}))}{\rho(X^{\hat{F}})}, \tag{3.43}$$

respectively.

In (3.43), $\rho(X^{\hat{F}})$ is the total risk exposure, while $\rho(\ell(X^{\hat{F}}))$ and $\rho(X^{\hat{F}} - \ell(X^{\hat{F}}))$ are risk exposure levels taken by the reinsurer and the insurer, respectively. Note that a distortion risk measure ρ is always comonotonic additive. For an admissible indemnity function ℓ, we have $\ell(X^{\hat{F}})$, $X^{\hat{F}} - \ell(X^{\hat{F}})$ and $X^{\hat{F}}$ are comonotonic random variables. Therefore, $\rho(\ell(X^{\hat{F}})) + \rho(X^{\hat{F}} - \ell(X^{\hat{F}})) = \rho(X^{\hat{F}})$ and $\mathrm{LR}_I^{\rho} + \mathrm{LR}_R^{\rho} = 1$ hold true.

In particular, if we take $\rho = \rho_w$ defined in (3.41), the $\rho_w(X^{\hat{F}})$ is Wang's premium for the full insurance, while $\rho(\ell(X^{\hat{F}}))$ is the Wang's premium of the insurance policy using the indemnity function ℓ. Then $\mathrm{LR}_I^{\rho_w}$ and $\mathrm{LR}_R^{\rho_w}$ show how the full premium is shared between two participants. In the following, we use several examples to investigate the LER^{ρ_w} in the worst-case scenarios.

Example 12 *From (3.43), it is easy to see that the LR^{ρ_w} is law-invariant. In the worst-case scenario, the worst-case distribution may be different from the reference distribution $\hat{F}$, and therefore, the associated LR^{ρ_w} deviated from the value given in (3.43) using the* *reference distribution. Suppose the reinsurer adopts an uncertainty set* $\mathcal{S}_R$*, and then defined the reinsurer's worst-case LR as*

$$\mathrm{LR}_R^{\rho\uparrow} = \frac{\rho(\ell(X^{F_R^*}))}{\rho(X^{F_R^*})}, \qquad where \qquad F_R^* = \underset{F \in \mathcal{S}_R}{\arg\max}\, \rho(\ell(X^F)). \tag{3.44}$$

Similarly, suppose the insurer adopts an uncertainty set $\mathcal{S}_I$*, and defined the insurer's worst-case LR as*

$$\mathrm{LR}_I^{\rho\uparrow} = \frac{\rho(X^{F_I^*} - \ell(X^{F_I^*}))}{\rho(X^{F_I^*})}, \qquad where \qquad F_I^* = \underset{F \in \mathcal{S}_I}{\arg\max}\, \rho(X^F - \ell(X^F)). \tag{3.45}$$

The worst-case distributions F_I^* *and* F_R^* *may be different even if we take same uncertainty set* $\mathcal{S}_I = \mathcal{S}_R$ *because the loss transformation function* $\ell(x)$ *and* $x - \ell(x)$ *are different.*

In this example, we take $\ell(x) = (x-d)^+$ *for* $d > 0$*. Then* $x - \ell(x) = \min\{x, d\}$*. In other words, the reinsurer take the stop-loss part while the insurer take the limited-loss part. Take* $\rho = \rho_w$*, where* $g(u) = \Phi(\Phi^{-1}(u) + 0.5)$*,* $0 \leqslant u \leqslant 1$*. Meanwhile, assume two uncertainty sets are the same and are given by* $\mathcal{S}_I = \mathcal{S}_R = \{F : W_2(F, \hat{F}) \leqslant 2\}$ *with the Pareto reference distribution* $\hat{F}(x) = 1 - \left(\frac{12}{x+12}\right)^4$*,* $x \geqslant 0$*. We first set a threshold for one participant's LR using the reference distribution, denoted by* $\mathrm{LR}^{\mathrm{ref}}$*. To meet this threshold, this participant can determine the value of* d *required in the reinsurance policy. Then the participant calculate the worst-case LR using* d*. Results for both the insurer and the reinsurer are summarized in the Table 3.2. The plots of* $\mathrm{LR}_I^{\uparrow}$ *VS* LR_I^{ref} *and* $\mathrm{LR}_R^{\uparrow}$ *VS* LR_R^{ref} *are in Figure 3.7 and Figure 3.8, respectively.*

Example 13 *Following the setup of Example 12, we are focused on the change of the reinsurer's worst-case LR against the change of* ε*, a parameter of the size of the uncertainty set. Instead of a more general loss function* l*, we set: a)* $d = 5, M = \infty$*, b)* $d = 0, M = 5$ *to look at the effect on stop-loss function and limited loss function, respectively. We keep the assumption of reference distribution in the Example 11. The numerical illustrations are given in Figure 3.9 and Figure 3.10 and the supporting data are in Table 3.3. In Figure 3.9, the reference LR's for Pareto and exponential distributions are 0.4981 and 0.4042, respectively. In Figure 3.10, the reference LR's for Pareto and exponential distributions are 0.5019 and 0.5958, respectively.*

Example 14 *Following the setup of Example 12, we are focused on the change of the reinsurer's worst-case LR against the change of deductible* d*, and limit* M *respectively, for a limited stop-loss reinsurance. We set* $d = 5$ *while* M *changes and* $M = 5$ *while* d *changes. We keep the settings in the previous examples and set* $\varepsilon = 2$*. The results are presented in Figure 3.11 and Figure 3.12, and precise values are summarized below in Table 3.4.*

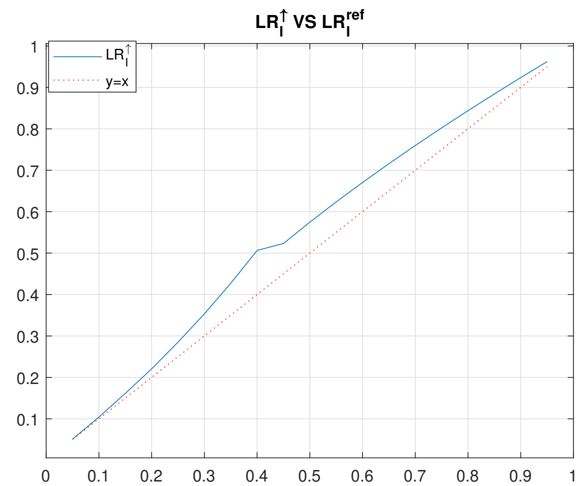


Figure 3.7: Example 12-$\mathrm{LR}_I^{\uparrow}$ vs LR_I^{ref}

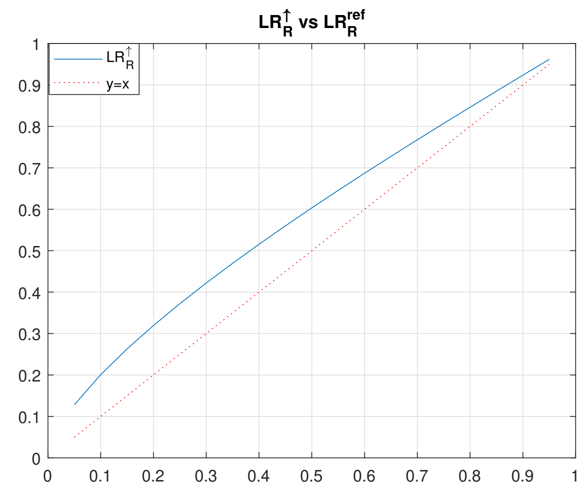


Figure 3.8: Example 12-$\mathrm{LR}_R^{\uparrow}$ vs LR_R^{ref}

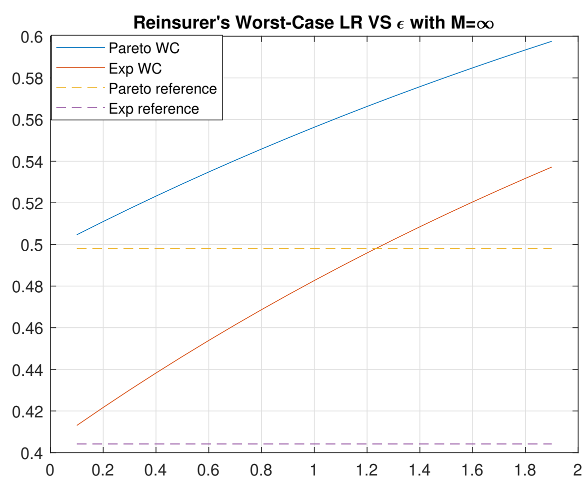


Figure 3.9: Example 13-$\mathrm{LR}_R^{\uparrow}$ vs ε with $d = 5, M = \infty$

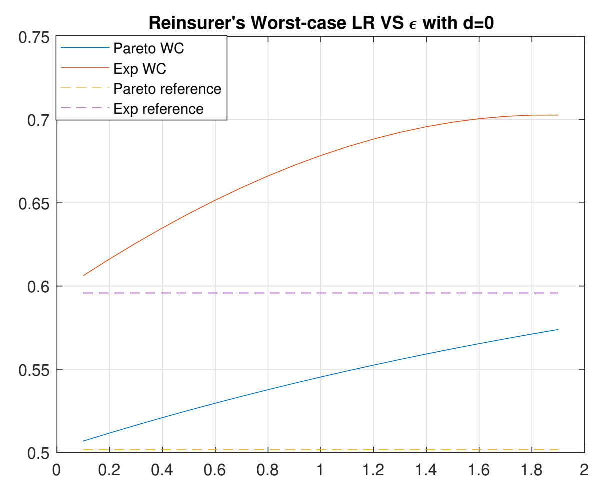


Figure 3.10: Example 13-$\mathrm{LR}_R^{\uparrow}$ vs ε with $d = 0, M = 5$

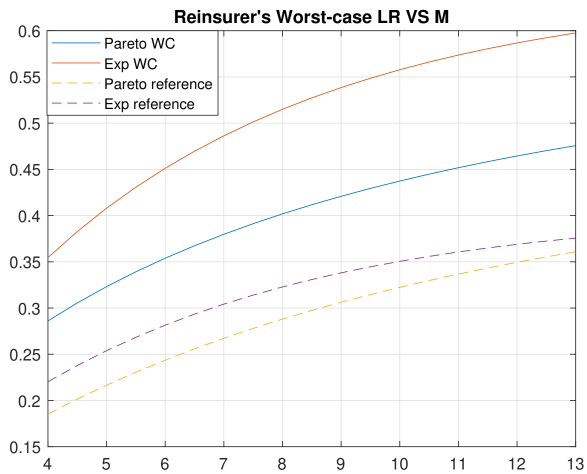


Figure 3.11: Example 14-$\mathrm{LR}_R^{\uparrow}$ vs M with $d = 5$

Figure 3.12: Example 14-$\mathrm{LR}_R^{\uparrow}$ vs d with $M = 5$

3.4 Concluding remarks

In this chapter, we focused on solving the worst-case problem $\sup_{F \in \mathcal{S}} \rho(l(X^F))$, l being a limited stop-loss function and $\mathcal{S}$ being the uncertainty set specified as a Wasserstein ball. In comparison to Chapter 2, in this chapter, we allow the order k of Wasserstein distance to be any real number no less than 1. We first looked at the special cases of the general problem, i.e., stop-loss function and limited loss function, and gave analytical solutions to the worst-case distribution and worst-case risk measure in both cases. For the stop-loss function part, our result applies when $k \geqslant 1$, which covers the result in [Guan et al., 2022]. For the limited loss function part, we presented analytical results for $k = 1$ and $k = 2$. We then provided several examples that illustrate our theoretical results for stop-loss, limited loss and limited stop-loss functions.

We then applied our theoretical results to the scenario of robustly pricing reinsurance treaties. We showed how each reinsurance contract parameter affects the reinsurance premiums decided in the worst-case. We also generalized Loss Elimination Ratio to ρ-based Loss Ratio that can help quantify the reinsurer's risk exposure level. We also looked at how this ratio is affected by the changes in deductible, limit and uncertainty level.

Table 3.1: Example 4.1-Notation ^{w} refers to the worst case while ^{r} refers to reference case.

ε	PAR^w	EXP^w	M	PAR^w	EXP^w	PAR^r	EXP^r	d	PAR^w	EXP^w
0.10	1.5328	1.6176	4.00	2.2986	2.4463	1.2623	1.3478	0.50	4.4066	4.5603
0.20	1.5908	1.6814	4.50	2.4667	2.6223	1.3723	1.4556	1.00	4.1752	4.3560
0.30	1.6488	1.7450	5.00	2.6207	2.7814	1.4748	1.5535	1.50	3.9419	4.1418
0.40	1.7068	1.8084	5.50	2.7625	2.9255	1.5702	1.6423	2.00	3.7160	3.9269
0.50	1.7647	1.8715	6.00	2.8933	3.0562	1.6593	1.7228	2.50	3.5015	3.7158
0.60	1.8226	1.9343	6.50	3.0143	3.1750	1.7425	1.7958	3.00	3.2997	3.5114
0.70	1.8804	1.9968	7.00	3.1268	3.2829	1.8203	1.8619	3.50	3.1112	3.3150
0.80	1.9381	2.0591	7.50	3.2313	3.3811	1.8932	1.9217	4.00	2.9356	3.1277
0.90	1.9957	2.1211	8.00	3.3287	3.4705	1.9615	1.9758	4.50	2.7723	2.9498
1.00	2.0532	2.1827	8.50	3.4197	3.5517	2.0256	2.0248	5.00	2.6207	2.7814
1.10	2.1106	2.2442	9.00	3.5046	3.6257	2.0858	2.0691	5.50	2.4799	2.6224
1.20	2.1678	2.3052	9.50	3.5842	3.6932	2.1424	2.1091	6.00	2.3491	2.4728
1.30	2.2250	2.3659	10.00	3.6587	3.7545	2.1957	2.1452	6.50	2.2277	2.3320
1.40	2.2820	2.4263	10.50	3.7287	3.8105	2.2459	2.1778	7.00	2.1145	2.1999
1.50	2.3389	2.4864	11.00	3.7945	3.8614	2.2933	2.2073	7.50	2.0094	2.0760
1.60	2.3956	2.5461	11.50	3.8565	3.9077	2.3380	2.2338	8.00	1.9114	1.9598
1.70	2.4522	2.6055	12.00	3.9148	3.9500	2.3802	2.2578	8.50	1.8199	1.8510
1.80	2.5085	2.6645	12.50	3.9697	3.9884	2.4202	2.2794	9.00	1.7345	1.7489
1.90	2.5647	2.7231	13.00	4.0216	4.0236	2.4580	2.2989	9.50	1.6547	1.6536

3.5 Appendix

3.5.1 Tables

Table 3.2: Example 4.2

LR_I^{ref}	LR_R^{ref}	d	$\mathrm{LR}_I^{\uparrow}$	$\mathrm{LR}_R^{\uparrow}$		LR_I^{ref}	LR_R^{ref}	d	$\mathrm{LR}_I^{\uparrow}$	$\mathrm{LR}_R^{\uparrow}$
0.05	0.95	0.3471	0.0509	0.9620		0.5	0.5	4.9705	0.5746	0.6032
0.1	0.9	0.7121	0.1043	0.9236		0.55	0.45	5.8186	0.6235	0.5600
0.15	0.85	1.0998	0.1612	0.8851		0.6	0.4	6.8041	0.6705	0.5155
0.2	0.8	1.5149	0.2206	0.8463		0.65	0.35	7.9723	0.7158	0.4696
0.25	0.75	1.9628	0.2849	0.8072		0.7	0.3	9.3930	0.7597	0.4219
0.3	0.7	2.4495	0.3532	0.7676		0.75	0.25	11.1805	0.8023	0.3721
0.35	0.65	2.9826	0.4271	0.7276		0.8	0.2	13.5399	0.8438	0.3194
0.4	0.6	3.5716	0.5064	0.6869		0.85	0.15	16.8889	0.8842	0.2629
0.45	0.55	4.2289	0.5232	0.6455		0.9	0.1	22.2741	0.9236	0.2007

Table 3.3: Example 4.3

Fix $d = 5$, $M = \infty$				Fix $d = 0$, $M = 5$		
ε	PAR	EXP		ε	PAR	EXP
0.10	0.5047	0.4131		0.10	0.5047	0.4131
0.20	0.5110	0.4217		0.20	0.5110	0.4217
0.30	0.5172	0.4301		0.30	0.5172	0.4301
0.40	0.5232	0.4382		0.40	0.5232	0.4382
0.50	0.5291	0.4462		0.50	0.5291	0.4462
0.60	0.5348	0.4539		0.60	0.5348	0.4539
0.70	0.5404	0.4614		0.70	0.5404	0.4614
0.80	0.5458	0.4687		0.80	0.5458	0.4687
0.90	0.5511	0.4758		0.90	0.5511	0.4758
1.00	0.5563	0.4827		1.00	0.5563	0.4827
1.10	0.5614	0.4894		1.10	0.5614	0.4894
1.20	0.5663	0.4959		1.20	0.5663	0.4959
1.30	0.5711	0.5023		1.30	0.5711	0.5023
1.40	0.5758	0.5085		1.40	0.5758	0.5085
1.50	0.5804	0.5145		1.50	0.5804	0.5145
1.60	0.5848	0.5204		1.60	0.5848	0.5204
1.70	0.5892	0.5262		1.70	0.5892	0.5262
1.80	0.5934	0.5317		1.80	0.5934	0.5317
1.90	0.5976	0.5372		1.90	0.5976	0.5372

Table 3.4: Example 4.4

Fix $d = 5$, $\varepsilon = 2$						Fix $M = 5$, $\varepsilon = 2$				
M	PAR^w	EXP^w	PAR^r	EXP^r		d	PAR^w	EXP^w	PAR^r	EXP^r
4.00	0.2860	0.3545	0.1853	0.2202		0.50	0.5445	0.6812	0.4608	0.5530
4.50	0.3055	0.3827	0.2015	0.2378		1.00	0.5123	0.6531	0.4224	0.5112
5.00	0.3232	0.4080	0.2165	0.2538		1.50	0.4820	0.6206	0.3871	0.4713
5.50	0.3392	0.4308	0.2306	0.2683		2.00	0.4538	0.5869	0.3551	0.4334
6.00	0.3539	0.4512	0.2436	0.2815		2.50	0.4277	0.5536	0.3259	0.3979
6.50	0.3674	0.4697	0.2558	0.2934		3.00	0.4034	0.5214	0.2995	0.3646
7.00	0.3798	0.4863	0.2673	0.3042		3.50	0.3810	0.4906	0.2756	0.3337
7.50	0.3913	0.5014	0.2780	0.3140		4.00	0.3602	0.4614	0.2540	0.3049
8.00	0.4019	0.5150	0.2880	0.3228		4.50	0.3410	0.4339	0.2343	0.2784
8.50	0.4117	0.5273	0.2974	0.3308		5.00	0.3232	0.4080	0.2165	0.2538
9.00	0.4209	0.5385	0.3063	0.3381		5.50	0.3066	0.3838	0.2004	0.2312
9.50	0.4294	0.5486	0.3146	0.3446		6.00	0.2912	0.3611	0.1857	0.2105
10.00	0.4374	0.5578	0.3224	0.3505		6.50	0.2769	0.3398	0.1723	0.1914
10.50	0.4448	0.5661	0.3298	0.3558		7.00	0.2636	0.3200	0.1601	0.1739
11.00	0.4518	0.5737	0.3367	0.3606		7.50	0.2511	0.3014	0.1490	0.1580
11.50	0.4584	0.5805	0.3433	0.3650		8.00	0.2395	0.2840	0.1388	0.1433
12.00	0.4645	0.5868	0.3495	0.3689		8.50	0.2286	0.2678	0.1295	0.1300
12.50	0.4703	0.5924	0.3554	0.3724		9.00	0.2184	0.2526	0.1209	0.1178
13.00	0.4757	0.5976	0.3609	0.3756		9.50	0.2089	0.2384	0.1131	0.1068

Chapter 4

The Pareto optimal deductible in the worst case with different risk measures

In Chapters 2 and 3, we solved the individual optimization problems (2.3), (2.4) and (3.2). For a given stop loss reinsurance contract, we determined the worst-case distribution of total loss X which leads to the largest risk measure values of the insurer's and reinsurer's parts of the loss. Obviously, the worst-case distribution depends on the given deductible d, and a change in the deductible may have reverse influences on the two parties' objective values. Motivated by this observation, in this chapter, we would like to further investigate the optimal deductible for the insurer and reinsurer, respectively. In a reinsurance contract, the choice of the deductible should be viewed as a result of the negotiation between the insurer and the reinsurer. As a result, the determination of the optimal deductible should take into consideration the joint perspectives from both parties, so that it can balance the objectives from different sides. To this end, in this chapter, we will propose a Pareto optimization problem incorporating both the insurer and the reinsurer's objectives and then determine the deductible d^* that can help reach the equilibrium between them in the sense of Pareto optimality.

4.1 Model setup and Pareto optimal solutions

To quantify the comprehensive losses for both the insurer and the reinsurer, in this chapter, we will introduce reinsurance premium as part of the model. In this chapter, the premium is determined using the mean-value principle, as the most adopted and one of the most tractable premium principles in the study of reinsurance. We assume that the reinsurer uses a pre-specified pricing distribution $\tilde{F}$ for the loss X to calculate the premium. The value of the reinsurance premium depends on the reinsurance deductible $d \geqslant 0$ chosen in the reinsurance policy. For simplicity, we write

$$\pi(d) = (1+\theta)\mathbb{E}[(X-d)_+] = (1+\theta)\int_d^\infty \tilde{S}(x)\,\mathrm{d}x, \qquad d \geqslant 0, \tag{4.1}$$

where $\theta > 0$ is the safety loading and $\tilde{S}$ is the survival function corresponding to $\tilde{F}$.

From results given in Chapter 2, the worst-case distribution among the uncertainty sets $\mathcal{S}_i$, $i = 1, 2$, may not have a close-form expression. The optimal deductible may not be mathematically tractable. In this chapter, we remove the Wasserstein distance constraints for both the insurer and the reinsurer. Equivalently, we consider a special case of $\mathcal{S}_i$ with $\varepsilon_i = \infty$, i.e.,

$$\mathcal{S}_i^\infty \triangleq \left\{ F : \int x\,\mathrm{d}F(x) = \mu_i, \int x^2\,\mathrm{d}F(x) = \mu_i^2 + \sigma_i^2 \right\}, \qquad i = 1, 2. \tag{4.2}$$

Given a stop-loss reinsurance contract I with deductible $d \geqslant 0$, the total retained loss for the insurer is $I(X) = X \wedge d + \pi(d)$. If the insurer adopts the risk measure ρ_1, then the insurer's optimization problem becomes

$$\min_{d \geqslant 0} \left\{ \sup_{F \in \mathcal{S}_1^\infty} \rho_1 \left((X^F \wedge d) + \pi(d) \right) \right\}. \tag{4.3}$$

Meanwhile, the reinsurer's total loss is $X - I(X) - \pi(d) = (X-d)_+ - \pi(d)$, and thus the reinsurer is interested in the following optimization problem

$$\min_{d \geqslant 0} \left\{ \sup_{G \in \mathcal{S}_2^\infty} \rho_2 \left((X^G - d)_+ - \pi(d) \right) \right\}. \tag{4.4}$$

Since the insurer and the reinsurer have conflicting interests in a reinsurance contract, intuitively the optimal deductibles for the both sides do not coincide. That means, an optimal deductible for one side may not be optimal, and even unacceptable for the other. It is important to find a "fair and acceptable" contract for two parties. To this end, we consider reinsurance solutions from the Pareto-optimal point of view under distributional uncertainty. We first define Pareto-optimal solutions under a general deductible reinsurance setting with model uncertainty.

Definition 4.1.1 *Let $(\rho_1, \rho_2, \pi, \mathcal{S}_1^\infty, \mathcal{S}_2^\infty)$ be a reinsurance setting with model uncertainty, where ρ_1 and ρ_2 are risk measures adopted by the insurer and the reinsurer, respectively. A deductible $d^* \geqslant 0$ is called Pareto-optimal with respect to $(\rho_1, \rho_2, \pi, \mathcal{S}_1^\infty, \mathcal{S}_2^\infty)$ if, for any $d \geqslant 0$ such that*

$$\sup_{F\in\mathcal{S}_1^\infty} \rho_1(X^F \wedge d + \pi(d)) \leqslant \sup_{F\in\mathcal{S}_1^\infty} \rho_1(X^F \wedge d^* + \pi(d^*))$$
$$\sup_{G\in\mathcal{S}_2^\infty} \rho_2((X^G - d)_+ - \pi(d)) \leqslant \sup_{G\in\mathcal{S}_2^\infty} \rho_2((X^G - d^*)_+ - \pi(d^*)),$$

the two inequalities must be equalities.

As can be found in literature, such as [Cai et al., 2017], a Pareto-optimal reinsurance deductible exists if there exists a deductible d minimizing a convex combination of the worst-case risk measures of the insurer and the reinsurer. The following proposition gives a sufficient condition for a deductible d to be Pareto-optimal with respect to the reinsurance setting $(\rho_1, \rho_2, \pi, \mathcal{S}_1^\infty, \mathcal{S}_2^\infty)$.

Proposition 4.1.1 *Given a reinsurance setting $(\rho_1, \rho_2, \pi, \mathcal{S}_1^\infty, \mathcal{S}_2^\infty)$, if d^* is an optimal solution to the problem*

$$\min_{d\geqslant 0}\left\{\delta \cdot \sup_{F\in\mathcal{S}_1^\infty} \rho_1\left(X^F \wedge d + \pi(d)\right) + (1-\delta)\cdot \sup_{G\in\mathcal{S}_2^\infty} \rho_2\left((X^G - d)_+ - \pi(d)\right)\right\}, \tag{4.5}$$

for some $\delta \in (0, 1)$, then d^* is a Pareto-optimal reinsurance deductible with respect to the setting $(\rho_1, \rho_2, \pi, \mathcal{S}_1^\infty, \mathcal{S}_2^\infty)$.

Proof. Assume d^* is a minimizer to the problem (4.5) for $\delta \in (0, 1)$. If d^* is not a Pareto-optimal deductible, then there exists $\tilde{d} \geqslant 0$, such that

$$\sup_{F\in\mathcal{S}_1^\infty} \rho_1(X^F \wedge \tilde{d} + \pi(\tilde{d})) \leqslant \sup_{F\in\mathcal{S}_1^\infty} \rho_1(X^F \wedge d^* + \pi(d^*))$$
$$\sup_{G\in\mathcal{S}_2^\infty} \rho_2((X^G - \tilde{d})_+ - \pi(\tilde{d})) \leqslant \sup_{G\in\mathcal{S}_2^\infty} \rho_2((X^G - d^*)_+ - \pi(d^*))$$

with at least one of the inequalities being strict. This implies

$$\begin{aligned}&\delta \sup_{F\in\mathcal{S}_1^\infty} \rho_1(X^F \wedge \tilde{d} + \pi(\tilde{d})) + (1-\delta) \sup_{G\in\mathcal{S}_2^\infty} \rho_2((X^G - \tilde{d})_+ - \pi(\tilde{d}))\\ <\, &\delta \sup_{F\in\mathcal{S}_1^\infty} \rho_1(X^F \wedge d^* + \pi(d^*)) + (1-\delta) \sup_{G\in\mathcal{S}_2^\infty} \rho_2((X^G - d^*)_+ - \pi(d^*)\end{aligned}$$

which contradicts the assumption that d^* is an optimal solution to the problem (4.5). ■

Optimization problems using uncertainty sets with fixed first two moments have been intensively investigated in the literature for many popular risk measures, including VaR and TVaR (see [Ghaoui et al., 2003], [Li, 2018]). In the context of insurance and reinsurance, the discussions on Pareto optimization problems with reinsurance premium are missing. To shed light on this area, we are going to determine Pareto optimal solutions to the problem (4.5) with respect to VaR and TVaR which are the two most popular risk measures for either party. In the following, we first present the VaR case in Section 4.2, then summarize the TVaR case in Section 4.3. Numerical illustrations of optimal solutions will be provided as well.

4.2 Optimal deductible with respect to worst-case VaR

In this section, we consider the Pareto optimization problem (4.5) with $\rho_i = \mathrm{VaR}_{\alpha_i}$ for $\alpha_i \in (0,1)$, $i = 1, 2$, i.e.,

$$\min_{d \geqslant 0} \left\{ \delta \cdot \sup_{F \in \mathcal{S}_1^\infty} \mathrm{VaR}_{\alpha_1}\left((X^F \wedge d) + \pi(d) \right) + (1-\delta) \cdot \sup_{G \in \mathcal{S}_2^\infty} \mathrm{VaR}_{\alpha_2}\left((X^G - d)_+ - \pi(d) \right) \right\}. \tag{4.6}$$

We first solve two inner optimization problems for the insurer and the reinsurer. It is well known that, given the first two moments, the value of VaR_α in the worst-case can be calculated by the closed-form

$$\mu + \sigma\sqrt{\frac{\alpha}{1-\alpha}} = \sup_{\mathbb{E}[X^F]=\mu, \mathrm{var}(X^F)=\sigma^2} \mathrm{VaR}_\alpha\left(X^F\right)$$

which is obtained at the worst-case distribution

$$F^{-1}(p) = \begin{cases} \mu - \sigma\sqrt{\frac{\alpha}{1-\alpha}}, & 0 < p \leqslant \alpha \\ \mu + \sigma\sqrt{\frac{\alpha}{1-\alpha}}, & \alpha < p < 1. \end{cases} \tag{4.7}$$

We refer [Ghaoui et al., 2003] and references there in.

Lemma 4.2.1 *Given $d \geqslant 0$, the following equations hold,*

$$\sup_{F \in \mathcal{S}_1^\infty} \mathrm{VaR}_{\alpha_1}(X^F \wedge d) = \left(\mu_1 + \sigma_1\sqrt{\frac{\alpha_1}{1-\alpha_1}} \right) \wedge d$$

$$\sup_{G \in \mathcal{S}_2^\infty} \mathrm{VaR}_{\alpha_2}((X^G - d)_+) = \left(\mu_2 + \sigma_2\sqrt{\frac{\alpha_2}{1-\alpha_2}} - d \right)_+ .$$

Furthermore, all equations hold at the worst-case distribution given by (4.7).

The proof to Proposition 4.2.1 can be found in Section 4.6. This result coincides with Case 2 of Proposition 1 in [Hu et al., 2015], wherein the supremum of the retained loss's VaR_α is found with respect to an identical uncertainty set to ours. Due to the fact that we do not assume any bound for the total loss X, our result coincides only with the second case of the proposition mentioned above, since they assume both an upper bound and a lower bound on the total loss, and only in the second case the bounds become ineffective. With this result, the optimization problem (4.3) now becomes

$$\min_{d \geqslant 0} \left\{ \left(\mu_1 + \sigma_1 \sqrt{\frac{\alpha_1}{1-\alpha_1}} \right) \wedge d + (1+\rho) \int_d^\infty \tilde{S}(x)\,\mathrm{d}x \right\}, \tag{4.8}$$

and the the optimization problem (4.4) becomes

$$\min_{d \geqslant 0} \left\{ \left(\mu_2 + \sigma_2 \sqrt{\frac{\alpha_2}{1-\alpha_2}} - d \right)_+ - (1+\rho) \int_d^\infty \tilde{S}(x)\,\mathrm{d}x \right\}. \tag{4.9}$$

Theorem 4.2.2 *Denote*

$$d_i = \mu_i + \sigma_i \sqrt{\frac{\alpha_i}{1-\alpha_i}}, \quad i = 1, 2, \qquad \text{and} \quad \tilde{x}_1 = \tilde{F}^{-1}\left(\frac{\rho}{1+\rho}\right).$$

The optimal deductible d_1^* to the insurer's problem (4.8) *is*

$$d_1^* = \begin{cases} \infty, & \text{if } d_1 \leqslant \tilde{x}_1 + (1+\rho) \int_{\tilde{x}_1}^\infty \tilde{S}(x)\,\mathrm{d}x \\ \tilde{x}_1, & \text{if } d_1 > \tilde{x}_1 + (1+\rho) \int_{\tilde{x}_1}^\infty \tilde{S}(x)\,\mathrm{d}x. \end{cases} \tag{4.10}$$

The optimal deductible d_2^* to the reinsurer's problem (4.9) *is*

$$d_2^* = \begin{cases} 0, & \text{if } d_2 < (1+\rho) \int_0^{d_2} \tilde{S}(x)\,\mathrm{d}x \\ d_2, & \text{if } d_2 \geqslant (1+\rho) \int_0^{d_2} \tilde{S}(x)\,\mathrm{d}x. \end{cases} \tag{4.11}$$

Following Theorem 4.2.2, we can calculate the insurer's and the reinsurer's minimal worst-case VaR.

1. For the insurer, the minimal worst-case VaR is $\min\left\{d_1, \tilde{x}_1 + (1+\rho) \int_{\tilde{x}_1}^\infty \tilde{S}(x)\,\mathrm{d}x\right\}$. Precisely,

- if $d_1 \leqslant \tilde{x}_1 + (1+\rho)\int_{\tilde{x}_1}^{\infty} \tilde{S}(x)\,\mathrm{d}x$, the insurer prefers no insurance (i.e., $d_1^* = \infty$) and the corresponding minimal worst-case VaR is

$$d_1 = \mu_1 + \sigma_1\sqrt{\frac{\alpha_1}{1-\alpha_1}};$$

- if $d_1 > \tilde{x}_1 + (1+\rho)\int_{\tilde{x}_1}^{\infty} \tilde{S}(x)\,\mathrm{d}x$, the insurer prefers the partial insurance with deductible $\tilde{x}_1 = \tilde{F}^{-1}\left(\frac{\rho}{1+\rho}\right)$ and the corresponding minimal worst-case VaR is

$$\tilde{x}_1 + (1+\rho)\int_{\tilde{x}_1}^{\infty} \tilde{S}(x)\,\mathrm{d}x.$$

2. For the reinsurer,

- if $d_2 < (1+\rho)\int_0^{d_2} \tilde{S}(x)\,\mathrm{d}x$, the reinsurer prefers the full insurance and the corresponding minimal worst-case VaR is

$$d_2 - (1+\rho)\tilde{\mu} = \mu_2 + \sigma_2\sqrt{\frac{\alpha_2}{1-\alpha_2}} - (1+\rho)\mathbb{E}[X^{\tilde{F}}],$$

- if $d_2 \geqslant (1+\rho)\int_0^{d_2} \tilde{S}(x)\,\mathrm{d}x$, the reinsurer prefers the partial insurance with deductible d_2 and the corresponding minimal worst-case VaR is

$$-(1+\rho)\int_{d_2}^{\infty} \tilde{S}(x)\,\mathrm{d}x.$$

With the help of Theorem 4.2.2, we are ready to solve the Pareto optimal deductible d^* to the problem (4.6). For simplicity, we write the objective function in the problem (4.6) as follows, for $d \geqslant 0$ and $\delta \in [0,1]$, let

$$h_\delta(d) = \delta\left(d_1 \wedge d + (1+\rho)\int_d^{\infty} \tilde{S}(x)\,\mathrm{d}x\right) + (1-\delta)\left((d_2-d)_+ - (1+\rho)\int_d^{\infty} \tilde{S}(x)\,\mathrm{d}x\right).$$

Obviously, $h_\delta(d)$ is continuous and differentiable a.e with respect to d on $[0,\infty)$.

Theorem 4.2.3 *Given a reinsurance setting* $(\mathrm{VaR}_{\alpha_1}, \mathrm{VaR}_{\alpha_2}, \pi, \mathcal{S}_1^{\infty}, \mathcal{S}_2^{\infty})$, *the optimal deductible* d^* *to the problem* (4.6) *is given below.*

(a) When $0 < \delta < \frac{1}{2}$,

(i) if $h_\delta(0) < h_\delta(d_2)$, then $d^* = 0$;

(ii) if $h_\delta(0) > h_\delta(d_2)$, then $d^* = d_2$;

(iii) if $h_\delta(0) = h_\delta(d_2)$, then d^* can be either 0 or d_2.

(b) When $\delta = \frac{1}{2}$,

(i) if $d_1 < d_2$, then d^* can be any value from the set $[d_2, \infty)$.

(ii) if $d_1 > d_2$, then d^* can be any value from the set $[0, d_2]$;

(iii) if $d_1 = d_2$, then d^* can be any value from the set $[0, \infty)$.

(c) When $\frac{1}{2} < \delta < 1$ and $d_1 < d_2$,

(i) if $(1+\rho)\tilde{S}(d_1) \geqslant 1$ or $h_\delta(\tilde{x}_1) \geqslant \delta d_1$, then $d^* = \infty$;

(ii) if $(1+\rho)\tilde{S}(d_1) < 1$ and $h_\delta(\tilde{x}_1) < \delta d_1$, then $d^* = \tilde{x}_1$.

(d) When $\frac{1}{2} < \delta < 1$ and $d_1 \geqslant d_2$,

(i) if $(1+\rho)\tilde{S}(d_2) < 1$ and $h_\delta(\tilde{x}_1) \leqslant \delta d_1$, then $d^* = \tilde{x}_1$;

(ii) if $1 \leqslant (1+\rho)\tilde{S}(d_2) \leqslant \frac{\delta}{2\delta-1}$ and $h_\delta(d_2) \leqslant \delta d_1$, then $d^* = d_2$;

(iii) if $(1+\rho)\tilde{S}(d_1) \leqslant \frac{\delta}{2\delta-1} \leqslant (1+\rho)\tilde{S}(d_2)$ and $h_\delta(\tilde{F}^{-1}(\frac{\delta-1+2\rho\delta-\rho}{(2\delta-1)(1+\rho)})) \leqslant \delta d_1$, then $d^* = \tilde{F}^{-1}(\frac{\delta-1+2\rho\delta-\rho}{(2\delta-1)(1+\rho)})$;

(iv) in all other cases we have $d^* = \infty$.

From the previous results, we can observe that the optimal deductible d^* depends on the reinsurance setting, especially parameters including $\mu_1, \mu_2, \sigma_1, \sigma_2, \tilde{F}, \alpha_1, \alpha_2$ and so on. It is noteworthy that the weight parameter, δ, can be taken as the weight given to the insurer in the negotiation process between the two parties. The larger δ is, the more negotiation power the insurer possesses and the more likely it is that the Pareto-optimal deductible will be closer to insurer's optimal deductible, and vice versa. A numerical demonstration of the above theorem is given below.

Example 15 *Let $\mu_1 = \sigma_1 = 5, \mu_2 = 4, \sigma_2 = 6, \alpha_1 = \alpha_2 = 0.95, \rho = 0.5, \tilde{F}(x) = 1 - (\frac{4}{x})^3, x \geqslant 4$. The joint worst-case* VaR *of the both parties in three situations are plotted below in Figure 4.1, where $\delta = 0, 0.25, 0.5, 0.75, 1$, respectively.*

δ	0	0.25	0.5	0.75	1
d^*	30.1534	30.1534	∞	4.5789	4.5789

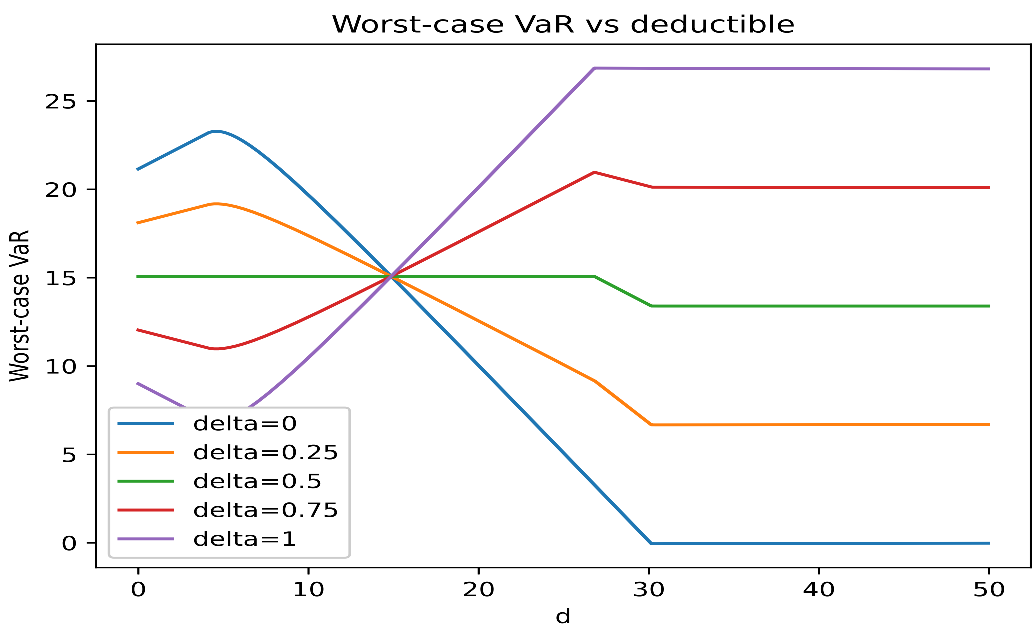


Figure 4.1: Joint worst-case VaR against d

However, one can easily observe the shortcoming of specifying a distribution, $\tilde{F}$, for premium pricing, that is the lacking of robustness. Since the pricing distribution is selected by the reinsurer to compensate for the component of risk it takes, it should take into account the potential errors that may occur during model estimations. To be conservative, the distribution for premium pricing should also be in worst-case. As a result, in the section, the premium principle becomes

$$\pi(d) = (1+\rho) \sup_{G \in \mathcal{S}_2^\infty} \mathbb{E}[(X^G - d)_+], \tag{4.12}$$

where ρ is the safety loading. The uncertainty set for the reinsurer, $\mathcal{S}_2^\infty$, is adopted here since the reinsurer decides the worst-case scenario in regard to deciding the premium.

After entering a stop-loss reinsurance contract with deductible $d \geqslant 0$, the insurer's risk exposure is $\mathrm{VaR}_{\alpha_1}(X \wedge d + \pi(d))$, and the reinsurer's risk exposure is $\mathrm{VaR}_{\alpha_2}((X-d)_+ - \pi(d))$.

With the result of Corollary 2.62, the worst-case version of reinsurance premium, (2.63), is equivalent to:

$$\pi(d) = \frac{(1+\rho)}{2} \left(\mu_2 - d + \sqrt{(\mu_2 - d)^2 + \sigma_2^2} \right). \tag{4.13}$$

Example 16 $\mu_1 = \sigma_1 = 5, \mu_2 = 4, \sigma_2 = 6, \alpha_1 = \alpha_2 = 0.95, \rho = 0.5$. *The joint worst-case* VaR *of the both parties in three situations are plotted below in Figure 4.2, where* $\delta = 0, 0.25, 0.5, 0.75, 1$*, respectively.*

δ	*0*	*0.25*	*0.5*	*0.75*	*1*
d^*	*30.1534*	*30.1534*	*30.9018*	*1.8787*	*1.8787*

4.3 The Pareto optimal deductible with respect to worst-case TVaR

Owing to the increasing popularity of TVaR as a tool for measuring risk exposures adopted by insurance institutions, we would like to investigate the Pareto optimal deductible under worst-case TVaR with the same uncertainty sets. That is, the optimization problem is

$$\min_{d \geqslant 0} \left\{ \delta \cdot \sup_{F \in \mathcal{S}_1^\infty} \mathrm{TVaR}_{\alpha_1} \big((X^F \wedge d) + \pi(d) \big) + (1-\delta) \cdot \sup_{G \in \mathcal{S}_2^\infty} \mathrm{TVaR}_{\alpha_2} \big((X^G - d)_+ - \pi(d) \big) \right\}, \tag{4.14}$$

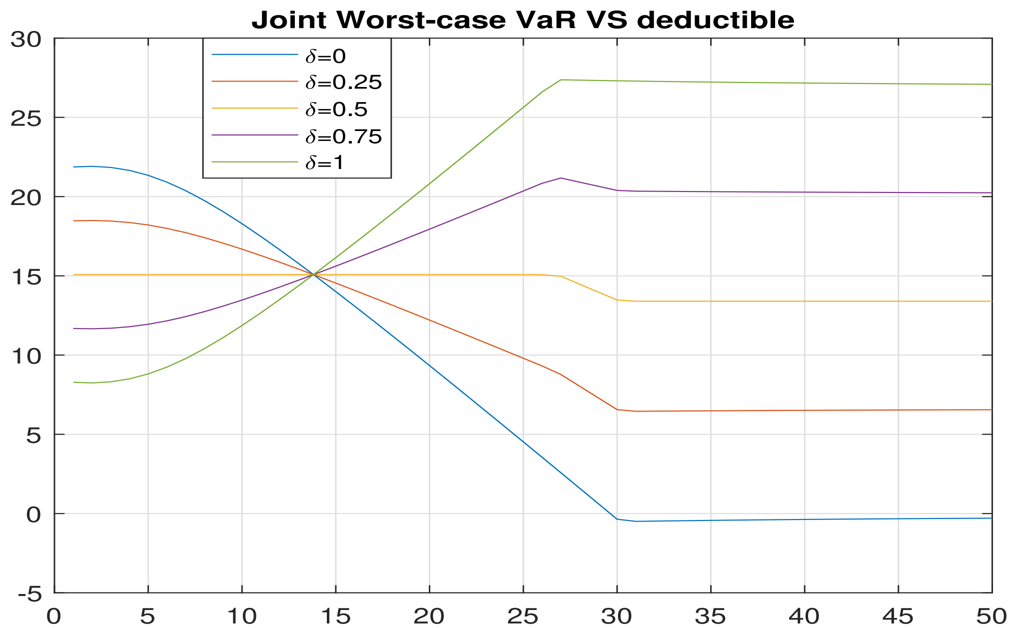


Figure 4.2: Joint worst-case VaR against d

where $\mathcal{S}_1^\infty$ and $\mathcal{S}_2^\infty$ follow the same definitions as in (4.2), $\delta \in (0,1)$, and $\pi(d)$ is the premium associated with deductible d. In this section, we again adopt the mean-value premium principle.

Since we have investigated the optimal deductible from the perspectives of the insurer and the reinsurer, respectively, we can now visit the Pareto-optimal deductible problem where the risk exposures of both sides are measured using TVaR and worst-case premium principle in 4.12. According to Proposition 4.1.1, $d^* \geqslant 0$ is an optimal deductible in Pareto sense if it minimizes the following objective function:

$$h_\delta(d) = \delta(\sup_{F\in\mathcal{S}_1^\infty} \mathrm{TVaR}_\alpha(X^F \wedge d) + \pi(d)) + (1-\delta)(\sup_{G\in\mathcal{S}_2^\infty} \mathrm{TVaR}_\alpha(X^G - d)_+ - \pi(d)), \tag{4.15}$$

or equivalently,

$$h_\delta(d) = \begin{cases} \delta(d\wedge d_1) + \frac{1+\rho}{2}(2\delta-1)(\mu_2 - d + \sqrt{(\mu_2-d)^2+\sigma_2^2}) + (1-\delta)(d_2-d), & 0 \leqslant d \leqslant d_3, \\ \delta(d\wedge d_1) + (\frac{1+\rho}{2}(2\delta-1) + (1-\delta)\frac{1}{2(1-\alpha_2)})(\mu_2 - d + \sqrt{(\mu_2-d)^2+\sigma_2^2}), & d > d_3, \end{cases} \tag{4.16}$$

where $\delta \in (0,1)$ is the negotiation power of the insurer.

The Pareto-optimal deductible problem, however, is intractable analytically. Hence, we would like to solve this kind of problems numerically and demonstrate it using an example.

Example 17 *Let $\mu_1 = \sigma_1 = 5, \mu_2 = 4, \sigma_2 = 6, \alpha_1 = \alpha_2 = 0.95, \rho = 0.5$, three situations of the joint worst-case* TVaR *of the both parties are plotted below in Figure 4.3 against deductible d, with $\delta = 0, 0.25, 0.5, 0.75, 1$, respectively.*

δ	*0*	*0.25*	*0.5*	*0.75*	*1*
d^*	∞	∞	*10.3599*	*1.8787*	*1.8787*

To this point, it would not bother too much to take a look at the Pareto optimality with TVaR and traditional mean-value premium principle.

Example 18 *Let $\mu_1 = \sigma_1 = 5, \mu_2 = 4, \sigma_2 = 6, \alpha_1 = \alpha_2 = 0.95, \rho = 0.5, \tilde{F}(x) = 1 - (\frac{4}{x})^3, x \geqslant 4$. Three situations of the joint worst-case* TVaR *of the both parties are plotted below in Figure 4.4 against deductible d, with $\delta = 0, 0.25, 0.5, 0.75, 1$, respectively.*

δ	*0*	*0.25*	*0.5*	*0.75*	*1*
d^*	∞	∞	*11.8583*	*4.5789*	*4.5789*

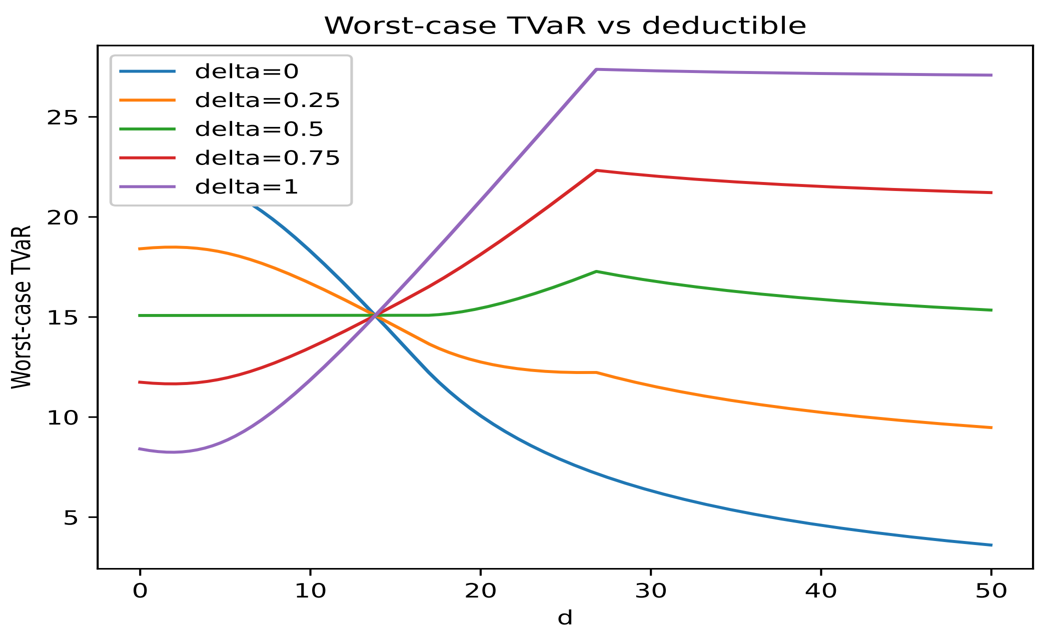


Figure 4.3: Joint worst-case TVaR vs d

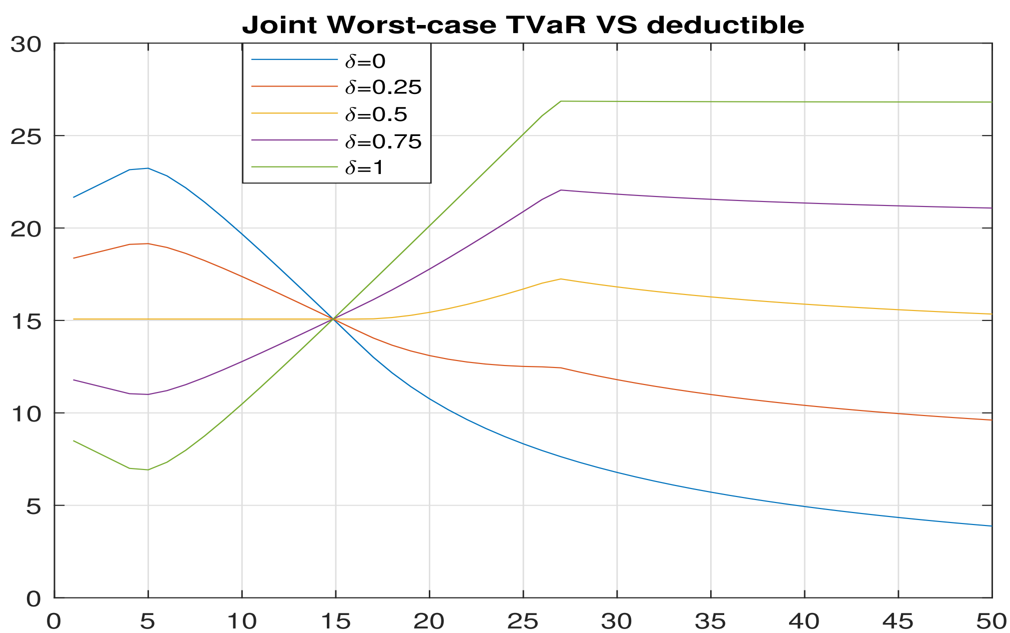


Figure 4.4: Joint worst-case TVaR vs d

4.4 The Pareto optimal deductible with respect to worst-case Wang's Premium

In this section, we investigate the Pareto optimality adopting the uncertainty sets and risk measures in Chapter 3. The optimization problem is hence:

$$\min_{d \geqslant 0} h_\delta(d),$$

where

$$h_\delta(d) = \left\{ \delta \cdot \sup_{F \in \mathcal{S}_1} \rho^{g_1}\left((X^F \wedge d) + \pi(d) \right) + (1-\delta) \cdot \sup_{G \in \mathcal{S}_2} \rho^{g_1}\left((X^G - d)_+ - \pi(d) \right) \right\}. \quad (4.17)$$

Herein, we take $\hat{F}(x) = 1 - (\frac{12}{x+12})^4$ and $\hat{G}(x) = 1 - e^{-x/4}$, $\delta \in (0,1)$, $\pi(d)$ being the premium associated with deductible d. Uncertainty sets for the insurer and reinsurer are

$$\mathcal{S}_1 = \{F : W_2(F, \hat{F}) \leqslant 0.5\} \text{ and } \mathcal{S}_2 = \{G : W_2(G, \hat{G}) \leqslant 1\}.$$

Note that here we assume the reinsurer's reference distribution, $\hat{G}$, is an exponential distribution, even though typically a heavy-tailed distribution is more likely to be selected by a insurance practitioner. Here we select different types of distributions for the both sides to illustrate an extreme case where there is substantial discrepancy between the parties concerning the reference distribution. This can happen in reality especially when the information is not shared between them.

In this section, we adopt the following four different non-robust premium principles to look at the differences they bring to the optimal deductible d^*.

$$\pi_1(d) = (1+\theta)\rho^{g_2}((X^{\hat{G}} - d)_+), \text{ where } g_2(u) = \Phi(\Phi^{-1}(u) + 0.1). \quad (4.18)$$

$$\pi_2(d) = (1+\theta)\mathbb{E}[(X^{\hat{G}} - d)_+]. \quad (4.19)$$

$$\pi_3(d) = (1+\theta)\mathrm{TVaR}_{0.9}((X^{\hat{G}} - d)_+). \quad (4.20)$$

$$\pi_4(d) = (1+\theta)[0.75\mathbb{E}[(X^{\hat{G}} - d)_+] + 0.25\mathrm{TVaR}_{0.9}((X^{\hat{G}} - d)_+)]. \quad (4.21)$$

Note that in the above setting, we assume that the insurer and the reinsurer adopt the same distortion risk measure, $\rho^{g_1}, g_1(u) = \Phi(\Phi^{-1}(u) + 0.5)$, to quantify their individual risks. As opposed to the general idea that the two counterparts can individually select their own risk metric, the assumption may seem lacking generality. Indeed, it serves as a

	$\delta = 0$	$\delta = 0.25$	$\delta = 0.5$	$\delta = 0.75$	$\delta = 1$
π_1	∞	∞	13.13	0	0
π_2	∞	∞	13.13	0	0
π_3	0	0	13.13	20.30	21.05
π_4	∞	∞	13.13	12.88	12.84

Table 4.1: d^* with different δ and π

simplification since we have already so many parameters and does not bring any speciality to the numerical results, due to the complex calculation.

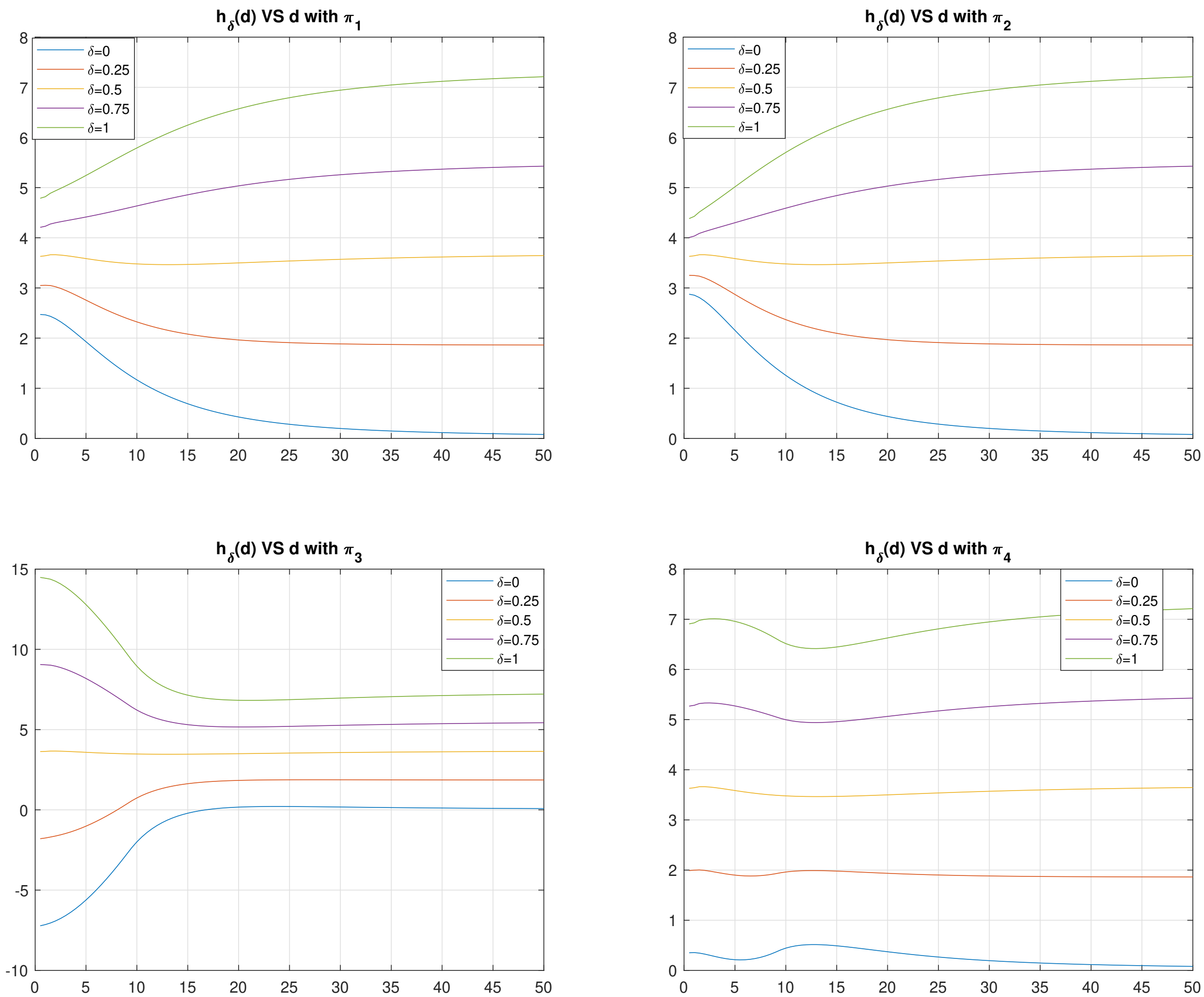
Among the above premium principles, π_4, as a linear combination of Mean-value Principle and TVaR Principle, is in the form of $\lambda\mathbb{E}+(1-\lambda)\mathrm{TVaR}_{0.9} = \mathbb{E}+(1-\lambda)(\mathrm{TVaR}_{0.9}-\mathbb{E}), \lambda \in (0, 1)$. The term $(1-\lambda)(\mathrm{TVaR}_{0.9}-\mathbb{E})$ can be viewed as a loading to the Expectation Premium Principle. The numerical example below sheds light on the optimal deductible in this setting with weight $\delta = 0, 0.25, 0.5, 0.75, 1$ and with the above four premium principles, respectively.

Example 19 *Let $\theta = 0.1$. The change of h_δ with respect to d is given in Figure 4.5. The optimal deductible corresponding to different weights δ and premium principles π are summarized in Table 4.1.*

From the numerical results, we can see the dramatic effects the choice of premium principle makes on the Pareto optimal deductible in the particular example. When we are using Wang's Premium (π_1), Expectation Premium (π_2) and Combined Premium (π_4), Pareto optimal deductible decreases with increasing weight put on the insurer. TVaR *Premium (π_3), however, deviates from this phenomenon. The reason is when premium is too high, the insurer tends not to buy and reinsurance and full coverage meets the reinsurer's best interest.*

When $\delta = 0$, the blue lines refer to the reinsurer's risk against deductible. Among the plots, Mean-value Premium and Wang's premium do not differ significantly, commonly returns ∞ as the reinsurer's optimal deductible. In contrast, TVaR *Premium principle brings substantially different result as mentioned above. As a linear combination of Mean-value Premium and* TVaR *Premium, π_4 gives the reinsurer the same optimal deductible as π_2.*

Figure 4.5: Joint worst-case Wang's Premium against d



Similarly is when $\delta = 1$. The high premium brought by TVaR *principle restricts the insurer from choosing a full coverage. However, the insurer's optimal deductible moves toward 0 when the premium puts more weight on the mean instead of* TVaR*, as depicted by the fourth plot of Figure 4.5.*

It is easy to notice that the Pareto optimal deductibles are the same for all four premium principles when $\delta = 0.5$. This is not coincident. When $\delta = 0.5$, one can easily transform (4.17) using the comonotonic-additivity property of distortion risk measures and find the premium term is ruled out from the formula of joint worst-case Wang's Premium.

In a bilateral cooperative negotiation model, Pareto-optimality is the state where no action can be taken that makes are party better off, without hurting the other party's interest. Even though we adhere to five concrete values of the weight δ, we restate the choice of δ can be arbitrarily from $[0, 1]$. With each choice giving us a Pareto-optimal deductible, the two parties' worst-case risk measures can form a Pareto frontier. To put the result into application, however, we should find a suitable point on the frontier that reflects the true bargaining power comparison between the two parties.

Broad researches have been carried out in the resource allocation problem of a bilateral cooperative negotiation model. In [Jazayeriy et al., 2011] is presented a Pareto-optimal algorithm in bilateral automated negotiation where the negotiation is modeled by "split the pie" game and alternating-offer protocol. [Bagga et al., 2020] present a novel negotiation model that allows an agent to learn how to negotiate during con- current bilateral negotiations in unknown and dynamic e-markets.

Even though this work does not expand the discussion in game theory, we still seek a reasonable weight allocation plan in our model, to facilitate the both parties in finding a balance point in their negotiation. We regard 0.75 as a reasonable value for δ, by giving more weights to the insurer, who is at the center of the insurance industry. One can refine the weight allocation strategy using extra information including the market condition and so on.

4.5 Concluding remarks

The process to reach a contract often accompanies negotiation. In this mainly numerically established chapter, we investigated how the balance point is reached between the two parties of a reinsurance contract, as well as how the balance moves according to the change in both parties' uncertainty recognition, risk metric, negotiation power, etc.. Generally, the movement of the balance point coincides with that when there is no model uncertainty, but counter-intuitive results may occur if, e.g., the premium is set too high.

4.6 Proofs

Proof of Lemma 4.2.1. Since both functions $x \wedge d$ and $(x-d)_+$ are increasing in x, it holds that

$$\sup_{F \in \mathcal{S}_1} \mathrm{VaR}_\alpha(X^F \wedge d) = \sup_{F \in \mathcal{S}_1} (\mathrm{VaR}_\alpha(X^F) \wedge d) = \left(\sup_{F \in \mathcal{S}_1} \mathrm{VaR}_\alpha(X^F) \right) \wedge d$$

$$\sup_{G \in \mathcal{S}_2} \mathrm{VaR}_\alpha(X^G - d)_+ = \sup_{G \in \mathcal{S}_2} (\mathrm{VaR}_\alpha(X^G) - d)_+ = \left(\sup_{G \in \mathcal{S}_2} \mathrm{VaR}_\alpha(X^G) - d \right)_+ .$$

According to Theorem 1 of [Ghaoui et al., 2003],

$$\sup_{F \in \mathcal{S}_1} \mathrm{VaR}_\alpha(X^F) = \mu_1 + \sigma_1 \sqrt{\frac{\alpha}{1-\alpha}}, \quad \text{and} \quad \sup_{G \in \mathcal{S}_2} \mathrm{VaR}_\alpha(X^G) = \mu_2 + \sigma_2 \sqrt{\frac{\alpha}{1-\alpha}}$$

achieved at the two-point distributions

$$\mathbb{P}\left(X = \mu_1 - \sigma_1 \sqrt{\frac{1-\alpha}{\alpha}} \right) = \alpha = 1 - \mathbb{P}\left(X = \mu_1 + \sigma_1 \sqrt{\frac{\alpha}{1-\alpha}} \right) = 1 - \alpha$$

and

$$\mathbb{P}\left(X = \mu_2 - \sigma_2 \sqrt{\frac{1-\alpha}{\alpha}} \right) = \alpha = 1 - \mathbb{P}\left(X = \mu_2 + \sigma_2 \sqrt{\frac{\alpha}{1-\alpha}} \right),$$

respectively. ■

Proof of Theorem 4.2.2.

- We first consider the insurer's optimization problem.
  Denote $f(d) = (d \wedge d_1) + (1+\rho) \int_d^\infty \tilde{S}(x)\,\mathrm{d}x$. f is continuous and differentiable almost everywhere on $[0, \infty)$. Its left derivative is given by

$$f'_-(d) = \mathrm{I}_{[0,d_1]}(d) - (1+\rho)\tilde{S}(d).$$

  (i) If $d_1 \leqslant \tilde{F}^{-1}(\frac{\rho}{1+\rho})$, $f'_-(d) \leqslant 0$ on $[0, \infty)$. Hence $d^* = \infty$ and $\lim_{d \to \infty} f(d) = d_1$.

(ii) If $\tilde{F}^{-1}(\frac{\rho}{1+\rho}) < d_1 \leqslant \tilde{F}^{-1}(\frac{\rho}{1+\rho}) + \Pi(\tilde{F}^{-1}(\frac{\rho}{1+\rho}))$, f decreases on $[0, \tilde{F}^{-1}(\frac{\rho}{1+\rho})), (\tilde{F}^{-1}(\frac{\rho}{1+\rho}) + \Pi(\tilde{F}^{-1}(\frac{\rho}{1+\rho})), \infty)$ and increases on $(\tilde{F}^{-1}(\frac{\rho}{1+\rho}), \tilde{F}^{-1}(\frac{\rho}{1+\rho}) + \Pi(d))$. However, since $f(\tilde{F}^{-1}(\frac{\rho}{1+\rho})) \geqslant d_1 = \lim_{d\to\infty} f(d)$, $d^* = \infty$.

(iii) If $d_1 > \tilde{F}^{-1}(\frac{\rho}{1+\rho}) + \Pi(\tilde{F}^{-1}(\frac{\rho}{1+\rho}))$, the monotonicity of f coincides with that in Case (ii). In this case, however, $f(\tilde{F}^{-1}(\frac{\rho}{1+\rho})) < d_1 = \lim_{d\to\infty} f(d)$. Hence, $d^* = \tilde{F}^{-1}(\frac{\rho}{1+\rho})$.

- Next we consider the reinsurer's optimization problem.
  Denote $g(d) = (d_2 - d)_+ - (1+\rho)\int_d^\infty \tilde{S}(x)\,\mathrm{d}x$. g is continuous and differentiable almost everywhere on $[0,\infty)$. Its left derivative is given by

$$g'_-(d) = -\mathrm{I}_{[0,d_2]}(d) + (1+\rho)\tilde{S}(d).$$

(i) If $d_2 \leqslant \tilde{F}^{-1}(\frac{\rho}{1+\rho})$, $g'_-(d) \geqslant 0$ on $[0,\infty)$. Hence $d^* = 0$ and $g(0) = d_2 - (1+\rho)\tilde{\mu}$.

(ii) If $d_2 > \tilde{F}^{-1}(\frac{\rho}{1+\rho})$ and $d_2 - (1+\rho)\int_0^{d_2} \tilde{S}(x)\,\mathrm{d}x < 0$, g increases on $[0, \tilde{F}^{-1}(\frac{\rho}{1+\rho})), (d_2, \infty)$ and decreases on $(\tilde{F}^{-1}(\frac{\rho}{1+\rho}), d_2)$. Plus, since $g(0) - g(d_2) = d_2 - (1+\rho)\int_0^{d_2} \tilde{S}(x)\,\mathrm{d}x < 0$, $d^* = 0$. Since $d_2 \leqslant \tilde{F}^{-1}(\frac{\rho}{1+\rho})$ can imply $d_2 - (1+\rho)\int_0^{d_2} \tilde{S}(x)\,\mathrm{d}x < 0$, we can combine the first two cases and remove the condition "$d_2 > \tilde{F}^{-1}(\frac{\rho}{1+\rho})$".

(iii) If $d_2 - (1+\rho)\int_0^{d_2} \tilde{S}(x)\,\mathrm{d}x \geqslant 0$, the monotonicity of g coincides with that in Case (ii). However, since $g(0) - g(d_2) = d_2 - (1+\rho)\int_0^{d_2} \tilde{S}(x)\,\mathrm{d}x \geqslant 0$, $d^* = d_2$.

■

Proof of Theorem 4.2.3.

(a) The left derivative of h_δ, $h'_\delta(d) = (1-2\delta)(1+\rho)\tilde{S}(d) + (\delta-1)\mathrm{I}_{[0,d_2]}(d) + \delta\mathrm{I}_{[0,d_1]}(d)$, for $d \geqslant 0$. When $\delta < \frac{1}{2}$,h_δ is increasing at 0, and d_2 is the only possible reflection point where the $h' - \delta$ changes from negative to positive. Hence, d_2 is the only possible local minimum for h_δ on $[0,\infty)$.

(b) When $\delta = \frac{1}{2}$, $h'_\delta = \frac{1}{2}(\mathrm{I}_{[0,d_1]}(d) - \mathrm{I}_{[0,d_2]}(d))$. The results hence follow.

(c) When $\delta > \frac{1}{2}$, $d_1 < d_2$, if $(1+\rho)\tilde{S}(d_1) \geqslant 1$, h_δ is decreasing on $[0,\infty)$. If otherwise $(1+\rho)\tilde{S}(d_1) < 1$, h_δ decreases on $[0, \tilde{F}^{-1}(\frac{\rho}{1+\rho})), (d_1, \infty)$ and increases on $(\tilde{F}^{-1}(\frac{\rho}{1+\rho}), d_1)$. Hence d_1 is the only local minimum for h_δ on $[0,\infty)$. In this case, if $h(\tilde{F}^{-1}(\frac{\rho}{1+\rho})) < \delta d_1 = \lim_{d\to\infty} h(d)$, $d^* = \tilde{F}^{-1}(\frac{\rho}{1+\rho})$; otherwise, $d^* = \infty$.

(d) When $\delta > \frac{1}{2}, d_1 \geqslant d_2$, if $(1+\rho)\tilde{S}(d_2) < 1$, h_δ decreases on $[0, \tilde{F}^{-1}(\frac{\rho}{1+\rho})), (d_1, \infty)$ and increases on $(\tilde{F}^{-1}(\frac{\rho}{1+\rho}), d_1)$. Hence $\tilde{F}^{-1}(\frac{\rho}{1+\rho})$ is the only local minimum, $d^* = \tilde{F}^{-1}(\frac{\rho}{1+\rho})$ if $h_\delta(\tilde{F}^{-1}(\frac{\rho}{1+\rho})) \leqslant \delta d_1 = \lim_{d\to\infty} h_\delta(d)$. The analysis in other cases is similar and hence omitted.

■

Chapter 5

Future works

In the future, we will be committed to improving our models in Chapters 2, 3 and 4 in the following directions:

1. The results we achieve can solve optimal deductible problems with specific model uncertainties. However, there are mounts of other types of reinsurance contracts in the reinsurance industry. For example, studies including [Tan et al., 2009] investigate optimal quota-share reinsurance with VaR and TVaR taken as criteria. Apart from this, [Sung et al., 2011] shows that an optimal reinsurance contract in the framework of Cumulative Prospect Theory can be one with upper and lower limits. Even though the approaches used in this work varies from ours, since ours uses risk measures as metrics, we can still see the significance of contracts with limits in the reinsurance industry.

   The above mentioned types of reinsurance contracts are relatively tractable in an optimization problem since there are at most two parameters to determine. For example, the upper and lower bounds for contracts with limits, the deductible for stop-loss contracts and the quota-share coefficient for quota-share reinsurance. More generally, an admissible reinsurance indemnity function I is assumed to satisfy the non-sabotage conditions, which basically means I is a non-decreasing function that is 1-Liptchiz. A typical way to deal with this kind of questions is to reduce the dimension of the problem. [Xu et al., 2019] can be taken as a good example. In the work, the authors prove the optimal insurance must be in a certain form, leaving only one parameter unknown. The problem then is converted into a one-dimensional optimization problem and becomes much more tractable. In the future, we would like to generalize our results by replacing the admissible set of reinsurance contracts with a more general one.

2. Other than these, we are interested in finding the optimal deductible in the following problem:
$$\min_{d\geqslant 0}\sup_{F\in\mathcal{S}(\mu,\sigma)}\rho^g((X^F-d)_+-(1+\theta)\mathbb{E}[(X^F-d)_+]), \tag{5.1}$$
where the uncertainty set
$$\mathcal{S}(\mu,\sigma)=\{F:[0,\infty)\mapsto[0,1]:\text{ is a distribution function },\\ \int_0^\infty x\,\mathrm{d}F(x)=\mu,\int_0^\infty x^2\,\mathrm{d}F(x)=\mu^2+\sigma^2\},$$
ρ^g is a risk measure and d is deductible. In this framework, the ceded-loss distribution and the pricing distribution are in the identical worst-case. This means the reinsurer decides its worst-case by considering the ceded loss and the premium together, which is much more reasonable and practical. This problem is dual to the problem considered in [Liu and Mao, 2021], which investigates a similar problem from the perspective of the insurer. By analyzing the above problem, we can make the optimal deductible problem with uncertainty more complete and general.

   The above raise problem, however, can not be solved using our approach given in this thesis, even if ρ^g is a distortion risk measure. The reason is that by letting $\rho^h=\rho^g-(1+\rho)\mathbb{E}$, the new "risk metric", ρ^h, has a weight function, γ^h that is not always non-negative. In this case, the results in this thesis can not be applied directly. We will come up with another way to tackle this problem.

3. Just as [Markowitz, 1968] investigate the problem of minimizing portfolio variance, we would like to shift our optimization problem into minimizing respective variances of the insurer and the reinsurer. The insurer's problem is formulated as below:
$$\min_{d\geqslant 0}\sup_{F\in S_1}\operatorname{var}(X^F\wedge d),$$
and the reinsurer's optimization problem:
$$\min_{d\geqslant 0}\sup_{F\in S_2}\operatorname{var}(X^F-d)_+.$$
In this problem, premium is not taken into account, since translation of random variables do not affect their variances. Solving this problem will help the risk bearers

minimize their variances in a more conservative way, hence reducing systematic risks in the financial system.

Bibliography

[Acerbi and Tasche, 2002] Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, 26(7):1487–1503.

[Allais, 1953] 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*, pages 503–546.

[Arrow, 1963] Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53:941–973.

[Arrow, 1996] Arrow, K. (1996). The theory of risk-bearing: small and great risks. *Journal of risk and uncertainty*, 12:103–111.

[Artzner et al., 1997] Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1997). Thinking coherent. *RISK*, 10:68–71.

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

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

[Bagga et al., 2020] Bagga, P., Paoletti, N., Alrayes, B., and Stathis, K. (2020). A deep reinforcement learning approach to concurrent bilateral negotiation. *arXiv preprint*, (2001.11785).

[Bernard et al., 2018] Bernard, C., Denuit, M., and Vanduffel, S. (2018). Measuring portfolio risk under partial dependence information. *Journal of Risk and Insurance*, 85(3):843–863.

[Bernard et al., 2020a] Bernard, C., Kazzi, R., and Vanduffel, S. (2020a). Range value-at-risk bounds for unimodal distributions under partial information. *Insurance: Mathematics and Economics*, 94:9–24.

[Bernard et al., 2020b] Bernard, C., Pesenti, S. M., and Vanduffel, S. (2020b). Robust distortion risk measures. *arXiv preprint arXiv:2205.08850*.

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

[Black and Litterman, 1992] Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43.

[Blanchet et al., 2022] Blanchet, J., Chen, L., and Zhou, X. Y. (2022). Distributionally robust mean-variance portfolio selection with wasserstein distances. *Management Science*, pages 6382–6410.

[Blanchet and Murthy, 2019] Blanchet, J. and Murthy, K. (2019). Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600.

[Borch, 1960] Borch, K. (1960). An attempt to determine the optimum amount of stop loss reinsurance. *Transactions of the 16th International Congress of Actuaries*, I:597–610.

[Borch, 1969] Borch, K. (1969). The optimal reinsurance treaty. *ASTIN Bulletin: The Journal of the IAA*, 5(2):293–297.

[Cai et al., 2020] Cai, J., Li, J., and Mao, T. (2020). Distributionally robust optimization under distorted expectations. *Available at SSRN 3566708*.

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

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

[Cai et al., 2008] Cai, J., Tan, K., Weng, C., and Zhang, Y. (2008). Optimal reinsurance under var and cte risk measures. *Insurance: Mathematics & Economics*, 43(1):185–196.

[Cai and Wang, 2021] Cai, J. and Wang, Y. (2021). Optimal capital allocation principles considering capital shortfall and surplus risks in a hierarchical corporate structure. *Insurance: Mathematics and Economics*, 100:329–349.

[Cai and Weng, 2016] Cai, J. and Weng, C. (2016). Optimal reinsurance with expectile. *Scandinavian Actuarial Journal*, 2016(7):624–645.

[Cont et al., 2010] Cont, R., Deguest, R., and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10(6):593–606.

[Di Guglielmo, 1977] Di Guglielmo, F. (1977). Nonconvex duality in multiobjective optimization. *Mathematics of Operations Research*, 2(3):285–291.

[Durrett, 2019] Durrett, R. (2019). *Probability: Theory and Examples*, volume 49. Cambridge university press.

[Ellsberg, 1961] Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, pages 643–669.

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

[Ghaoui et al., 2003] Ghaoui, L. E., Oks, M., and Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556.

[Ghossoub, 2019] Ghossoub, M. (2019). Optimal insurance under rank-dependent expected utility. *Insurance: Mathematics and Economics*, 87:51–66.

[Glasserman and Xu, 2014] Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58.

[Gneiting, 2011] Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

[Guan et al., 2022] Guan, Y., Jiao, Z., and Wang, R. (2022). A reverse expected shortfall optimization formula. *Available at SSRN 4092939*.

[Hu et al., 2015] Hu, X., Yang, H., and Zhang, L. (2015). Optimal retention for a stop-loss reinsurance with incomplete information. *Insurance: Mathematics and Economics*, 65:15–21.

[Jagannathan, 1977] Jagannathan, R. (1977). Minimax procedure for a class of linear programs under uncertainty. *Operations Research*, 25(1):173–177.

[Jazayeriy et al., 2011] Jazayeriy, H., Azmi-Murad, M., Sulaiman, N., and Udizir, N. I. (2011). Pareto-optimal algorithm in bilateral automated negotiation. *International Journal of Digital Content Technology and its Applications*, 5(3):1–11.

[Jiang, 2022] Jiang, W. (2022). Pareto-optimal insurance under heterogeneous beliefs and incentive compatibility. *Scandinavian Actuarial Journal*, pages 1–19.

[Jiang et al., 2019] Jiang, W., Ren, J., Yang, C., and Hong, H. (2019). On optimal reinsurance treaties in cooperative game under heterogeneous beliefs. *Insurance: Mathematics and Economics*, 85:173–184.

[Kahneman and Tversky, 2013] Kahneman, D. and Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the Fundamentals of Financial Decision Making: Part I*, pages 99–127. World Scientific.

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

[Linsmeier and Pearson, 1996] Linsmeier, T. J. and Pearson, N. D. (1996). *Risk Measurement: An Introduction to Value at Risk*. No. 1629-2016-134959.

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

[Liu and Mao, 2021] Liu, H. and Mao, T. (2021). Distributionally robust reinsurance with value-at-risk and conditional value-at-risk. *Insurance: Mathematics and Economics*, 107:393–417.

[Lo, 1987] Lo, A. W. (1987). Semi-parametric upper bounds for option prices and expected payoffs. *Journal of Financial Economics*, 19(2):373–387.

[Markowitz, 1968] Markowitz, H. M. (1968). *Portfolio Selection*. Yale University Press.

[Natarajan et al., 2010] Natarajan, K., Sim, M., and Uichanco, J. (2010). Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 20(4):695–731.

[Németh, 2003] Németh, A. B. (2003). Characterization of a Hilbert vector lattice by the metric projection onto its positive cone. *Journal of Approximation Theory*, 123(2):295–299.

[Newey and Powell, 1987] Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, pages 819–847.

[Ohlin, 1969] Ohlin, J. (1969). On a class of measures of dispersion with application to optimal reinsurance. *ASTIN Bulletin*, 5:249–266.

[Pesenti, 2022] Pesenti, S. M. (2022). Reverse sensitivity analysis for risk modelling. *Risks*, 10(7):141.

[Quiggin, 1982] Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

[Sion, 1958] Sion, M. (1958). On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176.

[Sung et al., 2011] Sung, K., Yam, S., Yung, S., and Zhou, J. (2011). Behavioral optimal insurance. *Insurance: Mathematics and Economics*, 49(3):418–428.

[Tan et al., 2009] Tan, K. S., Weng, C., and Zhang, Y. (2009). VaR and CTE criteria for optimal quota-share and stop-loss reinsurance. *North American Actuarial Journal*, 13(4):459–482.

[Wang and Wang, 2020] Wang, Q. and Wang, R.and Wei, Y. (2020). Distortion RiskMetrics on general spaces. *ASTIN Bulletin: The Journal of the IAA*, 50(3):827–851.

[Wang and Zitikis, 2021] Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the expected shortfall. *Management Science*, 67(3):1413–1429.

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

[Xu et al., 2019] Xu, Z. Q., Zhou, X. Y., and Zhuang, S. C. (2019). Optimal insurance under rank-dependent utility and incentive compatibility. *Mathematical Finance*, 29(2):659–692.

[Yaari, 1987] Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society*, pages 95–115.